SUPRAMOLECULAR STRUCTURE AND FUNCTION 9

Supramolecular Structure and Function 9

Edited by

Greta Pifat-Mrzljak
Rudjer Bošković Institute,
Zagreb, Croatia

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6465-4 (HB)
ISBN 978-1-4020-6466-1 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Preface

The frontiers in scientific research are difficult to define. Nowadays the new knowledge depends not only on new methods and concepts but also on the interaction with other fields of research. Biophysics offers a rational language for discussion among scientists of different disciplines. This is the general philosophy behind the organization of the Summer Schools organized by Rudjer Bošković Institute, Zagreb, Croatia and the Croatian Biophysical Society from 1981 on each third year.

Internationally recognized and successfully established nine international summer schools have been organized under the title «Supramolecular Structure and Function». The International Summer Schools on Biophysics have a very broad scope devoted to the structure-function relationship of biological macromolecules and to mayor biophysical techniques. The intention is to organize courses which provide advanced training at doctoral or postdoctoral level in biosciences. Sixteen main contributions presented here by prominent lecturers illustrate the principals, concepts and methods of biophysics coupled with molecular biology approaches. Since the great diversity of topics the book should be of interest for scientists of different disciplines as was the audience at the School.

The ninth Summer Schools, as Master Classes of UNESCO was supported by UNESCO and could be considered as a part of the mosaic forming the European Research Area (ERA) and the European Higher Education Area (EHEA). The organizers of the International Summer School on Biophysics think that the publication of this volume and its distribution within the scientific community will serve towards the objectives of expanding, sharing and providing easy access to scientific knowledge.

Other supporters of the School were IUPAB and EBSA reflecting international and European interest to bring together scientists of different profiles from all over the world. It is necessary to mention domestic financial supporters: Ministry of science, education and sport of the Republic of Croatia, Croatian Academy for Science and Art and The National Foundation for Science, Higher Education and Technological Development of the

Republic of Croatia which substantial supports enabled the participation of young domestic scientists.

The outcome of the ninth School is not only exchange of knowledge between lecturers and participants at the beautiful atmosphere of Red Island, near Rovinj, Croatia, not only this volume which will inform broader scientific community on the profile of Summer School and new biophysical achievements but, what we are very proud of, the Schools due to their quality, are included into the curricula of doctoral studies of some European universities.

Greta Pifat-Mrzljak

Contents

1. Predictions of Protein-Protein Interactions at the Atomic Scale 1
Shoshana J. Wodak, Raúl Mendez, and Marc Lensink

2. An Infrared Study of Fibril Formation in Insulin from Different Sources 21
Igor de la Arada and Jose Luis R. Arrondo

3. Some Basic Biomolecular NMR for Protein Structure Determination 33
Angela M. Gronenborn

4. Solid State NMR for Studying Membrane Proteins 45
Anthony Watts

5. Protein Complexes in SUMO Signaling 75
Christina-Maria Hecker and Ivan Dikic

6. Mass Spectrometry and Glycomics 89
Jasna Peter-Katalinić

7. Mechanisms of Supramolecular Assembly Exemplified by Microtubules and Amyloid Fibril Formation 103
Peter Bayley

8. Hepatitis B Virus (HBV): The Life-cycle and Assembly of a Complex Virus 131
P. Jonathan G. Butler

9. Photo-oxidative Stress in the Presence of a Water-soluble Derivative of C_{60}: ESR and AFM Assays 153
Bertrand Vileno, Andrzej Sienkiewicz, Małgorzata Lekka, Pierre R. Marcoux, and László Forró

10. Functional EPR Spectroscopy and Imaging of Nitroxides 181
Valery V. Khramtsov

11. Novel Leads for Selective Antibiotics Against Shigellosis by Virtual Screening, Crystallography and Synthesis 209
Bernhard Stengl and Gerhard Klebe

12. Virtual Ligand Screening: A Method to Discover New Drug Leads 251
Gerhard Klebe

13. From Solid State to Bio-Complexity: On the Emerging Science of Emergence 273
Davor Pavuna

14. MR Spectroscopy and the Early Detection of Cancer in Human Subjects 287
Ian C.P. Smith and Racquel Baert

15. Microarray and Single Cell Analysis Techniques in Bio-medical Fields 299
Wilhelm J. Ansorge

16. SOS Hypothesis and the Emergence of Integrative Biology 307
Miroslav Radman

Subject Index 315

Predictions of Protein-Protein Interactions at the Atomic Scale

SHOSHANA J. WODAK[1,2], RAÚL MENDEZ[3,4], AND MARC LENSINK[3]
[1]*Structural Biology Program, Hospital for Sick Children, 555 University Av. Toronto, Ontario M5G 1X8, Canada, Phone: +1 416 813 6351, Fax: +1 416 813 8755, E-mail: shoshana@sickkids.ca*
[2]*Departments of Biochemistry and medical Genetics and Microbiology, University of Toronto, 1 King's College Circle, Toronto, Ontario Canada, M5S 1A8*
[3]*Service de Conformation de Macromolécules Biologiques, et Bioinformatique, Centre de Biologie Structurale et Bioinformatique, CP 263, BC6, Université Libre de Bruxelles, Blvd du Triomphe, 1050 Bruxelles, Belgium*
[4]*Grup Biomatemàtic de Recerca. Institut de Neurociències. Unitat de Bioestadística. Facultat de Medicina. Universitat Autonòma de Barcelona. 08193 Bellaterra, Spain*

Abstract

The use of computational procedures to predict protein-protein interactions from sequence and structure information has attracted renewed interest in recent years with the realization that such procedures can fruitfully complement and guide experimental characterizations of these interactions. CAPRI (Critical Assessment of Predicted Interactions) the community-wide experiment for assessing blind predictions of protein complexes is playing an important role in fostering progress in protein docking algorithms – the computational procedures for predicting the 3D structure of a protein complex from those of the individual components. In this paper we provide our view of the major challenges that docking algorithms must meet. We then describe how current docking methods address these challenges by presenting a summary of results obtained in rounds 1-5 of the CAPRI experiment, and discuss the lessons learned from these results.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 1–20.

Keywords: protein interactions, protein complexes, docking, structure, predictions, homology modeling, side chain optimization, energy, calculations, conformational sampling

1. INTRODUCTION

Vital functions of the living cell are carried out by protein complexes, which can comprise anywhere between two to several tens different protein components. Characterizing these complexes and gaining understanding of their mechanism of action has therefore become a central theme in biology (Krogan et al. 2006; Gavin et al. 2006). It has been argued that in eukaryotic cells the full set of interacting proteins – the interactome – may be at least an order of magnitude larger than the set of protein coding genes (Claverie 2002; Uetz et al. 2000; Edwards et al. 2002).

Protein complexes are however still rather poorly represented in the Protein Data Bank (PDB) (Berman et al. 2000). Structural genomics programs have so far focused on large-scale structure determination of individual gene products, but not on larger complexes (Sali et al. 2003). Although this is already changing, with new efforts combining crystallography, NMR, and cryo electron microscopy being set up for analyzing multi-component assemblies, sampling of the interactome is likely to remain sparse for the near future, whereas the repertoire of individual protein structures will be increasingly well sampled.

Computational procedures capable of reliably generating structural models of multi-protein assemblies starting from the atomic coordinates of the individual components, the so-called 'docking' methods, should therefore play an important role in helping bridge the gap. But as with other predictive approaches, objective tests are needed to monitor their performance. The Critical Assessment of Predicted Interactions (CAPRI), a community wide experiment analogous to CASP (Moult 2005) was set up in 2001, with exactly this goal (Janin et al. 2003; Wodak and Mendez 2004; Mendez et al. 2003). In CAPRI, individual groups that develop docking procedures, predict the three dimensional structure of a protein complex from the known structures of the components. The predicted structure is subsequently assessed by comparing it to the experimental structure – the target – determined most commonly by X-ray diffraction, which is deposited with CAPRI prior to publication. The predictions are thus made blindly – without prior knowledge of the correct answer, and an independent team who has no knowledge of the identity of the predictors carries out the evaluation.

In the following we present a short overview of the basic challenges that docking procedures need to address. We then briefly review results obtained in the CAPRI docking predictions and summarize what these results have

taught us on the performance of current docking procedures and the directions of future developments.

2. THE CHALLENGES THAT DOCKING METHODS MUST MEET

Docking refers to the calculations whereby the three dimensional structure of a protein complex is computed starting from the atomic coordinates of its components (Fig. 2.1). The complex may be composed of several copies (subunits) of the same protein (homo-oligomer) or of one or more copies of different subunits (hetero-oligomer), and complexes amenable to docking calculations are those where the individual components also exist in free form.

Successful docking procedures must perform two main tasks: a) they must be able to single out specific association modes, those that are biologically relevant, from a large number of non-specific modes, and b) they need to efficiently sample a very large number of different association modes between the subunits, as well as to take into model changes in protein conformation that may occur upon association.

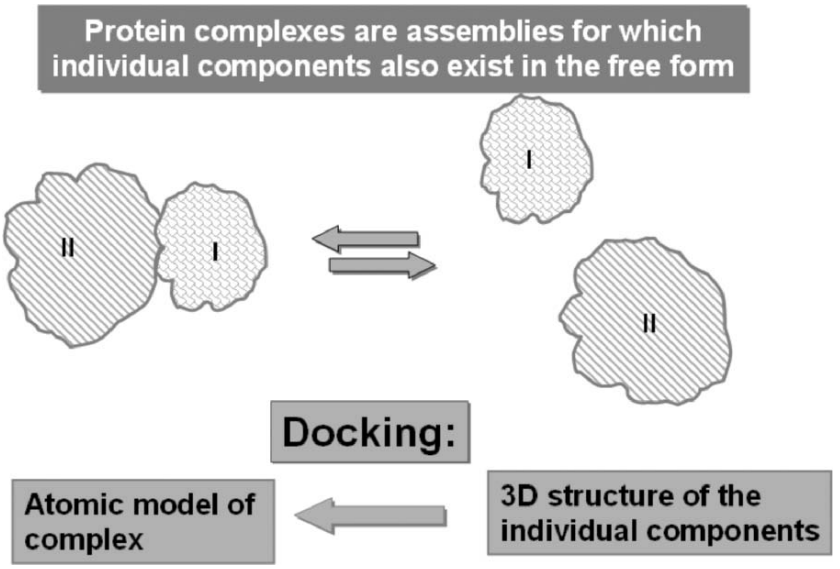


Figure 2.1. Defining docking. Docking is the operation whereby the detailed atomic model of a protein complex is computed from the atomic coordinates of its components. Docking can be applied to assemblies where the individual components also exist in the free form.

2.1. Specific versus non-specific associations

Differentiating between specific and non-specific association modes between two proteins is particularly challenging (for review see Wodak and Janin (2002) and references therein). This has been demonstrated by a number of studies that compared the characteristics of interfaces in known 3D structures of protein complexes from the Protein Data Bank (PDB) (Berman et al. 2000), taken to represent specific associations, with those of interfaces between neighbouring subunits in protein crystals, considered as representing non-specific associations. Some results of such studies (Janin and Rodier 1995; Janin 1997; Bahadur et al. 2004) are summarized in Fig. 2.2. This Figure plots the histogram of the area buried in interfaces belonging to known protein complexes on the one hand (blue bars) and the distribution of the buried areas in interfaces formed as a result of crystal contacts (green and red curves).

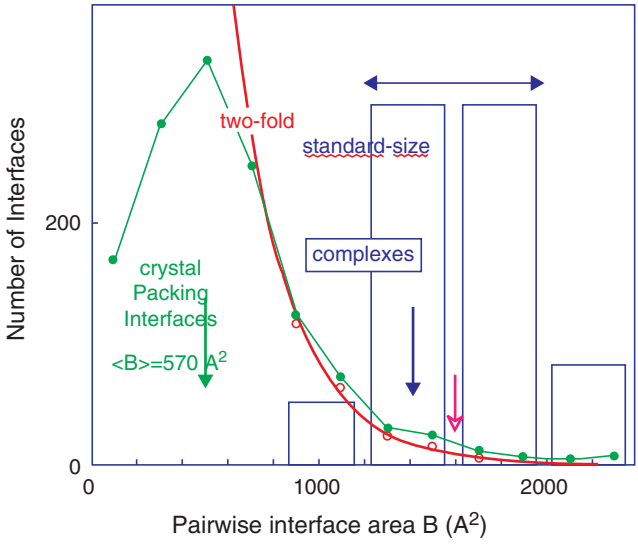


Figure 2.2. Distributions of surface areas buried in interfaces in complexes formed by homo- and hetero- dimers (blue bars), and in contacts formed between proteins in crystal structures between all neighbours (green curve) and between neighbours related by crystallographic twofold symmetry (red curve). The standard size of specific interfaces is ~1600 Å^2 whereas that of crystal contacts is much smaller (~570 Å^2). This figure illustrates results reported in references (Janin and Rodier 1995; Janin 1995; Lo Conte et al. 1999; Janin 1997; Bahadur et al. 2004) and is a gift by J. Janin.

The size of the buried area in an interface is related to the affinity of the association, more particularly it is related to the hydrophobic component of this affinity (Janin 1995). The larger the buried area the more favourable the contribution of the hydrophobic component will be towards the association.

Figure 2.2 shows that interfaces in known protein complexes tend to be larger (~1600Å^2 on average) than non-specific crystal interfaces (~570Å^2 on average) (Janin and Rodier 1995). However, the interface size taken alone cannot be used to reliably discriminate specific from non-specific associations because crystal contacts, especially those formed between subunits related by symmetry operation (for instance crystal dimers, where subunits are related by two fold symmetry) can form interfaces whose size may be as large as those occurring in specific complexes, as indicated by the long tails in the red and green curves of Figure 2.2.

Other parameters such as the number of hydrogen bonds formed between residues at the interface, and the number or fraction of contacts made between different types of atoms or residues, have also been surveyed in both specific and non-specific associations (see Lo Conte et al. 1999; Janin 1997; Bahadur et al. 2004). The observed ranges in these parameters tend however to vary widely making it a challenging task to rely on them for accurate classification.

On the other hand, interesting observations have been made on the manners in which interface atoms are distributed and packed at interfaces (Bahadur et al. 2004). In both homodimers and heterodimers, on average 34-36% of the interface atoms are fully buried. However, in the larger crystal packing interfaces (those with an interface area >800Å^2), the fraction of buried atoms is on average much lower (21%). Furthermore, in protein complexes, interface atoms tend to form large contiguous patches, whereas in crystal contacts interface atoms are organized in a large number of small patches, and are hence more dispersed across the interface as illustrated in Figure 2.3.

To single out specific interfaces from a very large number of non-specific ones, like the crystal contacts, docking procedures must use scoring schemes, or force-fields that are capable of capturing these differences, which might be more or less subtle depending on the system under study. An additional challenge for docking procedures is the role played by specific interactions with solvent molecules. It has been shown indeed, that specific interfaces may vary significantly in the number of water molecules that engage in hydrogen bonding interactions with interface residues (Rodier et al. 2005).

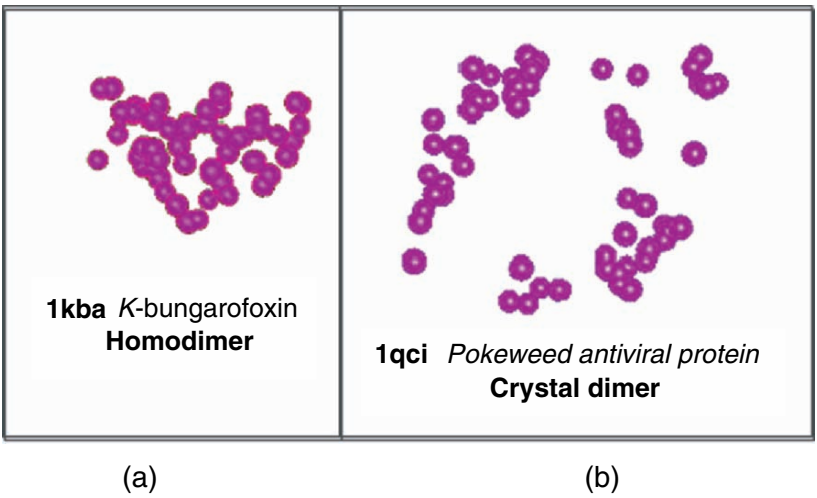


Figure 2.3. Differences in the packing of atoms in interfaces of specific and non-specific associations (Bahadur et al. 2004). A specific interface, that in the *K-bungarotoxin,* homodimer (PDB code 1kba) (a), and a non-specific interface in the *pokeweed antiviral protein,* crystal dimer (1qci) (b). Both interfaces include about 50 interface atoms (pink balls) in total, but these atoms form a large continuous patch in the specific interface whereas in the non-specific crystal contacts they are much more dispersed across the interface. Figures are a gift by J. Janin.

2.2. Sampling of association modes and conformational space

Docking being the search of a stable association mode between two or more proteins is by definitions an optimization problem, whose solution crucially depends on sampling the parameter space. Here the parameters are to a first approximation, the relative orientations and positions of the individual subunits that bring them into contact, as illustrated in Figure 2.4. Since the association between two proteins can in principle involve any portion of the surface of each protein, docking procedures must be able to efficiently explore a very large number of different associations modes between the interaction partners, which will increase as the size of the proteins, (and hence their surface areas) increases.

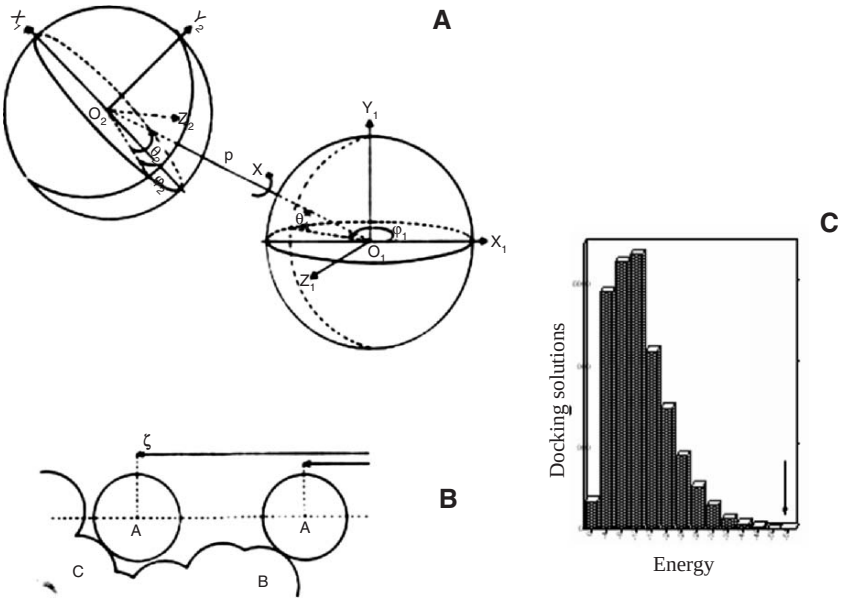


Figure 2.4. Illustration of the search procedure in the space of the rigid body parameters defining the relative orientation and position of the two protein subunits being docked. (a) The subunits are schematized as two spheres and their relative position and orientation is described using the polar coordinate system developed by Wodak and Janin (Wodak and Janin 1978). (b) Schematic representation of the inter subunits distance optimization procedure, carried out at each value of the 5 parameters (see (a)) that define the relative orientation of the subunits. (c) Distribution of interface areas encountered by sampling docking solutions between Trypsin and the Bovine Pancreatic Trypsin Inhibitor (BPTI), following Wodak and Janin (1978).

The task is further complicated by the fact that proteins tend to display appreciable conformational plasticity. Their association is therefore often accompanied by conformational changes. These can range from small adjustments of the side chain conformations at the newly formed interface or elsewhere in the protein, to large conformational transitions that involve whole segments or domains of the protein (Fig. 2.5). As will be seen below, present day docking procedures are quite efficient in sampling different association modes, but are still very much challenged by modelling conformational changes.

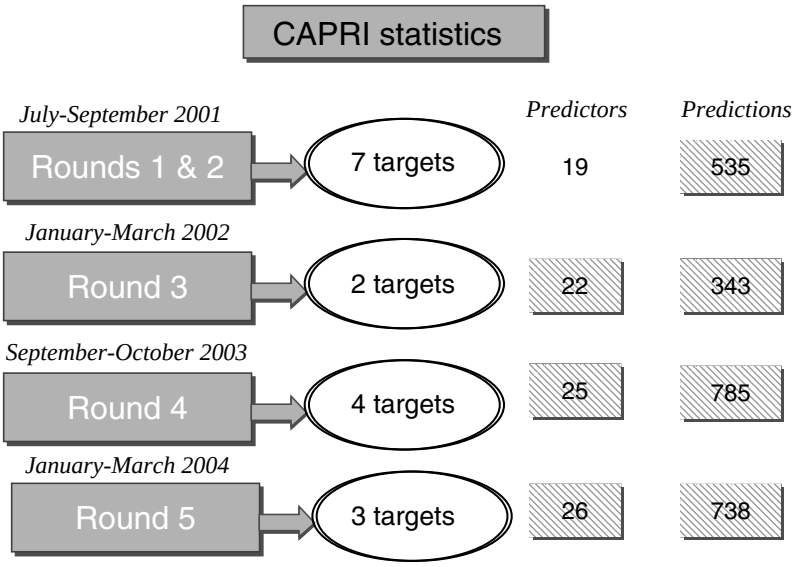


Figure 2.5. Statistics of the CAPRI rounds discussed in this chapter. For a detailed description of the targets, all the predicted complexes as well as the identity of the groups participating in the predictions, consult the CAPTI web site http://capri.ebi.ac.uk/.

3. THE CRITICAL ASSESSMENT OF PREDICTED INTERACTIONS (CAPRI): LESSONS LEARNED

In the following we briefly highlight results obtained in the first 5 rounds of CAPRI experiment and what they have taught us so far about the status of docking methods.

As summarized in Figure 2.5, these rounds took place between 2001-2004. Additional prediction rounds occurred since then but their results have not yet been publicly released at the time this manuscript was prepared. The 5 rounds discussed here included a total of 16 analyzed targets, for which 2400 predictions were submitted by 23 predictor groups, on average.

As already reported (Mendez et al. 2003; Mendez et al. 2005; Vajda 2005), the CAPRI targets comprised protein complexes with a wide range of different functional roles, including kinases and phosphatases and their interacting partners, a fair number of antibodies bound to various antigens, complexes involving metabolic enzymes and many more. While these targets covered the normal range in terms of the size and types (polar non-polar) of

interfaces observed in protein complexes, they represented different levels of difficulty for the prediction methods.

The easiest cases were those where the components proteins were small, and underwent little conformational adjustments upon complex formation. In such cases the number of candidate solutions to be sampled remained manageable and the conformational changes, mostly limited to side chain adjustments, could be readily modelled (see below). As the size of the protein components increases, and when association is accompanied by large conformational changes, docking methods will in general have a harder time coping. But these more difficult cases are those that drive progress in docking methods, as will be commented below.

3.1. Evaluating predicted complexes

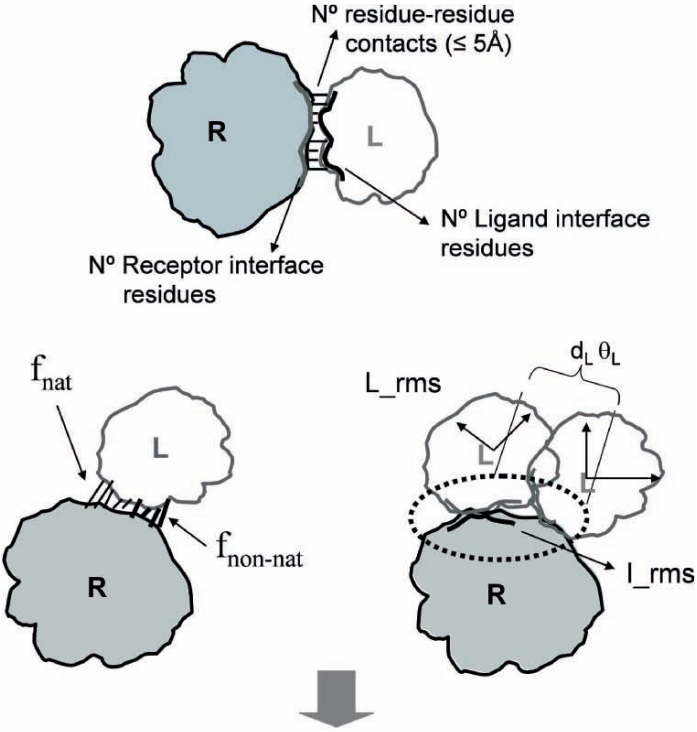
To evaluate the predicted complexes, these are systematically compared to the corresponding target structures using a set of quality measures devised by the independent assessors in close consultation with the CAPRI management groups and predictor community. The various computed parameters and quality criteria are illustrated in Figure 3.6, with further details available in Mendez et al. (2003).

The most important quantities are f_{nat}, L_rms, and I_rms. f_{nat} is defined as the number of native (correct) residue-residue contacts in the predicted complex divided by the number of contacts in the target complex. A pair of residues on different sides of the interface was considered to be in contact if any of their atoms were within 5Å. L_rms, assesses the global geometric fit between the 3D structures of the predicted and observed complexes. It is computed as the root mean square (rms) displacement of the ligand (the smaller of the 2 proteins) in the predicted versus target complexes after the receptors (the larger of the 2 proteins) were optimally superimposed (McLachlan 1982). Superimpositions were computed on backbone atoms (N,Cα,C,O).

I_rms, quantifies the fit in the interface region and is computed as the root mean square deviation, after optimal superimposition of the backbone atoms of interface residues only, in the predicted versus target complexes.

For this calculation the interface residues in the target are redefined as those having at least one atom within 10Å of an atom on the other molecule.

The quality of the predictions was assigned to one of 4 categories, high accuracy, medium accuracy, acceptable accuracy, and incorrect, according to the values of three parameters, f_{nat}, L_rms and I_rms, as summarized in Figure 3.6 and previously described (Mendez et al. 2003).



Ranking	Conditions based on Capri computed parameters
High	$f_{nat} \geq 0.5$ AND (L_rms ≤ 1.0 OR I_rms ≤ 1.0)
Medium	($f_{nat} \geq 0.3$ AND $f_{nat} < 0.5$) AND (L_rms ≤ 5.0 OR I_rms ≤ 2.0) OR $f_{nat} \geq 0.5$ AND L_rms > 1.0 AND I_rms > 1.0
Acceptable	($f_{nat} \geq 0.1$ AND $f_{nat} < 0.3$) AND (L_rms ≤ 10.0 OR I_rms ≤ 4.0) OR $f_{nat} \geq 0.3$ AND L_rms > 5.0 AND I_rms > 2.0
Incorrect	$f_{nat} < 0.1$ OR (L_rms > 10.0 AND I_rms ≤ 4.0)

Figure 3.6. Schematic illustration of the quality measures used to evaluate the predicted models. For each target, the CAPRI assessors computed the number of residue-residue contacts between the Receptor (R) and the Ligand (L), and for each of the components, the number of interface residues. The interface residues were defined as those that loose solvent accessible surface area upon association. For each predicted model, the following quantities were computed: the fractions f_{nat} of native and $f_{non\text{-}nat}$ of non native contacts in the predicted interface, the RMS displacement of the backbone atoms of the ligand (L_rms), the mis-orientation angle Θ_L and the residual displacement d_L of the ligand center of mass after the receptor in the model and experimental structures were optimally superimposed (McLachlan 1982), and lastly the I_rms, defined as the RMS displacement of the backbone atoms of the interface residues after they have been optimally superimposed. The threshold values of the parameters that were used to define the quality categories of the predicted complexes (incorrect, acceptable, medium and high accuracy) are listed in the bottom Table.

3.2. Examples of CAPRI prediction results

Figures 3.7 and 3.8 illustrate some informative cases from the CAPRI prediction results.

The very first CAPRI target was the complex between the protein kinase from *Lactobacillus casei* HprK, and its protein substrate Hpr (from *B. subtilis*) (Fieulaine et al. 2002) (Fig. 3.7a). This target represented a difficult docking problem. The free HprK protein is a hexamer and although the complex is known to involve one Hpr molecule per kinase subunit, the predictors had to make a choice between using the entire hexamer, a trimer only, or just the monomer, with the latter being a perilous choice since it

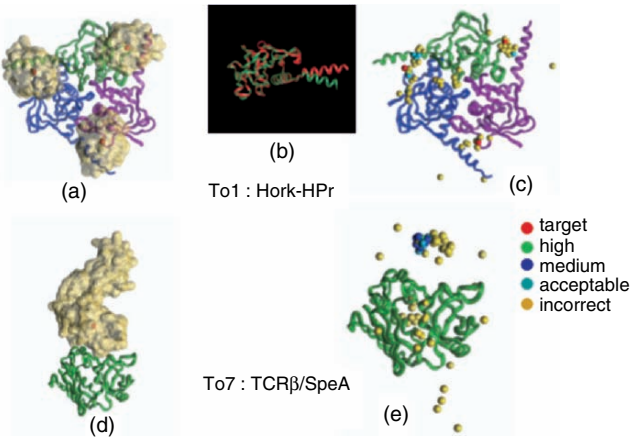


Figure 3.7. Prediction results for selected CAPRI targets: (a) Target 01 the complex between Hpr-K and Hpr (1kkl) (Fieulaine et al. 2002). Only one trimer of the kinase is shown (ribbon drawing) with three bound Hpr molecule (molecular envelope). The orange circles represent the geometric centres of the Hpr molecules; (b) Superimposed backbones of the bound and unbound kinase subunits, illustrating the differences in conformation between the two structures; (c) A pictorial overview of all the predictions provided for this target in Round 1 of CAPRI. Shown are the positions of the geometric centres in all the predicted complexes after the receptor backbones in these complexes have been superimposed onto their counterpart in the target. They are coloured according to the quality of the prediction as indicated in the Figure. For the acceptable predictions, the geometric centres of the ligand cluster more clearly around the observed positions (orange sphere), whereas in incorrect predictions ($f_{nat} \leq 0.1$) they sample a much wider range of positions, corresponding to random solutions. Some of the depicted solutions lie near the 3-fold axis of the kinase trimer or at the interface between the top and bottom trimer layers; (d) Target 07, the complex of T-cell receptor b-chain (ribbon drawing) and the superantigen Streptoccoccal pyrogenic exotoxin A (TCRb-SpeA) (Sundberg et al. 2002) (molecular envelope) (1lox); (e) Overview of the predictions for target 07 analyzed in CAPRI Round 1. All other details are as in (c).

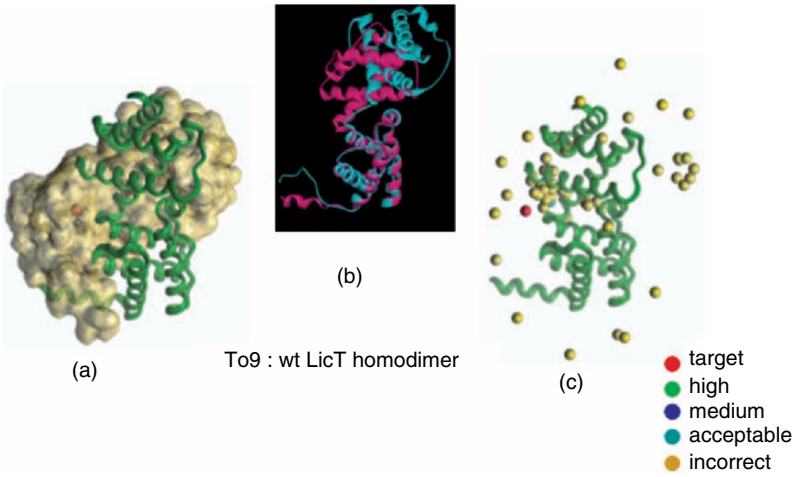


Figure 3.8. Prediction results for target 09, the LicT homodimer: (a) target 09 the *wt* LicT homodimer. One of the subunits is represented by a ribbon drawing whereas the other is depicted by its molecular envelope; (b) Superimposed backbone of the bound (wt) (1TLV) and the 'unbound' (LicT subunit of a mutant dimer) (1H99) used in the CAPRI experiments to calculate the wt homodimer, illustrating the large conformational differences between the two structures (backbone rmsd 12Å); (c) Pictorial overview of the CAPRI predictions for this target. All other details are as in Figure 3.7c.

excluded the possibility for the Hpr molecule to simultaneously interact with 2 kinase subunits, as it is in fact the case. A further difficulty arose from the fact that kinase conformations in the free and bound forms differed (Fig. 3.7b). The main chain of individual bound and unbound monomers superimpose with a RMSD of 1.96Å, with a significant movement the C-terminal helix, which makes contact with the Hpr molecule in the complex. The monomers move moreover relative to one another. In addition, the structures of the unbound HprK molecules available in the PDB had missing coordinates for the loop comprising residues 241-252, which is not uncommon. Predictors therefore needed to model the structure for this loop, to avoid the risk of missing solutions that involve interactions with it. On the other hand, there were some helpful clues as well. It was known that the phosphorylation site of the Hpr molecules was Ser-46, and not His-15, and hence that the kinase P-loop needed to be close to this residue in the cognate complex. Ample information on amino acid sequences of related kinases and Hpr molecules was also available.

Sixteen groups submitted a total of 63 analyzed predictions (predictions devoid of severe atomic clashes). Of these, only 8 were of acceptable quality (with $f_{nat} \geq 0.1$, L_RMS $\leq$ 10Å and I_RMS $\leq$ 4Å), and none were of medium or high accuracy (see pictorial representation in Fig. 3.7c). The best prediction had $f_{nat} = 0.2$, L_RMS = 7.5Å, and I_RMS = 2.45Å, and was thus of relatively poor quality. However, considering the large size of the receptor and the appreciable difference between the bound and unbound kinase conformations, the fact that some solutions approached, albeit imperfectly, the correct mode of association was deemed an encouraging result. It is however important to note that this limited success was primarily the fruit of using biochemical data (on the residues expected to be in contact) or clues from patches of conserved residues in multiple alignments of related kinases and Hpr proteins to guide or filter the docking predictions (see discussion below).

The other example of Figure 3.7 is target 07, the complex between the T-cell receptor β-chain and the superantigen Streptoccoccal pyrogenic exotoxin A (Sundberg et al. 2002) (Fig. 3.7d,e). Fourteen groups submitted 70 predictions for this target. Five were of high accuracy, 7 of medium accuracy, and 8 were qualified as acceptable. Interestingly, for this target more than any other, several of the groups submitted more than one correct prediction. The high accuracy predictions identified between 60-80% of the native contacts. Their L_RMS's were between 1.3-2.6Å and I_RMS's were all 1Å or less. Analysis of the prediction procedures used indicated that the groups that produced at least one high or medium accuracy prediction for target 07 used information on the known 3D structure of an homologuous complex from the PDB to guide, or filter, their solutions. The groups that did not use this information produced lower quality or incorrect predictions.

The third example, illustrated in Figure 3.8 is again a particularly challenging target, where docking procedures had to deal with both, conformational changes and symmetry operations. This was the wild type LicT homodimer (target 09) (Fig. 3.8a) (Graille et al. 2005). It involved the prediction of a homodimer using the structure of a subunit of a double mutant (H207D/H269D), which forms a different dimer, and furthermore differs substantially in conformation (backbonerms of 12Å) from the bound from (Fig. 3.8b). It is therefore not too surprising that the prediction results for this target were very poor, with only one acceptable model obtained by one group out of a total of 164 models submitted by 17 groups. Although the computed model is far from accurate ($f_{nat} = 0.2$, $f_{non\text{-}nat} = 0.9$, L_rms $\geq$ 9Å etc.) this is a quite remarkable result (Fig. 3.8c).

3.2. The docking methods used in CAPRI

A wide variety of docking methods have been used by CAPRI predictors, as summarized below. Further details about these methods can be found in Mendez et al. 2005 and in the publications by individual CAPRI predictors appearing in the 2 special issues of Proteins dedicated to CAPRI (Mendez et al. 2003; Mendez et al. 2005).

3.2.1. Force fields for scoring predicted complexes

CAPRI predictors have and are using a wide range of force fields or functions to score predicted complexes (for an overview, see Table III of Mendez et al. (2005)). A good number of methods rely on shape complementarity measures (see Chen and Weng 2003), one of the major measures used in early docking algorithms. Currently however, these measures are in general supplemented with other terms representing Van der Waals and electrostatic interactions, as well as solvent effects. The use of scoring functions that combine several of these terms is presently the rule. Some scoring functions also include terms representing the so-called knowledge-based potentials. These potentials include terms representing rotamer probabilities (probabilities of finding specific side chain conformations in known structures), scores derived from the frequencies of contacts between residue or atoms pairs in known structures, as well as experimental restraints (e.g. derived from NMR data). Several novel scoring functions with parameters computed from or fitted to known complexes, were also tested.

3.2.2. Sampling procedures and treatment of symmetry

Sampling the relative orientations of the ligand and receptor molecules is a key step of any docking procedure. It is usually the first step during which a very large number of candidate solutions must be efficiently and reliably evaluated so that very unlikely solutions can be eliminated, leaving a tractable number of candidates for subsequent analysis. For computational efficiency, the sampling step is therefore usually performed using a very simplified model of the protein structures and a correspondingly simple scoring function.

A good fraction of the methods used by CAPRI predictors, use FFT based algorithms, which together with the geometric hashing technique, remain the fastest and therefore most efficient procedures for sampling relative receptor/ligand orientations. Several of the alternative procedures, such as Monte Carlo or systematic searches using polar coordinates, seem

to be less efficient. A few novel methods such as Conformational Space Annealing (CSA) and the Molecular Interaction Field (MIF) have also been proposed (for an overview and literature references, see Table III (b) of Mendez et al. 2005).

In order to choose the subset of solutions that will undergo subsequent analysis, a majority of the groups cluster the solutions into families of similar structures and select representatives from the most populated clusters for further analysis. Some predictors also analyze the energy (score) spectrum of the solutions obtained in the first coarse sampling step, and look for one or a few solutions having significantly lower energy than the bulk of the remaining ones. Subsequent side chain building and global refinement steps concentrate on these solutions. When no such solutions are found, the coarse search step is repeated from different starting positions.

Lastly, a fair number of procedures also enable the docking of identical subunits into symmetrical assemblies. This is achieved by introducing specific dependencies between the rigid body degrees of freedom, whose effect is to reduce the number of independent parameters from 6 to 4.

3.2.3. Treatment of conformational flexibility

3.2.3.1. Side chain flexibility

Quite some progress has been achieved in recent years in the ability of protein structure prediction methods, both of the ab-initio and homology modelling category, to model side chain conformations. When the backbone conformation is that of the native protein or is close to it (≤1Å rms), side chain conformations of buried residues can be predicted quite accurately: to ~1Å rms of those in the crystal structure (Canutescu et al. 2003). Similar accuracy is now being achieved by many groups for modelled side chains in interfaces of docked complexes, provided the backbone conformations of the components have either changed little or have been accurately modelled.

Most docking procedures include a step for optimizing side chain conformations, but individual optimization procedures vary substantially (Gray et al. 2003; Rohl et al. 2004; Rohl et al. 2004; Jaramillo and Wodak 2005; Ogata et al. 2003; Fitzjohn and Bates 2003; Fernandez-Recio et al. 2003; Fernandez-Recio et al. 2004). Most commonly however, side chain flexibility is treated by performing a short step of Molecular Dynamics (MD) or energy minimization, both of which involve only a limited exploration of conformational space. Using these methods side chain conformations are in general modelled to within about 1-1.5Å rms, in all high accuracy solutions, those with I-rms (backbone rms) of 0.5Å or less.

3.2.3.2. Backbone flexibility

Adequate treatment of backbone flexibility remains a major challenge for all protein modelling tasks, including docking. In comparison with the earlier methods used in CAPRI, many more groups are currently addressing this challenge. This is often achieved by introducing a global refinement step as part of the procedure of side chain conformation sampling, which produces incremental backbone adjustments every time a set of new side chain conformations is introduced. Other methods use global refinement as a final step before ranking the docking solutions, enabling only small backbone adjustments.

Some methods perform docking on multiple conformations of the components, which are derived from a Principle Component Analysis coupled to MD simulations, from snapshots of MD simulations or from NMR models (see Mendez et al. 2003; Mendez et al. 2005). These latter methods can produce somewhat larger structural deformations than afforded by energy minimization, but these deformations are generally limited to those with low energy barriers that are accessible to individual components of the complex during very short time scales.

A few approaches are designed to handle conformational changes of any size, but preferably those involving relative movements of whole domains. When a subunit is known or suspected to undergo conformational change it is subdivided into domains or fragments, the fragments are docked independently and the docked fragments are assembled, in a similar fashion to some of the in-silico small molecule docking algorithms. (Wodak and Janin 1978). The down side of this strategy is that choices need to be made on how to best subdivide the protein structure, usually in absence of any evidence.

3.2.4. Exploiting prior knowledge

A very important ingredient for achieving successful docking in CAPRI has bee the use of prior knowledge on the protein regions that are likely to interact. Such knowledge includes indications on interacting residues from published biochemical or mutagenesis studies. With the increasing availability of information on protein sequences and 3D structures, however, predictors can themselves derive information on protein regions likely to interact by mapping patterns of sequence conservation across related proteins onto the 3D structure of the individual protein components. This information has been used with a remarkable success rate, either as ‘experimental restraints’ to guide the search or in order to filter out solutions, and has become an

integral component of docking procedures. However, in a few CAPRI targets, where the structure of the complex had some unexpected features, the use of prior knowledge to guide or filter solutions has hampered accurate predictions, and can therefore not be universally recommended.

4. CONCLUDING REMARKS

The successive analyses of CAPRI predictions have clearly demonstrated that progress is being achieved in protein-protein docking procedures. Although many of the CAPRI rounds included challenging targets where components undergo large conformational changes, or homology modelling of one of the components had to be performed, adequate prediction were made for all targets by at least one groups, and often by many. Furthermore, for a good number of targets, high accuracy models were produced (for a detailed account of all CAPRI results and rankings see Wodak and Mendez (2004) and Mendez et al. (2005)).

These encouraging results can be attributed to incremental improvements of scoring functions and to the manner in which additional information (biochemical or sequence conservation), is exploited in order to filter out the many false positive solutions invariably generated by all docking procedures.

It is also rewarding to see that many predictors try out new inventive approaches for tackling backbone and side chain flexibility, including the mixing and matching different scoring functions and sampling procedures. A particularly promising development has been the successful use of homology built models in docking. Several of the latter CAPRI targets for which results were published, involved modelling the structure of one of the components using as template the known structure of a related protein. This was successfully performed by a number of groups, leading to adequate predictions. Groups familiar with homology modelling methods had a clear advantage over classical dockers, but this advantage will fade away, as powerful web services for homology modelling become available.

CAPRI is clearly playing an important role in fostering these developments, of which will undoubtedly see more in the near future. The improvements in the performance of docking procedures that they are bringing and promise to bring are a strong incentive to continue with the CAPRI experiment. To maintain the momentum created by CAPRI a steady supply of new targets is needed, and x-ray crystallographers and NMR experts will therefore continued to be called upon to entrust CAPRI management with their structures prior to publication.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of the CAPRI management group, all CAPRI predictors, as well as all the crystallographers who contributed structures of complexes prior to publications. SJW acknowledges support from the Canada Institute for Health Research and the Hospital for Sick Children, Toronto. SJW and ML acknowledge support from the European Commission Project LSHG-CT-2004-503567.

REFERENCES

Bahadur, R.P., Chakrabarti, P., Rodier, F., Janin, J. (2004) A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.* **336**(4): 943-55.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**(1): 235-42.

Canutescu, A.A., Shelenkov, A.A., Dunbrack, R.L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**(9): 2001-2014.

Chen, R., Weng, Z. (2003) A novel shape complementarity scoring function for protein-protein docking. *Proteins* **51**(3): 397-408.

Claverie, J.M. (2001) Gene number. What if there are only 30,000 human genes? *Science* **291**(5507): 1255-1257.

Edwards, A.M., Kus, B., Jansen, R., Greenbaum, D., Greenblatt, J., Gerstein, M. (2002) Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* **18**(10): 529-536.

Fernandez-Recio, J., Totrov, M., Abagyan, R. (2003) ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **52**(1): 113-117.

Fernandez-Recio, J., Totrov, M., Abagyan, R. (2004) Identification of protein-protein interaction sites from docking energy landscapes. *J. Mol. Biol.* **335**(3): 843-865.

Fieulaine, S., Morera, S., Poncet, S., Mijakovic, I., Galinier, A., Janin, J., Deutscher, J., Nessler, S. (2002) X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr. *Proc. Natl. Acad. Sci. USA* **99**: 13437-41.

Fitzjohn, P.W., Bates, P.A. (2003) Guided docking: first step to locate potential binding sites. *Proteins* **52**(1): 28-32.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**(7084): 631-6.

Graille, M., Zhou, C.-Z., Receveur-Brechot, V., Collinet, B., Declerck, N., van Tilbeurgh, H. (2005) Activation of the LicT Transcriptional Antiterminator Involves a Domain Swing/Lock Mechanism Provoking Massive Structural Changes. *J. Biol. Chem.* **280**: 14780-14789.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**(1): 281-299.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J., Vajda, S., Vakser, I., Wodak, S.J. (2003) CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **52**(1): 2-9.

Janin, J., Rodier, F. (1995) Protein-protein interaction at crystal contacts. *Proteins* **23**(4): 580-7.

Janin, J. (1997) Specific versus non-specific contacts in protein crystals. *Nat. Struct. Biol.* **4**(12): 973-4.

Janin, J. (1995). 'Elusive affinities' *Proteins* **21**(1): 30-9.

Jaramillo, A., Wodak, S.J. (2005) Computational protein design is a challenge for implicit solvation models. *Biophys. J.* **88**(1): 156-171.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., Greenblatt, J.F. (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**(7084): 637-43.

Lo Conte, L., Chothia, C., Janin, J. (1999) The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* 285(5): 2177-98.

McLachlan, A. (1982) Rapid Comparison of Protein Structures. *Acta Crystallogr.* **A38**: 871-873

Mendez, R., Leplae, R., De Maria, L., Wodak, S.J. (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **52**(1): 51-67.

Mendez, R., Leplae, R., Lensink, M.F., Wodak, S.J. (2005) Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins* **60**(2): 150-69.

Moult, J. (2005) A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **15**(3): 285-9.

Ogata, K., Jaramillo, A., Cohen, W., Briand, J.P., Connan, F., Choppin, J., Muller, S., Wodak, S.J. (2003) Automatic sequence design of major histocompatibility complex class I binding peptides impairing CD8+ T cell recognition. *J. Biol. Chem.* **278**(2): 1281-1290.

Rodier, F., Bahadur, R.P., Chakrabarti, P., Janin, J. (2005) Hydration of protein-protein interfaces. *Proteins* **60**(1): 36-45.

Rohl, C.A., Strauss, C.E., Chivian, D., Baker, D. (2004) Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* **55**(3): 656-677.

Rohl, C.A., Strauss, C.E., Misura, K.M., Baker, D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.* 383: 66-93.

Sali, A., Glaeser, R., Earnest, T., Baumeister, W. (2003) From words to literature in structural proteomics. *Nature* **422**(6928): 216-225.

Sundberg, E.J., Li, H., Llera, A.S., McCormick, J.K., Tormo, J., Schlievert, P.M., Karjalainen, K., Mariuzza, R.A. (2002) Structures of Two Streptococcal Superantigens Bound to TCR beta Chains Reveal Diversity in the Architecture of T Cell Signaling. *Structure (Camb.)* **10**: 687-99.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**(6770): 623-627.

Vajda, S. (2005) Classification of protein complexes based on docking difficulty. *Proteins* **60**(2): 176-80.

Wodak, S.J., Janin, J. (2002) Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61**: 9-73.

Wodak, S.J., Mendez, R. (2004) Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* **14**(2): 242-249.

Wodak, S.J., Janin, J. (1978) Computer analysis of protein-protein interaction. *J. Mol. Biol.* **124**: 323-342.

An Infrared Study of Fibril Formation in Insulin from Different Sources

IGOR DE LA ARADA AND JOSE LUIS R. ARRONDO
Unidad de Biofísica (Centro Mixto CSIC-UPV/EHU) and Departamento de Bioquímica. Universidad del País Vasco. P.O. Box 644, E-48080 Bilbao, Spain, Phone: +34 946012485 Fax: +34 946013360, E-mail: joseluis.arrondo@ehu.es

Abstract

Amyloid fibrils are proteinaceous aggregates that can be formed in the process of degenerative diseases, such as Alzheimer's and Creutzfeldt-Jakob diseases. The process of fibril formation can also be observed, under appropriate conditions, in many proteins not involved in neurodegenerative diseases. Insulin, a peptide hormone consisting of two polypeptides linked together by two interchain and one intrachain disulfide bonds, is a model of fibril formation that has produced a wealth of biochemical and structural data making it an excellent model for amyloid studies. Insulin from different mammal species, such as human recombinant, bovine and porcine, has small differences in sequence that produce variations in the three-dimensional structure. Infrared spectroscopy, although it is not a high-resolution technique, it presents the advantages of fast-time response and wider applicability required for studying aggregated materials. The time-course of fibril formation can be followed looking at the appearance of a characteristic band in the region of β-sheet structure. Human insulin, with a different aminoacid in the N-terminal segment, has a lower time in fibril formation than bovine or porcine. The wavenumber and the percentage of the band corresponding to the fibril is different in bovine as compared with human and porcine insulin, what is

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 21–31.

associated with a change in aminoacids 8 and 10 located in the intrachain disulfide bond. The results show that even if the macromolecular structure of the fibrils is alike, the process is different depending on small changes in protein sequence.

Keywords: Amyloid, Fibrils, Infrared spectroscopy, Insulin, Protein structure

1. INTRODUCTION

1.1. Potein folding

One of the characteristics defining a living cell is the ability for self-assembling its molecular structure. The folding of polypeptide chains as compact and functional three-dimensional structures is one of the macromolecular examples of chains with a correct fold. In contrast with DNA where one strand guides the structure of the other, all of the information for protein folding must be contained within the primary structure (protein sequence). How a polypeptide chain adopts a specific three-dimensional structure from the primary structure is one of the more important challenges in Biochemistry. Generating biological activity is not the only biological process associated with polypeptide folding. We know now many other biological processes associated with protein folding including the trafficking of molecules to specific cellular locations and the regulation of cellular growth and differentiation. Only correctly folded proteins are not easily degraded in crowded biological environments and are able to interact selectively with their natural partners (Dobson 2003). Thus the failure of proteins to fold correctly, or to remain correctly folded, is the origin of a wide variety of pathological conditions.

The mechanism by which a polypeptide chain folds to a specific three-dimensional protein structure has until recently been based in the 'Thermodynamic Hypothesis', proposed by Anfinsen (Anfinsen 1973) where the functional native structure of a protein is determined by the full complement of interactions formed by the protein's amino acid sequence in a given environment. However, the total number of possible conformations would be so high, that the folding of a protein would take a very long time, not compatible with the cellular life. To overcome this problem it has been proposed that protein folding is based in a stochastic search of the many conformations accessible to a polypeptide chain (Dobson 2003; Jahn and Radford 2005). The road to the folded protein would follow an energy landscape where the selection of more favourable interactions among residues,

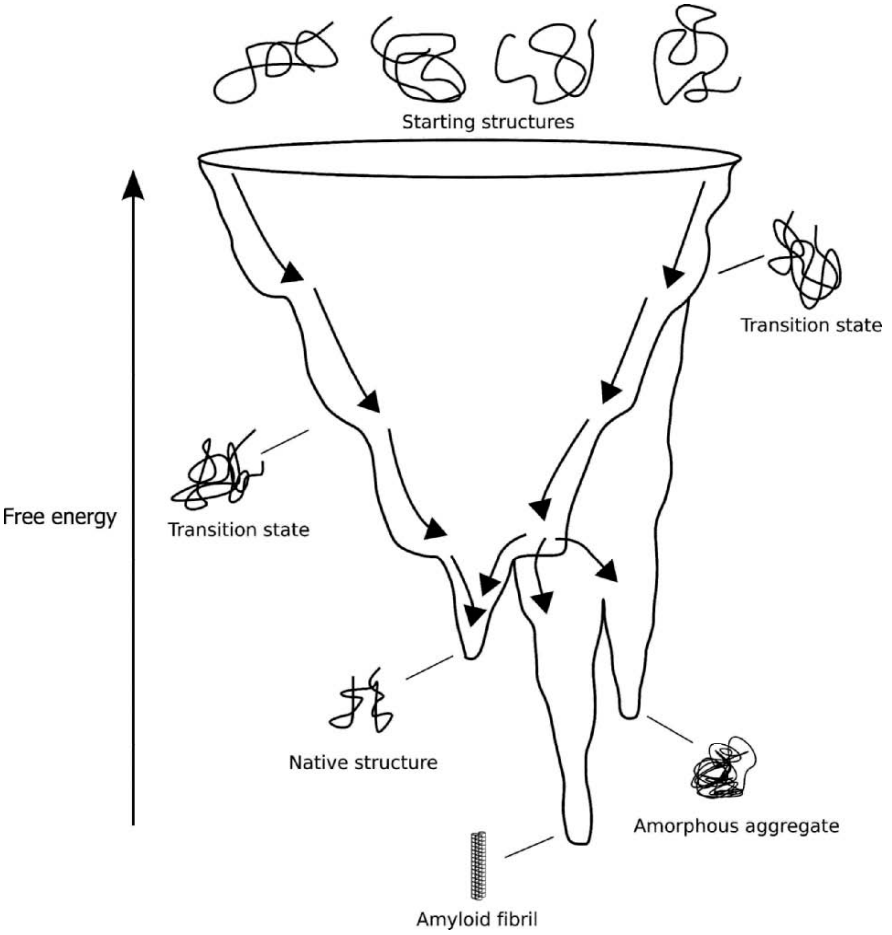


Figure 1.1. A schematic energy landscape for protein folding and/or aggregation. The multiple possible conformations are funnelled into a unique native structure, or they can converge into aggregates, fibrillar or not, that have a minimum energy level adapted from Dobson (2003) and Jahn and Radford (2005).

even if they are far apart in the aminoacid sequence, leads to a low free energy conformation by a trial and error process (Fig. 1.1). Moreover, if the energy surface or landscape has the right shape only a small number of all possible conformations need to be sampled by any given protein molecule during its transition from a random coil to a native structure. The landscape is determined by the aminoacid sequence and natural selection has enabled proteins to evolve so that they are able to fold rapidly and efficiently (Dobson 2003). In the road, several intermediates, or transition states, with different stability can be found, that can also lead to non-functional proteins. If misfolding is produced, the cellular machinery degrades the protein, but in

some cases these misfolded proteins form aggregates that can be deposed either in the same cell or in the extracellular space.

1.2. Amyloidogenesis

Amyloids are aggregates that were found in certain deposits found *post-mortem* in organs and tissues, which gave a positive reaction when stained with iodine (Virchow 1854), and it was thought that they were composed of starch-like material. Today it is known that protein constitutes the main component of amyloid deposits, although other constituents, such as glucosamineglycans are found as well. The formation of amyloid fibrils from soluble proteins through a misfolding process is a pathological condition associated with a variety of human diseases. In order to form these deposits, the proteins involved must convert from their native (folded or unfolded) states and aggregate into insoluble fibrils. The fibrils themselves and/or oligomers formed on the pathway to fibril formation are thought to be cytotoxic, and thereby directly responsible for disease initiation and progression. Amyloid-like fibrils can be obtained from a wide range of proteins and peptides, many of them not associated with any known disease (Stefani and Dobson 2003). The wide variety of proteins that give rise to similar ordered aggregates has led to the idea, following the energy landscape theory, that in appropriate conditions almost all the proteins can form amyloid-like fibrils.

Amyloid deposits were initially characterized by diagnostic staining methods and electron microscopy (Makin and Serpell 2005). Now a wide variety of biophysical techniques are used to characterize their structure including X-ray fibre diffraction, spectroscopic techniques, solid-state NMR, electron paramagnetic resonance (EPR) spectroscopy and neutron scattering (Tycko 2004). Amyloid fibrils are highly ordered molecular assemblies with a diameter of 7–10 nm. A typical X-ray fibre diffraction pattern has two characteristic signals, a sharp reflection at 4.7 Å along the same direction as the fibre and a more diffuse reflection at between 10 and 11 Å perpendicular to the fibre direction. Different spectroscopic methods have shown that all fibrillar amyloid assemblies are predominantly in β-sheet conformation. Another characteristic common to all amyloid aggregates is a clear green–gold birefringence upon staining with Congo red dye (Puchtler and Sweat 1965).

1.3. Amyloid formation and infrared spectroscopy

Infrared spectroscopy provides information on the conformation and dynamics of biomolecules measuring the absorption of infrared energy by the

vibrating chemical bonds. The energy absorbed is correlated with characteristic vibrational modes connected with the band frequency of the absorption maximum and its molecular conformation. Infrared spectroscopy was one of the first techniques used in the study of the conformation of biomolecules, but until the advenement of techniques based in the Michelson interferometer and the development of microprocessors, the information provided by the technique was scarce. The possibilities of the technique become apparent especially when applied to large proteins in turbid suspensions, as is often the case with aggregated or membrane proteins. Structural analysis usually implies a mathematical approach used to extract information contained in the composite bands, designated in IR spectroscopy as "amide bands", obtained from proteins. The amide I band, located between 1700 and 1600 cm^{-1} that arises mainly from the C = O stretching vibration is conformationally sensitive and the components arising from α-helix (1650-1653 cm^{-1}), canonical β-sheet (1630-1636 cm^{-1}) or aggregated components (1615-1630 cm^{-1}) are apart enough to distinguish α →β transitions such as the found in fibril formation. Commonly used methods of analysis imply narrowing the intrinsic bandwidths to visualise 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). Although IR spectroscopy lacks the high resolving power of X-ray crystallography or nuclear magnetic resonance (NMR), it presents the advantages of fast-time response and wider applicability required for studying aggregated materials.

Infrared spectroscopy has been widely used to study fibril formation in different protein and peptides including Alzheimer's peptides (Fabian et al. 1993), prions (Caughey et al. 1991; Gasset et al. 1993), insulin (Bouchard et al. 2000; Nielsen et al. 2001) and other systems (Dluhy et al. 2003; Guijarro et al. 1998; Zurdo et al. 2001).

2. INSULIN

2.1. Structure

Insulin is a small helical protein hormone consisting of two polypeptides, chain A ~21 residues and chain B ~30 residues, linked together by two interchain disulfide bridges. In addition, the A-chain contains an intrachain disulfide bond. In the bloodstream, insulin is present at very low concentrations. At these low concentrations, insulin is monomeric, which is its biologically active form. At higher concentrations, insulin assembles into dimers and, in the presence of zinc ions at neutral pH, into hexamers. When

insulin is heated at high temperature and low pH, a series of structural changes occur resulting in the formation of fibrillar structures. Insulin forms easily amyloid fibrils *in vitro* even if it is genetically distant from proteins usually associated with pathological aggregation in humans in spite of the very high concentration in secretory granules of pancreatic β-cells. Insulin has, however, on rare occasions been found to give amyloid deposits at the site of injection in diabetic subjects (Westermark 2005). The wealth of biochemical and structural data has made it an excellent model for amyloid studies.

Insulin from different mammal sources such as human (recombinant) bovine or porcine have a very similar sequence and structure (Fig. 2.2) and therefore they can be a very good model to study small differences that can give rise to different fibril formation paths. The small differences in sequence observed are enough to produce changes in the tertiary structure of the molecule and therefore to produce variations in fibril formation. It is interesting to note that changes in aminoacids 8 and 10 in bovine insulin, located in the intra-chain disulfide bond, produce a variation in the first helicoid segment whereas the change in one aminoacid in the N terminus of human insulin also affects the terminal random coil segment of the protein.

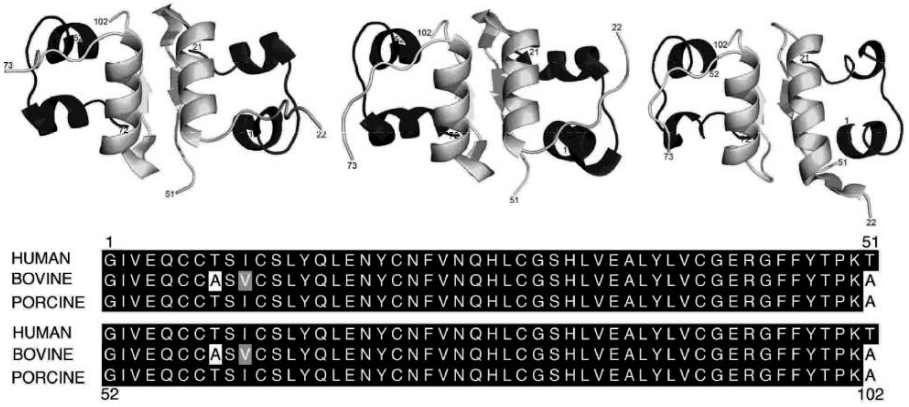


Figure 2.2. Sequence and three-dimensional structure of insulin from the three sources studied. The numbers in the Figure correspond to the aminoacids in the sequence. The Protein Data Bank codes are 1GUJ for human recombinant insulin, 1ZNI for bovine insulin and 2A3G for porcine insulin.

2.2. Fibril formation

Insulin fibril formation can be accomplished by lowering the pH of the solution below 2.0 and heating at 70°C. In these conditions, the normal pattern of insulin assembly (dimers → tetramers → hexamers) is disrupted and the monomeric species undergo conformational changes leading to the formation of insoluble amyloid fibrils. The structural study of insulin at pH 2.0 in different acid conditions suggests an increase in the susceptibility of the molecule to conformational change and dissociation and an enhancement of the rate of fibrillation because of the charged environment in which the attractive forces between the protein molecules are increased (Whittingham et al. 2002).

The infrared spectrum of a fibril has been characterized by bands in the amide I spectrum in the region corresponding to extended structures (1615-1630 cm^{-1}). The wavenumber of the band maximum is characterized by the dihedral angles and the hydrogen bonding of the C = O peptide carbonyl.

Previous infrared studies of insulin fibril formation (Bouchard et al. 2000) showed that before heating at pD 2.0 the amide I band is located around 1651 cm^{-1} indicating a prevailing α-helix structure. After increasing the temperature there is a shift of this band towards lower frequencies indicative of an increase in unordered structure and the appearance of a shoulder in the region of β-sheet structure. At longer times, it is observed the emergence of two bands at 1628 cm^{-1} and 1670 cm^{-1} corresponding to intermolecular β-sheet and β-turns.

The process of fibril formation can be followed looking at changes of the intensity maximum corresponding to the β-sheet structure. The time-intensity changes obtained after heating the samples at 70°C and pH ~1.5 from human recombinant, bovine and porcine insulin are shown in Fig. 2.3.

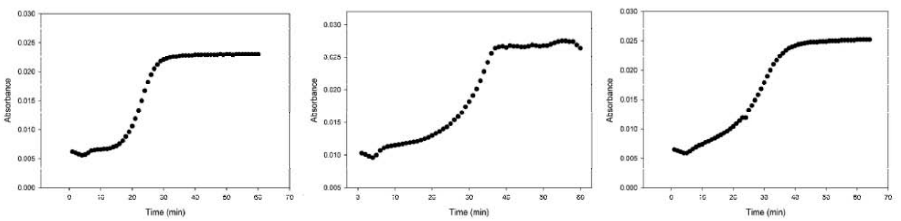


Figure 2.3. Time-course of the appearance of the band corresponding to the fibril in human insulin (left) at 1621 cm^{-1}, bovine insulin (centre) at 1627 cm^{-1} and porcine insulin (right) at 1622 cm^{-1}.

The kinetic of the process is cooperative. The lag time before the fibril is formed is consistent with the change in Amide I bandshape associated with the increase in unordered structure observed previously (Bouchard et al. 2000). The traces in Fig. 2.3 show a different time-course aggregation between human and bovine or porcine insulin. Thus, whereas the mid-point of the sigmoid transition is around 23 min in human insulin, it increases to around 32 min in bovine and porcine insulin. Looking at the aminoacid sequence shown in Fig. 2.2 it can be expected that the differences between human and bovine or porcine insulin is a change in aminoacid 51 that changes from threonine in human to alanina in bovine or porcine. Therefore, it can be assumed that the difference in the unordered structure in the N-terminal region helps the attractive forces that form the fibril.

2.3. Structure of the fibrils

The different characteristics of the fibrils can be studied by looking at the components of the amide I. Deconvolution of the original band shows that the major component of the β-sheet structure is different in bovine insulin when compared with human or porcine (Fig. 2.4). The weight of each component is obtained after band-fitting the original envelope. The mathematical solution to the decomposition may not be unique, but if restrictions are imposed such as the maintenance of the initial band positions in an interval of ± 1 cm^{-1}, the preservation of the bandwidth within the expected limits, or the agreement with theoretical boundaries or predictions, the result becomes, in practice, unique (Arrondo and Goñi 1999).

Band decomposition is depicted in Fig. 2.5 and the values corresponding are shown in Table 1.

There are several bands corresponding to vibrations from the peptide bond and one at around 1607 cm^{-1} corresponding to aminoacid side chains. In the three samples, two bands in the region 1610-1630 cm^{-1} can be observed. In bovine insulin the major component is around 1628 cm^{-1} whereas in human and porcine the major component is around 1620 cm^{-1}. In all the cases, no high frequency component is observed. Another component corresponding to extended chains is located at 1614-1619 cm^{-1}. There are also bands at higher wavenumbers that have been associated with turns (Arrondo et al. 1993). It is interesting to note that in bovine insulin the major component is the band around 1628 cm^{-1} whereas in human and porcine insulin is the band around 1619 cm^{-1}. If the structure is compared, bovine insulin has two different aminoacids in position 8 and 10 that are inside the segment defined by the intrachain disulfide bond between cysteines 6 and 11. The changes in

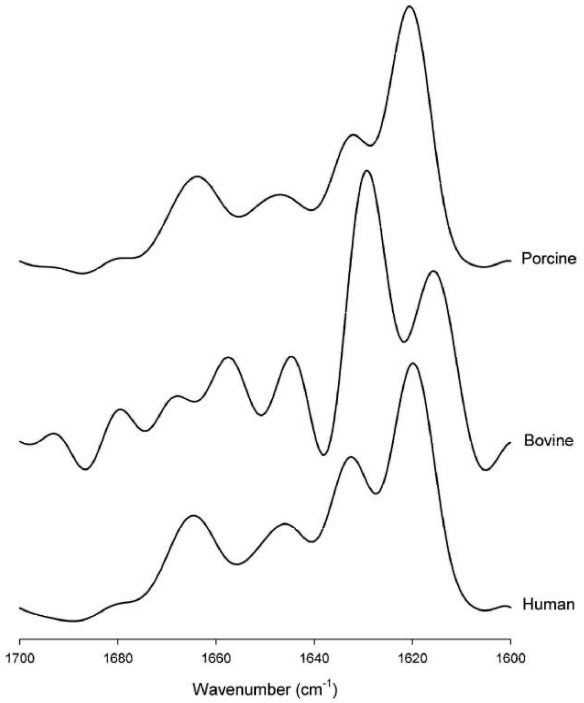


Figure 2.4. Amide I deconvolved spectra corresponding to human (bottom), bovine (middle) and porcine (upper) insulin. The deconvolving parameters are FWHH = 18 cm^{-1}) and a K = 2.0.

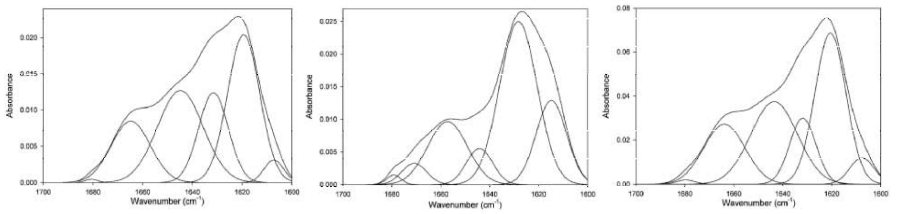


Figure 2.5. Band-fitting of the amide I envelope corresponding to the three insulin samples. Human (left), Bovine (middle) and porcine (right) are shown. The components have been obtained after an iterative process. The percentage values of the areas is shown in Table 1.

the aminoacids produce a different structure in the segment (Fig. 2.2) which can be accounted for the different infrared spectra. In the insulin fibril, the disulfide bonds are maintained what could be the cause of the two components observed.

Table 1. Percentage area values corresponding to the band-fitting of the Amide I derived from human, bovine or porcine insulin.

HUMAN		BOVINE		PORCINE	
Position (cm^{-1})	Area (%)	Position (cm^{-1})	Area (%)	Position (cm^{-1})	Area (%)
1680	0.5	1679	1	1680	0.5
-	-	1671	4	-	-
1664	17	1657	19.5	1664	19
1644	30	1644	7.5	1644	30
1631	18.5	1628	48	1631	13.5
1619	34	1614	20	1619	37

3. CONCLUSIONS

The process of fibril formation can be easily followed by infrared spectroscopy. In insulin from different sources, fibril formation is characterized by the appearance of bands in the region of extended structures without a high frequency component, similar to what has been found in parallel β-sheet in solids. The increase in intensity of the band can be used to follow the time course of fibril formation. A different value is obtained in human as compared with bovine or porcine insulin. The structural difference is a change in the N-terminal aminoacid of chain B pointing to a different tertiary structure. Band-decomposition of amide I showed a different shape between bovine as compared with human or porcine. In this case, the structural difference is due to aminoacids 8 and 10 that are located in the intrachain disulfide bond (6-11) of chain A. It can be concluded that even if the macroscopic aspect as seen by electron microscopy and the reactivity to Congo red is the same in the three samples studied, the changes in aminoacid composition that are associated with variations in three-dimensional structure produce alterations in the molecular structure of the fibril.

ACKNOWLEDGEMENTS

This work has been supported in part by grant BFU2006-14423 from DGICT.

REFERENCES

Anfinsen, C.B. (1973) Principles That Govern Folding of Protein Chains. *Science* **181:** 223-230.

Arrondo, J.L.R. and Goñi, F.M. (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **72:** 367-405.

Arrondo, J.L.R., Muga, A., Castresana, J., and Goñi, F.M. (1993) Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **59:** 23-56.

Bouchard, M., Zurdo, J., Nettleton, E.J., Dobson, C.M., and Robinson, C.V. (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci.* **9:** 1960-1967.

Caughey, B.W., Dong, A., Bhat, K.S., Ernst, D., Hayes, S.F., and Caughey, W.S. (1991) Secondary structure analysis of the scrapie-associated protein PrP 27-30 in water by infrared spectroscopy [published erratum appears in Biochemistry 1991 Oct 29; 30(43): 10600]. *Biochemistry* **30:** 7672-7680.

Dluhy, R.A., Shanmukh, S., Leapard, J.B., Kruger, P., and Baatz, J.E. (2003) Deacylated Pulmonary Surfactant Protein SP-C Transforms From alpha-Helical to Amyloid Fibril Structure via a pH-Dependent Mechanism: An Infrared Structural Investigation. *Biophys. J.* **85:** 2417-2429.

Dobson, C.M. (2003) Protein folding and misfolding. *Nature* **426:** 884-890.

Fabian, H., Szendrei, G.I., Mantsch, H.H., and 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.

Gasset, M., Baldwin, M.A., Fletterick, R.J., and Prusiner, S.B. (1993) Perturbation of the secondary structure of the scrapie prion protein under conditions that alter infectivity. *Proc. Natl. Acad. Sci. USA* **90:** 1-5.

Guijarro, J.I., Sunde, M., Jones, J.A., Campbell, I.D., and Dobson, C.M. (1998) Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95:** 4224-4228.

Jahn, T.R. and Radford, S.E. (2005) The Yin and Yang of protein folding. *FEBSD J.* **272:** 5962-5970.

Makin, O.S. and Serpell, L.C. (2005) Structures for amyloid fibrils. *FEBS J.* **272:** 5950-5961.

Nielsen, L., Frokjaer, S., Carpenter, J.F., and Brange, J. (2001) Studies of the structure of insulin fibrils by Fourier transform infrared (FTIR) spectroscopy and electron microscopy. *J. Pharm. Sci.* **90:** 29-37.

Puchtler, H. and Sweat, F. (1965) Congo Red as a Stain for Fluorescence Microscopy of Amyloid. *J. of Histochem.Cytochem.* **13:** 693-694.

Stefani, M. and Dobson, C.M. (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81:** 678-699.

Tycko, R. (2004) Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr. Opin. Struct. Biol.* **14:** 96-103.

Virchow, R. (1854) Zur Celluslose-Frage. *Virchows Arch.* **6:** 415-426.

Westermark, P. (2005) Aspects on human amyloid forms and their fibril polypeptides. *FEBS J.* **272:** 5942-5949.

Whittingham, J.L., Scott, D.J., Chance, K., Wilson, A., Finch, J., Brange, J., and Dodson, G.G. (2002) Insulin at pH 2: Structural analysis of the conditions promoting insulin fibre formation. *J. Mol. Biol.* **318:** 479-490.

Zurdo, J., Guijarro, J.I., Jimenez, J.L., Saibil, H.R., and Dobson, C.M. (2001) Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J. Mol. Biol.* **311:** 325-340.

This page intentionally blank

Some Basic Biomolecular NMR for Protein Structure Determination

ANGELA M. GRONENBORN
Rosalind Franklin Professor, Department of Structural Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA, Phone: (412) 648-9959, Fax: (412) 648-9008, E-mail: amg100@pitt.edu

Abstract

Atomic structures are critical for understanding biological processes at the molecular level. NMR has become has become the most effective and reliable methodology for three-dimensional structure determination of biological macromolecules in solution. The present overview describes the basic methodology. NMR structures depend primarily on Nuclear Overhauser Effect derived distance restraints between protons close in space (<6 Å), supplemented by ^{3}J scalar couplings that report on local dihedral angles. Large systems can be investigated using uniformly ^{13}C/^{15}N/^{2}H labeled molecules and sequential assignment strategies based solely on heteronuclear through-bond correlations have become the norm. In addition, a number of recent advances, such as using Residual Dipolar Couplings for refinement and relative positioning of domains are introduced.

Keywords: NMR, structure determination, heteronuclear, NOE, RDC

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 33–44.

1. INTRODUCTION

Traditionally, NMR determination of three-dimensional structures of biological macromolecules employs Nuclear Overhauser Effect (NOE) derived distance restraints supplemented by torsion angle restraints extracted from J-couplings (Wüthrich 1986, Clore and Gronenborn 1989). In general, using NOEs allows the identification of protons separated by less than 6Å since the size of the effect is proportional to the inverse sixth power of the distance between the protons. Despite the short range nature of the observed interactions, a large number of approximate distances are conformationally highly restrictive, particularly when they involve residues that are far apart in the sequence but close together in space. This led to structural NMR becoming highly successful, now standing side-by-side with X-ray crystallography as a technique to derive atomic resolution structures of biological macromolecules.

2. NMR METHODOLOGY AND PARAMETERS

The power of NMR over other spectroscopic techniques results from the fact that every proton gives rise to an individual resonance in the spectrum which can be resolved by 2, 3 and 4-dimensional techniques. In order to measure distances between these hydrogen atoms via NOEs, the corresponding resonances have to be identified unambiguously. Thus, each resonance has to be assigned to a unique atom in a specific monomer unit (amino acid for proteins, nucleotide for nucleic acids) in the macromolecule. This is achieved via a combination of through-bond and through-space correlations. For proteins, conventional sequential resonance assignment, relies on 2D homonuclear ^{1}H–^{1}H through-bond correlation experiments to identify amino acid spin systems coupled with 2D ^{1}H–^{1}H NOE experiments to identify through-space (<5 Å) sequential connectivities of the type CαH(i)–NH(i+1,2,3,4), NH(i)–NH(i±2) and CαH(i)–CβH(i+3) (Wüthrich, 1986). This works well for proteins up to about 100 residues, albeit with considerable effort. For larger proteins, the increased spectral complexity renders such 2D experiments no longer sufficient, and the spectral resolution has to be increased by increasing the dimensionality of the spectra (Clore and Gronenborn 1991a). Therefore, a sequential assignment strategy based solely on through-bond correlations involving large heteronuclear one-bond and two-bond couplings along the polypeptide chain was developed (Bax and Grzesiek 1993; Clore and Gronenborn 1991b). Using this suite of experiments backbone assignments

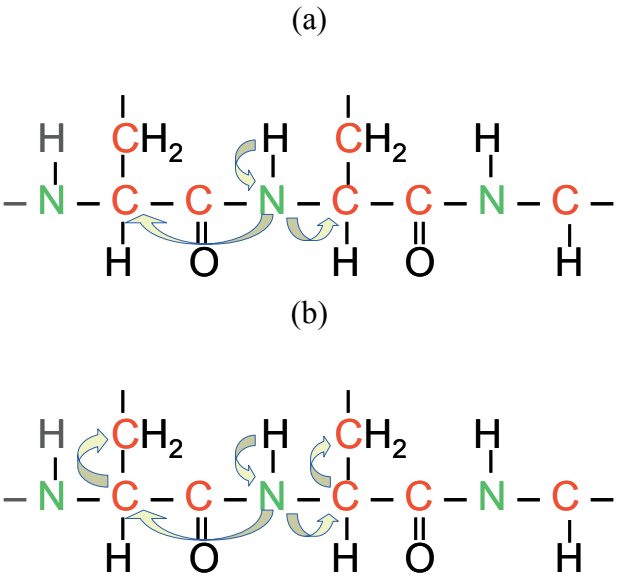


Figure 2.1. Correlations observed in HNCA and HNCACB.

for proteins have become extremely fast and efficient. The magnetization transfer active in two such experiments (HNCA and HNCACB) is illustrated in Figure 2.1.

For proteins with molecular masses greater than ~25 kDa, deuteration of non-exchangeable (carbon-attached) protons becomes necessary. This results in a significant reduction in linewidths brought about by the increase in transverse relaxation (T_2) times. For these large systems, transverse relaxation-optimized spectroscopy (TROSY) using constructive interference between dipole–dipole coupling and chemical shift anisotropy results in substantial line narrowing (Pervushin et al. 1997), especially in combination with perdeuteration. In this manner, one of the major impediments for applying NMR methods to big proteins, namely broad linewidths caused by slow molecular tumbling, is removed and the door is opened for solution NMR to be applicable to even more systems.

While backbone assignments can be achieved readily for the majority of proteins, even large ones, it may not always be possible to obtain complete sidechain proton assignments, either because of spectral overlap or the absence of through-bond connectivities. In such cases, specific deuteration can be extremely helpful.

For instance, labeling strategies in which the amides and methyl groups are protonated, while all other groups are deuterated have been devised (Gardner et al. 1997). Using appropriate precursors the side chains of particular residues may be uniformly ^{13}C-labeled or, alternatively, the ^{13}C labels can be restricted solely to specific methyl groups (Tugarinov et al. 2004).

While the panoply of 3D heteronuclear experiments is sufficient for the purposes of spectral assignment, structure determination requires unambiguous and reliable identification of NOE through-space interactions since NOE-derived distance constraints are the principal experimental parameters that are used for NMR structure calculations, supplemented by torsion angle constraints and chemical shifts. This is achieved by further extending the dimensionality of heteronuclear experiments. In this manner, each ^{1}H–^{1}H NOE interaction is specified by four chemical shift coordinates, the two protons giving rise to the NOE and the heavy atoms to which they are attached. Using the complete ^{1}H, ^{15}N and ^{13}C resonance assignments, analysis of 4D ^{15}N/^{13}C, ^{13}C/^{13}C and ^{15}N/^{15}N separated NOE spectra are recorded for identification of a large number of NOE interactions. A schematic illustration of 3- and 4D heteronuclear edited spectra is provided in Figure 2.2.

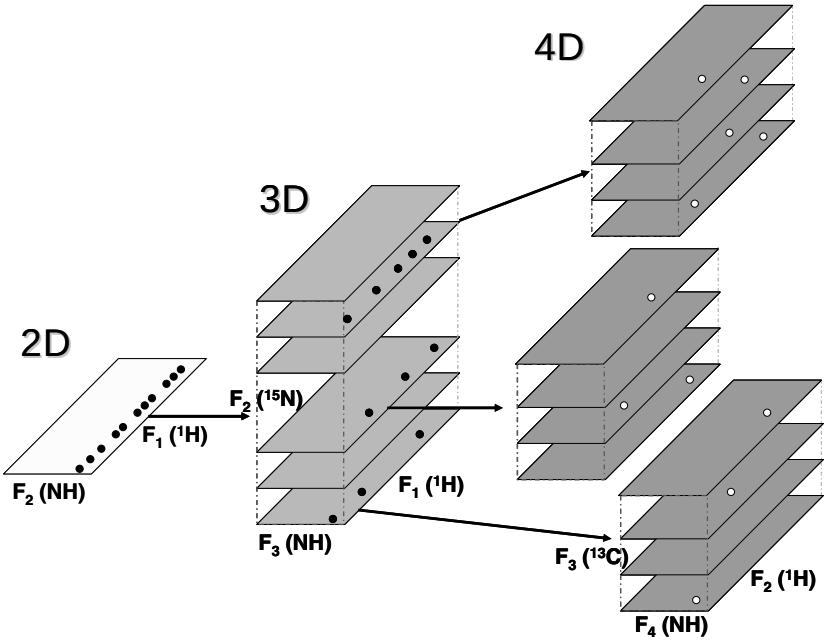


Figure 2.2. Concept of 2-, 3- and 4D heteronuclear edited NOE spectroscopy.

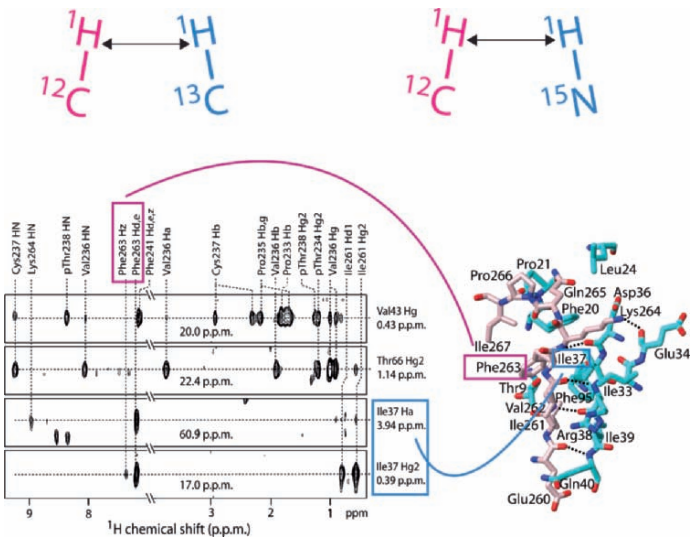


Figure 2.3. Intermolecular NOEs between labeled and unlabeled proteins in a protein-protein complex.

The above described concept of heteronuclear, multidimensional spectroscopy can be extended and modified for applications to protein complexes. To that end, complexes are prepared in which one component (protein or nucleic acid) is labeled with ^{15}N and ^{13}C and the partner is unlabeled (i.e. at natural isotopic abundance). Using a combination of heteronuclear filtering and editing allows one to record individual experiments in which only correlations within the labeled component or unlabeled partner are observed, thereby reducing the spectral complexity. In addition, correlations that are specific for inter-molecular contacts can also be unambiguously identified. The latter contacts are unique and can be obtained from either side of the interface by employing two complex samples with alternating labeled partners. An example of a 3D ^{13}C-edited, $^{12}C/^{14}N$-filtered NOE spectrum as well as the responsible contacts in the final structure of the protein-protein complex are illustrated in Figure 2.3.

Although structure determination by NMR has been highly successful over the last two decades, the use of only short distance information imposes limits on the attainable accuracy of NMR derived structures, especially for elongated structures in which small errors may propagate. In addition, for determining domain-domain orientations, frequently only few experimentally observable NOEs are available between structural elements, causing their relative orientation to remain ill defined. Examples of such systems include modular and multi-domain proteins and linear nucleic acids. Such shortcomings can be alleviated by inclusion of structural restraints that report on the orientation

relative to an external axis system. For example, the dependence of heteronuclear T_1/T_2 ratios on the rotational diffusion anisotropy axis (Tjandra et al. 1997a) or residual dipolar couplings (RDCs) that report on the orientation of inter-nuclear vectors relative to the molecular susceptibility tensor, or the magnetic field (Tjandra et al. 1997b) can be used for that purpose.

The use of magnetic field induced RDCs of proteins in solution was only realized with the advent of high field magnets and heteronuclear experiments, since their magnitude scales with the square of the magnetic field and novel experiments had to be devised to extract these extremely small couplings (Tjandra et al. 1996; Tolman and Prestegard 1996). Given the small degree of alignment of diamagnetic proteins in the magnetic field, RDCs of typically <0.2 Hz ensue, making their routine determination extremely difficult. Therefore, magnetic field induced alignment seemed initially only useful for paramagnetic proteins, nucleic acids and protein/nucleic acid complexes which exhibit magnetic susceptibility anisotropies of greater -10×10^{-34} m^3 per molecule.

This view changed dramatically with the observation that tunable degrees of molecular alignment can be achieved by placing the molecule under investigation into a dilute, aqueous liquid crystalline phase (Tjandra and Bax 1997; Prestegard 1998; Bax et al. 2001). Sufficiently high degrees of alignment can be obtained for almost any molecule in a variety of media. To date, at least a dozen of media for weakly aligning molecules have been described (Gronenborn 2002; Prestegard et al. 2004) and are routinely used in biomolecular NMR. In these media the degree of alignment is tuned such that the resulting one-bond $^1D_{NH}$ dipolar couplings range from 5-40 Hz, making them easy to detect. The most commonly measured RDCs for proteins are the five backbone one-bond ^{15}N-$^1H^N$, ^{15}N-$^{13}C'$, $^{13}C'$-$^{13}C^\alpha$ and $^1H^\alpha$-$^{13}C^\alpha$ and two-bond $^1H^N$-$^{13}C'$ residual dipolar couplings (deAlba and Tjandra 2002). In addition, one-bond $^{13}C^\alpha$-$^{13}C^\beta$ and $^1H^\beta$-$^{13}C^\beta$ and one-bond ^{13}C-^{13}C and ^1H-^{13}C couplings in methyl groups can also be measured (Ottiger and Bax 1999; Chou and Bax 2001).

A large number of different experiments are available for measuring RDCs, either by quantitative J correlation experiments using intensity analysis (Bax et al. 1994) or from the splittings in 2D and 3D coupled HSQC spectra (Ding and Gronenborn 2004).

3. STRUCTURE DETERMINATION

Irrespective of the algorithm used, any structure determination by NMR seeks to find the global minimum region of a target function, E_{tot}, given by:

$$E_{tot} = E_{cov} + E_{vdw} + E_{NMR}$$

where E_{cov}, E_{vdw}, and E_{NMR} are terms representing the covalent geometry (bonds, angles, planarity, and chirality), the non-bonded contacts, and the experimental NMR restraints, respectively. Currently available methodologies are based on simulated annealing (Nilges et al. 1988), metric matrix distance geometry (Havel and Wüthrich 1985), or minimization with a variable target function in torsion angle space (Braun, 1987). The currently most frequently used programs in the NMR community are Xplor/CNS (Brünger et al. 1998; Schwieters et al. 2002) or DYANA/CYANA (Güntert et al. 1997). Systematic bias arising from the different algorithms used for structure calculations can be introduced via the first two terms, E_{cov} and E_{vdw}, with E_{vdw} being associated with considerably more uncertainty than those specifying the covalent geometry. Given the numerous ways to represent E_{vdw} (for example, a simple van der Waals repulsion term or a complete empirical energy function including a van der Waals Lennard-Jones 6-12 potential, an electrostatic potential, and a hydrogen bonding potential) variability is introduced which naturally affects the final outcome. Nevertheless, the uncertainties associated with the covalent geometry and van der Waals terms will at most introduce errors of ~0.3 Å in the coordinates. The overriding determinant for the accuracy of an NMR structure resides in the number and quality of the experimental NMR restraints that enter into the third term, E_{NMR}. Depending on the parameters that are included in this term, commonly interproton distances, dihedral angles, RDCs, coupling constants, carbon and proton chemical shifts, differences in the precision of the ensemble and variability of the average will occur. The accuracy of NMR structures, on the other hand, will be affected mainly by errors in interproton distance restraints, caused by misassignments or errors in distance estimates. Errors due to misassignments can be avoided by carefully inspecting all available spectra and may become rare in the future with the introduction of robust automated assignment methodologies (Linge et al. 2001; Herrmann et al. 2002). Systematic errors in distance estimates can be avoided by converting NOE intensities into approximate interproton distance restraints (e.g., 1-8-2.7 Å, 1.8-3.3 Å, 1.8-5.0 Å, and 1.8-6.0 Å for strong, medium, weak, and very weak NOEs, respectively) with the lower bounds given by the sum of the van der Waals radii of two protons. These distance ranges are sufficiently generous to take into account untoward effects in the conversion of NOE intensities into distances and any systematic errors in the interproton distance restraints generally will be introduced only at the boundary of two distance ranges. Once 90% of the structurally useful NOEs have been assigned and incorporated into the restraints set, corresponding typically to an average of 15-20 restraints per residue (with >60% of the NOEs involving unique proton pairs), it is very unlikely that gross structural errors could persist since all experimental restraints are incorporated in an iterative fashion.

Although interproton distance restraints derived from NOEs are the most crucial parameters that allow the determination of the protein fold by NMR, direct refinement against other experimental NMR constraints is both feasible and desirable. Specifically three-bond coupling constants, secondary ^{13}C chemical shifts, ^{1}H chemical shifts and, in particular, RDCs allow to determine structures that are both, more accurate and better defined. It generally is advantageous to employ a two stage simulated annealing protocol: in the first stage, all conventional experimental constraints, such as NOE and H-bond based distance constraints, dihedral angle constraints, coupling constants and carbon chemical shifts are employed to calculate an ensemble of models. In the second stage RDCs are incorporated, which impose a tight restriction on the orientation of a bond, thereby providing qualitatively different 'long-range' constraints, resulting in an increase in the coordinate precision of the ensemble.

Like with all global optimization procedures it is essential to ensure that the global minimum of the target function is reliably and efficiently located,

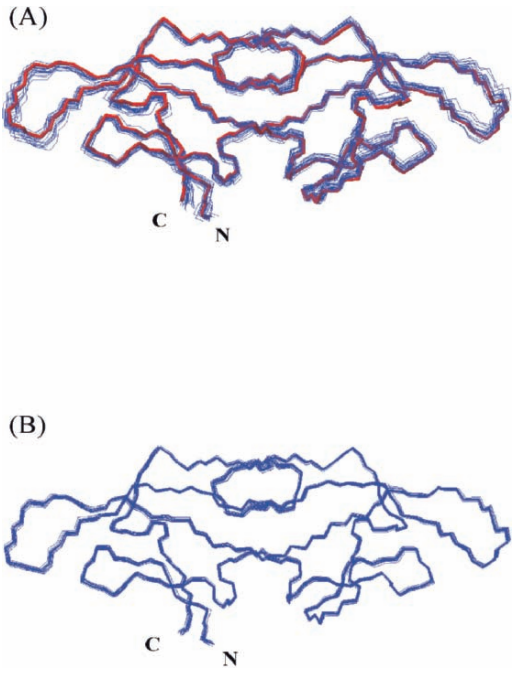


Figure 3.4. Best fit superpositions of backbone (N, Cα, C') atoms of 20 CV-N structures calculated without (A) and with residual dipolar coupling constraints (B). Individual conformers are shown in blue and the minimized average structure in red.

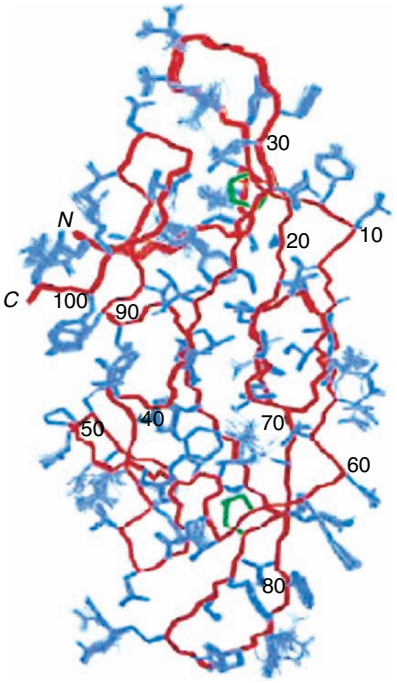


Figure 3.5. Overall structure of cyanovirin-N. The ensemble of the final 40 simulated annealing structures is displayed. The backbone is shown in red, the disulfide bridges in green, and all other side chains in blue.

without structures becoming trapped in deep local minima. Careful adjustments of the simulated annealing protocols including both high temperature and low temperature slow cooling steps have been devised and have proved to yield robust results.

Interestingly, significant improvements in the quality of the structures upon inclusion of all residual dipolar couplings relate also to the Ramachandran statistics, with an increase in percentage of residues in the most allowed regions to over 90% and the number of bad contacts reduced by more than 50%. Therefore, structures calculated even with only one-bond residual dipolar couplings exhibit superior packing characteristics. As an example, families of structures calculated without and with inclusion of NH residual dipolar couplings are displayed in Figure 3.4 for the potent HIV-inactivating protein cyanovirin-N, vividly demonstrating the improved precision.

The overall ensemble of structures is shown in Figure 3.5 and is representative of the currently attainable quality of NMR structures.

Recently a radically new structure determination principle, called Inferential Structure Determination (ISD), has been developed that uses Bayesian inference to derive a probability distribution representing the structure as well as its precision (Rieping et al. 2005). This method deals with the uncertainties in experimental data and shortcomings in computational models and statistically meaningful figures of merit for NMR structures can be derived relating to atomic uncertainties. It remains to be seen whether NMR structure determination in an entirely probabilistic framework will become widely used in the community.

ACKNOWLEDEGMENTS

I am indebted to all my collaborators and colleagues mentioned in the references who have developed the technologies described in this article and who have been a constant source of stimulating discussions. Particularly, Drs. Ad Bax, Marius Clore and Nico Tjandra provided invaluable intellectual contributions.

REFERENCES

Bax, A. and Grzesiek, S. (1993) Methodological Advances in Protein NMR. *Acc. of Chem. Res.* **26**: 131-138.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R., Zhu, G. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**: 79-105.

Bax, A., Kontaxis, G. and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods Enzymol.* **339**: 127-174.

Braun, W. (1987) Distance geometry and related methods for protein structure determination from NMR data. *Q Rev Biophys*, **19**: 115-157.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905-921.

Chou, J.J. and Bax, A. (2001) Protein side-chain rotamers from dipolar couplings in a liquid crystalline phase. *J. Am. Chem. Soc.* **123**: 3844-3845.

Clore G.M. and Gronenborn, A.M. (1989) Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol.* **24**(5): 479-564.

Clore, G.M. and Gronenborn, A.M. (1991a) Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**: 1390.

Clore G.M. and Gronenborn A.M. (1991) Applications of three- and four dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. NMR Spectrosc.* **23**: 43-92.

deAlba, E. and Tjandra, N. (2004) Residual dipolar couplings in protein structure determination. *Methods Mol. Biol.* **278**: 89-106.

Ding, K. and Gronenborn, A.M. (2004) Protein backbone 1H(N)-13Calpha and 15N-13Calpha residual dipolar and J couplings: New constraints for NMR structure determination. *J. Am. Chem. Soc.* **126**: 6232-6233.

Gardner, K.H., Rosen, M.K. and Kay, L.E. (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36**:1389-1401.

Gronenborn, A.M. (2002) The importance of being ordered: improving NMR structures using residual dipolar couplings. *C. R. Biol.* **325**: 957-966.

Gűntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**: 283-298.

Havel, T.F. and Wűthrich, K. (1985) An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J. Mol. Biol.* **182**: 281-294.

Herrmann, T., Guntert, P. and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**: 171-189.

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol.* **339**: 71-90.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**: 317-324.

Ottiger, M. and Bax, A. (1999) Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J. Biomol. NMR* **13**: 187-191.

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) Attenuated T2 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.

Prestegard, J.H. (1998) New techniques in structural NMR-anisotropic interactions. *Nat. Struct. Biol.* **5**: 517-522.

Prestegard, J.H., Bougault, C.M. and Kishore, A.I. (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* **104**: 3519-3540.

Rieping, W., Habeck, M. and Nilges, M. (2005) Inferential structure determination. *Science* **309**: 303-306.

Schwieters, C.D., Kuszewski, J., Tjandra, N. and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**: 66-74.

Tjandra, N., Grzesiek, S., and Bax, A. (1996). Magnetic field dependence of nitrogen-proton J splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Amer. Chem. Soc.* **118**, 6264-6272.

Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**: 1697.

Tjandra, N., Garrett, D.S., Gronenborn, A.M., Bax, A. and Clore, G.M. (1997a) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat. Struct. Biol.* **4**: 443-449.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997b). Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4**: 732-738.

Tolman, J.R., and Prestegard, J. H. (1996). A quantitative J-correlation experiment for the accurate measurement of one-bond amide ^{15}N-^{1}H couplings in proteins. *J. Magn. Reson*. Series B **112**, 245-252.

Tugarinov, V., Hwang, P.M. and Kay, L.E. (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Ann. Rev. Biochem.* **73**: 107-146.

Wüthrich, K. (1986) NMR of proteins and nucleic acids. John Wiley & Sons, New York.

Solid State NMR for Studying Membrane Proteins

ANTHONY WATTS
Biomembrane Structure Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK, Phone: +44 1865 275268, Fax: +44 1865 275234/275259, E-mail: anthony.watts@bioch.ox.ac.uk

Abstract

Solid state NMR is now established as a method for resolving structural information for large biomolecular complexes such as membrane-embedded proteins. In principle, there is no molecular weight limit to the use of the approach, although the complexity and volume of data is still outside complete assignment and structural determinations for any large ($M_r > \sim 40k$) complex, unless specific methods are used to reduce the information content. Such methods include specific residue type labelling, labelling of putative segments of a protein, probing ligand binding sites with labelled ligands, or examination of complexes made up of smaller, manageable units, such as oligomeric ion channels. Labelling possibilities often follow models from a bioinformatics approach. In all cases, and in common with most membrane studies, sample preparation is vital, and this activity alone can take considerable effort before NMR can be applied – peptide or protein production (synthesis or expression) followed by reconstitution into bilayers and/or chemical synthesis, and then resolution of suitable sample geometry, is still technically challenging. As experience is gained in the field, this development time should decrease. Here, a brief overview of the use of solid state NMR for membrane protein structural determinations will be presented.

Keywords: Solid state NMR, biomembranes, ion channels, GPCR, drug-receptor interactions, protein structure.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 45–73.

1. INTRODUCTION

From the ~35,000 atomic resolution structures in the protein data base, some 6,000 have unique folds but only ~40 are of membrane proteins. This lack of information about membrane proteins would not be so important if prediction methods were more effective. However, since this direct data is not available in the quantity desired, it is often suggested that membrane proteins present the last remaining major challenge in structural biology.

The lack of structural data does not parallel the intense interest in membrane proteins. Some 85% of all cellular signaling occurs through the plasma membrane, and membranes feature at some point in every described cell function. Coupled with this central role in cell function is the major drive in the pharmaceutical industry to exploit membrane proteins as new drug targets (Terstappen and Reggiani 2001) – about 70% of all known drugs target to ~5% of all membrane proteins, leaving a large number of undiscovered or uncharacterised targets still to be exploited. A structural biology approach to membrane proteins as drug targets is the basis of a new major initiative in the drug industry for the next decade, as an alternative to combinatorial chemistry screening programmes, and thus new methods for their structural resolution are urgently needed.

Membrane proteins vary in size from small, 20 – 30mers which can either be surface associated or span the membrane, such as melittin, gramicidin A and some fusion peptides, to large multi-complex proteins. Families are usually described in terms of the number of trans-membrane domains they (often putatively from predictive methods) possess in integral proteins (Arkin et al. 1997). Single TM peptides can then associate to form ion channels, and larger complexes, or even with other proteins (such as phospholamban and the Ca^{++}-ATPase). Interestingly, *crystallographic* structure determinations have been made mainly for very large (M_r > ~300k), multi subunit complexes (respiratory complexes, bacterial antennae complexes), with smaller proteins (7TMDs and less; M_r ~40k) being very much the exception in the data base – bR, rhodopsin, the potassium channel KcsA and the mechanosensitive channels MscL and MscS being the only examples at present. Conversely, it is this lower molecular weight range which is more accessible to solid and solution state NMR, in view of the complexity of the spectra for larger proteins. NMR approaches (such as TROSY) which have been applied to (deuterated) β-barrel proteins in solution, have potential for proteins rich in α-helices, and developments are still eagerly awaited (Wüthrich 1998).

Membrane associating peptides can be produced by solid phase peptide synthesis (SPPS) or expression, and the method chosen is dictated by the type of structural information required. Uniform labelling of a peptide is best

accomplished by expression in a bacterial vector (*E. coli*) with fusion proteins (often glutathione transferease, GST). For ^{15}N labelling, ^{15}N ammonium chloride and for ^{13}C, ^{13}C sodium acetate metabolites (Goto and Kay 2000) or ^{13}C-glucose is incorporated in a minimal growth medium (Castellani et al. 2002), respectively. Selective labelling can be achieved by supplementing the bacterial expression growth medium with certain (non-scrambled) labelled amino acids, or with protected residues in SPPS.

Rather than total structures, the alternative approach of viewing specific details or regions of large, membrane-bound, functionally competent integral membrane proteins, such as drug or ligand binding regions, is now becoming possible through the use of solid state NMR methods (Watts, 2005). For ligands, detection levels are in the 30 – 300nmole range of binding sites (usually involving ~1 – 30mgs of target protein) and NMR methods have been developed to select out specifically only bound ligands whilst at their site of action, unimpaired by exchange phenomena or solution state, unbound information. Selectivity of information content is achieved by the use of isotopes positioned specifically at places of interest either in the ligand or target where individual amino acid residues can be labelled. NMR visible isotopes such as ^{2}H, ^{13}C, ^{15}N or ^{19}F have all been exploited, yielding different kinds of information depending upon the system or the kind of information required. It could be argued that it is the detail of side chains, residue locations and local structure within a membrane protein which is central to understanding function, detail which can be resolved rather well using solid state NMR. Ligand binding and functional descriptions, such as we have them, invariably involve groups of residues from several secondary structure elements, implying that how these vital residues are located with respect to each other, is of major interest. Here, the ways in which solid state NMR can be used to make contributions to membrane protein structures and ligand interactions with membrane targets will be briefly reviewed, and more comprehensive discussions can be found at (Watts et al. 2004; Watts 2005; Williamson et al. 2004; Hong 2006).

2. SOLID STATE NMR METHODOLOGY

2.1. Sample form

Solid state NMR opens up ways for resolving both specific distance and torsion angle constraints between sites of interest within a protein, or for overall structural determinations. Specific distance measurements depend upon predetermined labelling sites and these are introduced chemically (say

between residues in a peptide, or moieties in a ligand), and can be at isolated or multiple locations. The distance range depends upon the strength of the dipolar coupling, with ^{13}C-pairs potentially giving a distance range of up to 0.7nm at good precision (± <0.03nm) in ideal cases – i.e. rigid systems.

A number of solid state NMR methods have been specifically tailored to determine complete structures of membrane proteins (Watts et al. 2004). Such structures have so far been confined to smaller proteins (usually small transmembrane peptides) or larger proteins in which specific labelling has been possible, although new labelling strategies based on intein technology do open up the possibility of structural determinations of defined domains of larger proteins.

Both oriented and random membrane dispersions have been used for study by solid state NMR (Davis and Auger 1999; de Groot 2000; Griffin 1998; Smith et al. 1996). The choice of solid state NMR sample configuration that can be used on a sample may be constrained by the nature or state of the sample. One approach relies on the use of either mechanically (on plates) or magnetically (in bicelles) oriented samples.

Since membranes are complexes that are anisotropic both functionally and structurally, orientational constraints can be extracted from *macroscopically oriented membranes* by exploiting the anisotropic interactions revealed in solid state NMR spectra. Although the angular information obtained often cannot be used to resolve a complete structure, the orientational constraints resolved can nevertheless be very useful, and are perhaps the only direct information available. In this case, these parameters can be incorporated into models or used with other data to refine still further structural models.

Macroscopically oriented membranes give NMR spectra from which the spectral anisotropy inherent in the chemical shifts, quadrupolar interactions and dipolar couplings, is exploited to yield molecular orientations. Another methodology is to use magic angle spinning (MAS) on an unoriented, random dispersion. In MAS, the sample container (rotor) is set spinning (0.5-12kHz) about an axis which subtends an angle of 54.74°, the magic angle, with respect to the spectrometer field B_0 (Andrew et al. 1959; Lowe 1959). Finally, the combined approach of magic angle oriented sample spinning (MAOSS) first proposed by C. Glaubitz and A. Watts (Glaubitz and Watts 1998), where the sample is oriented and spun slowly (0.5-4kHz) at the magic angle, can be used to obtain orientational information whilst benefiting from the higher spectral resolution of MAS. The various sample geometries used and then the applicable solid state NMR methods generally available have been described elsewhere (Watts et al. 2004). New developments on every front, from protein production and labelling, sample geometry, pulse methods, spectral representation and modelling, are still on-going, and

the number of proteins available is increasing, with little shortage of membrane proteins for study, although getting sufficient amounts and successful labelling potential are still major obstacles.

Model membranes containing proteins or peptides can be made using conventional reconstitution approaches (using dialysis, centrifugation-dilution, co-solubilization followed by solvent removal, for example), which vary depending upon the protein and lipid. Often a fairly rigid environment is required in structural studies (motion defeats spectral anisotropy); therefore saturated lipids (in the gel phase) are preferred (dimyristoyl ($C_{14:0}$), dipalmitoyl ($C_{16:0}$), and distearoyl ($C_{18:0}$)), and bilayer forming zwitterionic phosphatidylcholines are popular. However, it is sometimes necessary to have a mobile lipid environment and to work in the liquid crystalline phase to achieve well-resolved spectra, in particular in oriented samples (Marassi et al. 1997). For proteins with functional characteristics, it may be found that, under some conditions, function is lost in such bilayers, and a judgement needs to be made about whether this is tolerable for NMR studies. Additional restriction of molecular motion can be achieved at low hydration (<100wt% water), which can be obtained by air-drying or incubation in a controlled atmosphere. Since limiting hydration for most lipids is ~30wt%, levels below this are difficult to achieve and maintain (lipids are very hygroscopic).

For sensitivity reasons, few natural membranes contain sufficient protein of one type to be of use in direct structural studies. Some cell membranes, however, like mitochondrial inner-membranes, ATPase or receptor enriched membranes and purple membrane of *Haloarchaea*, are comprised of around 25-75% of one particular membrane protein as a protein-lipid complex. In the rather unusual case of purple membranes in which bacteriorhodopsin is the major (>95%) protein component, the purple membrane is obtained by harvesting *Halobacterium* culture, lysing the cells and purifying using a sucrose gradient (Oesterhelt and Stoeckenius 1974). Another exceptional natural membrane is the disk membranes from retinal rods which contain ~95% rhodopsin, but here, incorporation of labels can only be achieved in a recombinant protein (using bacculovirus or HEK cells) followed by purification and then reconstitution (Creemers et al. 1999; Eilers et al. 1999). In many cases, however, the proteins under study may not be sufficiently abundant. Consequently, they may need to be expressed or synthesized, which is in itself a major hurdle for the amounts (≥ 1-10 mg for ~30kDa) required in structural biology.

2.2. Solid state NMR experimental approaches

The most commonly used method which can be applied to randomly oriented samples is magic angle sample spinning (MAS), whereas static NMR

methods are applied to oriented samples. There are additional solid state NMR experiments which are required when it is possible to assign spectra, and their application to large membrane proteins for total structures is becoming viable often with nanocrystals of protein. For additional information, the reader is referred to recent reviews (Baldus 2002; Davis and Auger 1999; Griffin 1998; Straus et al. 1998).

In magic-angle spinning (MAS) NMR methods, the spatial rotation of the sample introduces time-dependence to the anisotropic spin interactions, such as CSA, quadrupolar interactions, homonuclear dipole-dipole couplings and heteronuclear dipole-dipole couplings, all of which are averaged out more efficiently as the sample spinning frequency increases. Due to the periodic time-dependence of the spin interactions, the broad static lineshape breaks up into a centre band at the isotropic position (ω_{iso}) and a set of spinning sidebands separated by the spinning frequency (ω_r). As the spinning frequency increases, the time averaging is more effective which leads to a decrease in the sideband intensities, and an increase in the centre band intensity. Now, the broad static lineshape, due to anisotropy of the spin interactions, breaks up into a centre band at the isotropic position and a set of spinning sidebands separated by the spinning frequency (< ~4kHz; nucleus dependent). As the spinning frequency increases (~>7-10kHz), the sideband intensities decrease, and an increase in the centre band intensity is observed giving much enhanced spectral intensities and resolution over static samples.

The advantage of MAS is that both resolution and sensitivity are greatly increased. The disadvantage is that the spectrum loses all the molecular geometry information. However, full spectral assignments have been achieved using correlation spectroscopy in MAS experiments of extensively or uniformly ^{13}C, ^{15}N labelled peptide/proteins (Creemers et al. 2002; Straus et al. 1998; Yao and Hong 2001; Yao et al. 2001).

Proton coupling to rare spins complicates the observation of the rare spins, such as ^{13}C and ^{15}N through spectral broadening. It is therefore usually necessary to remove the heteronuclear dipolar couplings between protons and the observed rare spins by introducing strong rf irradiation at the proton resonance frequency, so called *high power proton decoupling*. Continuous-wave (CW) decoupling is a routine scheme for decoupling of the heteronuclear dipolar spin interactions. The rf field induces a fast rotation of the proton spin states averaging out their interaction with the rare spins (Mehring 1983). For fast MAS, more sophisticated pulse sequences (such as TPPM, C12, R24 and XiX) achieve a better decoupling efficiency (Bennett et al. 1995; Carravetta et al. 2000; Detken et al. 2002; Eden and Levitt 1999).

In another exploitation of high γ nuclei, the sensitivity of rare spin species (*S*-spins) with low γ, such as ^{13}C and ^{15}N, can be enhanced by transferring magnetization from abundant spin species (*I*-spins) with high γ, such as ^{1}H,

to the S-spins. Cross-polarization (CP) is the most widely used method to transfer polarization between unlike spin species through the heteronuclear dipolar couplings (Pines et al. 1973). Magnetization transfer is achieved when the strengths of the two fields match the Hartmann-Hahn (HH) condition $|\gamma_I B_{1I}| = |\gamma_S B_{1S}|$, where γ_I and γ_S are the magnetogyric ratios of the *I*-spins and *S*-spins, respectively (Hartmann and Hahn 1962). The enhancement of the *S*-spin magnetization is roughly proportional to the ratio of the two magnetogyric ratios $|\gamma_I|/|\gamma_S|$. Methods have also been developed where one of the rf fields is ramped in a linear fashion (Hediger et al. 1994; Metz et al. 1994; Peersen and Smith 1993) or adiabatically (Hediger et al. 1994). These methods can improve the stability and reproducibility of HH-CP, especially under fast MAS conditions.

Fast MAS leads to high resolution and sensitivity in solid-state NMR, which is the basic requirement for the sequential assignment of protein spectra. However, all anisotropic spin interactions which can be used to extract molecular geometry information are, in principle, averaged out by fast MAS. Therefore, in order to recover the anisotropic spin interactions in the presence of MAS, a range of recoupling techniques have been developed. Generally, there are two approaches to reintroduce anisotropic dipolar interactions, either mechanically, where the recoupling is achieved through sample rotation, or by rf pulse driven methods where the recoupling is achieved by applying rf pulse trains. Thus, magic-angle spinning NMR recoupling techniques have been extensively used to give a selective restoration of informative anisotropic nuclear spin interactions which, in turn, can yield structural information about distance and orientational constraints (Bennett et al. 1994; Dusold and Sebald. 2000).

Internuclear distance measurements can be made from homonuclear recoupling approaches, where rotational Resonance (R^2) can be induced when the sample spinning frequency ω_r matches the isotropic chemical shift difference ($\Delta\omega$) of a pair of coupled homonuclear spin S_j and S_k under the condition

$$\Delta\omega = \left|\omega_j^{iso} - \omega_k^{iso}\right| = n\omega_r$$

where *n* is a small integer (Levitt et al. 1990; Peersen and Smith 1993; Raleigh et al. 1988). Under this condition, the homonuclear dipole-dipole interaction is recoupled which leads to spectral broadening and splitting as well as an exchange of longitudinal magnetisation between the two spins in a two-dimensional experiment (Colombo et al. 1988; Levitt et al. 1990; Raleigh et al. 1988). Magnetisation transfer rates are much higher for doubly-labelled samples containing closely coupled spin pairs compared to those with distant spins. From the change in intensity of the two labelled sites, the dipolar coupling constant (b_{jk}) can be extracted by numerical

simulation for the best fit of the experimental data (Levitt et al. 1990; Peersen et al. 1995), and then lead to a distance determination. R^2 and modified R^2 methods have been applied extensively to structural determination of biological membranes (Ahmed et al. 2000; Feng et al. 1999; Glaubitz et al. 2000; Lam et al. 2002; Lam et al. 2001; Middleton et al. 2000; Middleton et al. 1997; Nomura et al. 1999; Peersen and Smith 1993; Smith et al. 1995), by measuring inter- and intramolecular distances of membrane bound proteins and peptide-inhibitor complexes since the dipolar coupling is $\alpha 1/r^3$.

The rotational echo double resonance (REDOR) experiment has been introduced by Gullion and Schaefer in 1989 (Gullion 1998; Gullion and Schaefer 1989a; Gullion and Schaefer 1989b) as a method for heteronuclear distance determination. Since then it has proven its robustness and applicability to biological systems in numerous applications for up to 1MDa protein complexes (Goetz et al. 1999). The aim of the REDOR experiment is to prevent the heteronuclear dipolar coupling from being averaged by magic angle spinning by rotor synchronized rf pulses. By fitting the data with calculated 'REDOR-curves' a dipolar coupling and hence a distance can be determined. As $\Delta S/S_0$ depends on the dipolar coupling in the form of the product $N\ T_R\ b_{IS}$, the dephasing follows a 'universal' curve, representing an identical behaviour which depends on the dipolar coupling only through scaling along the abscissa. A number of experimental improvements have been introduced to the basic REDOR experiment and have lead to a variety of REDOR derivatives. Also, the use of composite pulses has been shown to increase the performance of dephasing (Raleigh et al. 1988; Sack et al. 1999). It is thus possible to adapt the pulse sequence to sample or hardware specific circumstances and, for example, apply most of the pulses on the channel with the better inversion efficiency.

Typically the REDOR experiment is used for distance measurements. It can, however, also provide orientational information. If the MAS frequency chosen is sufficiently slow, REDOR dephasing will affect the individual sidebands differently. While the integral intensity of the difference spectrum S_0-S depends only on the dipolar coupling strength, the sideband pattern of the difference spectrum then also reflects the relative orientation of the chemical shift anisotropy tensor (or quadrupolar interaction tensor) of the observed spin and the internuclear vector. The analysis of the spinning sidebands thus provides additional structural information in addition to the internuclear distance (Goetz and Schaefer 1997) and in membranes (Glaubitz et al. 2000).

Torsion angle constraints may be obtained by estimating the relative orientation of two anisotropic spin interactions. Generally, there are two approaches used to determine torsion angles. The first approach is called

two-dimensional (2D) exchange NMR spectroscopy. The second approach is based on multiple-quantum (MQ) coherence excitation and both exploit anisotropic spin interactions which are sensitive to their relative orientation in the system (Bower et al. 1999; Costa et al. 1997; Feng et al. 1997; Feng et al. 1996; Feng et al. 2000; Feng and Gregor 1997; Gregory et al. 1997; Hong et al. 1997a; Hong et al. 1997b; Huster et al. 2000a; Huster et al. 2000b; Schmidt-Rohr 1996). In either way, estimates of the torsion angles ϕ and ψ for adjacent peptide bonds or neighbouring nuclei in ligands can be made to within ± ~10°.

By aligning the sample on a solid support which has itself a specific direction with respect to the external magnetic field, orientationally dependent parameters such as the chemical shift, quadrupolar interactions or the dipolar coupling can be measured for each labelled position, such as a peptide plane in the membrane protein. Thus, solid state NMR experiments of oriented systems can be used to resolve both local and global structures from an observation of the spectral anisotropy.

The polarization inversion spin exchange at the magic angle (PISEMA) (Ramamoorthy et al. 1999) experiment yields a two-dimensional correlation of the chemical shift of spin S ($S = {}^{13}C, {}^{15}N$) with the heteronuclear IS ($I={}^{1}H$) dipolar coupling. Structural information is obtained by relating experimentally determined chemical shift/dipolar splitting pairs for a given peptide plane to four possible solutions of (α_{PL}, β_{PL}). These four solutions arise because of the symmetry of the interactions. By finding pairs of angles for all peptide planes it was possible, for instance, to solve for a family of structures for gramicidin A (Kim et al. 2001) and M2 from the nicotinic acetylcholine receptor (Opella et al. 1999).

The global arrangement of helices can be estimated by fitting experimentally determined chemical shift/dipolar splitting pairs with calculated parameters for given (α_{PL}, β_{PL}) angles for a set of peptide planes. In order to relate (α_{PL}, β_{PL}) for each peptide plane to a global orientation of the helix relative to the external magnetic field, a specific geometry must be defined. One possibility is to assume that the helix is ideal, i.e. that the (Φ, Ψ) torsion angles are identical throughout (Marassi and Opella 2000; Wang et al. 2000). The implications of this choice of torsion angles and other parameters, including motion, used to obtained the calculated chemical shift/dipolar splittings have been recently investigated in great detail (Fares et al. 2002; Straus et al. 2003). For the regular α-helix structure in uniformly labelled single-pass peptides with ^{15}N in the backbone, the PISEMA experiment leads to wheel like patterns called "PISA (Polarization Index Slant Angle) wheels" (Marassi and Opella 2000; Wang et al. 2000). From such wheels it is possible to determine the secondary structure, the tilt and rotational angles of α-helices or β-sheets with respect to the magnetic field, and in favorable

cases sequential assignment of all resonances on the basis of a single unique resonance assignment (Denny et al. 2001; Marassi 2001).

In order for an NMR experiment to be useful when applied to membrane peptides and proteins, it is important that the method yields well resolved spectra, and the PISEMA experiment offers comparatively high resolution. Line narrowing is achieved by incorporating a proton homonuclear decoupling sequence during the *IS* dipolar evolution in t_1 which eliminates the contribution of the homonuclear dipolar interaction.

3. MEMBRANE PROTEIN STRUCTURES DETERMINED BY SOLID STATE NMR

3.1. Single pass peptides

Single transmembrane domain peptides are, as a class of protein, the most abundant in number in many life-forms, as predicted from all known genomes (Arkin et al. 1997). Although many will be leader sequences used in processing and targeting to a specific location, single TMD peptides do form an important class of proteins functioning alone or through interactions with similar or identical units to form complexes such as ion channels, or with larger proteins and acting as regulators of function. It is generally assumed that single TMD peptides adopt an α-helical structure in the membrane. Even if this is the general rule, kinks and helical tilts do have a significant bearing on function and mechanism - solid state NMR methods which exploit spectral anisotropy and give distance measurements can make significant contributions to this knowledge base, helping functional descriptions.

Complete structures by solid state NMR of a number of trans-membrane peptides, small single pass proteins or trans-membrane helices of larger membrane proteins have recently been elucidated. Examples deposited in the Protein Data Bank (PDB) include gramicidin A, the M2 segment of the nicotinic acetylcholine receptor and of the proton channel of influenza, and the fd bacteriophage coat protein. Most of the structural information for these systems was obtained from orientational constraints from oriented samples of uniformly ^{15}N-labelled peptides or proteins (Marassi and Opella 2000).

Viral ion channels have been a focus of solid state NMR structure determinations. The structure of the M2 trans-membrane segment of the influenza H^+ channel has been obtained from solid state NMR (Wang et al. 2001). It was found to be formed by four almost ideal α-helices, tilted at

approximately 25° with respect to the membrane normal (Tian et al. 2002). A histidine residue pointing towards the centre of the M2 bundle plays a crucial role in the pH dependent gating of the proton channel. Extensive use was made of the PISEMA experiment to provide the necessary orientational constraints. However, to identify intermolecular connectivities and hence the oligomeric state of the bundle, and to determine the position of functionally important side-chains, these orientational studies were combined with REDOR distance measurements (Nishimura et al. 2002).

A similar strategy was adopted in the study of the nicotinic acetylcholine receptor, where a transmembrane segment (M2) was investigated individually (Montal and Opella 2002). It was found to form pentamers even without the context of the entire protein, thus enabling insight into the structure on the basis of a smaller peptide. A funnel-like structure, formed by a bundle of straight helices tilted at 12° with respect to the membrane normal was determined this way using, again oriented samples and the PISEMA approach. Several other examples are now available and the method is becoming more accessible (Watts et al. 2004).

3.2. Polytopic proteins

Polytopic membrane proteins have more than one transmembrane domain (TMD) connected by loops which themselves may have structural elements. Examples of single chain and multi-complex proteins exist, although the tractability of total structures for polytopic proteins is hampered by assignment and spectral complexity for uniformly ^{15}N labelled proteins, as shown from simulations of PISEMA spectra for rhodopsin and porin (Fig. 3.1). Thus, most approaches are designed to simplify the information content or focus on one part, such a prosthetic group structure and its binding site, or through the use of selective labelling.

3.2.1. Bacteriorhodopsin

Bacteriorhodopsin (bR) is one of the most comprehensively studied proteins due largely to its stability, ease of production with, specifically in NMR studies, selective isotope labelling. It works as a light-driven proton pump of *Haloarchaea* that converts light energy into an electrochemical proton gradient from the cytoplasmic to the extracellular side. This process is driven by photoisomerization of chromophore retinal from all-*trans* to the 13-*cis*, 15-*anti* configuration. This proton transport process is cyclic and some residues in the proton pathway are protonated/deprotonated as the counter ions. A number of high-resolution X-ray crystal structures of bR have been

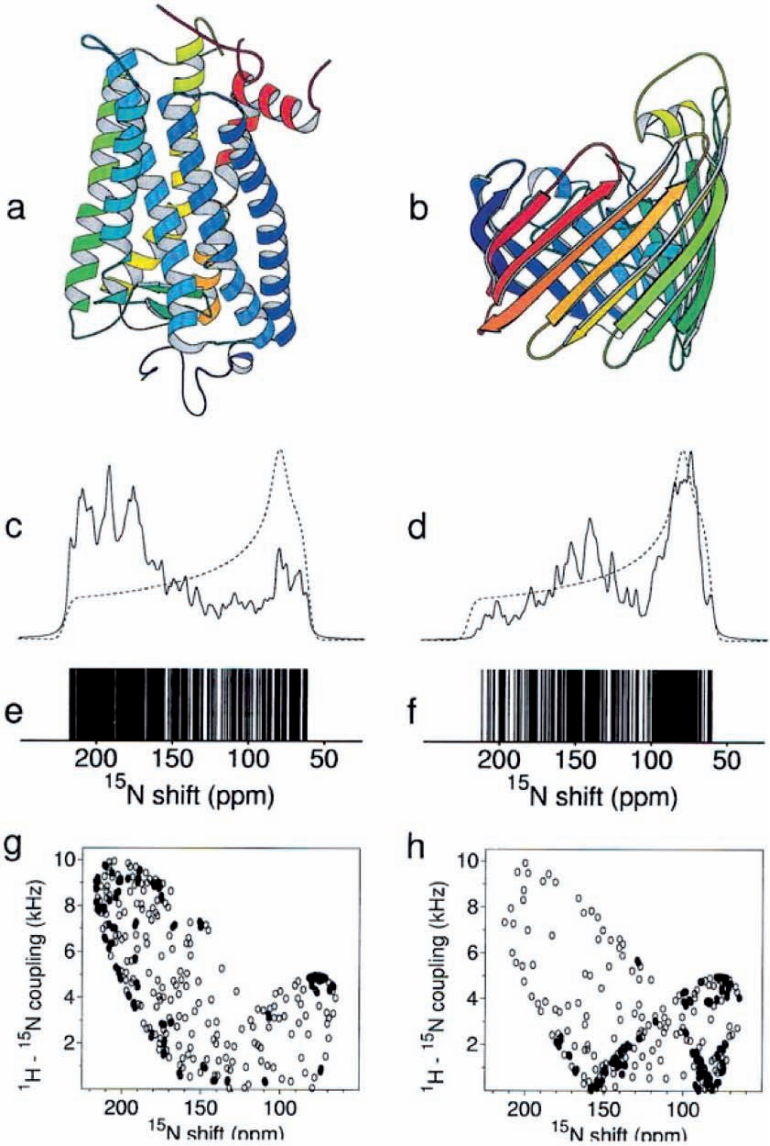


Figure 3.1. SIMPSON simulations of one-dimensional ^{15}N spectra and ^{15}N chemical shift/^{15}N -^{1}H dipolar PISEMA spectra for two membrane proteins under ideal conditions in macroscopically oriented membranes with their average α-helix and β-strand axes parallel to the external field to show the expected complexity of the spectra. Molecular structures of bovine rhodopsin (1FB8) (A) and the porin OmpA (2POR) (B) and their simulated 1D spectra (solid line) and powder spectra (dashed line). Stick plots of the resonance spectral line positions (E, F) and from the simulated spectra (C, D), and the simulated PISEMA spectra (G, H) with open and closed ellipses representing resolved and unresolved resonances, respectively. Adapted from (Vosegaard and Nielsen 2002).

determined at up to 1.43 Å resolution for the ground state, and of some functional photo-intermediates (Edman et al. 1999; Faham and Bowie 2002; Lanyi and Schobert 2002; Luecke et al. 1999; Royant et al. 2001; Sass et al. 2000; Schobert et al. 2002). However, solid state NMR studies have given complementary information such as accurate conformational determinations of local active sites, not only for the ground state but also some other photo-intermediates trapped by adjusting the temperature, pH or using mutants (Herzfeld and Lansing 2002).

Since bR forms a two-dimensional hexagonal crystalline lattice arranged in trimeric units in the purple membrane (PM), it orients well on glass plates (Fitter et al. 1999; Seiff et al. 1985). Bacteriorhodopsin oriented on glass plates preserves the directional quality of membrane protein structure and also maintains the structural and functional integrity of the protein under a wide range of pH, temperature, humidity, or chemical environments (Oesterhelt et al. 1991). Wide line ^{2}H NMR on oriented PM has been used to determine the orientation of a specifically labelled molecular segment relative to the magnetic field (B_0), that is parallel to the membrane normal when the glass plates oriented perpendicular to B_0. From the splittings of the quadrupole couplings in the ^{2}H NMR spectra, the C-CD_3 angle with respect to the magnetic field was measured. Specifically, deuterium-labelled at C16 and 17 methyls on the cyclohexene ring incorporated into the oriented protein were found to be oriented at angles of 94±2° and 75± 2° with respect the membrane normal in the ground state (Ulrich et al. 1992). Furthermore, results from deuterated methyl groups at C18, C19 and C20 showed that the C-CD_3 bond vectors were oriented at 37±1°, 40±1° and 32±1° respectively (Ulrich et al. 1994). The M intermediate trapped using guanidium hydrocholoride at –60°C showed the angle between the C-CD_3 bond and the magnetic field at the C19 methyl group changed to 44±2°, which is consistent with a slight upward tilting of the polyene chain. Additional line broadening compared with that of the ground state suggested some two state heterogeneity (Ulrich et al. 1995). ^{2}H NMR also revealed that the C9-CD_3 bond changes by 7° from the ground to the M state (Moltke et al. 1999). Magic angle spinning has also successfully applied to the study of purple membrane pellet to determine local conformation around the labelled positions in bR. Rotational resonance (R^2) was used to determine the distances of 4.1±<0.4 Å from the C8-C18 positions and 3.3-3.5±<0.4 Å for the average C8-C16/C8-C17 positions for selectively labelled retinal in the ground state, showing that retinal is in a 6-s-*trans* conformation (McDermott et al. 1994). The same approach was also used to determine the distance between ^{13}C14 in retinal and the ^{13}C-Lys216, which was found to be 4.1±0.3 Å in bR_{568} and 3.9±0.1 Å in the thermally trapped M_{412} state. This demonstrated that the C = N bond is *anti* in these photostates. The distance of 3.0±0.2 Å between

these same two groups in bR_{555} showed a change to a C = N *syn* bond in this state (Lakshimi et al. 1993; Thompson et al. 1992).

For structural studies of bR, one type of amino acid can be labelled selectively. How ever since bR has 29 alanines, 21 valines and 9 methionine residues and many are located in helices, the NMR signals may be overlapped because of similar environments and orientations.

^{15}N spectra are also sensitive to conformational changes, although the isotropic chemical shifts are weakly dependent on local secondary structure because the amide nitrogen chemical shift is affected by the neighbouring amino acids and existence or strength of hydrogen bonds (Shoji et al. 1990). Recently, assignments have been reported of ^{15}N CP-MAS spectra of [^{15}N]Met-bR. This was achieved by resonance knockouts using mutants and by substitution of easily exchangeable protons to deuterium located in residues located in loops and at the surface of the protein (Mason et al. 2003). Further

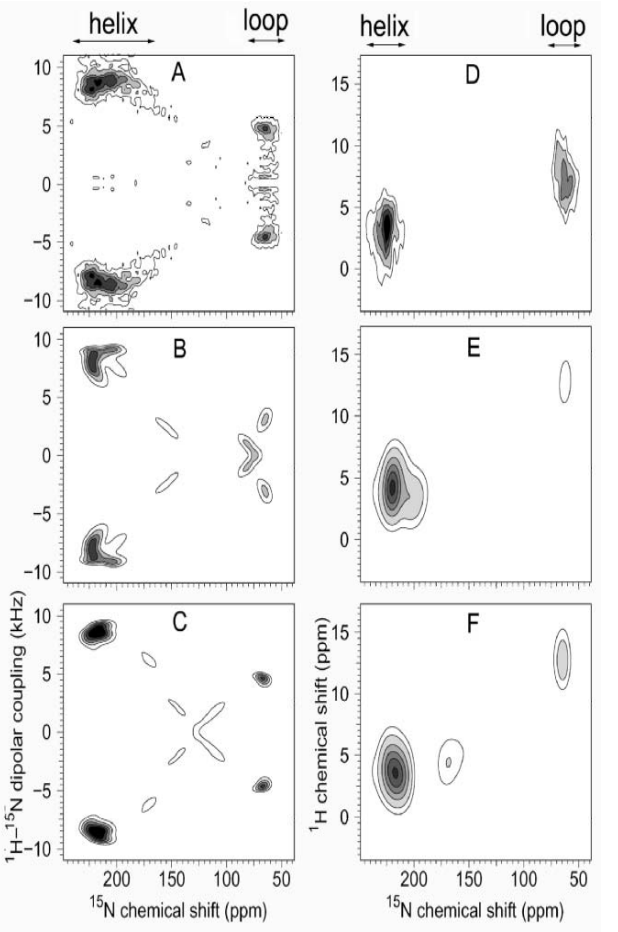


Figure 3.2. Experimental PISEMA (A) and ^{15}N-^1H HETCOR (D) spectra of [^{15}N-Met]-bacteriorhodopsin. Simulated PISEMA (B) and simulated HETCOR (E) spectra (which are half the experimental symmetrical spectra), using SIMPSON/SIMMOL and structural parameters from the x-ray structure 1C3W.pdb. Simulated PISEMA (C) and simulated HETCOR (F) spectra using the structural parameters for the membrane bound protein in 1FBB.pdb. Resonances from helix and loop positioned labelled residues are well resolved and show that the NMR spectra are rather sensitive to the local protein structure (Kamihira et al. 2005).

examination of 2D HETCOR MAS spectra showed some further signal separation in the 1D spectrum (Mason et al. 2003), whereas PISEMA and ^{1}H-^{15}N HETCOR and MAOSS methods have been applied to uniaxially oriented PMs to determine structural and orientational constraints of some of the Met residues (Kamihira et al. 2005; Mason et al. 2003). The spectra allow clear distinction of helix and loop resonances (Fig. 3.2A,D) and using SIMPSON (Bak et al. 2000) and SIMMOL (Bak et al. 2002) simulation software the helix signal was deconvoluted to provide information about helix tilt angles and structural constraints. Using a combination of experimental and simulated spectra, it was estimated that the extracellular section of helix B has a tilt of less than 5° from the membrane normal, in agreement with most crystal structures. Furthermore, the experimental solid-state NMR spectra were directly compared with recent X-ray and electron crystal structures, 1C3W (Luecke et al. 1999) (Fig. 3.2B,E) and 1FBB (Subramaniam and Henderson 2000) (Fig. 3.2C,F) respectively, showing that the NMR spectral methods are extremely sensitive to local conformations and crystal differences (Kamihira et al. 2005).

3.3. Ligand-protein interactions

3.3.1. Orientation and dynamics of a channel agonist and blocker

Examples of agonist and blocking ligands within ion channels have been studied using solid state NMR methods and natural membranes. The mechanisms whereby channels are both activated and blocked are of potential drug design importance since 50 – 85% of all drug targets for the next decade are membrane proteins, and many of these are channels (Terstappen and Reggiani 2001). The new information being gained about channels from various methods, is now helping in mechanistic descriptions, and hence the potential for designing new ligands, either for new means of activation or of inhibition.

3.3.2. Agonist binding to a ligand-gated ion channel

The nicotinic acetylcholine receptor is a ligand-gated ion channel (~290k) for which acetylcholine is the agonist. It is a pentameric ($\alpha_2\beta\delta\gamma$) protein with the agonist binding pocket in the two α-subunits (Unwin 2005). The receptor is a major component of the electric organ of electric fish such as *Torpedo marmorata*, from which enriched membranes containing ~30 – 35% of the total protein is the receptor. These membranes are used extensively for pharmacological characterisations is a model for the mammalian receptor.

Electron microscopy methods are being used for structural resolution of the receptor to ~4 – 5 Å (Unwin 2005) and recently a high-resolution crystal structure of a soluble bacterial acetylcholine binding protein was resolved using x-ray diffraction (Brejc et al. 2001). Activation by the agonist at a site some 60-70 Å above the membrane surface produces a significant structural rearrangement (rotation) of the five M2 trans-membrane segments forming the pore, to permit rapid Na^+ ion translocation though the channel and synapse activation. Simple chemical changes within the quaternary ammonium of the agnostic cause major reduction in activation, whereas relatively major acetyl modifications can be used to probe the agonist-binding site, although it might be expected that these details might be perturbed by such modification(s).

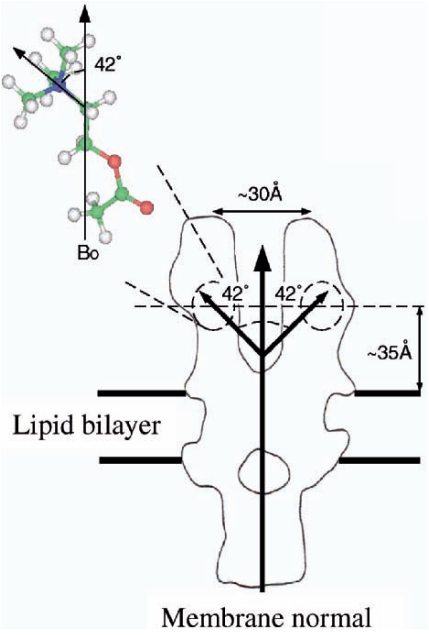


Figure 3.3. The solid state NMR derived orientation of the acetylcholine whilst in its binding site in the membrane bound receptor. Although the agonist is constrained in its site while resident, the trimethylammonium group undergoes rapid rotation around its long axis whilst in a cation-π interaction with an aromatic binding pocket.

As a direct probe of the agonist binding site, we have synthesized bromoacetylcholine with ^2H and acetylcholine with ^{13}C in the $-N^+(CH_3)_3$ group, to probe directly the binding site (Williamson et al. 1998; Williamson et al. 2001). By exploiting the quadrupolar anisotropy of deuterium in receptor rich membranes oriented on glass slides, the orientation of the binding site has been determined to be 42±5° with respect to the membrane normal (Fig. 3.3). Although this information does not contain details about location of the binding site within the protein, it is not inconsistent with current views of the orientation of the agonist-binding site from spray freezing trapping EM studies. Whilst at the binding site, the trimethyl group undergoes fast (<μs) motion around the molecular long axis even at rather low (–60°C) temperatures, as revealed by the deuterium spectral line-widths.

In earlier studies we showed the selective binding and detection of ~30 – 40nmoles of ligand binding

sites in acetylcholine receptor-rich membranes (Williamson et al. 1998). In that work, the constrained agonist was observed using delayed dephasing, cross polarisation to detect only bound agonist in this freely exchanging ligand-receptor system. This approach selects out constrained ligands from any in solution, although the conditions were set such that the amount of free ligand was small and agonist hydrolysis prevented by addition of FPP. The direct observation of such a freely exchanging ligand in a receptor is novel and the results showed that the environment within the binding site is commensurate with an aromatic binding pocket (rather than an $-N^{+}(CH_3)_{3-}$ anion) interaction. This result agrees with later studies that suggest that Trp149 of the α-subunit is the aromatic component of the interaction (Zhong et al. 1998), in which case the $-N^{+}(^{13}CH_3)_3$ group in the non-perturbing ^{13}C NMR spin-probe makes van der Waals contact with the six-membered ring of the indole sidechain and influences the measured chemical shift (Williamson et al. 1998). Also, correlations between *ab initio* quantum mechanical predictions of cation binding abilities and of inhibition determinations for acetylcholine at the receptor for a series of tryptophan derivatives mutants, confirm further the cation-π interaction (Zhong et al. 1998).

The mechanism of cation-π interactions in initiating an activation process must clearly be capable of mobilising sufficient energy, in this case to effect the rotation of the M2 helices, to open the channel. This energy might be already inherently bound up in the system and released through agonist binding, or generated directly from ligand binding itself. In bacterial chemoreceptors, for example, ligand binding is followed by major conformational changes triggering protein-protein contacts some 400 Å distant from the site of ligand binding, deep in the cell. How this, and other similarly major conformational changes can occur following small ligand binding, is not well described for any membrane bound receptor to date, but presumably such dramatic rearrangements would require some multiples of kT to drive these conformational changes. Cation-π interactions of several (3 – 10) kcal/mole stabilise protein structure through side-chain interactions (Gallivan and Dougherty 1999), although in these cases the cumulative interactions are important for protein stability rather than any one single binding. This does, however, suggest that rearrangements locally within a binding site can occur when a new component is introduced, but quantification of this is necessary if the mechanism is to be understood fully.

3.3.3. Inhibition of ion translocation

The primary P-type ion translocators coupled to ATPase hydrolysis are large, ubiquitous and membrane bound proteins. The high resolution,

crystallographic structure of the Ca^{++}-Mg^{++}-ATPase is now available (Toyoshima et al. 2000) and is proving useful for modelling the disposition of the transmembrane segments, on the assumption that primary sequence directed structural modelling is applicable and valid for the other ATPases. All known ATPases are well-characterised biochemically and pharmacologically, and so there is a wide range of ligands, including reversible and irreversible inhibitors.

3.3.4. Irreversible inhibition of the Na^+,K^+-ATPase by Ouabain

Digitalis (ouabain, a cardiac glycoside), is an extract of the flower fox-glove, and is probably one of the best and most ancient of all known poisons. It is an irreversible inhibitor of the Na^+/K^+-ATPase which is ubiquitous in cell membranes, with a 1:1 molar stoichiometric binding ratio and k_D's which are in the < nmolar range.

Firstly, the distance between nuclei positioned in the sugar and in steroid ring using ^{19}F (dephase) to ^{13}C (observe, heteronuclear dipolar recoupling (REDOR) has been determined at ~9 Å when at the ATPase binding site (one ligand/monomer). This then fixes the orientation of the two moieties at close to right angles to each other. Secondly, ^{13}C labelling was carried out to reveal a significant chemical shift in the spectral positions for the rhamnose labelled, but not steroid labelled, part of ouabain. This implies some electronic interaction with a specific part of the protein, possibly through the Tyr108 which is located in the putative transmembrane region of the H1 according to topology mapping (Shull et al. 1985). Lastly, deuterium was used as a dynamic sensor for the bound ouabain, showing a differential motional freedom for each part of the ligand, with the sugar being rather mobile and the steroid being significantly more constrained at 5°C in natural membranes. Taken together with the distance and chemical shift constraints, and site directed mutagenisis studies of function, this dynamic information leads to a model for the binding of the oubain on the protein, with a helix disposition (from the Ca^{++}-ATPase structure) and size of the ligand incorporated into the model.

3.3.5. Reversible inhibition of the H^+/K^+-ATPase interaction with substituted imidazo-pyridines

A range of reversible substituted imidazole pyridines (SIPS) and irreversible inhibitors are available as a result of the need to design both types of inhibitors for potential treatment of peptic ulcers, and indeed, significant revenues are generate through the use of these world-wide. We have labelled a series of

SIPS with NMR-visible isotopes and used them in investigations of bound ligand structure. In these studies, both homo- and heteronuclear dipolar recoupling methods have been used to determine internuclear distances to good precision. For these kinds of ligands, the two planar parts are linked with a single torsion angle across the ether bond. In solution, presumably all excursions of the two planar systems with respect to each other are explored. It was found early on from precise interatomic distance measurements by solid state NMR that the drug is almost planar when it occupies the binding site (Fig. 3.4), with both rings in a very similar plane and hence flat (Middleton et al. 1997). This is rather surprising since it might be expected that conformational energy is required to induce conformational changes within the protein. The tacit implication here then is that energetics of binding and interaction with the protein are more subtle than simply conformationally driven.

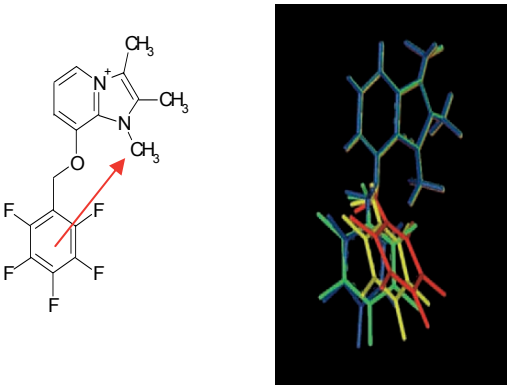


Figure 3.4. The distance between the fused rings (arrow) has been determined to be 6.10 – 5.81 Å from solid state NMR heteronuclear (^{19}F dephase, ^{13}C observe) dipolar recoupling methods for labelled inhibitor when at its site of action in the H^+,K^+-ATPase in gastric membranes in which ~35% (~250nmoles) of the total protein is the target. The possible excursions for the pyridine ring are shown (right) within theses distance constraints for an essentially planar ligand whilst at the target site of action.

NMR line-widths imply a relatively fixed environment of the ligand with respect to the protein, and chemical shift changes (when compared with an aqueous environment) suggest electronic interactions play a part of the binding process. Indeed, this class of molecule has now been modelled into the binding site through π-π ring stacking involving possibly Phe126 (Munson and Sachs 1988), or other local aromatic side chain. It is unclear whether this stacking process has the resultant effect of initiating the subsequent changes, presumably conformational as suggested by spin label ESR studies (Middleton et al. 1995), which lead to blocking of ion translocation.

3.3.6. Ligand activation of 7TMD, GPCRs

The photoreceptor protein, rhodopsin, is probably the most intensely studied seven transmembrane domain (7TMD), GTP-coupled receptor (GPCR).

Although not a conventional 7TMD, GPCR in that it possesses a prosthetic group, retinal, it has provided some indications about structure and secondary element reorganisation induced upon activation, in this case by light. Direct and indirect studies have revealed exposure of loops to activate transducin although major changes in helical content do not occur. These studies have been made in the absence of the crystal structure of rhodopsin which was resolved only recently for the dark-adapted state. Although many aspects of function are being discovered, the generality and relevance of the activation process and subsequent structural changes for other ligand binding GPCRs, is not yet fully known.

Rhodopsin has been studied extensively by solid state NMR methods to resolve, primarily, the retinal structure and changes upon activation, whilst in membranes (Feng et al. 1997; Gröbner et al. 2000; Smith et al. 1990). Chemical shifts and orientational parameters have been resolved to identify both intramolecular distances and orientation of the retinal at its binding site. The *cis-trans* nature of the β-ionone ring has been the source of much discussion. Although high enough resolution crystal structures are not yet available (Palczewski et al. 2000), it has been shown that ring flipping may occur (Singh et al. 2001) and the promiscuity of the retinal binding site around the ring does permit extensive movement of the chromophore during activation (Borhan et al. 2000). Indeed, a mixed population (74%:26%, with 74% being *cis*) of different structures of retinal may exist in retinal membranes (Spooner et al. 2002), but not in reconstituted saturated lipid bilayers (Gröbner et al. 2000). Importantly, some 36kcals.mole^{-1} of energy are released upon photon incidence on retinal, enough to then alter the protein structure sufficiently to lead to transducin activation. This is unlike other GPCRs where ligand binding precedes activation.

The neurotensin receptor is activated by the 13-mer peptide, neurotensin, with the final six residues (Arg-Arg-Pro-Tyr-Ile-Leu) having activation capability similar to the full-length peptide. No solution structure of either neurotensin or the hexamer have been found or known to exist, but it is assumed that some structural requirements exist for the bound ligand. The active receptor has been expressed as a recombinant protein in *E. coli* and the ligand bound to the receptor studied. By incorporating NMR-visible isotopes into this hexamer peptide, it has been possible to resolve which of the residues, and which parts of residues, are vital for binding to the receptor (Williamson et al. 2002). As expected, chemical shifts occur for most sites within the agonist, especially the carboxy terminus and the Tyr residue. Although the binding site residues are not known, heuristic studies driven by bioinformatics, and site-directed mutagenisis studies have suggested the

involvement of a significant number of aromatic residues in the putative Helix 3 – 4 loop region (Pang et al. 1996). Also, a direct determination of the full structure of the ligand has not yet been possible, but some suggestions that this might take a Type 1 turn when bound seems likely. Solid state NMR seems one of the direct methods for resolving this important information. Additionally, the significant chemical shifts observed for bound agonist give strong indications that electronic interactions play a major role in the activation process, again something which NMR is well suited to resolving.

ACKNOWLEDGEMENTS

The Medical Research Council (UK), BBSRC, CLRC, Bionanotechnology IRC and EPSRC are thanked for their support for grants to AW, and HEFCE, Magnex and Varian for equipment support.

REFERENCES

Ahmed, Z., D.G. Reid, A. Watts, and D.A. Middleton. (2000) A solid-state NMR study of the phospholamban transmembrane domain: local structure and interactions with Ca(2+)-ATPase. *Biochim. Biophys. Acta* **1468**(1-2): 187-198.

Andrew, E.R., A. Bradbury, and R.G. Eades. (1959) Removal of dipolar broadening of Nuclear Magnetic Resonance Spectra of Solids by Specimen Rotation. *Nature* **183**(4678): 1802-1803.

Arkin, I.T., A.T. Brünger, and D.M. Engelman. (1997) Are there dominant membrane protein families with a given number of helices? *Protein: Structure, Function and Genetics* **28**: 465-466.

Bak, M., J.T. Rasmussen, and N.C. Nielsen. (2000) SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Res.* **147**: 296-330.

Bak, M., R. Schultz, T. Vosegaard, and N.C. Nielsen. (2002) Specification and visualization of anisotropic interaction tensors in polypeptides and numerical simulations in biological solid-state NMR. *J. Magn. Res.* **154**(1): 28-45.

Baldus, M. (2002) Correlation experiments for assignment and structure elucidation of immobilized polypeptides under magic angle spinning. *Progress in Nuclear Magn. Res. Spectros.* **41**: 1-47.

Bennett, A.E., L.R. Becerra, and R.G. Griffin. (1994) Frequency-selective heteronuclear recoupling in rotating solids. *J. Chem. Phys.* **100**(2): 812-814.

Bennett, A.E., C.M. Rienstra, M. Auger, K.V. Lakshmi, and R.G. Griffin. (1995) Heteronuclear Decoupling in Rotating Solids. *J. Chem. Phys.* **103**(16): 6951-6958.

Borhan, B., M.L. Souto, H. Imai, Y. Shichida, and K. Nakanishi. (2000) Movement of Retinal Along the Visual Transduction Path. *Science* **288**: 2209-2212.

Bower, P.V., N. Oyler, M.A. Mehta, J.R. Long, P.S. Stayton, and G.P. Drobny. (1999) Determination of torsion angles in proteins and peptides using solid state NMR. *J. Amer. Chem, Soc.* **121**(36): 8373-8375.

Brejc, K., W.J.v. Dijk, R.V. Klaassen, M. Schuurmans, J.v.d. Oost, A.B. Smit, and T.K. Sixma. (2001) Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**: 269-276.

Carravetta, M., M. Eden, X. Zhao, A. Brinkmann, and M.H. Levitt. (2000) Symmetry principles for the design of radiofrequency pulse sequences in the nuclear magnetic resonance of rotating solids. *Chem. Phys. Lett.* **321**: 205-215.

Castellani, F., B.v. Rossum, A. Diehl, M. Schubert, K. Rehbein, and H. Oschkinat. (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**: 98.

Colombo, M.G., B.H. Meier, and R.R. Ernst. (1988) Rotor-driven spin diffusion in natural abundance 13C spin systems. *Chem. Phys. Lett.* **146**: 189-196.

Costa, P.R., J.D. Gross, M. Hong, and R.G. Griffin. (1997) Solid-state NMR measurement of Psi in peptides: a NCCN 2Q-heteronuclear local field experiment. *Chem. Phys. Lett.* **280**(1-2): 95-103.

Creemers, A.F., S. Kiihne, P.H. Bovee-Geurts, W.J. DeGrip, J. Lugtenburg, and H.J. de Groot. (2002) (1)H and (13)C MAS NMR evidence for pronounced ligand-protein interactions involving the ionone ring of the retinylidene chromophore in rhodopsin. *Proc. Nat. Acad. Sci. USA* **99**(14): 9101-9106.

Creemers, A.F.L., C.H.W. Klaassen, P.H.M. Bovee-Geurts, R. Kelle, U. Kragl, J. Raap, W.J. de Grip, J. Lugtenburg, and H.J.M. de Groot. (1999) Solid State ^{15}N NMR Evidence for a Complex Schiff Base Counterion in the Visual G-Protein-Coupled Receptor Rhodopsin. *Biochemistry* **38**(22): 7195-7199.

Davis, J.H., and M. Auger. (1999) Static and magic angle spinning NMR of membrane peptides and proteins. *Progress Nucl. Magn. Res. Spectros.* **35**: 1-84.

de Groot, H.J. (2000) Solid-State NMR Spectroscopy Applied to Membrane Proteins. *Curr. Opin. Biotechnol.* **10**: 593-600.

Denny, J.K., J. Wang, T.A. Cross, and J.R. Quine. (2001) PISEMA powder patterns and PISA wheels. *J. Magn. Res.* **152**(2): 217-226.

Detken, A., E.H. Hardy, M. Ernst, and B.H. Meier. (2002) Simple and efficient decoupling in magic-angle spinning solid-state NMR: the XiX scheme. *Chem. Phys. Lett.* **356**(3-4): 298-304.

Dusold, S., and A. Sebald. (2000) Double-Quantum Filtration under Rotational-Resonance Conditions: Numerical Simulations and Experimental Results. *J. Magn. Res.* **145**: 340-356.

Eden, M., and M.H. Levitt. (1999) Pulse sequence symmetries in the nuclear magnetic resonance of spinning solids: Application to heteronuclear decoupling. *J. Chem. Phys.* **111**(4): 1511-1519.

Edman, K., P. Nollert, A. Royant, H. Belrhali, E. Pebay-Peyroula, J. Hajdu, R. Neutze, and E.M. Landau. (1999) High-resolution x-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* **401**: 822-826.

Eilers, M., P.J. Reeves, W. Ying, H.G. Gobind Khorana, and S.O. Smith. (1999) Magic angle spinning NMR of the protonated retinylidene Schiff base nitrogen in rhodopsin: Expression of ^{15}N-lysine- and ^{13}C-glycine-labeled opsin in a stable cell line. *Proc. Nat. Acad. Sci. USA* **93**(2): 487-492.

Faham, S., and J.U. Bowie. (2002) Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**(1): 1-6.

Fares, C., F.J. Sharom, and J.H. Davis. (2002) ^{15}N, ^{1}H heteronuclear correlation NMR of gramicidin A in DMPC-d_{67}. *J. Am. Chem. Soc.* **124**: 11232-11233.

Feng, X., M. Eden, A. Brinkmann, H. Luthman, L. Ericksson, A. Graslund, O.N. Antzutkin, and M.H. Levitt. (1997) Direct Determination of a peptide torsion angle psi by double-quantum solid-state NMR. *J. Am. Chem. Soc.* **119**: 12006-12007.

Feng, X., Y.K. Lee, D. Sandstroem, M. Eden, H. Maisel, A. Sebald, and M.H. Levitt. (1996) Direct determination of a molecular torsional angle by solid-state NMR. *Chem. Phys. Lett.* **257**(3,4): 314-320.

Feng, X., P.J. Verdegem, M. Eden, D. Sandstrom, Y.K. Lee, G. Bovee, G. de, J. Lugtenburg, H.J. de Groot, and M.H. Levitt. (2000) Determination of a molecular torsional angle in the metarhodopsin-I photointermediate of rhodopsin by double-quantum solid-state NMR. *J. Biomol. NMR* **16**(1): 1-8.

Feng, X., P.J.E. Verdegem, Y.K. Lee, M. Helmle, S.C. Shekar, H.J.M. de Groot, J. Lugtenburg, and M.H. Levitt. (1999) Rotational resonance NMR of ^{13}C2-labelled retinal quantitative internuclear distance determination. *Solid State Nucl. Magn. Res.* **14**: 81-90.

Feng, Y., and P. Gregor. (1997) Cloning of a novel member of the G protein-coupled receptor family related to peptide receptors. *Biochem. Biophys. Res. Commun.* **231**(3): 651-654.

Fitter, J., R.E. Lechner, and N.A. Dencher. (1999) Interactions of hydration water and biological membranes studied by neutron scattering. *J. Phys. Chem.* **103**: 8036-8050.

Gallivan, J.P., and D.A. Dougherty. (1999) Cation-pi Interactions in Structural Biology. *Proc. Natl. Acad. Sci. USA* **96**: 9459-9464.

Glaubitz, C., G. Gröbner, and A. Watts. (2000) Structural and orientational information of the membrane embedded M13 coat protein by 13C-MAS NMR spectroscopy. *Biochim. Biophys. Acta* **1463**: 151-161.

Glaubitz, C., and A. Watts. (1998) Magic angle-oriented sample spinning (MAOSS): A new approach toward biomembrane studies. *J. Magn. Res.* **130**: 305-316.

Goetz, J., B. Poliks, D. Studelska, M. Fischer, K. Kugelbrey, A. Bacher, M. Cushman, and J. Schaefer. (1999) Investigation of the binding of fluoroluminazes to the 1-MDa capsid of luminaze synthase by ^{15}N{^{19}F} REDOR NMR. *J. Am. Chem. Soc.* **121**: 7500-7508.

Goetz, J.M., and J. Schaefer. (1997) Orientational information in solids from REDOR sidebands. *J. Magn. Res.* **129**: 222-223.

Goto, N.K., and L.E. Kay. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struc. Biol.* **10**: 585-592.

Gregory, D.M., M.A. Mehta, J.C. Shiels, and G.P. Drobny. (1997) Determination of local structure in solid nucleic acids using double quantum nuclear magnetic resonance spectroscopy. *J. Chem. Phys.* **107**(1): 28-42.

Griffin, R.G. (1998) Dipolar recoupling in MAS spectra of biological solids. *Nature Struct. Biol.* (NMR supplement): 508-512.

Gröbner, G., I.J. Burnett, C. Glaubitz, G. Chol, A.J. Mason, and A. Watts. (2000) Observation of light-induced structural changes of retinal within rhodopsin. *Nature* **405**: 810-813.

Gullion, T. (1998) Introduction to rotational-echo, double-resonance NMR. *Concepts in Magn. Res.* **10**(5): 277-289.

Gullion, T., and J. Schaefer. (1989a) Detection of weak heteronuclear dipolar coupling by rotational-echo double-resonance nuclear magnetic resonance. *Adv. Nucl. Magn. Res.* **13**: 57-83.

Gullion, T., and J. Schaefer. (1989b) Rotational-echo double resonance NMR. *J. Magn. Res.* **81**: 196-200.

Hartmann, S., and E.L. Hahn. (1962) Nuclear double resonance in the rotating frame. *Phys. Rev.* **128**: 2042-2053.

Hediger, S., B.H. Meier, N.D. Kurur, G. Bodenhausen, and R.R. Ernst. (1994) NMR cross polarization by adiabatic passage through the Hartmann-Hahn condition (APHH). *Chem. Phys. Lett.* **223**(4): 283-288.

Herzfeld, J., and J.C. Lansing. (2002) Magnetic resonance studies of the bacteriorhodopsin pump cycle. *Ann. Rev. Biophys. Biomol. Struc.* **31**: 73-95.

Hong, M. (2006) Solid-State NMR Studies of the Structure and Dynamics of Disordered and Membrane-Bound Peptides and Proteins. *Acc. Chem. Res.* **39**: 176-183.

Hong, M., J.D. Gross, and R.G. Griffin. (1997a) Site-resolved determination of peptide torsion angle g from the relative orientations of backbone N-H and C-H bonds by solid-state NMR. *J. Phys. Chem. B* **101**(30): 5869-5874.

Hong, M., J.D. Gross, C.M. Rienstra, R.G. Griffin, K.K. Kumashiro, and K. Schmidt-Rohr. (1997b) Coupling amplification in 2D MAS NMR and its application to torsion angle determination in peptides. *J. Magn. Res.* **129**(1): 85-92.

Huster, D., K. Arnold, and K. Gawrisch. (2000a) Strength of Ca(2+) binding to retinal lipid membranes: consequences for lipid organization. *Biophys. J.* **78**(6): 3011-3018.

Huster, D., S. Yamaguchi, and M. Hong. (2000b) Efficient b-sheet identification in proteins by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **122**(46): 11320-11327.

Kamihira, M., T. Vosegaard, A.J. Mason, S.K. Straus, N.C. Nielsen, and A. Watts. (2005) Structural and orientational constraints of bacteriorhodopsin in purple membranes determined by oriented-sample solid-state NMR spectroscopy. *J. Struc. Biol.* **149**: 7-16.

Kim, S., J. Quine, and R. Cross. (2001) Complete Cross-Validation and R-Factor Calculation of a Solid-State NMR Derived Structure. *J. Am. Chem. Soc.* **123**(30): 7292-7298.

Lakshimi, K.V., M. Auger, J. Raap, J. Lugtenburg, R.G. Griffin, and J. Herzfeld. (1993) Internuclear distance measurement in a reaction intermediate: solid-state ^{13}C NMR

rotational resonance determination of the Shiff base configuration in the M photo-intermediate of bacteriorhodopsin. *J. Am. Chem. Soc.* **115**: 8515-8516.

Lam, Y.-H., C.J. Morton, and F. Separovic. (2002) Solid-state NMR Conformational Studies of a Melittin-inhibitor Complex. *Eur. Biophys. J.* **31**: 383-388.

Lam, Y.-H., S.R. Wassall, C.J. Morton, R. Smith, and F. Separovic. (2001) Solid-state NMR structure determination of melittin in a lipid environment. *Biophys. J.* **81**: 2752-2761.

Lanyi, J., and B. Schobert. (2002) Crystallographic structure of the retinal and the protein after deprotonation of the Schiff base: the switch in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **321**(4): 727-737.

Levitt, M.H., D.P. Raleigh, F. Creuzet, and R.G. Griffin. (1990) Theory and simulations of homonuclear spin pair systems in rotating solids. *J. Chem. Phys.* **92**(11): 6347-6364.

Lowe, I.J. (1959) Free Induction Decay of Rotating Solids. *Phys. Rev. Lett.* **2**(7): 285-287.

Luecke, H., B. Schobert, H.T. Richter, J.P. Cartailler, and J.K. Lanyi. (1999) Structure of bacteriorhodopsin at 1.55Å resolution. *J. Mol. Biol.* **291**(4): 899-911.

Marassi, F.M. (2001) A simple approach to membrane protein secondary structure and topology based on NMR spectroscopy. *Biophys. J.* **80**(2): 994-1003.

Marassi, F.M., and S.J. Opella. (2000) A solid-state NMR index of helical membrane protein structure and topology. *J. Magn. Res.* **144**: 150-155.

Marassi, F.M., A. Ramamoorthy, and S.J. Opella. (1997) Complete resolution of the solid-state NMR spectrum of a uniformly ^{15}N-labeled membrane protein in phospholipid bilayers. *Proc. Nat. Acad. Sci. USA* **94**: 8551-8556.

Mason, A.J., S.L. Grage, C. Glaubitz, S.K. Strauss, and A. Watts. (2003) Identifying anisotropic constraints in multiply labelled membrane proteins by ^{15}N MAS NMR. *Biophys. J.* **86**: 1610-1617.

McDermott, A.E., F. Creuzet, R. Gebhard, K. van der Hoef, M.H. Levitt, J. Herzfeld, J. Lugtenburg, and R.G. Griffin. (1994) Determination of internuclear distances and the orientation of functional groups by solid-state NMR: rotational resonance study of the conformation of retinal in bacteriorhodopsin. *Biochemistry* **33**(20): 6129-6136.

Mehring, M. (1983) Fluck E, Diehl P, Kosfeld R, eds. *Principles of High Resolution NMR* Solids. Springer, New York.

Metz, G., X. Wu, and S.O. Smith. (1994) Ramped-amplitude cross polarization in magic-angle-spinning NMR. *J. Magn. Res.* **110**: 219-227.

Middleton, D.A., S. Rankin, M. Esmann, and A. Watts. (2000) Structural insights into the binding of cardiac glycosides to the digitalis receptor revealed by solid-state NMR. *Proc. Nat. Acad. Sci. USA* **97**(25): 13602-13607.

Middleton, D.A., D.G. Reid, and A. Watts. (1995) The conformations of a functional spin-labelled derivative of gastric H/K-ATPase investigated by EPR spectroscopy. *Biochemistry* **34**: 7420-7429.

Middleton, D.A., R. Robins, X. Feng, M. Levitt, I.D. Spiers, C. Schwalbe, D.G. Reid, and A. Watts. (1997) The conformation of an inhibitor bound to the gastric proton pump. *FEBS Lett.* **410**: 269-274.

Moltke, S., I. Wallat, N. Sakai, K. Nakanishi, M.F. Brown, and M.P. Heyn. (1999) The angles between the C(1)-, C(5)-, and C(9)-methyl bonds of the retinylidene chromophore and the membrane normal increase in the M intermediate of bacteriorhodopsin: direct determination with solid-state (2)H NMR. *Biochemistry* **38**(36): 11762-11772.

Montal, M., and S.J. Opella. (2002) The structure of the M2 channel-lining segment from the nicotinic acetylcholine receptor. *Biochim. Biophys. Acta* **1565**(2): 287-293.

Munson, K.B., and G. Sachs. (1988) Inactivation of H+,K+-ATPase by a K+ competitive photoaffinity inhibitor. *Biochemistry* **27**: 3932-3938.

Nishimura, K., S.G. Kim, L. Zhang, and T.A. Cross. (2002) The closed state of a H^+ channel helical bundle combining precise orientational and distance restraints from solid state NMR-1. *Biochemistry* **41**: 13170-13177.

Nomura, K., K. Takegoshi, T. Terao, K. Uchida, and M. Kainosho. (1999) Determination of the Complete Structure of a Uniformly Labeled Molecule by Rotational Resonance Solid-State NMR in the Tilted Rotating Frame. *J. Am. Chem. Soc.* **121**: 4064-4065.

Oesterhelt, D., C. Brauchle, and N. Hampp. (1991) Bacteriorhodopsin: a biological material for information processing. *Quart. Rev. Biophys.* **24**: 425-478.

Oesterhelt, D., and W. Stoeckenius. (1974) Isolation of the cell membrane of *Halobacterium halobium* and its fractionation into red and purple membrane. *Meth. Enzymol.* **31**: 667-678.

Opella, S.J., F.M. Marassi, J.J. Gesell, A.P. Valente, Y. Kim, M. Oblatt-Montal, and M. Montal. (1999) Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. *Nature Struct. Biol.* **6**: 374-379.

Palczewski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **277**: 687-690.

Pang, Y.P., B. Cusack, K. Groshan, and E. Richelson. (1996) Proposed ligand binding site of the transmembrane receptor for neurotensin (8-13). *J. Biol. Chem.* **271**: 15060-15068.

Peersen, O.B., M. Groesbeek, S. Aimoto, and S. Smith. (1995) Analysis of rotational resonance magnetization exchange curves from crystalline peptides. *J. Am. Chem. Soc.* **117**: 7228-7237.

Peersen, O.B., and S.O. Smith. (1993) Rotational resonance NMR of biological membranes. *Concepts in Magn. Res.* **5**: 303-317.

Pines, A., M.G. Gibby, and J.S. Waugh. (1973) Protein-enhanced NMR of dilute spins in solids. *J. Chem. Phys.* **59**: 569-590.

Raleigh, D.P., M.H. Levitt, and R.G. Griffin. (1988) Rotational resonance in solid state NMR. *Chem. Phys. Lett.* **146**: 71-76.

Ramamoorthy, A., C.H. Wu, and S.J. Opella. (1999) Experimental aspects of multi-dimensional solid-state NMR correlation spectroscopy. *J. Magn. Res.* **140**: 131-140.

Royant, A., K. Edman, T. Ursby, E. Pebay-Peyroula, E.M. Landau, and R. Neutze. (2001) Spectroscopic characterization of bacteriorhodopsin's L-intermediate in 3D crystals cooled to 170K. *Photochem. Photobiol.* **74**(6): 794-804.

Sack, I., A. Goldbourt, S. Vega, and G. Buntkowsky. (1999) Deuterium REDOR: principles and applications for distance measurements. *J. Magn. Res.* **138**: 54-65.

Sass, H., G. Buldt, R. Gessenich, D. Hehn, D. Neff, J. Schlesinger, J. Berendzen, and P. Ormos. (2000) Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature* **40**: 649-653.

Schmidt-Rohr, K. (1996) Double-Quantum Solid-state NMR technique for determining torsion angles in polymers. *Macromolecules* **29**(11): 3975-3981.

Schobert, B., J. Cupp-Vickery, V. Hornak, S. Smith, and J. Lanyi. (2002) Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. *J. Mol. Biol.* **321**(4): 715-726.

Seiff, F., I. Wallat, P. Ermann, and M.P. Heyn. (1985) A neutron diffraction study on the location of the polyene chain of retinal in bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **82**: 3227-3231.

Shoji, A., T. Ozaki, T. Fujito, K. Deguchi, S. Ando, and I. Ando. (1990) ^{15}N chemical shift tensors and conformation of solid polypeptides containing ^{15}N-labeled L-alanine residues by ^{15}N NMR. 2. Secondary structure reflected in sigma22. *J. Am. Chem. Soc.* **112**: 4693-4697.

Shull, G.E., A. Schwartz, and J.B. Lingrel. (1985) Amino-acid sequence of the catalytic subunit of the (Na+ + K+)ATPase deduced from a complementary DNA 691. *Nature* **316**: 691-695.

Singh, D., B.S. Hudson, C. Middleton, and R.B. Birge. (2001) Conformation and Orientation of the Retinyl Chromophore in Rhodopsin: A Critical Evaluation of Recent NMR Data on the Basis of the Theoretical Calculations Results in a Refined Picture Consistent with all Experimental Data. *Biochemistry* **40**: 4201-4204.

Smith, S.O., K. Aschheim, and M. Groesbeek. (1996) Magic angle spinning NMR spectroscopy of membrane proteins. *Quart. Rev. Biophys.* **29**: 395-449.

Smith, S.O., I. Palings, M.E. Miley, J. Courtin, H.J. de Groot, J. Lugtenburg, R. A. Mathies, and R.G. Griffin. (1990) Solid-state NMR studies of the mechanism of the opsin shift in the visual pigment rhodopsin. *Biochemistry* **29**: 8158-8164.

Smith, S.O., O.B. Peersen, S. Yoshimura, and S. Aimoto. (1995) Determination of peptide structure in membranes by rotational resonance NMR. *Peptide Chem.* **32**: 109-112.

Spooner, P.J.R., J.M. Sharples, M.A. Verhoeven, J. Lugtenburg, C. Glaubitz, and A. Watts. (2002) Relative orientation between the beta-ionone ring and the polyene chain for the chromophore of rhodopsin in native membranes. *Biochemistry* **41**: 7549-7555.

Straus, S.K., T. Bremi, and R.R. Ernst. (1998) Experiments and strategies for the assignment of fully ^{13}C/^{15}N-labelled polypeptides by solid state NMR. *J. Biomol. NMR* **12**: 39-40.

Straus, S.K., W. Scott, and A. Watts. (2003) Assessing the effects of time- and spatial-averaging in ^{15}N chemical shift/^{15}N-^{1}H dipolar correlation solid state NMR experiments. *J. Biomol. NMR* **26**: 283-295.

Subramaniam, S., and R. Henderson. (2000) Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* **406**(6796): 653-657.

Terstappen, G.C., and A. Reggiani. (2001) *In silico* research in drug discovery. *TRENDS in Pharmacol. Sci.* **22**(1): 23-26.

Thompson, L.K., A.E. McDermott, J. Raap, C.M. van der Wielen, J. Lugtenburg, J. Herzfeld, and R.G. Griffin. (1992) Rotational resonance NMR study of the active site structure in bacteriorhodopsin: conformation of the Schiff base linkage. *Biochemistry* **31**(34): 7931-7938.

Tian, C., K. Tobler, R.A. Lamb, L.H. Pinto, and T.A. Cross. (2002) Expression and initial structural insights from solid state NMR of the M2 proton channel from influenza A virus. *Biochemistry* **41**: 11294-11300.

Toyoshima, C., M. Nakasako, H. Nomura, and H. Ogawa. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**: 647-655.

Ulrich, A.S., M.P. Heyn, and A. Watts. (1992) Structure determination of the cyclohexene ring of retinal in bacteriorhodopsin by solid-state deuterium NMR. *Biochemistry* **31**: 10390-10399.

Ulrich, A.S., I. Wallat, M.P. Heyn, and A. Watts. (1995) Re-orientation of retinal in the M-photointermediate of bacteriorhodopsin. *Nature Struct. Biol.* **2**(3): 190-192.

Ulrich, A.S., A. Watts, I. Wallat, and M.P. Heyn. (1994) Distorted structure of the retinal chromophore in bacteriorhodopsin resolved by ^{2}H-NMR. *Biochemistry* **33**: 5370-5375.

Unwin, N. (2005) Refined structure of the nicotinic acetylcholine receptor at 4A resolution. *J. Mol. Biol.* **346**: 967-989.

Vosegaard, T., and N.C. Nielsen. (2002) Towards high-resolution solid-state NMR on large uniformly ^{15}N- and [^{13}C, ^{15}N]-labeled membrane proteins in oriented lipid bilayers. *J. Biomol. NMR* **22**: 225-247.

Wang, J., J. Denny, C. Tian, S. Kim, Y. Mo, F. Kovacs, Z. Song, K. Nishimura, Z. Gan, R. Fu, J.R. Quine, and T.A. Cross. (2000) Imaging membrane protein helical wheels. *J. Magn. Res.* **144**: 162-167.

Wang, J.F., F. Kovacs, and T.A. Cross. (2001) Structure of the transmembrane region of the M2 protein H^{+} channel. *Protein Sci.* **10**: 2241-2250.

Watts, A. (2005) Solid state NMR in drug design and discovery for membrane embedded targets. *Nature Rev. Drug Discovery* **4**: 555-568.

Watts, A., S.K. Straus, S. Grage, M. Kamihira, Y.-H. Lam, and Z. Xhao. (2004) Membrane protein structure determination using solid state NMR. *In:* Downing K, ed., *Methods in Molecular Biology – Techniques in Protein NMR*. Humana Press, New Jersey, pp. 403-474.

Williamson, P.T.F., S. Bains, C. Chung, R. Cooke, and A. Watts. (2002) Probing the environment of neurotensin whilst bound to the neurotensin receptor by solid state NMR. *FEBS Letts.* **518**: 111-115.

Williamson, P.T.F., G. Gröbner, P.J.R. Spooner, K.W. Miller, and A. Watts. (1998) Probing the agonist binding pocket on the nicotinic acetylcholine receptor: a high resolution solid state NMR approach. *Biochemistry* **37**: 10854-10859.

Williamson, P.T.F., B.H. Meier, and A. Watts. (2004) Structural and functional studies of the nicotinic acetylcholine receptor by solid-state NMR. *Eu. Biophys. J.* **33**: 247-254.

Williamson, P.T.F., J.A. Watts, G.H. Addona, K.W. Miller, and A. Watts. (2001) Dynamics and orientation of N+(CD3)3-bromoacetylcholine bound to its binding site on the nicotinic acetylcholine receptor. *Proc. Nat. Acad. Sci. USA* **98**: 2346-2351.

Wüthrich, K. (1998) The second decade – into the third millenium. *Nature Struct. Biol.* **5**: 492-495.

Yao, X.L., and M. Hong. (2001) Dipolar filtered ^{1}H-^{13}C heteronuclear correlation spectroscopy for resonance assignment of proteins. *J. Biomol. NMR* **20**: 263-274.

Yao, X.L., K. Schmidt-Rohr, and M. Hong. (2001) Medium- and long-distance ^{1}H-^{13}C heteronuclear correlation NMR in solids. *J. Magn. Res.* **149**: 139-143.

Zhong, W., J.P. Gallivan, Y. Zhang, L. Li, H.A. Lester, and D.A. Dougherty. (1998) From ab initio quantum mechanics to molecular neurobiology: A cation-1 binding site in the nicotinic receptor. *Proc. Nat. Acad. Sci. USA* **95**: 12088-12093.

This page intentionally blank

Protein Complexes in SUMO Signaling

CHRISTINA-MARIA HECKER[1] AND IVAN DIKIC[1,2]
[1]*Institute of Biochemistry II, Goethe University Hospital, Theodor-Stern-Kai 7, D-60590 Frankfurt (Main), German, Phone: +49 69 6301 83647, Fax: +49 69 6301 5577, E-mail: ivan.dikic@biochem2.de, corresponding author*
[2] *Mediterranean Institute for Life sciences, Mestrovicevo setaliste bb, 21 000 Split, Croatia*

Abstract

SUMO is a small ubiquitin-related modifier implicated in control of various cellular processes including gene transcription, cell cycle, DNA repair and apoptosis. Here we describe details of the SUMO molecular machineries implicated in the formation of signaling networks that underlie the specificity in these biological processes. SUMO signaling is also altered during development of human diseases, including cancer pathogenesis, and these alterations can be explored as a possible drug target.

Keywords: SUMO, Ubiquitin, NFκB, PCNA, RanGAP1, RanBP2, nucleocytoplasmic transport, reptin, KAI1, PTP1B, cancer pathogenesis

1. INTRODUCTION

Posttranslational protein and lipid modifications are versatile cellular strategies to respond dynamically to intracellular or environmental clues involved in the control of cellular processes (Seet et al. 2006). A modification that has attracted great attention is protein ubiquitylation, which occurs in a

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 75–87.

three-step process leading to the attachment of the small protein ubiquitin (Ub) to lysine residues of a substrate protein (Weissman 2001). Ub is activated in an ATP-dependent manner by a Ub-activating enzyme (E1), and is then transferred to a Ub-conjugating enzyme (E2) via a thioester bond. A Ub-protein ligase (E3) specifically attaches Ub to the ε-amino group of a lysine residue in the target protein (Hershko and Ciechanover 1998). While only a few E1 enzymes are known, humans have more than 20 different E2s. In humans there are about 500-1,000 different E3 ligases, which provide specificity (Hicke et al. 2005). Ubiquitylation reactions are reversed by deubiquitylating enzymes (DUBs), of which several types have been described so far (Hershko and Ciechanover 1998).

The types of Ub modifications are diverse, as well as their functions. In the simplest form, a single Ub molecule is attached, which is defined as monoubiquitylation. It is involved in several processes including endocytosis, endosomal sorting, histone regulation and DNA repair (Haglund and Dikic 2005). Alternatively, a substrate that contains several lysine (Lys) residues may be tagged with multiple single Ub molecules, which plays a role in endocytosis (Haglund and Dikic 2005). Furthermore, Ub itself has seven Lys residues that can be involved in chain formation *in vivo*. Ub chains linked via Lys48 and 63 of Ub are the best characterized so far. Linkage via Lys48 is well known to represent a signal for proteasomal degradation of modified substrates. Lys63 linked chains seem to have completely different, degradation-independent functions involved in processes like DNA repair and protein sorting (Haglund and Dikic 2005).

A class of specific ubiquitin-binding domains (UBDs) mediates most of the effects of protein ubiquitylation. UBDs can be found in enzymes that catalyze ubiquitylation or deubiquitylation, or in Ub receptors that recognize and interpret signals from Ub-conjugated substrate proteins. UBDs are structurally diverse and are found in proteins that contain different structural features and have different biological functions (Hicke et al. 2005).

1.1. SUMO and SUMO paralogues

Just few years ago, only Ub seemed to be an important protein modifier. Today we know that several proteins have similarities both in the sequence and in the three-dimensional structure to ubiquitin. They are named ubiquitin-like modifiers (UBLs) and fulfill other cellular functions than targeting proteins for degradation (Table 1). The best-characterized UBL protein so far is SUMO (Small Ubiquitin-related Modifier). It is present in all eukaryotic cells and is highly conserved from yeast to humans (Melchior 2000). Whereas invertebrates have only a single SUMO gene, three members of the

Table 1.1. Summary of misregulated proteins that are involved in ubiquitin or SUMO modification. For detailed references please see (Hoeller et al. 2006).

Modifier	Pathway	dysregulated protein	Type of dysregulation	Substrate	Modification	Tumour type/disease
Ubiquitin		Mdm2	SNP in the promotor region	p53	polyubiquitylation	non-small cell-lung cancer soft tissue carcinoma colorectal cancer
	cell cycle	HAUSP	downregulation	p53 Mdm2	de-ubiquitylation	non-small cell-lung cancer
		SCF(Skp2)	upregulation of SCF	p27 (Kip)	polyubiquitylation	malignant melanoma lymphoma
		APC		cyclin B, securin	polyubiquitylation	colorectal cancer
	DNA repair	FANCL	defect of PHF9	FANCD2	monoubiquitylation	Fanconi anemia-ralated cancers
	NFκB signaling	CYLD	mutation	NEMO	de-ubiquitylation	cylindromatosis
		IAP2	mutation	Bcl10	polyubiquitylation	MALT lymphomas
	RTK signaling	Cbl	dysregulation	receptor tyrosine kinases	multiple monoubiquitylation	lymphoma, AML, gastric carcinoma
SUMO		SENP1	chromosomal translocation		de-sumoylation	infantile teratoma thyroid oncocytic adenoma
	cell cycle	pRB	mutation	pRB	sumoylation	retinoblastoma tumour
				p14ARF	sumoylation	melanoma
		PIAS3	upregulation		sumoylation	lung, breast, prostate, colon, rectum, brain tumour
		PIASy	downregulation		sumoylation	myelodysplastic syndrome
		SUMO1	dysregulation		sumoylation	anaplastic large cell lymphoma
		SUMO2, UBA2	dysregulation		sumoylation	hepatocellular carcinoma
	apoptosis	UBC9	upregulation	bcl2		ovarian tumour

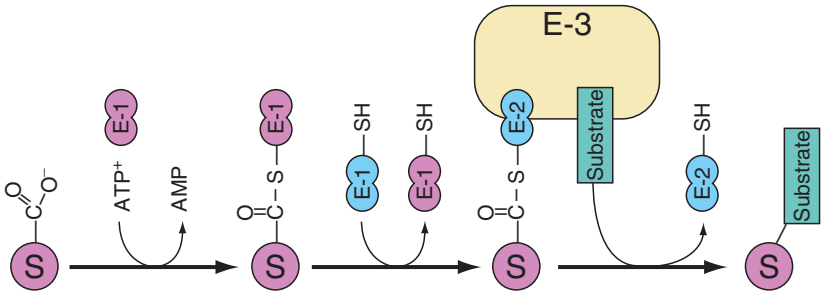


Figure 1.1. Schematic representation of the sumoylation process: SUMO is activated by an E1 enzyme and conjugated to an E2 enzyme via a reactive thiolester bond. E3 enzymes mediated substrate specificity.

SUMO family have been described in vertebrates: SUMO1 and its close paralogues SUMO2 and SUMO3. All SUMO proteins are covalently attached to target proteins through an isopeptide bond by a mechanism similar to that of ubiquitylation, which involves E1, E2 and E3 enzymes (Fig. 1.1) (Melchior 2000). In mammals, only one E1 enzyme and one E2 enzyme are known so far. Interestingly, the E2 enzyme UBC9 is not only able to form a reactive thiolester with SUMO, but also binds SUMO non-covalently. Several E3 ligases for SUMO proteins are known so far, including the PIAS family proteins, Pc2, RanBP2 and TOPORS (Ulrich 2005b). E3 ligases for sumoylation enhance the conjugation process rather than being required for its mediation, as it is generally during the ubiquitylation reactions.

The first description of the reversible SUMOylation was of RanGAP1, a protein localized at the nuclear pore complex (Mahajan et al. 1997). A series of subsequent functional studies has revealed that reversible sumoylation of proteins is instrumental in the control of fundamental cellular processes (Hay 2005; Seeler and Dejean 2003). Conjugation with SUMO has been shown to be involved in the regulation of signal transduction, DNA repair, stress response or nuclear-cytoplasmic transport (Gill 2003; Muller et al. 2004). Although different SUMO paralogues share common properties and conjugation machineries in cells they also have some distinct functions. For example, the promyelocytic leukemia protein (PML) is conjugated to all three SUMO paralogs (Kamitani et al. 1998), whereas RanGAP1 is preferentially modified by SUMO1 and Topoisomerase II with SUMO2/3 during mitosis (Saitoh and Hinchey 2000). Furthermore, the distribution of the SUMO paralogues within cells seems to be different. SUMO1 is uniquely found within the nucleoli, the nuclear envelope and cytoplasmic foci, whereas

SUMO2/3 are accrued on chromosomes at an earlier point in the nuclear reformation process (Ayaydin and Dasso 2004).

1.2. SUMO binding proteins

While our understanding of the mechanisms of SUMO protein conjugation are quite advanced, much remains to be understood on how these modifications are translated into different biological responses. Consistent with the role of Ub binding proteins it was proposed that many if not most functions can be mediated via SUMO binding partners (Haglund and Dikic 2005). Indeed, several studies have shown that the functional properties of SUMO isoforms *in vivo* to a great extent reflect their ability to mediate distinct protein-protein interactions thus forming multimeric signaling complexes (Hay 2005; Seeler and Dejean 2003). This is based on the ability of SUMO paralogues to engage in non-covalent binding to other proteins containing specific motifs that recognize SUMOs, called SUMO-interacting motifs (SIMs), also known as SUMO-binding domains (SBDs) (Hecker et al. 2006; Minty et al. 2000; Song et al. 2004). SIMs are present in a great variety of proteins (Hecker et al. 2006; Minty et al. 2000; Song et al. 2004). Biophysical studies of the SUMO-interacting motif (SIM) revealed that the small hydrophobic region is an essential determinant of SUMO recognition (Hecker et al. 2006; Minty et al. 2000; Song et al. 2004). In the following paragraphs we will describe how sumoylation of specific targets controls cell functions and indicate cases in which these functions are dependent on SUMO binding proteins.

2. CELLULAR FUNCTIONS CONTROLLED BY THE SUMO SYSTEM

Genetic studies in different model organisms underline the crucial function of the SUMO system for normal cell functions. Targeted disruption of the SUMO E2 enzyme Ubc9 gene in mice results in early embryonic lethality and is associated with severe defects in nuclear morphology (Nacerddine et al. 2005). In the budding yeast *S. cerevisiae* temperature-sensitive mutants in the genes for the SUMO system exhibit arrest at the G2/M boundary of the cell division cycle (Hooker and Roeder 2006). The importance of deconjugating SUMO from target proteins is also illustrated by genetic data in the yeast *S. cerevisiae* and in the plant Arabidopsis showing that SUMO deconjugation is needed for viability and in the control of flowering time, respectively (Johnson and Blobel 1997).

2.1. Nucleo-cytoplasmic transport

The first identified target for sumoylation, the GTPase activating protein RanGAP1, also provided the first link between nucleocytoplasmic transport and SUMOylation. RanGAP1 is a key component of the RanGTPase cycle, which serves as the driving force for directional movement through the nuclear pore. Vertebrate RanGAP1 is highly enriched at the nuclear pore, where it forms a stable complex with RanBP2 (Mahajan et al. 1997; Mahajan et al. 1998; Matunis et al. 1996; Matunis et al. 1998). This association depends on RanGAP1 sumoylation. RanBP2 was shown to be the E3 ligase that mediates the sumoylation of RanGAP1. Despite the existence of at least two cytoplasmic SUMO isopeptidases, RanGAP1 is efficiently constitutively modified *in vivo*. The presence of a SUMO E3 ligase at the nuclear pore complex provides the bases for a speculative model in which modification of certain targets is coupled to their shuttling into and out of the nucleus. Yet, it is still not clear if sumoylation precedes import or if import precedes sumoylation. Finally, sumoylation of mammalian RanGAP1 may even affect transport of proteins that are not themselves sumoylated, but contain a SUMO binding domain.

2.2. Transcriptional regulation

Many of the known SIM-containing proteins identified to date have nuclear functions, which is consistent with accepted functions of SUMO, such as inhibition of transcription. Covalent binding of SUMO to various transcription factors has been shown to promote their interaction with co-repressors (Gill 2005). For example, sumoylation of MBD1, a methyl-CpG-binding protein, is in part responsible for the recruitment of the SIM-containing MCAF1 (MBD1-containing chromatin-associated factor 1), which is a component of a chromatin-remodeling complex also containing a histone methyltransferase, SETDB1 (Uchimura et al. 2006). In addition, MBD1 is known to bind SETDB1 independently of its sumoylation status (Lyst et al. 2006 and references therein).

Sumoylation of transcription factors, cofactors or proteins involved in chromatin remodelling have been shown to modulate the transcriptional activity (Melchior 2000) and to regulate signaling pathways, such as the Wnt pathway, which is linked to various human cancers like colon, hepatocellular carinoma, leukaemia or melanoma (Moon et al. 2004; Reya and Clevers 2005). After stimulation with the Wnt ligand, β-catenin enters the nucleus and recruits a chromatin remodelling complex to activate transcription

(Willert and Jones 2006). Among other proteins, reptin was identified to be a part of the β-catenin chromatin remodelling complex (Bauer et al. 2000). Sumoylation of reptin causes repression of the expression of the metastasis suppressor gene KAI1 and results in promotion of tumour metastasis (Kim et al. 2006) (Fig. 2.2). Abrogation of SUMO modification of reptin promotes the expression of KAI1 and thus inhibits the invasive activity of cancer cells (Kim et al. 2006).

The Peroxisome proliferator-activated receptor -γ (PPAR-γ) has essential roles in adipogenesis and glucose homeostasis. It has the ability to repress the transcriptional activation of inflammatory response genes in macrophages. This PPAR-γ-dependent repression is initiated by ligand-induced sumoylation of its ligand-binding domain. This modification targets PPAR-γ to the NCoR complexes associated with the promoter. Thus, the recruitment of the ubiquitylation/19S proteasome machinery, that normally mediates the signal-dependent removal of corepressor complexes, is prevented. As a result, NCoR complexes are not cleared from the promoter and target genes are maintained in a repressed state (Pascual et al. 2005).

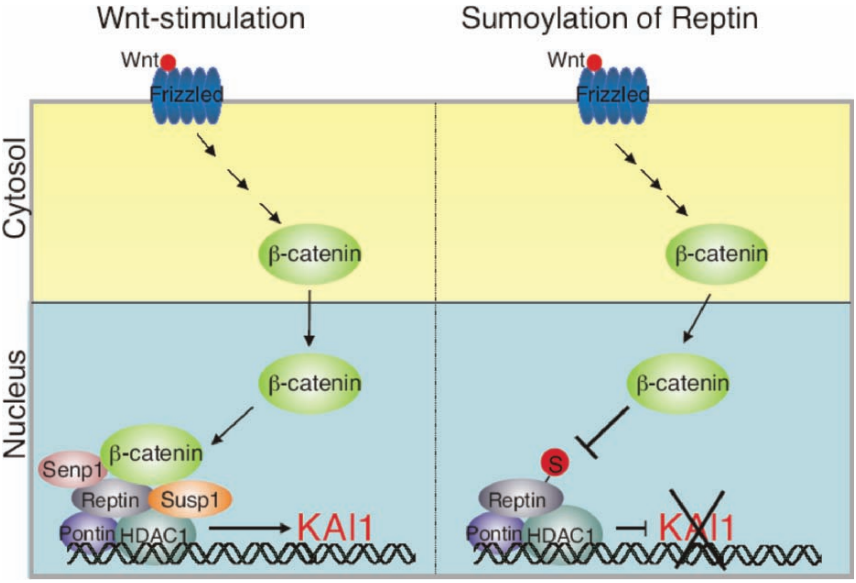


Figure 2.2. Transcriptional repression via sumoylation: Upon Wnt stimulation, β-catenin is translocated to the nucleus. It binds reptin and induces transcription of the tumor suppressor KAI1. When reptin is sumoylated, binding of β-catenin to reptin is abolished and thus expression of KAI1.

2.3. Regulation of intranuclear localization

PML nuclear bodies (NBs) are nuclear vesicle like structures that have been implicated in processes such as transcriptional regulation, genome stability, response to viral infection, apotosis and tumor suppression. A lot of proteins found in the PML NBs are sumoylated (Seeler and Dejean 2003). Furthermore, components of the sumoylation machinery are also localized in PML NBs. However, it is PML (promyelocytic leukemia protein), which is the major and essential component of PML NBs. It contains a SUMO interacting motif and is sumoylated itself. Thus, by PML sumoylation and noncovalent binding of PML to sumoylated PML and other sumoylated proteins, it constitutes a nucleation event and a platform for subsequent recruitment of sumoylated or SUMO-binding proteins (Shen et al. 2006).

3. INTERPLAY BETWEEN Ub AND SUMO SIGNALING

In many cases, there is an active interplay between ubiquitin and SUMO in the regulation of individual proteins and/or cellular pathways. For example, several targets can be modified either by Ub or by SUMO (Ulrich 2005a). Emerging data show that Ub and SUMO cross-talk plays an important role in stress and DNA repair response both by regulating cell cycle arrest (NEMO) and by controlling translesion DNA synthesis (PCNA) (Hoeller et al. 2006).

3.1. NFκB pathway

It has become apparent that modification systems often communicate and jointly affect the properties of common proteins, in some cases by being targeted to the same site. For example, two key regulators in the NF-κB pathway are alternatively modified by SUMO and Ub (Desterro et al. 1998; Huang et al. 2003). SUMOylation protects the NF-κB inhibitor IκBα from degradation by blocking its ubiquitylation at the same lysine residue. Moreover, sequential SUMOylation/ubiquitylation triggers nucleocytoplasmic shuttling and activation of the IκB kinase regulator NEMO in response to DNA damage (Ulrich 2005a).

3.2. Regulation of PCNA

The crucial role of a Ub/SUMO switch in DNA repair processes became evident by work on the yeast PCNA protein (Hoege et al. 2002; Pfander et al.

2005b). PCNA is a target for SUMOylation, mono-ubiquitylation and Lys63-linked multiubiquitylation. Ubiquitylation of PCNA is essential in post-replicative DNA repair by the RAD6 pathway, where multi-ubiquitylation is required for the error-free branch, while mono-ubiquitylation triggers error-prone repair. In yeast it has been shown that during S-phase sumoylated PCNA recruits the DNA helicase Srs2 that prevents the assembly of the homologous recombination machinery (Papouli et al. 2005; Pfander et al. 2005a; Ulrich et al. 2005). This effect in turn appears to pave the way for Ub-dependent damage bypass that would otherwise be less favourable. Thus, SUMO and Ub do not act antagonistically but rather cooperatively to co-ordinate the appropriate response to DNA damage.

3.3. Interplay at the enzymatic level

An interplay between Ub and SUMO modification systems is also apparent at the level of the enzymatic machinery mediating either SUMOylation or ubiquitylation. For example, in *S. cerevisiae* Ubc9 physically interacts with the Rad18 and Rad5 RING-type ubiquitin ligases of the Rad6 pathway, indicating that enzymes for SUMO and Ub can communicate in a common regulatory unit (Hoege et al. 2002). Along this line SUMOylation of the E2-25K E2 Ub-conjugating enzyme has been demonstrated (Pichler et al. 2005). A further extension of this concept is the involvement of a specific enzymatic component in both pathways. The human RING-finger protein TOPORS represents the first example of a dual-specificity ligase, which can promote both ubiquitylation and SUMOylation of p53 (Rajendra et al. 2004; Weger et al. 2005).

Many protein tyrosine kinases and phosphatases are regulated by the Ub system (Seet et al. 2006). Protein-tyrosine phosphatase 1B (PTP1B) is among well-studied phoshatases, which is a critical negative regulator of insulin and leptin signal transduction. In a recent study it was demonstrated that the SUMO E3 ligase PIAS (protein inhibitor of activated STAT1) acts as a PTP1B-interacting protein that reduces its catalytic activity and subsequent insulin signaling. Sumoylation of PTP1B was suggested to induce a conformational change that reduces the phosphatase's ability to interact with substrates. Interestingly, PTP1B sumoylation is dependent on its cellular localization and is most prominently modified in the endoplasmic reticulum and nuclear envelope (Dadke et al. 2007).

4. SUMO IN CANCER PATHOGENESIS

The observed misregulation of components of the SUMO system in human malignancies further illustrates that the integrity of the SUMO pathway is

essential for normal cell function (Table 1). For example, altered expression of Ubc9 in human tumors leads to a growth advantage by modulating apoptotic pathways and may be related to cancer metastasis (for review see (Baek 2006)). A recent case report further links SUMO1 haploinsufficiency to the formation of cleft lip and palate pointing to a role of SUMO in specific developmental programs (Alkuraya et al. 2006).

Due to its common deregulation in cancer development, targeting the SUMO system in cancer therapeutics emerges as a promising approach. Possible strategies involve inhibiting the SUMO activation or conjugation, SUMO ligase activity, or inhibiting the function of the E2 enzyme Ubc9 (Gutierrez and Ronai 2006; Mo and Moschos 2005).

5. CONCLUSIONS AND FUTURE PERSPECTIVES

Since the discovery of SUMO in 1996 a large number of cellular proteins have been reported to undergo SUMO modification (current estimate around several hundred cellular SUMO target proteins) (Rosas-Acosta et al. 2005; Vertegaal et al. 2004). The modification by SUMO appears to alter protein functions by regulating formation of large protein complexes via its ability to mediate protein/protein interactions, intracellular localization, enzymatic activity and transcriptional events. A complete knowledge of the SUMO signaling pathways and in particular the interplay between sumoylation and other post-translational modifications such as ubiquitylation, phosphorylation and acetylation, will provide a basis for better understanding of physiological and pathological cellular processes.

REFERENCES

Alkuraya, F.S., Saadi, I., Lund, J.J., Turbe-Doan, A., Morton, C.C., and Maas, R.L. (2006) SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* **313**: 1751.

Ayaydin, F., and Dasso, M. (2004) Distinct in vivo dynamics of vertebrate SUMO paralogues. *Mol. Biol. Cell* **15**: 5208-5218.

Baek, S.H. (2006) A novel link between SUMO modification and cancer metastasis. *Cell Cycle* **5**: 1492-1495.

Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbacher, U., Aragnol, D., Kemler, R., and Pradel, J. (2000) Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J.* **19**: 6121-6130.

Dadke, S., Cotteret, S., Yip, S.C., Jaffer, Z.M., Haj, F., Ivanov, A., Rauscher, F., 3rd, Shuai, K., Ng, T., Neel, B.G., and Chernoff, J. (2007) Regulation of protein tyrosine phosphatase 1B by sumoylation. *Nat. Cell Biol.* **9**: 80-85.

Desterro, J.M., Rodriguez, M.S., and Hay, R.T. (1998) SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. *Mol. Cell* **2**: 233-239.

Gill, G. (2003) Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. *Curr. Opin. Genet. Dev.* **13**: 108-113.

Gill, G. (2005) Something about SUMO inhibits transcription. *Curr. Opin. Genet. Dev.* **15**: 536-541.

Gutierrez, G.J., and Ronai, Z. (2006) Ubiquitin and SUMO systems in the regulation of mitotic checkpoints. *Trends Biochem. Sci.* **31**: 324-332.

Haglund, K., and Dikic, I. (2005) Ubiquitylation and cell signaling. *EMBO J.* **24**: 1-7.

Hay, R.T. (2005) SUMO: a history of modification. *Mol. Cell* **18**: 1-12.

Hecker, C.M., Rabiller, M., Haglund, K., Bayer, P., and Dikic, I. (2006) Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281**: 16117-16127.

Hershko, A., and Ciechanover, A. (1998) The ubiquitin system. *Annu. Rev. Biochem.* **67**: 425-479.

Hicke, L., Schubert, H.L., and Hill, C.P. (2005) Ubiquitin-binding domains. *Nat. Rev. Mol. Cell Biol.* **6**: 610-621.

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**: 135-141.

Hoeller, D., Hecker, C.M., and Dikic, I. (2006) Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nat. Rev. Cancer* **6**: 776-788.

Hooker, G.W., and Roeder, G.S. (2006) A Role for SUMO in meiotic chromosome synapsis. *Curr. Biol.* **16**: 1238-1243.

Huang, T.T., Wuerzberger-Davis, S.M., Wu, Z.H., and Miyamoto, S. (2003) Sequential modification of NEMO/IKKgamma by SUMO-1 and ubiquitin mediates NF-kappaB activation by genotoxic stress. *Cell* **115**: 565-576.

Johnson, E.S., and Blobel, G. (1997). Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J. Biol. Chem.* **272**: 26799-26802.

Kamitani, T., Nguyen, H.P., Kito, K., Fukuda-Kamitani, T., and Yeh, E.T. (1998) Covalent modification of PML by the sentrin family of ubiquitin-like proteins. *J. Biol. Chem.* **273**: 3117-3120.

Kim, J.H., Choi, H.J., Kim, B., Kim, M.H., Lee, J.M., Kim, I.S., Lee, M.H., Choi, S.J., Kim, K.I., Kim, S.I., et al. (2006) Roles of sumoylation of a reptin chromatin-remodelling complex in cancer metastasis. *Nat. Cell. Biol.* **8**: 631-639.

Lyst, M.J., Nan, X., and Stancheva, I. (2006) Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J.* **25**: 5317-5328.

Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**: 97-107.

Mahajan, R., Gerace, L., and Melchior, F. (1998) Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J. Cell Biol.* **140**: 259-270.

Matunis, M.J., Coutavas, E., and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **13***:* 1457-1470.

Matunis, M.J., Wu, J., and Blobel, G. (1998) SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J. Cell Biol.* **140**: 499-509.

Melchior, F. (2000). SUMO–nonclassical ubiquitin. *Annu. Rev. Cell Dev. Biol.* **16**: 591-626.

Minty, A., Dumont, X., Kaghad, M., and Caput, D. (2000) Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. *J. Biol. Chem.* **275**: 36316-36323.

Mo, Y.Y., and Moschos, S.J. (2005) Targeting Ubc9 for cancer therapy. *Expert Opin. Ther. Targets* **9**: 1203-1216.

Moon, R.T., Kohn, A.D., De Ferrari, G.V., and Kaykas, A. (2004) WNT and beta-catenin signalling: diseases and therapies. *Nat. Rev. Genet.* **5**: 691-701.

Muller, S., Ledl, A., and Schmidt, D. (2004) SUMO: a regulator of gene expression and genome integrity. *Oncogene* **23**: 1998-2008.

Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P.P., and Dejean, A. (2005) The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* **9**: 769-779.

Papouli, E., Chen, S., Davies, A.A., Huttner, D., Krejci, L., Sung, P., and Ulrich, H.D. (2005) Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol. Cell* **19**: 123-133.

Pascual, G., Fong, A.L., Ogawa, S., Gamliel, A., Li, A.C., Perissi, V., Rose, D.W., Willson, T.M., Rosenfeld, M.G., and Glass, C.K. (2005). A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-gamma. *Nature* **437**: 759-763.

Pfander, B., Moldovan, G.L., Sacher, M., Hoege, C., and Jentsch, S. (2005a) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**: 428-433.

Pfander, B., Moldovan, G.L., Sacher, M., Hoege, C., and Jentsch, S. (2005b) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **136**: 428-433.

Pichler, A., Knipscheer, P., Oberhofer, E., van Dijk, W.J., Korner, R., Olsen, J.V., Jentsch, S., Melchior, F., and Sixma, T.K. (2005) SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat. Struct. Mol. Biol.* **12**: 264-269.

Rajendra, R., Malegaonkar, D., Pungaliya, P., Marshall, H., Rasheed, Z., Brownell, J., Liu, L.F., Lutzker, S., Saleem, A., and Rubin, E.H. (2004) Topors functions as an E3 ubiquitin ligase with specific E2 enzymes and ubiquitinates p53. *J. Biol. Chem.* **279**: 36440-36444.

Reya, T., and Clevers, H. (2005) Wnt signalling in stem cells and cancer. *Nature* **434**: 843-850.

Rosas-Acosta, G., Russell, W.K., Deyrieux, A., Russell, D.H., and Wilson, V.G. (2005) A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. *Mol. Cell Proteomics* **4**: 56-72.

Saitoh, H., and Hinchey, J. (2000). Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **275**: 6252-6258.

Seeler, J.S., and Dejean, A. (2003) Nuclear and unclear functions of SUMO. *Nat. Rev. Mol. Cell Biol.* **4**: 690-699.

Seet, B.T., Dikic, I., Zhou, M.M., and Pawson, T. (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**: 473-483.

Shen, T.H., Lin, H.K., Scaglioni, P.P., Yung, T.M., and Pandolfi, P.P. (2006) The mechanisms of PML-nuclear body formation. *Mol. Cell* **24**: 331-339.

Song, J., Durrin, L.K., Wilkinson, T.A., Krontiris, T.G., and Chen, Y. (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc. Natl. Acad. Sci. USA* **101**: 14373-14378.

Uchimura, Y., Ichimura, T., Uwada, J., Tachibana, T., Sugahara, S., Nakao, M., and Saitoh, H. (2006) Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. *J. Biol. Chem.* **281**: 23180-23190.

Ulrich, H.D. (2005a) Mutual interactions between the SUMO and ubiquitin systems: a plea of no contest. *Trends Cell Biol.* **15**: 525-532.

Ulrich, H.D. (2005b) SUMO modification: wrestling with protein conformation. *Curr. Biol.* **15**: R257-259.

Ulrich, H.D., Vogel, S., and Davies, A.A. (2005) SUMO keeps a check on recombination during DNA replication. *Cell Cycle* **4**: 1699-1702.

Vertegaal, A.C., Ogg, S.C., Jaffray, E., Rodriguez, M.S., Hay, R.T., Andersen, J.S., Mann, M., and Lamond, A.I. (2004). A proteomic study of SUMO-2 target proteins. *J. Biol. Chem.* **279**: 33791-33798.

Weger, S., Hammer, E., and Heilbronn, R. (2005) Topors acts as a SUMO-1 E3 ligase for p53 *in vitro* and *in vivo*. *FEBS Lett.* **579**: 5007-5012.

Weissman, A.M. (2001) Themes and variations on ubiquitylation. *Nat. Rev. Mol. Cell Biol.* **2**: 169-178.

Willert, K., and Jones, K.A. (2006) Wnt signaling: is the party in the nucleus? *Genes Dev.* **20**: 1394-1404.

This page intentionally blank

Mass Spectrometry and Glycomics

JASNA PETER-KATALINIĆ
Institute for Medical Physics and Biophysics, Biomedical Analysis Department, University of Münster, Robert-Koch-Str. 31, D-48149 Münster, Germany, Phone: (+49)251-83-52308, Fax: +49 251 8355 140, E-mail: jkp@uni-muenster.de

Abstract

Mass spectrometry (MS) in biochemical and biophysical research of complex carbohydrates has to meet a number of challenges if compared to other biomacromolecules, like proteins and nucleic acids. MS, as an universal and indispensible tool for analysis of biological samples after introduction of soft ionization techniques, like Fast Atom Bombardment (FAB), electrospray (ESI) and matrix-assisted laser desorption ionization (MALDI), allows a high degree of flexibility to develop protocols for investigations to be conducted in a systems biology way. Glycomics arises as a discipline encompassing evaluation and discovery of carbohydrate structure and function. Intact oligomers, analyzed in the gas phase as complex mixtures, render mass over charge (m/z) values reflecting molecular ion patterns which can be assigned to defined molecular compositions. Essential requirements for such MS-based protocols are high sensitivity, high resolution and high mass accuracy. Besides, the protocols to be performed on instruments fitted with fragmentation facilities add sets of data crucial for interpretation by sequencing. Ability for sequencing are given by distinct procedures for fragmentation, which deliver compatible data sets for identification and assignment of single components from complex mixtures. Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS is rendering the highest quality of data with respect to its high mass accuracy and resolving power. By hyphenation of FT-ICR MS with chips and robots development of high throughput mass spectrometry for

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 89–102.

biomedical applications became available. Generation of computer-assisted calculations opens in combination with monosaccharide building block analysis a tool for rapid *de-novo* assignment. Validation of glycomics procedures is carried out within the HUPO HGPI initiative to ensure the quality of analytical data in glycomics life science projects.

Keywords: Mass spectrometry, glycomics, accurate mass determination, carbohydrate sequencing, HUPO – HGPI

Abbreviations:
Hex: hexose
HexNAc: *N*-acetylhexosamine
dHex: deoxyhexose
NeuAc: *N*-acetylneuraminic acid

1. MASS SPECTROMETRY IN GLYCOMICS

Mass spectrometry paved its way into life sciences during the last two decades in a more powerful way upon development of new ionization methods, electrospray (ESI) and matrix-assisted laser desorption/ionization (MALDI), which allow a pragmatic and robust transfer of polar macromolecular species from the solid or liquid to the gas phase (Bruins 1997; Hillenkamp and Karas 2007). Coupling such ion sources to more and more potent mass analyzers, like quadrupole time-of-flight (QTOF), linear and threedimensional ion traps (IT), orbitrap and Fourier transform ion cyclotron resonance (FT ICR) cells, contributed continuously to identification and structure elucidation of relevant small and large molecules in complex biological systems. Detailed insight into molecule structure is available on the basis of the fragmentation behavior of molecular ions, implying different mechanisms of activation and followed by analysis of diagnostic fragment ions and their patterns. Addressing molecular diversity in cellular systems and increasing technological capacities to reduce a number of steps in sample preparation, experiments can currently be performed in a systems biology way to encompass a number of structural parameters in a parallel or a subsequent acquisition mode.

Efficient, highly accurate structural information can be obtained by designing experiments, in which different sets of data are obtained by finely tuned experimental conditions, making advantage of different types of ion activation. In high resolution analyzers, like FT-ICR sustained off-resonance irradiation collision-induced dissociation (SORI-CID), infrared multiphoton induced dissociation (IRMPD) and electron capture dissociation (ECD), are running under different mechanisms providing complementary information on structural parameters and their uniqueness (Marshall et al. 1998).

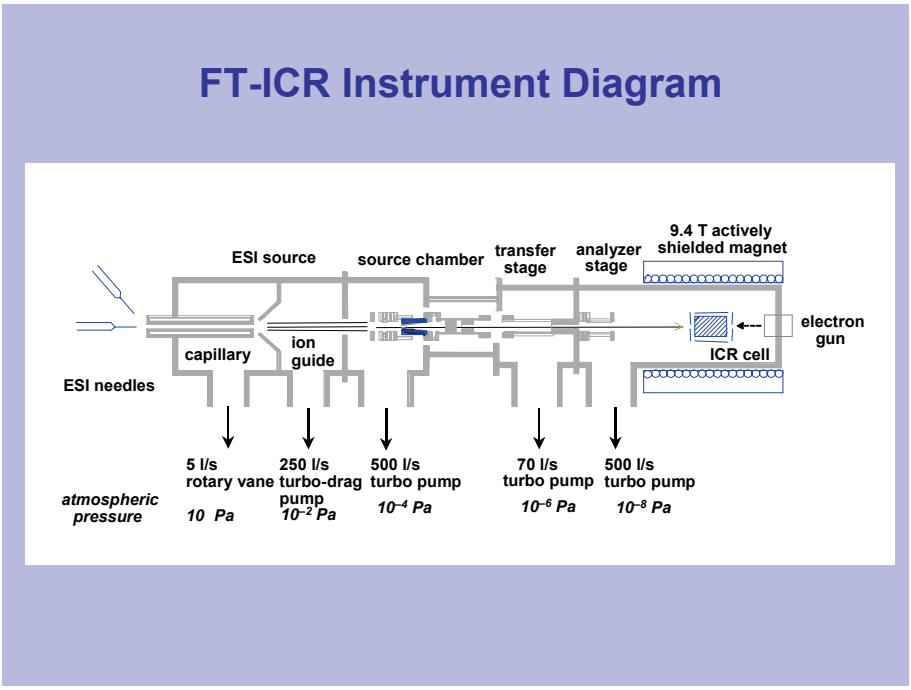


Figure 1.1. Apex II FT-ICR mass spectrometer equipped with a 9.4 T magnet (Courtesy Bruker Daltonik, Bremen, Germany) in the Biomedical Analysis Department, Institute for Medical Physics and Biophysics, University of Muenster, Germany.

New emerging fields of systems biology are in particular proteomics, glycomics, lipidomics and metabolomics. Deciphering human proteome and its application to diagnosis, prognosis and therapy of disease is the mission of the world wide Human Proteome Organisation (HUPO), which include the Plasma, Liver and Brain Proteome projects along with the Antibody and Proteome Standards Initiatives.

The recently initiated Disease Glycomics/Proteome Project Initiative (HGPI) is focussing on characterization of the disease-related glycomes using two complementary approaches: (a) to define standards and the corresponding methods and technologies for deciphering the disease glycome and (b) to decipher the function of single complex carbohydrate structures as specific biomarkers.

For mass spectrometry, as a central technology for discovery, screening and monitoring of abundant and rare complex carbohydrate molecules in purified samples and mixtures, different MS off- and on-line systems must be carefully designed to address human diseases, like cancer, inflammation, life-style-, neurodegenerative and hereditary diseases, and to provide robust data platforms for glycomics in HGPI. Twenty laboratories from all over the

world contributed data obtained with different MS approaches and technologies for structural glycomics (Wada et al. 2007). EurocarbDB contain a large collection of carbohydrate structures obtained by NMR and MS experiments (http://www.dkfz.de/spec/EuroCarbDB_forum/).

The main obstacle toward a rational and rapid access to glycomics is the non-template structure of complex carbohydrates in living systems. Polysaccharides (glycans) have a greater potential for structural diversity than nucleic acids or proteins. Statistically, a vast array of monosaccharide building blocks can be assembled in almost limitless combinations. For single glycomes, the limits can be set by combining the knowledge on biosynthesis and structural carbohydrate chemistry. Human glycosylation pathways, where the sugar chain is attached to the protein by an N-link to asparagine, were in the meantime explored in numerous singular studies and the basic *N*-glycan architecture was shown to be well suited for *in-silico* calculations (https://glycosuite.proteomesystems.com/glycosuite/glycodb). The other large group of glycans, *O*-glycans, is much more difficult to assign, because of the less conserved biosynthetic pathways, where single monosaccharides are added to the growing carbohydrate chains in steps controlled by glycosyl transferases and substrate specificity according to the cell traffic.

For biomarker discovery glycans could represent promissing specific targets, since glycosylation patterns largely differ between different type of cells or organs.

2. GLYCOPROTEOM ANALYSIS

Glycosylation is an ubiquitous posttranslational modification of proteins, where glycans of high structural diversity may carry relevant biological information. Usually, a single glycosylation site can contain a number of glycoforms. A glycan pool may represent a highly heterogeneous mixture like in the case of the secretory IgA where at least 25 different *O*-glycans are expressed on a single *O*-glycosylation site in its hinge region (Rademacher et al. 1988) or in the case of the humane erythrocyte CD59 antigen containing over 100 glycans at its single *N*-glycosylation site (Lloyd 2000). For the carbohydrate part of glycoproteins a variety of biological functions have been proposed. The blood group antigens are composed of oligosaccharides bound to proteins or lipids (Lloyd 2000), glycans are involved in the proper folding and the quality control of newly synthesized glycoproteins in the endoplasmic reticulum (Williams 2006; Mitra et al. 2006) and in control the serum lifetime of glycoproteins (Mitra et al. 2006; Paulson 1989), *N*- and *O*-glycans play a role in the polarized biosynthetic sorting of proteins (Rasmussen 1992), carbohydrates are supposed to participate in the sperm-egg binding (Clark and Dell 2006), just to mention a few.

Two procedures for structural analysis of glycoproteins by means of mass spectrometry are widely used. One involves the enzymatic or chemical release of the intact oligosaccharide prior to either in-gel or in-solution proteolytic digest. *N*-linked oligosaccharide moieties can be released as glycosylamines from the polypeptide chain by use of Peptide-*N*-glycosidase F (*PNGase* F) (Šagi et al. 2005; Müthing et al. 2003; Kunneken et al. 2004; Imre et al. 2005). Simultaneously the formerly *N*-glycosylated Asn residue will be converted to aspartic acid (Asp) by action of the enzyme which leads to a mass shift in the remaining peptide of roughly 1 Da.

The benefit of a separate analysis of the carbohydrate and the protein portion by mass spectrometry is on one hand the characterization of the overall glycan pattern and on the other hand the determination of the previously occupied glycosylation sites by virtue of the mass shift. However, by use of this strategy no assignment of an individual oligosaccharide to a certain glycosylation site can be made.

An other method of choice for analysis of glycoproteins is the separation of intact proteolytic glycopeptides by HPLC prior to mass spectrometric analysis (Wuhrer et al. 2005; Pohlentz et al. 2005). This technique provides information on the glycan structure as well as on the glycosylation site and should be applied in case glycopeptides are not directly detectable in a proteolytic glycopeptide/peptide mixture. For structure elucidation of glycolproteins the analysis of proteolytic glycopeptides is highly advantageous because determination of the glycan structure and the glycosylation site can be achieved in the same experiment. However, during purification procedures succeeding in-gel or in-solution digests of (glyco)proteins losses of proteolytic (glyco)peptides frequently occur in the course of protein identification investigations. In order to prevent the losses and to achieve a maximum of structure information a more efficient and rapid method has been recently developed by taking advantage of a proteolytic digestion in an online procedure with MS analysis. Purified (glyco)proteins or mixtures are mixed with a protease in the electrospray capillary and the resulting peptides and glycopeptides simultaneously analyzed by nanoESI MS and MS/MS (Henning et al. 2007; Steiner et al. 2006) (Fig 2.2).

For O-glycans there is no enzyme available that can be universally applied to de-O-glycosylate proteins. β-elimination as a chemical method is still a single useful method for analysis of O-glycosides (Hanisch et al. 2001; Packer et al. 1998). Under reducing conditions the reaction yields the oligosaccharides as sugar alcohols and results in the formation of unique amino acid residues derived from the formerly *O*-glycosylated Ser and Thr residues. A detailed review on *O*-glycosylation of proteins and its analysis has been recently published (Peter-Katalinić 2005).

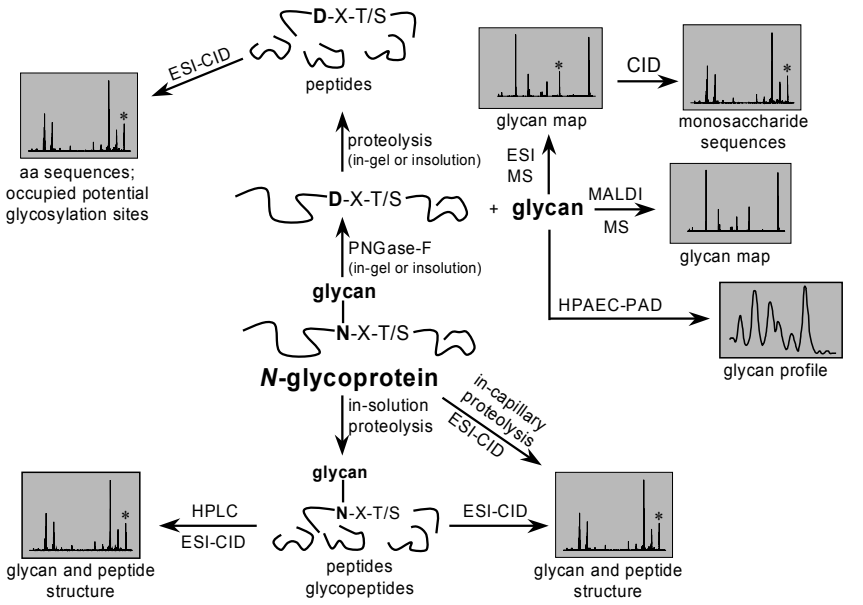


Figure 2.2. Workflow of methods for glycoproteome analysis, established in the Biomedical Analysis Department, Institute of Medical Physics and Biophysics, University of Münster, Germany (reproduced from Henning et al. (2007), with permission).

Taking into account, that glycans are frequently located on the surface of glycoproteins and that protease cleavage sites around these glycosylation sites are accessible to enzymes, it may be useful to run the in-capillary digest using virtually native proteins i.e. without reductive cleavage of the disulfide bonds. As an example the nanoESI mass spectrum obtained during an in-capillary tryptic digest of ribonuclease B is shown in Figure 2.3.

A number of glycopeptide ions could be directly detected during data acquisition. Besides proteolytic (glyco-)peptide ions in the early phase of the in-capillary digest molecular ions derived from the intact protein ($[M+7H]^{7+}$ and $[M+8H]^{8+}$) that already reflect the glycosylation pattern were found. In the course of the analysis the relative intensity of these intact protein ion species decreased for the benefit of the (glyco-)peptide ions.

This phenomenon is frequently observed during in-capillary digests of smaller proteins (Henning et al. 2007). The sequence coverage of the detected tryptic peptides was as low as 53%. Incubation with dithiothreitol prior to in-capillary digest lead to a significantly faster degradation of the protein, however, the relative abundances of glycopeptide ions are decreased in favor of non-glycosylated peptide ions.

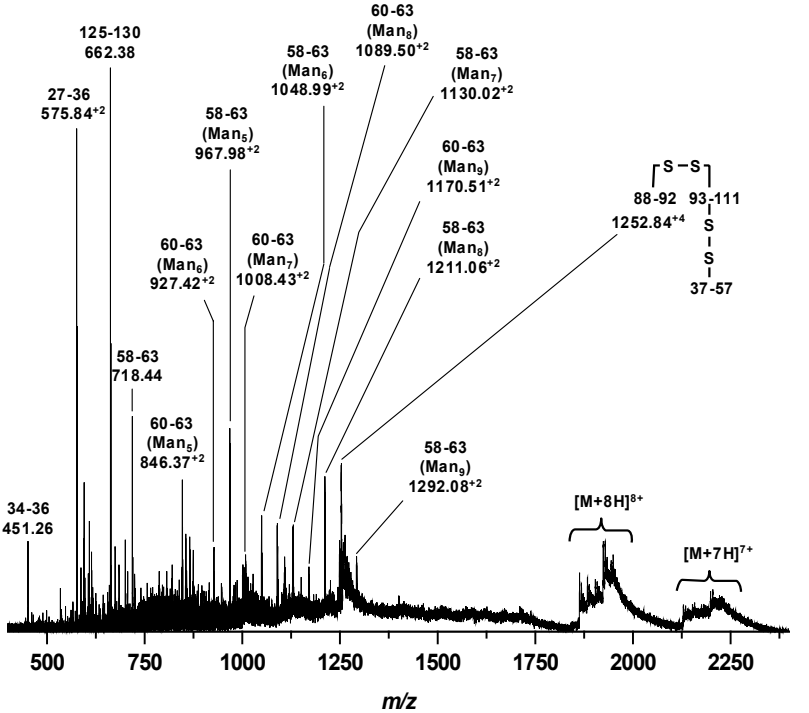


Figure 2.3. NanoESI Q-TOF mass spectrum obtained from an in-capillary tryptic digest of ribonuclease B (RNAse B/trypsin 25:1; reproduced from Henning et al. (2007) with permission).

3. APPLICATION TO MEDICINE: CONGENITAL DISORDERS OF GLYCOSYLATION

Congenital disorders of glycosylation (CDG) are defined as inherited metabolic diseases caused by decreased activity of enzymes, transporters or other functional proteins responsible for the glycosylation processing pathway of glycoconjugates (Keir et al. 1999; Yarema and Bertozzi 2001). For the first time CDGs were reported in 1980 (Jaeken et al. 1980), where the decreased serum thyroxin-binding globulin and increased arylsulfatase A activity in two patients with familial psychomotor retardation had been observed. In the following years a number of human disorders related to *N*-glycosylation pathway were identified, where the clinical pictures of patients are frequently diffuse and difficult for diagnostic prognosis. The types of CDG diseases have been classified into two groups, according to the region of the defected glycosylation pathway (Freeze 2001). The defects in group I are described as those involved in the assembly of the dolichol-phosphate linked oligosaccharide precursor (Carchon et al. 1999) in the cytosol and the endoplasmic

reticulum (Jaeken et al. 1997; Jaeken et al. 1998; Imbach et al. 1999; Korner et al. 1999; Schachter and Jaeken 1999; Dupre et al. 2000), those in the CDG-II group in the processing of the glycans in the ER and the Golgi (Quintin et al. 1990). In most of the presently known CDG cases evidence was obtained for abnormalities in the *N*-glycosylation pathway of glycoproteins, though only little is known about those of the *O*-glycosylation (Quintin et al. 1990).

Aberrant glycosylation is in general investigated by SDS-gel electrophoresis of the patient's plasma to explore the appearance of the transferrin band, which is usually considered to be abnormal, if split in two or more diffuse bands, indicating the underglycosylation on one or both *N*-glycosylation sites. Our attention, however, is turned toward glycopatterns in urine as a potential alternate source of glycoconjugate biomarkers (Linden et al. 1989; Froesch et al. 2003; Froesch et al. 2004; Vakhrushev et al. 2004; Zamfir et al. 2004) because of earlier observations that *O*-linked glycans and *O*-glycosylated amino acids are largely accumulated in the urine of CDG patients. A high degree of heterogeneity concerning the type, number and values of molecular ions must be addressed by sophisticated analytical strategies to include MS mapping and MS^2 sequencing and allow differential alignment of the type, number, abundance and size of molecular species involved.

The best specifications with respect to resolving power, mass accuracy, high sensitivity and dynamic range are found in the FT-ICR mass analyser (Marshall et al. 1998). ESI-FT-ICR MS represents a perfect tool for analysis of complex analyte mixtures as previously shown for crude oil samples (Hughley et al. 2002), humic and fulvic acids from various sources (Kramer et al. 2004; Stenson et al. 2003) or peptide mixtures from digestion procedures (Bergquist et al. 2002). Taking advantage of FTICR MS technology single urine samples from patients suffering from CDG were submited to glycoconjugate analysis for *de-novo* assignment of free oligosaccharides and *O*- and *N*-glycosylated sialylated peptides (Vakhrushev et al. 2006).

In the APEX II FT-ICR mass spectrometer (Bruker Daltonik, Bremen, Germany), equipped with a 9.4 T superconducting magnet (Magnex Scientific Ltd., Oxford, UK) and an Infinity™ cell (Caravatti and Allemann 1991), gas-phase ions were generated by nano-electrospray ionisation in an Apollo ESI source in the positive ion mode. Typical source parameters were capillary voltages of – 800 V and capillary exit voltages of 65 V. Ions were accumulated in the hexapole located behind the second skimmer of the ion source for 0.1 – 1 s and then transferred into the cell of the ICR instrument. Trapping was achieved by application of a "sidekick" at trapping potentials of 1 V. Mass spectra were acquired in the broadband mode in the mass range from m/z 200 to 2000 with 512 k data points. The time-domain signals were zerofilled once and apodized by a quadratic sine bell function prior to Fourier transformation.

A computer simulation based on the building block monosaccharide ratio analysis of a given *m/z* value detected for an ionic species with a given charge state has been developed, for different combinations of glycan building blocks, i.e., desoxyhexoses, hexoses, *N*-acetylhexosamines, and *N*-acetyl neuraminic acid. In the second step of calculations the amino acids Ser and Thr, and other possible sugar modifications such as phosphates and sulfates, were introduced. Applying basic principles of *N*- and *O*-glycan biosynthetic assembling it is possible to assume that the ratio between numbers of hexoses and *N*-acetylhexosamines might indicate the type of the glycan involved. This can be illustrated in the case of *N*-glycans, where the pentasaccharide $Man_3GlcNAc_2$ is a common core for all types of *N*-glycans. The presence of a $Hex_3HexNAc_2$ unit found in the species under inspection should indicate the presence of an *N*-glycan. A "high-mannose" type *N*-glycan can be rather easily identified according to the Hex/HexNAc ratio, 2:1 or higher, whereas for the hybrid type *N*-glycan this ratio decreases, for the complex type the lowest ratio is observed. In *O*-glycans the ratio is 1:1 or lower. Along, fast and reliable distinction of a glycan-linked glycoconjugates from free oligosaccharides has been developed.

In electrospray MS experiments overlapping of precursor ions due to the close proximity of the components' *m/z* values, mostly because of their different charge states, represents a limiting factor for a successful identification. The high resolving power of the FT-ICR MS instrument used (~100.000 in the broadband mode at *m/z* 1000) allows, however, detection of previously unknown molecular entities in spectra with dense ion coverage. In a single urine CDG fraction more than 300 singly, doubly and triply charged distinct ionic species corresponding to approximately 150 molecular species were detected, where about half of them have been directly assigned in the first cycle at the mass accuracy of less than 3 ppm. The *de-novo* identified components were free and amino acid-linked *N*- and *O*-glycans beside some of their phosphated and/or sulphated *O*-glycopeptide counterparts.

In Figure 3.4 the zoomed mass range at *m/z* 674-677 is depicted to illustrate the dense coverage of ions in glycomics experiments, where three distinct molecular species are overlapping: an at the reducing end truncated disialylated, galactosylated, biantennary *N*-glycan species detected as triply charged ions $[NeuAc_2Hex_5HexNAc_3]^{3+}$ at *m/z* 674.2433; singly charged non-sialylated tetrasaccharide ions that have been formed upon elimination of water $[dHexHex_2HexNAc(-H_2O)+H]^+$ at *m/z* 674.2515 and singly charged sialylated trisaccharide ions $[NeuAcHexHexNAc+H]^+$ at *m/z* 675.2470. The urinary trisaccharide NeuAcHexHexNAc could be correlated to the one found in the urine of a patient suffering from mannosidosis to be a NeuAc(α2-6)Man(β1-4)GlcNAc trisaccharide (van Pelt et al. 1990). Similar correlation

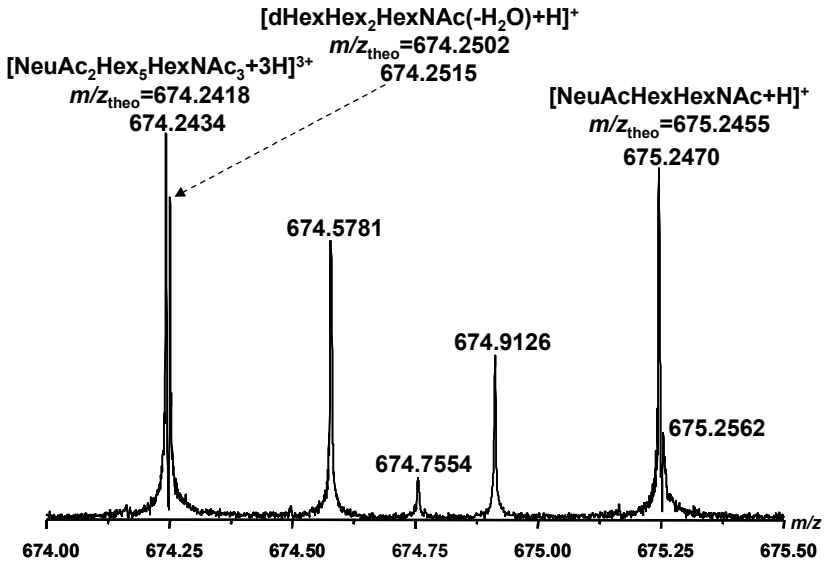


Figure 3.4. Expansion the mass range at *m/z* 674.0-675.5 region from the positive ion mode nanoESI FT-ICR MS of the fraction M5 obtained from urine of the patient KL after gel permeation chromatography (reproduced from Vakhrushev et al. (2006) with permission).

can be established to glycosylated amino acids isolated from the urine of two patients suffering from Schindler's disease (α-*N*-acetylgalactosaminidase deficiency) first identified by one- and two-dimensional 500 MHz ^{1}H-NMR spectroscopy and FAB mass spectrometry as NeuAc(α2-3)Gal(β1-3)[NeuAc(α2-6)]GalNAcα1-R (R = Ser, Thr) with MW 1052.3656 Da and 1066.3813 Da, respectively (Linden et al. 1989), and reinvestigated by a number of additional MS techniques (Froesch et al. 2003; Froesch et al. 2004; Zamfir et al. 2004; Peter-Katalinić et al. 1994; Bindila et al. 2004). A free truncated *N*-glycan NeuAc(α2-3)Gal(β1-4)GlcNAc(β1-2)Man(α1-3)[NeuAc(α2-3)Gal(β1-4)GlcNAc(β1-2)Man(α1-6)]Man(β1-4)GlcNAc, previously found in urine of a patient suffered from sialidosis (van Pelt et al. 1991), could be correlated to the glycan molecular species of MW 2019.7036 Da, depicted as triply charged ions (Fig. 3.4).

This approach allows the construction of a data bank of glycoconjugate structures in biological samples from different sources helpful for diagnostic assignments in human disease. High resolution broadband mass spectra can therefore be generally considered for glycoscreening of complex mixture samples in a single stage. Particular enhancement of identification is provided by computer-assisted calculations in combination with monosaccharide

building block analysis and their non-carbohydrate modifications, such as amino acids, phosphates and sulfates.

4. HGPI: GENERAL VALIDATION OF ANALYTICAL STRATEGIES IN THE HUMAN GLYCOPROTEOME INITIATIVE

N-linked oligosaccharides in transferrin and immunoglobulin G were qualitatively and quantitatively analyzed in 20 laboratories by different mass spectrometric and hyphenated methods, demonstrating that MS-based analysis is efficient and reproducible (Wada et al. 2007). To establish technical standards and demonstrate capabilities of different strategies the glycans were enzymatically detached and analyzed as a native mixture or after permethylation. Human transferrin contains 2 *N*-glycosylation sites, at ^{432}N and ^{630}N, whereas in the human IgG there is a single conserved one, at ^{297}N in the Fc region. All glycans are of the complex type and show structural heterogeneity. Major objective for both glycoproteins is to keep the relatively labile constituents, like sialic acid and fucose moieties, conserved on the sugar chain during the sample preparation and ionization process in the ESI or MALDI ion source. Keeping the nozzle-skimmer potential low, it was possible to keep the sialic acids intact. Also, derivatization by a single methylation or by permethylation was shown to be beneficial for preservation of the glycan structure intact. Fragmentation by MS/MS was shown to deliver sufficient fragment data for the sequence assignment and in some cases to determine the linkage specificity.

The glycopeptide analysis for identification of site-specific oligosaccharide structures is an essential information, which helps to understand the protein folding and functions. In optimizing conditions for the glycopeptide electrospray analysis low energy CID experiment efficient diagnostic cleavage patterns can be obtained (Šagi et al. 2002; Hofsteenge et al. 2001; Maček et al. 2001). Alternatively, applying the principle of non-ergodic processes in the Electron Capture Dissociation experiment on the Fourier Transform ICR-MS analyzer, straight-foreward data on the attachment sites can be obtained (Mormann et al. 2004). Finally, introduction of microfluidic devices, like chips, as hyphenation to MS, opens best perspectives for highly efficient and rational methods to assay glycomics in systems biology combating human diseases (Zamfir and Peter-Katalinić 2001; Froesch et al. 2004; Zamfir et al. 2004). Although the predictions made for new horizons and technical requirements in analytical glycobiology are already fully achieved

(Peter-Katalinić 1994), the high goals for functional and clinical glycomics will need in the future even stronger integration of all mass spectrometric strategies beside significant bioinformatics support.

ACKNOWLEDGEMENTS

Financial support by the Deutsche Forschungsgemeinschaft for the FT-ICR MS instrument (DFG PE 415/14-1) and the Sonderforschungsbereich 492 "Extracellular Matrix: Biogenesis, Assembly and Cellular Interactions (project Z2 to JPK 2000-2002, 2003-2005, and 2006-2008) is greatly acknowledged. I am indebted to all former and present lab members, in particular to Dr. Laura Bindila, Dr. Martin Froesch, Stefanie Henning, nee Kölbl, Dr. Boris Maček, Dr. Michael Mormann, PD Dr. Gottfried Pohlentz, Dr. Dijana Šagi, Dr. Sergei Vakhrushev, Doz. Dr. Željka Vukelić, and Prof. Dr. Alina Zamfir, for their excellent contributions, in part presented in this work. The HUPO HGPI is directed by Prof. Dr. Naoyuki Taniguchi and Dr. Yoshinao Wada is the coordinator of the HGPI pilot study (both University of Osaka Medical School, Japan).

REFERENCES

Bergquist, J., Palmblad, M., Wetterhall, M., Hakansson, P., Markides, K.E. (2002) *Mass Spectrom. Rev.* **21**: 2-15.

Bindila, L., Froesch, M., Lion, N., Vukelic, Z., Rossier, J.S., Girault, H.H., Peter-Katalinić, J., Zamfir, A.D. (2004) *Rapid. Commun. Mass Spectrom*. **18**: 2913-2920.

Bruins, A.P. (1997) In: R. Cole (ed.), *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation Applications*, John Wiley & Sons, Inc. p. 107-136.

Caravatti, P., Allemann, M. (1991) *Org. Mass Spectrom.* **26**: 514-518.

Carchon, H., Van Schaftingen, E., Matthijs, G., Jaeken, J. (1999) *Biochim. Biophys. Acta* **1455**: 155-165.

Clark, G.F., Dell, A. (2006) *J. Biol. Chem.* **281**: 13853-13856.

Coddeville, B., Carchon, H., Jaeken, J., Briand, G., Spik, G. (1998) *Glycoconjugate J.* **15**: 265-273.

Dupre, T., Barnier, A., de Lonlay, P., Cormier Daire V., Durand, G., Codogno, P., Seta, N. (2000) *Glycobiology* **10**: 1277-1281.

Freeze, H.H. (2001) *Glycobiology* **11**: 129R-143R.

Froesch, M., Bindila, L., Baykut, G., Allen, M., Peter-Katalinić, J., Zamfir, A. (2004) *Rapid Commun Mass Spectrom*. **18**: 3084-3092.

Froesch, M., Bindila, L., Vukelić, Z., Peter-Katalinić, J., Zamfir, A. (2004) *Glycoconjugate J.* **21**: 398.

Froesch, M., Bindila, L., Zamfir, A., Peter-Katalinić, J. (2993) *Rapid Commun. Mass Spectrom.* **17**: 2822-2832.

Hanisch, F.G., Jovanović, M., Peter-Katalinić, J. (2001) *Anal. Biochem.* **290**: 47-59.

Henning, S., Peter-Katalinić, J., Pohlentz, G. (2007) In: Lj. Paša-Tolić and M. Lipton (eds.), *Meth. Mol. Biol*., in press.

Henning, S., Peter-Katalinić, J., Pohlentz, G. (2007) *J. Mass Spectrom.,* in press.

Hillenkamp, F., Karas, M. (2007) In: F. Hillenkamp, J. Peter-Katalinić (eds.), *MALDI MS*: *A Practical Guide to Instrumentation, Methods and Applications*, Wiley-VCH Verlag GmbH & Co KgaA, p. 1-28.

Hofsteenge, J., Huwiler, K.G., Maček, B., Hess, D., Lawler, J., Mosher, D.F., Peter-Katalinić, J. (2001) *J. Biol. Chem*. **276**: 6485-6498.

http://www.dkfz.de/spec/EuroCarbDB_forum/

https://glycosuite.proteomesystems.com/glycosuite/glycodb

Hughey, C.A., Rodgers, R.P., Marshall, A.G. (2002) *Anal. Chem*. **74**: 4145-4149.

Imbach, T., Burda, P., Kuhnert, P., Wevers, R.A., Aebi, M., Berger, E.G., Hennet, T. (1999) *Proc. Natl Acad. Sci. USA* **96**: 6982-6987.

Imre, T., Schlosser, G., Pocsfalvi, G., Siciliano, R., Molnar-Szollosi, E., Kremmer, T., Malorni, A., Vekey, K. (2005) *J. Mass Spectrom.* **40**: 1472-1483.

Jaeken, J., Matthijs, G., Barone, R., Carchon, H. (1997) *J. Med. Genet*. **34**: 73-76.

Jaeken, J., Matthijs, G., Saudubray, J.M., Dionisi-Vici, C., Bertini, E., de Lonlay, P., Henri, H., Carchon, H., Schollen, E., Van Schaftingen, E. (1998) *Am. J. Hum. Genet.* **62**: 1535-1539.

Jaeken, J., Vanderschueren-Lodeweyckx, M., Casaer, P., Snoeck, L., Corbeel, L., Eggermont, E., Eeckels, R. (1980) *Pediat. Res.* **14**: 179.

Keir, G., Winchester, B.G., Clayton, P. (1999) *Annu. Clin. Biochem.* **36**: 20-36.

Korner, C., Knauer, R., Stephani, U., Marquardt, T., Lehle, L., von Figura, K. (1999) *EMBO J.* **18**: 6816-6822.

Kramer, R.W., Kujawinski, E.B., Hatcher, P.G. (2004) *Environ. Sci. Technol*. **38**: 3387-3395.

Kunneken, K., Pohlentz, G., Schmidt-Hederich, A., Odenthal, U., Smyth, N., Peter-Katalinić, J., Bruckner, P., Eble, J.A. (2004) *J. Biol. Chem.* **279**: 5184-5193.

Linden, H., Klein, R., Egge, H., Peter-Katalinić, J., Dabrowski, J., Schindler, D. (1989) *Biol Chem. Hoppe Seyler* **370**: 661-672.

Lloyd, K.O. (2000) *Glycoconj. J.* **17**: 531-541.

Maček, B., Hofsteenge, J., Peter-Katalinić, J. (2001) *Rapid Commun. Mass Spectrom*. **15**: 771-777.

Marshall, A.G., Henrickson, C.L., Jackson, G.S. (1998) *Mass Spectrom. Rev.* **17**: 1-35.

Mitra, N., Sinha, S., Ramya, T.N.C., Surolia, A. (2006) *Trends Biochem. Sci.* **31**: 156-163.

Mormann, M., Maček, B., Gonzalez de Peredo, A., Hofsteenge, J., Peter-Katalinić, J. (2004) *Int. J. Mass Spectrom.* **234**: 11-21.

Müthing, J., Kemminer, S.E., Conradt, H.S., Šagi, D., Nimtz, M., Karst, U., Peter-Katalinić, J. (2003) *Biotechnol. Bioeng.* **83**: 321-334.

Packer, N.H., Lawson, A., Jardine, D.R., Redmond, J.W. (1998) *Glycoconj. J.* **15**: 737–747.

Paulson, J.C. (1989) *Trends Biochem. Sci.* **14**: 272-276.

Peter-Katalinić, J. (1994) *Mass Spectrom. Rev.* **13**: 77-98.

Peter-Katalinić, J. (2005) *Methods Enzymol*. **405**: 139-171.

Peter-Katalinić, J., Williger, K., Egge, H., Green, B., Hanisch, F.-G., Schindler, D. (1994) *J. Carbohydr. Chem*. **13**: 445-456.

Pohlentz, G., Kölbl, S., Peter-Katalinić, J. (2005) *Proteomics* **5**: 1758-1763.

Quintin, E., Gladen, A., Roden, L., Kresse, H. (1990) *Proc. Natl Acad. Sci. USA* **67**: 1342-1346.

Rademacher, T.W., Parekh, R.B., Dwek, R.A. (1988) *Ann. Rev. Biochem.* **57**: 785-838.

Rasmussen, J.R. (1992) *Curr. Op. Struct. Biol.* **2**: 682-686.

Šagi, D., Kienz, P., Denecke, J., Marquardt, T., Peter-Katalinić, J. (2005) *Proteomics.* **5**: 2689-2701.

Šagi, D., Peter-Katalinić, J., Conradt, H.S., Nimtz, M. (2002) *J. Am. Soc. Mass. Spectrom*. **13**: 1138-1148.

Schachter, H., Jaeken, J. (1999) *Biochim. Biophys. Acta* **1455**: 179-192.

Steiner, K., Pohlentz, G., Dreisewerd, K., Berkenkamp, S., Messner, P., Peter-Katalinić, J., Schaffer, C. (2006) *J. Bacteriol.* **188**: 7914-7921.

Stenson, A.C., Marshall, A.G., Cooper, W.T. (2003) *Anal. Chem*. **75**: 1275-1284.

Vakhrushev, S., Mormann, M., Peter-Katalinić, J. (2006) *Proteomics* **6**: 983-992.

Vakhrushev, S.Y., Zamfir, A., Peter-Katalinić, J. (2004) *J. Am. Soc. Mass Spectrom.* **15**: 1863-1868.

van Pelt, J., Dorland, L., Duran, M., Hokke, C., Kamerling, J., Vliegenthart, J. (1990) *J. Biol. Chem.* **265**: 19685-19689.

van Pelt, J., Kamerling, J., Bakker, H., Vliegenthart, J. (1991) *J. Inherit. Metab. Dis*. **14**: 730-740.

Wada, Y., Azadi, P., Costello, C.C., Dell, A. Dwek, R. Geyer, H., Geyer, K., Kakehi, K., Karlsson, N.G., Kato, K., Kawasaki, N., Khoo, K.-H., Kim, S., Kondo, A., Lattova, E., Mechref, Y., Miyoshi, E., Nakamura, K., Narimatsu, H., Novotny, M.V., Packer, N.H., Perreault, H., Peter-Katalinić, J., Pohlentz, G., Reinhold, V.N., Rudd, P.M., Suzuki, A., Taniguchi, N. (2007) *Glycobiology* **17**: 411-422.

Williams, D.W. (2006) *J. Cell Sci.* **119**: 615-623.

Wuhrer, M., Deelder, A.M., Hokke, C.H. (2005) *J. Chromatogr B Analyt. Technol. Biomed. Life Sci.* **825**: 124-133.

Yarema, K.J., Bertozzi, C.R. (2001) *Genome Biol.* **2**: 0004.1-0004.10.

Zamfir, A., Peter-Katalinić, J. (2001) *Electrophoresis* **22**: 2448-2457.

Zamfir, A., Sterling, A., Vakhrushev, S., Niebel, H.J., Allen, M., Peter-Katalinić, J. (2004) In: A.E. Ashcroft, G. Brenton, J.J. Monaghan, (eds.), *Advances in Mass Spectrometry,* Vol. 16, Elsevier, Amsterdam, CD-ROM Suppl. No. 629.

Zamfir, A., Vakhrushev, S., Sterling, A., Niebel, H.J., Allen, M., Peter-Katalinić, J. (2004) *Anal. Chem.* **76**: 2046-2054.

Mechanisms of Supramolecular Assembly Exemplified by Microtubules and Amyloid Fibril Formation

PETER BAYLEY
Division of Physical Biochemistry, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK, Phone: +44 20 8816 2085, Fax: +44 20 8906 4477, E-mail: pbayley@nimr.mrc.ac.uk

Abstract

The recent dramatic progress in the application of diffraction and spectroscopic techniques of structural biology has provided high resolution structural information on many individual proteins and their complexes. In cases where such proteins form biological polymers, this information provides a firm basis for consideration of structural factors affecting supramolecular assembly. However information from additional sources and techniques is frequently necessary in defining the mechanisms of the assembly process. In this article, structural and kinetic evidence on the assembly of microtubules and the formation of amyloid fibrils is presented, and compared as examples of two biophysical processes with contrasting biological roles with respect to cellular health and survival.

Keywords: amyloids, dynamic instability, oligomers, prion protein, protofilaments, tubulin

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 103–129.

1. MICROTUBULE ASSEMBLY

Microtubules are one of the major polymeric constituents of the eukaryotic cytoskeleton, with unique dynamic properties. In dividing cells, an extensive cytoplasmic microtubule array serves as a temporary substrate for motor-protein driven vesicle motility throughout interphase: during mitosis this array disassembles rapidly, and reforms with other protein components into the more condensed structure of the mitotic spindle. After metaphase (organisation of the duplicated chromosomes), anaphase (separation of chromosomes to centrosomal poles) and telophase (the separation towards cell division), the microtubule network of the daughter cells is re-established. Hence the dynamics and kinetics of the process are of paramount importance (Kirschner and Mitchison 1986). *In vivo,* microtubules are complexed with microtubule associated proteins, a diverse family including structural proteins, vesicle transporting motor proteins, regulators of dynamic activity, and end-specific proteins responsible for specific cytoplasmic targeting, (Carvalho et al. 2003).

Microtubules are composed of the tubulin αβ heterodimer. The two subunits of tubulin (Tu), each Mr ~50kDa, are closely homologous. They differ in that α-Tu contains a non-dissociating, structural GTP (non-exchangeable) while β-Tu binds GTP reversibly, and whose hydrolysis to GDP provides a source of energy for microtubule assembly and dynamics. Structurally, microtubules are hollow cylindrical structures of ~25 nm external diameter, comprising ~12 to 16 protofilaments, each composed of the longitudinally oriented αβ heterodimer aligned approximately parallel to the long axis. They can exist as either an A- or B-lattice, which differ in the lateral registration of adjacent protofilaments. The 13-protofilament A-lattice and the 16-protofilament B-lattice are perfect helices; all others contain a seam or lateral mismatch. The microtubule is a polar structure, with the ends identified as 'α-out' or minus-end and 'β-out' or plus end. Microtubule motors are generally end-specific with respect to their direction of motion (e.g. the plus-end directed motors, reviewed in Wu et al. 2006).

High resolution structures of tubulin in 2D (Zn)-sheets have been obtained by electron crystallography, by X-ray diffraction of (one) complex of tubulin with an effector protein, and protofilament structure has been resolved from frozen hydrated electron microscopy of assembled microtubules (Nogales and Wang 2006a). Intermediates in assembly and disassembly observed by imaging techniques have also recently been reviewed (Nogales and Wang 2006b).

The process of microtubule assembly of tubulin (generally purified from animal brain) is readily studied *in vitro* under seeded or non-seeded conditions, under suitable buffer conditions (plus GTP and in the absence of Ca^{2+}) and at temperatures 25-37°C; the time course is monitored by light

scattering of bulk solutions, or by fluorescence, dark-field or DIC microscopy of individual microtubules. Seeds can be either from biological sources (e.g. axonemes) or from preformed microtubules stabilized by chemical cross-linking.

1.1. Unique experimental features of the supramolecular assembly of tubulin

1. Assembly into microtubules is dependent on Tu-GTP (and is also supported by non-hydrolysable GTP analogues).
2. During assembly, the GTP is hydrolysed to GDP: microtubules are therefore predominantly composed of Tu-GDP.
3. A critical concentration of Tu-GTP (Cc) remains in equilibrium with the polymer.
4. If solution GTP is exhausted, or the temperature is lowered to ~10°C, microtubules disassemble to Tu-GDP.
5. At protein concentrations > Cc, microtubules grow continuously. At concentrations below Cc (by dilution) microtubules disassemble rapidly.
6. At concentrations close to the Cc, spontaneous transitions occur between growth (G) and shortening (S) phases of random lengths. This is called microtubule Dynamic Instability (Mitchison and Kirschner 1984). This is a unique intrinsic property of the assembled microtubule in this concentration range, implying a unique structural property of this supramolecular assembly. The transitions G to S and S to G are sometimes referred to colloquially as 'catastrophe' and 'rescue'.
7. At high protein concentrations [Tu-GTP] slow oscillations in the total mass of assembled microtubules can also be observed: the slow, rate limiting regeneration of Tu-GTP from the exchange reaction

   Tu-GDP + GTP ↔ Tu-GTP + GDP)

   is then coupled to a synchronised regrowth of the microtubule mass.

1.2. Characteristics of seeded tubulin assembly

Seeded assembly (also called 'nucleated assembly') involves the elongation of a pre-exisiting seed by longitudinal addition. The concentration of seeds/nuclei can be quantified either by counting the microtubule number C_N, or measuring the lengths of a microtubule population. Typically $C_N \sim 10^{-9}$M. If C_N remains constant during growth, assembly follows a first order process i.e. exponential growth, with observed rate constant $k_{obs} = \{k_+ . C_N\}$. Typically $k_{obs} \sim 10^{-2}\ s^{-1}$, corresponding to a bimolecular rate constant for addition, $k_+ = \sim 10^7\ M^{-1}s^{-1}$.

Thus the seed (nucleus) acts effectively as catalyst of conversion of monomer to polymer.

On average one molecule GTP is hydrolysed to GDP + phosphate (Pi) per Tu-GTP added to the lattice whose Tu-GDP content increases (n →n+1) and the process can be approximated as:

Tu-GTP + [TuGTP: (Tu-GDP)(n)] = [Tu-GTP: (Tu-GDP)(n+1)] + Pi

The rate of dissociation (k_-) is observed to be in the range ~10 to 100 s^{-1}.

At the critical concentration Cc, the on-rate (k_+.Cc) and off-rate (k_-) are equal:

Hence $Cc = k_-/k_+ = 10^{-5}$ to 10^{-6} M. For tubulin, Mr = 10^5 Daltons the experimental critical concentration Cc will be typically in the range 0.1 to 1.0 mg/ml.

1.3. Characteristics of tubulin self-assembly

In this case the concentration of nuclei has to be built up from monomers, (conc [M]).

This is a slow process, evidenced by a characteristic lag-phase. The length of the lag-phase is proportional to $[M]^x$, where the nucleating species is M_x. Once nuclei are formed (in a highly co-operative process), if the initial elongation rate >> nuclei formation, then, as before, C_N remains constant. Thus k_{obs} is constant, and in the simplest case, exponential growth will also be found, following the lag-phase. The direct experimental observation of the nucleating species is difficult, since, once formed, it is rapidly elongated, however it can be deduced from the dependence of lag-phase on monomer protein concentration (Oosawa and Asakura 1975).

Experimentally seeded assembly (once seeds are characterised) is more readily quantified. The criterion that the rate constant for assembly is proportional to both the seed concentration and the monomer concentration is characteristic of a simple nucleated assembly. If this is not met, then more complex mechanisms have to be invoked, usually involving slower off-pathway kinetic processes.

1.4. Modelling of microtubule dynamic instability

The unique phenomenon of dynamic instability – the spontaneous transitions between random phases of growth and shortening of an assembled microtubule – has attracted much interst in terms of modelling and kinetic simulation. In the model of Chen and Hill (1985), the hydrolysis of Tu-GTP to Tu-GDP is taken to be slow compared to the initial growth rate of the microtubule (~k_{obs}). A limited amount of Tu-GTP would then accumulate in

a "cap" at the microtubule end (as proposed by Mitchison and Kirschner 1984), while the body of the microtubule is composed of Tu-GDP. In this model, GTP hydrolysis is initiated at random sites within the cap, and, by appropriate, but arbitrary choice of hydrolysis rates, is propagated co-operatively. If the Tu-GTP cap is entirely lost, the microtubule shortens. Bimolecular addition of Tu-GTP can reverse the process and restore growth on a stochastic basis. Spontaneous transitions from G → S and S → G are then seen at concentrations of free Tu-GTP close to the Cc.

We subsequently proposed a simplification of the limiting case in which a terminal Tu-GTP is hydrolysed to Tu-GDP synchronously with the longitudinal addition of Tu-GTP to the same protofilament (Martin et al. 1993). In this 'lateral cap model', the cap is then limited to a single layer of Tu-GTP during growth, but can be lost stochastically to enter the shortening phase, and restored by stochastic addition so that the microtubule grows. This is an assumption, but one which appears quite realistic for representing the transitions between states of growth (Tu-GTP cap present) and shortening (Tu-GTP cap absent) in the protein concentration range close to the experimental Cc. On this model, the stochastic processes are governed by affinities (and rates) for addition of Tu-GTP or loss of Tu-GTP or TU-GDP varying according to the current Tu-GTP or TuGDP content of individual binding sites, and determined by the sum of the affinities of specific geometric lattice interactions. The limits of the rate constants are bounded by the value of the experimental Cc (for Tu-GTP and Tu-GDP), and the affinities are greatest in a fully Tu-GTP environment, and lowest in a fully Tu-GDP site.

Modelling was initially performed for the symmetrical 13-protofilament A-type lattice, as well as the B-type lattice which is now favoured by structural evidence. The B-lattice necessarily contains a structural irregularity (a 'seam') for all except the minor case of the 16-protofilament lattice. The examination of the symmetrical A-lattice was examined to resolve whether either an extended cap or the presence of a seam was necessary for the phenomenon of dynamic instability.

The approach is given in detail in the original literature (Martin et al. 1993). Briefly, the binding site for an incoming Tu-GTPmolecule is defined as xy in terms of the nucleotide state of tubulin as Tu-GTP or Tu-GDP in sites x and y in the the two adjacent protofilaments, related by the lattice symmetry. The treatment depends upon assigning the affinity of the binding site for Tu-GTP, according to free-energies associated with specific relative positions in the lattice (3-start, 5- start, 8-start etc) summed over all interactions at a given site. For modelling the addition of either Tu-GTP or Tu-GDP at either end, only four limiting affinities are required (per microtubule end) for either Tu-GTP or Tu-GDP binding. For the 13 protofilament, A-type microtubule lattice, typical values for the affinities (Kd) for the

binding of Tu-GTP to the four types of sites at the beta-out end of the are 0.2 μM (TT), 3 μM (TD), 4 μM (DT) and 50 μM (DD). The corresponding affinities for the binding of Tu-GDP are 1 μM, 14 μM, 14 μMand 180 μM, respectively. (Corresponding values for the alpha-out end are 3 μM, 6 μM, 9 μM, 16 μM for Tu-GTP, and 50 μM, 95 μM,95 μM, 180 μM for Tu-GDP, respectively). Note the general pattern of highest affinity for the TT site through to lowest affinity for the DD site in all cases. The overall range for a given ligand is constrained by experimentally determined values of the critical concentration.

The (constant) value for k_{on} (the bimolecular rate constant) means each affinity translates into the appropriate site dissociation rate constant, k_{off}. The set of k_{on} & k_{off} rate constants is used for a Monte-Carlo simulation of microtubule length vs time throughout an extended period of growth/shortening events. The lattice energies are then refined to achieve best fit to experimental data of Cc and specifically to reproduce the observed strong concentration dependence of the transition frequencies as a function of [Tu-GTP]. This formal model is therefore an example of the more general situation in a macromolecular assembly showing how the local environment of the binding site determined by the state of liganding (GTP/GDP) can regulate the equilibrium and dynamic properties of the assembly process.

The original publication (Martin et al. 1993) illustrates the pronounced random transitions between phases of growth (G) and shortening (S). Microtubules assembled from pure tubulin plus GTP are seen to be relatively rigid linear assemblies. The dynamic length changes are most clearly seen in conditions of nucleated growth from a biological seed (e.g. axonemal fragment) or from chemically cross-linked, sonicated preformed microtubules. The observation of two superimposed microtubules, one shortening while the other grows through the same point in the field, indicates that the dynamics are an intrinsic property of the microtubule lattice rather than the effect of local solution conditions. The typical measured length vs time plots, show multiple G → S and S → G transitions over a period. The transition behaviour exists in a concentration range around the critical concentration: above this growth is effectively continuous, as is shortening at concentrations below Cc. It is an important test of modelling to reproduce the experimentally observed strong dependence of the transition frequencies on protein concentration within this narrow range.

Numerical simulations using the Monte-Carlo method simulate the experimentally observed dynamic instability for tubulin microtubules, as well as for the inhibition of dynamics and growth by anti-mitotic drugs at concentrations substoichiometric to the microtubule number concentration.

The simulations show that the phenomenon of microtubule dynamic instability can be reproduced by a relatively simple thermodynamic model

involving the choice of a set of affinities and associated kinetic constants related to a specific and fully symmetrical lattice. Other common lattice configurations (determined by the A/B type of lateral association of protofilaments, and the protofilament number) involve a lateral mismatch of protofilaments. In these cases, this 'seam' may well modulate the precise route of assembly and disassembly, but the fundamental transition behaviour is accounted for by the unique GTPase of tubulin under conditions of assembly, and the stochastic nature of the additions and subtractions from the multiplicity of binding sites at the microtubule end. This multiplicity is therefore an intrinsic geometric feature of the supramolecular structure, and also incorporates the biochemical state of the guanine nucleotide (GTP/GDP) due to the specific events GTP hydrolysis in adjacent sites. In fact the Tu-GTP-cap that is postulated to play an important kinetic role in the dynamic properties has never been directly observed or quantified, and remains a working hypothesis. More recently, structural proposals have been made in which the stabilising properties of terminal residues of the microtubule are attributed to the morphological appearance of the local lattice observed under conditions of continuous growth or shortening.

The biological role of microtubule dynamic instability, dependent on the multi-protofilament character of the assembled microtubule, appears potentially to be the provision of a 'search and find' mechanism, acting together with specific microtubule associated proteins in ensuring microtubule attachment to subcellular structures. In addition, the biphasic nature of the microtubule stability underlying the phenomenon contributes to the efficient large scale spatial reorganisations of the microtubule system that are integral to the regulation of different stages of the cell cycle.

1.5. Structural aspects of microtubules: new developments

The thermodynamic description of microtubule assembly considers properties of the growing and shrinking lattice. The combination of electron microscopy of frozen hydrated microtubules, and electron crystallography of sheet and other structures has recently provided a fuller picture of events that can occur at the microtubule ends. It is well-established that microtubules in a rapidly disassembling state can show individual protofilaments splitting off the ends of the microtubule in curved structures (ram's horns) with a diameter of some 50 nm, (Nogales and Wang 2006b). This, together with evidence from structures of the Tu-GDP molecule complexed with drugs and effector proteins, is interpreted as indicating a preference for a curved individual protofilament composed of Tu-GDP, (approximately 12 degrees between the axial vectors of the subunits of an αβ-heterodimer). By contrast, the model

for the microtubule lattice (created by recruitment of Tu-GTP) is of essentially straight protofilaments, deriving from reconstruction of the microtubule lattice from "Zn-sheets" of tubulin. This implies that the disassembly of the bulk microtubule to give Tu-GDP will be spontaneous (i.e. exergonic) as individual tubulin dimers adopt a curved conformation In the case of the non-hydrolysable analogue Tu-GMP-PCP, this angle is reduced to about 6 degrees (Nogales and Wang 2006a).

Additional model building shows how a partially formed extended sheet observed by cryo-EM under conditions of rapid assembly might undergo closure to form a completed cylindrical lattice. Likewise, the disassembly processs can be envisaged to provide energy to move a microtubule associated protein in the direction of the disassembly. The observations of structural intermediates (e.g. curved protofilament sheets seen under fast assembly, and the peeling protofilaments seen under rapid disassembly) indicate that there are additional structural states that may contribute by modulating the intrinsic dynamic properties of the microtubule lattice. These projections arise from detailed analysis of the observable structures that are necessarily static; the extent to which such states contribute to the observable kinetics of assembly/disassembly is unclear. A complementary view is given by consideration of the contribution of lattice interactions in these processes (Buey et al. 2005) emphasizing that the conformation in the lattice and in intermediate structures is determined by the magnitude of lateral (and longitudinal) interactions prevailing under different states of nucleotide content, determined by the course of nucleotide hydrolysis. Also a more mechanical view of the stabilizing of the end of microtubule lattices by elastic forces has been presented, (Janosi et al. 2002), consistent with the Tu-GTP cap model.

1.6. Overview

Microtubule assembly provides a rigorous example of a nucleated assembly mechanism: in seeded assembly, the growth rate is proportional to the concentration of seeds, and the growth rate of individual microtubules is proportional to the concentration of free Tu-GTP. In addition, the assembly process demonstrates an additional unique feature, due to the involvement of GTP hydrolysis, of a supramolecular lattice whose stability can be regulated via the ligand GTP. This in turn gives rise to the unique behaviour, close to the overall steady state of the system, where random transitions between growth and shortening of individual microtubules can occur. High-resolution structural studies show that ligand-induced differences exist in the conformation of Tu-GTP and Tu-GDP as seen in assembled states, and these have a significant contribution to the stability of the microtubule lattice. Also the

microtubule accommodates significant polymorphism in terms of protofilament number, and, to a degree, A/B lattice type. The kinetic criteria regarding the dependence of bulk properties on seed concentration and the observation of changes in the length of individual microtubules as a function of time provide useful guides to determining the mechanism of supramolecular assembly that are readily applicable to other polymeric systems.

2. AMYLOID FIBRIL FORMATION

Amyloid structures have long been recognised as aberrant intra-cellular assemblies visualised as deposits in a range of tissues and pathological conditions (see Westermark 2005 for a historical perspective). They have come to recent prominence due to parallel developments in a) new concepts in protein folding (Dobson 2001) and b) the identification of structural characteristics common to the amyloid fibril state (Serpell et al. 2000b).

The amyloid state of a protein is typified by an increased content of β-sheet structure, specifically the 'cross-β' structure, identified by a 4.6-4.8 A meridional line in X-ray fibre diffraction, that is characteristic of the separation of β-strands oriented perpendicular to the fibre axis, (Serpell et al. 2000a). This state can be adopted by an extremely wide range of proteins and peptides, many without disease associations. The conformational conversion is often indicated by the appearance of β-structure by optical spectroscopic techniques such as circular dichroism (CD) and, more specifically, Fourier Transform Infra Red (FTIR) where absorbance in the 1620 cm^{-1} region is associated with amyloid as opposed to other β-structures. For globular proteins, the transformation inot amyloid frequently involves treatment with agents (heat, acid, limited denaturant, mechanical agitation) to induce a partial denaturation or adoption of the molten globule state. Natively unfolded proteins and peptides may similarly readily form amyloid structures. Fibril growth may require long time scales (even to weeks or months), and fibrils show a variety of dimensions, dependent on the source, and the growth conditions. The basic protofilament appears to have diameter ~2.5 to 5 nm, but fibrils are generally multiple protofilament structures, with a 10 nm diameter being common. At the highest resolution, these fibrils can be seen to have a twisted, striated substructure, associated with the bundling of multiple protofilaments. Thus the term "amyloid" refers to a family of structures, rather than a single specific conformation. The major amyloid fibril structures deduced by a range of biophysical techniques have recently been reviewed, (Makin and Serpell, 2005).

2.1. Amyloids and disease

Particular interest lies in the disease-associated amyloid conditions and their specific amyloidogenic proteins and peptides, namely Alzheimer's disease (Aβ protein 1-40 and 1-42 produced by secretase proteolysis of APP); Parkinson's disease (α-synuclein; NAC (non-amyloid component); Huntington's disease (huntingtin); and the transmissible spongiform encephalopathies, CJD, BSE, scrapie, and kuru (prion protein, Prp). The recognition that non-Mendelian inheritance in yeast is associated with the yeast prion proteins (Sup35; Ure2 etc) is an extension of the 'prion' concept: in this case the presence of prions can have a beneficial rather than disease asociated effect (Wickner et al. 2004; Tuite and Cox 2006).

Homologies between relatively short β-forming regions of amyloid disease proteins are readily identified. For example GAVV occurs in α-synuclein and Prp, and VGGV/VGGL in Aβ (Alzheimer's disease) and Prp. Such proteins may contain more than a single potential amyloid-forming sequence, e.g. Prp; Jamin et al. 2002). Extensive studies have been performed with synthetic peptides as models for amyloid formation, and modifications as small as point mutations, either synthetic or species dependent, can enhance or eliminate amyloid formation. For example, islet amyloid polypeptide (IAPP or amylin), is a characteristic extracellular morphological marker of type 2 diabetes in humans, but the corresponding peptide from rodent IAPP does not form fibrils, attributed to specific sequence differences, (Westermark 1990). Thus while amyloid formation is evidently a widespread phenomenon, the evidence suggests that it appears in many structural variants, and hence presumably involves a variety of mechanisms. This suggests a process, driven in part by the stability of interchain and intra-chain beta sheet interactions, but also involving additional sequence specific factors.

2.2. Examples of amyloid forming peptides

2.2.1. Aβ amyloid

Solid-state NMR of appropriately labelled synthetic peptides provides uniquely detailed structures for amyloidogenic peptides, with identification of residues involved in intermolecular and intramolecular cross-strand interactions. This is exemplified by the recent structural analysis of amyloid prepared from the Aβ1-40 peptide fragment, (Petkova et al. 2006). The analysis of the conformation of the structured residues 9-40 indicates an in-register parallel β-sheet structure for strand alignment, with a central fold in the sheet determining the overall hairpin structure maintained by side-chain interactions. A number of alternative possible structures are proposed that

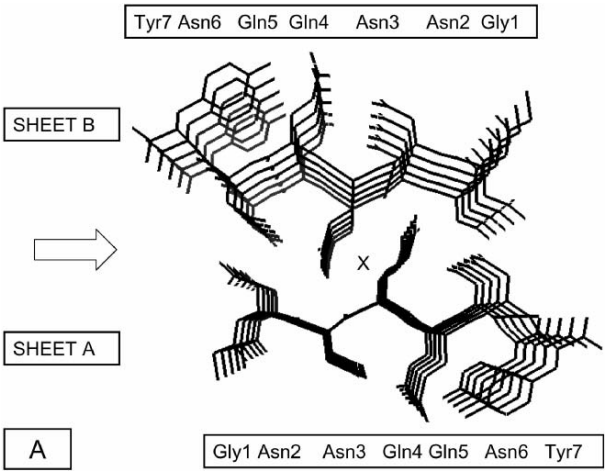
differ in the precise definition of the bend which affects the degree of overlap of two folds of the sheet, and in possible variations in the degree of staggering in the intermolecular arrangements within the folded sheet (i.e. the two components of a hairpin may reside in adjacent layers of the three-dimensional array). Further, the stacking of adjacent folded sheets may involve either parallel or anti-parallel alignment at the quaternary interface. These considerations indicate the high sensitivity of the assembly mechanism to individual side-chain interactions involved in forming the folded sheet, and the subsequent interaction of folded sheets. Combining deuterium-hydrogen exchange nmr data and side-chain packing constraints from pair-wise mutations affecting assembly of Aβ42, Luhrs et al. (2005) have proposed a specific structural model, based on an intramolecular hairpin, comprising residues 18-26 and 31-42, each of which stacks to form an in-register β-sheet, giving a fibril of width ~4.5 nm.

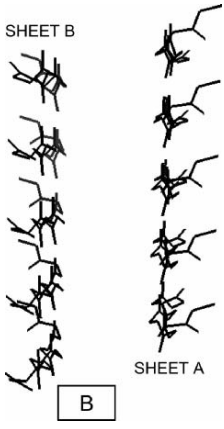
In addition, Petkova et al. (2004). observed that the amyloid fibril of Aβ peptide formed under different conditions of agitation comprised an anti-parallel alignment in forming the basic β-sheet, and similar conflicting findings have been reported for other shorter peptides of Aβ. The type of β-sheet is not necessarily determined uniquely at the level of short (n = 7-9) residues. Amyloid-forming peptides are often 'sticky', tending to aggregate, and solutions, particularly at the high concentration or dehydrated state dictated by some physical studies, may contain some material capable of nucleating the assembly. Hence there is the possibility of some transmission of the past history of a sample, again indicating the sensitivity of the process of assembly to possible extraneous factors.

2.2.2. Sup35 heptapeptide

In view of the above sensitivity of conformation to solution conditions, it is not surprising that there have been few examples of high-resolution crystal structures of amyloidogenic sequences. Nonetheless, crystallisation of the N-terminal heptapeptide sequence GNNQQNY from the fibril-forming domain of yeast Sup35 amyloid-forming protein was recently achieved, (Nelson et al., 2005). Micro crystals were obtained (a few micrometers in length, and smaller cross section) which uniquely allowed the structure determination to be made to < 2Å resolution. This structure (Fig. 2.1) shows a parallel β-sheet formed from the short peptide, with elongation of the sheet occurring perpendicular to the long axis of the crystals. Two-anti-parallel β-sheets interact to form a 'dry' interface which lacks inter chain hydrogen bonding, and no water molecules could be detected. Side-chains of residues N2, Q4 and N6 interdigitate with corresponding residues in the antiparallel sheet, with which they engage in van der Waals interactions. By contrast the

interface between the adjacent molecular pair of sheets is hydrated (i.e. the 'wet' interface.) This sequence was too short to form amyloid fibrils as such, but the alignment, which parallels the Aβ structure described above in some respects, allows indentification of the important interactions leading to the higher order antiparallel docking between two sheets. Each heptapeptide molecule is seen to form 5 backbone CO-NH hydrogen bonds and 4 amide-amide hydrogen bonds (Asn-Asn or Glu-Glu 'stacks') to its two neighbouring peptides in the same sheet. Again this illustrates the combined effect of hydrogen bonding within the β-sheet, plus the specific interactions of appropriately located side chains i.e. the sequence specificity is a major determinant in this structure.





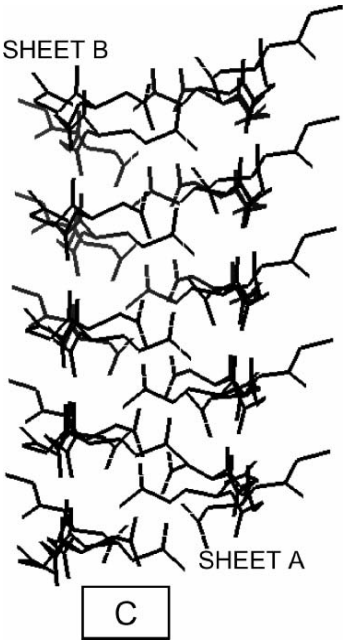


Figure 2.1. The globular domain of Prp, the prion protein, residues 123-230, illustrating the three α-helices and two β-strands (129-131 and 161-163). A single disulphide bridge links residue 179 at the N-terminus of α-helix 2 with residue 214 of α-helix 3. The ovine protein (black) determined by X-ray, (1uw3.pdb, Haire et al. 2004) is superimposed on the human Prp protein (grey) determined by nmr (1hjm.pdb, Calzolai and Zahn, 2003). The largest deviations are seen in the loop joining β2 and α-helix 2. In the full-length protein, this globular domain is preceded by an unstructured sequence with residues 23-90 containing the octapeptide repeats, and additional residues containing the 'cytotoxic peptide', residues 106-126.

These examples illustrate the versatility that derives from the β-sheet structure. In sheets formed form either parallel or anti-parallel alignment of individual strands, amino acid residues protrude from both upper and lower sides of the sheet. The registration of the peptides determines the regular spatial characteristics of both of these surfaces, and the heptapeptide structure above indicates that this can be sufficient to allow a set of strong interactions to be repeated through the interface without engagement of interfacial hydrogen bonding or polar interactions, both of which are seen to enforce the parallel alignment of the strands in a 'polar zipper'. A similar structure has been reported from solid-state nmr of the fast-folding WW-domain of protein human CA150, which is deposited together with huntingtin in Huntington's disease, and which forms a β-hairpin as a 'steric zipper'

(Ferguson et al. 2006). Sheet-sheet interactions involving salt bridges and aromatic ring stacking have been reported in a designed 12-residue peptide sequence that gave crystals resulting in a high-resolution structure, (Makin et al. 2005).

The capacity to adopt parallel β-sheet structure (as opposed to the anti-parallel alignment found in the classic silk fibroin structure) has led to suggestions that the parallel β-helix, as identified in some globular proteins, may be a basic structure for amyloid. However the above caveat on extrapolating structural propensities of amyloid-forming proteins from the structures of relatively short β-sheet forming peptides indicates the need for structural studies on the proteins themselves, to supplement what may be learned from peptide model systems. This reservation may also be taken to indicate a potential role in amyloid formation for the sidechains of residues not directly involved in sheet formation, in determining the type of β-sheet that is formed. Interactions involving these residues will contribute to the overall stability of a sheet structure, and the specific registration of the strands. Clearly the nature of the nucleating species has a strong effect on the eventual type of amyloid structure, and this may differ according to *in vitro* versus *in vivo* conditions. In the case of yeast prions, where the amyloid is physically transmitted between generations, the specific template can be inherited to ensure a precise transmission of the structural information. (Tanaka et al. 2006). However it seems improbable that this mechanism could apply in the more complex mammalian systems, where the mechanism of *in vivo* transmission to neuronal tissues remains unclear.

2.3. Studies of the mechanism of amyloid formation in Solution

A number of model amyloid systems (two of which are discussed here) have been examined in detail, and illustrate some of the features of the 'nucleated assembly' presented in the context of microtubule assembly.

2.3.1. Light chain amyloidosis: kinetics and off-pathway intermediates

This disease results from overexpression of the protein LEN, the immunoglobulin variable domain, also known as the Bence-Jones light chain, resulting in the appearance of amyloid. Experimentally, fibril formation from LEN occurs at pH 7 *in vitro* with vigorous shaking (for 36h), (Souillac et al. 2003). Unusually decreased protein concentration increases the rate of fibril formation, and this effect, also shown by the presence of non-denaturing concentrations of urea, is the result of dissociation of the normally dimeric

protein to the monomer. This forms a nucleating species via an intermediate, I, and polymer is formed by the classical mechanism. At pH 2, the dimer, D converts into an intermediate (D*) that combines with I and inhibits fibril formation. The rate of formation of nuclei is therefore reduced and the lag phase shows little dependence on the protein concentration. This behaviour is typical of the formation of off-pathway intermediates, and the kinetics are then less easily interpretable in terms of individual molecular species. In the case of LEN, the dimer is the stable form *in vivo*: functionally it may be seen as a protector against fibril formation, but if partial unfolding occurs, the intermediate species may be generated, leading to the formation of amyloid.

2.3.2. Aβ protein: amyloid assembly observed by total internal reflectance fluorescence microscopy (TIRF)

The technique of TIRF allows observation of the behaviour of single fluorescent molecules, and is particularly suited to studies of the kinetics of growth of individual polymers, (Ban et al. 2003). The principle depends upon the incidence of the excitation laser light at a glass-water interface at an angle greater than the critical angle for reflectance. An evanescent wave is generated in the aqueous phase whose intensity falls off rapidly, a typical characteristic distance being ~100 Å, depending on wavelength, refractive index etc. This limits the excitation field to a volume close to the glass surface, and avoids passage of the excitation though the aqueous phase which invariable creates background fluorescence and seriously detracts from the signal-noise ratio of images. Unidirectional growth of beta2 microglobulin was visualized directly by use of the dye Thioflavin T (ThT), whose fluorescence is strongly enhanced on binding to amyloid fibrils. However, the rates observed for individual fibres (in the range of a few um per hour) were different for individual fibrils, implying a heterogeneity of the polymeric product. In recent work (Ban et al. 2004), individual Aβ fibrils were allowed to attach to the surface to act as seeds for further growth in the presence of the Aβ protein. The results showed a convincingly linear growth rate for a period of ~1 hour for a number of fibrils within the same field, and the rate was roughly constant at ~15um per hr. This constancy is somewhat surprising, since closer examination of the images shows the presence of substructures (possibly protofilaments, or bundles of the same, at the end of the growing fibril. This would suggest a potential variability of independent growth of these substructures if this was a simple elongation reaction: an alternative would be that molecules encountering the growing end have to undergo a slow process before futher additions occur. These authors examined the effects of D-L amino-acid substitutions in peptides related to Ab, and the technique would be well-suited to monitoring directly the inhibition of amyloid growth. This demonstration of fibril growth, even with

the limited spatial resolution attainable by this fluorescence microscopy, appears a very promising application of TIRF to the study of amyloid formation in general. Other current fluorescence methods allow monitoring the (sub-second time-resolution) movement of isolated individual fluorophores to a precision of ~10Å, i.e. much below the classical λ/2 limit, and such techniques maybe able to help define individual events occurring at the ends of growing assemblies.

2.3.3. Amyloid assembly studied by atomic force microscopy (AFM)

In parallel with the above techniques, recent developments in AFM, including time-resolved AFM (Stolz et al. 2000), show great promise in examining structural properties at high resolution. The indications are that amyloid formation may involve multiple methods of association, possible in a hierarchical mechanism that will be specific to individual cases. Many amyloid fibrils typically have 10 nm diameter (e.g. the immunoglobulin light chains, (Kelly 1996), with a substructure comprising multiple protofibrils or protofilaments, (Ionescu-Zanetti et al. 1999). More recently, a twisted protofibril of 1.8 nm was identified I by AFM *in vitro* assembly of IAPP (amylin), with subsequent production of twisted 2.7 nm fibrils. The latter mechanism may involve addition of one or more extra strands, or possibly the *de novo* addition of units to the basic protofibril structure (Goldsbury et al. 2005).

In conclusion, the formation of amyloid fibrils appears to be a complex and variable process with potential for a multiplicity of pathways, depending on factors such as solution conditions, degree of mechanical agitation, and the nature of the source material. This multiplicity may account for the variety of supramolecular amyloid assemblies that have been identified to date. The amyloid state has, as its typical signature, the cross-β structure that reflects the fundamental step of fibril formation, establishing the fundamental strand-strand interactions that propagate into β-sheets. Sheet-sheet interactions extend the assembly into the third dimension, as exemplified both by the crystal structures of short peptides, and also the effectively linear growth of protofibrils. This growth appears able to proceed by an end-addition mechanism, and multiple fibrils can then give rise to protofilaments that eventually comprise the multiple aligned structure of the amyloid filaments. In addition, many naturally occurring amyloid fibrils derive from proteins that are only partially in the β-form, and the way in which non-β or globular domains are accommodated in the regular structure provides a further source of variation in the ultimate appearance of the supramolecular assembled form.

2.4. The prion protein, Prp

The prion protein occupies a special position amongst the amyloid proteins in that the diseases with which it is associated are transmissible between animals. Prp is the cellular indicator of the TSE diseases – scrapie in sheep, bovine spongiform encephalopathy, (BSE), Gerstmann-Straussler-Scheinker syndrome, (GSS), and Creutzfeldt-Jakob Disease (CJD, and its new variant, nvCJD), (Weissmann 2005). In an infected animal, the normal cellular Prp-c (which is partially α-helical) converts into a proteinase K resistant, aggregated β-form, Prp-res, that is characteristic of the disease state ("scrapie") Prp-Sc, and gives rise to amyloid deposits in the brain and other tissues. Prp-c is a glycosylated protein with a C-terminal GPI lipid-directed membrane anchor. Its normal function is unclear, but may act to reduce oxidative stress, since the protein is known to bind Cu ions. Unlike Prp-sc, Prp-c is PK-sensitive. This form (either as full-length Prp23-231) and Prp89-231, the globular domain that lacks a large N-terminal, largely unstructured (octapeptide-repeat) sequence, has been subject to intense structural scrutiny, most successfully by NMR (Wuthrich and Riek, 2001). Bv contrast Prp-sc has not been assignable to a specific structure, though it is known to involve an increase in β-structure due in part to a conformational conversion of some the helical structure of Prp-c.

The crystal structure of the globular domain of ovine Prp-c has also been resolved (Haire et al. 2004 JMB), and is compared in Figure 2.2 with the NMR structure of the corresponding domain of human Prp, (Calzolai and Zahn 2003 JBC). Both approaches show similar secondary structure (helices α1, α2 and α3, and two short β-strands, β and β2). Helices α2 and α3 are joined by a single disulphide bridge. The two structures differ most in the 'loop' region between β2 and α2, that appears more structured in the crystal structure. Many approaches, including the use of peptide sub-fragments of the sequence, have been used to try to identify what part (or indeed parts) of the structure may be responsible for the so-far unexplained conformational conversion behaviour. For example, β1 (129-231) is close to the 'cytotoxic' sequence 106-126 that also includes a palindomic sequence; Met129 is a sensitive allelic variant for nvCJD; and the short β1- β2 intramolecular β-sheet make a lattice contact with the β1- β2 sheet of its neighbour in the crystal lattice. Comparison with the structure for a dimeric (disulphide exchanged) form of the Hu-Prp (Knaus et al. 2001) suggest a possible further region 190-195 (a highly conserved Thr-rich sequence) that could in principle act as an initiation point for further β-structure formation. However, in spite of numerous relevant peptide sequences that can be converted into amyloid-like fibrils, the structural details of the interconversion process in the globular domain or full-length protein remain unresolved.

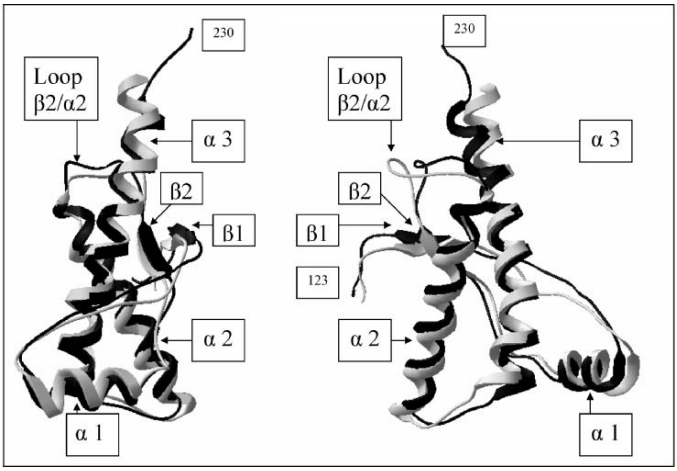


Figure 2.2. Amyloid-like structure of the heptapeptide of the yeast prion protein, Sup35, constructed from 1yjp.pdb. Panel A shows two β-sheets (A and B), both formed from a parallel array of β-strands. The hydrogen bonding is in the direction of the axis perpendicular to the figure. The two sheets are then stacked in anti-parallel orientation, and in projection present a two-fold appearance (about axis X). Side chains of Asn2, Gln4 and Asn 6 interdigitate, forming a dry 'interface', maintained by hydrogen bonding between side-chains and peptide backbones. Panel B shows the orthogonal view (in direction of the arrow of panel A) of the staggered registration of the β-strands (without side-chains). Panel C shows the side-chains of sheets A and B that contribute to the dry interface. This configuration has been termed a 'polar zipper', (Nelson et al. 2005).

Advantage has been taken of the fact that when Prp-sc from an infected animal was treated with proteolytic enzymes under appropriate conditions, reducing it to the 89-231 sequence (Prp27-30), it was found to form 'Prp-rods' in a spontaneous process. This material subsequently formed an ordered, extended 2-dimensional array that was examined by electron crystallography, and compared with a similar (isomorphous) preparation from a recombinant mutant Prp 89-231(Δ141-176) i.e. with a 36-residue deletion. (Wille et al. 2002; Govaerts et al. 2004). The resulting low-resolution difference map was interpreted as indicating a left-handed β-helical structure for residues 141-176 in the Prp-sc. (The β-helix structure occurs naturally in some globular proteins e.g. the pectin lyase-like family as an extended parallel β-helix composed of alternating beta-turn motifs, with an approximately 3-fold symmetry about the helical axis). However, the assembly of the disc-shaped structures into an amyloid has not been observed experimentally. Also, residues 141-176 include those comprising the helix α1 in Prp-c. Other more recent studies

have argued that α1 does not readily convert to β-structure in an aggregated form of truncated Prp, though this was deduced by comparing the c-terminal truncation 23-159 with 23-144 (both being truncations associated with a GSS-like pathological state), rather than from within the context of the full-length protein, (Watzlawik et al. 2006). Other approaches, e.g. molecular dynamics have argued for the beta structure growing from the β1 region, and involving residues up to but not including helix α1. Since Supattapone et al. (1999) showed that helix α1 is not essential for infectivity, it can be suggested that no single sequence within Prp may be exclusively responsible – in keeping with the known ability of a wide range of protein sequences to adopt an amyloid-like structure.

2.4.1. The formation of amyloid-like fibrils of Prp

Prp fibrils can be formed *in vivo* via Prp-sc as described above. It has been shown that intermolecular disulphide formation is not responsible for fibril formation (Welker et al. 2002). *In vitro* fibrils form quite readily after treatment of Prp with excess reducing agent, cleaving the single intramolecular disulphide bond and inducing the b-conformation. This approach was used recently by Tattum et al. (2006) to produce Prp "beta-fibrils" which formed after 3 months at pH 3). The resulting fibrils were found to have only a low cysteine content, suggesting that an intra molecular disulphide had reformed. More generally, fibril formation from Prp-c (without reduction and at neutral pH) can be promoted using mild denaturants (1M guanidinium hydrochloride plus 3M urea (Bocharova et al. 2005), or 1M guanidinium hydrochloride alone (Apetrie et al. 2005). Limited unfolding by brief heat treatment at low pH <5, causes formation of multiple β-oligomeric intermediates that are more resistant to conversion to fibrils (Rezaei et al. 2005). The requirement for partial destabilisation of the monomeric Prp suggests that some form of conformational conversion governs the rate of the process leading to fibril formation *in vitro*. *In vivo*, this conversion may be either assisted or restricted by interactions of the Prp with additional molecules, though, to date, no specfic interaction has been identified unambiguously.

The fibrils of Prp 89-231 prepared by Tattum et al. (2006) exhibit high β-content (and no a-helix). Cryoelectron microscopy revealed fibrils with a protofilament substructure. The density is interpreted in terms of a ladder-like structure, with a basic unit (length 60 Á) stacked in cross-beta orientation, with β-srands perpendicular to the long axis. Substantial additional mass links the two protofilaments, and presumably comprises non-β material, forming the rungs of the ladder. The protofilaments show different

degrees of twist, the variable pitch ranging from 850 Å (short) to 1200 Å (long) indicating a degree of freedom about the overall structure. Thus this represents a completely distinct amyloid form of Prp , compared to that proposed by Govaerts et al. (2004) possibly due to the latter being cellular as opposed to recombinant material, and the fibrils of both cases being produced under very different solution conditions.

2.4.2. The 'protein only' hypothesis and prion 'strains'

The identification of the prion protein as being a component common to a range of pathological conditions, and the fact that the agent used experimentally to transmit disease appears to be without sufficient nucleic acid to convey information by orthodox genetic processes led to the 'protein-only' hypothesis for transmission of disease proposed as a principle by Griffiths (1967), and later developed by Prusiner (1993) and Prusiner (1998) for the specific case of the prion protein being the cause of the transmissible disease. Two general molecular mechanisms of (non-genetic) transmission in particular have been considered – 'template directed assembly' of Prp into Prp-Sc, in which the monomeric Prp is considered to be capable of spontaneously adopting one of numerous potentially infective conformations, which can then be individually imparted to other components of an oligomeric or polymeric assembly; and a 'seeded-nucleation assembly', in which a small amount of inoculated or physically transmitted material dictates the specific conformation of the assembled polymer. While the latter mechanism appears to account for the transmission of yeast prions of different phenotypes to daughter cells at cell division, convincing biophysical evidence for the former mechanism of transmission via multiple Prp conformations conferring the infectious properties of each Prp strain is still lacking, (Aguzzi 2004).

Extensive *in vitro* studies have been performed on the conversion reaction of the prion protein, and in principle the kinetics of the process should yield useful information about underlying mechanisms. However, the general case is that, as with other amyloid systems previously described, a simple nucleated assembly process is rarely observed, due to the multiplicity of associated forms, and the variation of the types of oligomers with *in vitro* solution conditions. In addition to fibrils, oligomers of various sizes are formed as relatively stable non-fibrillar products of assembly, as well as being implicated kinetically as off-pathway intermediates. Hence the simple relationships of seeded assembly – rates dependent on the concentration of both seeds and the monomeric species – are rarely found (Baskakov and Bocharova 2005). The formation of Prp fibrils *in vitro* is often a very slow process, and effectively irreversible. This adds further to the difficulty in

treating the experimental phenomenon within the concepts of orthodox physical principles of protein assembly.

The *in vitro* studies are clearly fundamental to understanding the pathological states with which amyloid is associated, however remote these may appear to be from the idealised assembly reactions. In the case of mammalian subjects, while inoculation and hence direct transfer of infected material into host animals, is frequently used, disease transmission can also follow an oral route, with the disease later manifested in a wide range of neuronal and other tissues (Cashman and Caughey 2004). The pathway is clearly more complex, and the manifestation of "strain" phenotypes make the complex nature of the infectious agent even more difficult to assign. A number of distinct strains of scrapie prions can be derived from sheep isolates. The strains differ by the length of the incubation times (prior to death) in various transgenic mouse lines, and by the lesion patterns with which they are associated in the affected brains. Most significantly, (and allowing for the existence of some species barriers between donor and host), different prions strains can be propagated within a single specific mouse as host which is homozygous for Prp (i.e. only a single Prp sequence is present). In terms of the 'protein only' hypothesis (that the donated Prp is the causative agent of the disease state), this means that the single polypeptide sequence (of the host protein) is apparently 'able to mediate different strain phenotypes', (Weissmann 1999). While the term 'mediate' does not imply a specific mechanistic interpretation for prion strains, this idea is further developed in the 'conformational hypothesis'. This proposes that each strain is associated with a different conformation of PrP-Sc and that this conformation is conserved in the conversion of the host PrP-c of the host into Prp-Sc. This remains a hypothesis since the definition of Prp-Sc is not fully defined (e.g. both PK resistant and sensitive types have been identified), and the conformation has not been defined to sufficient resolution to identify strain dependent conformational differences.

2.5. What is the transmissible 'infectious' agent in the prion diseases?

2.5.1. The origins of prion disease

The association of assembled forms of Prp with the infected state has made Prp-Sc a prime candidate for being the agent of infection. Prp-Sc is characterised as diffuse highly aggregated material *in vitro*, which is proteinase-K resistant on SDS gels, and which is shown *in vitro* to be transformed into

amyloid-like prion 'rods' following proteolysis (McKinley et al. 1991). Prp-Sc is thought to form *in vivo* when misfolded protein accumulates to such an extent that the proteasome cannot dispose of it (Dimcheff et al. 2003). Proteasome inhibitors stimulate PrpSc formation, (Ma et al. 2002). This accumulation occurs during the course of the intracellular trafficking involved in internalisation of Prp-c from the plasma membrane into low pH endosomes (Nunziante et al. 2003).

However: *in vivo* evidence questions whether Prp-Sc is the most appropriate indicator of infection, for example: a) little or no accumulation of PrpSc is seen post mortem in some pathogenic cases (such as Familial CJD; b) the converse situation is recognised of of extensive PrpSc accumulation, but no pathogenesis; c) the early effects of prion infection causes loss of synaptic function, destruction of hippocampal synapses and loss of cognitive processes (learning) significantly earlier than the appearance of amyloid-like material (Cunningham et al. 2003).

Hence, is PrpSc/amyloid the end-product of the degenerative process? If so, is the active agent an intermediate in amyloid formation eg an oligomeric form of Prp? Two recent approaches have attempted to identify the "infectious" agent. The first examines the possibility of making "synthetic prions". In many cases where oligomeric or fibrillar Prp species were tested for infectivity, negative results were found. However, fibrils were recently made *in vitro* from recombinant (non-glycosylated, non-GPI anchored) Prp, (Legname et al. 2005). Injection of this material into the brain of transgenic mice caused pathological symptoms and death. Further passage of this brain material into similar mice transmits the pathological results. However: the mice used have high degree of overexpression of Prpc, and hence might be considered 'abnormal' in sensitivity. Also the specific level of infectivity was much lower than found with *in vivo* derived PrpSc material. Thus, while this has been an important finding, the interpretation of these experiments is still somewhat ambiguous. In a different approach, PrpSc material isolated from infected animals was subjected to a new semi-preparative method of chromatography (flow-field flow fractionation) to separate oligomeric Prp components as a function of size (Silveira et al. 2005). Certain fractions of intermediate molecular size tested positive for ability to effect the *in vitro* conversion of Prpc to PK resistance, and for the incubation time following transmission into transgenic mice. The most infective fractions were sharply defined as having a size ~300-600kD (i.e. about 15-30 monomers), assessed by light scattering. The precise composition of this naturally produced material is not fully known, nor is it clear whether Prp is acting by itself, or in the context of additional components. This is the subject of a recent review considering the evidence for components additional to Prp that are involved in the transmissible disease (Caughey and Baron 2006).

2.5.2. Prp Oligomers: implications for therapy

Potential inhibitors. The above work follows numerous reports from other amyloid forming systems that oligomeric species of the respective proteins, and their proteolytic fragments are frequently cytotoxic, though not transmissible. In the case of the prion diseases, they suggest that Prp oliogmer formation may be a valid target for therapy? A further suggestion is therefore that the amyloid fibril form may be the cells attempt to negate the toxicity of oligomers, and the presence of fibrils could denote the unsuccessful aresult of this potentially protective action. If this is he case, then agents that suppress formation of amyloid fibrils might result in the accumulation of oligomeric species. For example an agent that inhibits the lateral associations typical of amyloid, might increase the concentration of small potentially toxic oligomers, such as those identified in other systems as being off-pathway intermediates.

Amongst the possible current experimental approaches, two are notable: High throughput screening of libraries of large numbers of compounds that may inhibit *in vitro* assembly and cytotoxicity, or 'curing' of scrapie-infected cell culture. (Boshuizen et al. 2004; Breydo et al. 2005; Caughey et al. 2006). Similarly, antibodies against other amyloid-forming proteins such as Aβ have had some success, but have resulted in serious side effects in clinical trials due to neuroinflammation, and alternative vaccination techniques are now being explored (Okura et al. 2006).

For future work, the need for a better understanding of Prp oligomer formation and action is indicated. In my group we have aimed at a rational design of oligomers of defined size to identify nucleating or toxic species. The key aims are a) inhibition of the fibril formation from the wt Prp sequence *in vitro*; b) prevention of Prp oliogmer formation from monomeric Prp; and c) inhibition of cytotoxic effects of specific synthetic oligomers of Prp. In parallel with this, and based on other amyloid-forming systems, we are attempting the rational design of modified peptide blockers of beta structure formation to identify inhibitors of the fundamental steps of oligomerisation. At present, the precise sequence (or sequences) involved in Prp amyloid formation have not been identified. Many peptide fragments of the Prp sequence (such as the "cytotoxic peptide" sequence Prp 106-126) have been shown to adopt soluble beta and amyloid fibril structure *in vitro*.

3. CONCLUSIONS

The two protein systems examined here – microtubules and amyloid – represent strikingly different examples of biological supramolecular assembly. Microtubules are a unique and specific component of the cytoskeletal system, with a fundamental role in the dynamics of cell division, and in providing

the spatial pathways for intracellular transport. They depend on a strongly conserved class of tubulin proteins with a fundamental role attached to the different properties of the alpha and beta tubulin molecules that defines their polar structure, and confers the essential and unique property of dynamic instability on the multi-protofilament assembly. By contrast, the amyloid state is an aggregated state that can be adopted by a very wide range of proteins – often under non-physiological conditions, and by mechanisms that are frequently more complicated than a simple nucleated growth process, possibly due to the more extensive three-dimensional nature of amyloid-forming interactions. There is therefore relatively little overall sequence specificity that defines amyloid formation in general, though there is a common fundamental requirement for some fraction of the sequence being able to form beta-sheet associations, possible representing only a sub-fraction of the total sequence. This offers some prospect of being able to target potential inhibitors of both the formation of amyloid, and also the reversal of pre-existing amyloid-like aggregates. Naturally occurring amyloid is usually associated with disease, though not its cause. In many cases it appears to be the end-product of a process involving abnormal protein expression, or malfunction of cellular regulatory processes. In the special case of the Transmissible Spongiform Encephalopathies (TSE's), the disease pathway is still unresolved, though it is becoming more unlikely that the prion protein is the sole agent of disease. Future developments will inevitably seek to complement considerations from the purely structural aspects of supramolecular assembly with those more physiological approaches that address the cellular, subcellular and membrane interactions involved in the transmission of prion disease within the mammalian system.

ACKNOWLEDGEMENTS

With thanks to colleagues at NIMR, especially Steve Smerdon for assistance with graphics; Stephen Martin for his collaborations on microtubules; and Angus Nash, Oxana Polyakova and Igor Bronstein, for their interest in amyloids.

REFERENCES

Aguzzi, A. (2004) Understanding the diversity of prions. *Nat. Cell. Biol.* **6**: 290-2.

Apetri, A.C., Vanik, D.L. and Surewicz, W.K. (2005) Polymorphism at residue 129 modulates the conformational conversion of the D178N variant of human prion protein 90-231. *Biochemistry* **44**: 15880-8.

Ban, T., Hamada, D., Hasegawa, K., Naiki, H. and Goto, Y. (2003) Direct observation of amyloid fibril growth monitored by thioflavin T fluorescence. *J. Biol. Chem.* **278**: 16462-5.

Ban, T., Hoshino, M., Takahashi, S., Hamada, D., Hasegawa, K., Naiki, H. and Goto, Y. (2004) Direct observation of abeta amyloid fibril growth and inhibition. *J. Mol. Biol.* **344**: 757-67.

Baskakov, I.V. and Bocharova, O.V. (2005) *In vitro* conversion of Mammalian prion protein into amyloid fibrils displays unusual features. *Biochemistry* **44**: 2339-48.

Bocharova, O.V., Breydo, L., Parfenov, A.S., Salnikov, V.V. and Baskakov, I.V. (2005) *In vitro* Conversion of Full-length Mammalian Prion Protein Produces Amyloid Form with Physical Properties of PrP(Sc). *J. Mol. Biol.* **346**: 645-59.

Boshuizen, R.S., Langeveld, J.P., Salmona, M., Williams, A., Meloen, R.H. and Langedijk, J.P. (2004) An *in vitro* screening assay based on synthetic prion protein peptides for identification of fibril-interfering compounds. *Anal. Biochem.* **333**: 372-80.

Breydo, L., Bocharova, O.V. and Baskakov, I.V. (2005) Semiautomated cell-free conversion of prion protein: Applications for high-throughput screening of potential antiprion drugs. *Anal. Biochem.* **339**: 165-73.

Buey, R.M., Barasoain, I., Jackson, E., Meyer, A., Giannakakou, P., Paterson, I., Mooberry, S., Andreu, J.M. and Diaz, J.F. (2005) Microtubule interactions with chemically diverse stabilizing agents: thermodynamics of binding to the paclitaxel site predicts cytotoxicity. *Chem. Biol.* **12**: 1269-79.

Calzolai, L. and Zahn, R. (2003) Influence of pH on NMR structure and stability of the human prion protein globular domain. *J. Biol. Chem.* **278**: 35592-6.

Carvalho, P., Tirnauer, J.S. and Pellman, D. (2003) Surfing on microtubule ends. *Trends Cell Biol.* **13**: 229-37.

Cashman, N.R. and Caughey, B. (2004) Prion diseases – close to effective therapy? *Nat. Rev. Drug Discov.* **3**: 874-84.

Caughey, B. and Baron, G.S. (2006) Prions and their partners in crime. *Nature* **443**: 803-10.

Caughey, B., Caughey, W.S., Kocisko, D.A., Lee, K.S., Silveira, J.R. and Morrey, J.D. (2006) Prions and transmissible spongiform encephalopathy (TSE) chemotherapeutics: A common mechanism for anti-TSE compounds? *Acc. Chem. Res.* **39**: 646-53.

Chen, Y.D. and Hill, T.L. (1985) Monte Carlo study of the GTP cap in a five-start helix model of a microtubule. *Proc. Natl. Acad. Sci. USA* **82**: 1131-5.

Cunningham, C., Deacon, R., Wells, H., Boche, D., Waters, S., Diniz, C. P., Scott, H., Rawlins, J.N. and Perry, V.H. (2003) Synaptic changes characterize early behavioural signs in the ME7 model of murine prion disease. *Eur. J. Neurosci.* **17**: 2147-55.

Dimcheff, D.E., Portis, J.L. and Caughey, B. (2003) Prion proteins meet protein quality control. *Trends Cell Biol.* **13**: 337-40.

Dobson, C.M. (2001) The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**: 133-45.

Ferguson, N., Becker, J., Tidow, H., Tremmel, S., Sharpe, T.D., Krause, G., Flinders, J., Petrovich, M., Berriman, J., Oschkinat, H. et al. (2006) General structural motifs of amyloid protofilaments. *Proc. Natl. Acad. Sci. USA* **103**: 16248-53.

Goldsbury, C., Frey, P., Olivieri, V., Aebi, U. and Muller, S.A. (2005) Multiple assembly pathways underlie amyloid-beta fibril polymorphisms. *J. Mol. Biol.* **352**: 282-98.

Govaerts, C., Wille, H., Prusiner, S.B. and Cohen, F.E. (2004) Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc. Natl. Acad. Sci. USA* **101**: 8342-7.

Griffith, J.S. (1967) Self-replication and scrapie. *Nature* **215**: 1043-4.

Haire, L.F., Whyte, S.M., Vasisht, N., Gill, A.C., Verma, C., Dodson, E.J., Dodson, G.G. and Bayley, P.M. (2004) The crystal structure of the globular domain of sheep prion protein. *J. Mol. Biol.* **336**: 1175-83.

Ionescu-Zanetti, C., Khurana, R., Gillespie, J.R., Petrick, J.S., Trabachino, L.C., Minert, L.J., Carter, S.A. and Fink, A.L. (1999) Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy. *Proc. Natl. Acad. Sci. USA* **96**: 13175-9.

Jamin, N., Coic, Y.M., Landon, C., Ovtracht, L., Baleux, F., Neumann, J.M. and Sanson, A. (2002) Most of the structural elements of the globular domain of murine prion protein form fibrils with predominant beta-sheet structure. *FEBS Lett.* **529**: 256-60.

Janosi, I.M., Chretien, D. and Flyvbjerg, H. (2002) Structural microtubule cap: stability, catastrophe, rescue, and third state. *Biophys. J.* **83**: 1317-30.

Kelly, J.W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior. *Curr. Opin. Struct. Biol.* **6**: 11-7.

Kirschner, M. and Mitchison, T. (1986) Beyond self-assembly: from microtubules to morphogenesis. *Cell* **45**: 329-42.

Knaus, K.J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W.K. and Yee, V.C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat. Struct. Biol.* **8**: 770-4.

Legname, G., Nguyen, H.O., Baskakov, I.V., Cohen, F.E., Dearmond, S.J. and Prusiner, S.B. (2005) Strain-specified characteristics of mouse synthetic prions. *Proc. Natl. Acad. Sci. USA*.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D. and Riek, R. (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. USA* **102**: 17342-7.

Ma, J., Wollmann, R. and Lindquist, S. (2002) Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**: 1781-5.

Makin, O.S., Atkins, E., Sikorski, P., Johansson, J. and Serpell, L.C. (2005) Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. USA* **102**: 315-20.

Makin, O.S. and Serpell, L.C. (2005) Structures for amyloid fibrils. *FEBS J.* **272**: 5950-61.

Martin, S.R., Schilstra, M.J. and Bayley, P.M. (1993) Dynamic instability of microtubules: Monte Carlo simulation and application to different types of microtubule lattice. *Biophys. J.* **65**: 578-96.

McKinley, M.P., Meyer, R.K., Kenaga, L., Rahbar, F., Cotter, R., Serban, A. and Prusiner, S.B. (1991) Scrapie prion rod formation *in vitro* requires both detergent extraction and limited proteolysis. *J. Virol.* **65**: 1340-51.

Mitchison, T. and Kirschner, M. (1984) Dynamic instability of microtubule growth. *Nature* **312**: 237-42.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R. and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**: 773-8.

Nogales, E. and Wang, H.W. (2006a) Structural mechanisms underlying nucleotide-dependent self-assembly of tubulin and its relatives. *Curr. Opin. Struct. Biol.* **16**: 221-9.

Nogales, E. and Wang, H.W. (2006b) Structural intermediates in microtubule assembly and disassembly: how and why? *Curr. Opin. Cell. Biol.* **18**: 179-84.

Nunziante, M., Gilch, S. and Schatzl, H.M. (2003) Prion diseases: from molecular biology to intervention strategies. *Chembiochem.* **4**: 1268-84.

Okura, Y., Miyakoshi, A., Kohyama, K., Park, I.K., Staufenbiel, M. and Matsumoto, Y. (2006) Nonviral A{beta} DNA vaccine therapy against Alzheimer's disease: Long-term effects and safety. *Proc. Natl. Acad. Sci. USA*. **103**: 9619-24.

Oosawa, F. and Asakura, S. (1975) *Thermodynamics of the polymerization of proteins.* Academic Press, London.

Petkova, A.T., Buntkowsky, G., Dyda, F., Leapman, R.D., Yau, W.M. and Tycko, R. (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J. Mol. Biol.* **335**: 247-60.

Petkova, A.T., Yau, W.M. and Tycko, R. (2006) Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry* **45**: 498-512.

Prusiner, S.B. (1993) Transgenetic investigations of prion diseases of humans and animals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **339**: 239-54.

Prusiner, S.B. (1998) Prions. *Proc. Natl. Acad. Sci. USA* **95**: 13363-83.

Rezaei, H., Eghiaian, F., Perez, J., Doublet, B., Choiset, Y., Haertle, T. and Grosclaude, J. (2005) Sequential generation of two structurally distinct ovine prion protein soluble oligomers displaying different biochemical reactivities. *J. Mol. Biol.* **347**: 665-79.

Serpell, L.C., Blake, C.C. and Fraser, P.E. (2000a) Molecular structure of a fibrillar Alzheimer's A beta fragment. *Biochemistry* **39**: 13269-75.

Serpell, L.C., Sunde, M., Benson, M.D., Tennent, G.A., Pepys, M.B. and Fraser, P.E. (2000b) The protofilament substructure of amyloid fibrils. *J. Mol. Biol.* **300**: 1033-9.

Silveira, J.R., Raymond, G.J., Hughson, A.G., Race, R.E., Sim, V.L., Hayes, S.F. and Caughey, B. (2005) The most infectious prion protein particles. *Nature* **437**: 257-61.

Souillac, P.O., Uversky, V.N. and Fink, A.L. (2003) Structural transformations of oligomeric intermediates in the fibrillation of the immunoglobulin light chain LEN. *Biochemistry* **42**: 8094-104.

Stolz, M., Stoffler, D., Aebi, U. and Goldsbury, C. (2000) Monitoring biomolecular interactions by time-lapse atomic force microscopy. *J. Struct. Biol.* **131**: 171-80.

Supattapone, S., Bosque, P., Muramoto, T., Wille, H., Aagaard, C., Peretz, D., Nguyen, H.O., Heinrich, C., Torchia, M., Safar, J. et al. (1999) Prion protein of 106 residues creates an artifical transmission barrier for prion replication in transgenic mice. *Cell* **96**: 869-78.

Tanaka, M., Collins, S.R., Toyama, B.H. and Weissman, J.S. (2006) The physical basis of how prion conformations determine strain phenotypes. *Nature* **442**: 585-9.

Tattum, M.H., Cohen-Krausz, S., Khalili-Shirazi, A., Jackson, G.S., Orlova, E.V., Collinge, J., Clarke, A.R. and Saibil, H.R. (2006) Elongated oligomers assemble into mammalian PrP amyloid fibrils. *J. Mol. Biol.* **357**: 975-85.

Tuite, M.F. and Cox, B.S. (2006) The [PSI+] prion of yeast: a problem of inheritance. *Methods* **39**: 9-22.

Watzlawik, J., Skora, L., Frense, D., Griesinger, C., Zweckstetter, M., Schulz-Schaeffer, W.J. and Kramer, M.L. (2006) Prion protein Helix1 promotes aggregation but is not converted into beta-sheet. *J. Biol. Chem.* **281**: 30242-50.

Weissmann, C. (1999) Molecular genetics of transmissible spongiform encephalopathies. *J. Biol. Chem.* **274**: 3-6.

Weissmann, C. (2005) Birth of a prion: spontaneous generation revisited. *Cell* **122**: 165-8.

Welker, E., Raymond, L.D., Scheraga, H.A. and Caughey, B. (2002) Intramolecular versus intermolecular disulfide bonds in prion proteins. *J. Biol. Chem.* **277**: 33477-81.

Westermark, P. (2005) Aspects on human amyloid forms and their fibril polypeptides. *FEBS J.* **272**: 5942-9.

Westermark, P., Engstrom, U., Johnson, K.H., Westermark, G.T. and Betsholtz, C. (1990) Islet amyloid polypeptide: pinpointing amino acid residues linked to amyloid fibril formation. *Proc. Natl. Acad. Sci. USA* **87**: 5036-40.

Wickner, R.B., Edskes, H.K., Ross, E.D., Pierce, M.M., Baxa, U., Brachmann, A. and Shewmaker, F. (2004) Prion genetics: new rules for a new kind of gene. *Annu. Rev. Genet.* **38**: 681-707.

Wille, H., Michelitsch, M.D., Guenebaut, V., Supattapone, S., Serban, A., Cohen, F.E., Agard, D.A. and Prusiner, S.B. (2002) Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. USA* **99**: 3563-8.

Wu, X., Xiang, X. and Hammer, J.A., 3rd. (2006) Motor proteins at the microtubule plus-end. *Trends Cell Biol.* **16**: 135-43.

Wuthrich, K. and Riek, R. (2001) Three-dimensional structures of prion proteins. *Adv. Protein Chem.* **57**: 55-82.

This page intentionally blank

Hepatitis B Virus (HBV): The Life-cycle and Assembly of a Complex Virus

P. JONATHAN G. BUTLER
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK,
Phone: +44 1223 402296, E-mail: pjgb@mrc-lmb.cam.uk

Abstract

The overall structure of hepatitis B virus (HBV) is briefly described, followed by a discussion of its infectious life-cycle. This starts with entry into the cell, transport of the partially-double stranded viral DNA to the nucleus, its uncoating and "repair" by cellular polymerases to produce the covalently-closed circular DNA. Transcription and translation of the viral genes, by cellular factors, is discussed, together with the roles of the viral proteins and full-length "pre-genomic" RNA, and the assembly of the latter, together with the viral polymerase, into immature cores.

Subsequent synthesis of the viral DNA, which precedes and causes a structural transition, from the immature to the mature core, is described. The structures of both the immature and mature core are discussed, together with a hypothesis of how this transition is brought about, and how the maturation is necessary for interaction with the surface protein, to allow budding and production of further infectious virions.

Keywords: HBV, Hepatitis B virus, Viral cores, Structure, Virus assembly

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 131–151.

1. INTRODUCTION

Hepatitis B virus (HBV) is a major human pathogen, infecting over 70% of the population, in areas of high endemicity, by the age of 40. It causes not only a serious acute illness but also, in over 10% of cases, leads to carrier status, with chronic infection which can result in cirrhosis and primary liver cancer. Even in low endemicity areas, up to 20% will be infected during their lifetime and 1–2% will be carriers. HBV is the prototype hepadnavirus, with related viruses infecting both mammals and birds. HBV consists of an inner nucleocapsid or "core", containing the viral nucleic acid and polymerase, surrounded by a lipid envelope, in which are the virally encoded surface proteins.

Hepadnaviruses are *para*-retroviruses and, unlike the classic retroviruses in which the virion contains RNA and reverse transcription only occurs in the newly infected cell (Varmus 1988), reverse transcription occurs in the virus-producing cell, from the "pre-genomic RNA", before the now mature cores are enveloped to yield virus. HBV contains an open circular DNA genome, which is partially double-stranded (a full length minus-strand and a variable length plus-strand, typically about ¾ length) (Delius et al. 1983), and contains overlapping protein reading frames over much of its length (Fig. 1.1) – corresponding to a high information density. The life-cycle and assembly are complicated (Ganem 1991; Ganem and Varmus 1987; Nassal 1996) and are outlined below, before discussing the maturation of the core in some detail, and only a brief overview of the most salient points will be given here. Much of what is known about viral assembly has been learnt from studies on other hepadnaviruses, as these are more readily manipulated than the human pathogen.

As with a classic retrovirus, assembly of the cores occurs around a complex of the pre-genomic RNA and the viral polymerase. The major difference is that with HBV, unlike a classic retrovirus, there is strong selection against envelopment of the newly synthesised cores, until the polymerase has synthesised at least the minus-strand of the DNA. This has been predicted to result in some structural change in the core which allows its packaging (Summers and Mason 1982). This observation was confirmed and extended to show that the RnaseH activity of the polymerase was not needed, so the DNA plus-strand is not directly necessary (Gerelsaikhan et al. 1996; Wei et al. 1996). The latter authors also quantitated the immature DNA-containing cores, showing that they comprised 4% in the virions, but 18% in the cytoplasm of the producer cells. However, with the RnaseH mutant, the virion fraction rose to 50%. These observations were partially

explained when it was shown (Perlman and Hu 2003) that the immature DNA was in fact one strand of an RNA/DNA hybrid and that the important factors are the presence of one DNA strand and a largely double-stranded nucleic acid. This suggests that the structural transition requires a relatively stiff double-stranded nucleic acid chain, rather than a more flexible single strand.

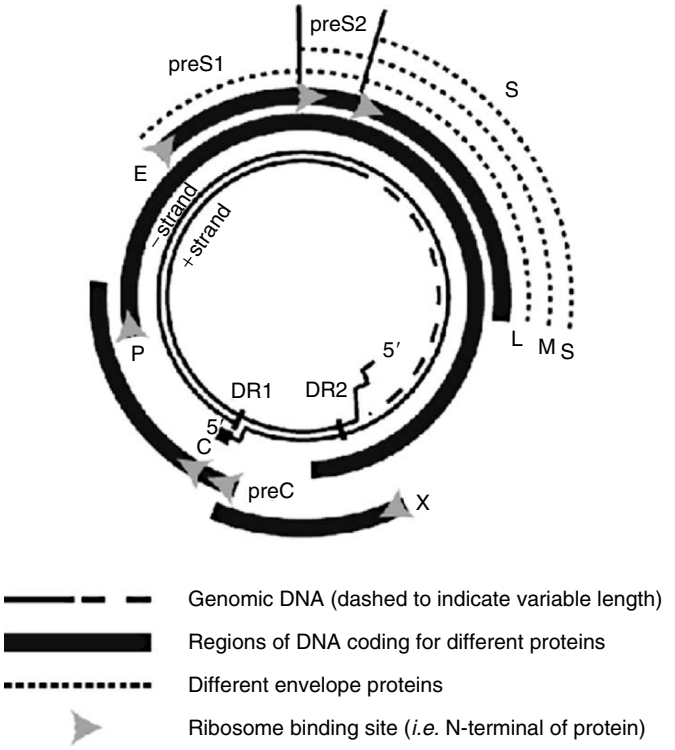


Figure 1.1. Genomic map for hepatitis B virus. The partially double-stranded DNA is shown (the minus- and plus-strands are indicated), with the variably extended region of the plus-strand shown dashed, the direct repeats (DR1, which occurs at each end of the minus-strand, and DR2) marked and the RNA primer shown as a zig-zag. Outside this are shown the coding regions, with the ribosome binding sites (from which the proteins will be synthesised) shown as grey arrows. These overlap over most of the genome and are labelled C (and preC) for the core (and pre-core), P for the polymerase, E for the envelope proteins (with the large, medium and small surface proteins shown outside as dotted lines, to show the preS1 and preS2 at the N-termini) and X for protein X.

2. THE INFECTIOUS CYCLE

2.1. Infection and establishment of HBV component synthesis in infected cell

HBV appears to infect only hepatocytes and, although the viral receptor has not been identified, the receptor binding domain is on the pre-S1 region (see below) of the surface protein (Neurath et al. 1986) which is exposed on the surface of the virion. Infection occurs by fusion of the viral membrane and cellular plasma membrane, with the core being released into the cytoplasm. It is probable that the intact core is transported to the nucleus, before loss of the core protein to expose the partially double-stranded viral DNA. The nick in the minus-strand and the gap in the plus-strand are then repaired by normal cellular enzymes, to produce the covalently-closed circular DNA in the nucleus.

Covalently-closed circular DNA is transcribed by cellular RNA polymerase, from four promoters, two for surface protein, one for protein X, and one for both the core protein and polymerase (Fig. 1.1). As there is only a single poly-A site, the four mRNAs produced all end with the same 3'-terminus. The longest transcript, coding for the core protein and polymerase (pol), is transcribed through its promoter and contains 110% of the genome, with a repeated sequence at its ends. As well as acting as an mRNA, this is also the pre-genomic RNA which is packaged into immature cores, together with pol. Translation of the various mRNAs is somewhat unusual, in that several of them are read from more than one ribosome binding site, which may be either in reading phase, to produce co-terminal proteins with a common C-terminus, or in different reading-phases, to produce distinct proteins. The different protein products are considered below.

2.2. The specific proteins coded in the HBV genome

2.2.1. Envelope proteins

mRNAs for the envelope proteins are transcribed from two promoters, with the shorter mRNA having two ribosome binding sites, but all the sites are in the same reading phase so that the C-terminal sequence of the shortest protein (S) is repeated in both the middle (M) and long (L) proteins, with the addition of N-terminal pre-S2 and pre-S1+pre-S2 regions respectively S and M proteins are typical membrane proteins, with a Type I and an internal Type II signal, so that they are co-translationally inserted into the membrane of the endoplasmic reticulum, with their N-termini in the lumen.

A sequence in the pre-S1 region of the L protein, called the "cytosolic anchorage determinant", interacts with a cellular chaperonin (Hsc70) (Prange et al. 1999) and retains pre-S1, pre-S2 and the most N-terminal transmembrane helix of S in the cytosol during translation. However, post-translationally approximately half of the L chains refold, with these sequences being translocated across the membrane (Bruss et al. 1994), possibly again with the assistance of chaperonins. This dual morphology of L is important, since the pre-S1 region is exposed on the surface of virions (Heermann et al. 1984) and binds the cellular receptor, thus determining the ability of the virus to enter and infect fresh hepatocytes (Ishikawa and Ganem 1995); but it is also the region of the envelope protein which interacts with cores (Gerhardt and Bruss 1995), to enable their envelopment and the production of mature virions, and therefore must be exposed in the cytosol, which corresponds topologically to the inside of the viral envelope, as well.

2.2.2. Protein X

Undoubtedly the least understood protein of HBV is protein X (Hbx), which is not incorporated into the virions, but has been suggested to be involved in many control processes in the infected cell, all of which would result in moderately increased synthesis of viral proteins and also possible inhibition of apoptosis of the infected cell (Bouchard and Schneider 2004).

Probably the two best established direct interactions of Hbx are with the transcriptional activator CREB (Maguire et al. 1991; Williams and Andrisani 1995) and in stimulating release of Ca^{++} from mitochondria (Oh et al. 2003), leading to activation of Src kinases and many nuclear transcription factors. Other interactions have been suggested, but the evidence for these is still relatively weak.

2.2.3. Polymerase

The polymerase of HBV (pol) is typical of a retrovirus, in having both reverse transcriptase and RNase H activities, and it also contains a terminal protein (TP) domain with a tyrosine residue which acts as the primer for addition of the first deoxynucleotide triphosphate during synthesis of the minus-strand of the viral DNA. Pol binds to a specific site ("ε") which is located just 3' to the 5'-direct repeat (DR1) on the pre-genomic RNA. The RNA/pol complex is encapsidated by core protein, to produce the immature cores which subsequently mature with the synthesis of the viral DNA. These processes are considered in more detail below.

2.2.4. Core protein gene products

The core protein gene has two ribosome binding sites, the downstream one resulting in the core protein, while the upstream one is in the same reading phase and results in the addition of a "pre-core" sequence at the N-terminus, resulting in the HbeAg. HbeAg is a secreted protein and it is thought that its main effect is to act as a T cell tolerogen and regulate the immune response against the intracellular capsid (Chen et al. 2003). This HbeAg-mediated immune regulation may predispose to chronicity during perinatal infections and also prevent severe liver injury during adult infections. Significantly for a patient who becomes a carrier, loss of HbeAg increases the likelihood of liver damage and progression to hepatocarcinoma.

The major product from this gene is the core protein (Hbc), which forms the coat of the viral cores (or "capsids"). Core protein expressed in *E. coli* spontaneously forms spherical shells which closely resemble the cores (Cohen et al. 1982), even to the extent of encapsidating RNA (Birnbaum and Nassal 1990; Hatton et al. 1992; Ulrich et al. 1993; Zheng et al. 1992), although not DNA, and this RNA has been shown to contain a significant proportion of partially degraded mRNA for the core protein (Roseman et al. 2005). The structure of the core particles has been investigated in some detail and is discussed below, including the events involved in maturation of cores so that they will interact with the surface protein to form the virions.

2.3. Assembly of cores and DNA synthesis

2.3.1. Core assembly

Assembly of the immature core starts with binding of pol to a short, kinked hairpin element located on the pre-genomic RNA just downstream of the 5'-DR1 sequence (Fig. 2.2A). This DR sequence also occurs twice, in the same orientation, at the 3'-terminus of pre-genomic RNA, where it is known as DR1 and DR2 (Fig. 2.2B, 1). The pol/RNA complex acts as the nucleus for assembly of core protein, and under physiological conditions core protein does not assemble into significant numbers of empty capsids. Probably due to the size of the complex, cores containing pre-genomic RNA plus pol are largely T = 4, with 240 core protein subunits (Roseman et al. 2005). These RNA-containing cores are immature and are not readily enveloped, rather pol is activated by binding to ε and encapsidation (Tavis and Ganem 1996), and synthesis of the DNA minus-strand, which is complementary to the pre-genomic RNA, commences.

2.3.2. Synthesis of minus-strand DNA

Immature cores are assembled with pol bound to ε near the 5'-terminus of the pre-genomic RNA. DNA synthesis starts copying the specific sequence in ε, with the first deoxynucleotide triphosphate added directly onto a tyrosine residue in the TP domain of pol (Fig. 2.2A), and proceeds through the 5'-DR1 sequence to the 5'-terminus of the pre-genomic RNA, where it must stall for want of any further template (Fig. 2.2B, 1). At this point the template shift typical of any retrovirus occurs, with the primer DNA shifting from the 5'-DR1 onto the 3'-DR1 (Fig. 2.2B, 2), allowing further elongation of the primer (Fig. 2.2B, 3). As the minus-strand DNA grows, the RNase H function of pol will start to degrade the pre-genomic RNA which has already been reverse transcribed, and is therefore now in an RNA/DNA double helix which is the substrate for RNase H (Fig. 2.2B, 4).

When pol reaches the 5'-terminus of the pre-genomic RNA for the second time, it will have generated a minus-strand DNA with terminal repeats in the same orientation; and also a short RNA fragment, remaining from the RNase H degradation of the template, will still be hybridised to the 3'-DR1 sequence (of the minus-strand) (Fig. 2.2B, 5). This is the initial primer for the plus-strand synthesis.

2.3.3. Synthesis of plus-trand DNA

Synthesis of the DNA plus-strand starts with a folding of the single-strand DNA, bringing the ends close together (Fig. 2.2C,1) and allowing the RNA-fragment primer to shift from the 3'-DR1 onto DR2 of the minus-strand (Fig. 2.2C, 2). It is a matter of speculation whether the bound pol TP domain at the 5'-terminus of the minus-strand may obstruct binding of the primer onto the DR1 immediately alongside DR2, but in practice primer extension then occurs along the template which includes this DR1 up to its 5'-terminus, to generate a short segment of plus-strand with two complements of the DR sequence in tandem (Fig. 2.2C, 3). A template jump then occurs, with the 3'-DR1 sequence on the minus-strand displacing the 5'-DR1, to circularize the DNA complex and release the 5'-tail with its attached pol (Fig. 2.2C, 4). It can be hypothesised that the presence of this terminal protein may be responsible for a weaker bonding between the complementary sequences of the DR, thus not only steering the primer onto DR2, but also facilitating the template jump between DR1's.

With the 3'-end of the primer now bound to the 3'-DR1 of the minus-strand, plus-strand elongation can readily occur, again catalysed by the reverse transcriptase function of pol (Fig. 2.2C, 5). These cores are still

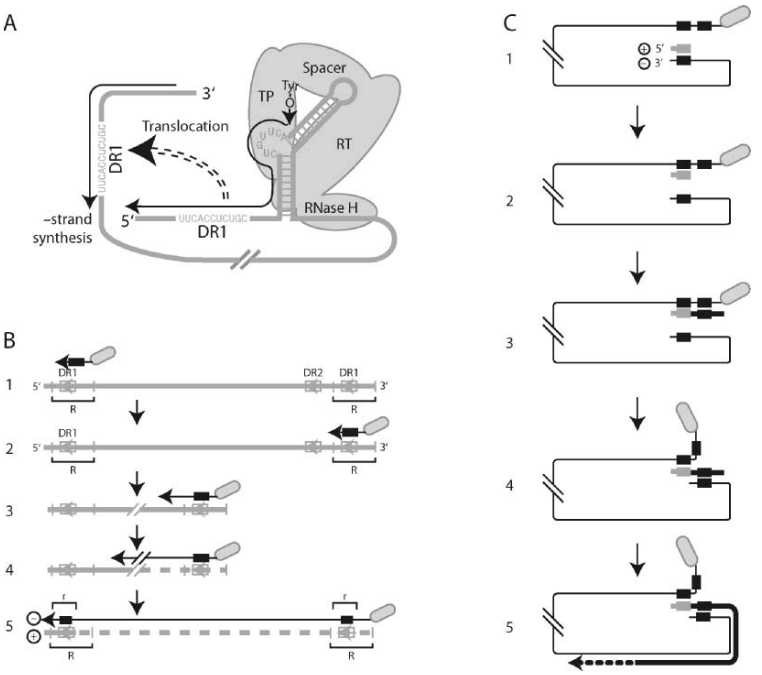


Figure 2.2. Synthesis of minus- and plus-strands of HBV genome from pre-genomic RNA by pol. Pol is shown in light density, RNA in medium density and DNA in black, with a thin line for the minus-strand and a thick line for the plus-strand. Degraded RNA is shown by a broken line in B; the variable extension of the plus-strand DNA as a dotted line in C.

A: Priming of minus-strand synthesis.

Pol binds to the ε signal on pre-genomic RNA and starts synthesis along the 5'-DR1, using a specific tyrosine residue in its TP domain as primer. Subsequently the short DNA fragment is translocated onto the 3'-DR1, allowing synthesis along the entire genome, with a short terminal repeat.

B: minus-strand synthesis and degradation of pre-genomic RNA.

1: priming on 5'-DR1, as described in A above;

2: result of translocation of short minus-strand (primer) fragment from 5'- to 3'-DR1;

3: start of elongation from primer on 3'-DR1;

4: continuing elongation of minus-strand, with degradation of pre-genomic RNA by RNase H function of pol;

5: end product of minus-strand synthesis, with single-stranded DNA and a short fragment of RNA, complementary to DR sequence, left from RNase H degradation.

C: plus-strand synthesis, with circularization of HBV genome.

1: product from minus-strand synthesis, with short primer of RNA hybridised to 3'-DR1, folding to bring termini close together;

2: primer translocation from 3'-DR1 (of the minus-strand) onto DR2, which is located close, and 3'-, to the 5'-DR1;

3: extension of primer, producing plus-strand, from DR2 through the 5'-DR1;

immature and do not normally interact with the cytoplasmic tail of the surface protein, until significant double-stranded DNA has been formed. However, when this has reached the stage of somewhere between one-half and three-quarters of the genome (Delius et al. 1983) envelopment does occur, with the cores budding into the endoplasmic reticulum to form virions, and thus terminating the supply of deoxynucleotide triphosphates and further plus-strand elongation.

3. MATURATION AND ENVELOPMENT OF CORES

3.1. Balance between immature and mature cores

It is evident that HBV particles contain partially double-stranded DNA (Delius et al. 1983), rather than RNA or even single-stranded DNA, and the hypothesis that control of the interaction between cores and the surface protein would depend upon a structural change in the core, brought about by the DNA synthesis, was originally proposed by Summers and Mason (Summers and Mason 1982). It had also been observed that common missense mutations in the core protein of HBV, whether naturally occurring (in patients who have become carriers) or engineered in the laboratory, lead to changes in the envelopment and budding of cores (Table 1) (although the precise effect of a mutation depends upon the HBV strain) and the phenol-type is not due to either deficiency in DNA synthesis (Yuan et al. 1999) or instability in the cores (Newman et al. 2003). In particular, mutation at codon 97 (which is close to the bend in the bent helix in the spike – see discussion below) results in significant loss of the selectivity for encapsidation of only DNA containing cores (Yuan et al. 1999) which can, in part at least, be compensated for by other core mutations, which on their own reduce the envelopment of cores (Table 1). The importance of the detailed interaction between the core protein and the pre-S1 region of the surface protein is shown by a compensating mutation, at residue 119 in the L-surface protein (Le Pogam and Shih 2002), which restores the selectivity. Perhaps surprisingly, mutations of residues near the tip of the spikes have minimal

4: template shift, with 3'-DR1 displacing 5'-DR1, to circularize DNA and position primer close to 3'-terminus of minus-strand;

5: plus-strand extension from primer, generating variety of lengths of resulting plus-strand, depending upon length when envelopment of core occurs, which will exclude further deoxynucleotide triphosphates and thus stop elongation.

Table 1. Residues affecting core budding. Naturally occurring mutations are shown in bold type, others are laboratory derived.

Mutation	Effect of mutation on budding of core
I97L	Immature cores in *pseudo*-virions, containing single-stranded DNA (Yuan et al. 1999)
P5T	Very low secretion of virus, with very few cores budding (Le Pogram et al. 2000); corrects for I97L mutation (Chua et al. 2003)
L60V	Very low secretion of virus, with very few cores budding (Le Pogram et al. 2000)
L95A	Intracytoplasmic cores which do not become enveloped (Ponsel and Bruss 2003)
K96A	Intracytoplasmic cores which do not become enveloped (Ponsel and Bruss 2003)
L60A	Intracytoplasmic cores which do not become enveloped (Ponsel and Bruss 2003)
P130T	Low secretion of virus, with few cores budding; corrects for I97L mutation (Yuan and Shih 2000)

effect upon encapsidation of the cores, while a number of residues in the shell do affect this interaction, but not core formation (Ponsel and Bruss 2003).

It is thus evident that point mutations in the core protein appear to be able to tip the balance point of nucleic acid content at which a core ceases to remain largely cytoplasmic and becomes able to interact with the cytoplasmic tail of the surface protein and bud into the endoplasmic reticulum, acquiring a lipoprotein membrane. If the core contains a correct complement of DNA, rather than RNA, and is therefore truly mature, the result is an infectious virion, otherwise the *pseudo*-virion is non-infectious as the core was immature and did not contain the correct nucleic acid, even though its surface had interacted as though it were mature.

3.2. The structures of immature and mature cores

3.2.1. The immature core

The earliest detailed work on the structure of cores was carried out on cores which spontaneously assemble from protein expressed in *E. coli*. The structure was determined by electron cryomicroscopy and cores were found to contain both T = 3 and T = 4 particles (see Fig. 3.3, left hand columns) with 2-fold spikes projecting out from the underlying shell, which still has holes which would allow small molecules to enter or leave the inside of the core (Crowther et al. 1994). In these *E. coli*-produced cores, the proportion of T = 3 and T = 4 forms is strongly determined by both the strain of HBV and

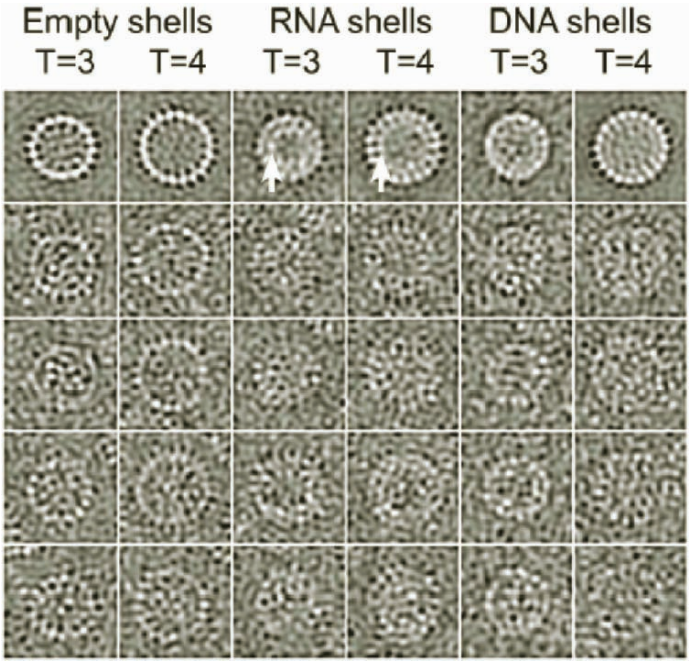


Figure 3.3. Electron cryomicroscopic images of various core structures (from reference Roseman et al. (2005)). The top row shows averaged aligned images, to show the general features with high contrast and reduced noise. The other rows show individual particles. Both T = 3 and T = 4 particles are shown for empty shells, produced in *E. coli* from core protein truncated at residue 149, for RNA-containing shells, again produced in *E. coli* but from full length protein, and for DNA-containing shells, isolated from mature virions. Arrows mark the clearly visible inner shell of density seen most clearly in RNA-containing cores. Box size = 480 Å.

also the precise point of the C-terminal truncation (Zlotnick et al. 1996). The structures of duck hepatitis B virus cores and of HBV cores, both with the protein expressed in *E. coli* and also isolated from infected livers, were shown to be substantially similar (Kenney et al. 1995), with a ratio of ~13:1 for T = 4 to T = 3 cores in the liver material. Extension of the resolution of the structures, again by electron cryomicroscopy, allowed the fold of the polypeptide chain to be traced and showed that the spikes consist of 2 pairs of α-helices, folded back at the tip of the spike and with one helix slightly bent (Böttcher et al. 1997; Conway et al. 1997), with a further α-helix lying in the shell. It was also shown that the highly basic C-terminal tail lies inside the shell, in a position to interact with the negatively charged nucleic acid (Zlotnick et al. 1997). Cores, with the protein truncated at residue 149, were crystallised (Wynne et al. 1999) and the structure solved by X-ray crystallography to allow an atomic model of the CW strain of HBV cores to be built (Wynne et al. 1999), which closely confirmed the secondary structure deduced from the electron cryomicroscopy (Böttcher et al. 1997) (Fig. 3.4h).

More recently cores containing RNA (Fig. 3.3, central columns) have also been determined to a resolution of 7.6 Å by electron cryomicroscopy (Fig. 3.4a) (Roseman et al. 2005) and it was shown that their structure was very similar to that of the empty cores, which had been studied previously, with the obvious exception that they contain an additional shell of material, inside the capsid shell (Fig. 3.3, central columns; Fig. 3.4a and b). This was ascribed to the RNA and the basic C-terminal peptide tails, which are thought to bind, and partially neutralise, the negative charge of the nucleic acid. Interestingly, more of the RNA in these cores than might have been expected was found to be comprised of partially degraded core protein mRNA (Roseman et al. 2005).

The inner shell is linked to the shell of the capsid by a number of slender links (Fig. 3.4b and c), located around the icosahedral 5-fold and 2-fold (locally *pseudo*-6-fold) axes. When the trace of the peptide backbone, determined from the X-ray crystallographic structure, was mapped into the density for the RNA-containing core, determined by electron cryomicroscopy, it was found that the positions of the most C-terminal ordered residues in the truncated protein (T142 in the A subunit and L143 in the B subunit) were located at the outer ends of the links (Fig. 3.4d), strongly supporting the hypothesis that these links may correspond to the C-terminal peptide tails, extending in from the outer capsid to interact with the nucleic acid.

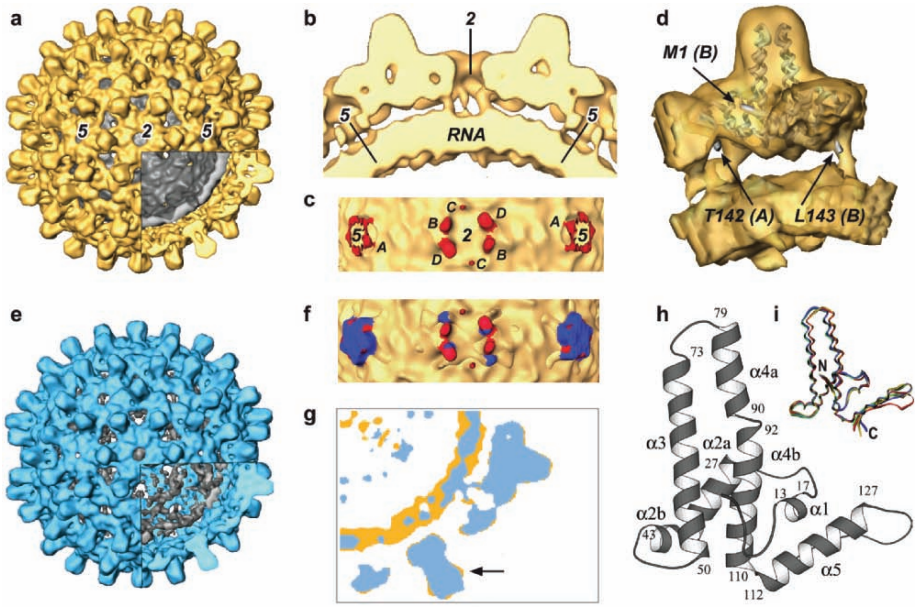


Figure 3.4. Maps of the T = 4 RNA- and DNA-containing core particles (from reference Roseman et al. (2005)). (a) the RNA-containing map with the outer parts of the protein coloured yellow and a segment of this cut away to show the inner shell (coloured grey), containing the RNA and parts of the C-terminal tails of the proteins; (b) a central slab from the map, with the 2-fold and 5-fold axes indicated, showing the slender links between the inside of the outer protein shell and the outside of the inner, RNA-containing shell; (c) radial view of the outside of the inner shell (yellow) with the links highlighted in red. This shows the symmetry of 5 links around each 5-fold axis and 6 links around each 2-fold (locally 6-fold) axis, corresponding to the T = 4 *quasi*-equivalence of the core protein subunits. The identity of the subunit for each link is labelled A, B, C and D; (d) one protein dimer (the A-B dimer) of the RNA-containing core, with the ribbon trace for the peptide chains (from the crystallographic structure) superimposed. The positions of the most C-terminal ordered residues in the crystal (T142(A) and L143(B)) are shown, together with the N-terminal methionine of the B chain; (e) the DNA-containing map with the outer parts of the protein coloured blue and a segment of this cut away to show the inner shell (coloured grey), containing the DNA and parts of the C-terminal tails of the proteins. This shows the close similarity to the RNA-containing map shown in (a); (f) superposition of link region from the RNA- and DNA-containing maps, showing the nucleic acid containing shell (yellow) and the links for the RNA (red) and DNA (blue); (g) superimposed central sections of the RNA- (yellow) and DNA-containing (blue) maps. The inner shell is stronger and better defined for the RNA than the DNA, and parts of the dimer spikes protrude sideways, showing a small change in the structure of the spike (arrowed); (h) secondary structure fold of the core protein (Wynne et al. 1999); (i) variations in the fold of the four independent subunits in the core. Subunits are aligned in their central hydrophobic regions and show two classes of conformation of the helical hairpins, with each dimer comprised of one monomer from each class.

3.2.2. The mature core

A major problem in determining the structure of mature, DNA-containing cores is obtaining suitable amounts of material for such study. The only *in vitro* system which might yield these are transfected tissue culture cells (since there are no cell lines known to be susceptible to infection), but these do not secrete sufficient amounts of virus, while one cannot be sure that DNA-containing cores isolated from their cytoplasm are correctly matured. It has therefore been necessary to isolate virus from blood of HBV carriers and then to strip off the lipoprotein membrane, to obtain authentic viral cores which are therefore, by definition, mature and capable of envelopment (Fig. 3.3, right hand columns).

One major feature is that >98% of the mature cores are T = 4, suggesting that the T = 3 cores are an artefact of assembly and not directly relevant to the assembly of virions. Using these cores, the structure of the mature, DNA-containing core has been determined, again by electron cryomicroscopy, to a resolution of 8.8 Å (Fig. 3.4e) (Roseman et al. 2005).

3.2.3. Comparison of RNA- and DNA-containing cores

As would be expected, the structures of the cores are clearly very similar, whether they contain RNA or DNA, with no change in the overall fold of the peptide chain. However a detailed comparison does show up a number of potentially significant differences (of the order of a few angstroms). The most obvious is a change in the profile of the spikes projecting from the shell of the capsid (compare Fig. 3.4a and e, also Fig. 3.4g), with those of the RNA-containing cores being bulbous or knob-shaped and wider than those in the DNA-cores, which have a more uniform and regular spike. This change in profile is readily seen when the central sections through the spikes in the maps are overlaid (Fig. 3.4g) and on comparing the changes statistically, using Student's *t* test, a number of significant features can be described.

The change in shape of the spikes, mentioned above, results from an alteration in the kink in one of the α-helices in the hairpin (see Fig. 3.4h for numbering and indication of structure), with helix-4 of subunit A being straighter in the DNA structure (for definition of the four independent subunits in the T = 4 core see reference (Wynne et al. 1999)). The close contact between all four helices in the spike means that compensating changes are required in the other helices, and changes in orientation of helix-3 of the A subunit and helix-4 of the B-subunit are seen. In the case of helix-4, the changes (and straightening) are more apparent towards the inside of the shell in subunit B, rather than towards the tip as in subunit A, and result in changes in helix-5 of the B-subunit, which forms part of the capsid

shell. There is therefore a direct interconnection between changes in this shell and those in the spike, and this interaction could, of course, be brought about in either direction, i.e. it is probable that the changes in the spike result from changes in the shell, rather than the other way round. Thus the helices in the spike can be likened to the blades of a pair of scissors which close somewhat on going from the RNA-containing core to the DNA-containing one, due to movement of the handles.

Even in the RNA-containing cores the A-B and C-D dimers in the spikes do not have strict 2-fold symmetry (they are on neither crystallographic nor icosahedral 2-fold axes), but rather occur in 2 classes of fold, with one monomer of each spike belonging to each class and being related by a 178° (rather than 180°) rotation (Fig. 3.4i). In the DNA-containing cores, the spike density deviates even further from a strict *quasi*-symmetry (Fig. 3.4g), with the A-B and C-D dimers differing further. There is additional density in the region of amino acids 140–145 of subunit-A, around the strict 5-fold axis (Fig. 3.4f), indicating that this region is more ordered than with RNA, with the interactions of helix-5 being stabilised.

A further, and possibly very important feature, is the presence of strong, significant additional density in the DNA structure in externally accessible pockets on either side of the spikes, at the interface between the subunits forming each dimer (Fig. 3.5). The structure determined by X-ray crystallography (Wynne et al. 1999) shows that these pockets are bounded by residues P5, K96 and I97 from one subunit, together with L60 from the other, all of which residues have been shown to affect budding of cores if mutated (Table 1 and references therein).

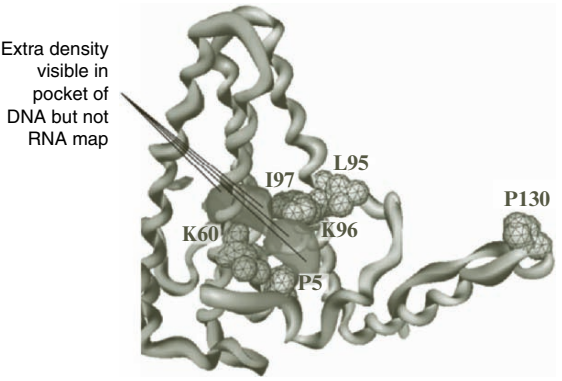


Figure 3.5. The hydrophobic pocket in the spike of the cores (adapted from reference (Roseman et al. 2005)). The functionally important residues (see Table 1) are marked, as is extra density present in the DNA- but not the RNA-containing map (shaded dark).

4. OVERVIEW OF ENVELOPMENT

4.1. Introduction

We now have a clear picture of the mature, DNA-containing cores which are capable of interacting with the surface protein, inserted into the membrane of the endoplasmic reticulum, to bring about the envelopment of the core and its budding into the reticulum, from where it can be exported from the cell by the usual exocytotic pathway. These core shells are dominantly T = 4, containing 240 core protein subunits, and with only a tiny minority of T = 3 shells. This is only slightly more tilted in favour of T = 4 compared to unenveloped cores extracted from infected liver, which also show a majority of T = 4 (Kenney et al. 1995), in contrast to the *in vitro* situation with some core proteins (particularly truncated variants) expressed in *E. coli*. importance of the detail of this interaction is highlighted by the finding that a point mutation in the pre-S1 region of the surface protein (A119F) can compensate for the I97L mutation of the core protein, and restore normal budding of only mature, DNA-containing cores (Le Pogam and Shih 2002).

4.2. Possible model for the triggering of maturation

One hypothesis of how the maturation might be triggered is based on work on the packaging of DNA into the heads of phage ϕ29. As with most phages, the DNA is packaged by a portal motor, driven by ATP hydrolysis, and it has been shown that this motor has to work increasingly hard against the bending, electrostatic and entropic forces due to close packing of the DNA into the head (Smith et al. 2001). A somewhat similar situation comes about in HBV as the pre-genomic RNA is reverse transcribed into the double-stranded DNA. During the normal pol activity, the RNA template is degraded as the DNA minus-strand is synthesised, so there is little change in nucleic acid packing density at this stage, and only as the DNA plus-strand is synthesised will double-stranded nucleic acid be formed and the density rise. We would suggest that this increasing DNA stiffness and charge density may be the trigger for the change in core structure, by exerting an outwards force onto the inner shell of the core protein capsid.

Analysis of the size of plus-strands of individual virions shows that these are typically between one-half and three-quarters of the full genome, with never less than 650–700 bases of single-stranded DNA remaining (Delius et al. 1983), i.e. in practice the transition has always occurred by this stage. Such a DNA content would give a packing density of ~70% of that found in ϕ29 and, at this density, there is considerable internal force (Smith et al.

2001), particularly when account is taken of the smaller internal radius of HBV which would result in greater bending of the double-stranded DNA, and this could well be sufficient to trigger the change in structure. This hypothesis is supported by the observation of packaging of RNA/DNA hybrids in HBV mutants with an inactivated RNase H activity (Gerelsaikhan et al. 1996), again yielding a substantially double-stranded nucleic acid in the core.

5. CONCLUSION

The picture that emerges is of a delicate balance between two structures for the core; the one containing single-stranded nucleic acid and immature, in the sense that it will not interact with the surface protein to promote budding, while the other contains significant double-stranded nucleic acid (double-stranded DNA in the infectious virus) and has matured and now proceeds to interact with the surface protein and bud, to yield virus. This balance is clearly quite fine, with many point mutations in the core protein (particularly around the pocket on the spike) being able to upset it (Table 1), and it can also be influenced by the surface protein itself, as shown by the effect of the pre-S1 A119F mutation. It is probable that the effect of these mutations (and external interactions) is expressed by a change in the internal pressure necessary to tip the cores between their states, so that the outside of the core influences the inside, as well as the obvious effect of the nucleic acid change altering the structure so that it now binds better to the surface protein.

The diverse nature and position of the functional mutations suggests that the core is a finely tuned mechanism for controlling the appropriate envelopment in response to the nucleic acid which it contains. Disturbing this balance, for example possibly by targeting the pocket, could provide an antiviral strategy in addition to the nucleotide analogues currently used to inhibit the polymerase.

6. SUMMARY

HBV is a major human pathogen, which leads to significant morbidity and mortality in many parts of the world. It is a small enveloped virus, with a partially double-stranded, circular DNA genome encapsidated together with the viral polymerase in a T = 4 icosahedral core. During entry into the cell the viral envelope is lost and the DNA is then transported to the nucleus, probably still in the intact core before this is disassembled. Normal nuclear enzymes then complete the nicked and partial DNA strands, to produce the viral covalently-closed circular DNA, which is the template for cellular

RNA polymerases to synthesise the viral mRNAs and also the pre-genomic RNA. These are exported from the nucleus and the mRNAs translated using the host machinery.

Proteins encoded in the genome include the envelope proteins, which are a series of C-terminally redundant membrane proteins, inserted into the membrane of the endoplasmic reticulum co-translationally; the viral polymerase, which is a reverse transcriptase with RNase H activity and also a TP domain which acts as the primer for reverse transcription; the core protein, which forms the viral core encapsidating the pre-genomic RNA and the polymerase; and also protein X, which is poorly characterised, but does appear to serve both to increase synthesis of HBV components in the cell and also reduce the tendency for apoptosis of the infected hepatocyte. A similar protective function may be served by the e-antigen, which is core protein with a pre-core extension at its N-terminus and a C-terminal truncation, and which appears to reduce the immune response to core protein.

Pol binds to a specific ε site on the pre-genomic RNA and this complex is then encapsidated by the core protein, to produce immature cores. Pol then acts to reverse transcribe the RNA into the minus-strand of DNA, concomitantly degrading the RNA template, and then uses the last RNA fragment as a primer for synthesis of the plus-strand DNA from the minus-strand template. These steps involve a number of template switches and jumps, which differ in detail from those occurring with a classical retrovirus, and generate the viral DNA genome. As the plus-strand is synthesised, a "pressure" inside the core capsid is generated by the bending, electrostatic and entropic forces due to closer packing of the double-stranded DNA and this tips the balance between two delicately poised structures for the core, with this falling when between one-half and three-quarters of the plus-strand has been synthesised, so that the core changes from the immature form to the mature. Some mature cores are transported into the nucleus, where they act like newly infecting cores, thus producing further copies of covalently-closed circular DNA and there is a slow build-up of this in the nucleus during an infection.

The mature core will also interact productively with the cytoplasmic tails of the surface protein, which is in the membrane of the endoplasmic reticulum, leading to budding of the core into the reticulum to give envelopment. This process removes the core from the cytosol and thereby prevents any further access of deoxynucleotide triphosphates, so that elongation of the plus-strand DNA stops, leaving the partially double-stranded state. These nascent virions are then transported to the plasma membrane of the cell, presumably along the usual exocytotic pathway from the endoplasmic reticulum, and are secreted as free virus into the blood.

Clearly this is a complex life-cycle which, while similar to that of a classic retrovirus, is in fact probably still more complex. While ~90% of

patients suffering from an acute infection will clear the virus and recover fully, a significant 10% fail to clear it and go on to a carrier status, which can result in cirrhosis and primary liver cancer. Despite the availability of an effective vaccine (Murray et al. 1984), there is therefore still a need for better therapeutic agents, to treat the large numbers of carriers and to try to reduce the one million deaths per annum.

REFERENCES

Birnbaum, F. and Nassal, M. (1990) Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* **64**: 3319-30.

Böttcher, B., Wynne, S.A. and Crowther, R.A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature (Lond.)* **386**: 88-91.

Bouchard, M.J. and Schneider, R.J. (2004) The enigmatic X gene of Hepatitis B virus. *J. Virol.* **78**: 12725-34.

Bruss, V., Lu, X., Thomssen, R. and Gerlich, W.H. (1994) Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO J.* **13**: 2273-79.

Chen, M.T., Billaud, J.-N., Sällberg, M., Guidotti, L.G., Chisari, F.V., Jones, J., Hughes, J. and Milich, D.R. (2004) A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc. Nat. Acad. Sci.* **101**: 14913-18.

Chua, P.K., Wen, Y.-M. and Shih, C. (2003) Coexistence of Two Distinct Secretion Mutations (P5T and I97L) in Hepatitis B Virus Core Produces a Wild-Type Pattern of Secretion. *J. Virol.* **77**: 7673-76.

Cohen, B.J. and Richmond, J.E. (1982) Electron microscopy of hepatitis B core antigen synthesized in *E. coli*. *Nature (Lond.)* **296**: 677-78.

Conway, J.F., Cheng, N., Zlotnick, A., Wingfield, P.T., Stahl, S.J. and Steven, A.C. (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature (Lond.)* **386**: 91-94.

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V. and Pumpens, P. (1994): Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**: 943-50.

Delius, H., Gough, N.M., Cameron, C.H. and Murray, K. (1983) Structure of the Hepatitis B virus genome. *J. Virol.* **47**: 337-43.

Ganem, D. (1991) Assembly of hepadnaviral virions and subviral particles. *Curr. Topics Microbiol. Immunobiol.* **168**: 61-83.

Ganem, D. and Varmus, H.E. (1987) The molecular biology of he hepatitis B virus. *Ann. Rev. Biochem.* **56**: 651-93.

Gerelsaikhan, T., Tavis, J. E. and Bruss, V. (1996) Hepatitis B virus nucleocapsid envelopment does not occur without genomic DNA synthesis. *J. Virol.* **70**: 4269-74.

Gerhardt, E. and Bruss, V. (1995) Phenotypic mixing of rodent but not avian hepadnavirus surface proteins into human hepatitis B virus particles. *J. Virol.* **69**: 1201-08.

Hatton, T., Zhou, S. and Standring, D.N. (1992) RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. *J. Virol.* **66**: 5232-41.

Heermann, K.H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H. and Gerlich, W. H. (1984) Large surface proteins of hepatitis B virus containing the pre-*s* sequence. *J. Virol.* **52**: 396-402.

Ishikawa, T. and Ganem, D. (1995) The pre-S domain of the large viral envelope protein determines host range in avian hepatitis B viruses. *Proc. Nat. Acad. Sci.* **92**: 6259-63.

Kenney, J.M., von Bonsdorff, C.-H., Nassal, M. and Fuller, S.D. (1995) Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. *Structure* **3**: 1009-19.

Le Pogam, S. and Shih, C. (2002) Influence of a putative intermolecular interaction between core and the pre-S1 domain of the large envelope protein on hepatitis B virus secretion. *J. Virol.* **76**: 6510-17.

Le Pogam, S., Yuan, T.T.-T., Sahu, G.K., Chatterjee, S. and Shih, C. (2000) Low-Level Secretion of Human Hepatitis B Virus Virions Caused by Two Independent, Naturally Occurring Mutations (P5T and L60V) in the Capsid Protein. *J. Virol.* **74**: 9099-105.

Maguire, H.F., Hoeffler, J.P., Kuno, K., Akiyama, M., Ikeda, B., Matshushima, K. and Murakami, S. (1991) Hepatitis B virus X protein transactivates human interleukin-8 gene through acting on nuclear factor κB and CCAAT/enhancer-binding protein-like cis elements. *J. Biol. Chem.* **266**: 13759-63.

Murray, K., Bruce, S.A., Hinnen, A., Wingfield, P., van Erd, P.M.C.A., de Reus, A. and Schellekens, H. (1984) Hepatitis B virus antigens made in microbial cells immunise against viral infection. *EMBO J.* **3**: 645-50.

Nassal, M. (1996) Hepatitis B virus morphogenesis. *Curr. Topics Microbiol. Immunobiol.* **214**: 297-337.

Neurath, A.R., Kent, S.B., Strick, N. and Parker, K. (1986) Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. *Cell* **46**: 429-36.

Newman, M., Suk, F.-M., Cajimat, M., Chua, P.K. and Shih, C. (2003) Stability and morphology comparisons of self-assembled virus-like particles from wild-type and mutant human hepatitis B virus capsid proteins. *J. Virol.* **77**: 12950-60.

Oh, J.C., Jeong, D.L., Kim, I.K. and Oh, S.H. (2003) Activation of calcium signalling by hepatitis B virus-X protein in liver cells. *Exp. Mol. Med.* **35**: 301-09.

Ostapchuk, P., Hearing, P. and Ganem, D. (1994) A dramatic shift in the transmembrane topology of a viral envelope glycoprotein accompanies hepatitis B viral morphogenesis. *EMBO J.* **13**: 1048-57.

Perlman, D. and Hu, J. (2003) Duck hepatitis B virus virion secretion requires a double-stranded DNA genome. *J. Virol.* **77**: 2287-94.

Ponsel, D. and Bruss, V. (2003) Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. *J. Virol.* **77**: 416-22.

Prange, R., Werr, M. and Löffler-Mary, H. (1999) Chaperones involved in Hepatitis B virus morphogenesis. *Biol. Chem.* **380**: 305-14.

Roseman, A.M., Berriman, J.A., Wynne, S.A., Butler, P.J.G. and Crowther, R.A. (2005) A structural model for maturation of the hepatitis B virus core. *Proc. Nat. Acad. Sci.* **102**: 15821-26.

Smith, D.E., Rans, S.J., Smith, S.B., Grimes, S., Anderson, D.L. and Bustamante, C. (2001) The bacteriophage Φ29 portal motor can package DNA against a large internal force. *Nature (Lond.)* **413**: 748-52.

Summers, J. and Mason, W.S. (1982) Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**: 403-15.

Tavis, J.E. and Ganem, D. (1996) Evidence for activation of the hepatitis B virus polymerase by binding of its RNA template. *J. Virol.* **70**: 5741-50.

Ulrich, R., Meisel, H., Soza, A., Krüger, D.H., Ladhoff, A.-M., Borisova, G. and Pumpen, P., Eds. (1993) Characterization of chimeric core particles of hepatitis B virus containing foreign epitopes. Vaccines-93, Modern approaches to new vaccines including prevention of AIDS. New York: Cold Spring Harbor Laboratory Press.

Varmus, H. (1988) Retroviruses. *Science (Wash.)* **240**: 1427-35.

Wei, Y., Tavis, J.E. and Ganem, D. (1996) Relationship between viral DNA synthesis and virion envelopment in hepatitis B viruses. *J. Virol.* **70**: 6455-58.

Williams, J.S. and Andrisani, O.M. (1995) The hepatitis B virus X protein targets the basic region-leucine zipper domain of CREB. *Proc. Nat. Acad. Sci.* **92**: 3819-23.

Wynne, S.A., Crowther, R.A. and Leslie, A.G. (1999) The crystal structure of the human hepatitis B virus capsid. *Molec. Cell* **3**: 771-80.

Wynne, S.A., Leslie, A.G.W., Butler, P.J.G. and Crowther, R.A. (1999) Crystallisation of hepatitis B virus core protein shells: determination of cryoprotectant conditions and preliminary X-ray characterization. *Acta Cryst.* **D55**: 557-60.

Yuan, T.T.-T., Sahu, G.K., Whitehead, W.E., Greenberg, R. and Shih, C. (1999) The mechanism of an immature secretion phonotype of a highly frequent naturally occurring missense mutation at codon 97 of human hepatitis B virus core antigen. *J. Virol.* **73**: 5731-40.

Yuan, T.T.-T. and Shih, C. (2000) A Frequent, Naturally Occurring Mutation (P130T) of Human Hepatitis B Virus Core Antigen Is Compensatory for Immature Secretion Phenotype of Another Frequent Variant (I97L). *J. Virol.* **74**: 4929-32.

Yuan, T.T.-T., Tai, P.-C. and Shih, C. (1999) Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen. *J. Virol.* **73**: 10122-28.

Zheng, J., Schödel, F. and Peterson, D.L. (1992) The structure of hepadnaviral core antigens. *J. Biol. Chem.* **267**: 9422-29.

Zlotnick, A., Cheng, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J. and Wingfield, P.T. (1996) Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**: 7412-21.

Zlotnick, A., Cheng, N., Stahl, S.J., Conway, J.F., Steven, A.C. and Wingfield, P.T. (1997) Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. *Proc. Nat. Acad. Sci.* **94**: 9556-61.

This page intentionally blank

Photo-oxidative Stress in the Presence of a Water-soluble Derivative of C_{60}: ESR and AFM Assays

BERTRAND VILENO[a], ANDRZEJ SIENKIEWICZ[a], MAŁGORZATA LEKKA[a,b], PIERRE R. MARCOUX[c] AND LÁSZLÓ FORRÓ[a]

[a] *Institute of Physics of Complex Matter, Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland, Phone: +41 21 693 4306, Fax: +41 21 693 4470, E-mail: laszlo.forro@epfl.ch*

[b] *The Henryk Niewodniczański Institute of Nuclear Physics, Polish Academy of Sciences, ul. Radzikowskiego 152, 31–342 Kraków, Poland*

[c] *LETI/DTBS/Equipe commune CEA – bioMérieux CEA, 17, rue des Martyrs 38054 Grenoble cedex9, France*

Abstract

Reactive oxygen species (ROS) are a family of oxygen-based highly reactive molecules, which are implicated in numerous processes of cellular metabolism, where maintaining the redox equilibrium is a prerequisite for normal functioning. Under certain conditions, however, ROS can damage important parts of cells, including proteins, membranes and DNA. At the cellular level, such pathogenic scenario, also called oxidative stress (OS), results from an imbalance between the pro-oxidative capacity of ROS and cell ability to neutralize them with antioxidants. OS-induced cellular damage has been implicated in numerous human pathologies, from brain disorder (like Alzheimer and Parkinson's diseases), to various forms of cancer (like skin melanoma) and eye pathologies (cataract, macular disease), as well as in the ageing process. OS can also be purposely induced, for e.g. in diseased tissues. This is the basis of a novel therapeutic modality called photodynamic therapy (PDT), which enables one to selectively eradicate diseased cells and tissues.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 153–180.

PDT is emerging as a promising therapeutic technique to cure various types of cancer, including skin cancers and is also used in many non-tumor related medical, dermatological and cosmetic applications. To eradicate the diseased tissues, PDT employs tissue-selective light-harvesting molecules, called photosensitizers (PS), and light of the appropriate wavelength.

This work presents our efforts to characterize a novel class of PS that might find applications in photo-oxidation of biological materials. In particular, we discuss photophysical and photochemical properties of a water-soluble derivative of fullerene C_{60}, fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$. The efficiency of this custom-made PS for the formation of singlet oxygen was verified *in vitro* by electron spin resonance (ESR). The effects of the photo-oxidative stress generated in the presence of the newly-developed fullerol were also investigated on living cells (neurons) by atomic force microscopy (AFM). In these *in situ* AFM experiments, we were capable of probing the cellular stiffness of individual nerve cells exposed to the photo-oxidative stress. AFM measurements revealed that even very short-time exposures (less than 5 min) to the photo-oxidative stress result in a significant drop in stiffness of living neurons. The AFM-detected alterations in the cellular elasticity suggest that OS might induce early changes in the distribution and organization of important cellular components, like, for e.g. focal adhesion points and actin filaments. These results also point to possible applications of water soluble fullerols in performing biological oxidations.

Keywords: Reactive oxygen species, Electron spin resonance, Atomic force microscopy, Singlet oxygen, Oxidative stress, Single cell mechanics.

1. INTRODUCTION

1.1. Reactive oxygen species (ROS)

Oxidation corresponds to the loss of one or more electrons by a molecule, atom, or ion. It is accompanied by reduction - the uptake of one or more electrons by other reaction partners. Oxidation is present in everyday, "macroscopic", life, from iron rusting and hydrocarbon combustion, to physiological processes used by living organisms. In particular, in all species, which live by consuming oxygen, a crucial aspect of energy conversion is cellular respiration, in which dioxygen (O_2) is reduced to water. In living cells, the most common biological reductions of O_2 occur by enzymatically controlled, stepwise, transfers of single electrons. At their numerous intermediate stages, these processes generate the whole variety of reactive species, which reveal both reductive and oxidative properties. The primary products of these reactions, i.e. oxygen-derived reactive compounds, are called reactive

oxygen species (ROS). ROS include (but are not limited to): hydroxyl radical ($OH^{\bullet}$), superoxide ($O_2^{\bullet-}$), electronically-excited dioxygen, i.e. singlet oxygen ($^1\Delta_g$), peroxynitrite ($ONOO^{\bullet-}$) or hydrogen peroxide (H_2O_2). For example, superoxide radical $O_2^{\bullet-}$ is produced by a series of 1-electron transfers in the mitochondrial electron transfer chain. Other enzymes capable of producing $O_2^{\bullet-}$, are xanthine oxidase, NADPH oxidases and cytochrome P450(s). Hydrogen peroxide (H_2O_2) is produced by a wide variety of enzymes including monoxygenases and oxidases. ROS play various roles in numerous biological processes. However, abnormal redox states and especially oxidative states, involving increased concentrations of ROS, invariably lead to impaired physiological functions. Therefore, any disturbance of the equilibrium between the cellular level of ROS and the antioxidant cell capability may result in various pathogenic processes, as summarized in Figure 1.1. Thus, oxidative stress (OS) originates basically from an acute imbalance between the production of ROS and antioxidant capacities of living cells and organisms. In living cells, this imbalance can result from deficient antioxidant capacity, such as the lack of active antioxidants (AO), or from exogenous or endogenous stressor. If not regulated properly, the excess of ROS can damage a cell, disrupting its normal function and may lead its destruction.

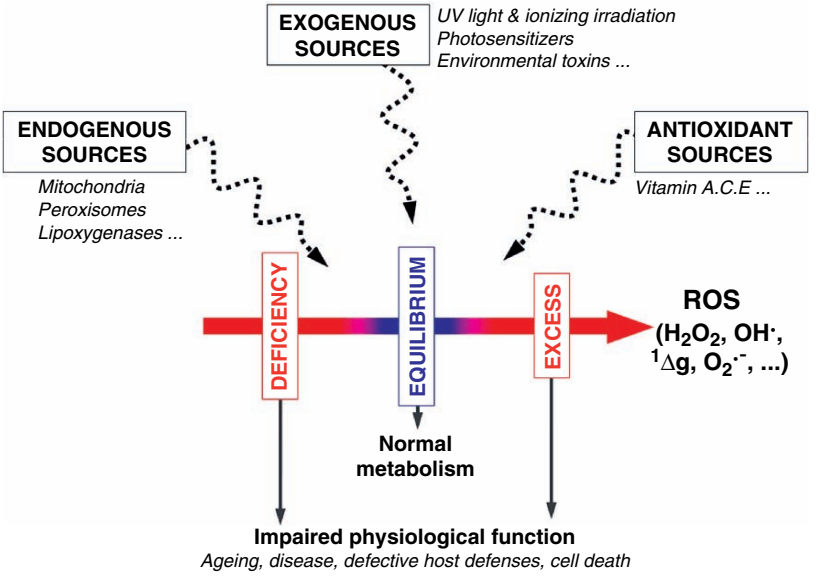


Figure 1.1. Schematic representation of various cellular sources of pro-oxidants (ROS) and antioxidants (AO), as well as mechanisms of cellular responses to the oxidative stress (OS). ROS generation can be induced in both endogenous, including normal metabolism, and exogenous mechanisms. Shifting the equilibrium of the homeostatic set by disturbing the balance between cellular levels of ROS and AO may impair cell functions, thus leading to cell death and numerous human pathologies (Finkel and Holbrook 2000).

Evidence is growing that oxidative stress may be central to many human pathologies, including neurodegenerative brain diseases, hypertension, and pathogenesis of many cancers, as well as to the ageing process (Finkel and Holbrook 2000). However, oxidative stress can also play a positive role. Especially, photo-induced ROS generated in the presence of selected photo-catalysts, herein referred to as photosensitizers (PS), have already found numerous applications.

In particular, light-induced generation of ROS in living tissues has been adopted by a novel medical modality, which is called photodynamic therapy (PDT) (Dolmans et al. 2003). PDT, which is now becoming a powerful therapeutic tool, enables one to eradicate tumors by inducing oxidative stress conditions selectively in diseased tissues. PDT is a treatment that uses photo-sensitizing drugs, i.e. PS, which selectively accumulate in diseased tissues and visible light of low intensity. When PS are exposed to a specific wave-length of light, they produce various forms of ROS, including singlet oxygen, thus initiating a cascade of processes leading to eradication of diseased tissues. Over the last 20 years, PDT has matured and has been optimized for the treatment of various types of cancers. The ever growing number of application domains of PDT is based on its advantages over other therapies, which include its high selectivity and low invasiveness, as well as low intrinsic toxicity of the photosensitizing drugs. The latter feature, together with short lifetimes of ROS generated within the illuminated tissues, minimize potential side effects of the PDT treatment. Thus, in many cases, PDT might be advantageous in comparison to classical treatments, such as chemo- and radio-therapies, which, more often than not, affect also healthy tissues. In parallel to the selective eradication cancer tissues, PDT has also found other applications. PDT has now become a method of choice to cure age-related macular degeneration, in cosmetic removal of port wine stains, dilated facials vessels, wrinkles, as well as in quasi-definitive hair removal.

Recent applications of the photo-oxidative processes also include the design of novel materials, like, for e.g. self-cleaning surfaces. These newly-developed systems photo-sensitize ROS in the presence of surface-immobilized PS, thus enabling: high-level cleaning of surfaces from organic dirt and bacteria - a feature that is useful for sterilization of medical devices (Maness et al. 1999), cleaning air conditioning filters (Fujishima et al. 2000), purification of waste waters in flow-type photoreactors (Kobayakawa et al. 1998), manufacturing of self-cleaning glass (Guan 2005), and surfaces requiring a germ-free environment (e.g. surfaces for food preparation) (Sunada et al. 1998). Furthermore, PS have also found applications in novel

technologies of fully biodegradable products. In particular, it has been shown that PS embedded in different types of substrates can produce ROS upon UV-light irradiation or long-time sun exposure, thus triggering degradation of various coating and packaging materials. In the case of biodegradable plastics, the action of ROS severs polymer chains into smaller, wettable fragments. The oxidized molecules are then degraded by the microbial growth, leading to carbon dioxide and water as final products of degradation. Clearly, all these emerging applications require the design of novel PS with optimized and controllable properties for ROS generation.

1.2. Mechanism of photo-generation of ROS

PS molecules act as light-harvesting antennas that serve to capture incident photons. Targeted systems containing PS are illuminated with light of appropriate energy, which corresponds to the photosensitizer's characteristic absorption bands. The energy uptake from an incoming photon excites the PS molecule to its electronically excited singlet states ($^1S_n^*$). Relaxation of short-lived $^1S^*_n$ states occurs *via* a rapid process of internal conversion (IC), which, within a few tenths of a nanosecond populates the lowest excited singlet state of the PS, $^1S_1^*$. Then, the effective intersystem crossing mechanism (ISC) yields the excited triplet state ($^3S^*$) of the photosensitizer molecule. Although $^1S_1^*$ might give rise to a radiative decay of its energy excess via fluorescence, it is customarily accepted that the major deactivation channel for $^1S_1^*$ occurs mostly through the ISC mechanism, which effectively populates $^3S^*$ states. The excited triplets are long-lived states. Depending on environment and the type of PS, the lifetime of $^3S^*$ states might be in the range of microseconds to milliseconds. This favors the energetically rich excited triplet state of a photosensitizer to react with other molecules from their nearest neighborhood. It is generally accepted that the long-lived excited triplet gives rise to two major types of photophysical processes, which are defined as Type I and Type II mechanisms. Type I mechanism proceeds through electron or hydrogen extraction/transfer from/to the excited triplet state of the PS, whereas Type II mechanism is based on a direct energy transfer from the excited triplet state of the photosensitizer ($^3S^*$) to the triplet ground state of molecular oxygen. The Type I mechanism yields different ROS, like, for example, $OH^\bullet$ or $O_2^{\bullet-}$, whereas the Type II mechanism leads to the formation of singlet oxygen. These two major photo-oxidative routes can be schematically explained in the frame of the Jabłoński-Kasha diagram, as shown in Figure 1.2.

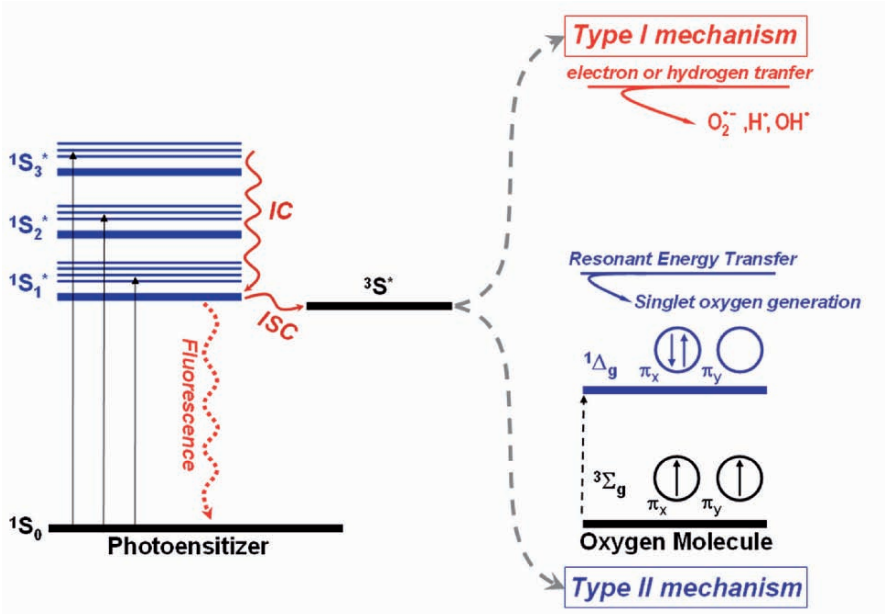


Figure 1.2. Jabłoński-Kasha diagram showing ROS generation *via* Type I or II photoreaction mechanisms. The light-induced excitation populates the short-lived singlet excited states ($^1S_n^*$) of the photosensitizer (PS). This process is then followed by a rapid internal conversion (IC) to the lowest singlet excited state ($^1S_1^*$). Then, the energy from the excited $^1S^*_1$ is transferred *via* the intersystem crossing mechanism (ISC) to a long-lived excited triplet state ($^3S^*$). In Type I mechanism, electron or hydrogen transfer leads to free radical generation. In type II mechanism, molecular oxygen in its triplet ground state (^3S_g) effectively quenches the PS triplet state ($^3S^*$) through a resonant energy transfer, thus yielding singlet oxygen ($^1\Delta_g$).

2. PHOTOSENSITIZERS IN AQUEOUS SOLUTIONS: A WATER-SOLUBLE DERIVATIVE OF C_{60}

2.1. From fullerenes…

Fullerenes, originally called buckminsterfullerenes, are a new class of all-carbon molecules. Since their discovery in the mid-eighties, fullerenes have been the subject of much research as a "third phase" of carbon for new materials science and nanotechnology (Kroto et al. 1985). Fullerenes are carbon nanoparticles that have characteristic spherical or close-spherical cage structures. Fullerene C_{60}, the most abundant member of the family, is a precisely defined molecule composed of 60 carbon atoms arranged in a soccer ball-shaped structure. The cage of C_{60} is made up of 12 pentagons and

20 hexagons and reveals an icosahedral symmetry. The cage of fullerene C_{70}, a larger, but substantially less abundant member of the fullerene family, is composed of 70 carbon atoms arranged in a rugby ball-shaped structure. The close to spherical symmetry and extended π-bond conjugation present in C_{60} in C_{70} confer on these molecules a number of unique properties. In particular, due to the condensed aromatic rings present in C_{60} and C_{70}, both pristine fullerenes are very good absorbers of visible and UV light. It has been found that C_{60} and C_{70} possess a very effective ISC mechanism, which converts electronic excitations to the long-living triplet state ($^1S_n^* \rightarrow {}^3S^*$) with a quantum yield close to unity. The triplet lifetimes are in the range from tens of microseconds for C_{60} to tens of milliseconds for C_{70} (Arbogast et al. 1991; Fraelich and Weisman 1993). Moreover, since the excited triplet states of pristine fullerenes have sufficiently high energy, of 1.6 eV and 1.48 eV, for C_{60} and C_{70}, respectively, these both molecules photosensitize singlet oxygen with high yield in organic solvents (Soemloewocz et al. 2000). Furthermore, in contrast to many organic photosensitizers, pristine fullerenes and their derivatives generally exhibit high resistance to photo-bleaching (Kasermann and Kempf 1998). Therefore, fullerenes are being investigated for a wide array of potential applications, including solar energy conversion (Brabec et al. 2001), coating materials (Colvin 2003), semiconductor technologies (Garaj et al. 2003), and bio-medical applications (Ros and Prato 1999).

2.2. …to fullerols

Despite their excellent photosensitizing properties, biological applications of pristine fullerenes have been restricted due to their extreme hydrophobic character. While pristine fullerenes are soluble in aromatic solvents, they are quasi insoluble in water. However, the system of conjugated double bonds of fullerenes engenders a high chemical reactivity allowing covalent derivatization of the cage with hydrophilic addends, leading to a water-soluble molecule. Recently, novel C_{60}-based water-soluble compounds have become available and are being explored for their biological compatibility and potential applications. The first suggestion that water-soluble fullerenes might find applications in medicine came in 1993 from Friedman et al. who demonstrated that the bis(monosuccinimide) derivative of p,p'-bis(2aminoacethyl)-diphenyl-C_{60} was an efficient HIV protease inhibitor (Friedman et al. 1993). Over the last two decades, water-soluble fullerenes have been studied as contrast agents in magnetic resonance imaging (MRI) (Wharton et al. 2001), sensitizers for performing photo-induced DNA cleavage (Tokuyama et al. 1993), or as candidates for sensitizing biological oxidations (Tabata and Ikada 1999). In some cases, photo-toxic properties of

fullerene-based compounds have been identified as features useful for therapeutics (Kasermann and Kempf 1998; Prato 1997; Wang et al. 2004).

In particular, it has been shown that water-soluble derivatives of C_{60}, with various functional groups (e.g.: -OH, -COOH, -NH_2) attached to the fullerene cage, are promising candidates for biomedical applications (Bosi et al. 2003). Poly-hydroxylated fullerenes, fullerols, seem to be particularly intriguing, since depending on the way of their use, they might behave either as efficient antioxidants, thus reducing ROS levels in cells and tissues (Dugan et al. 1996) or to be effective producers of ROS under illumination with visible of UV light (Vileno et al. 2006). Moreover, it is also generally thought that hydroxylation of the fullerene cage reduces the net generation of singlet oxygen (Guldi and Prato 2000). A recent differential toxicity study of some water-soluble fullerene species demonstrated that changes in the fullerene cage structure directly affected the cytotoxicity *in vitro*. In particular, water-soluble nano-C_{60} demonstrated significant cytotoxicity to tissue culture cells, but a highly-hydroxylated water-soluble fullerene, $C_{60}(OH)_{24}$, produced no evidence of cytotoxicity under the same conditions (Sayes et al. 2004).

To systematically study the photophysical and photochemical properties of water-soluble fullerenes, we developed a novel type of a highly-hydroxylated water-soluble C_{60}-drivative, fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$. The chemical structure of this molecule is shown in Figure 2.3. The synthesis of the novel fullerol was inspired by the previous work of Chiang et al. (Chiang et al. 1992). This approach is based on the oxidation of C_{60} under strongly acidic conditions. A detailed description of the synthesis and spectroscopic characterization of fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ were published elsewhere (Vileno et al. 2006).

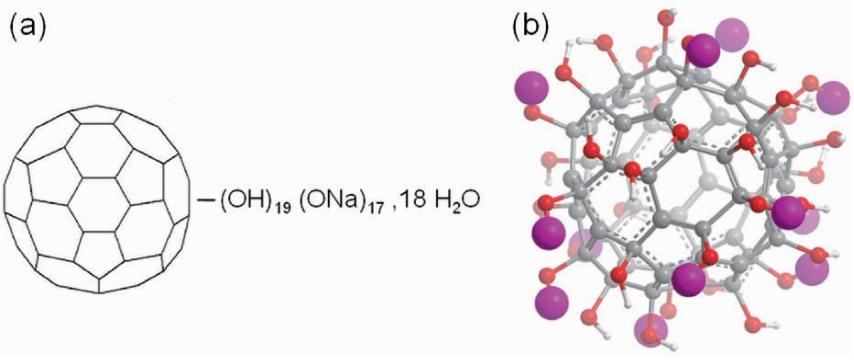


Figure. 2.3. (a) Schematic illustration of the chemical structure of the custom-made poly-hydroxylated fullerene, fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$. (b) Ball and stick representation of the fullerol molecule.

3. ESR - A TOOL FOR STUDYING PHOTOPHYSICAL PROPERTIES OF FULLEROLS

3.1. Introduction to ESR

Electron spin resonance (ESR) is a branch of radio spectroscopy that relies on the behavior of an electron in the magnetic field. ESR enables one to study paramagnetic, ferromagnetic and antiferromagnetic spin systems and is widely applied in many domains of the contemporary science, ranging from physics and chemistry to biology and medicine (Poole Jr. 1967).

If an electron is placed in the external magnetic field, its magnetic moment will align either *parallel* or *antiparallel* to the field. This leads to the Zeeman energy splitting that is proportional to the strength of the applied magnetic field (H_0). Schematic representation of the behavior of an unpaired electronic spin in the presence of the external magnetic field is shown in Figure 3.4a. Absorption of electromagnetic radiation having an appropriate frequency (ν_0) induces a resonant transition from the lower level to the upper level. Although the resonance condition can be fulfilled for a variety of values of ν_0 and H_0, typical ESR measurements are performed at magnetic fields of ~0.35 T and in the range of the electromagnetic radiation of ca. ~9 GHz, i.e. in the microwave X-band range. This resonant absorption of microwaves incident on a spin system under study can be detected in the ESR experiment. Usually, the ESR line is recorded as a first derivative of the absorption signal, as shown in Figure 3.4b (bottom panel). Such a simple ESR spectrum consisting of just one resonant feature is observed for paramagnetic centers having one unpaired electronic spin (S = 1/2), for which all the other magnetic interactions, which are different from that with the external field H_0, can be neglected. In particular, such spectrum can be acquired for one of the most widely used standards in conventional EPR spectroscopy, i.e. 2,2-diphenyl-1-picrylhydrazyl (DPPH). The room-temperature EPR spectrum of polycrystalline DPPH at ~9.5 GHz, shown in Figure 3.4b (bottom panel), consists of a single narrow line (linewidth ΔH_{pp} = 1.6 G), appearing at the spectroscopic factor $\boldsymbol{g}$ = 2.0036 (Al'tshuler and Kozyrev 1964). In ESR, the spectroscopic $\boldsymbol{g}$ factor encodes information about the electronic structure of paramagnetic species. Unpaired electrons in paramagnetic centers are usually associated with their host atoms. Therefore, the angular momentum of an unpaired electronic spin may be also affected by the spin-orbit coupling. This modifies the spectroscopic $\boldsymbol{g}$ factor and enables one to distinguish between different paramagnetic centers. Moreover, if the atom which the unpaired electron is associated with has a non-zero nuclear spin, the resulting magnetic field will also be modified. This. leads

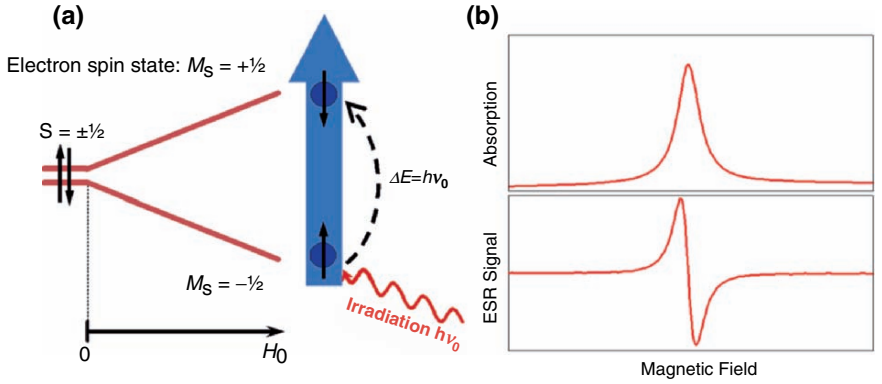


Figure 3.4. (a) Energy level diagram showing Zeeman splitting for a free electron spin system in the presence of the external magnetic field H_0. Absorption of electromagnetic radiation of appropriate energy ($\Delta E = h\nu_0$) induces a resonant transition from the lower level to the upper level. (b) Field modulation and phase sensitive detection convert the absorption curve (upper panel) to its first derivative (lower panel), as acquired for the well-known ESR standard, 2,2-diphenyl-1-picrylhydrazyl (DPPH), at ambient conditions.

to the phenomenon of *hyperfine coupling*, which, depending on its nature and strength, results in ESR line broadening or splits the ESR line into sets of separate features, thus yielding *hyperfine structure*.

It is noteworthy that the numerical result of the double integration of the ESR line is directly proportional to the number of paramagnetic centers in the investigated sample. Thus, in parallel to the observed resonance line intensities, in the ESR experiment one can also get information concerning the number of unpaired spins (paramagnetic centers) in investigated samples.

3.2. Singlet oxygen detection by ESR

In contrast to many other ROS, the electronically excited form of dioxygen, singlet oxygen, is not a free radical (see Figure 1.2). It can be trapped, however, by using an appropriate scavenger. Such a method of singlet oxygen detection by ESR was introduced by Lion *et al*. (Lion et al. 1976). This approach is based on the reaction between a diamagnetic substrate, water-soluble 2,2,6,6-tetramethyl-4-piperidinol (TMP-OH), and singlet oxygen. The process yields a paramagnetic product, 4-hydroxy-2,2,6,6-tetramethy-piperidinyloxy (TEMPOL), which is a stable nitroxide radical (Figure 3.5). TEMPOL is a well-known stable nitroxide radical, which is often used in other spectroscopic techniques, for e.g. to scavenge short-lived free radical species (Meng et al. 2003; Cuzzocrea et al. 2000).

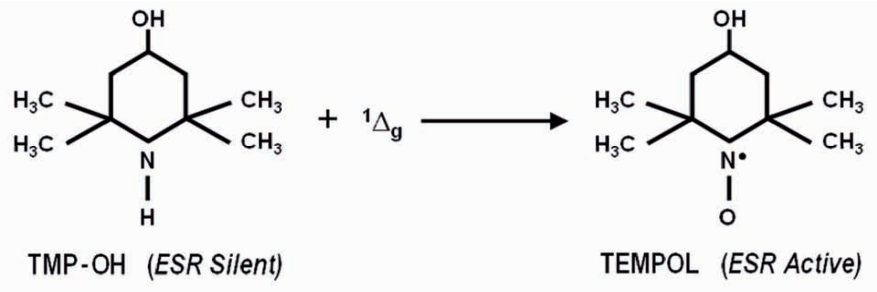


Figure 3.5. Reactive scheme of singlet oxygen detection by ESR. The ESR-silent substrate, TMP-OH, scavenges singlet oxygen ($^1\Delta_g$), which results in the formation of the paramagnetic product, TEMPOL.

It has been demonstrated that the reaction presented in Figure 3.5 is highly specific to the formation of singlet oxygen (Ando et al. 1997). In TEMPOL, the single unpaired electron ($S = 1/2$) resides on the nitrogen atom. Due to this, TEMPOL in water solution has a characteristic ESR spectrum consisting of three equidistant and equi-intense lines. The Zeeman splitting diagram in an external magnetic field for $S = 1/2$ spin system interacting with the nuclear spin $I_N = 1$ of nitrogen ^{14}N and the corresponding ESR spectrum of TEMPOL in water is shown in Figure 3.6.

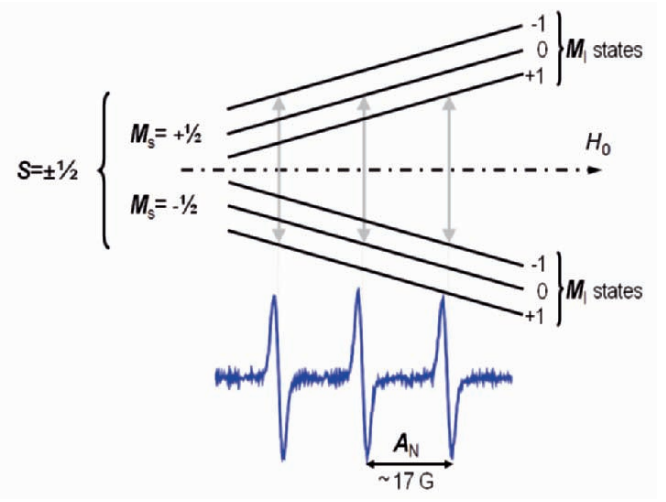


Figure 3.6. The splitting diagram and the corresponding ESR spectrum of the resonant transitions for 0.2 mM TEMPOL in H_2O at ambient temperature. The unpaired electronic spin ($S = \pm ½$) interacts with the nuclear spin of the nitrogen atom ^{14}N ($I_N = 1$), having $2\times I_N + 1 = 3$ nuclear spin states, $M_I = -1, 0, +1$. This magnetic dipole-dipole interaction yields three spectral transitions (satisfying $\Delta M_S = 1$ and $\Delta M_I = 0$ selection rules), which results in a characteristic three-line hyperfine ESR pattern of TEMPOL in water.

3.3. Experimental setup

The samples were prepared by supplementing aqueous solutions of the custom-made fullerol, $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$, with singlet oxygen scavenger, TMP-OH. The concentration of TMP-OH was chosen to be at large excess compared to other substrates, i.e. fullerol and diluted molecular oxygen. Such experimental conditions ensured a zeroth-order kinetic of the observed processes. The solutions were bubbled during 25 min in the dark with gaseous nitrogen or oxygen, to ensure the oxygen-depleted or oxygen-saturated conditions, when appropriate. Small aliquots (~7 μL) of prepared solutions, containing TMP-OH and fullerol were transferred into thin-walled (0.6 mm ID and 0.84 mm OD) quartz capillary tubes (sample height of 25 mm) and sealed on both ends. The thin-walled quartz capillary tubes were then positioned inside regular 2.9 mm ID and 4.0 mm OD ESR sample tubes. This experimental arrangement enabled us to easily handle small aliquots of measured solutions as well as to work with real small samples volumes, thus lowering the usage of the photosensitizer. Such experimental set-up also resulted in negligible loss of the ESR cavity quality factor, thus preserving high spectrometer sensitivity.

To induce the formation of singlet oxygen, samples were exposed to the white light from a halogen source (150 W halogen lamp) outside the ESR cavity at stabilized temperature of 21 $\pm$ 1^{0}C. Samples were illuminated in a specially designed sample holder, where light was delivered using four glass fibers. The arrangement of glass fibers around the sample provided a uniform illumination over the entire sample volume. To follow the ROS-generation as a function of illumination time, this procedure was repeated several times, as schematically shown in Figure 3.7.

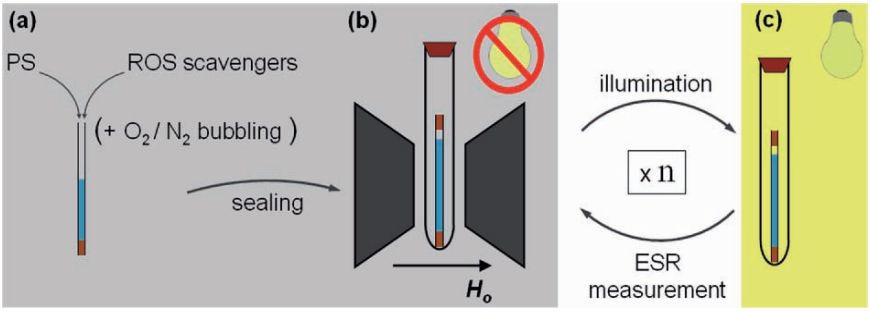


Figure 3.7. Schematics of the experimental arrangement for $^1\Delta_g$ detection by ESR. (a) Sample preparation, including: O_2/N_2 bubbling, sample sealing, aliquots transfer to the ESR capillaries, and the final transfer of sample capillaries to the ESR cavity. (b) ESR measurement. (c) Illumination outside the ESR cavity. After each exposure to light, the samples were quickly transferred back to the cavity. Subsequently, ESR spectra were acquired at stabilized temperature of 21 $\pm$ 1°C . The procedures (a) and (b) were performed in the dark.

3.4. Photo-generation of ROS in the presence of fullerols: an ESR assay

The typical evolution of the ESR signal of TEMPOL as a function of time for photosensitization of singlet oxygen in oxygen-saturated H_2O containing 0.5 mM of fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ and 25 mM of TMP-OH is shown in Figure 3.8a. As can be seen in this Figure, for increasing illumination times the ESR of TEMPOL markedly increases, thus pointing to rapid generation of singlet oxygen. The corresponding plot of the ESR signal intensity, i.e. the double-integrated ESR signal of TEMPOL, as a function of illumination time is shown in Figure 3.8b. As expected for the experimental conditions used in this work, the plot of the ESR signal intensity *vs.* time is linear. This points to the zeroth-order kinetics, for which the overall reaction rate depends of the quantum yield of singlet oxygen generation by the photosensitizer and is also controlled by the diffusion rates of the substrates. In the experiments described in this work, the oxygen diffusion rate is the major reaction rate-limiting factor. However, as can be seen in Figure 3.8b, at certain time of exposure to light (denoted by T), the linear plot of the ESR signal intensity as a function of time changes its increasing character and, after reaching a relatively narrow plateau, the ESR signal intensity starts to decay. We interpret this phenomenon as a rapid change (turnover) of the photosensitization mechanisms from Type II ($^1\Delta_g$ generation) to Type I (free radical generation). In particular, Type I mechanism might result in the formation of hydroxyl radicals ($OH^\bullet$), which are very efficient reductants of TEMPOL (Takeshita et al. 2002). Similar process of the photo-induced decay of TEMPOL has previously been reported by Lion et al. (Lion et al. 1980) and can be summarized by the following scheme:

$$\text{TMP-OH} \xrightarrow{^1\Delta_g} \text{TEMPOL} \underset{dark}{\overset{OH^\bullet}{\rightleftharpoons}} \text{diamagnetic product}$$

Next, the overall evolution of both photosensitizing processes was monitored for three different concentrations of $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ in aqueous solution, i.e. for 74μM, 739 μM, and 1.48 mM. The results are shown in Figure 3.9. Phosphate buffered saline (PBS), an aqueous buffer commonly used in cell biology (pH = 7.4), was used as a solvent in this study.

Because of its physiological neutrality, PBS was used in all ESR experiments reported here, as well as in the subsequent study of the oxidative stress on living cells. It is also worth noting that the PBS salinity did not influence the yield of singlet oxygen in this study.

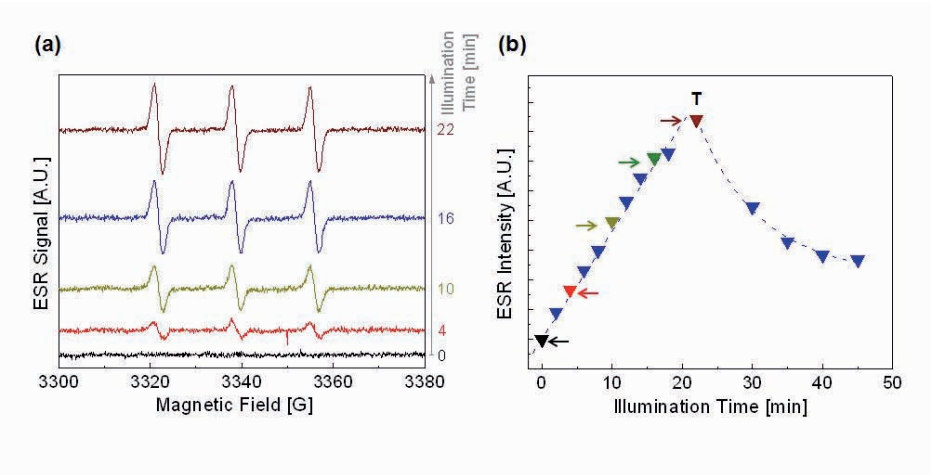


Figure 3.8. Evolution of the ESR signal of TEMPOL as a function of illumination time. (a) ESR spectra acquired for increasing time exposures to visible light. (b) ESR signal intensity at different illumination times. The process of $^1\Delta_g$-photosensitization was performed in oxygen-saturated water solution containing 0.5 mM of $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ and 25 mM of TMP-OH. The point denoted by T corresponds to the turnover between Type I and Type II photo-processes. The points denoted by arrows correspond to the ESR spectra shown in (a).

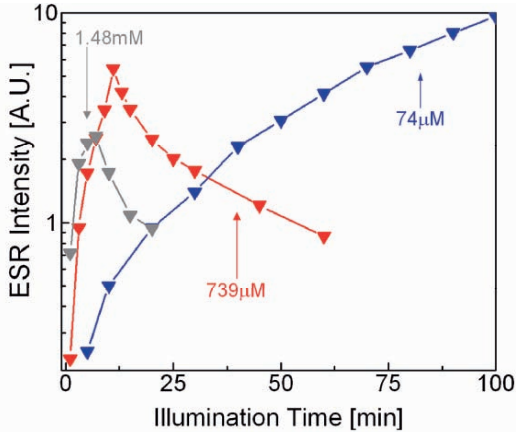


Figure 3.9. Evolution of the ESR signal intensity of TEMPOL as a function of illumination time for photo-sensitization of $^1\Delta_g$ in oxygen-saturated PBS solutions containing 74 μM, 739 μM and 1.48 mM of fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$.

As expected, for short exposures to light, the initial rates of TEMPOL production (proportional to the rates of singlet oxygen generation), were proportional to the concentrations of fullerol in the investigated solutions. In contrast, the total intensity of TEMPOL, which corresponds to the total consumption of molecular oxygen during the Type II photosensitization process, was found to be inversely proportional to the fullerol concentration.

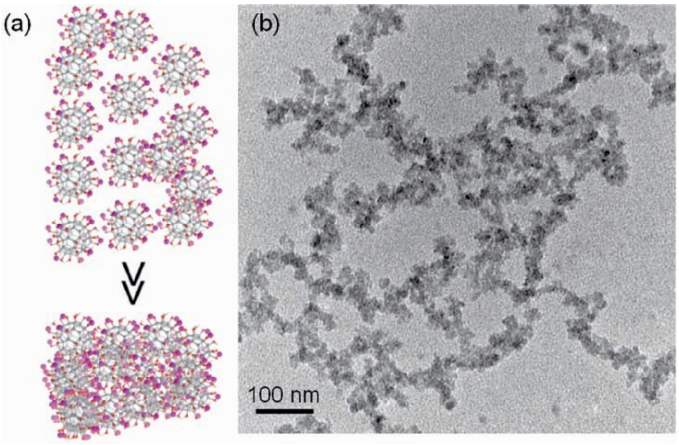


Figure 3.10. (a) Schematics of the lowering of the active surface of fullerol molecules due to the formation aggregates. (b) Cryo-electron microscopy image of the aggregated fullerol molecules (for 1.47 mM solution of fullerol in PBS).

As can also be seen in Figure 3.9, for low concentrations of the photo-sensitizer (<100 μM), the ESR signal intensity of TEMPOL gradually increased and eventually saturated upon prolonged illumination (> 200 min). This rather counter-intuitive behavior can be explained, at least partially, in terms of fullerol aggregation (Vileno et al. 2006). Aggregation of fullerols, which is favored at higher concentrations, reduces the active surface of PS molecules for an efficient energy transfer to dioxygen molecules ($^3O_2$). Schematic representation of the formation of fullerol clusters is shown in Figure 3.10a. The cryo-electron microscopy image of the frozen aqueous solution of fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ reveals the presence of large size aggregates (Figure 3.10b).

4. MECHANICAL PROPERTIES OF LIVING CELLS UNDER OXIDATIVE STRESS: AN AFM STUDY

4.1. Introduction to AFM on living cells

Living cells are complex systems where biochemical and bio-mechanical processes are closely linked. Numerous functions and properties of cells, such as their shape, physical robustness, internal structure, transport of intra-cellular constituents, cell motility and communication with other cells, depend on a network of cytoskeletal filaments and molecular motors (Stossel 1993). They interact with the plasma membrane and other cellular constituents

via numerous proteins. It is known that external or internal stimuli affect the mechanical properties of cells linked to the cytoskeleton and lead to their alterations. This might result, for e.g., in the cell's ability to adhere and/or to deform (Thoumine and Ott 1997). Evidence has now accumulated that quantifying changes in the local mechanical properties of cells might be very helpful in understanding the influence of different environmental factors on living cells.

Scanning probe microscopy has been identified as a promising tool to characterize biological systems at scales of nanometers to microns. In particular, atomic force microscopy (AFM) and its derivatives, such as force recognition AFM, are very well suited to the characterization of biological systems (Morris et al. 1999). AFM probes the interaction of a tiny sharp tip in contact with the investigated specimen to provide information about its topography and elastic properties (Moy et al. 1994; Lekka et al. 1999). Thus, AFM, unlike other high-resolution microscopic techniques, provides not only the high resolution topographic images (Langer et al. 2000), but also enables one to investigate mechanical properties of living objects, like, for e.g., elasticity (Walch et al. 2000) and adhesion (Lehenkari and Horton 1999) of living cells under nearly physiological conditions. Recently, AFM has successfully been used to study the mechanical response of living cells to different external and internal stimuli. In particular, it has been shown that changes in the cytoskeleton can be evaluated based on cell stiffness measured by AFM (Radmacher 1997; Lekka et al. 2001; Rotsch and Radmacher 2000). Therefore, in this work, to check the phototoxic effects of the newly-developed fullerol on living neurons we implemented AFM technique. The changes in the cell elasticity of individual neurons exposed to the oxidative stress were measured *in situ*, i.e. the formation of ROS, including singlet oxygen, was photosensitized directly in an AFM "liquid-cell" setup.

In the standard topography mode, the AFM's probing tip, with the apex curvature ranging from 2 to 50 nm, is located at the end of a cantilever having the typical length of 100-200 μm. The cantilever acts as a tiny spring with a spring constant *k*. The interaction forces between the tip and the investigated surface cause deflection of the cantilever, which is recorded by a position-sensitive photo detector. Knowledge of the cantilever spring constant enables one to estimate the magnitude of forces causing its deflection. As the cantilever is scanned over the sample surface, its deflection is directly related to the sample's topography (Binning et al. 1986).

In contrast, in the force-spectroscopy mode, instead of being laterally scanned over the sample, the AFM tip approaches and retracts from the surface. From the obtained force versus distance dependence (also called a force-distance curve or, simply, a force curve), the tip-sample interaction as

a function of tip-sample distance can partly be extracted. A single force-distance curve consists of two parts: the one recorded during approaching of the tip to the sample surface (*approach curve*) and the other one, collected during opposite motion (*retract curve*).

The shape of the force curve depends on physical and chemical properties of the two interacting surfaces. In the *contact* regime, when the AFM tip is held less than a few Ångström from the sample surface, the interatomic force between the AFM probe and the sample is repulsive, since the electronic clouds of atoms of both the tip and the sample are overlapping and repulsing. The further approach results in the cantilever bending, the character of which depends on the mechanical properties of the investigated sample. Force curves recorded during cantilever approach are usually applied to characterize the elastic properties of the investigated sample by analyzing them using contact mechanics models (Landau and Lifchitz 1999; Cappella and Dietler 1999; Kopycinska-Müller et al. 2006). The illustration of the AFM experiment is shown in Figure 4.11.

In solid mechanics, the elasticity or stiffness of a given material is often expressed in terms of its modulus of elasticity or the Young modulus, which is defined as the ratio between the mechanical stress and deformation (Landau and Lifchitz 1987). In this work, to quantify the mechanical response of cells to the oxidative stress, the Young's modulus of individual cells (E_{cell}) was derived based on AFM measurements. E_{cell} was evaluated by applying Sneddon's elastic contact mechanics approach (Sneddon 1965), which was derived from the classical Hertz mechanical model (Hertz 1881). This

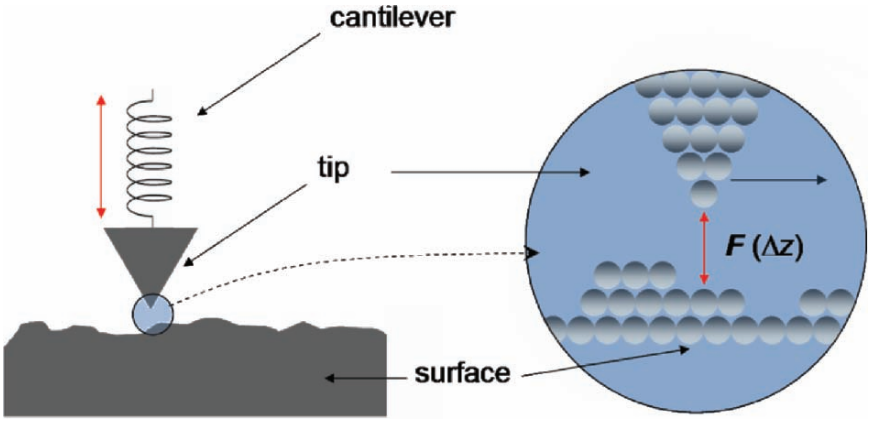


Figure 4.11. Schematics of the AFM experiment. A sharp and tiny AFM tip is brought into contact with the sample surface. The interacting force between the atoms of AFM tip and surface causes the cantilever deflection which can be converted into force value when a spring constant of the cantilever is known.

latter model was originally developed to describe the elastic behavior of two spheres in contact. Sneddon refined the classical Hertzian approach to include also other shapes. In particular, Sneddon determined the relation between the load of axisymmetric indenters of different shapes and the resulting penetration depth in an elastic half space. This approach was then adapted to AFM by Weisenhorn et al. (Weisenhorn et al. 1993) and revised by Radmacher et al. (Radmacher et al. 1995). Since the number of atoms in the apex of the AFM tip that make contact with the sample during indentation is rather large (>> 100), the classical theory of elasticity can be applied.

Very often, the commercially available AFM tips have the shape of a four-sided pyramid. Therefore, in theoretical models describing their mechanical interaction with investigated samples, the shape of the AFM tip is often approximated either by a cone or by a paraboloid. Schematic representations of the cell surface responses indented by either the conical or parabolic-shape AFM tips are shown in Figure 4.12. The theoretical formulae of the force-indentation relations depending of the assumed AFM tip shape are presented below.

For the conical tip, as shown in Figure 4.12a:

$$F(\Delta z) = \frac{2E^*}{\pi \, tan\alpha} \Delta z^2 \tag{4.1}$$

where F is the force, Δz the indentation depth, α the semi-vertical angle describing the cone, and E^* the relative Young's modulus of the tip-cell system, defined by:

$$\frac{1}{E^*} = \frac{1-\eta_{cell}^2}{E_{cell}} + \frac{1-\eta_{tip}^2}{E_{tip}} \tag{4.2}$$

with η being the Poisson ratio (which relates shear stress to compression stress), E_{tip} and E_{cell} are the Young's moduli of the tip and the cell, respectively. Assuming that $E_{tip} >> E_{cell}$ we obtain:

$$E^* \sim \frac{E_{cell}}{1-\eta_{cell}^2} \tag{4.3}$$

and

$$F(\Delta z) = \frac{2\,E_{cell}}{\pi\,(1-\eta_{cell}^2)\,tan\alpha} \Delta z^2 \tag{4.4}$$

For the parabolic-shaped tip (see Figure 4.12b), the following equation can be obtained:

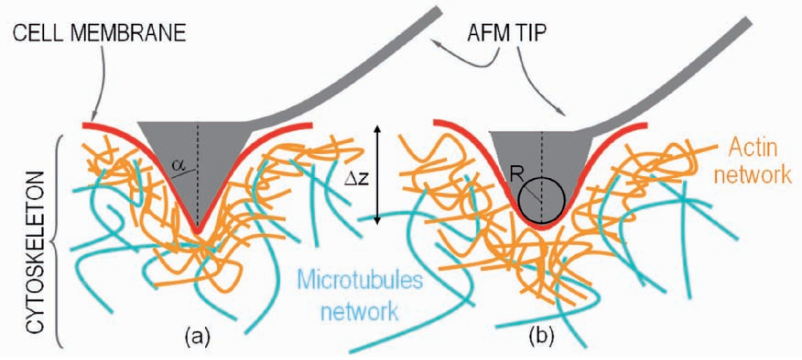


Figure 4.12. Schematics of cell indentation with: conical (a) and parabolic (b) shaped AFM tips.

$$F(\Delta z) = \frac{4\sqrt{R}}{3} \frac{E_{cell}}{1-\eta_{cell}^2} \Delta z^{1.5} \qquad 4.5$$

where R is the radius of curvature at the tip apex.

Both discussed models represent a relatively good approximation of indentation of cells by the AFM probe in real experiments. However, there is no simple criterion to choose one model over the other. In this study, the most appropriate criterion for choosing the theoretical model was the best fit (i.e. having the smallest value of the fitting parameter χ^2) to the force-distance curves.

To determine the Young's modulus of individual cells (E_{cell}) we followed the experimental procedure that is schematically depicted in Figure 4.13a (Weisenhorn et al. 1992). First, to get a calibration force curve, the surface of the glass coverslip was indented. Afterwards, a topographic AFM image of the studied cell was obtained (see Figure 4.13b). Finally, the investigated cell was indented by the AFM tip. During indentation, the cantilever deflection (related to the applied load force) was recorded as a function of the relative surface position. This yielded the force-distance curve for the investigated cell. Both force curves are shown in Figure 4.13a. In this diagram, the y-axis represents both the deflection of the cantilever and the applied force. The latter one can be derived by multiplying the deflection d by the cantilever's spring constant k using the Hook's law: $F = k\,d$. When the cantilever is far away from the sample surface, there is no significant force, which could cause its bending. This is represented by the horizontal straight line in Figure 4.13a. After the contact with the surface is established (point C), the cantilever pronouncedly deflects during the further approach due to the repulsion between the tip and the surface. For non-deformable surfaces, for e.g. the surface of the glass coverslip (denoted by the red spot in

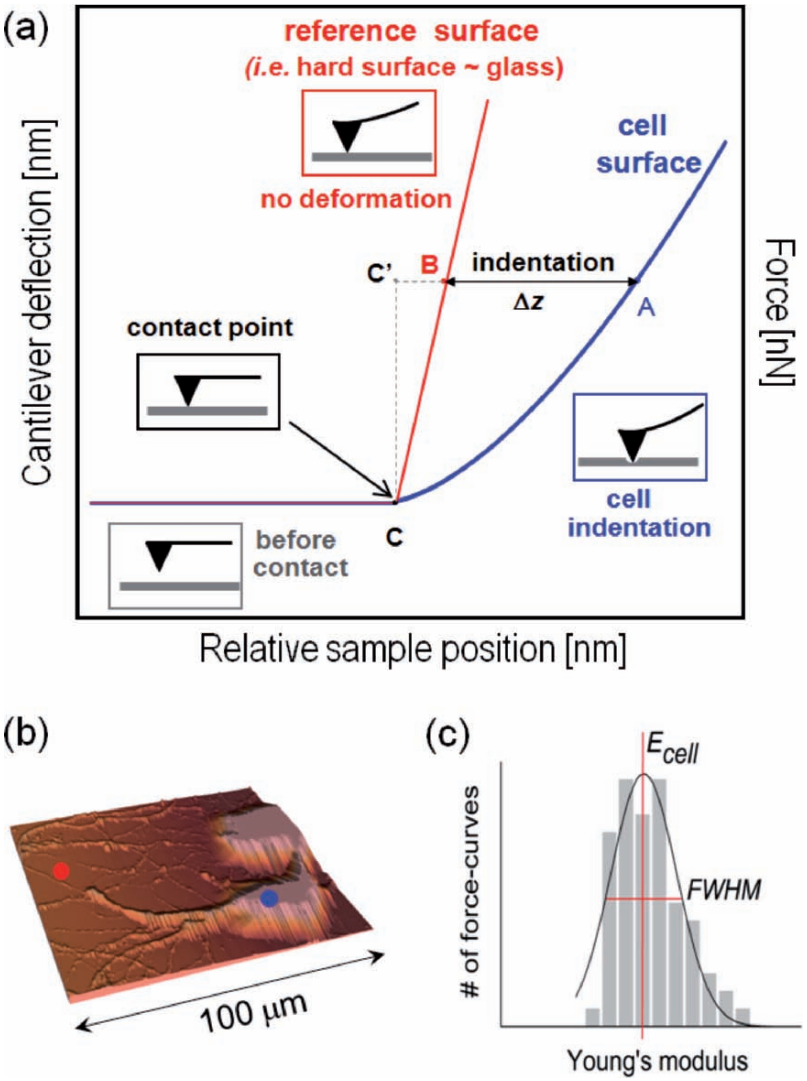


Figure 4.13. (a) Diagram of two force-distance curves measured on soft (blue curve) and a hard (red curve) surfaces. The subtraction of these two curves (A - B) determines the indentation as a function of the applied force, which yields the so-called force *vs.* indentation curve. (b) The AFM image of a living cell. The force-distance curves were collected at the hard surface (location denoted by red spot) and the soft surface, i.e. cell (location denoted by blue spot). (c) Gaussian distribution (black solid line) fitted to a histogram of the estimated Young's modulus values.

Figure 4.13b), this deflection linearly depends on the relative surface position (the straight red line in Figure 4.13a). For soft surfaces, like cells, the cantilever deflection is smaller due to the tip penetration into cell, and the

resulting force curve has a non-linear character, thus revealing compliance of the sample (the blue curve in Figure 4.13a). The difference between the curves obtained for the hard surface (calibration) and the soft substrate (cell) determines the magnitude of the AFM tip indentation and is denoted by Δz in Figure 4.13a. The dependence of the load force applied to a cell on the resulting indentation was used for calculation of the cell stiffness. The actual values of E_{cell} were derived using the best fitting model of the AFM tip (equations 4.4 or 4.5). To obtain a statistically reliable set of data, 15 to 50 curves were taken by indentation of the surface of one cell. The force curves were recorded within the area depicted by the blue spot in Figure 4.13b. The same procedure was repeated for 10 to 15 different cells, for each experimental condition. Typical indentation depth was of *ca.* 500 nm. All the resulting data were used for preparing the histogram. The calculations of E_{cell} values were done assuming the Poisson ratio equal to 0.5 (cells were assumed to be perfectly isotropic and incompressible). The average value of the Young's modulus was derived from the histogram shown in Figure 4.13c by fitting Gaussian function. The spread of the Young modulus values, i.e. the standard deviation of E_{cell} that is defined as a half of the full width at half maximum of the fitting Gaussian curve (denoted as *FWHM* in Figure 4.13c), originates from different factors, like uncertainties concerning the tip radii, the spring constants of cantilevers, the contact area, sample heterogeneity, and approximations of the theoretical model used for analysis. Therefore, it is rather difficult to determine the absolute value of the Young's modulus. However, while studying the mechanical response of cells to the oxidative stress, measurements of the relative changes in cellular stiffness were sufficient.

4.2. Sample preparation and experimental assay

Rat primary hippocampal neurons were grown on glass coverslips according to Steiner et al. (Steiner et al. 2992). To induce the photo-oxidative stress conditions on neurons the custom-made fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ was used as photosensitizer. Physiological pH was ensured by dissolving the photosensitizer in PBS. Then, the stock solution was oxygen-bubbled for 30 min. To perform AFM measurements, the culture plate with cells, immersed in PBS solution supplemented with fullerol, was positioned in the AFM "liquid-cell" setup. *In-situ* illumination was done with visible light, using a halogen lamp (150 W). Light was delivered directly *via* a flexible fiber-optic light guide to the Petri dish positioned on the AFM scanner. Schematic illustration of the AFM experiment is shown in Figure 4.14.

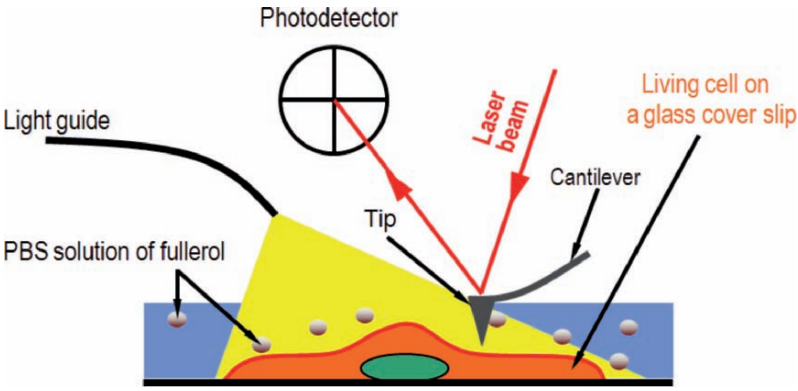


Figure 4.14. Schematic illustration of the AFM experiment with *in situ* generation of ROS. A flexible light guide delivers white light to a living cell attached to the bottom of the Petri dish, which is positioned in the liquid-cell module of the AFM scanner.

4.3. Mechanical response of cells to the photo-oxidative stress

Rat primary hippocampal neurons cells were exposed to OS generated in the presence of fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ under illumination with visible light for 3, 6 and 12 min. Two different photosensitizer's concentrations, 739 μm and 74 μm, were used. Control measurements were regularly performed during the experiment. No marked changes in the calculated Young modulus values were observed for neurons in the absence of the photosensitizer (i.e. in pure PBS) under illumination with white light for 60 min. Similarly, there were no marked changes in the cell stiffness for neurons incubated for 60 min in the dark in PBS supplemented with fullerol. However, as can be seen in Figure 4.15, the combined action of fullerol and light resulted in a rapid decrease of the cellular stiffness. A drop of ~50% in the Young's modulus values was observed already after 3 min of illumination. Interestingly, this marked drop in the cell elasticity was found independent of fullerol concentration. As can be seen in of Figure 4.15, after 3 min of illumination with white light, almost the same decrease of the cell stiffness was found for 739 μM and 74 μM concentrations of fullerol. For longer exposures (*ca.* 12 min), however, this drop in the cell elasticity was more pronounced for neurons incubated in the oxidative stress conditions generated by the higher concentration of fullerol (739 μM). As shown in section 3.4, at both concentrations, fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$

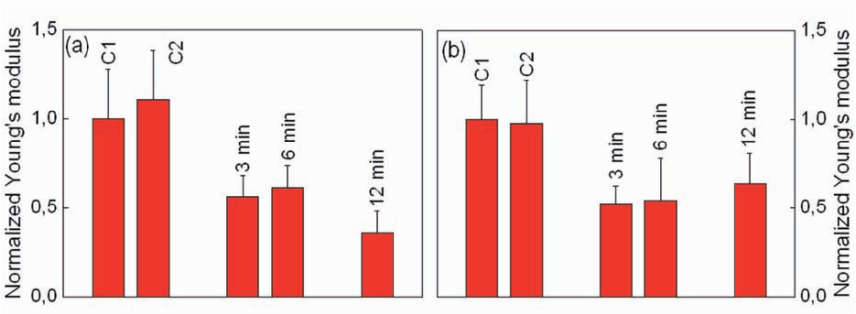


Figure 4.15. Comparison of the evolution of the Young's modulus measured for exposed to the photo-oxidative stress in the presence of the custom-made fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ at two concentrations (a) 739 μM and (b) 74 μM. The columns denoted by C1 and C2 correspond to the control measurements. The control measurements were performed in PBS in the absence of fullerol (C1) and in PBS supplemented with fullerol in the dark (C2), respectively. All the data were normalized to the control value denoted by C1. The Young's modulus values determined for C1 were: 5.2 ± 1.4 kPa for fullerol concentration of 739 μM (a) and 4.4 ± 0.9 kPa for fullerol concentration of 74 μM (b).

generated $^1\Delta_g$ via the Type II photosensitization pathway. Although under the experimental conditions used in this study the formation of ROS other than singlet oxygen can not be excluded, one can assume that the photo-oxidative stress on neurons was mostly mediated by singlet oxygen.

Alterations in the Young's modulus of cells determined *via* AFM elasticity measurements have often been attributed to the reorganization of the cytoskeleton (Rotsch and Radmacher 2000). Taking into account a rather shallow range of action of ROS, the rapid drop in the cell stiffness observed in this study might be explained by damages occurring to the elements close to the cellular periphery like cellular membrane, focal adhesion points, changes in the actin network and damage to the membrane-bound ends of microtubules. There is a growing body of evidence that singlet oxygen can target actin filaments, thus leading to their disruption (Conlon et al. 2002). The changes in the actin filament network can influence the structure and function of other parts of the cytoskeleton, which might result in the overall cell elasticity change observed in this study for cells exposed to the oxidative stress.

5. CONCLUSIONS AND PERSPECTIVES

The major objective of this work was to characterize photophysical properties a novel water-soluble derivative of fullerene C_{60}. The highly-derivatized C_{60}-derivative was obtained *via* multiple hydroxylation of the C_{60} cage. The synthesized fullerol $C_{60}(OH)_{19}(ONa)_{17}\times 18H_2O$ revealed a very

high water-solubility (up to 6 mg/mL), as well as a strong tendency to aggregate. Fullerol clusters were observed in aqueous solutions at low, micromolar, concentrations. By all evidence, as revealed in the ESR portion of this study, the novel fullerol generated singlet oxygen in aqueous solutions. Thus, the well-known ability of pristine fullerenes to generate singlet oxygen was not suppressed by a massive covalent derivatization of the C_{60} cage.

The novel compound was used as a photosensitizer to generate oxidative stress on living cells (neurons). The deleterious effect of the oxidative stress on cells was studied by an *in-situ* AFM experiment, where the reactive oxidative species (ROS), including singlet oxygen, were photosensitized directly under the AFM tip. The AFM measurements revealed a significant decrease of the cell stiffness already after short exposures of living neurons to the photo-oxidative stress. The changes in the cell elasticity of individual neurons were associated with OS-induced damages to cellular and subcellular targets located in close vicinity to the photosensitizer molecules, i.e. elements at the cellular periphery, like cellular membrane, focal adhesion points, and membrane-bound ends of microtubules or fibers of the actin network. In particular, we emphasize the OS-mediated damages to the actin filament network. Actin fibers interact with other parts of the cytoskeleton. Therefore, the early oxidative damage to the actin fibers may continue to propagate inside the cell, thus resulting in the overall change of cell elasticity. It is also worth noting that AFM responded to early OS-induced changes occurring to the mechanical properties of cells. This might bear a potential for further applications of AFM for studying the dynamics of oxidative stress in living cells.

In this study, we also demonstrate that the highly-hydroxylated fullerene derivative, may not necessarily behave as a "free radical sponge", as has been suggested in several recent reports (Sayes et al. 2004; Isakovic et al. 2006). In contrast, under certain conditions, it can also generate the strong oxidizing agent, singlet oxygen. Therefore, the photo-biological risk associated with the use of water-soluble derivatives of C_{60} might involve, in part, the formation of singlet molecular oxygen *via* Type II photoreaction mechanism. This is of particular importance for fullerols, given their wide-spread applications in products designed for direct human use and consumption. Fullerols have strong assets for their perspective biological and biomedical applications, such as their low toxicity in the dark, high photo-stability (i.e. lack of photo-bleaching), and sensitivity to photosensitize singlet oxygen within a particularly wide spectrum of illumination, ranging from UV to visible light. Therefore, potential applications of fullerols might include photo-generation of ROS in aqueous environments, for e.g. in general sterilization or for selective treatment of bacterial infections *via* photodynamic inactivation (Caminos et al. 2006).

REFERENCES

Al'tshuler, S.A. and Kozyrev, B.M. (1964) In: Poole, Jr., C.P. ed. *Electron Paramagnetic Resonance,* Academic Press, New York.

Ando, T., Yoshikawa, T., Tanigawa, T., Kohno, M., Yoshida, N. and Kondo, M. (1997) Quantification of singlet oxygen from hematoporphyrin derivative by Electron Spin Resonance. *Life Sciences* **61**(19): 1953–1959.

Arbogast, J.W., Darmanyan, A.P., Foote, C.S., Rubin, Y., Diederich, F.N., Alvarez, M.M., Anz, S.J. and Whetten, R.L. (1991) Photophysical properties of C_{60}. *J. Phys. Chem.* **95**(1): 11–12.

Binnig, G., Quate, C.F. and Gerber, C. (1986) Atomic Force Microscope. *Phys. Rev. Lett.* **56**(9): 930–933.

Bosi, S., Da Ros, T., Spalluto, G. and Prato, M. (2003) Fullerene derivatives: an attractive tool for biological application. *Eu. J. Med. Chem.* **38**: 913-923.

Brabec, C.J., Shaheen, S.E., Fromherz, T., Padinger, F., Hummelen, J.C., Dhanabalan, A., Janssen, R.A.J. and Sariciftci, N.S. (2001). Organic photovoltaic devices produced from conjugated polymer/methanofullerene bulk heterojunctions. *Synthetic Metals* **121**(1-3): 1517–1520.

Caminos, D.A., Spesia, M.B. and Durantini, E.N. (2006) Photodynamic inactivation of *Escherichia coli* by novel *meso*-substituted porphyrins by 4-(3-*N,N,N* trimethylammonium-propoxy)phenyl and 4-(trifluoromethyl)phenyl groups. *Photochem. Photobiol. Sc.* **5**: 56–65.

Cappella, B. and Dietler, G. (1999) Force-distance curves by atomic force microscopy. *Surface Science Reports* **34**: 1–104.

Chiang, L.Y., Upasani, R.B. and Swirczewski, J.W. (1992) Versatile nitronium chemistry for C_{60} fullerene functionalization. *J. Am. Chem. Soc.* **114**(26): 10154–10157.

Colvin, V.L. (2003) The potential environmental impact of engineered nanomaterials. *Nature Biotechnol.* **21**(10): 1166–1170.

Conlon, K.A., Rosenquist, T. and Berrios, M. (2002) Site-directed photochemical disruption of the actin cytoskeleton by actin-binding rose bengal-conjugates. *J. Photochem. Photobio. B: Biology*, **68**(2-3): 140–146.

Cuzzocrea, S., McDonald, M.C., Filipe, H.M., Costantino, G., Mazzon, E., Santagati, S., Caputi, A.P. and Thiemermann, C. (2000) Effects of tempol, a membrane-permeable radical scavenger, in a rodent model of carrageenan-induced pleurisy. *Eu. J. Pharmacol.* **390**: 209–222.

Da Ros, T. and Prato, M. (1999) Medicinal chemistry with fullerenes and fullerene derivatives. *Chem. Commun.* **8**: 663–669.

Dolmans, D.E.J.G.J., Fukumura, D. and Jain, R. K. (2003) Photodynamic therapy for cancer. *Nature Rev. Cancer* **3**(5): 380–387.

Dugan, L.L., Gabrielsen, J.K., Yu, S.P., Lin, T.S. and Choi, D.W. (1996) Buckminsterfullerenol free radical scavengers reduce excitotoxic and apoptotic death of cultured cortical neurons. *Neurobiol. Disease* **3**: 129–135.

Finkel, T., and Holbrook, N.J. (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* **408**(6809): 239–247 .

Fraelich, M.R. and Weisman, R.B. (1993) Triplet-states of C_{60} and C_{70} in solution – longintrinsic lifetimes and energy pooling. *J. Phys. Chem.* **97**(43): 11145–11147.

Friedman, S.H., Decamp, D.L., Sijbesma, R.P., Srdanov, G., Wudl, F. and Kenyon, G.L. (1993) Buckyball based inhibitors of the HIV protease – 2nd-generation design and model studies. *FASEB J.* **7**(7): A1184–A1184.

Fujishima, A., Rao, T.N. and Tryk, D.A. (2000) Titanium dioxide photocatalysis. *J. Photochem. Photobiol. C: Photochem. Rev.* **1**: 1–21.

Garaj, S., Kambe, T., Forro, L., Sienkiewicz, A., Fujiwara, M. and Oshima, K. (2003) Polymer phase of the tetrakis(dimethylamino)ethylene-C_{60} organic ferromagnet. *Phys. Rev. B*, **68** (14): 144430–7.

Guan, K., (2005). Relationship between photocatalytic activity, hydrophilicity and self-cleaning effect of TiO_2/SiO_2 films. *Surface and Coatings Technol.* **191**: 155– 160.

Guldi, D. and Prato, M. (2000) Excited-State Properties of C_{60} Fullerene Derivatives. *Acc. Chem. Res.* **33**(10): 695–703.

Hertz, H. (1881) Über die Berührung fester elastischer Körper. *J. Reine und Eingewandte Mathem.* **92**: 156–171.

Isakovic, A., Markovic, Z., Todorovic-Markovic, B., Nikolic, N., Vranjes-Djuric, S., Mirkovic, M., Dramicanin, M., Harhaji, L., Raicevic, N., Nikolic, Z. and Trajkovic, V. (2006) Distinct cytotoxic mechanisms of pristine versus hydroxylated fullerene. *Toxicol. Sci.* **91**(1): 173–183.

Kasermann, F. and Kempf, C. (1998). Buckminsterfullerene and photodynamic inactivation of viruses. *Rev. Med. Virol.* **8**(3): 143–151.

Kobayakawa, K., Sato, C., Sato, Y. and Fujishima, A. (1998) Continuous-flow photoreactor packed with titanium dioxide immobilized on large silica gel beads to decompose oxalic acid in excess water. *J. Photochem. Photobiol. A: Chem.* **118**(1): 65–69.

Kopycinska-Müller, M., Geiss, Roy H. and Hurley, Donna C. (2006) Contact mechanics and tip shape in AFM-based nanomechanical measurements. *Ultramicroscopy* **106**: 466–474.

Kroto, H.W., Heath,, J.R., O'Brien, S.C., Curl, R.F. and Smalley, R.E. (1985) C_{60}: Buckminsterfullerene. *Nature* **318**: 162–163.

Landau, L. and Lifchitz, E. (1987) *Theory of elasticity*. Pergamon Press, Oxford, U.K.

Langer, M.G., Koitschev, A., Haase, H., Rexhausen, U., Horber, J.K.H. and Ruppersberg, J.P. (2000) Mechanical stimulation of individual stereocilia of living cochlear hair cells by Atomic Force Microscopy. *Ultramicroscopy* **82**(1-4): 269–278.

Lehenkari, P.P. and Horton, M.A. (1999) Single integrin molecule adhesion forces in intact cells measured by Atomic Force Microscopy. *Biochem. Biophys. Res. Comm.* **259**(3): 645–650.

Lekka, M., Laidler, P., Ignacak, J., Labedz, M., Lekki, J., Struszczyk, H., Stachura, Z. and Hrynkiewicz, A.Z. (2001) The effect of chitosan on stiffness and glycolytic activity of human bladder cells. *Biochim. Biophys. Acta-Mol. Cell Res.* **1540**(2): 127–136.

Lekka, M., Lekki, J., Marszalek, M., Golonka, P., Stachura, Z., Cleff, B. and Hrynkiewicz, A.Z. (1999) Local elastic properties of cells studied by SFM. *Applied Surface Science* **141**: 345–349.

Lion, Y., Delmelle, M. and Van de Vorst, A. (1976) New method of detecting singlet oxygen production. *Nature* **263**: 442–443.

Lion, Y., Gandin, E. and Vandevorst, A. (1980) On the production of nitroxide radicals by singlet oxygen reaction – An Electron Paramagnetic Resonance study. *Photochem. Photobiol.* **31**(4): 305–309.

Maness, P.C., Huang, Z., Smolinski, S., Jacoby, W., Blake, D. and Wolfrum, E. (1999) Photosterilization and photomineralization of microbial cells with titanium dioxide. *Photochem. Photobiol.* **69**: 64S–65S.

Meng, S., Cason, G.W., Gannon, A.W., Racusen, L.C. and Mannign, Jr., R.D. (2003) Oxidative stress in Dahl salt-sensitive hypertension. *Hypertension* **41**: 1346–1352.

Morris, V.J., Gunning, A.P., Kirby, A.R. (1999) In: *Atomic Force Microscopy for Biologists*, Imperial College Press, London.

Moy, V.T., Florin, E.L. and Gaub, H.E. (1994) Intermolecular forces and energies between ligands and receptors. *Science* **266**: 257-259.

Poole Jr., C.P. (1967) In: *Electron Spin Resonance: A Comprehensive Treatise on Experimental Techniques*, Interscience Publishers, NY. p. 257.

Prato, M. (1997) Fullerene chemistry for materials science applications. *J. Materials Chem.* **7**: 1097-1109.

Radmacher, M., Fritz, M., and Hansma, P.K. (1995) Imaging soft samples with the Atomic Force Microscope – gelatin in water and propanol. *Biophys. J.* **69**(1): 264–270.

Radmacher, M. (1997) Measuring the elastic properties of biological samples with the afm. *IEEE Engineering Med. Biol. Magazine*, **16**(2): 47–57.

Rotsch, C. and Radmacher, M. (2000) Drug-induced changes of cytoskeletal structure and mechanics in fibroblasts: An Atomic Force Microscopy study. *Biophys. J.* **78**(1): 520–535.

Sayes, C.M., Fortner, J.D., Guo, W., Lyon, D., Boyd, A.M., Ausman, K.D., Tao, Y.J., Sitharaman, B., Wilson, L.J., Hughes, J.B., West, J.L. and Colvin, V.L. (2004) The differential cytotoxicity of water soluble fullerenes. *Nano Lett.* **4**: 1881-1887.

Sienkiewicz, A., Garaj, S., Białkowska-Jarowska, E. and Forró, L. (2000). Singlet oxygen generation by C_{60} and C_{70} – An ESR study. In Eds. Kuzmany/Finck/Mehring, *Winterschool on Electronic Properties of Novel Materials*, volume 544 of AIP Conference Proceedings, Melvile New-York, pp. 63–66.

Sneddon, I.N. (1965) The relation between load and penetration in the axisymmetric Boussinesq problem for a punch or arbitrary profile. *Inter. J. Engineering Sci.* **3**: 47–57.

Steiner, P., Sarria, J.C.F., Glauser, L., Magnin, S., Catsicas, S., and Hirling, H. (2002) Modulation of receptor cycling by neuron-enriched endosomal protein of 21 KD. *J. Cell Biol.* **157**(7): 1197–1209.

Stossel, T.P. (1993) On the crawling of animal cells. *Science* **260** (5111): 1086–1094.

Sunada, K., Kikuchi, Y., Hashimoto, K. and Fujishima, A. (1998) Bactericidal and detoxification effects of TiO_2 thin film photocatalysts. *Environ. Sci. Technol.* **32**(5): 726–728.

Tabata, Y. and Ikada, Y., (1999) Biological functions of fullerene. *Pure Appl. Chem.* **71**(11): 2047–2053.

Takeshita, K., Saito, K., Ueda, J., Anzai, K., Ozawa, T. (2002) Kinetic study on ESR signal decay of nitroxyl radicals, potent redox probes for *in vivo* ESR spectroscopy, caused by reactive oxygen species. *Biochim. Biophys. Acta* **1573**: 156– 164.

Thoumine O. and Ott, A. (1997) Comparison of the mechanical properties of normal and transformed fibroblasts. *Biorheology* **34**(4-5): 309–326.

Tokuyama, H., Yamago, S., Nakamura, E., Shiraki, S. and Sugiura, Y. (1993) Photoinduced biochemical-activity of fullerene carboxylic-acid. *J. Am. Chem. Soc.* **115**(17): 7918–7919.

Vileno, B., Marcoux, P.R., Lekka, M., Sienkiewicz, A., Feher, T. and Forró, L. (2006) Spectroscopic and photophysical properties of a highly derivatized C_{60} fullerol. *Advan. Funct. Materials* **16**(1): 120–128.

Walch, M., Ziegler, U. and Groscurth, P. (2000) Effect of streptolysin on the microelasticity of human platelets analyzed by Atomic Force Microscoppy. *Ultramicroscopy* **82**(1-4): 259–267.

Wang, S., Gao, R., Zhou, R. and Selke, M. (2004) Nanomaterials and singlet oxygen photo-sensitizers: potential applications in photodynamic therapy. *J. Materials Chem.* **14**: 487-493.

Weisenhorn, A.L., Khorsandi, M., Kasas, S., Gotzos, V. and Butt, H.J. (1993) Deformation and height anomaly of soft surfaces studied with an AFM. *Nanotechnology* **4**: 106–113.

Weisenhorn, A.L., Maivald, P., Butt, H.J. and Hansmam P.K. (1992) Measuring adhesion, attraction, and repulsion between surfaces in liquids with an Atomic Force Microscope. *Phys. Rev. B*, **45**(19): 11226–11232.

Wharton, T., Kini, V.U., Mortis, R.A. and Wilson, L.J. (2001) New non-ionic, highly water-soluble derivatives of C_{60} designed for biological compatibility. *Tetrahedron Lett.*, **42**(31): 5159–5162.

Functional EPR Spectroscopy and Imaging of Nitroxides

VALERY V. KHRAMTSOV
Dorothy M. Davis Heart & Lung Research Institute, Division of Pulmonary, Critical Care & Sleep Medicine, The Ohio State University, Columbus, OH 43210, USA, Phone: +1 614-688-3664, Fax: +1 614-293-4799, E-mail: khramtsov-1@medctr.osu.edu

Abstract

Absorption and fluorescent probes are particularly effective for studies at the cellular and subcellular levels, while magnetic resonance spectroscopy has the advantage of *in vivo* applications in animals and humans. Over the past decade, despite formidable technical problems, significant progress has been made regarding *in vivo* EPR techniques that have much higher sensitivity than NMR for the same probe concentration. However the potential of the EPR-based techniques is still far from being maximally defined, predominantly because of the requirement for exogenous spin probes. Among paramagnetic probes, the nitroxides, stable organic free radicals, provide a wide range of available structures variable in solubility and tissue redistribution, spectral and functional sensitivity, ability to be targeted and their lifetimes in living tissues. Particularly an ability to follow by EPR specific chemical reactivity of the nitroxides towards biologically relevant compounds provides unique functionality to *in vivo* EPR spectroscopy and imaging. In this contribution we reviewed applications of EPR spectroscopy and imaging of the nitroxides for the probing local chemical environment, including oxygen, redox state, thiols, pH, and nitric oxide, with particular emphasis on *in vivo* applications.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 181–208.

Keywords :EPR spectroscopy, nitroxides, EPR oximetry, spin pH probes, *in vivo* EPR, thiol-specific spin labels, glutathione, redox state, EPR imaging, PEDRI (Proton-Electron Double Resonance Imaging)

1. INTRODUCTION

Fluorescence and magnetic resonance spectroscopies in combination with functional molecular probes are powerful instruments in biochemical and biomedical research. Absorption and fluorescent probes might be considered particularly effective for studies at the cellular and subcellular levels, while magnetic resonance spectroscopy has the advantage of *in vivo* applications in animals and humans. Broad clinical applications of NMR and MRI are possible due to the existence of endogenous NMR-sensitive nuclei, up to 110 M concentration in the case of water protons. EPR spectroscopy has tremendous advantage over NMR in sensitivity by three orders of magnitude or more. However two main factors limit biomedical EPR applications: (i) low depth of penetration of the microwaves in aqueous samples and (ii) absence of endogenous paramagnetic species. Over the past decade, despite formidable technical problems, significant progress has been made regarding *in vivo* EPR techniques. Low-field L-band EPR spectrometers, designed to increase the depth of microwave penetration and decrease non-resonant losses, has become commercially available. Moreover even lower-field (down to 300 MHz) home-made RF-EPR spectrometers, as well as instrumentation for spatial and spectral-spatial EPR imaging of radicals have been constructed (Lurie 2001; He et al. 2002). Along with the progress in low-field RF-EPR, other EPR-based techniques such as Longitudinally-Detected EPR (LODEPR) (Nicholson et al. 1998), proton electron double-resonance imaging (PEDRI) (Lurie et al. 2002) and dynamic nuclear polarization (DNP) (Lurie et al. 1991) have been developed for *in vivo* applications. However the potential of the EPR-based techniques is still far from being maximally defined, predominantly because of the requirement for exogenous spin probes. EPR oximetry using oxygen-sensitive paramagnetic materials such as India ink, lithium phthalocyanine particulates and derivatives of triarylmethyl free radical is probably the most practical biomedical *in vivo* EPR application developed till present time (Swartz 2004). Among other paramagnetic probes, nitroxides represent a wide class of organic radicals varying in stability, spectral properties and functionality. The range of biological applications using stable nitroxides is immense and involves, for example, therapeutic agents, MRI contrast agents, probes for measurement of redox state (Kuppusamy et al. 2002), oxygen (Baker et al. 1997), pH (Mader et al. 1997;

Khramtsov et al. 2000), NO (Akaike et al. 1993; Joseph et al. 1993; Woldman et al. 1994), thiols (Berliner et al. 2001; Khramtsov et al. 2004a), and structural studies of biological macromolecules (Hubbell et al. 2000).

The first stable di-*tert*-alkyl nitroxide, 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) (Lebedev and Kayanovskii 1959) and TEMPO derivatives substituted at atom C-4 of heterocycle (Lebedev et al. 1961; Neiman et al. 1962) have been reported almost fifty years ago. Several years later EPR spectral sensitivities of the nitroxides to the local environment, including viscosity (Edelstein et al. 1964) and polarity (Il'yasov 1962), were reported. Moreover, the stability of N-O fragment allowed chemical reactions of the nitroxides which do not involved radical center (Neiman et al. 1962). McConnell with coauthors greatly stimulated nitroxides applications establishing EPR spin labeling technique based on the ability of the nitroxides to be bound to biologically relevant macromolecules and sensitivity of their EPR spectra to molecular motion and microenvironment (Stone et al. 1965; Griffith and McConnell 1966). Two decades later combination of EPR spin labeling with site-directed mutagenesis reactions resulted in establishment of the site-directed spin labeling (SDSL) approach widely used in the studies of protein structure and dynamics (Altenbach et al. 1989; Hubbell et al. 1996; Hubbell et al. 2000). Sensitivity of the EPR spectra of the nitroxides to Heisenberg spin exchange (Molin et al. 1980) or a magnetic dipole-dipole interactions (Rabenstein and Shin 1995) provides instrument for the studies of intermolecular distances, and localization of paramagnetic species such as oxygen or paramagnetic ions.

Chemical reactions of the nitroxides may provide additional functionality to the EPR spin probes beyond modifying biological macromolecules. Apparently the most known reaction of the nitroxides, its reduction to EPR-silent hydroxylamines, is largely responsible for the biodegradation of the nitroxides in living tissues (Kocherginsky and Swartz 1995) and significantly limits their applications. On other side, the rate of the nitroxides reduction can provide important data on redox state in living tissues (Kuppusamy et al. 2002). The reactions of the nitroxides which do not involve radical center but other functional groups in the vicinity of NO-fragment also can be studied by EPR (Khramtsov and Volodarsky 1998). The corresponding functionally-enhanced nitroxides for detection of specific biologically relevant molecules, among them spin pH-, SH-, and NO-sensitive probes, were developed (Khramtsov and Volodarsky 1998; Khramtsov et al. 2004a). In this contribution we reviewed applications of EPR spectroscopy and imaging of the nitroxides for the probing local chemical environment, including oxygen, redox state, thiols, pH, and nitric oxide, with particular emphasis on *in vivo* EPR applications.

2. NITROXIDES AS FUNCTIONAL EPR PROBES

2.1. EPR oximetry

Electron paramagnetic resonance oximetry is one of the most promising and rapidly developing techniques for measurement of oxygen in living tissues (Swartz 2004). Initially, the paramagnetic probes for EPR oximetry were almost exclusively nitroxides (Backer et al. 1977; Lai et al. 1982; Froncisz et al. 1985; Bacic et al. 1989; Chan et al. 1989; Hyde and Subszynski 1989). The physical basis of the spin label oximetry method depends on Heisenberg exchange between nitroxide radical spin labels and molecular oxygen, which is a stable diradical. Both the longitudinal (T_1) and transverse (T_2) relaxation time of the nitroxide labels can be rather close to the bimolecular collision rate of oxygen dissolved in water under standard conditions and be significantly affected by the encounter rate, which gives the method a good sensitivity. According to early work of Pake and Tuttle (1959) (Pake and Tuttle 1959), the EPR line exchange broadening is proportional to radical-radical collision rate, w. The encounter rate is governed by the Smoluchowski equation, $w = 4\pi \cdot r \cdot D \cdot [0_2]$ (r is the interaction distance, and D is the diffusion constant of oxygen, which is assumed to be much greater than the diffusion constant of the nitroxide), and, therefore, is proportional to oxygen concentration. First observation of the linewidth broadening of the nitroxide TEMPONE (4-oxo-2,2,6,6-tetramethylpiperidine-1-oxyl) by dissolved oxygen in various solvents was reported by Povich in 1975 (Povich 1975). Two years later the T_2 oximetry method was introduced by Backer et al. (Backer et al. 1977). The authors applied EPR oximetry using the nitroxide label to measure oxygen concentration during biochemical reactions and were able to detect mitochondrial respiration in a volume of about 1 μl and cellular respiration in the samples containing about 100 liver cells. The T_1-sensitive EPR oximetry methods were introduced by Hyde et al. (Froncisz et al. 1985; Hyde et al. 1990). For most biological applications in low viscosity solutions, T_2-sensitive methods are preferred because they are instrumentally easy. However for highly viscous environments or spin-labeled macromolecules, T_1-sensitive methods might have an advantage. This is because for the nitroxides the $T_2 << T_1$, and for the optimal sensitivity the collision rate needs to be comparable to the relaxation time.

An advantage of EPR spin label oximetry is that it is based on pure physical interaction between the label and oxygen molecules and does not interfere with the oxygen metabolism. Polarographic methods are known to distort the distribution of oxygen in the samples (Albanese 1973) causing repetitive measurements to be problematic (Gnaiger 2001). The effect of

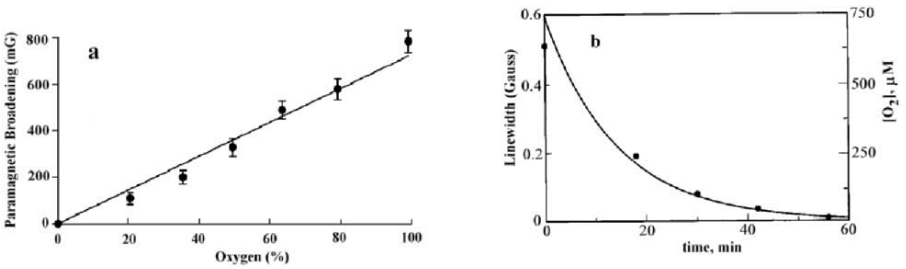


Figure 2.1. (a) EPR line-broadening as a function of oxygen partial pressure at 25°C measured in a saline solution of 0.25 mM nitroxide label, ^{15}N-labeled 4-oxo-2,2,5,5-tetramethyl-pyrrolinyl-d_{16}-N-oxy (perdeuterated tempone, PDT) equilibrated with mixtures of nitrogen and oxygen gases (Velan et al. 2000); (b) Measurements of oxygen concentration in ischemic heart performed from the changes in L-band EPR linewidth (Kuppusamy et al. 1994) of TEMPO spin label. The heart was infused with 1 mM TEMPO in Krebs bicarbonate-buffered perfusate solution for 10 min to reach steady state, then it was cardiopleged and subjected to a global ischemia at 24°C. The linewidth decreased by 0.45 G over a 60-min duration of ischemia, and approached the linewidth observed in the absence of oxygen. This corresponds to a drop in myocardial oxygen concentration from 625 μM to near 0 μM.

oxygen on the broadening of the nitroxide linewidth is equal to approximately 0.1 G for air-equilibrated (21% oxygen) aqueous solutions (Fig. 2.1). This is only about 10% from the typical linewidth of the nitroxides, about 1 Gauss in the absence of oxygen. For the samples with low oxygen tension and low viscosity an application of the nitroxides with the resolved components of the superhyperfine structure with the individual linewidth about 0.2 G allows to increase the sensitivity to oxygen (Backer et al. 1977; Sarna et al. 1980; Lai et al. 1982). Similar increase in sensitivity is achieved by application of perdeuterated nitroxides with the linewidth of about 0.2 Gauss (Halpern et al. 1994; Kuppusamy et al. 1994; Gallez et al. 1996a; Velan et al. 2000). Encapsulation of the nitroxides in lipophilic environment such as proteinaceous microspheres filled with organic liquid provides another way to enhance sensitivity to oxygen which is more soluble in organic solvents than in water (Liu et al. 1994).

A well-developed chemistry of the nitroxides (Volodarsky et al. 1994; Hideg et al. 2005) allows manipulation of their structure and properties, including presence of charged, ionizable, hydrophilic or hydrophobic groups which determine label solubility and ability to penetrate biomembranes. For example, small neutral nitroxide will be distributed throughout the intracellular and extracellular environments while a charged nitroxide will not cross the plasma membrane and thus can be used to measure the concentration of oxygen in the extracellular compartment (Baker et al. 1997). Moreover, nitroxides can be linked to carrier molecules (Gallez et al. 1993) or incorporated in liposomes (Glockner et al. 1991; Glockner et al. 1993) to

achieve a tissue or organ selectivity. They can be injected into the animal intraperitoneally or intravenously, distributed into a variety of pharmacological compartments, and structurally modified to target a specific distribution compartment. Figure 2.1(b) demonstrates an example of EPR oximetry application for monitoring oxygen depletion in ischemic myocardium obtained in isolated rat heart (Zweier and Kuppusamy 1988; Kuppusamy et al. 1994) after the infusion of TEMPO spin label. The spectral data from the images of the cardiopleged heart showed a gradual decrease in the linewidth at each spatial location over a 60 min duration of ischemia by 0.45 G down to the linewidth observed in the absence of oxygen. The authors demonstrate an ability of the technique noninvasively report spatially resolved myocardium oxygenation. To obtain oxygen mapping the spectral–spatial imaging was applied which contains a complete spectral profile, as a function of field, at each spatial voxel element. Spectral–spatial imaging can be performed in one, two or three spatial dimensions using one to three sets of gradient coils resulting in 2-, 3- or 4D spectral–spatial images, respectively (Kuppusamy and Zweier 2004). Because the spatial and spectral dimensions are fully separable, information about local linewidth, and hence local oxygen concentration, can be derived independently from local spin density. The continuous wave EPR imaging technique, CW EPRI, fits better for the nitroxides imaging while pulsed Fourier transform, FT EPRI, is limited to narrow line probes with linewidth of the order of 100 mG or less (Murugesan et al. 1997).

The oxygenation status of normal and tumor tissues is important parameter being in part responsible for the effectiveness of cancer therapy. Halpern et al. (Halpern et al. 1994) demonstrated an efficiency of low-field 250 MHz EPR oximetry in combination with hydrophilic and partially deuterated nitroxide probe (3-carbamoyl-2,2,5,5-tetraperdeuteromethyl-3-pyrrolinyl-1-^{15}N-oxy) to report oxygen concentration in the body water of murine FSa and NFSa fibrosarcomas 7 cm deep in the tissues of a living animal. An oxymetric 2D (1spectral/1 spatial) spectral image of the tumor was obtained allowing, in principle, direct assessment of tumor hypoxia to determine the usefulness of radiation and chemotherapy adjuvants directed to hypoxic cell compartments. Kuppusamy et al. (Kuppusamy et al. 1998) applied hydrophobic probe, PDT, to access oxygenation level in lipophilic compartments with higher oxygen solubility in the the RIF-1 (radiation induced fibrosarcoma) tumor model in mice. The data showed a substantial difference in the level of oxygenation between normal (muscle) and tumor tissues being 3-fold lower for the tumor.

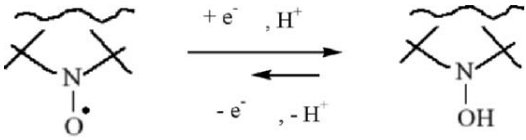
Nitroxides were first paramagnetic probes used for EPR oximetry and still keeping advantages in variability in structure, solubility and ability to be targeted. Nitroxides were also the first compounds that were tested in

dynamic nuclear polarization to probe the oxygen environment (Guiberteau and Grucker 1997). However, comparatively fast reduction of the nitroxides in EPR-silent hydroxylamine, as well as multiplet spectral pattern with broad spectral lines limit their applications, particularly for EPR oxygen mapping. The applications of other oxygen-sensitive paramagnetic materials include lithium phthalocyanine particulates (Liu et al. 1993; Presley et al. 2006), carbonaceous materials (chars, coals, carbon blacks) (Clarkson et al. 1998), and derivatives of triarylmethyl free radical (Andersson et al. 1996; Ardenkjaer-Larsen et al. 1998; Krishna et al. 2002). It should be noted that particulate probes such as lithium phthalocyanine and synthetic char are suitable for measurements of oxygen partial pressure, whereas soluble probes such as nitroxides and trityl compounds measure dissolved oxygen concentration (Swartz 2004; Presley et al. 2006).

2.2. Redox- and thiol-sensitive nitroxide probes

Chemical reduction of NR to EPR-silent hydroxylamines in many cases is an unfavorable factor that significantly limits their applications in biological systems (Kocherginsky and Swartz 1995). On other side, EPR-measured rates of NR reduction provide information on tissue redox status (Kuppusamy et al. 2002; Matsumoto et al. 2004; Yamada et al. 2004). NR being introduced in the biologically relevant systems are predominantly observed in the radical and hydroxylamine forms (Swartz and Timmins 2000). The reduction of the nitroxides to hydroxylamines by cells is primarily intracellular and, therefore the rate of the nitroxide reduction in great extent is determined by intracellular redox capacity. The intracellular thiols, and particularly the redox couple of glutathione, GSH, and its disulfide form, GSSG, is considered as the major regulator of the intracellular redox state (Schafer and Buettner 2001). The oxidation-reduction status of blood glutathione is especially used in investigations involving oxidative stress and free radical pathologies. Low GSH, high GSSG, and a lower GSH/GSSG ratio have been found in blood from patients with various pathologies (Yelinova et al. 1996; Herzenberg et al. 1997; Samiec et al. 1998; Rossi et al. 2002). Because blood GSH concentrations may reflect the status of glutathione in other less accessible tissues, measurement of both GSH and GSSG in blood often is used as an index of whole-body glutathione status and is a useful indicator of disease risk in humans.

Scheme 2.1 illustrates nitroxide/hydroxylamine redox couple. For the most biologically relevant experimental situations equilibrium is strongly shifted towards hydroxylamine formation, and therefore nitroxides undergoes reduction to EPR-silent product (Fig. 2.2). Hence the EPR-measured



Scheme 2.1. Illustration of the nitroxide/hydroxylamine redox couple. Commonly, for the most biologically relevant samples one-electron reduction of the nitroxides prevails and the equilibrium is strongly shifted towards hydroxylamine form.

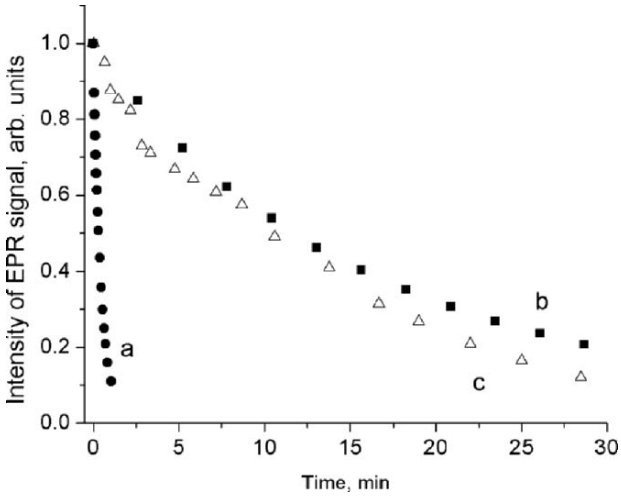


Figure 2.2. The kinetics of the loss of EPR signal of the nitroxide, 4-amino-2,2,5,5-tetramethyl-3-imidazoline-1-yloxy (ATI), in various reducing samples: (a) in the presence of 1 mM ascorbic acid in 0.1 M Na-phosphate buffer, pH 7.5, T = 24°C, initial radical concentration was 50 μM; (b) in the rat blood after mixing 20 μl of 0.5 mM nitroxide solution in 0.1 M sodium phosphate buffer, pH = 7.5, containing 0.1 mM DTPA, with 180 μl of untreated rat blood at ambient conditions, T = 24°C; (c) in isolated ischemic rat heart. The solution of 5 mM nitroxide (2 ml) in 0.9% NaCl in perfused buffer was administered in the heart on the onset of ischemia. Rat heart was perfused in the non-recirculating Langendorff mode by Krebs-Henseleit solution at 37°C. The measurements during ischemia were performed at 25°C.

rate of reduction depends on overall tissue redox status allowing to differentiate normal and pathological states (Kuppusamy and Krishna 2002; Matsumoto et al. 2004; Yamada et al. 2004).

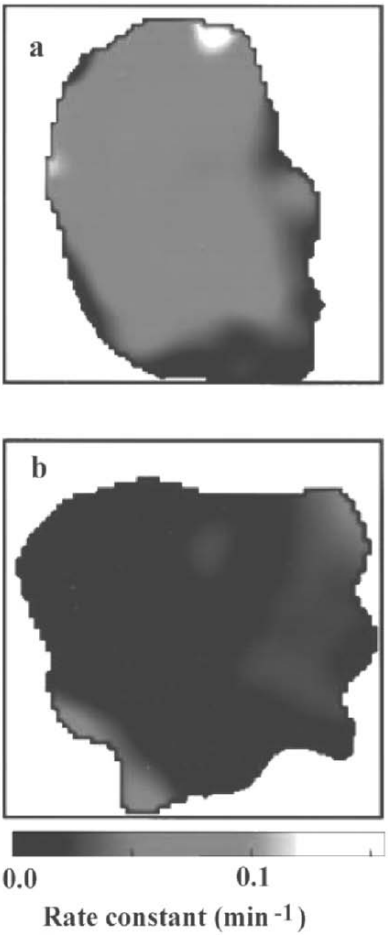


Figure 2.3. Redox mapping of tumor. 2-D spatial mapping of pseudo-first order rate constants of the nitroxide 3-CP (3-carbamoyl-2,2,5,5-tetramethylpyrrolidine-N-oxyl) reduction obtained in RIF-1 tumors implanted in the upper leg of mouse and measured using *in vivo* L-band (1.2 GHz) EPR imaging (Kuppusamy and Krishna 2002). The rate of reduction was significantly slower in tumors of mice treated with 2.25 mmol/kg of L-buthionine-S,R-sulfoximine (BSO), a GSH depleting agent, for 6 hours.

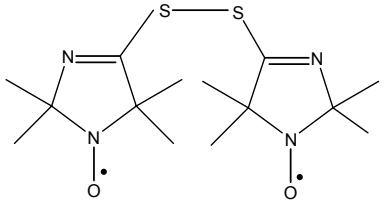
Figure 2.3 demonstrates application of L-band EPRI for redox mapping of the tumor in living mice (Kuppusamy and Krishna 2002; Kuppusamy et al. 2002) using 3-CP nitroxide (3-carbamoyl-2,2,5,5-tetramethylpyrrolidine-N-oxyl). Significant decrease in the rate of the loss of EPR signal of the probe after mice treatment by GSH depleting agent, BSO, clearly demonstrates a central role of glutathione in the reduction of the nitroxide. Note that appreciable chemical reduction of NR by glutathione does not occur over a few hours (Finkelstein et al. 1984; Glebska et al. 2003; Bobko et al. 2007). However glutathione can significantly contribute in the reduction of NR in biological systems indirectly by acting as a secondary source of reducing equivalents (Takeshita et al. 1999; Kuppusamy et al. 2002; Bobko et al. 2007). To perform direct quantitative measurements of GSH, the

application of a SH-specific spin probes is required (Khramtsov et al. 1989b; Weiner et al. 1991; Khramtsov et al. 2004a).

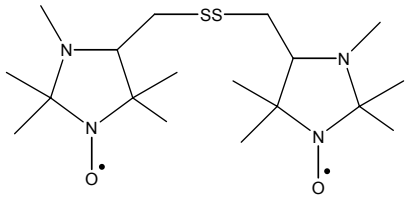
The thiol-specific reagents include such reactive functionalities as maleimides, iodoacetates, organomercurials, alkylthiosulfonates, and disulfides. The first maleimide nitroxide label was proposed by Griffith and McConnell (Griffith and McConnell 1966) thirty years ago in their pioneering work in the EPR spin labeling field. Maleimide and iodoacetamido nitroxides of piperidine, pyrrolydine, pyrrolinepirrolydine and pirroline types were applied in numerous biophysical applications as thiol-specific spin labels (Hensley et al. 1994; Gabbita et al. 1998; Alonso et al. 2001) (Marsh and Henderson 2001). However at elevated pH they can react with amines or may be hydrolyzed to non-reactive products, therefore compromising labeling specificity. Isothiocyanate - (Konczol et al. 1998), alkylhalide - (Balthasar 1971; Popova et al. 1982; Hideg et al. 2005), indanedione- (Roopnarine et al. 1993) and even organomercurial-containing nitroxides (Hideg et al. 2005) were synthesized and may have advantages in some specific applications for the labeling of thiols. The alkylthiosulfonate spin labels developed by Hideg et al. (Berliner et al. 1982; Hideg et al. 2005) are distinguished by their extremely rapid reactivity under mild conditions, high selectivity for sulfhydryl groups, and the general reversibility of the formation of disulfide bonds upon the addition of thiols such a β-mercaptoethanol or dithiothreitol. The synthesis of the methanethiosulfonate spin label (Berliner et al. 1982), greatly facilitated the development of an important area of spin labeling applications, site-directed spin labeling, SDSL (Hubbell et al. 2000).

An EPR spectroscopy in combination with thiol-specific nitroxides allows determination of the accessible thiol groups in various biological macromolecules, such as human plasma low-density lipoproteins (Kveder et al. 2003) and erythrocyte membranes (Soszynski and Bartosz 1997). The EPR characterization of the protein sulfhydryl groups is based on the immobilization of the nitroxide upon binding to the macromolecular structure and normally requires purification of the sample from the unbound label. The measurement of glutathione and other low-molecular-weight thiols using these thiol-specific nitroxides is difficult due to insignificant EPR spectral changes of the label upon binding. The development of the biradical disulfide labels allowed overcoming the latter limitation (Khramtsov et al. 1989b; Khramtsov et al. 1997). The principal advantage of the biradical disulfide reagents is the large EPR spectral changes that accompany their reaction with low-molecular-weight thiols such as GSH (Fig. 2.4).

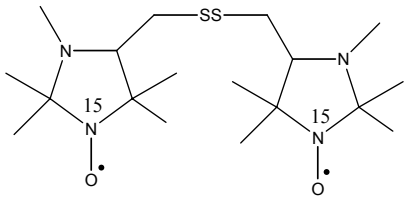
Disulfide biradicals of imidazoline R_1SSR_1 (Khramtsov et al. 1989b) and imidazolidine R_2SSR_2 (Khramtsov et al. 1997) types (Scheme 2.2 for the structures) have been synthesized and applied for thiols detection by EPR in numerous biological modalities.



R_1SSR_1



R_2SSR_2



$^{15}N\text{-}R_2SSR_2$

Scheme 2.2. Structures of the disulfide biradical reagents, R_1SSR_1, R_2SSR_2 and its ^{15}N-substituted analog, $^{15}N\text{-}R_2SSR_2$.

Disulfide biradicals, being paramagnetic analogs of widely used Ellman's reagent (Ellman 1959), react with thiols, BSH, via the reversible reaction of thiol-disulfide exchange:

$$RS\text{-}SR + BSH <=> RSSB + RSH \qquad [1]$$

EPR spectra of the biradicals are significantly affected by spin exchange between two radical fragments. Figure 2.4 demonstrates the typical changes of the EPR spectra of the biradicals R_2SSR_2 and its isotopically-substituted derivative upon an addition of GSH. The decrease of biradical spectral components and simultaneous increase of the monoradical ones is consistent with the splitting of the disulfide bond of the biradical and the formation of two monoradicals with conventional triplet EPR spectra or doublet EPR spectrum for ^{15}N-substituted analogue. In the case of GSH and cysteine, similar thiol-induced EPR changes were observed (Khramtsov et al. 1997) supporting the very small influence of the nature of the low-molecular weight thiol, BSH, on the magnetic resonance parameters of the resulting monoradical, RS-SB.

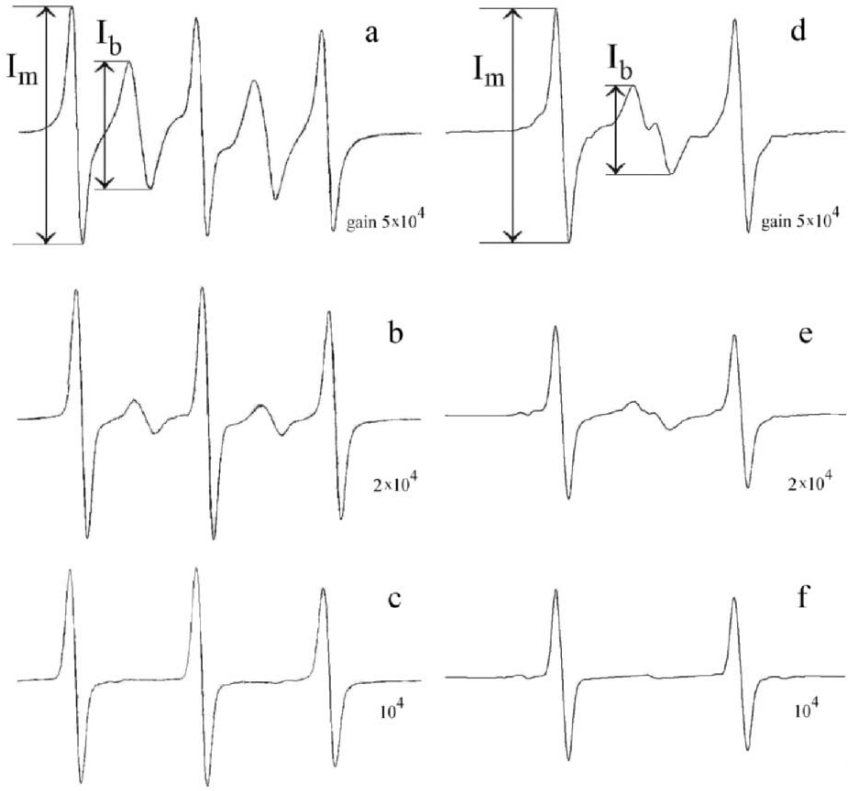


Figure 2.4. X-band EPR spectra of 0.1 M aqueous solutions of the disulfide labels R_2SSR_2 (a-c) and ^{15}N-R_2SSR_2 (d-f) alone and after various additions of the GSH: 30 μM (b), 10 μM (e) and 1 mM (c, f). The spectrometer settings were as following: total scan range, 50 G; microwave power, 20 mW (a-c) or 10 mW (d-f); modulation amplitude, 0.8 G (a-c) or 0.5 G (g-k); and gain indicated near the spectra.

The biradicals R_1SSR_1 and R_2SSR_2 being lipophilic compounds diffuse easily across cellular membranes. The biradical R_1SSR_1 reacts with intracellular glutathione at physiological pH within a few seconds, therefore providing a fast and reliable EPR approach for determination of GSH in optically nontransparent samples. The approach is based on a dominant contribution of the GSH in the intracellular pool of the fast-reacting thiols. The authors (Khramtsov et al. 1989b; Weiner et al. 1991; Balcerczyk and Bartosz 2003; Balcerczyk et al. 2003; Bratasz et al. 2003; Bratasz et al. 2006) used R_1SSR_1 label to measure GSH and/or total thiols in various cells and cellular homogenates. The sensitivity of the EPR approach allows detection of the intracellular GSH in a few hundred cells (Weiner et al. 1991). The EPR studies of the thiols in human and rat blood show an increased level of

the oxidized GSH in the blood plasma under oxidative stress conditions, including some human pathologies such as kyphoscoliosis (Yelinova et al. 1996). The application of the above approach to thiols measurements in the isolated perfused rat hearts showed GSH depletion related to ischemia/ reperfusion oxidative damage of the myocardium (Nohl et al. 1995). Note that the basis of EPR detection of thiols using R_1SSR_1 is its fast reaction with GSH and other low-molecular-weight thiols. This approach is based on the measurement of EPR spectral changes after complete binding of all reactive thiols to the label according to reaction (1). This requires an excess of the label over thiols. The method, being simple and reliable for *in vitro* systems, cannot be used *in vivo* due to its invasiveness because vital thiols are consumed during the measurement.

The biradical R_2SSR_2 reacts with thiols in about four orders of magnitude slower than R_1SSR_1, e.g. the bimolecular rate constant $k_1(R_2SSR_2) = 0.26$ $M^{-1}s^{-1}$ (Khramtsov et al. 1997) and $k_1(R_1SSR_1) \approx 5\times10^3$ $M^{-1}s^{-1}$ (Khramtsov et al. 1989). This complicates the use of the R_2SSR_2 reagent due to possible incomplete reaction with the thiols and, therefore, requirement of a longer incubation time. On the other hand, the "kinetic" approach for measuring thiols using lower concentrations of the label, $[R_2SSR_2] \ll [GSH]$, becomes possible. This approach eliminates the requirement of total consumption of the thiols during the measurement, therefore making it noninvasive and feasible for *in vivo* use. Figure 2.5 shows the increase of the peak intensity, I_m, of the biradical R_2SSR_2 measured by L-band EPR spectroscopy in rat blood during convenient time scale of about ten minutes. The analysis of the initial part of the kinetics allowed estimation of GSH concentration (Khramtsov et al. 2004a). This approach has been used to demonstrate increased depletion of GSH in the rat blood in an animal model of stress-sensitive arterial hypertension (Yelinova et al. 1999). The potential of R_2SSR_2 label for *in vivo* application is enhanced for isotopically-substituted analogues (Roshchupkina et al. 2006). Thus ^{15}N-R_2SSR_2 has higher spectral sensitivity and a simplified EPR spectrum (Fig. 2.4), which especially important for imaging applications.

2.3. EPR spectroscopy and imaging of pH

Local pH is among the most important parameters in the biochemistry of living organisms. Stable nitroxides of the imidazoline and imidazolidine types have been shown to be useful spin probes for EPR spectroscopy and imaging of pH (Khramtsov et al. 2000; Khramtsov 2005) due to the large effect of pH on their EPR spectra. Scheme 2.3 illustrates the chemical origin of the pH effect on EPR spectra of the imidazoline radical, ATI.

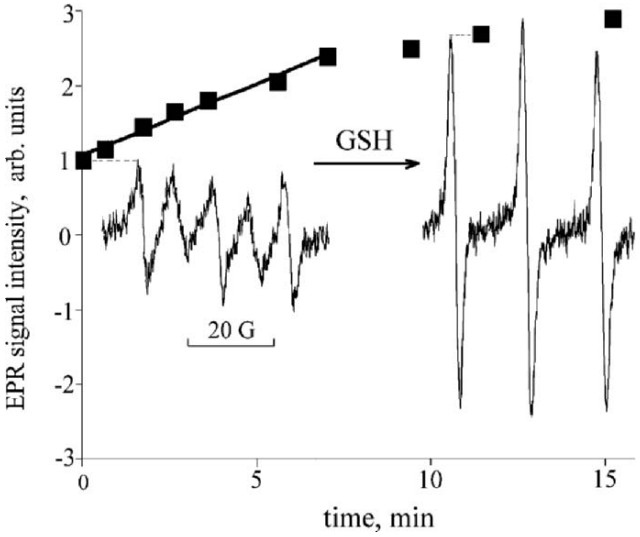
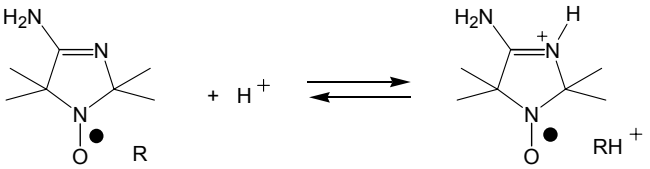


Figure 2.5. The kinetics of the increase of the monoradical spectral peak intensity, I_m (see Fig. 2.4), of the biradical R_2SSR_2 measured in blood from a Sprague-Dawley rat. The spectra are L-band EPR spectra of the 0.1 ml blood sample measured immediately (left) and 12 min after addition of 5 μl of the biradical (right) to achieve a final concentration 50 μM. Acquisition time 10.5 s. An increase of the monoradical component is due to the splitting of the disulfide bond of the R_2SSR_2 radical in accordance with equation [1] and corresponding formation of two monoradicals, RSH and RSSB with practically identical triplet EPR spectrum.

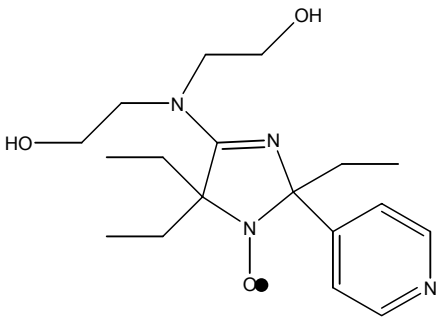


Scheme 2.3. Reversible protonation of the nitrogen atom N-3 of the ATI nitroxide results in an EPR-detected difference in hyperfine splitting, a_N and g-factor (Δa_N = 0.8 G and Δg = 0.0002) between R and RH+ forms (Khramtsov et al. 1982).

The pH dependence of the relative contribution of the protonated (RH^+) and unprotonated (R) forms of the probe in the EPR spectrum provides the basis for the application of these nitroxides for pH determination. One of the great strengths of the technique is that it is ratiometric, or in other words, the pH measurement is independent of the concentration of the probe but rather depends on the ratio $[RH^+]/[R]$. In general spectral simulation is required for accurate $[RH^+]/[R]$ determination.

In practice, two convenient spectral parameters can be used as markers of pH. The first one is the ratio of peak intensities of RH^+ and R spectral components resolved upon detection by high-frequency EPR and partly resolved in X (9.5 GHz)-band EPR spectra (Khramtsov et al. 2004a). The second one is nitrogen hyperfine splitting, a_N, measured as a distance between unresolved spectral components and being used almost exclusively as a highly sensitive pH marker in numerous applications (Fig. 2.6(b)). Note that sensitivity of the latter experimental parameter to pH strongly depends both on EPR frequency and spectrometer settings (e.g. modulation amplitude) and can be optimized (Khramtsov et al. 2004b).

Up to the present time we have developed a wide set of pH-sensitive nitroxides, with different ranges of pH sensitivity (centered on pK_a of the probe), labeling groups, and lipophilicity (Khramtsov and Weiner 1988; Kirilyuk et al. 2004; Kirilyuk et al. 2005; Voinov et al. 2005). These spin pH probes, together with low-field EPR spectroscopy, offer unique opportunities for non-invasive pH assessments in living animals in compartments with widely varying pH ranges. The potential applications are enormous, as tumors may have specific pH values compared to surrounding tissues, local skin treatments can have specific effects on local skin pH at various levels, local areas of infection or inflammation (such as in the pleural space, the peritoneum or the spinal cord) can exhibit specific localized reductions in pH allowing infection to be imaged and localized. L-band EPR spectroscopy (1.2 GHz) using spin pH probes has been shown to be a valuable tool for *in vivo* monitoring of microacidity in rodents (Gallez et al. 1996b; Mader et al. 1996; Khramtsov et al. 2000; Foster et al. 2003; Potapenko et al. 2006). ATI spin probe ($pK_a = 6.1$) was recently applied for the monitoring of ischemia-induced acidosis in isolated rat hearts (Khramtsov et al. 2002; Khramtsov 2005). Spin pH probes were used for *in vivo* studies stomach acidity in rats using L-band EPR, LODEPR, and Field-Cycled (FC)-DNP with its imaging analogue FC-PEDRI, operating at EPR frequencies as low as 120 MHz (Foster et al. 2003). Scheme 2.4 shows the structure of the nitroxide API with improved properties optimized for monitoring stomach acidity (Potapenko et al. 2006), and it demonstrates state-of-the-art structural design using synthetic power of organic chemistry of the nitroxides. Indeed, the presence of several hydrophilic functions, namely pyridine, nitroxyl, hydroxy, and amino groups increased aqueous solubility of the API probe preventing its membrane localization and redistribution from the stomach. To increase stability of the probes towards reduction, and therefore its lifetime *in vivo*, the bulky ethyl groups in the vicinity of the NO fragment were introduced. In addition to protonatable nitrogen N-3, a second protonatable group, pyridine was incorporated into the structure of the probes extending its range of pH sensitivity. Figure 2.6 demonstrates low-field 300 MHz EPR spectrum of the



Scheme 2.4. The chemical structure of API probe, 4-[bis(2-hydroxyethyl)amino]-2-pyridine-4-yl-2,5,5-triethyl-2,5-dihydro-1H-imidazol-oxyl, synthesized to optimize nitroxide properties for the measurement of stomach acidity *in vivo* (Potapenko et al. 2006).

API probe, calibration curve for hyperfine splitting, a_N, and typical examples of API application for real-time monitoring of the stomach acidity of living rats (Potapenko et al. 2006). Improved probe characteristics allowed authors to follow *in vivo* the drug-induced perturbation of the stomach acidity and its normalization afterwards during 1 h or longer period of time by low-field EPR based techniques. The results show the applicability of the techniques for monitoring drug pharmacology and disease in the living animals.

Low depth of microwave penetration does not allow X-band application *in vivo* but was found to be very effective for noninvasive pH mapping in small samples *in vitro*, e.g. inside biodegradable polymers used as drug delivery systems (Mader et al. 1997). Spectral-spatial X-band EPR using imidazolidine spin pH probe has been applied for noninvasive direct and depthspecific measurement of pH within rat and human skin obtained from cosmetic surgery (Kroll et al. 2001). The penetration depth for L-band frequency, 1.2 GHz, in aqueous samples exceeds in almost ten times that for X-band frequency (9.9 GHz) being somewhat greater than 1cm. This allows pH mapping by L-band EPR spectroscopy *in vivo* in small animals such as mice. However loss in EPR sensitivity at low frequency aggravated by necessity of acquisition of numerous spectral projections with a variable strength of field gradients significantly limits spatial and temporal resolution of the approach. Application of low-field EPR spectra detection at high modulation amplitude provides significant improvement in sensitivity, allowing easy conversion of the 'position' image of the low- or high-field spectral component (Fig. 2.6(a)) in 'pH map' as shown in Figure 2.7. 3 D-image (1-spectral/2-spatial) of phantom sample demonstrates good spatial (0.2 mm) and functional (0.2 pH units at pH close to pK of the radical) resolution. pH mapping by low-field EPR approach may find applications in biomedical field for continuous and non-invasive spatiotemporal pH assessment *in vivo*.

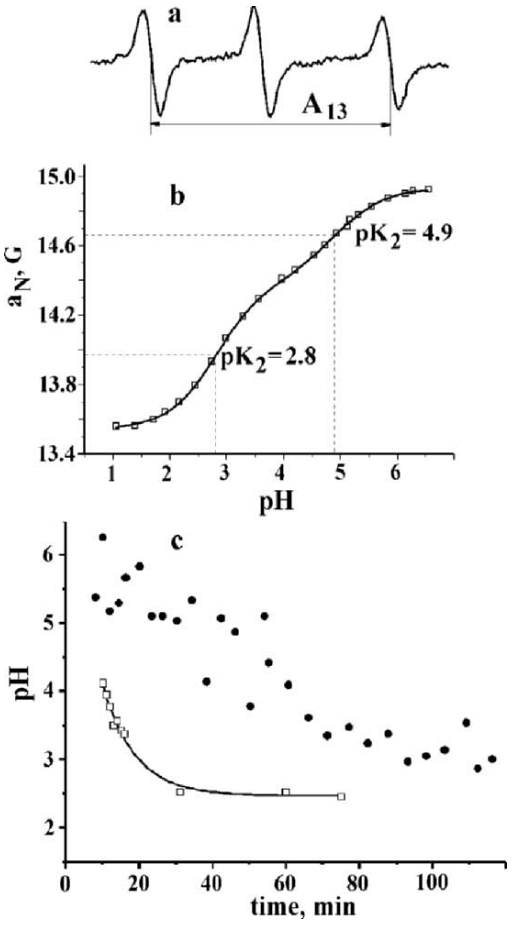


Figure 2.6. (a) 300 MHz EPR spectrum of 2 ml aqueous solution of 2 mM API nitroxide. Spectrometer settings were as follows: microwave power, 5 mW; modulation amplitude, 1 G. Note that low- and high-field components are not equidistantly positioned from the central line due to the deviation from high magnetic field approximation. Therefore, the Breit-Rabi (Breit and Rabi 1931) effect has to be taken into account for the accurate calculation of hyperfine interaction constant, a_N. Alternatively, a half of the distance between low- and high-field components, $A_{13}/2$, can be used for a_N estimation with a good accuracy (about 0.1 G for the API nitroxide at magnetic field of 100 G); (b) The dependence of hyperfine splitting, a_N, on pH calculated from corresponding 300 MHz EPR spectra of the API nitroxide as $A_{13}/2$ (see Fig. 2.6a). The solid line was calculated according the standard titration equation for the compound with two ionizable groups (Potapenko et al. 2006). Note convenient pH sensitivity range from pH 1.8 to pH 6, ideally optimized for the monitoring of the stomach acidity; (c) Time dependencies of pH changes in the stomach measured by LODEPR spectrometer operated at 304 MHz EPR excitation frequency, after giving 3 ml of gavage containing 5 mM nitroxide API alone (●) or with 50mM bicarbonate (□). After dosing the rats were placed in the LODEPR coil assembly and a series of spectra were obtained. The pH values from each spectrum were determined using corresponding calibration curve for a_N.

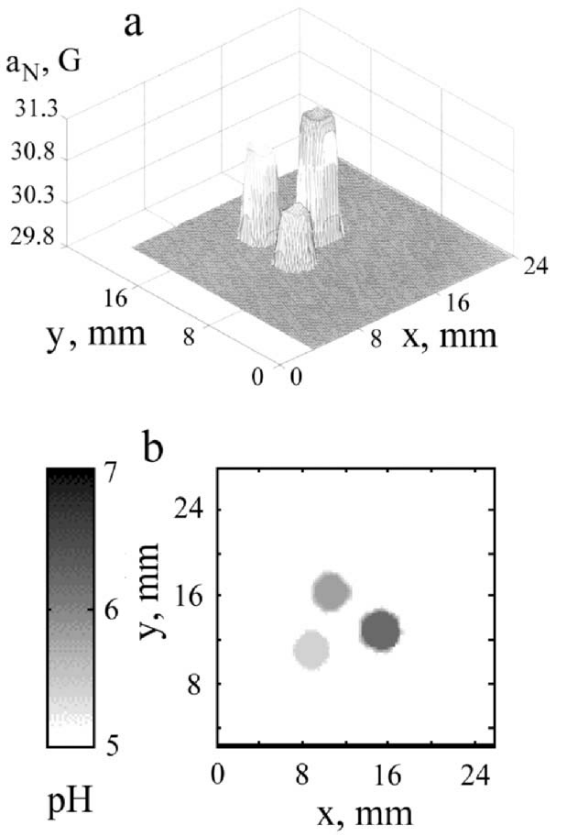
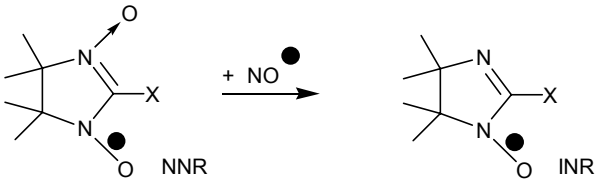


Figure 2.7. Cross-sectional 3D (1-spectral/2-spatial) image of a phantom prepared using capillary tubes of 2 mm diameter filled with 1 mM solutions of the radical ATI prepared at different pH: (a) spatial distribution of the hyperfine interaction splitting, A_{13}, determined from position mapping of the low-field peak of L-band EPR spectrum; (b) pH map. The data acquisition parameters were: acquisition time, 5 min; projections, 64; maximum gradient, 30 G/cm; field of view, 24 mm.

2.4. NO-sensitive nitroxides

The measurement of nitric oxide *in vivo* is of paramount importance due to its key role as a signaling molecule in numerous physiological and patho-physiological processes. The methods for NO detection in biological systems are limited because of its short half-life and its low concentration in living systems. Fe(II)-dithiocarbamate complexes (Mordvintcev et al. 1991), widely used for NO detection by EPR spectroscopy, have an important advantage of relative stability of the paramagnetic adduct of the reaction with NO.

However, disadvantages of the approach include the preparation of metal-chelator complexes from two components, which are accompanied by toxicity of both the iron and dithiocarbamate moieties. Therefore, application of a simple low-molecule weight organic molecule, 2-imidazoline nitroxide, or nitronyl nitroxide (NNR), as a paramagnetic trap for NO seems to be very attractive. The scavenging of NO by NNR has been used for its detection by EPR spectroscopy and to antagonize biological actions of nitric oxide (Akaike et al. 1993; Joseph et al. 1993; Woldman et al. 1994; Blasig et al. 2002; Cao and Reith 2002). The detection of NO using NNR is based on the specific radical-radical reactions accompanied by transformation to imino nitroxides, INR, and followed by dramatic changes in EPR spectra (Scheme 2.5 and Fig. 2.8).



Scheme 2.5. The reaction of the NNR with NO forming INR. Corresponding EPR spectral changes are shown in Fig. 2.8 for NNR1 (X = 2,6-difluorophenyl).

The physiological effects of NNR cannot be entirely explained by their specific radical-radical reactions with nitric oxide due to the fast reduction of NNR to its corresponding hydroxylamines. Recently we performed mechanistic studies of the NNR reaction with NO in reducing environment (Bobko et al. 2004). Figure 2.8 demonstrate an ability of the NNR to react with NO in the presence of ascorbate as reducing agent. It has been shown that equilibrium between radical, NNR, and its hydrohylamine is normally strongly shifted towards diamagnetic hydroxylamine, therefore no initial EPR spectra were observed in the presence of ascorbate (Fig. 2.8(a)). However NO generation and its consequent reaction with NNR results in accumulation of paramagnetic NNR and appearance of the corresponding EPR spectra (Fig. 2.8(b)). During the reaction, NNR is transformed to INR followed by further EPR spectra changes (Fig. 2.8(c,d)). After the reaction was complete, INR was reduced in its diamagnetic form by ascorbate (Fig. 2.8(e)). The corresponding scheme of the reactions (for details see ref. (Bobko et al. 2004)) for the first time provides a plausible mechanism explaining the antagonistic action of NNR against NO in a reducing environment, a phenomenon well documented *in vivo*. Note that the presence of fluorine atoms in the structure of NNR1 allowed us to visualize

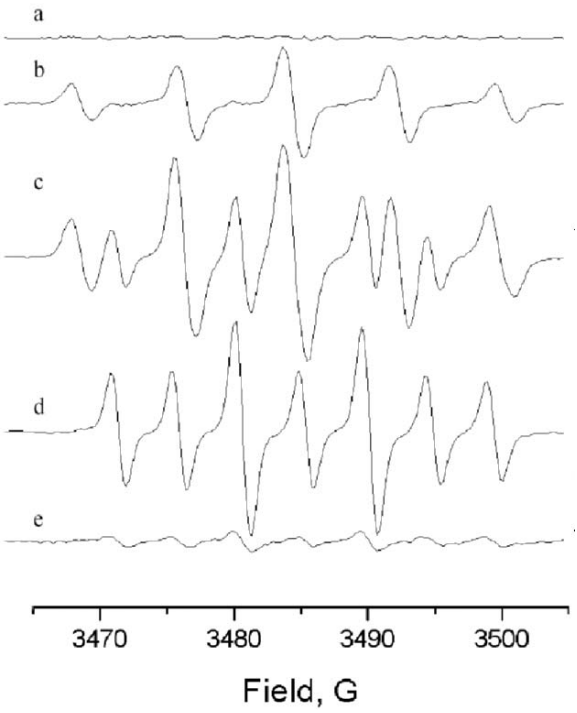


Figure 2.8. Time evolution of the EPR spectrum of the mixture of 3.33 mM NNR1 (X = 2,6-difluorophenyl, Scheme 5), 3.33 mM ascorbate, 33.3 mM SNP (sodium nitroprusside, light-sensitive NO donor), in 0.1 M sodium phosphate, pH 7.5, 0.1 mM DTPA, 8.5% DMSO before (a) and after 3 min (b), 6 min (c), 9 min (d) and 20 min (e) of the photolysis. The spectrometer settings were as follows: microwave power 10 mW, modulation amplitude 0.8 G, scan time 5 s.

diamagnetic hydroxylamine product by ^{19}F NMR spectroscopy providing additional proof of the redox-sensitive mechanism of the reaction of NNR with NO (Bobko et al. 2004).

The applications of NNR for NO detection by EPR in biological systems with physiologically low rates of NO generation is limited due to the extremely fast reduction of NNR, and particularly of INR, into diamagnetic EPR-silent product (Woldman et al. 1994; Haseloff et al. 1997; Bobko et al. 2004), so equilibrium radical concentration is lower than EPR detection level. Alternative approach for NO detection using NNR was recently developed and is based on application of fluorinated traps, such as NNR1, in combination with ^{19}F NMR spectroscopy. An *in vivo* accumulation of diamagnetic hydroxylamine of INR, product of the reaction of NNR with NO in reducing environment (Fig. 2.8), was measured by ^{19}F NMR in normotensive and

hypertensive rats. The NMR data were found to positively correlate with the levels of nitrite/nitrate evaluated by Griess method and negatively correlate with the blood pressure (Bobko et al. 2005). While ^{19}F NMR spectroscopy allows *in vivo* evaluation of NO production and provides the basis for in vivo NO imaging, the drawback of the method is low intrinsic NMR sensitivity. Therefore the development of the NNR structures with higher stability for the EPR detection in living tissues is desirable. Several attempts by our group and others to protect NNR against reduction by encapsulation in liposomal (Woldman et al. 1994) or dendrimeric (Rosen et al. 2003) structures resulted in significant increase of their lifetime but still further work has to be done to obtain NNR probes stable enough for *in vivo* applications.

3. CONCLUSION

In recent decade functional EPR spectroscopy and imaging applications move closer to biomedical field. The bottleneck of *in vivo* EPR applications is the requirement of appropriate exogenous paramagnetic probes. Among other probes, the nitroxides, stable organic free radicals, provide a wide range of available structures variable in solubility and tissue redistribution, spectral and functional sensitivity, ability to be targeted and their lifetimes in living tissues. Particularly an ability to follow by EPR specific chemical reactivity of the nitroxides towards biologically relevant compounds such as oxygen, thiols and other reducing species, protons (pH) and nitric oxide provide unique functionality to *in vivo* EPR spectroscopy and imaging. The nitroxides were the first compounds applied for EPR oximetry due to their ability to participate in spin exchange with diradical molecule of oxygen. The most known reaction of the nitroxides, its reduction to EPR-silent hydroxylamines, provides information on redox state in living tissues. Moreover, specific biradical nitroxides with intramolecular disulfide bonds were developed, and can be used for probing thiol redox state, e.g. intracellular glutathione concentration. The reversible reactions of the specific nitroxides, termed spin pH probes, with protons allowed EPR spectroscopy and imaging of pH. These spin pH probes already found applications in organ-specific real-time EPR measurements of pH in living animals. Finally, nitronyl nitroxides specifically react with nitric oxide, providing EPR tool for the detection of this extremely important in biological systems molecule which is very difficult to measure. One of the key disadvantages of the nitroxides applications for *in vivo* EPR is their limited stability in living tissues. Recent development of tetramethyl ‘substituted nitroxides provides one of the possible future directions in overcoming this limitation.

ACKNOWLEDGEMENTS

This work was partly supported by grants from NIH KO1 EB03519.

REFERENCES

Akaike, T., Yoshida, M., Miyamoto, Y., Sato, K., Kohno, M., Sasamoto, K., Miyazaki, K., Ueda, S. and Maeda, H. (1993) Antagonistic action of imidazolineoxyl N-oxides against endothelium-derived relaxing factor/NO through a radical reaction. *Biochemistry* **32**(3): 827-32.

Albanese, R.A. (1973) On microelectrode distortion of tissue oxygen tensions. *J. Theor. Biol.* **38**(1): 143-54.

Alonso, A., dos Santos, W.P., Leonor, S.J., dos Santos, J.G. and Tabak, M. (2001) Stratum corneum protein dynamics as evaluated by a spin-label maleimide derivative: effect of urea. *Biophys. J.* **81**(6): 3566-76.

Altenbach, C., Flitsch, S.L., Khorana, H.G. and Hubbell, W.L. (1989) Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. *Biochemistry* **28**(19): 7806-12.

Andersson, S., Radner, F., Rydbeek, A., Servin, R. and Wistrand, L.-G. (1996) Free radicals, Nycomed Imaging AS. U.S. Patent 5, 530, 140.

Ardenkjaer-Larsen, J.H., Laursen, I., Leunbach, I., Ehnholm, G., Wistrand, L.G., Petersson, J.S. and Golman, K. (1998) EPR and DNP properties of certain novel single electron contrast agents intended for oximetric imaging. *J. Magn. Reson.* **133**(1): 1-12.

Bacic, G., Nilges, M.J., Magin, R.L., Walczak, T. and Swartz, H.M. (1989) *In vivo* localized ESR spectroscopy reflecting metabolism. *Magn. Reson. Med.* **10**(2): 266-72.

Backer, J.M., Budker, V.G., Eremenko, S.I. and Molin, Y.N. (1977) Detection of the kinetics of biochemical reactions with oxygen using exchange broadening in the ESR spectra of nitroxide radicals. *Biochim. Biophys. Acta* **460**(1): 152-6.

Baker, J.E., Froncisz, W., Joseph, J. and Kalyanaraman, B. (1997) Spin label oximetry to assess extracellular oxygen during myocardial ischemia. *Free Radic. Biol. Med.* **22**(1-2): 109-15.

Balcerczyk, A. and Bartosz, G. (2003) Thiols are main determinants of total antioxidant capacity of cellular homogenates. *Free Radic. Res.* **37**(5): 537-41.

Balcerczyk, A., Grzelak, A., Janaszewska, A., Jakubowski, W., Koziol, S., Marszalek, M., Rychlik, B., Soszynski, M., Bilinski, T. and Bartosz, G. (2003) Thiols as major determinants of the total antioxidant capacity. *Biofactors* **17**(1-4): 75-82.

Balthasar, W. (1971) Spin labeling studies of D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **22**(2): 158-65.

Berliner, L.J., Grunwald, J., Hankovszky, H.O. and Hideg, K. (1982) A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal. Biochem.* **119**(2): 450-5.

Berliner, L.J., Khramtsov, V., Fujii, H. and Clanton, T.L. (2001) Unique *in vivo* applications of spin traps. *Free Radic. Biol. Med.* **30**(5): 489-99.

Blasig, I.E., Mertsch, K. and Haseloff, R.F. (2002) Nitronyl nitroxides, a novel group of protective agents against oxidative stress in endothelial cells forming the blood-brain barrier. *Neuropharmacology* **43**(6): 1006-14.

Bobko, A.A., Bagryanskaya, E.G., Reznikov, V.A., Kolosova, N.G., Clanton, T.L. and Khramtsov, V.V. (2004) Redox-sensitive mechanism of no scavenging by nitronyl nitroxides. *Free Radic. Biol. Med*. **36**(2): 248-58.

Bobko, A.A., Kirilyuk, I.A., Grigor'ev, I.A., Zweier, J.L. and Khramtsov, V.V. (2007) Reversible Reduction of Nitroxides to Hydroxylamines: the Roles for Ascorbate and Glutathione. *Free Rad. Biol. Med.* **42**(3).

Bobko, A.A., Sergeeva, S.V., Bagryanskaya, E.G., Markel, A.L., Khramtsov, V.V., Reznikov, V.A. and Kolosova, N.G. (2005) 19F NMR measurements of NO production in hypertensive ISIAH and OXYS rats. *Biochem. Biophys. Res. Commun*. **330**(2): 367-70.

Bratasz, A., Khramtsov, V.V. and Kuppusamy, P. (2003) A modified Tietze assay for the determination of thiols in intact cells and tissues. *Free Radic. Biol. Med*. **35** (Suppl. 1): S147.

Bratasz, A., Weir, N.M., Parinandi, N.L., Zweier, J.L., Sridhar, R., Ignarro, L.J. and Kuppusamy, P. (2006) Reversal to cisplatin sensitivity in recurrent human ovarian cancer cells by NCX-4016, a nitro derivative of aspirin. *Proc. Natl. Acad. Sci. USA* **103**(10): 3914-9.

Breit, G. and Rabi, I.I. (1931) Measurement of nuclear spin. *Phys. Rev*. **38**: 2082-2083.

Cao, B.J. and Reith, M.E. (2002) Nitric oxide scavenger carboxy-PTIO potentiates the inhibition of dopamine uptake by nitric oxide donors. *Eur. J. Pharmacol*. **448**(1): 27-30.

Chan, H.C., Glockner, J.F. and Swartz, H.M. (1989) Oximetry in cells and tissues using a nitroxide-liposome system. *Biochim. Biophys. Acta* **1014**(2): 141-4.

Clarkson, R.B., Odintsov, B.M., Ceroke, P.J., Ardenkjaer-Larsen, J.H., Fruianu, M. and Belford, R.L. (1998) Electron paramagnetic resonance and dynamic nuclear polarization of char suspensions: surface science and oximetry. *Phys. Med. Biol*. **43**(7): 1907-20.

Edelstein, N., Kwok, A. and Maki, A.H. (1964) Effects of hydrostatic pressure on linewidths of free radical in solutions. I. Anisotropic region. *J. Chem. Phys*. **41**: 179-183.

Ellman, G.L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys*. **82**(1): 70-7.

Finkelstein, E., Rosen, G.M. and Rauckman, E.J. (1984) Superoxide-dependent reduction of nitroxides by thiols. *Biochim. Biophys. Acta* **802**: 90-98.

Foster, M.A., Grigor'ev, I.A., Lurie, D.J., Khramtsov, V.V., McCallum, S., Panagiotelis, I., Hutchison, J.M., Koptioug, A. and Nicholson, I. (2003) *In vivo* detection of a pH-sensitive nitroxide in the rat stomach by low-field ESR-based techniques. *Magn. Reson. Med*. **49**(3): 558-67.

Froncisz, W., Lai, C.S. and Hyde, J.S. (1985) Spin-label oximetry: kinetic study of cell respiration using a rapid-passage T1-sensitive electron spin resonance display. *Proc. Natl. Acad. Sci. USA* **82**(2): 411-5.

Gabbita, S.P., Subramaniam, R., Allouch, F., Carney, J.M. and Butterfield, D.A. (1998) Effects of mitochondrial respiratory stimulation on membrane lipids and proteins: an electron paramagnetic resonance investigation. *Biochim. Biophys. Acta* **1372**(2): 163-73.

Gallez, B., Bacic, G., Goda, F., Jiang, J., O'Hara, J.A., Dunn, J.F. and Swartz, H.M. (1996a) Use of nitroxides for assessing perfusion, oxygenation, and viability of tissues: in vivo EPR and MRI studies. *Magn. Reson. Med*. **35**(1): 97-106.

Gallez, B., Debuyst, R., Demeure, R., Dejehet, F., Grandin, C., Van Beers, B., Taper, H., Pringot, J. and Dumont, P. (1993) Evaluation of a nitroxyl fatty acid as liver contrast agent for magnetic resonance imaging. *Magn. Reson. Med*. **30**(5): 592-9.

Gallez, B., Mader, K. and Swartz, H.M. (1996b) Noninvasive measurement of the pH inside the gut by using pH-sensitive nitroxides. An *in vivo* EPR study. *Magn. Reson. Med*. **36**(5): 694-7.

Glebska, J., Skolimowski, J., Kudzin, Z., Gwozdzinski, K., Grzelak, A. and Bartosz, G. (2003). Pro-oxidative activity of nitroxides in their reactions with glutathione. *Free Radic. Biol. Med*. **35**(3): 310-6.

Glockner, J.F., Chan, H.C. and Swartz, H.M. (1991) In vivo oximetry using a nitroxide-liposome system. *Magn. Reson. Med*. **20**(1): 123-33.

Glockner, J.F., Norby, S.W. and Swartz, H.M. (1993) Simultaneous measurement of intracellular and extracellular oxygen concentrations using a nitroxide-liposome system. *Magn. Reson. Med*. **29**(1): 12-8.

Gnaiger, E. (2001) Bioenergetics at low oxygen: dependence of respiration and phosphorylation on oxygen and adenosine diphosphate supply. *Respir. Physiol*. **128**(3): 277-97.

Griffith, O.H. and McConnell, H.M. (1966) A nitroxide-maleimide spin label. *Proc. Natl. Acad. Sci. USA*. **55**: 8-11.

Guiberteau, T. and Grucker, D. (1997) Dynamic nuclear polarization imaging in very low magnetic fields as a noninvasive technique for oximetry. *J. Magn. Reson*. **124**(1): 263-6.

Halpern, H.J., Yu, C., Peric, M., Barth, E., Grdina, D.J. and Teicher, B.A. (1994) Oxymetry deep in tissues with low-frequency electron paramagnetic resonance. *Proc. Natl. Acad. Sci. USA*. **91**(26): 13047-51.

Haseloff, R.F., Zollner, S., Kirilyuk, I.A., Grigor'ev, I.A., Reszka, R., Bernhardt, R., Mertsch, K., Roloff, B. and Blasig, I.E. (1997) Superoxide-mediated reduction of the nitroxide group can prevent detection of nitric oxide by nitronyl nitroxides. *Free Radic. Res*. **26**(1): 7-17.

He, G., Samouilov, A., Kuppusamy, P. and Zweier, J.L. (2002) *In vivo* imaging of free radicals: applications from mouse to man. *Mol. Cell Biochem*. **234-235**(1-2): 359-67.

Hensley, K., Carney, J., Hall, N., Shaw, W. and Butterfield, D.A. (1994) Electron paramagnetic resonance investigations of free radical-induced alterations in neocortical synaptosomal membrane protein infrastructure. *Free Radic. Biol. Med*. **17**(4): 321-31.

Herzenberg, L.A., De Rosa, S.C., Dubs, J.G., Roederer, M., Anderson, M.T., Ela, S.W., Deresinski, S.C. and Herzenberg, L.A. (1997) Glutathione deficiency is associated with impaired survival in HIV disease. *Proc. Natl. Acad. Sci. USA*. **94**(5): 1967-72.

Hideg, K., Kalai, T. and Sar, C.P. (2005) Recent results in chemistry and biology of nitroxides. *J. Heterocycl. Chem.* **42**: 437-450.

Hubbell, W.L., Cafiso, D.S. and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labeling. *Nat. Struct. Biol*. **7**(9): 735-9.

Hubbell, W.L., McHaourab, H.S., Altenbach, C. and Lietzow, M.A. (1996) Watching proteins move using site-directed spin labeling. *Structure* **4**(7): 779-83.

Hyde, J.S., Jin, J.-J., Felix, J.B. and Hubbell, W.L. (1990) Advances in spin label oximetry. *Pure & Applied Chem.* **62**: 255-260.

Hyde, J.S. and Subszynski, W.K. (1989) Spin label oximetry. In: Berliner, L.J. and Reubens, J. eds., *Spin Labeling: Theory and Application*. Plenum Press, New York, Vol. **8**, pp. 399-425.

Il'yasov, A.V. (1962) Solvent effects in the EPR spectra of certain free radicals. *J. Struct. Chem*. **3**(1): 84-86.

Joseph, J., Kalyanaraman, B. and Hyde, J.S. (1993) Trapping of nitric oxide by nitronyl nitroxides: an electron spin resonance investigation. *Biochem. Biophys. Res. Commun*. **192**(2): 926-34.

Khramtsov, V.V. (2005) Biological imaging and spectroscopy of pH. *Curr. Org. Chem.* **9**: 909-923.

Khramtsov, V.V., Grigor'ev, I.A., Foster, M.A. and Lurie, D.J. (2004a) *In vitro* and *in vivo* measurement of pH and thiols by EPR-based techniques. *Antioxid. Redox Signal*. **6**(3): 667-76.

Khramtsov, V.V., Grigor'ev, I.A., Foster, M.A., Lurie, D.J. and Nicholson, I. (2000) Biological applications of spin pH probes. *Cell Mol. Biol.* **46**(8): 1361-74.

Khramtsov, V.V., Grigor'ev, I A., Foster, M.A., Lurie, D.J., Zweier, J.L. and Kuppusamy, P. (2004b) Spin pH and SH probes: enhancing functionality of EPR-based techniques. *Spectroscopy* **18**: 213-225.

Khramtsov, V.V., Grigor'ev, I.A., Kirilyuk, I.A., Ilangovan, G. and Kuppusamy, P. (2002) *In vivo* EPR measurement of tissue acidosis during myocardial ischemia using pH-sensitive nitroxides. *Free Rad. Biol. Med.* **33**(Suppl. 2): S423-S424.

Khramtsov, V.V., Panteleev, M.V. and Weiner, L.M. (1989a) ESR study of proton transport across phospholipid vesicle membranes. *J. Biochem. Biophys. Methods* **18**(3): 237-46.

Khramtsov, V.V. and Volodarsky, L.B. (1998) Use of imidazoline nitroxides in studies of chemical reactions. ESR measurements of the concentration and reactivity of protons, thiols and nitric oxide. In: Berliner, L.J., ed., *Spin labeling. The next Millennium.* Plenum Press, New York, Vol. 14, p. 109-180.

Khramtsov, V.V. and Weiner, L.M. (1988) Proton exchange in stable nitroxyl radicals: pH sensitive spin probes. In: Volodarsky, L.B., ed., *Imidazoline nitroxides*. CRC Press, Boca Raton, FL., Vol. 2, 37-80.

Khramtsov, V.V., Weiner, L.M., Grigor'ev, I.A. and Volodarsky, L.B. (1982) Proton exchange in stable nitroxyl radicals. EPR studies of the pH of aqueous solutions. *Chem. Phys. Lett*. **91**: 69-72.

Khramtsov, V.V., Yelinova, V.I., Glazachev Yu, I., Reznikov, V.A. and Zimmer, G. (1997) Quantitative determination and reversible modification of thiols using imidazolidine biradical disulfide label. *J. Biochem. Biophys. Methods* **35**(2): 115-28.

Khramtsov, V.V., Yelinova, V.I., Weiner, L.M., Berezina, T.A., Martin, V.V. and Volodarsky, L.B. (1989b) Quantitative determination of SH groups in low- and high-molecular-weight compounds by an electron spin resonance method. *Anal. Biochem*. **182**(1): 58-63.

Kirilyuk, I.A., Bobko, A.A., Grigor'ev, I.A. and Khramtsov, V.V. (2004) Synthesis of the tetraethyl substituted pH-sensitive nitroxides of imidazole series with enhanced stability towards reduction. *Org. Biomol. Chem*. **2**(7): 1025-30.

Kirilyuk, I.A., Bobko, A.A., Khramtsov, V.V. and Grigor'ev, I.A. (2005) Nitroxides with two pK values–useful spin probes for pH monitoring within a broad range. *Org. Biomol. Chem*. **3**(7): 1269-74.

Kocherginsky, N. and Swartz, H.M. (1995) *Nitroxide spin labels. Reactions in biology and chemistry*, CRC Press, Boca Raton, FL.

Konczol, F., Lorinczy, D. and Belagyi, J. (1998) Effect of oxygen free radicals on myosin in muscle fibres. *FEBS Lett*. **427**(3): 341-4.

Krishna, M.C., English, S., Yamada, K., Yoo, J., Murugesan, R., Devasahayam, N., Cook, J. A., Golman, K., Ardenkjaer-Larsen, J. H., Subramanian, S. and Mitchell, J.B. (2002) Over-hauser enhanced magnetic resonance imaging for tumor oximetry: coregistration of tumor anatomy and tissue oxygen concentration. *Proc. Natl. Acad. Sci. USA* **99**(4): 2216-21.

Kroll, C., Hermann, W., Stosser, R., Borchert, H.H. and Mader, K. (2001) Influence of drug treatment on the microacidity in rat and human skin–an *in vitro* electron spin resonance imaging study. *Pharm. Res*. **18**(4): 525-30.

Kuppusamy, P., Afeworki, M., Shankar, R.A., Coffin, D., Krishna, M.C., Hahn, S.M., Mitchell, J.B. and Zweier, J.L. (1998) *In vivo* electron paramagnetic resonance imaging of tumor heterogeneity and oxygenation in a murine model. *Cancer Res*. **58**(7): 1562-8.

Kuppusamy, P., Chzhan, M., Vij, K., Shteynbuk, M., Lefer, D.J., Giannella, E. and Zweier, J.L. (1994) Three-dimensional spectral-spatial EPR imaging of free radicals in the heart: a technique for imaging tissue metabolism and oxygenation. *Proc. Natl. Acad. Sci. USA* **91**(8): 3388-92.

Kuppusamy, P. and Krishna, M.C. (2002) EPR Imaging of tissue redox status. *Current topics in Biophysics* **26**: 29-34.

Kuppusamy, P., Li, H., Ilangovan, G., Cardounel, A.J., Zweier, J.L., Yamada, K., Krishna, M. C. and Mitchell, J.B. (2002) Noninvasive imaging of tumor redox status and its modification by tissue glutathione levels. *Cancer Res*. **62**(1): 307-12.

Kuppusamy, P. and Zweier, J.L. (2004) Cardiac applications of EPR imaging. *NMR Biomed*. **17**(5): 226-39.

Kveder, M., Krisko, A., Pifat, G. and Steinhoff, H.J. (2003) The study of structural accessibility of free thiol groups in human low-density lipoproteins. *Biochim. Biophys. Acta* **1631**(3): 239-45.

Lai, C.S., Hopwood, L.E., Hyde, J.S. and Lukiewicz, S. (1982) ESR studies of O2 uptake by Chinese hamster ovary cells during the cell cycle. *Proc. Natl. Acad. Sci. USA* **79**(4): 1166-70.

Lebedev, O.A. and Kayanovskii, S.N. (1959) Catalytic oxidation of aliphatic amines with hydrogen peroxide. *Trudy po Khimii i Khim. Technologii (Gorkii)* **8**: 649-652.

Lebedev, O.A., Khidekel, M.L. and Razuvaev, G.A. (1961) Isotopic analysis of nitrogen by electron paramagnetic resonance method. *Dokladi Akademii Nauk S.S.S.R.* **140**: 1327-1331.

Liu, K.J., Gast, P., Moussavi, M., Norby, S.W., Vahidi, N., Walczak, T., Wu, M. and Swartz, H.M. (1993) Lithium phthalocyanine: a probe for electron paramagnetic resonance oximetry in viable biological systems. *Proc. Natl. Acad. Sci. USA* **90**(12): 5438-42.

Liu, K.J., Grinstaff, M.W., Jiang, J., Suslick, K.S., Swartz, H.M. and Wang, W. (1994) *In vivo* measurement of oxygen concentration using sonochemically synthesized microspheres. *Biophys. J*. **67**(2): 896-901.

Lurie, D.J. (2001) Free radical imaging. *Br. J. Radiol*. **74**(885): 782-4.

Lurie, D.J., Li, H., Petryakov, S. and Zweier, J.L. (2002) Development of a PEDRI free-radical imager using a 0.38 T clinical MRI system. *Magn. Reson. Med*. **47**(1): 181-6.

Lurie, D.J., Nicholson, I. and Mallard, J.R. (1991) Low-field EPR measurements by field-cycled dynamic nuclear polarisation. *J. Magn. Res.* **95**: 405-409.

Mader, K., Gallez, B., Liu, K.J. and Swartz, H.M. (1996) Non-invasive in vivo characterization of release processes in biodegradable polymers by low-frequency electron paramagnetic resonance spectroscopy. *Biomaterials* **17**(4): 457-61.

Mader, K., Nitschke, S., Stosser, R. and Borchert, H.H. (1997) Non-destructive and localized assessment of acidic microenvironments inside biodegradable polyanhydrides by spectral spatial electron paramagnetic resonance imaging. *Polymer* **38**(19): 4785-4794.

Marsh, D. and Henderson, P.J. (2001) Specific spin labelling of the sugar-H(+) symporter, GalP, in cell membranes of Escherichia coli: site mobility and overall rotational diffusion of the protein. *Biochim. Biophys. Acta* **1510**(1-2): 464-73.

Matsumoto, K., Krishna, M.C. and Mitchell, J.B. (2004) Novel pharmacokinetic measurement using electron paramagnetic resonance spectroscopy and simulation of *in vivo* decay of various nitroxyl spin probes in mouse blood. *J Pharmacol. Exp. Ther*. **310**(3): 1076-83.

Molin, Y.N., Salikhov, K.M. and Zamaraev, K.I. (1980) In: S*pin Exchange. Principles and Applications in Chemistry and Biology*. Springer-Verlag, Berlin, New York, 242.

Mordvintcev, P., Mulsch, A., Busse, R. and Vanin, A. (1991) On-line detection of nitric oxide formation in liquid aqueous phase by electron paramagnetic resonance spectroscopy. *Anal. Biochem*. **199**(1): 142-6.

Murugesan, R., Cook, J.A., Devasahayam, N., Afeworki, M., Subramanian, S., Tschudin, R., Larsen, J.A., Mitchell, J.B., Russo, A. and Krishna, M.C. (1997) *In vivo* imaging of a stable paramagnetic probe by pulsed-radiofrequency electron paramagnetic resonance spectroscopy. *Magn. Reson. Med*. **38**(3): 409-14.

Neiman, M.B., Rozatzev, E.G. and Mamedova, Y.G. (1962) Free radical reactions involving no unpaired electrons. *Nature* **196**(3): 472-474.

Nicholson, I., Robb, F.J., McCallum, S.J., Koptioug, A. and Lurie, D.J. (1998) Recent developments in combining LODESR imaging with proton NMR imaging. *Phys. Med. Biol*. **43**(7): 1851-5.

Nohl, H., Stolze, K. and Weiner, L.M. (1995) Noninvasive measurement of thiol levels in cells and isolated organs. *Methods Enzymol*. **251**: 191-203.

Pake, G.E. and Tuttle, T.R. (1959) Anomalous loss of resolution of paramagnetic resonance hypetfine structure in liquids. *Phys. Rev. Lett*. **3**: 423-425.

Popova, V.I., Leonova, I.N., Weiner, L.M. and Salganik, R.I. (1982) Interaction of the substrate analogue of cytochrome P-450 and mixed function oxidases. *Biochem. Pharmacol*. **31**(11): 1993-8.

Potapenko, D.I., Foster, M.A., Lurie, D.J., Kirilyuk, I.A., Hutchison, J.M., Grigor'ev, I.A., Bagryanskaya, E. G. and Khramtsov, V.V. (2006). Real-time monitoring of drug-induced changes in the stomach acidity of living rats using improved pH-sensitive nitroxides and low-field EPR techniques. *J. Magn. Reson*. **182**(1): 1-11.

Povich, M.J. (1975) Electron Spin Resonance Oxygen Broadening. *J. Phys. Chem*. **79**: 1106-1109.

Presley, T., Kuppusamy, P., Zweier, J.L. and Ilangovan, G. (2006) Electron paramagnetic resonance oximetry as a quantitative method to measure cellular respiration: a consideration of oxygen diffusion interference. *Biophys. J*. **91**(12): 4623-31.

Rabenstein, M.D. and Shin, Y.K. (1995) Determination of the distance between two spin labels attached to a macromolecule. *Proc. Natl. Acad. Sci. USA* **92**(18): 8239-43.

Roopnarine, O., Hideg, K. and Thomas, D.D. (1993) Saturation transfer electron parametric resonance of an indane-dione spin-label. Calibration with hemoglobin and application to myosin rotational dynamics. *Biophys. J*. **64**(6): 1896-907.

Rosen, G.M., Porasuphatana, S., Tsai, P., Ambulos, N.P., Galtsev, V.E., Ichikawa, K. and Halpern, H.J. (2003) Dendrimeric-Containing Nitronyl Nitroxides as Spin Traps for Nitric Oxide: Synthesis, Kinetic, and Stability Studies. *Macromolecules* **36**: 1021-1027.

Roshchupkina, G.I., Bobko, A.A., Reznikov, V.A. and Khramtsov, V.V. (2006) Thiol-specific nitroxides for site-directed spin labeling and EPR measurement of thiol content. XXIInd International Conference on Magnetic Resonance in Biological Systems, Goettingen, Germany.

Rossi, R., Milzani, A., Dalle-Donne, I., Giustarini, D., Lusini, L., Colombo, R. and Di Simplicio, P. (2002) Blood glutathione disulfide: *in vivo* factor or *in vitro* artifact? *Clin. Chem*. **48**(5): 742-53.

Samiec, P.S., Drews-Botsch, C., Flagg, E.W., Kurtz, J.C., Sternberg, P., Jr., Reed, R.L. and Jones, D.P. (1998) Glutathione in human plasma: decline in association with aging, age-related macular degeneration, and diabetes. *Free Radic. Biol. Med*. **24**(5): 699-704.

Sarna, T., Duleba, A., Korytowski, W. and Swartz, H. (1980) Interaction of melanin with oxygen. *Arch. Biochem. Biophys*. **200**(1): 140-8.

Schafer, F.Q. and Buettner, G.R. (2001) Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radic. Biol. Med*. **30**(11): 1191-212.

Soszynski, M. and Bartosz, G. (1997) Decrease in accessible thiols as an index of oxidative damage to membrane proteins. *Free Radic. Biol. Med*. **23**(3): 463-9.

Stone, T.J., Buckman, T., Nordio, P.L. and McConnell, H.M. (1965) Spin-labeled biomolecules. *Proc. Natl. Acad. Sci. USA* **54**(4): 1010-7.

Swartz, H.M. (2004) Using EPR to measure a critical but often unmeasured component of oxidative damage: oxygen. *Antioxid. Redox Signal*. **6**(3): 677-86.

Swartz, H.M. and Timmins, G.S. (2000) In: Rhodes, C.J., ed., The metabolism of nitroxides in cells and tissues. *Toxicology of the human environment: the critical role of free radicals*. Taylor & Francis Inc., London, New York, pp. 91-111.

Takeshita, K., Hamada, A. and Utsumi, H. (1999) Mechanisms related to reduction of radical in mouse lung using an L-band ESR spectrometer. *Free Radic. Biol. Med*. **26**(7-8): 951-60.

Velan, S.S., Spencer, R.G., Zweier, J.L. and Kuppusamy, P. (2000) Electron paramagnetic resonance oxygen mapping (EPROM): direct visualization of oxygen concentration in tissue. *Magn. Reson. Med*. **43**(6): 804-9.

Voinov, M.A., Polienko, J.F., Schanding, T., Bobko, A.A., Khramtsov, V.V., Gatilov, Y.V., Rybalova, T.V., Smirnov, A.I. and Grigor'ev, I.A. (2005) Synthesis, structure and X-band (9.5 GHz) characterization of the new series of pH-sensitive probes: N-N-disubstituted 4-amino-2,2,5,5-tetramethyl-3-imidazoline 1 Oxyls. *J. Org. Chem.* **70**: 9702-9711.

Volodarsky, L.B., Reznikov, V.A. and Ovcharenko, V.I. (1994) *Synthetic chemistry of stable nitroxides*. CRC Press, Boca Raton, Fl.

Weiner, L.M., Hu, H. and Swartz, H.M. (1991) EPR method for the measurement of cellular sulfhydryl groups. *FEBS Lett*. **290**(1-2): 243-6.

Woldman, Y., Khramtsov, V.V., Grigor'ev, I.A., Kiriljuk, I.A. and Utepbergenov, D.I. (1994) Spin trapping of nitric oxide by nitronylnitroxides: measurement of the activity of no synthase from rat cerebellum. *Biochem. Biophys. Res. Commun*. **202**(1): 195-203.

Yamada, K., Inoue, D., Matsumoto, S. and Utsumi, H. (2004) *In vivo* measurement of redox status in streptozotocin-induced diabetic rat using targeted nitroxyl probes. *Antioxid. Redox Signal*. **6**(3): 605-11.

Yelinova, V., Glazachev, Y., Khramtsov, V., Kudryashova, L., Rykova, V. and Salganik, R. (1996) Studies of human and rat blood under oxidative stress: changes in plasma thiol level, antioxidant enzyme activity, protein carbonyl content, and fluidity of erythrocyte membrane. *Biochem. Biophys. Res. Commun*. **221**(2): 300-3.

Yelinova, V.I., Khramtsov, V.V. and Markel, A.L. (1999) Manifestation of oxidative stress in the pathogenesis of arterial hypertension in ISIAH rats. *Biochem. Biophys. Res. Commun*. **263**(2): 450-3.

Zweier, J.L. and Kuppusamy, P. (1988) Electron paramagnetic resonance measurements of free radicals in the intact beating heart: a technique for detection and characterization of free radicals in whole biological tissues. *Proc. Natl. Acad. Sci. USA* **85**(15): 5703-7.

Novel Leads for Selective Antibiotics Against Shigellosis by Virtual Screening, Crystallography and Synthesis

BERNHARD STENGL AND GERHARD KLEBE
Institute of Pharmaceutical Chemistry, Philipps-University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, Phone: +49 6421 282 1313, Fax: +49 6421-28 28994, E-mail: Klebe@mailer.uni-marburg.de

Abstract

Eubacterial tRNA-guanine transglycosylase (TGT) is involved in the hyper-modification of cognate tRNAs leading to an exchange of guanine 34 at the wobble position in the anticodon loop by preQ$_1$, as part of the biosynthesis of queuine. Mutation of the *tgt*-gene in *Shigella flexneri* results in a significant loss of pathogenicity of the bacterium, revealing TGT as a prospective target for the design of potent drugs against Shigellosis. The X-ray structure of *Zymomonas mobilis* TGT in complex with preQ$_1$ was used to search for putative inhibitors, initially with the computer program LUDI. Furthermore, the recognition properties of the protein binding site have been used to derive a protein-based pharmacophore which served as a prerequisite for virtual screening based on molecular similarity and docking. This strategy retrieved several novel scaffolds potentially matching with the substrate recognition site. Iterative design has been applied to reveal significantly larger inhibitors with improved binding properties addressing further polar

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 209–249.

residues in the binding pocket and filling specifically a small hydrophobic cavity. The protein performs several conformational adaptations upon ligand binding which are in agreement with the required substrate promiscuity of the enzyme. Water molecules accommodated in the binding pocket have been detected as important either for mediating interactions between protein and ligand or to bridge interactions between polar groups of the protein. In the latter case, replacement of these waters is detrimental to ligand binding. Addressing the U33 binding pocket reveals a substantial increase in ligand binding affinity, also due to the formation of charge-assisted hydrogen bonds.

Keywords: Shigella Dysentery, Leads Compound Discovery, Drug Discovery and Design, Protein-Ligand Interactions, Docking, Scoring.

1. INTRODUCTION: THE DISEASE AND POSSIBLE TREATMENTS

Shigellosis or bacillary dysentery is caused by bacteria belonging to the genus *Shigella*. Oral uptake of contaminated drinking water or food possibly initiates an infection. The bacteria pass through the stomach, the small intestine and finally reach the colon where they invade the colon epithelium and the mucosa. This results in a destructive recto-colitis which is responsible for the dysenteric symptoms like watery diarrhoea, fever, intestinal cramps and emission of mucopurulent and blood stools (Sansonetti 2001a). Due to infiltration of inflammatory cells in the colon epithelium mucosal ulcers are often observed (Mathan and Mathan 1991). In the absence of an effective treatment, secondary complications may occur like septicaemia, pneumoniae or haemolytic uremic syndrome (Bennish 1991).

Shigellosis is a global burden with an estimated annual number of 165.7 million episodes. Only 1.5 million cases occur in industrialized countries, many of them endemically in children day care centres, custodial institutions or as traveller's diarrhoea. Almost no fatalities are observed. The vast majority of 163.2 million episodes take place in developing countries, often epidemically, with a death rate of 1.1 million. Most of the episodes occur in displaced populations. Civil war in areas with dense population and natural disasters (flooding, drought) often cause hunger and poverty forcing people to leave their home. Mostly, these displaced people are crowded in areas with poor sanitation and insufficient supplies of clean water. Concomitant malnutrition is widely spread. Children are affected most severely from such conditions. Thus, 69% (112.6 million) of all episodes and 61% (660.000) of all fatalities involve children under 5 years of age (Kotloff et al. 1999). Particularly in Africa the interference with HIV/AIDS is a serious problem.

For effective treatment clean water, sufficient food supply and hygiene conditions avoiding the fecal – oral spreading route are inevitable. This includes washing of hands as well as the use of separate dishes and towels for patients and other members of the household. Additionally, the control of houseflies by bait and trap strategies significantly reduces the spreading of *Shigella*. Houseflies are a key vector for the faecal contamination of human food (Cohen et al. 1991).

Antibiotics can be used to treat shigellosis and reduce the period of bacterial excretion from the patient. In developing countries they are used to stop epidemic spreading (Jennison and Verma 2004). A major problem for shigellosis treatment is the extraordinary ability of *Shigella* to acquire plasmid-encoded resistance to antimicrobial drugs that constituted the first-line therapy. Sulfonamides, tetracycline, ampicilline and trimethoprim-sulfamethoxy-azole were highly efficient drugs two decades ago, but are becoming more and more ineffective (Kotloff et al. 1999). For trimethoprim the resistance rate rose from 3% in 1975 to 98% in 1988 (Heikkila et. al. 1990). A survey from Israel reported for the period 1991 - 2000 high resistances to trimethoprim (94%) and ampicillin (85%), significantly increased resistances to tetracycline (23% to 87%) and emerging resistances to quinolones (0.5-2%) (Ashkenazi et al. 2003). During an epidemic outbreak in Zaire in 1994 a *Shigella* strain was identified that was resistant to all commonly used antibiotics (Goma Epidemiology Group 1995). These findings underline the need to monitor resistance and to develop new, innovative antibiotics that maintain the ability of successful treatment.

A further important goal is the development of vaccines that prevent infection. Since the 1940s this has been attempted with little success. Current research approaches using deeper insight into *Shigella* pathogenicity, however, made promising progress. But still these vaccines do not yet fulfil efficacy and safety requirements for the treatment of humans (Jennison and Verma 2004).

2. CELLULAR AND MOLECULAR PATHOGENICITY

Shigellae are very infective. Only 10 – 100 bacteria can cause a disease in an adult. After passage through the stomach and the small intestine they are capable to invade the colon epithelium. The passage is not possible through the apical side of the epithelial cells (Fig. 2.1). Invasion occurs indirectly via M cells. They are specialized to transport antigens (bacteria, etc.) through the colon epithelium to present them to macrophages which are in

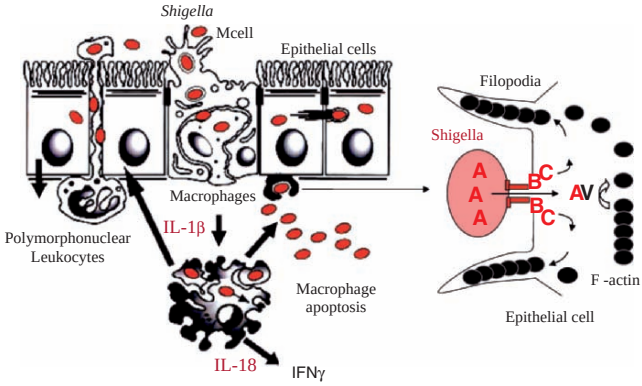


Figure 2.1. a) Pathway of epithelial colonization and disintegration – figure modified from Sansonetti (2001a); b) macropinocytotic uptake of *Shigella*, A: ipaA, B: ipaB, C: ipaC, V: vinculin, S: Src kinase, red: *Shigella* virulence factors, black: host cell proteins.

tight association with them. Both are located in follicle-associated epithelia (FAE). The FAE overly the mucosa-associated lymph nodes that are responsible for intestinal immunity. *Shigella* have the unusual capacity to enter the M cell associated macrophages without being damaged. After phagocytotic uptake by the macrophage, they escape from the phagosome.

Inside the macrophages they induce apoptosis. Upon apoptotic death inflammatory interleukins (IL-1β and IL-18) are produced and released into the mucosa after lysis, together with the intact *Shigella*. The bacteria then enter the colon epithelial cells from the basal side. Once inside these cells they propagate and spread from cell to cell – well protected from immune cells. Infected epithelial cells release IL-8. This interleukin, together with the two above mentioned ones, attracts macrophages and polymorphonuclear leucocytes (PMN) from subepithelial tissues. The latter disrupt the integrity of the epithelial barrier and facilitate further *Shigella* invasion from the colon. Thus, in a sort of 'snowball effect' the initial inflammatory reaction is amplified and results in an uncontrolled tissue destabilization. Nevertheless, in contrast to macrophages, PMN are able to kill *Shigella* inside their phagosomes. Together with NK cells and T lymphocytes, immune cells attracted by IL-18, they eventually cope with the infection. In addition IL-18 triggers the production of the interferon IFN-γ which activates the immune system. (Sansonetti 2001a; Sansonetti 2001b; Jennison and Verma 2004).

In order to enable cell entry, intracellular mobility, cell spreading and induction of apoptosis *Shigella* produces virulence factors. These virulence factors are used to reprogram the cellular machinery of epithelial as well as

immune cells by activating innate transport and signalling pathways (Fernandez and Sansonetti 2003; Van Nhieu et al. 2000).

The first step of invasion into an epithelial cell is the interaction of a tube like type III secretion apparatus with the host cell membrane (Fig. 2.1). A pore is formed on the tip of the secretion apparatus by the virulence factors IpaB and IpaC. Exposure of IpaC into the host cell cytoplasm activates a Src tyrosine kinase. A further signalling cascade results in actin dependent filopodia and lamellipodia formation. These structures form in the surrounding of the secretion apparatus tip and finally enclose *Shigella*, resulting in macropinocytotic uptake. To provide lamellipodia with enough actin and to avoid polymerization in the secretion apparatus contact region, IpaA is secreted through the secretion tube porus into the host cell cytoplasm. Within the cell IpaA binds to vinculin and available F-actin from adjacent cytoskeleton elements of the host cell is depolymerised.

After the uptake of *Shigella* into the host cell the macropinocytotic vacuole lyses and the bacterium is released into the cytoplasm. To enable intracellular mobility the virulence factors IcsA is produced. It is located in the bacterial cell wall, exposed to the host cell cytoplasm. Binding of the host cell proteins N-WASP and ARP2/3 results in actin polymerization that pushes the bacteria through the cytoplasm.

The described virulence factors are only the most prominent ones among a large set of further virulence factors involved in generating pathogenicity.

3. REGULATION OF PATHOGENICITY

The virulence factor genes are located on a 214-kb virulence plasmid which was isolated and sequenced from *S. flexneri*. Coding sequences are scattered all over the plasmid. One block of 30 kb shows a particular dense pattern of genes and is called pathogenicity island (PAI). The *mxi*/*spa*- and *ipa*-loci found in this region code for proteins necessary to establish the type III translocon and to allow cell entry (Sansonetti 2001a). The expression of virulence genes is organized hierarchically (Dorman and Porter 1998). VirF and VirB are the key transcription activators for virulence gene expression. The *virF* and *virB* genes are plasmid encoded. Expression of virF directly activates the transcription of virulence factor genes like *icsA* as well as the transcription of the *virB*-gene. The *virB*-gene product then activates the transcription of the *mxi*/*spa*- and *ipa*-genes. Thus, VirF is in the centre of pathogenicity regulation (Fig. 3.2).

Remarkably, it is not the absence or presence of VirF that regulates virulence. Instead, a threshold level for VirF exists above which virulence factors are produced. The amount of VirF depends on various environmental

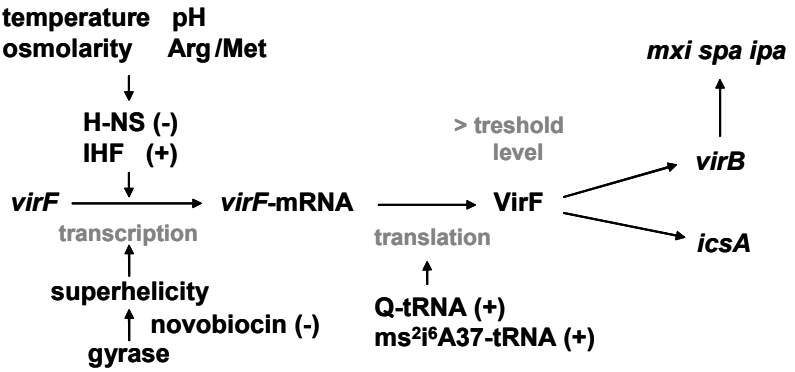


Figure 3.2. Regulation of VirF expression.

and internal factors. Such environmental factors are pH, osmolarity, temperature and nutrition factors. Below 30°C *Shigella* is not virulent. The *virF*-gene transcription is fully activated at 37°C, pH 7.4, physiological osmolarity and in the presence of free amino acids (arginine/methionine). The *virF*-gene promoter is positively regulated by the transcription factor IHF and negatively by H-NS. Both transcription factors are encoded by the chromosome and their expression is sensitive to the mentioned environmental factors (Dorman and Porter 1998; Durand et al. 2000; Durand and Björk 2003).

Apart from this 'classical' regulation the virF level also depends on an accurate transcriptional and translational machinery (Durand et al. 2000). Each intervention affecting this machinery may also influence VirF expression. The antibiotic novobiocin for instance inhibits gyrase, thus decreasing negative superhelicity in DNA. This results in decreased virF levels. On the translational level the expression of virF can be influenced as well. Efficient translation of *virF*-mRNA at the ribosome requires the presence of modified tRNA molecules. Modification of tRNA bases is very common in nature (Björk 1996). Two modifications were demonstrated to exhibit major influence on the *virF* mRNA translational speed. In specific tRNA molecules the highly modified nucleosides queuosine in position 34 (the anticodon wobble position) or 2-methylthio-N^6-isopentenyladenosine (ms^2i^6A37) in position 37 (adjacent to the anticodon) have to be present. *Shigella* mutants lacking one of these modifications show significantly reduced virulence. This was tested in mutational studies where gene knock-out in the tRNA modification pathway resulted in the absence of these modifications. The first step of A37 modification is catalyzed by the *miaA* gene product. Mutation of the *miaA* gene reduces the VirF level to 10%, and the

haemolytic activity to 10-20% compared to the wild type (Durand et al. 1997). In tRNA, position 34 the *tgt*/(*vacC*)-gene product is involved in the formation of the modified base queuine. Mutation of the *tgt*-gene reduces both, virF level and haemolytic activity, to 50-60% of the wild type (Durand et al. 1994; Durand et al. 2000). Thus, tRNA modifying enzymes could represent promising targets for the development of antibiotics. Inhibition of such specific tRNA modification steps should result in significantly reduced virulence of *Shigella*.

4. EUBACTERIAL TGT AND ITS FUNCTION

For functional analyses of the *tgt*/(*vacC*)-gene product *Escherichia coli* can be used, as *Shigella* can be considered as an 'enteroinvasive *E. coli*' (EIEC) subtype (Escobar-Pámaro et al. 2004). In *E. coli* the *tgt*-gene codes for the tRNA – guanine transglycosylase (TGT). This enzyme catalyzes a crucial step in a tRNA modification pathway targeting the wobble position 34 of four specific tRNAs (Okada and Nishimura 1979). The modification finally results in the formation of the functional base queuine, required for *virF*-translation (Fig. 4.3a). A crystal structure of this protein is available for the structurally very similar *Zymomonas mobilis* TGT (Romier et al. 1996). Thus, the TGT from the latter species can be used for structure-based drug design to develop potent inhibitors and finally to test the hypothesis whether a significant virulence reduction can be achieved.

TGT catalyzes the base exchange of guanine in position 34 by the modified base $preQ_1$, a queuine precursor. The base exchange follows a ping-pong reaction mechanism resulting in the irreversible incorporation of $preQ_1$ (Goodenough-Lashua and Garcia 2003). In a first step, tRNA binds to TGT and G_{34} is cleaved off the tRNA. In the intermediate reaction state tRNA ribose 34 is covalently bound to TGT. In a second, reverse reaction step, $preQ_1$ replaces G_{34} in the active site and is incorporated into the tRNA (Figs. 4.3b, 4.4a-d).

The comparison of available crystal structures of *Z. mobilis* TGT in complex with tRNA substrate, $preQ_1$ and a guanine-type inhibitor allowed deducing single steps of the catalytic reaction. They enabled to develop a detailed, new functional model for the course of the base exchange reaction. According to this model previous assumptions, discussing Asp102 as the catalytic nucleophile were corrected and functional roles were assigned for previously unconsidered active site residues (Stengl et al. 2005).

Following the reaction pathway, TGT recognizes the tRNA substrate via the 'zinc binding' site and G_{34} is specifically recognized in the active site by Asp102, Asp156, Gln203 and Gly230 (Fig. 4.4a-d). A polar contact towards

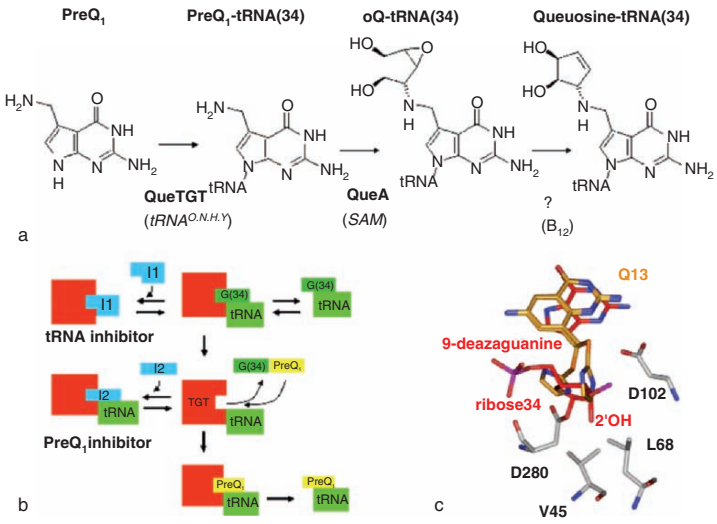


Figure 4.3. a) Queuine modification pathway in Eubacteria; b) Inhibition modes along the TGT ping-pong reaction pathway; c) structural superposition of covalent TGT-tRNA intermediate stabilized by 9-deazaguanine and **Q13** binding to uncomplexed TGT; the **Q13** substituent would interfere with the ribose binding site (chemical formula in Tab. 5.5).

the peptide NH group of Ala232 is mediated via a water molecule (W1). G_{34} is buried in the active site *via* hydrophobic stacking interactions with Met260 and Tyr106.

Asp280, instead of the previously assumed Asp102, located adjacent to the ribose ring 34, acts as catalytic nucleophile. This residue is well kept in position by Tyr258 and Gly261. Tyr258 itself is arrested and kept in position through several hydrophobic interactions (Met43, Leu100, Met153, Phe199, Met260, Met278). Asp280, Gly261 and Tyr258 as well as its neighbouring hydrophobic residues are conserved in eubacterial and eukaryotic TGTs emphasizing their particular role to guarantee accurate adjustment of the nucleophile (Xie et al. 2003).

The Asp280 carboxylic oxygen attacks the C1 carbon of ribose 34 in an S_N2 reaction and pulls the ribose ring towards Asp280 (Xie et al. 2003). The ribose performs a 40° rotational movement anchored by adjacent ribose phosphate groups P_{34} and P_{35}. The rotation causing the rupture of the covalent bond between C1 and G_{34} is controlled by Asp102 with its carboxy group H-bonded to the 2'OH-group of the rotating ribose. The 2'OH group is finally released towards a hydrophobic cleft formed by Val45 and Leu68 (Stengl et al. 2005). In this orientation the polar group cannot form any H-bond to the enzyme and experiences only weak and rather unfavourable

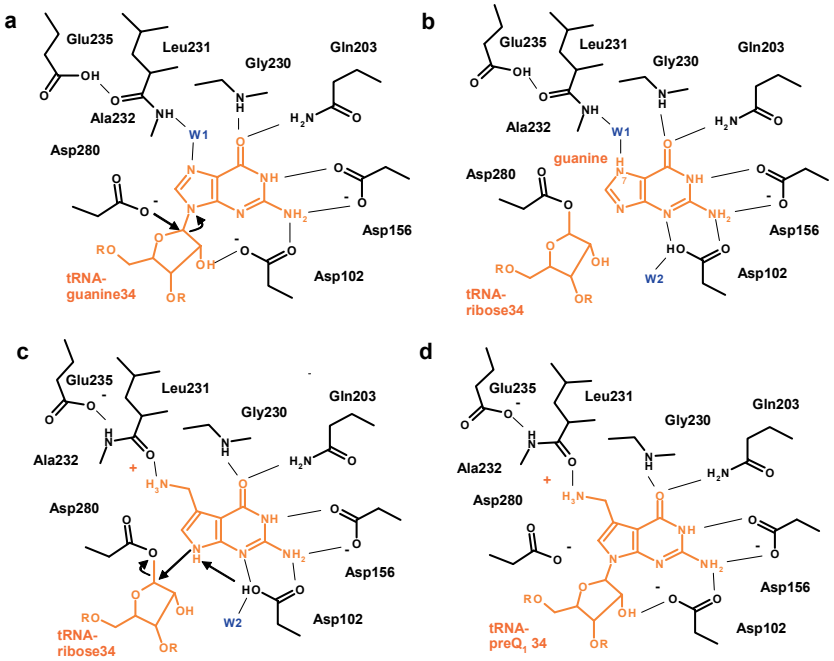


Figure 4.4. Base exchange mechanism in eubacterial QueTGT.

interactions (Fig. 4.3c). Supposedly, the unfavourable intermediate occupancy of the hydrophobic pocket through the 2'OH group serves as a kind of tense spring state and stores energy for the conformational movements required for the reverse reaction step during the ping-pong reaction pathway. Thus, this geometry guarantees a sterically favoured but electrostatically unfavoured intermediate state orientation.

Guanine is reprotonated in the binding pocket after cleavage from tRNA, either in position 7 or in position 9. Reprotonation in position 7 could be supported by water molecule W1, which is then forming a hydroxy-anion (Fig. 4.4b). The fact that the methyl group of Ala232 in the amide-exposing conformation reduces the available space of the binding pocket and perfectly shields this water molecule, speaks in favour of this position (Stengl et al. 2005). Alternatively, reprotonation in position 9 is possible as well, then supported by Asp102 serving as general acid. Another water molecule, W2, bridging Asp102 with Gln107 at the upper rim of the binding pocket would be suited to shuffle a proton into the active site (Xie et al. 2003).

After reprotonation guanine and W1/OH^- leave the binding pocket and are replaced by $preQ_1$. To be able to accommodate $preQ_1$, the Leu231-Ala232 peptide bond supposedly undergoes a flip which is controlled by

Glu235 acting as acid/base system (Fig. 4.4b-c) (Stengl et al. 2005). $PreQ_1$ is bound in a similar fashion as guanine except of the amino methyl group that directly interacts with the flipped Leu231 carbonyl oxygen. For activation $preQ_1$ is deprotonated in position 9 by Asp102 then acting as a general base. The proton is shuffled from the active site supposedly mediated by water molecule W2. In a reverse S_N2 reaction step the now activated $preQ_1$ nucleophilically attacks the C1 carbon of ribose 34. Upon product formation the covalent bond towards Asp280 is cleaved and the ribose 34 2'OH group is pushed out of its unfavourable environment, stabilized via H-bond formation with Asp102 (Fig. 4.4c-d). In this process the Asp102 carboxy group rotates and expels W2 from the active site followed by the final release of the $preQ_1$-modified tRNA from TGT. Accordingly, Asp102 in addition to its function as general acid/base imposes a strong directional driving force thus controlling the structural changes upon product formation. Finally, the modified tRNA is released from the binding pocket and TGT is ready for a new base exchange cycle.

5. DESIGN, SYNTHESIS AND CHARACTERIZATION OF DIFFERENT TGT INHIBITORS

5.1. Functional and structural analysis of TGT inhibitors

5.1.1. Determination of inhibition constants

The recent discovery that TGT follows a ping-pong reaction mechanism required modifications of the formerly applied inhibition assay. Therefore, the inhibition constants of some previously developed compound series had to be redetermined.

In a ping-pong reaction, which follows a two step reaction, each single step can be inhibited separately (Fig. 4.3b). To inhibit the first reaction step a putative inhibitor has to compete with the binding of the tRNA molecule to TGT. Therefore, the size and shape of such an inhibitor is only determined by the properties of the TGT binding pocket. In terms of size, an appropriate inhibitor must extent sufficiently into the recognition area used to bind tRNA in order to prevent binding of the latter to TGT. Nevertheless, also the second step of the reaction can be inhibited. After the covalent attachment of tRNA to TGT which significantly reduces the available space of the binding pocket, only small compounds of similar size to the natural substrate $preQ_1$ can be accommodated in the remaining pocket. These observations result in a size-dependent inhibition model (Fig. 4.3c). To record both inhibitory contributions, identified as competitive (first step) and uncompetitive (second

step) with respect to tRNA binding, resulting in non-competitive inhibition, the previously applied assay procedure had to be redesigned. Trapping and SDS-PAGE experiments allow to identify uncompetitive inhibitors. These compounds stabilize the covalent TGT-tRNA complex which can be observed as a retarded band in SDS-PAGE due to its larger molecular mass. Inhibition constants can then be determined by an inhibition assay protocol permitting separation of competitive and uncompetitive contributions. In the formerly applied assay, both inhibitory contributions were recorded as one superimposed signal. Therefore, partly altered inhibition constants were determined for some of the previously investigated compound series. The new assay setup was published in Meyer et al. 2006.

5.1.2. TGT binding pocket conformations

Multiple crystal structure analyses of the enzyme with bound substrates and substrate analogues have shown that the TGT binding pocket performs some pronounced adaptations upon substrate binding. Additionally, deviating binding pocket geometries are also relevant for the binding of different inhibitor classes. The combinations of two local adaptive sites result in four binding pocket geometries relevant for inhibitor binding which are systematically presented in Figure 5.5a-d All four conformations have been observed in crystal structures of pyridazindione- and quinazolinone-based inhibitor classes in complex with TGT (Grädler et al. 2001; Brenk et al. 2003a; Brenk et al. 2004; Meyer et al. 2004). The first adaptation results from the Leu231/Ala232 peptide switch, which has already been described during the base exchange reaction (Fig. 4.4c). Either a carbonyl donor or an amide acceptor functionality is exposed towards the binding pocket and modifies the size of the pocket. The second adaptation refers to Asp102. In uncomplexed structures, this residue points out of the binding pocket and is involved in an H-bonding network with Asn70 and Thr47. Upon substrate binding of, e.g., $preQ_1$ the side chain of Asp102 rotates into the binding pocket and forms two hydrogen bond with the substrate. As a consequence, the size of the guanine recognition pocket is significantly decreased, distinct active site water molecules are expelled and the properties exposed to the binding site are fundamentally altered. Furthermore, the rotational movement of Asp102 provokes a reorientation/disordering of the side chains of Asn70 and Thr47 (Brenk et al. 2003; Brenk et al. 2003b). In the following chapters the discovery of various inhibitor classes including pyridazinediones and quinanzolinones, as well as their interference with TGT binding pocket geometries will be presented.

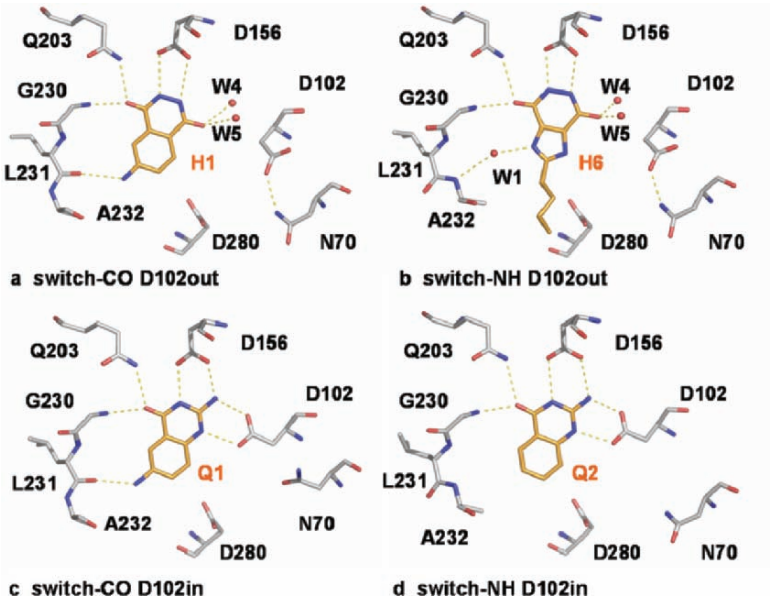


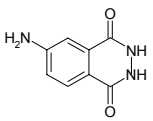
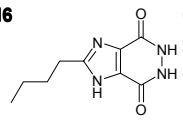
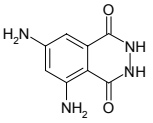
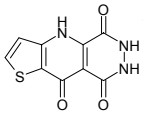
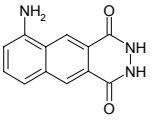
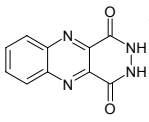
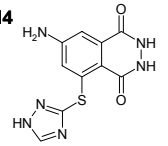
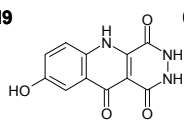
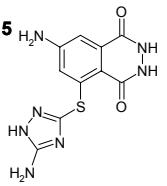
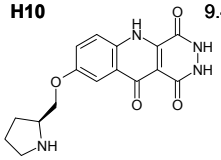
Figure 5.5. a-d Binding pockets of TGT relevant for inhibitor binding differ in the Leu231/Ala232 peptide switch and in the orientation of Asp102 (in/out).

5.2. Pyridazindione-based inhibitor series

In the very first strucure-based design attempt performed on *Z. mobilis* TGT, the pyridazindione compound **H1** ('H' for hydrazide) has been discovered. The *de novo* design program LUDI was used as starting point for the discovery of TGT inhibitors (Grädler et al. 2001). **H1** exhibits an inhibition constant of 8.3 μM, following the former assay, and the crystal structure in complex with TGT was successful determined (Fig. 5.5a). Due to its lead-like scaffold the structure was used for further design. Modifications resulted in compound **H2** and **H3** with similar binding properties as **H1** (Tab. 5.1).

All compounds bind to the Leu231 carbonyl exposing/Asp102out conformation with the waters W4 and W5 present in the binding pocket (Fig. 5.5a). Attempts, to address the catalytic nucleophile Asp280 by substituting the scaffold **H1** had only limited success. Crystal structures of **H4** and **H5** in complex with TGT revealed that the triazole-based substituents were not well oriented in this pocket. Both compounds turned out to be weaker binders of TGT (Grädler et al. 2001).

Table 5.1. Pyridazindione-based inhibitors.

H1		**75 ± 9** 8.3 ± 0.4*	1.95Å 1ENU	**H6**		**62 ± 40** 83 ± 18*	2.1Å 1N2V
H2		**73 ± 13** 0.3 ± 0.1*	1.95Å 1F3E	**H7**		5.0 ± 1.2*	unpub. (ln01)
H3		0.3 ± 0.1*	1.95Å unpub. (ug01)	**H8**		5.6 ± 0.4*	---
H4		54 ± 14*	2.1Å unpub. (hd03)	**H9**		0.7 ± 0.1*	---
H5		38 ± 1*	1.4Å unpub. (hd04)	**H10**		9.4 ± 0.1*	---

K_{ic} in [μM] with average error; (*): former assay; bold: modified assay; crystal structures: PDB-code and maximum resolution

Searches for alternative pyridazindione-type leads in the NNC-database revealed **H6** and **H7** as the most promising hits (Brenk et al. 2003a). With an inhibition constant of 83 μM, following the former assay, **H6** was a ten-fold weaker binder than **H1** (Tab. 5.1). Nevertheless, a well defined crystal structure in complex with TGT could be obtained with **H6** (Brenk et al. 2003a). Surprisingly, it shows an altered binding mode compared to **H1**. The binding pocket adopts the Ala232 amide exposing/Asp102out conformation and an interstitial water molecule (W1) is bridging towards Ala 232 (Fig. 5.5b). Of **H7** no crystal structure in complex with TGT could be obtained, probably due to the low solubility of the compound. Nevertheless, **H7** was used as lead for a structure-based design approach (Brenk et al. 2003c). Modifications of the scaffold resulted in **H8** and **H9** with similar or slightly improved affinity. Substitution of **H8** resulted in **H10**, but no stronger binding was observed. Due to the low solubility in water, of **H8** – **H10** no crystal structures in complex with TGT could be achieved.

Inhibition constants for **H1**, **H2** and **H6** were revalidated following the newly designed assay procedure. Surprisingly, it turned out that uncompetitive inhibition is not relevant for **H1**, **H2**, and **H6**, although similar molecular dimensions as for $preQ_1$ are given. This may result from the fact that in the tRNA-bound complex Asp102 is rotated into the active site (Fig. 5.5c,d). The pyridazindiones, however, require Asp102 to point out of the binding pocket (Fig. 5.5a,b). The redetermination of competitive inhibition constants revealed that the pyridazindiones were overestimated in their binding properties by the previously used assay. The affinity differences, observed for **H1** and **H6**, resulting from the Leu231/Ala232 peptide switch could not be confirmed. Instead, the new data suggests that the orientation of the Leu231/Ala232 peptide switch has only minor influence on binding affinity.

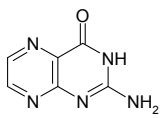
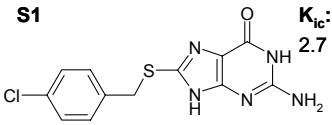
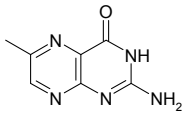
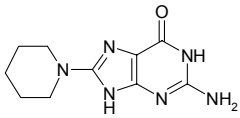
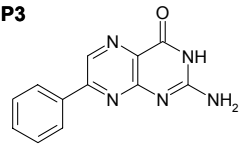
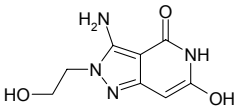
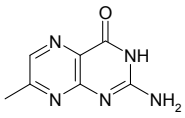
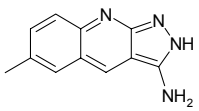
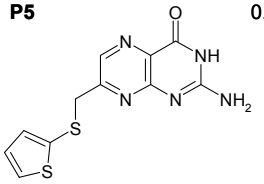
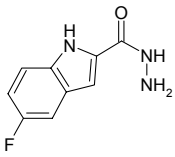
5.3. Pteridines and virtual screening hits

A virtual screening campaign was performed on *Z. mobilis* TGT based on the crystal structure of TGT·**H6** and TGT·**H1** to create a composite pharmacophore hypothesis (Brenk et al. 2003a). It included the binding pocket information for the Leu231/Ala232 peptide switch in both orientations. Nine compounds were finally selected and tested for TGT inhibition (Tab. 5.2). Pteridine-based ligands were retrieved which showed micromolar inhibition (Tab. 5.2; **P1** - **P3**; 'P' for pteridine). Further groups of compounds had a guanine like scaffold (**S1**–**S3**) or only low similarity to the natural substrates (**S4**; **S5**). They all were significantly weaker binders compared to the pteridines. For all compounds no crystal structure in complex with TGT could be obtained, most probably due to their low solubility.

Pteridine **P1** was used as scaffold for a structure-based design approach (Brenk et al. 2003c). First **P4** was synthesized and tested. Although a slight affinity decrease was observed, this compound provided a starting point for further substitutions in position 7 to address the hydrophobic ribose 34 binding pocket formed by Val45 and Leu68. Of all tested compounds **P5**, substituted with a hydrophobic thiophen substituent showed the best affinity. Also for these compounds crystal structure analyses in complex with TGT were not successful.

Due to the structural similarity of **P1** with guanine, this compound was tested for non-competitive inhibition. In a trapping experiment, **P1** was identified to stabilize the covalent TGT-tRNA intermediate. With an uncompetitive inhibition constant of 0.7 μM it is one of the most potent uncompetitive inhibitors known to date. The competitive inhibition constant of 2.2 μM indicates that a guanine-type scaffold structure is more favourable

Table 5.2. Pteridine-based inhibitors and virtual screening hits.

P1		**K_{ic}: 2.2 ± 0.6** **K_{iu}: 0.7 ± 0.2** 0.6 ± 0.2*	**S1**		**K_{ic}: 51 ± 13** 2.7 ± 0.3*
P2		0.25 ± 0.1*	**S2**		37 ± 7*
P3		3.8 ± 0.1*	**S3**		8.1 ± 1.0*
P4		5.6 ± 1.7*	**S4**		156 ± 36*
P5		0.45 ± 0.05*	**S5**		72 ± 5*

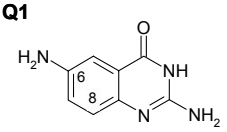
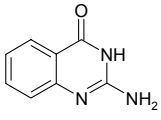
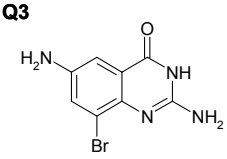
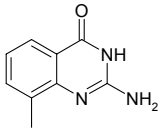
K_{ic} in [μM] with average error; (*): former assay; bold: modified assay;

for binding than the pyridazindione scaffold. Compared to the pyridazindione scaffold **H1** a gain of affinity by a factor of 40 with respect to the competetive inhibition constant is observed for **P1**. For compound **S1** the trapping experiment revealed that uncompetitive inhibition is not relevant. With respect to competitive inhibition (51 μM) this guanine derivative experiences a significant loss in affinity compared to **P1**.

5.4. Quinazolinones

The most potent TGT inhibitors were derived from the class of quinazolinones (3*H*-chinazolin-4-on). This class was designed in a collaborative project with the group of Prof. F. Diederich, ETH Zürich. Quinazolinones are characterized by inhibition constants in the lower micromolar range and they are suited for crystal structure analysis in complex with TGT (Meyer et al.

Table 5.3. Competitive and uncompetitive inhibition constants of small-sized inhibitors.

Q1		**K_{ic}: 1.5 ± 0.4** **K_{iu}: 0.6 ± 0.2** 0.35 ± 0.12*	**Q2**		**K_{ic}: 2.1 ± 0.5** **K_{iu}: 1.7 ± 0.4** 0.02-0.05*
Q3		9.1 ± 1.0*	**Q4**		**K_{ic}: 3.7 ± 0.9** **K_{iu}: 19.1 ± 4.8** 7.0 ± 1.8*

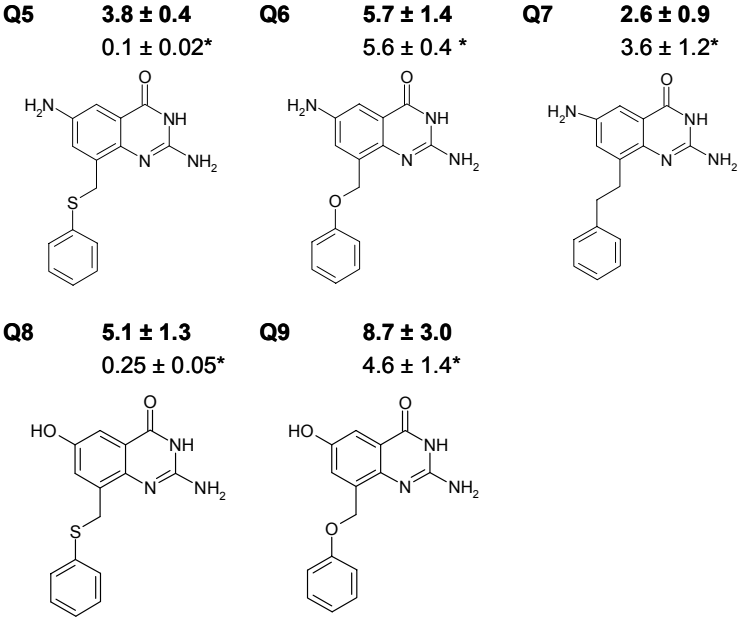
K_{ic} in [μM] with average error; (*): previous assay; bold: modified assa

2002; Brenk et al. 2004, Meyer et al. 2004). The lead structure **Q1** was designed as the combination of structural elements of **H1** and $preQ_1$ (Tab. 5.3). With an inhibition constant of 0.35 μM, following the former assay, **Q1** appeared as a potent inhibitor. Crystal structure analysis of TGT·**Q1** revealed a binding mode similar to **TGT·H1** (Fig. 5.5a,c). However, one major structural difference could be observed. Asp102 is rotated into the binding pocket and engages in a bifurcated hydrogen bond to **Q1**. This binding mode resembles that of $preQ_1$. Upon rotation of Asp102 the water molecules W4 and W5 are expelled from the binding site.

Q2 lacks the 6-amino group of **Q1**. It was synthesized to probe for the relevance of the exocyclic amino group for binding (Meyer et al. 2004). With an inhibition constant of 20-50 nM, following the previous assay, it was a surprisingly tight binder, although in structural terms this could not be explained. The crystal structure of TGT·**Q2** revealed a deviating binding mode compared to **Q1** (Fig. 5.5d). This compound binds to the peptide switch with the amide oriented to the binding pocket. In contrast to the crystal structure of TGT·**H6** the interstitial water molecule W1 is not found in this structure.

Compounds **Q1** and **Q2** were used as starting points to further explore the binding site properties and to develop more potent inhibitors (Meyer et al. 2002; Brenk et al. 2004, Meyer et al. 2004). Decorations at position 8 were intended to address the hydrophobic pocket formed by Val45 and Leu68. **Q3** and **Q4** served as starting points for a series of compounds modified in this position (Tab. 5.3). Repulsive interactions of the 8-methyl

Table 5.4. 6-Aminoquinazolinones – series with different spacers.



K_{ic} in [μM] with average error; (*): previous assay; bold: modified assay

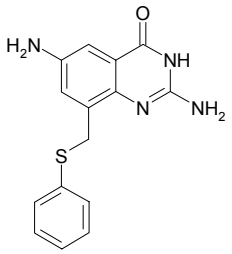
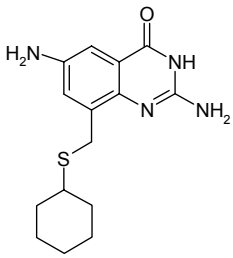
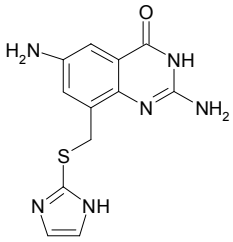
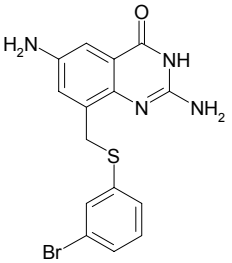
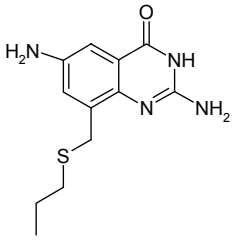
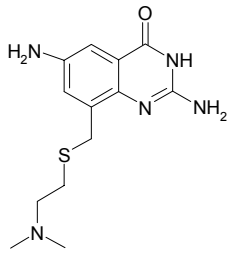
group of **Q4** with Asp102 were identified in the TGT·**Q4** crystal structure. With an inhibition constant of 7μM, following the previous assay, **Q4** exhibited a significant loss of affinity (Meyer et al. 2004).

Further substituted compounds in position 8 can be classified in two series with respect to the spacer and side chain properties. In the first series compounds with different spacers have been synthesized. Ethyl, sulfanylmethyl and phenoxymethyl spacers have been tested to link the scaffold with a phenyl substituent (Tab. 5.4). The second series is based on 6-aminoquinazolinone with a sulfanylmehtyl spacer. In this series side chains with aliphatic, aromatic and polar substituents have been investigated (Tab. 5.5).

Modifications at position 6 of the quinazolinone scaffold were synthesized to study the relevance of the peptide switch for inhibitor binding (Tab. 5.6). For synthesis quinazolinone with a phenylsulphanylmehtyl substituent was used as core fragment. TGT was identified to accept a broad variety of functionalities in this position (-CN, $-NH_2$, -OH, -H, -Br, CH_2NH_2).

Crystal structure analyses of substituted inhibitors in complex with TGT, however, was successful only in some cases. Most of the substituted compounds exhibit low solubility. Only in case of **Q12**, **Q13**, **Q14** and **Q15**

Table 5.5. 7-Quinazolinones with aromatic, aliphatic and polar side chains.

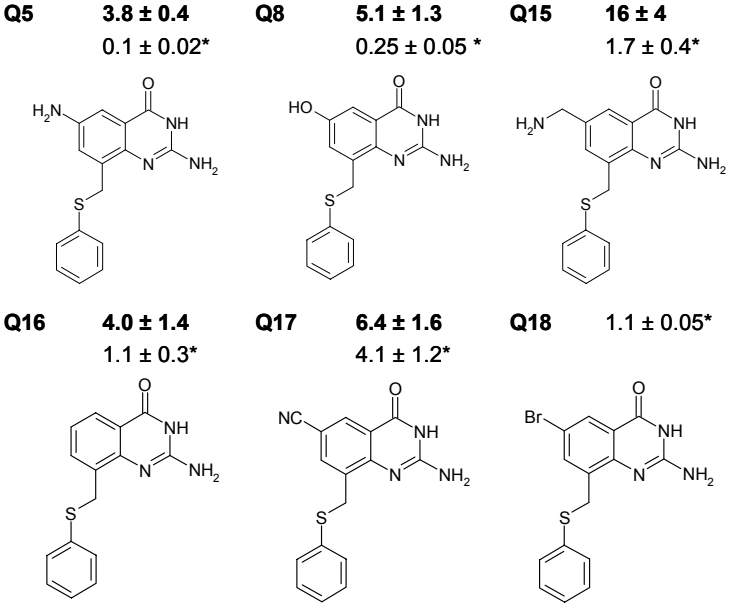
Q5	**3.8 ± 0.4** 0.1 ± 0.02*	**Q11**	5.4 ± 0.6 *	**Q13**	**34 ± 9** 1.4 ± 0.6*
					
Q10	**23 ± 25** 0.6 ± 0.1*	**Q12**	**14 ± 4** 7.7 ± 1.6*	**Q14**	**22 ± 6** 3.5 ± 0.9*
					

K_{ic} in [µM] with average error; (*): previous assay; bold: modified assay

crystal structures complexed by an inhibitor could be determined as a result of enhanced solubility (Brenk et al. 2004). Although the quinazolinone scaffold is well defined in the binding pocket in all structures, the ligand side chains are not properly ordered. While for the spacer split conformations could be observed in some cases, additional side chains are disordered in all cases (Fig. 5.6).

The revalidation of inhibition constants for most compounds appeared essential: Firstly, to estimate the relevance of uncompetitive inhibition for small sized inhibitors, and secondly, to quantify the impact of compound solubility with respect to inhibition. By adding the detergent *Tween 20* to the assay solution the water solubility of low soluble substituted quinazolinones was enhanced. The revalidated affinity data was published in Meyer et al. 2006.

Table 5.6. Quinazolinones – modifications in position 6 to address the peptide switch by Leu231/Ala232.



K_{ic} in [µM] with average error; (*): previous assay; bold: modified assay

5.4.1 Non-competitive inhibition by small-sized quinazolinones

Ligands **Q1** and **Q2** were identified by the trapping experiment as uncompetitive inhibitors. The K_{iu} constants indicate decreasing power to stabilize the covalent complex from **Q1** (0.6 µM) to **Q2** (1.7 µM) (Tab. 5.3). This descending order is also reflected by the intensities of the retarded bands in the SDS-PAGE of the trapping experiment. Facing the competitive inhibition constants (K_{ic}), **Q2** (2.1 µM) shows a similar binding potency compared to **Q1** (1.5 µM). From these findings it can be concluded that the previously reported, very low inhibition constants of **Q1** (350 nM) and **Q2** (20-50nM), (Meyer et al. 2004), compared to substituted compounds, resulted from neglecting the uncompetitive inhibition contribution in the former assay.

Q1 and **Q2** exhibit a similar uncompetitive inhibitory potency as **P1** (K_{iu}: 0.7 µM; Tab. 5.2). Obviously, this results from their similar chemical structures allowing to stabilize the covalent TGT-tRNA intermediate state. Also the competitive inhibition constants for **P1**, **Q1** and **Q2** are in a similar range. Compared to the pyridazindiones **H1** and **H6** (K_{ic}: 60 - 80 µM) they gain affinity by a factor of 40-60. The comparison of the respective crystal structures in Figure 5.5 indicates that a guanine-type scaffold structure is more favourable for binding by allowing Asp102 to rotate into the binding pocket and providing a better binding geometry for Asp156.

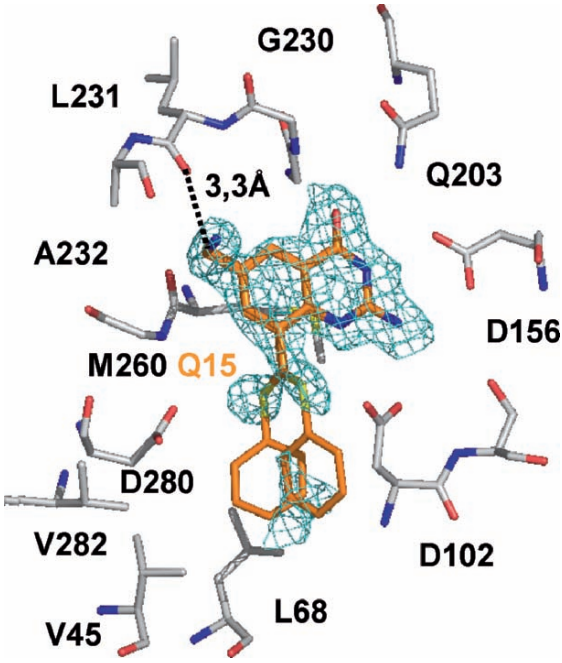


Figure 5.6. Crystal structure of **Q15** in complex with TGT contoured at 1.0σ in the 2 $|F_o|$ - $|F_c|$ density map (adapted from Brenk *et al*. (2004)).

To evaluate the impact of 8-substitution at the quinazolinone scaffold, **Q4** was reevaluated as well (Tab. 5.3). It also shows non-competitive inhibition, but the competitive contribution dominates (K_{ic} = 3.7 µM vs. K_{iu} = 19.1 µM). Consistently, the retarded covalent band in the respective trapping experiment is less pronounced than for **Q2**. In structural terms this observation can be explained by spatial conflicts with the attached methyl group in **Q4** most likely interfering with the ribose ring 34 of the covalently bound tRNA. This effect is even more pronounced for inhibitors bearing larger side chains at this position, which was confirmed by trapping experiments (cf. Fig. 4.3c). Therefore, the contribution of uncompetitive inhibition is further reduced and can be neglected for substituted compounds.

5.4.2. 8-Substituted quinazolinones

All substituted quinazolinones were identified as primarily competitive inhibitors *via* trapping experiments. The revalidation of inhibition constants revealed that they were modulated in a non-linear fashion, most probably due to contributions of non-specific inhibition resulting from low solubility (Meyer et al. 2006). In consequence, the previously drawn assumptions

about structure – activity relationship had to be revised and corrected also for substituted inhibitor series (Brenk et al. 2004; Meyer et al. 2004; Meyer et al. 2002).

Especially the previously reported large difference in binding affinity between S-, O- and C-atoms in the linker, that prompted to postulate a sulphur effect, is no longer evident (Meyer et al. 2002). Instead **Q7**, the compound with an ethylene linker, turned out to be the most potent inhibitor of this series, however with only a slight affinity advantage (Tab. 5.4).

Additionally, the revalidation showed that compounds with substituents other than unsubstituted phenyl rings experience reduced affinity (Tab. 5.5). Whereas in the previous assay no preference for aromatic, aliphatic and polar substituents could be identified, the redetermined values indicate a preference for the phenyl-substituted compounds. Compared to **Q5**, aliphatic substituents (**Q12**) and polar substituents (**Q13**, **Q14**) loose affinity by factors of four to nine. But also further substitution of the phenyl ring of **Q5** resulted in significantly weaker binding (**Q10**).

With respect to modifications in position 6 the enzyme accepts a diverse set of functionalities without showing dramatic changes in binding (Tab. 5.6). The structural basis for this promiscuity is associated with the Leu231/ Ala232 peptide switch. The only exception in the quinazolinone series is compound **Q15** with a four-fold affinity loss compared to **Q5**. However, due to the enhanced solubility, most likely resulting from the charged amino methyl group, crystal structure analysis of the compound in complex with TGT could be successfully performed (Brenk et al. 2004). It indicates repulsive van der Waals contacts of the 6-methyl group with the carbonyl group of Leu231 to be responsible for the affinity loss (Fig. 5.6).

5.4.3. 7-Amino-quinazolinones

The knowledge from the revalidated quinazolinones served as a starting point for a new design approach based on quinazoliones. Comparison of **Q1** and **Q2** clearly revealed that the presence of the 6-amino group is not associated with a significant gain in affinity (Tab. 5.6). Possibly, this is due to the structural adaptability of the binding pocket exemplified by the Leu231/Ala232 peptide switch. Additionally, substitution of the quinazolinone scaffold in position 8 clearly resulted in an affinity loss, due to repulsive interactions of the substituent with Asp102 (Meyer et al. 2004). To circumvent direct repulsion, a set of compounds was developed based on the quinazolinone scaffold **Q2**, substituted in position 7 (Fig. 5.7). An amine function was thought to serve as anchor fragment to address the hydrophobic ribose 34 pocket formed by Val45 and Leu68, avoiding the repulsive interactions observed for substitution in position 8.

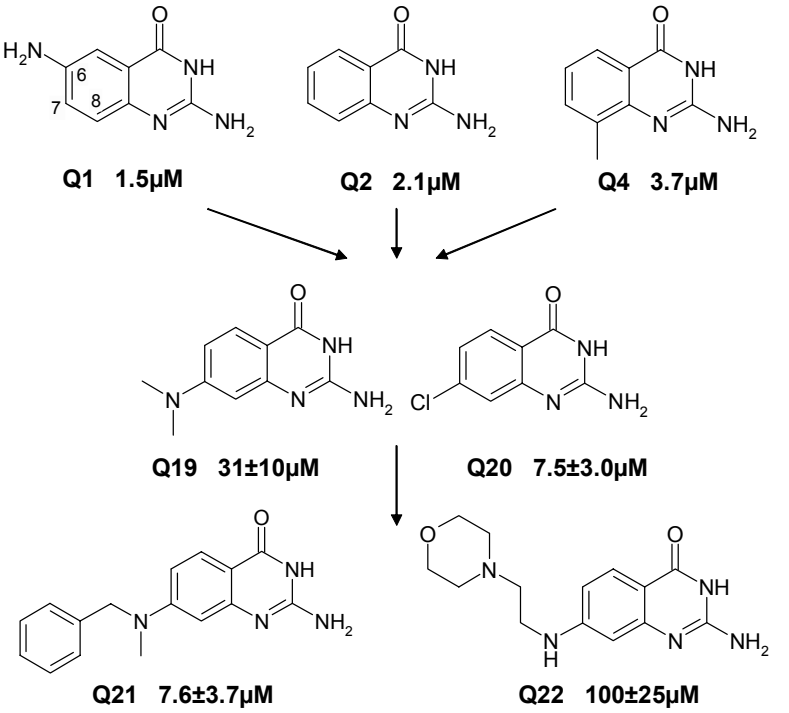


Figure 5.7. Design strategy and inhibition constants (K_{ic}) for 7-substituted quinazolinones.

All newly designed compounds **Q19**-**Q22** were tested for uncompetitive inhibition in trapping experiments. No stabilization of the covalent complex could be observed. Thus, this series represents a set of predominantly competitive inhibitors. Substitution in 7-position seems to interfere significantly with the covalently bound ribose 34 in the TGT-tRNA complex, impeding uncompetitive inhibition. Concerning competitive inhibition all 7-substituted inhibitors showed significantly reduced affinity compared to the parent structure **Q2**. In case of **Q19** affinity decreased by a factor of 15, however phenyl substitution in case of **Q21** could restore affinity by a factor of 4.

Crystal structure analyses have been successfully performed with **Q19** and **Q21** as binary complexes with *Z. mobilis* TGT, resulting in crystals diffracting up to a maximum resolution of 2.0–2.15 Å. In the crystal structure of TGT·**Q19,** the ligand is sufficiently well defined in the G_{34} binding pocket (Fig. 5.8a). The most pronounced difference electron density is observed for the dimethylamino substituent next to Asp280, while the quinazolinone scaffold is less well defined.

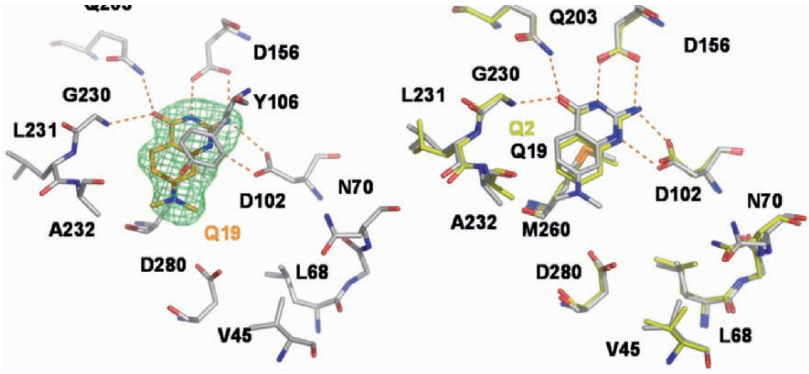


Figure 5.8. a) Crystal structure of TGT·**Q19** with the ligand contoured at 2.5σ in the |Fo| - |Fc| difference electron density map refined in the last step excluding ligand coordinates; b) structural superimposition of the active sites of TGT·**Q19** and TGT·**Q2**.

In the crystal structure of TGT·**Q21** only the quinazolinone scaffold is well defined in the difference electron density. The phenyl substituent is likely scattered over multiple conformations resulting in a rather diffuse difference electron density. In the last refinement cycle, an average B-value of 93.1 Å^2 is assigned to its orientation presented in Figure 5.9a. Due to the higher resolution (2.0 Å) and the slightly better quality of the diffraction data (R-factor: 21.5% and R_{free}-factor: 30.3%) the resulting average B-values for the protein (43.5 Å^2) and the 7-amino-quinazolinone scaffold of the ligand (55.3 Å^2) adopt more reasonable values compared to TGT·**Q19**. Additionally, it was possible to locate two crystal water molecules in the binding pocket, exhibiting average B-values of 45.0 Å^2. The presence of these two water molecules adjacent to the hydrophobic ribose 34 pocket formed by Val45 and Leu68 clearly indicates that the phenyl substituent is exposed to the solvent sticking towards the ribose 33 binding site.

The comparison of the binding mode of **Q19** with the binding mode of **Q2** reveals almost identical orientations for the ligands and the active site residues (Fig. 5.8b). Both compounds are bound to a protein conformer with the Leu231/Ala232 peptide switch in its amide exposing geometry. Thus, from the TGT·**Q19** structure it is difficult to extract obvious reasons for the observed affinity loss.

The comparison of TGT·**Q21** with the high resolution structure of TGT·**Q1** (1.2 Å maximum resolution), however, provides an explanation for

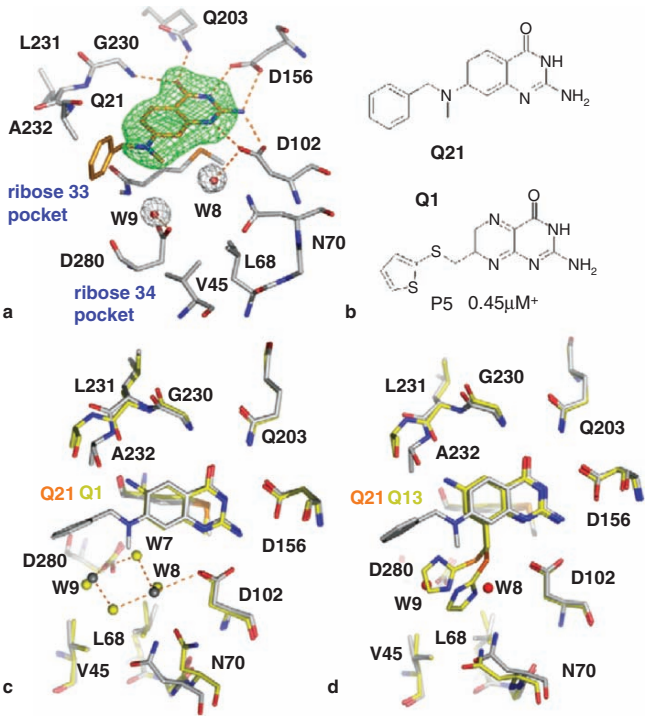


Figure 5.9. a) Crystal structure of TGT·**Q21** with the ligand contoured at 2.0 σ in the |Fo| - |Fc| difference electron density map; b) structural similarity of the quinazolinone inhibitor **Q21** and the pteridine inhibitor **P5** (*K_i referring to former assay); c) structural superimposition of the active sites of TGT·**Q21** and TGT·**Q1**; d) structural superimposition of the active sites of TGT·**Q21** and TGT·**Q13**.

the affinity loss of the 7-substituted quinazolinones (Fig. 5.9c). The obvious difference in TGT·**Q1** is that the Leu231/Ala232 peptide switch is present in the carbonyl exposing conformation. The important difference, however, seems that binding of **Q21** expels a cluster of water molecules mediating interactions between Asp102 and Asp280 from the active site. In TGT·**Q1** these polarizable waters, particularly W7 and W8, are thought to buffer and compensate for the negative charge of the nucleophile Asp280 and Asp102 which recognizes the substrate and catalyzes the proton abstraction. In TGT·**Q21** W7 is expelled from the active site as its distance to the ligand's methyl group would only amount 2.6 Å. In consequence, Asp280 looses its H-bonding contact to Asp102 *via* W8.

Crystal structure analyses of TGT·**Q2** and TGT·**Q4** had already shown that the 8-methyl-substituted **Q4** produces unfavourable contacts to the carboxy group of Asp102 and additionally disrupts the water network (Meyer et al. 2004). Subtracting the uncompetitive inhibitory contributions reveals an almost two-fold affinity decrease of **Q2** (2.1 μM) compared to **Q4** (3.7 μM). Therefore, it can be concluded for TGT that besides unfavourable ligand – protein contacts perturbation of the crystal water network in between Asp280 and Asp102 is detrimental to binding. Both factors have to be compensated by favourable contacts experienced in other regions of the binding pocket.

This assumption is underlined by the superposition of TGT·**Q21** with TGT·**Q13** (Fig. 5.9d). In the latter structure the imidazole substituent penetrates into the hydrophobic ribose 34 pocket and all crystal water molecules are expelled from this site. The binding mode of the most potent substituted quinazolinone **Q7** (2.6 μM) should be similar to that of TGT·**Q13** (34 μM), however the former exposes a phenyl substituent into this pocket. Unfortunately, no crystal structure of **Q7** could be determined. Its competitive inhibition constant is similar to that of **Q1**. Supposedly, the free enthalpy price paid for the replacement of the water cluster next to Asp280 and Asp102 is virtually compensated by additional favourable contacts formed by the added side chain. Obviously, in case of **Q21** replacement of the water cluster is avoided and its phenyl substituent orients towards the ribose 33 binding site which opens to the solvent. Hydrophobic stacking of the **Q21** phenyl substituent with the methyl group of Ala231 is thought to stabilize this orientation and explains the four-fold affinity gain compared to **Q19**.

The binding mode observed for **Q21** also provides an explanation for the surprisingly low binding constant of **P5** (Fig. 5.9b). With 0.45 μM it is the strongest binder from the pteridine series, although this inhibition constant still refers to the former protocol (Tab. 5.2) (Brenk et al. 2003). Due to the low solubility of **P5** no crystal structure in complex with TGT could be determined. **P5** was designed to bind in the hydrophobic ribose 34 pocket, but the structural similarity of **P5** and **Q21** suggests alike binding modes. Both substituents (thiophen and phenyl ring) are expected to orient towards the ribose 33 binding site minimizing the interference with the water cluster.

In summary, the design concept to substitute the quinazolinone skeleton at position 7 was not as successful as expected with respect to inhibitory potency. However, from the affinity data and the crystal structures in complex with TGT valuable information for further design can be extracted. Firstly, substitution in position 7 hampers uncompetitive binding. All 7-substitued derivatives predominantly compete with tRNA binding. Secondly, the water network in between Asp280 and Asp102 is an important structural

element and should only be replaced by polar interactions. Thirdly, addressing the ribose 33 binding pocket instead of the ribose 34 pocket might result in stronger binding due to favourable interactions with Ala232.

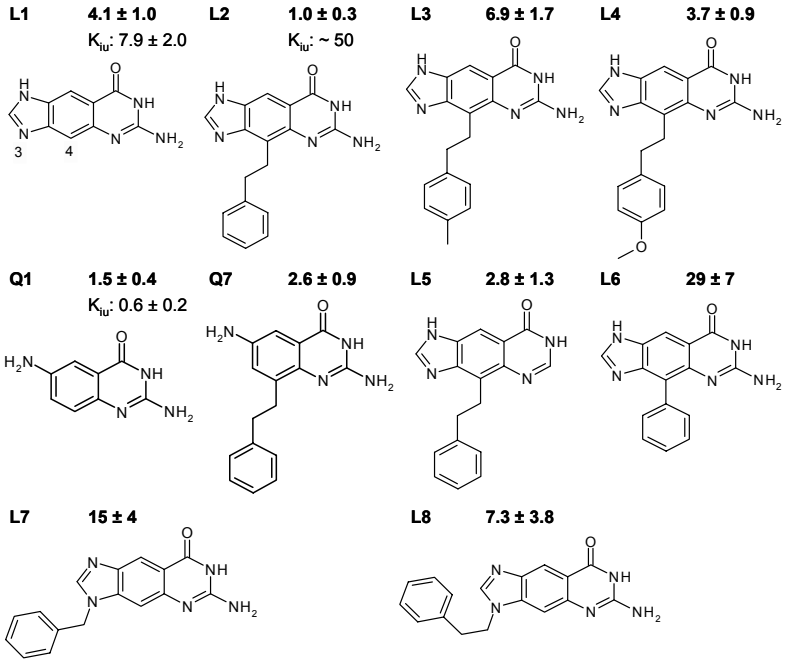
5.5. *lin*-Benzoguanines

The 6-amino-quinazolinone scaffold **Q1** was used as a starting point for the development of a new class of TGT inhibitors. Extended by an imidazole moiety *lin*-benzoguanine (6-aminoimidazo(4,5-*g*)quinazolin-8(7*H*)-one) **L1** was developed as putative scaffold (Tab. 5.7). In case of the quinazolinones, the most potent binders were decorated with aromatic side chains (Tab. 5.4 and 5.5). Therefore, systematic substitutions in 3- and 4-position by aromatic substituents attempted to address the hydrophobic ribose 34 binding pocket formed by Val45, Leu68 and Val282 adjacent to the G_{34} binding pocket (Tab. 5.5). For the 7-amino-quinazolinones it has already been demonstrated, that addressing this pocket is not favourable, due to the perturbation of a water cluster (previous chapter). But *lin*-benzoguanines and 7-amino-quinazolines were developed and analyzed at the same time. Therefore, the considerations concerning the water cluster equally apply also to the *lin*-benzoguanines (chapter 5.5.1). The crystal structures with the substituted *lin*-benzoguanines gave insight into the adaptability of the ribose 34 binding pocket and allowed to investigate the structural basis of induced-fit adaptations observed upon tRNA binding (chapter 5.5.2). Based on the results of quinazolinones and 4-substituted *lin*-benzoguanines, a new series of *lin*-benzoguanines substituted in position 2, was developed and will be presented in chapter 5.5.3.

5.5.1. 3- and 4-substituted lin-Benzoguanines

All *lin*-benzoguanine based inhibitors were tested in trapping experiments for uncompetitive inhibition. Uncompetitive inhibition is only relevant for the basic scaffold **L1** (K_{iu}: 7.9 µM) although this value is already significantly reduced compared to **Q1** (K_{iu}: 0.6 µM) (Tab. 5.7). Obviously, the extended scaffold interferes to some degree with ribose 34 covalently bound in the tRNA-TGT intermediate complex. To assess a possible uncompetitive contribution of substituted *lin*-benzoguanine inhibitors, the K_{iu} of **L2** has been determined. It is separated from competitive tRNA binding by a factor of 50 (Tab. 5.7). Obviously, for substituted compounds uncompetitive inhibition is still possible but on much reduced level. In consequence, as this contribution appears only marginal, which was also indicated by the trapping experiments, only K_{ic} values were determined for the remaining derivatives.

Table 5.7. Competitive inhibition constants of *lin*-benzoguanine based inhibitors in [μM] with average error; bold: modified assay.



Comparison of competitive inhibition constants (K_{ic}) for *lin*-benzoguanine **L1** (4.1 μM) with respect to 6-aminoqinazolinone **Q1** (1.5 μM) shows a slight decrease of inhibitory potency with respect to the former. Substitution of **L1** in position 4 as realized in **L2**, results in a slightly improved affinity (Tab. 5.7). With a K_{ic} of 1.0 μM **L2** is the most potent TGT inhibitor of this series and slightly stronger than the equally substituted 6-amino-quinazolinone **Q7** (2.6 μM). Compound **L5** lacks the C(6)-NH_2 group compared to ligand **L2.** Its three-fold loss in affinity demonstrates the beneficial contribution of this function for binding as it mediates polar contacts to Asp156 and Asp102. Additional substitutions of the phenyl ring in **L2** by *p*-Me (**L3**, 6.9 μM) or *p*-OMe (**L4**, 3.7 μM) appears detrimental to binding. Immediate attachment of the phenyl group, realized in **L6**, results in a substantial loss of activity (29 μM). Similarly, the N(3)-substituted *lin*-benzoguanines experience a affinity loss (**L7**, 15.4 μM; **L8**, 7.3 μM). Structural evidence for these trends will be discussed in the following chapters. Details for synthesis and inhibition constants recently have been published by us (Meyer et al. 2006).

The assumed binding mode of *lin*-benzoguanines could be confirmed by crystal structure analysis of **L1** in complex with TGT at a resolution of 1.7Å.

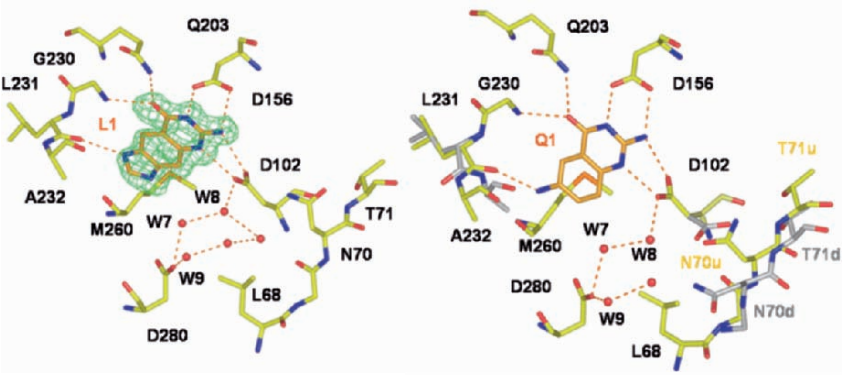


Figure 5.10. a) Crystal structure of **L1** in the binding pocket of TGT determined at 1.7Å resolution and contoured at 2.4σ in the |Fo| - |Fc| density map refined excluding ligand coordinates in the last refinement cycle; b) binding mode of **Q1** in the binding pocket of TGT; coloured in grey: apo conformation, coloured in yellow: inhibitor bound conformation.

The structure was solved and refined by Ruth Brenk (PhD Thesis, 2003). **L1** is well defined in the G_{34} binding pocket and in tight contact with the protein forming 7 H-bonds and 60 van der Waals interactions. 94% of its solvent accessible surface is buried upon complexation (Fig. 5.10a).

The observed binding mode is similar to that of the natural substrate $preQ_1$ or the initial lead **Q1** (Fig. 5.10b). The residual electron density of **L1** is better defined compared to **Q1**. Both, **L1** and the residues forming the G_{34} binding pocket are properly ordered whereas for **Q1** several split side chain conformations were observed (Brenk et al. 2004). It had been concluded that **Q1** occupied not all binding pockets in the crystal upon soaking due to reduced ligand solubility. Thus, the diffraction pattern, recorded as average for the entire crystal, comprises a superimposition of the ligand-bound and the ligand-free binding pockets. The former was refined to 60% occupancy. It shows, superimposed with the unoccupied pocket, for Asp102, Asn70 (Gly69-His73), Leu231 and Ala232 two alternative conformations. This gives rise to the interpretation that upon ligand binding Asp102 has to rotate towards the ligand and Asn70 moves concertedly together with the Val45-His73 backbone strand by 2 Å. A similar conformation is also found in the tRNA-complexed TGT structures. In this orientation Asn70 stabilizes one of the phosphate groups anchoring ribose 34 in the binding pocket. Furthermore, the Leu231/Ala232 peptide bond is flipped to expose its carbonyl oxygen to H-bond the 6-amino group of **Q1**. In TGT·**L1** the ligand is fully occupied. Binding pocket residues adopt only one conformation similar to

TGT·**Q1** in the ligand-bound state. Asp102 recognizes the ligand via a double hydrogen bond. The movement of Asn70 significantly extends the volume of the adjacent hydrophobic cavity formed by Val45 and Leu68. A similar cluster of crystal waters, particularly W7, W8 and W9, stabilize Asp102 and Asp280 in both structures (Fig. 5.10a,b).

In consequence, further design concentrated on filling the enlarged hydrophobic cavity, simultaneously expelling the water cluster. Crystal structure analyses of TGT·**L2** and TGT·**L3** showed that this goal has been successfully achieved with the aromatic substituents occupying this cavity (Fig. 5.11a,c). In both structures, the basic *lin*-benzoguanine scaffold is located equally as in TGT·**L1**. Surprisingly, no significant gain in affinity was observed for the substituted compounds **L2** - **L4** (Tab. 5.7). Obviously, the expected entropic gain does not pay entirely for the enthalpic price of the replacement of the water cluster next to Asp280 and Asp102 (Fig. 5.10a), as already described in detail for the 7-amino-quinazolinones in chapter 5.4.3. The remaining cost virtually compensates for the additional contacts formed by the introduced ligand side chain (60 van der Waals interactions in **L1** vs. 86 – 96 in **L2** and **L3**). All substituted derivatives cross this region of the binding pocket with hydrophobic, hardly polarizable portions unable to form any polar interactions.

For **L7** and **L8** the chemical structure provides a further explanation for the slight affinity loss compared to **L1** (Tab. 5.7). In **L1**–**L4** N(1) is protonated and the Leu231/Ala232 peptide switch is present in the conformation allowing H-bonding *via* the Leu231 carbonyl group (Fig. 5.11a,c). Substitution at N(3) of the imidazole moiety of **L7** and **L8** would require N(1) to be present in deprotonated state, at least as long as this position of the ligand remains uncharged and N(1) does not pick-up a proton. In this unprotonated state, N(1) is not expected to induce a flip of the peptide switch and the amide nitrogen of Ala232 would remain exposed to the binding pocket. But in this conformation the methyl group of Ala232 limits space at the upper rim of the binding pocket for ligand accommodation (see TGT·**Q2** in Fig. 5.5d). The *lin*-benzoguanine scaffold, however, requires extended volume in this region compared to **Q2**, thus proper hosting of **L7** and **L8** appears sterically unfavourable.

To obtain further evidence for this hypothesis, **L8** was placed into the difference electron density of the TGT·**L8** crystal structure in which the bound ligand was unfortunately only partly visible. **L8** was energetically minimized using the force-field program MOLOC. The crystallographically indicated and force-field minimized orientations are very similar (Fig. 5.11e). The superimposition of TGT·**L8** with the binding mode of TGT·**L2**

supports these assumptions. As a result of the Leu231/Ala232 peptide switch present in the NH-exposing orientation, the **L8** scaffold is shifted and rotated inside the binding pocket compared to **L2** (Fig. 5.11f).

Analysis of TGT·**L2** and TGT·**L3** crystal structures reveals an unexpected conformational splitting for both, ligand and active site residues. Both structures were fully refined and resolved to a resolution of 1.58 Å, thus allowing the interpretation of partial occupancy (Fig. 5.11a,c). With respect to a best plane defined by the *lin*-benzoguanine skeleton, the 4-substituents orient either above (up-conformation = u) or below this plane (down-conformation = d). Both orientations are observed side-by-side in the same crystal suggesting virtually equivalent energy content. Both orientations have already been reported for substituted quinazolinone-type inhibitors (e.g. **Q15** – Fig. 5.6) (Brenk et al. 2004). However, in these examples structural disorder was too pronounced and the difference electron densities of ligand substituents and active site residue side chains were not well enough defined to unambiguously relate these split conformations to specific structural adaptations of the binding pocket.

In TGT·**L2** the d-conformation is slightly higher populated (occupancy of 56%) and a low average B-value of 23.1 $Å^2$ is found for the phenyl substituent. The u-conformation was refined to an occupancy of 44% with a slightly increased average B-value of 27.8 $Å^2$ (Fig. 5.11a,b). The two observed conformations of the substituents correlate with two different binding site conformers involving the Asn70 backbone stretch from Gly69 to His73. They differ by 1.4 Å with respect to averaged C_α rmsd. In Figure 5.11b the shift of the Gly69 – His73 backbone is visualized by the most pronounced shift of 1.9 Å experienced by Thr71. For the d-conformation, Asn70 (N70d) is found in virtually the same position as in uncomplexed TGT. In the u-conformation Asn70 (N70u) adopts a position similar to that in TGT·**L1** or TGT complexed by tRNA. The carboxyamine group of Asn70 performs a parallel stacking with the phenyl ring of the substituent. Accordingly, both conformations are largely identical to either the ligand-bound or ligand-free conformations observed in TGT·**Q1** (Fig. 5.10b). But in contrast to this structure, in TGT·**L3** both protein conformers are accommodating the ligand.

In the crystal structure of TGT·**L3**, u- and d-conformation split differently. Electron density suggests that the aromatic substituent favours u-conformation, however with reduced accuracy, as indicated by the rather diffuse side chain electron density (Fig. 5.11c). With an average B-value of 45.6 $Å^2$ the substituent exhibits a significantly higher value compared to the remaining

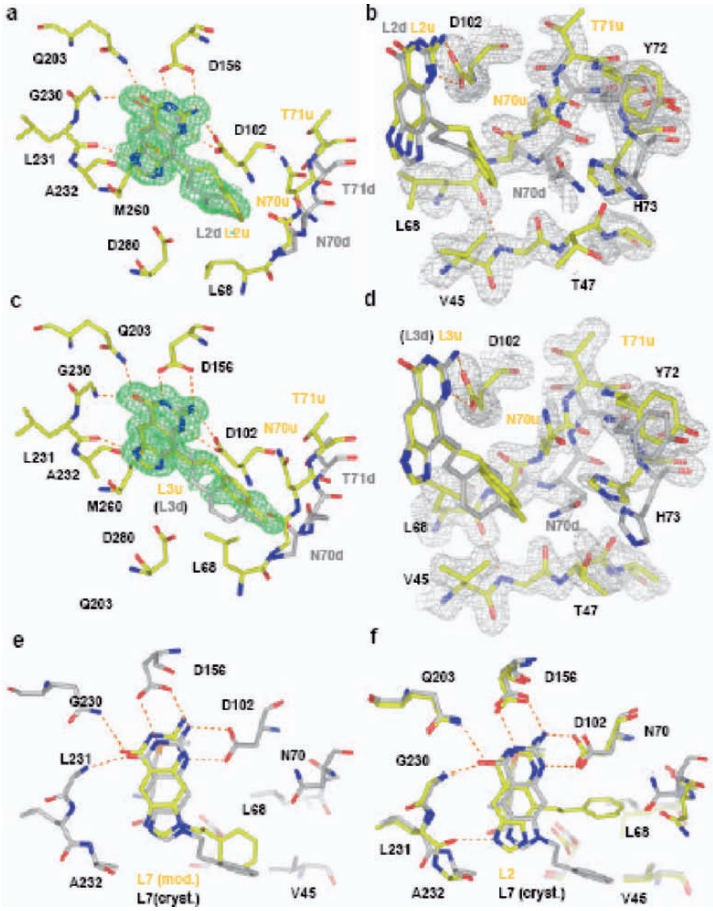


Figure 5.11. a) Crystal structure of TGT·**L2** in the binding pocket of TGT contoured at 2.4σ in the |Fo| - |Fc| density map. The phenyl substituent adopts two conformations (grey = d, yellow = u) paralleled by two conformations for N70; b) electron density of the TGT·**L2** N70 loop contoured at 1.0σ in the 2 |Fo| - |Fc| density map. Associated with the split conformations of **L2** the residues G69-H73 adopt two arrangements; c) crystal structure of TGT·**L3** contoured at 2.4σ in the |Fo| - |Fc| density map. The phenyl substituent was only refined in up-conformation (yellow) although the binding pocket conformation also suggests presence of the down-conformation (grey); d) electron density of the TGT·**L3** N70 loop contoured at 1.0σ in the 2|Fo| - |Fc| density map. Residues G69–H73 adopt two almost equally distributed conformations; e) structural superimposition of the assumed binding mode of TGT·**L8** indicated by the partly occupied binding pocket (grey) and minimization with MOLOC (yellow); f) structural superimposition of the assumed binding mode of TGT·**L8** and the crystal structure of TGT·**L2**.

scaffold atoms (21.3 $Å^2$). The ligand is to 90% buried from solvent accessibility, forming 7 H-bonds and 92 van der Waals contacts. Surprisingly Asn70 is found almost equally distributed in u- and d-conformation (Fig. 5.11d). Consulting TGT·**L1** and TGT·**L2** it can be concluded that the Asn70d conformer is only observed once the *lin*-benzoguanine scaffold is substituted. Therefore, we assume that also for **L3**, the d-conformation of the side chain is present, however distributed over multiple states, thus resulting in a rather blurred and inaccurately defined electron density. In consequence, this conformation was not included in the refinement. Nevertheless, we added its possible geometry in Figure 5.11c,d to show its putative contacts to Asn70d. Supposedly unfavourable contacts experienced by the additional para-methyl group in the d-conformation with Asn70d result in the less-well defined binding mode. Directly facing the orientation of Asn70d in TGT·**L2** and TGT·**L3** shows that this residue is further pushed out of space as a consequence of accommodating the extended para-methyl-phenyl substituent in **L3** (Fig. 5.11a,c).

5.5.2. Induced fit adaptations and importance of active site water

In the crystalline inhibitor complexes determined so far pronounced induced-fit adaptations for e.g. Asp102/Asn70 and Leu231/Ala232 could be observed resembling those experienced by TGT upon tRNA binding. In this context the question arose whether the adaptations result as specific responses to the properties of each individual inhibitor or whether they point to structural adaptations potentially required by the protein in order to accomplish its functional task? If the latter assumption is correct, it can be assumed that these adaptations correspond to low energy transformations required for efficient functioning of the enzyme. Most of the studied inhibitors place their substituent in two alternative conformations into the hydrophobic ribose 34 binding pocket. With respect to the strand hosting Asn70 the protein responds differently to both placements: One geometry relates closely to the uncomplexed, the other to the tRNA-bound situation. For three other residues which are also associated with this structural rearrangement (Tyr106/Gln107 and Thr47) conformations within the ligand bound crystal structures are half-way in-between the uncomplexed and tRNA-bound situation. From this finding we can conclude for inhibitor binding: TGT is highly adaptive, however this multiplicity is obviously required to fulfil its functional role. A ligand to be bound selects one of the binding competent conformers of the protein and stabilizes this state in the complex. Very similar considerations have been suggested to rationalize the binding properties of different inhibitors to aldose reductase, (Urzhumzev et al. 1997; Howard et al. 2004; Sotriffer et al. 2004), or trypsin mutants (Rauh et al. 2002; Rauh et al. 2003).

A further conformational adaptation within the TGT binding pocket leads to another important aspect of ligand binding and catalytic activity: water molecules tightly associated with the binding pocket. The Leu231/Ala232 peptide flip controls substrate recognition and promiscuity (Stengl et al. 2005). Depending on the orientation of the peptide switch either the carbonyl- or the amide-group is exposed to the binding pocket. In two uncomplexed TGT crystal structures exhibiting both orientations, due to deviating pH conditions used for crystallization, these functional groups are H-bonded to two distinct water molecules in deviating positions. Upon ligand binding, they either get replaced or mediate an interaction between the bound ligand and the exposed peptide bond functionality. Further water molecules are indicated as important for the substrate or ligand binding process. One water molecule serves as a surrogate for an alternative protein side chain conformer that allows tRNA binding in the correct orientation. Another water molecule is only found in the active site when the substrate tRNA is bound, as its accommodation requires a conformational adaptation of the binding pocket. As this water molecule is likely to receive a proton from $preQ_1$ during catalysis its presence is highly relevant for the enzymatic reaction (Fig. 4.4, W2). In the ligand bound crystal structure of TGT·**L1** two waters, W7 and W8 (Fig. 5.10a) are bridging between Asp102 and Asp280, the catalytic nucleophile. Attempts to replace them by hydrophobic ligand portions were geometrically successful; however, they did not parallel the expected increase in binding affinity. Despite the assumed entropic benefit for water release, obviously the price for their replacement by a hydrophobic substituent appears detrimental to binding. Possibly, the waters compensate for the negative charge of Asp280 and should be replaced by polarizable ligand functionalities that are able to form polar interactions.

In summary, these findings underline the structural and functional relevance of water molecules associated with the structural adaptability of the binding pocket. The knowledge from this inhibitor series should be considered in further design approaches, particularly in new attempts trying to address the hydrophobic ribose 34 binding pocket.

5.5.3. 2-Amino-*lin*-benzoguanines

Soaking of an impure sample of **L1** resulted in a crystal structure with a difference electron density corresponding to the chemical structure of **L1** (Fig. 5.12c) and an additional difference density peak next to the 2-position. Unfortunately, the chemical structure of the by-product could not be identified from the sample, as it was present in a too small amount. However, from the observations it could be concluded that the 2-substituted by-product of *lin*-benzoguanine displayed a potent TGT inhibitor. Although present only in traces, it was capable to accumulate in TGT crystals. This observation

suggested that substitution of lin-benzoguanine **L1** in position 2 should result in prospective lead candidates. The 2-methyl and 2-amino derivatives **L9** and **L10** turned out to be significantly more potent as the unsubstituted parent structure **L1** (Tab. 5.8).

Subsequently, **L10** was extended by a phenyl (**L11**) and a naphtyl (**L12**) moiety. To identify their inhibition modes, trapping experiments were performed. In contrast to **L1**, no stabilization of the covalent TGT-tRNA intermediate could be observed. Thus, 2-substituted *lin*-benzoguanines represent competitive inhibitors of tRNA, binding to uncomplexed TGT. The competetive inhibition constant (K_{ic}) for all compounds of the new series are in the nanomolar range (Tab. 5.8).

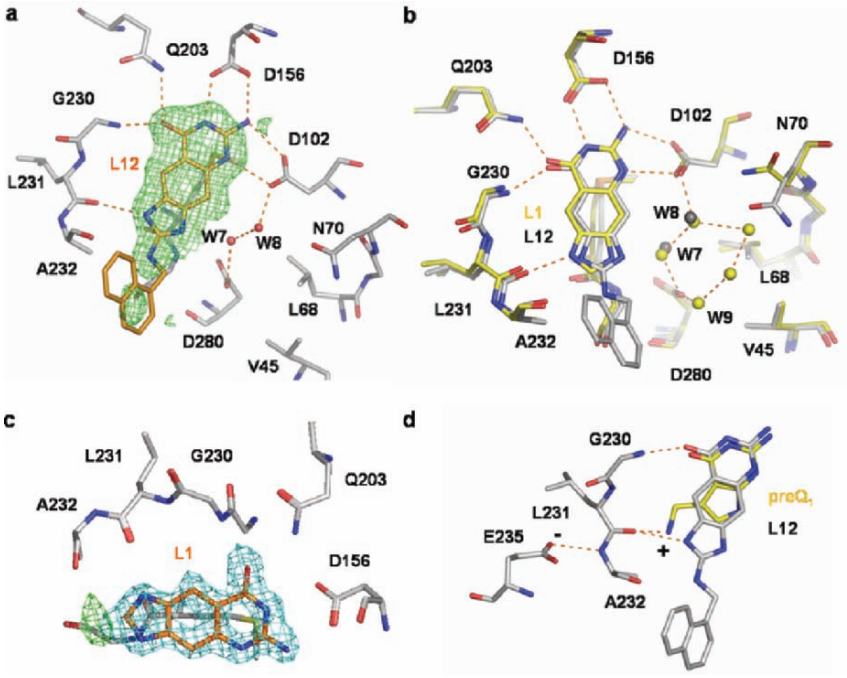
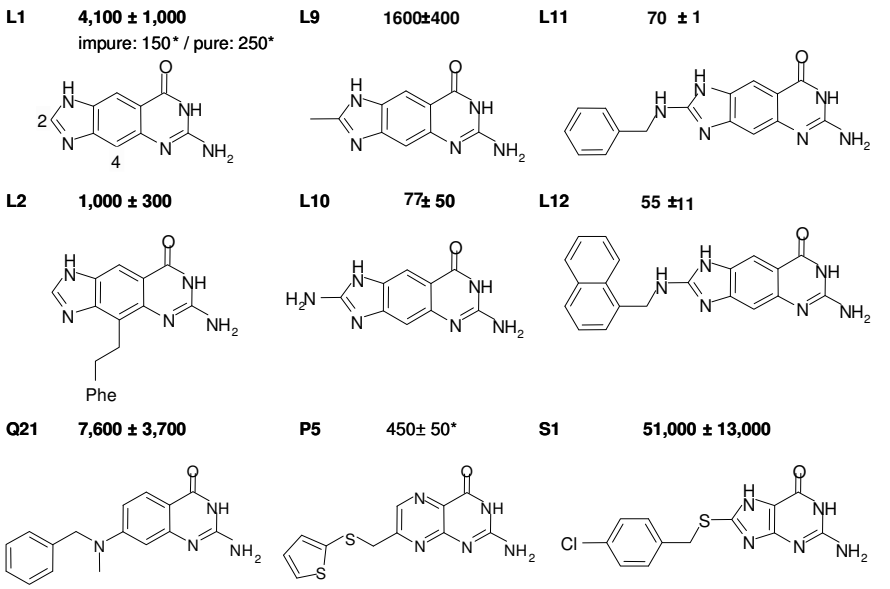


Figure 5.12. a) Crystal structure of TGT·**L12** with the ligand contoured at 2.0σ in the |Fo| - |Fc| difference electron density map refined in the last cycle excluding ligand coordinates to test the model; b) structural superimposition of the active sites of TGT·**L1** and TGT·**L12**; c) crystal structure of TGT·**L1** (impure sample) with the ligand contoured at 1.0σ in the 2 |Fo| - |Fc| density map; additional difference electron density contoured at 2.5σ in the |Fo| - |Fc| difference electron density map indicates the presence of a small 2-substituted derivative of **L1** in the binding pocket; d) structural superposition of the Leu231/Ala232 peptide switch of TGT·preQ$_1$ and TGT·**L12** indicates charge assisted binding.

Crystal structure analysis was successfully performed with **L12** as binary complex with *Z. mobilis* TGT. In TGT·**L12** only the *lin*-benzoguanine scaffold is well defined (Fig. 5.12a). It forms seven H-bonds to residues of the binding pocket and an extended H-bond to the carbonyl-group of Ala232 *via* the endocyclic 2-amino group (distance: 3.6 Å). The naphtyl substituent is scattered over multiple conformations, sticking into the direction of the solvent exposed ribose 33 binding pocket. This is indicated by the rather diffuse difference electron density making it difficult to properly locate the substituent.

An explanation for the affinity gain of the 2-substituted *lin*-benzoguanines might be given by considering charge-assisted hydrogen bonding. This is suggested taking possible pK_a shifts of relevant functional groups in account. For preQ$_1$ the positive charge on the exocyclic amino group supposedly assists binding (Fig. 5.12d). The carbonyl group of Leu231 is stabilized in its orientation by the negatively charged side chain of Glu235. Thus, the Leu231/Ala232 peptide bond projects to some extend its negative charge into the binding pocket. The p*K*a value for an exocyclic amino-methyl group falls into the range of 10 (cf. pK_a of 1-amino-1-phenyl-methane: 10.5; Fig. 5.13). Thus, preQ$_1$ is likely protonated under physiological pH and capable of compensating the negative charge of Glu235 mediated *via* the peptide

Table 5.8. Competitive inhibition constants of 2-amino-*lin*-benzoguanines in comparison with other relevant compounds.



in [nM]; (*): previous assay

bond. In general charge assistance is thought to contribute significantly to binding affinity (Böhm and Klebe, 1996). In contrast, for **L1** no charge assisted H-bond is expected. The pK_a of the imidazole moiety in *lin*-benzoguanine is 4.5 whereas the assay was performed at a pH of 7.3. Thus, **L1** binds most likely deprotonated. Similar considerations hold for the aniline-type 6-amino group of **Q1** (pK_a of 2-amino-naphthaline: 4.1; Fig. 5.13). Substitution of the *lin*-benzoguanine scaffold in position 2 is expected to significantly affect the pK_a value of the imidazole moiety. In case of **L9**, a shift towards higher pK_a is expected as suggested by the pK_a differences of benzimidazole/2-methyl-benzimidazole (5.4 vs. 6.1) or imidazole/2-methyl-imidazole (7.0 vs. 7.8). The pK_a value of the imidazole moiety will be further shifted by the adjacent negatively charged Glu235. This might result in protonation of the imidazole moiety at a pH of 7.3. The resulting charge assisted H-bond could explain the ten-fold affinity gain going from **L1** (4100 nM) to **L9** (300 nM).

For 2-amino-*lin*-benzoguanine an even more pronounced pK_a shift can be expected, as indicated by the comparison of 2-amino-benzimidazole/benzimidazole (5.4 vs. 7.5). The additional amino group generates a guanidinium-type

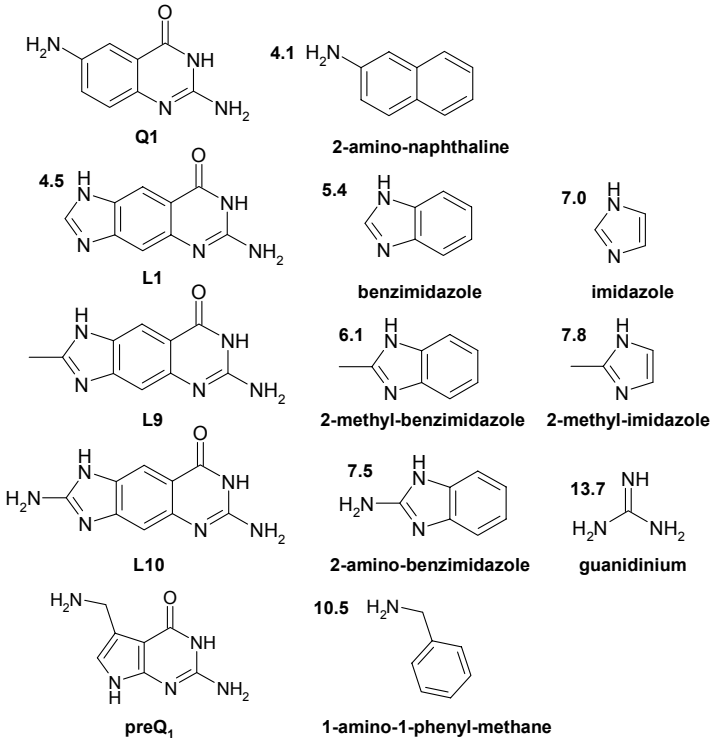


Figure 5.13. pK_a values for nitrogen groups in relevant scaffold structures; values for fragments taken from (Brown et al. 1955; Albert et al. 1948; Bruice and Schmir, 1958).

functionality in the *lin*-benzoguanine scaffold (pK_a of guanidinium: 13.7; Fig. 5.13). Thus, 2-amino-*lin*-benzoguanine should bind in its protonated form and mimic the charge assisted binding mode of preQ$_1$. Compared to **L9**, for **L10** an additional H-bond can be formed to the carbonyl-group of Ala232. This provides an explanation for additional, more than two fold affinity gain of **L10** (125 nM).

The distinct orientation of the substituent towards the ribose 33 binding site marks the significant structural difference of this inhibitor series, resulting from the substitution of the *lin*-benzoguanine scaffold in position 2 (Fig. 5.14b). All previously designed TGT inhibitors were substituted to address the hydrophobic ribose 34 binding pocket formed by Val45 and Leu68 (Grädler et al. 2001; Brenk et al. 2003; Meyer et al. 2002). In the present approach it was intended to direct the substituents towards the ribose 33 binding pocket, flanked by Ala232 and Gly261, to provide a scaffold that will allow further extension towards the uracil 33 binding pocket. Substitution at the 2-amino group, realized by **L11** and **L12**, had only some influence on affinity. Considering the TGT.**L12** crystal structure, it can be concluded that these substitutents experience only little contact with the protein and remain largely solvent exposed. Thus, no significant contribution to binding can be expected.

Compounds that might adopt binding modes similar to **L12** have been identified in previous inhibitor series (Brenk et al. 2003, Brenk et al. 2003). E.g. **P5**, **S1** and **Q21** possess substituents to be oriented towards the ribose 33 binding site (Tab. 5.8). For **Q21** this binding mode was confirmed crystallographically, although unexpected, as the design intended to address the ribose 34 pocket. In TGT·**Q21** the basic scaffold is well defined in the electron density, whereas the phenyl substituent is not properly ordered. The structural superposition of TGT·**Q21** and TGT·**L12** suggests criteria to correlate affinity with structure (Fig. 5.14a). **Q21** binds 100-fold weaker than **L12** (75 nM). One explanation for this drop is the missing charge assistance for **Q21**. The Leu231/Ala232 peptide switch adopts the amide exposing conformation and Glu325 will be uncharged. Furthermore, **Q21** places its N-methyl group in a region that interferes with the water cluster. The phenyl substituent of **Q21** contributes significantly to binding. Compared to the unsubstituted scaffold **Q19** (31,000 nM) a four-fold stronger binding is experienced. Structurally this can be explained by hydrophobic stacking of the phenyl ring with the methyl group of Ala232. In TGT·**L12** a similar contact is found, however, it involves to some degree also the 2-amino group. This might explain the only marginal gain in affinity compared to the unsubstituted **L10** (Tab. 5.8). Obviously, the charge assistance and the unperturbed water network between Asp102 and Asp280 seem to be the key factors for the superior binding of **L12**.

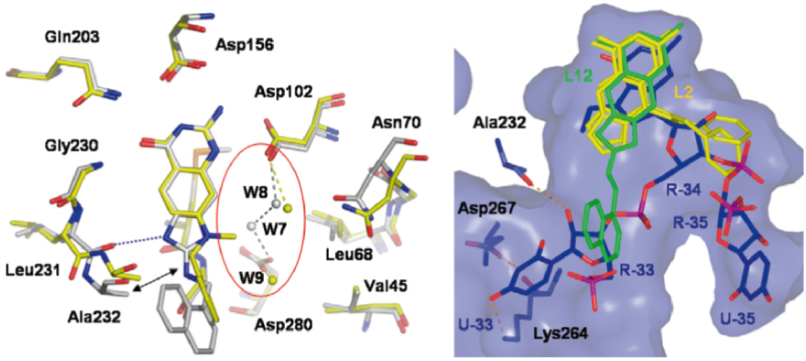


Figure 5.14. a) Structural superposition of the active sites of TGT·**L12** and TGT·**Q21**; b) structural superposition of the binding modes of 4-substituted **L2** and 2-substitued **L12** in the binding pocket of TGT bound to preQ$_1$-tRNA.

Similar arguments hold for the comparison of 2-substituted and 4-substituted *lin*-benzoguanines (Fig. 5.14b). In case of the 4-substitution the water network between Asp280 and Asp102 is expelled from the binding pocket without providing appropriate structural compensation. Most presumably also charge assisted binding is not given for these derivatives.

In summary, the 2-amino-*lin*-benzoguanine scaffold provides a promising starting point for the further design of even more potent nanomolar inhibitors. Therefore, the phenyl and naphtyl substituents should be replaced by better suited substituents specifically addressing the ribose 33 binding site. This might even allow to specifically address the uracil 33 binding pocket in future design steps. Due to their affinity such compound series might also be suited for cell based *Shigella* assays to test virulence reduction under *in-vivo* conditions.

ACKNOWLEDGEMENT

The authors are grateful for the entire TGT developing team for many years of fruitful and productive work. From Marburg we would like to thank Dr. Ulrich Grädler, Dr. Ruth Brenk, Dr. Klaus Reuter, Tina Ritschel, Naomi Tidten and Hans-Dieter Gerber. From ETH Zürich our sincere thanks go to the group of Prof. François Diederich, particularly Dr. Emanuel Meyer and Dr. Simone Hörtner. Constructive collaboration with the group of Prof. George Garcia (Univ. Michigan) is gratefully acknowledged.

REFERENCES

Albert, A., Goldacre, R. and Phillips, J. (1948) The strength of heterocyclic bases. *J. Chem. Soc*, 2240-2248.

Ashkenazi, S., Levy, I., Kazaronovski, V. and Samra, Z. (2003). Growing antimicrobial resistance of *Shigella* isolates. *J. Antimicrob. Chemother.* **51**: 427-429.

Bennish, M.L. (1991) Potentially lethal complications of shigellosis. *Rev. Infect. Dis.* **13**: S319-S324.

Böhm, H. and Klebe, G. (1996) What Can We Learn From Molecular Recognition in Protein-Ligand Complexes for the Design of New Drugs? *Angew. Chem. Int. Ed. Engl.,* **35**: 2588-2614.

Brenk, R. Virtuelles Screening, strukturbasiertes Design und Kristallstrukturanalyse von Inhibitoren der tRNA-Guanin Transglykosylase, ein Target der Bakterienruhr (PhD Thesis). Philipps-University of Marburg; 2003.

Brenk, R., Naerum, L., Gradler, U., Gerber, H.D., Garcia, G.A., Reuter, K., Stubbs, M.T. and Klebe, G. (2003a) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem*. **46**: 1133-43. 64.

Brenk, R., Stubbs, M.T., Heine, A., Reuter, K and Klebe, G. (2003b) Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. *Chembiochem*. **4**: 1066-77.

Brenk, R., H.-D. G, Kittendorf, J.D., Garcia, G.A., Reuter, K. and Klebe, G. (2003c) From Hit to Lead: *De Novo* Design Based on Virtual Screening Hits of Inhibitors of tRNA-Guanine Transglycosylase, a Putative Target of Shigellosis Therapy. *Helv. Chim. Acta*. **86**: 1435-1452.

Brenk, R., Meyer, E.A., Reuter, K., Stubbs, M.T., Garcia, G.A., Diederich, F. and Klebe, G. (2004) Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. *J. Mol. Biol*. **338**: 55-75.

Brown, H.C. (1955) In: Braude, E.A., Nachod, F.C., eds. *Determination of Organic Structures by Physical Methods*. New York: Academic Press.

Bruice, T.C. and Schmir, G.L. (1958) Imidazole Catalysis. II. The Reaction of Substituted Imidazoles with Phenyl Acetates in Aqueous Solution *J. Am. Chem. Soc.* **80**: 148.

Björk, G. (1996) Stable RNA modification. In: Neidhardt, F.C., Curtiss I.R., Ingraham, J.L., Lin, C.C.C., Low, J.K.B., Magasanik, Bea, eds. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2 ed. Washington, DC: American Society for Microbiology Press, 861-886.

Cohen, D., Green, M., Block, C., Slepon, R., Ambar, R., Wassermann, S.S. and Levine, M.M. (1991) Reduction of transmission of shigellosis by control of houseflies (Musca domestica). *Lancet*. **337**: 993-997.

Dorman, C.J. and Porter, M.E. (1998) The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. *Molecular Microbiology*. **29**: 677-684.

Durand, J.M., Okada, N., Tobe, T., Watarai, M., Fukuda, I., Suzuki, T., Nakata, N., Komatsu, K., Yoshikawa, M. and Sasakawa, C. (1994) vacC, a virulence-associated chromosomal locus of *Shigella flexneri*, is homologous to tgt, a gene encoding tRNA-guanine transglycosylase (Tgt) of *Escherichia coli* K-12. *J. Bacteriol.* **176**: 4627-4634.

Durand, J.M., Björk, G.R., Kuwae, A., Yoshikawa, M. and Sasakawa, C. (1997) The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of *Shigella flexneri* is required for expression of virulence genes. *J. Bacteriol.* **179**: 5777-5782.

Durand, J.M., Dagberg, B., Uhlin, B.E. and Bjork, G.R. (2000) Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the virF gene. *Mol. Microbiol.* **35**: 924-935.

Durand, J.M. and Björk, G.R. (2003) Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of *Shigella flexneri*: a possible role in adaptation of virulence. *Mol. Microbiol.* **47**: 519-527.

Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.B., Bui, H., Le Bouguénec, C. and Denamur, E. (2004) A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in *Escherichia coli*. *Molecular Biology and Evolution*. **21**: 1085-1094

Fernandez, M.I. and Sansonetti, P.J. (2003). *Shigella* interaction with intestinal epithelial cells determines the innate immune response in shigellosis. *Int. J. Med. Microbiol.* **293**: 55-67.

Goodenough-Lashua, D.M. and Garcia, G.A. (2003) tRNA-guanine transglycosylase from E. coli: a ping-pong kinetic mechanism is consistent with nucleophilic catalysis. *Bioorg. Chem*. **31**: 331-44.

Grädler, U., Gerber, H.D., Goodenough-Lashua, D.M., Garcia, G.A., Ficner, R., Reuter, K., Stubbs, M.T. and Klebe, G. (2001) A New Target for Shigellosis: Rational Design and Crystallographic Studies of Inhibitors of tRNA-guanine Transglycosylase. *J. Mol. Biol*. **306**: 455-467.

Goma Epidemiology Group. (1995) Public health impact of Rwandan refugee crisis: what happened in Goma, Zaire, in July, 1994? *Lancet*. **345**: 339-344.

Heikkila, E. (1990) Increase of trimethoprim resistance among *Shigella* species, 1975-1988: analysis of resistance mechanism. *J. of Infectious Diseases*. **161**: 1242-1248.

Howard, E.I., Sanishvili, R., Cachau, R.E., Mitschler, A., Chevrier, B., Barth. P., Lamour, V., Van Zandt, M., Moras, D., Schneider, T.R., Joachimiak, A. and Podjarny, A. (2004) Ultrahigh Resolution Drug Design I: Details of Interactions in Human Aldose Reductase–Inhibitor Complex at 0.66 Å. *Proteins: Structure, Function and Bioinformatics*. **55**: 792-804.

Jennison, A.V. and Verma, N.K. (2004) *Shigella flexneri* infection: pathogenesis and vaccine development. *FEMS Microbiology Reviews*. **28**: 43-58.

Kotloff, K., Winickoff, J., Ivanoff, B., Clemens, J., Swerdlow, D., Sansonetti, P., Adak, G. and Levine, M. (1999) Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization (WHO Bull.)*. **77**: 651-666.

Mathan, M.M. and Mathan, V.I. (1991) Morphology of rectal mucosa of patients with shigellosis. *Rev. Infect. Dis.* **13**: S314-S318.

Meyer, E.A., Brenk, R., Castellano, R.K., Furler, M., Klebe, G. and Diederich, F. (2002) De Novo Design, Synthesis, and *In Vitro* Evaluation of Inhibitors for Prokaryotic tRNA-Guanine Transglycosylase: A Dramatic Sulfur Effect on Binding Affinity. *Chembiochem*. **3**: 250-253.

Meyer, E.A., Furler, M., Diederich, F., Brenk, R. and Klebe, G. (2004) Synthesis and *In Vitro* Evaluation of 2-Aminoquinazolin-4(3*H*)-one-Based Inhibitors for tRNA-Guanine Transglycosylase (TGT). *Helv. Chim. Acta*. **87**: 1333-1356.

Meyer, E.A., Donati, N., Guillot, M., Schweizer, W.B., Diederich, F., Stengl, B., Brenk, R., Reuter, K. and Klebe, G. (2006) Synthesis, Biological Evaluation, and Crystallographic Studies of Extended Guanine based (*lin*-Benzoguanine) Inhibitors for tRNAGuanine Transglycosylase (TGT). *Helv. Chim. Acta*. **89**: 573-597.

Okada, N. and Nishimura, S. (1979) Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from *Escherichia coli*. *J. Biol. Chem*. **254**: 3061-3066.

Rauh, D., Reyda, S., Klebe, G. and Stubbs, M.T. (2002) Trypsin mutants for structure-based drug design: expression, refolding and crystallisation. *J. Biol. Chem.* **383**: 1309-1314.

Rauh, D., Klebe, G., Sturzebecher, J. and Stubbs, M.T. (2003) ZZ made EZ: influence of inhibitor configuration on enzyme selectivity. *J. Mol. Biol.* **330**: 761-770.

Romier, C., Reuter, K., Suck, D. and Ficner, R. (1996) Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. *EMBO J*. **15**: 2850-2857.

Sansonetti, P.J. (2001a) Rupture, invasion and inflammatory destruction of the intestinal barrier by Shigella, making sense of prokaryote-eukaryote cross-talks. *FEMS Microbiology Reviews*. **25**: 3-14.

Sansonetti, P.J. (2001b) Microbes and microbial toxins: paradigms for microbial-mucosal interactions III. Shigellosis: from symptoms to molecular pathogenesis. *Am. J. Physiol. Gastrointest. Liver Physiol*. **280**: G319-G323.

Stengl, B., Reuter, K. and Klebe, G. (2005) Mechanism and Substrate Specificity of tRNA–Guanine Transglycosylases (TGTs): tRNA Modifying Enzymes from the Three Different Kingdoms of Life Share a Common Catalytic Mechanism. *Chembiochem*. **6**: 1-15.

Sotriffer, C., Krämer, O. and Klebe, G. (2004) Probing flexibility and 'induced-fit' phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations *Proteins: Structure, Function and Genetics*. **56**: 52-66.

Urzhumtsev, A., Tete-Favier, F., Mitschler, A., Barbanton, J., Barth, P., Urzhumtseva, L., Biellmann, J.F., Podjarny, A. and Moras, D. (1997) A 'specificity' pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. *Structure*. **5**: 601-612.

Van Nhieu, G.T., Bourdet-Sicard, R., Duménil, G., Blocker, A. and Sansonetti, P.J. (2000) Bacterial signals and cell responses during *Shigella* entry into epithelial cells. *Cellular Microbiology*. **2**: 187-193.

Xie, W., Liu, X. and Huang, R.H. (2003) Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol*. **10**: 781-788.

This page intentionally blank

Virtual Ligand Screening: A Method to Discover New Drug Leads

GERHARD KLEBE
Institute of Pharmaceutical Chemistry, Philipps-University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, Phone: +49 6421 282 1313, Fax: +49 6421-28 28994, E-mail: Klebe@mailer.uni-marburg.de

Abstract

In Virtual Ligand Screening (VS) lead compounds are selected, in contrast to High-Throughput Screening (HTS), by predicting their binding mode to a particular target receptor under consideration on the computer. Key prerequisite is the knowledge about the spatial and energetic criteria responsible for protein-ligand binding. The concepts and prerequisites to perform VS are summarized, given limitations are analyzes and explanations are sought for why we still face these limitations. Target selection, analysis and preparation are discussed, followed by considerations about the compilation of candidate ligand libraries. Then the tools and strategies of a VS campaign are reflected, along with the accuracy of scoring and ranking of the screening results.

Keywords: Leads Compound Discovery, Screening Technologies, Protein-Ligand Interactions, Docking, Scoring, Affinity Prediction.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 251–272.

1. INTRODUCTION

In the last two decades experimental high-throughput screening and combinatorial chemistry were aggressively developed to overcome the lead discovery bottleneck in drug development. Using sophisticated large-scale automation, it was anticipated that this would generate an unprecedented number of novel leads resulting in a substantial leap of novel drug entities launched to the market per year. Experience showed, however, that the opposite was the case (Bolten and deGregorio 2002; Fishman and Porter 2005). Frequently the discovered hits could not be validated and further optimized into actual leads and preclinical candidates. Thus, the initial euphoria surrounding these approaches has subsided owing to the disappointingly low hit rates and significant costs involved (Lahana 1999); Langer and Krovat 2003).

This situation fuelled the consideration and development of alternative techniques. The expression "virtual screening" was coined in the late nineties, however, the techniques involved are much older. In an effort to show that searching for potential lead candidates on the computer is a serious alternative to HTS, the term virtual screening (VS) was picked up by the community. In contrast to HTS, which is largely phenomenological and technology-driven, in VS compounds are selected by predicting their binding to a target receptor on the computer. The compounds to be studied do not necessarily exist and their "testing" does not consume valuable substance material. Experimental deficiencies, such as limited solubility, aggregate formation or any sort of influence that possibly interferes with experimentally applied assay conditions do not matter in the initial computational screen. In contrast to HTS, VS requires as key prerequisite knowledge about the spatial and energetic criteria responsible for the binding of a particular candidate ligand to the target receptor under investigation. In consequence, either the three-dimensional structure of the target is given by crystal structure analyses, NMR or sophisticated homology modelling or, at least, a rigid reference ligand with known bioactive conformation mapping out the putative receptor binding site must be available (Langer and Krovat 2003). This defines VS as a knowledge-driven approach.

As mentioned, the methods involved in VS are much older than the approach itself. First attempts to find ligands by docking or mapping onto ligand-based pharmacophore hypotheses are the generic prototypes of a virtual ligand screening approach. However, nobody used this expression in those days, likely because computers and algorithms were not fast enough to allow for large scale applications. Focussing on approaches that actually make use of an available target structure, one of the first systems to be studied by docking was HIV protease. Initial versions of the program DOCK, developed over many years in the lab of I.D. Kuntz (Kuntz et al.

1982; DesJarlais et al. 1986), tried to dock rigid entries from the Cambridge Crystallographic database into the protein, focussing primarily on shape complementarity and later considering chemical complementarity. In 1990, the Kuntz group retrieved the neuroleptic drug haloperidol from a docking screen as potential lead, but the compound would have had to be administered at very high dose level to be effective, far beyond a toxicologically tolerable concentration (DesJarlais et al. 1990). Nevertheless, it gave rise to some ideas how to develop a 15 μM derivative (Rutenber et al. 1993). Later at Dupont-Merck a 3D database search retrieved a substituted terphenyl derivative as putative lead for HIV protease. Further optimization via six- and seven-membered rings resulted in the class of cyclic ureas that replace the structural water in the protease addressing at the same time both catalytic aspartates via appropriately placed hydroxy functionalities (Lam P.Y. et al. 1994). Since these early virtual screening attempts, a plethora of case studies has been performed and the list of success stories is steadily growing. Despite the fact that the approach is still a young discipline, it has already been reviewed in other contributions (Lyne 2002; Barril et al. 2004); Shoichet 2004; Kitchen et al. 2004; Jain 2004), most recently in a comprehensive overview by Kubinyi (Kubinyi 2006). If one excludes purely retrospective studies, where only known actives were retrieved from a sample of anticipated non-actives to show the potential of a method to enrich on putatively active molecules, nowadays about 50 targets have been documented as studied and reports on the discovery of mostly micro molar binding ligands in a truly predictive fashion are available (Kubinyi 2006). Being still under development and far from mature, the number of strategies followed in VS is nearly as large as the number of reported screening campaigns.

2. TARGET SELECTION, VALIDATION AND ANALYSIS

Being a knowledge-driven approach, the scope of VS strongly depends on the amount and quality of information available about the system under investigation. Clearly, studies for which the structure of the target receptor is available will be of advantage compared to situations where only a rigid reference ligand is known. Since structures of the target receptors are increasingly available, this contribution solely concentrates on such situations. The target receptor could be any macromolecular biomolecule, either a protein, RNA or DNA. For the latter cases, only recently the development of appropriate docking and scoring tools has been initiated, and accordingly VS applications on such targets are still rare (Charifson et al. 2004; Filikov et al.

2000). In contrast, the scope of applications addressing proteins ranges from enzymes to GPCRs and ion channels (Kubinyi 2006).

The first issue to be discussed is the druggability of the selected target. Does the selected protein exhibit a binding pocket that can be successfully addressed by small molecule ligands? Clearly, characteristics such as pocket size and geometry, surface complexity and roughness, exposure of recognition properties and their complementarity in shape and polarity with respect to a putative druglike ligand, are of importance. A pragmatic, but hardly general approach would be correlating gene families. If one member of such a family is able to bind a drug, some chances might exist that also other members are able to bind druglike ligands with related physicochemical properties (Hopkins and Groom 2002). However, the setup for a VS study requires more conclusive information about the actual binding-site architecture. Recently, Hajduk et al. suggested a number of very decisive indices that help to discriminate druggable from nondruggable binding pockets (Hajduk et al. 2005). For druggability the total surface area and a portion of a polar contact area below 75 $Å^2$ appear to be beneficial along with an appropriate pocket compactness, surface roughness and complexity. This suggests that there is an optimal size and composition of a protein binding pocket that is best suited to recognize and accommodate small organic ligands. However, the analysis also showed that no single index consistently dominates the correlation. This fact points to the complexity of the interrelationship and explains why at present our concepts to predict druggability are still very rudimentary.

The selection of an appropriate 3D geometry for the target is another important issue when setting up a VS run. The most powerful method to learn about the spatial structure of proteins is crystal structure analysis. The accuracy and reliability of this method strongly depends on the resolution of the diffraction data. An alternative experimental technique is NMR; there the accuracy of the structure determination depends strongly on the local distribution of NOE (**N**ulcear **O**verhauser **E**ffect) distance information. In the past, significant methodological enhancements in homology modelling have improved the quality of protein models (Fiser and Sali 2003; Wallner and Elofsson 2005; Hillisch et al. 2004); accordingly, several successful VS screening campaigns have been reported based on model-built protein structures (Bissantz et al. 2003; Evers and Klebe 2004; Evers and Klabunde 2005). Homology modelling strongly depends on the availability of related proteins for which a crystal structure has been determined. The model will be particularly precise in those areas where the homology with the experimentally determined references is high. This is usually given in the conserved core regions. However, functional binding sites, to be addressed by VS, are generally located in loop regions where even among homologues

of a gene family significant differences are experienced. To enhance the accuracy of model-built structures in regions next to binding sites, new algorithms have been suggested that consider the putative binding orientation of known ligands during the homology modelling process. They feed the information about the binding properties of possibly bound ligands back into the homology building process. Improved binding site geometries can be expected from such approaches (Schafferhans and Klebe 2001; Evers et al. 2003).

Another obstacle that complicates VS attempts is molecular flexibility. Both ligands and proteins possess internal degrees of freedom and can adopt various conformational states. With respect to the receptor, several methods have been described to simulate flexibility (Teague 2003). For VS it is important to obtain an estimate on the binding competent conformers of the protein. Once such a view has been collected, a VS run can either be performed by considering the flexibility of the protein on the fly or by addressing an ensemble of several rigid receptor conformations (Knegtel et al. 1997; Leach 1994; Claussen et al. 2001; Wei et al. 2004). Binding competent conformers can either be sampled by exhaustive conformational searches e.g., using molecular dynamics simulations (Carlson and McCammon 2000), or examining multiple conformational states, observed in crystal structures with different ligands bound, to obtain insight into the relevant conformations (Sotriffer et al. 2004; Barril and Morley 2005). The latter approach appears quite tempting because it can be assumed that a sample of experimentally observed protein conformers agrees to stable low free-energy states.

Even though the crystal structure is usually considered as "gold standard" to learn about the geometry of a protein, it is highly in question how representative a single structure determination really is. McGovern and Shoichet reported on increasing information decay whether the geometry of a ligand-bound or ligand-free crystal structure is used or a model-built one is considered (McGovern and Shoichet 2003). We recently observed, depending on the soaking protocol applied, different conformational states of the protein aldose reductase complexed by zopolrestat. In one structure the protein forms an H-bond to the ligand via one of its binding site-exposed amide bonds (Steuber et al. 2006). In a second structure, the amide bond is rotated off from the binding site and the same ligand cannot form this H-bond any longer. Such changes have significant impact on docking and scoring results in VS. Depending on the crystallization conditions ligands have been observed to adopt reversed binding modes (Stubbs et al. 2002) or induced conformational adaptations of the protein are observed (Rauh et al. 2003). Increasing the concentration of a ligand in the soaking buffer will enhance the chances to accommodate it in the binding pocket. However, most likely

soaking depicts some kind of kinetic trap for ligands, in consequence surprising differences to co-crystallization have been observed (several ligands bound at a time, multiple binding modes (Steuber et al. 2006; Rauh et al. 2003)). Before selecting a particular crystal structure as reference for an VS run, detailed analysis of parameters such as population of the bound ligand, B-factors next to the binding site, or consistency of the H-bond network is advisable. In addition, most programs used for the actual computer screens require properly defined protonation states of the active-site residues. This is by no means a trivial problem, since local dielectric conditions can modulate pKa values of functional groups by several orders of magnitude. Even more complex and at present difficult to calculate are pKa shifts of residues in due course of ligand binding (Czodrowski et al. 2006b; Czodrowski et al. 2006a). Isothermal titration calorimetry (ITC) (Jelesarov and Bosshard 1999) measurements can make such changes apparent and most likely they occur more frequently than we appreciate at present (Gohlke and Klebe 2002). They can easily turn an acceptor functional group into a donor or a charge-assisted H-bond to a neutral one. These changes matter significantly at various validation steps of a VS run. We have recently observed that the protonation states in the binding pocket of inhibitor-bound aldose reductase change depending on whether the cofactor is present in oxidized or reduced state (Steuber et al. 2006). With respect to protein flexibility the decision has to be taken whether a VS run should consider one single protein conformer or multiple binding-competent states of the receptor. Different strategies with respect to docking and scoring will be the consequence (s. below). Experience shows that conformational adaptations of a protein binding pocket are usually related to structural modulations that a protein knows about to fulfil its functional role (Sotriffer et al. 2004). Those adaptations must correspond to low energy states, otherwise dramatic short comings in the functional performance would be the consequence. Accordingly, in particular enzymes that operate on a large palette of structurally diverse substrates (cf. aldose reductase, short chain dehydrogenases) or perform substantial conformational adaptations in due course of the catalytic mechanism (cf. kinases) will experience multiple binding-competent states that have to be considered in VS. Thus, a profound understanding of the functional properties of a protein might be the best option to predict conformational adaptability of a protein. Usually, such information is hardly given at the beginning of a drug development project when VS is consulted. Since proteins occur in gene families of closely related members, the analysis across different entries of the family might unravel some of the flexibility properties inherently given within the fold of the family. However, it might also be that some family members display rigid solutions as

required by their function whereas others possess pronounced flexible behaviour to properly accomplish their function (Reyda et al. 2003).

After the validation of the relevance of the selected target structure, its binding site properties should be mapped prior to blindly starting a VS run. Several tools have been described to elucidate the "hot spots" of binding in a particular binding pocket (Gohlke et al. 2000). These methods are either based on thoroughly parameterized force fields (Goodford 1985) or on well-selected empirical information (Verdonk et al. 1999; Gohlke et al. 2000). Most importantly, this analysis has to be performed on all multiple conformational states of a binding pocket, since this will provide a composite picture of how the molecular recognition properties of a binding site might change upon protein adaptation. Finally, a very crucial decision with respect to the setup of the protein reference for a VS run concerns the consideration of water molecules in a binding site. The analysis of several thousand crystal structures of ligand-protein complexes using the waterbase module in Relibase (Hendlich et al. 2003; Günther et al. 2003) revealed in about two thirds of all cases a water molecule to be involved in ligand binding, frequently mediating contacts between protein and ligand. Thus, any approach based on the prediction of binding modes of putative candidate ligands as required in VS, has to take water into consideration. One possible approach, at least to learn about firmly bound waters, is the analysis of crystallographic data with respect to the repetitive occurrence of waters in structurally related binding sites (e.g. in a gene family) or multiple structure determinations of the same protein with many distinct ligands (Günther et al. 2003). If in such examples water appears recurrently in all structures this might be a strong indication that this water is firmly bound and could be considered in a VS run as integral part of the target structure.

3. SELECTION OF CANDIDATE LIGANDS

As HTS, VS needs a thoughtfully designed and thoroughly compiled sample of small molecule candidate ligands for screening. Pharmaceutical companies will primarily screen their own propriatory compound collections, giving advantage that detected hits will be exclusive and cover a chemistry well established at their site. However, are such sample pools strongly biased? How well is the available chemical spaced covered? These considerations have lead big pharma companies to complement their in-house collection by compounds offered through commercial suppliers. Starting with the ACD as initial prototype (Web site: http://www.mdli.com), today more than 10 million unique purchasable compounds are offered. How well do they cover chemical space and which portion of this space represents druglike molecules? Some dramatic numbers have been proposed how many organic molecules

are principally imaginable under the constraints of being druglike (H, C, N, O, P, S, Cl, Br, MW <500Da (Bohacek et al. 1996)). How diverse should the entries of a screening database be? What physicochemical properties have to be met by the candidates to guarantee sufficient bioavailability? The concept of a well-balanced and homogeneously populated space of diverse druglike molecules appears very tempting; however, nobody can properly define what descriptors to use as coordinates on the axes of such a compound space. What is diverse in this context? As the criterion 'similarity', the expression 'diversity' is a relative measure that relates to a reference point. In VS the reference point is the target and the affinity difference of two ligands with respect to this structure. However, for one target a correctly placed methyl group might have a dramatic effect on binding affinity whereas for another it will be tolerated without any difference. Ligand-based diversity scores will most probably rank the methylated and unsubstituted derivative as 'highly similar'.

Is there an optimal size for candidate ligands to be submitted to screening? In the past, especially combinatorial chemistry has allowed pharmaceutical companies to develop screening libraries of several million candidate molecules, thus in principle the sheer number of test compounds is no longer an issue for them. Although large in total count, the individual members of such libraries are usually rather large in terms of size and molecular weight. They are mostly in the range of typical drug-size molecules since they were produced in other drug development projects. However, if they turn out as micro molar screening hit they still require optimization by two or three orders of magnitude in affinity toward the target under consideration. This step has to be accomplished while maintaining molecular size and ADME properties. In consequence, it involves stripping down to the core skeleton while decorating with novel well-tailored side chains. This is a challenging task, however, often difficult to achieve. Experience from drug optimization programs has shown that small core fragments ("privileged templates") known to bind with significant affinity are ideal starting points for further optimization (Erlanson et al. 2004; Rees et al. 2004). Accordingly, the compounds selected for VS should leave some room for optimization, thus matching with the range of so-called lead-like molecules (Oprea 2002).

Verdonk et al. (2004) have presented a very insightful discussion on approaches to be taken constructing databases for VS, at least to validate the involved methods, e.g., in terms of enrichment rates of known binders. Such considerations are extremely important for methods development, however, in an actual drug development project the medicinal chemist will ask at the end of the day for novel leads attractive to pursue with synthesis and optimization. Impressive enrichment rates of known actives will not be of convincing evidence to him.

In light of these considerations, what is a suitable compound collection? Some general criteria have to be matched either by in-house proprietary compounds or substances offered by commercial suppliers. Compounds can be validated with respect to drug-likeness or lead-likeness considering Lipinski's rule-of-five (or rule-of-three for lead-likeness) (Lipinski et al. 1997; Convreve et al. 2003). Very recently, Yvonne Martin tested this frequently applied rule with respect to experimentally determined rat bioavailability (Martin 2005). Obviously, the rule-of-five has predictive ability for neutral and positively charged molecules; however, anions obey different rules and the polar surface area allows for some predictive power. Similar to the considerations about druggability of binding sites, criteria for bioavailability are multifactorial. John Irwin and Brian Shoichet took the initiative to setup the freely available database ZINC with validated compounds for VS (Irwin and Shoichet 2005). It is built from 2D compound information, generates 3D coordinates and curates, if possible, from stereo- and regioisomeric ambiguities. Multiple states with respect to protonation, charges and tautomers are enumerated. However, as described for proteins, similar conditions might alter the latter properties of candidate molecules upon protein binding. To avoid wasted effort, insoluble, reactive and aggregating compounds have been eliminated. The rule-of-five is a straight-forward filter to discard compounds with putatively undesired properties from the screening sample. Depending on the strategy pursued in the subsequent VS campaign, multiple conformers can be precalculated and stored as separate entries in the database.

Limiting the search sample to purchasable compounds in VS is very pragmative as rapid testing of screening hypotheses can be conceived (Irwin and Shoichet; Baurin et al. 2004). However, VS can also scan over virtual compound libraries, and synthesis can be postponed to later stage, considering only the most prospective hits. The group of Reymond attempted to generate all possible organic molecules up to 160 Da under the constraints of defined chemical stability and synthetic feasibility (Fink et al. 2005). This database contains 13.9 million entries. Possibly such a sample could be used for fragment screening. For larger compounds exhaustive sampling of possible skeletons will end up in a combinatorial explosion. However, proper design criteria, defined by the architecture of the binding site used in VS as target, might guide the generation of target-tailored virtual libraries for VS. In particular, considering the criteria of combinatorial chemistry and parallel synthesis, such VS strategies can help to actually synthesize only the most promissing entries of a large virtual combinatorial library.

4. TOOLS AND STRATEGIES FOR A VIRTUAL SCREENING CAMPAIGN

Principally two strategies can be followed in a VS campaign: forward or backward filtering of hits obtained by docking. The most crucial step in VS is the docking of candidate molecules to the target protein (Shoichet et al. 2002; Muegge et al. 2001; Abagyan and Totrov 2001). In forward filtering, various criteria are used to reduce the initial data sample, which might comprise of several millions of test compounds, to the several hundred or thousand most promising candidates to be docked (Grüneberg et al. 2002). In backward filtering, all entries from the candidate sample are docked to the target protein and filter criteria are subsequently applied to rank the generated docking solutions. Nowadays the speed of computers is no longer the limiting factor in selecting the strategy, even though the forward technique requires less computational resources. This is mainly due to the fact that at each hierarchical filtering step, a significant amount of the original data sample is discarded, however, with decreasing amount of compounds; the filtering becomes computationally increasingly demanding and sophisticated. Usually flexible docking is the computationally most intensive step of all, thus the fewer candidates to be considered here the more effort can be spent in controlling, validating and assessing docking results. Forward filtering eliminates compounds initially according to simple descriptors such as molecular weight, number of rotatable bonds, log P, or crude shape descriptors, such as the ellipticity of the overall structure. Subsequently, information about the receptor's binding site is exploited. Once a hot-spot analysis of the most-likely anchoring positions in the binding pocket has been performed a protein-based pharmacophore can be derived (Brenk et al. 2003). This sets the constraints for the minimal requirement of functional groups to be matched by putative ligands (e.g., number of H-bond donors, acceptors or hydrophobic groups). Molecules satisfying such criteria can be retrieved by any database engine capable for functional group substructure search. Once the topographical arrangement of the protein-based pharmacophore is incorporated into the search and the remaining candidates are requested to match this pattern the study can be further focused. UNITY (UNITY Chemical Information Software, version 4.1, Tripos, Inc.: St. Louis, MO) and CATALYST (Moebius et al. 2001) are prototypes supporting this screening step. An alternative to these tools is FeatureTrees, which can retrieve molecules of similar topology in feature space (Rarey and Dixon 1998). In this context "features" are considered as similar types of functional groups or molecular building blocks. Similarity can be considered as a further filter criterion in VS. Matching with the protein-based pharmacophore already regards similarity criteria in very generic terms. Nevertheless, also

information about known binders can be used for filtering, even though the danger exists that the search could be significantly biased toward already known chemistry. Finally, docking is pursued, usually considering only 1-10% of the initial sample collection. An advantage of this forward filtering approach is that it allows more elaborate docking protocols, e.g., taking multiple protein conformational states into account or reflecting a protein-based pharmacophore as restraint in docking can be performed. Most importantly, this hierarchical filtering strategy allows one to track the performance at the various filter levels by human intervention, and particularly visual inspection of docking solutions remains feasible. Nevertheless, the danger exists that due to biased filters and preconceived concepts some unexpected and novel chemistry could be discarded at early filtering steps.

Backwards filtering starts with high-throughput docking and analyses the generated docking modes as subsequent step. Especially challenging, since docking returns multiple solutions for most candidate molecules, strongly discriminative and reliable scoring functions must be available to analyze the computed results in fairly automated fashion, as visual inspection for this many diverse compounds is hardly feasible. However, it is in question whether the presently existing scoring functions are reliable enough to succeed with such heroic demands (s. section 5). Nevertheless, docking solutions can be filtered with respect to achieved matching of the protein-based pharmacophore, contact complementarity of protein and ligand surfaces or remaining residual unoccupied voids along the protein-ligand interface. Most probably nature avoids voids in molecular assemblies, accordingly the latter criterion could be a powerful indicator for irrelevant binding modes or the putative accommodation of interstitial water molecules.

Docking is the crucial step in VS. The seminal program DOCK, originally described in 1982 by Kuntz and coworkers (Kuntz et al. 1982), has evolved as first virtual screening tool. Later other programs have been successfully applied in VS, such as GOLD (Jones et al. 1997), FlexX (Rarey et al. 1996), Glide (Friesner et al. 2004; Morris et al. 2004) or Autodock (Morris et al. 1998), to name only the most popular prototypes. They have been recently reviewed (Abagyan and Totrov 2001; Halperin et al. 2002; Brooijmans and Kuntz 2003; Cummings et al. 2005; Schulz-Gasch and Stahl 2003; Kellenberger et al. 2004; Perola et al. 2004). All follow slightly different concepts, thus allowing to better incorporate aspects such as flexibility of ligand and receptor or restraining the docking search engine to particular regions in configuration space (e.g. mapping a protein-based pharmacophore) (Joseph-McCarthy et al. 2003). An even more challenging, but carelessly disregarded aspect in docking is the appropriate consideration of water molecules. As indicated, water molecules are in about 2/3 of the known protein-ligand complexes involved in binding. However, most

docking tools ignore them simply because conclusive concepts of how to consider them correctly are missing. If structural evidence is given, pre-placement of water molecules in a docking run is a feasible strategy (Günther et al. 2003). The docking tool Slide (Schnecke and Kuhn 1999) treats pre-placed water molecules in a way that still allows their replacement by ligand atoms in docking. The particle concept in FlexX (Rarey et al. 1999) permits the placement of waters on the fly in due course of the generation of individual docking solutions. In the docking tool GOLD water molecules can be switched on or off and they can spin around their principle axes to achive good contacts with a docked ligand (Verdonk et al. 2005). Recently, the popular docking tools DOCK and FlexX have been equipped by features that allow docking on a preconceived pharmacophore (Hindle et al. 2002; Fradera et al. 2002). These concepts drive docking solutions particularly into regions either frequently trapped by other bound ligands or featured by complementary methods as particularly relevant for binding. The program AutoDock (Morris et al. 1998) performs docking on a pre-calculated grid storing potential values from any sort of interaction field (Österberg et al. 2002). In the original implementation Lennard-Jones and Coulomb potential values are used. Sotriffer et al. replaced them by knowledge-based potentials originally implemented into DrugScore (Sotriffer et al. 2002). They have been proven powerful to discriminate and rank among multiple ligand poses. The potential grid approach in AutoDock also allows to average across the fields produced by various protein frames. This allows regarding conformational degrees of freedom of a protein. Also adapted fields optimized with respect to the binding properties of some known actives in the comparative molecular field-type approach AFMOC (Gohlke and Klebe 2002) can be used as target potential values in AutoDock (Radestock et al. 2005). The latter docking tool performs multiple stochastic searches on the potential hypersurface, accordingly the frequency of occurrence of certain docking solutions can be used as an additional figure-of-merit for their relevance (Sotriffer et al. 1996).

The issue of an optimal ligand size for screening has been addressed in section 3. The complexity of the docking problem increases with the size of the ligand and its number of rotatable bonds. Thus, smaller molecules in the typical range of fragments should be simpler to dock. Perplexingly, present experience indicates the opposite to be the case. Fragments are easily scattered all over a binding site by docking, only in case that the binding site itself is rather restricted in size, reliably successful docking can been reported (Wei et al. 2001).

VS runs are usually monitored and validated by comparing the performance of a set of known actives with a large number of "randomly" picked compounds. The actives are pooled with the random entries. All

compounds are submitted to the selected VS protocol and the performance ranks of the known actives with respect to the remaining pool are converted into enrichment plots. Such doing is essential to keep control over the achievements of VS. However, the choice of the random compounds libraries is crucial and can strongly affect the enrichments obtained (Verdonk et al. 2004). Accordingly, it has to be examined whether the pooled randomly picked decoy structures are actually non-binders. In a real life scenario, one has to note that merging known actives with a set of candidate ligands will hopefully result in a gradually declining enrichment rate at the subsequent hierarchical filter steps, since novel actives, retrieved by VS, will populate on prominent ranks and dilute the set of predefined known actives.

5. SCORING, RANKING AND VALIDATION OF VS DOCKING RESULTS

Independent of the actual VS strategy applied, docking and subsequent scoring of the suggested solutions is the key performance-determining factor in VS. The discriminative power of the applied scoring function is of utmost importance to rank and hopefully enrich potentially active binders at the top of the list of docking solutions. In consequence, a myriad of scoring functions has been developed over the past years (Stahl and Rarey 2001; Wang et al. 2003; Ferrara et al. 2004). Approaches have been taken not simply to rely on one single function but to consult the consensus picture of several scoring schemes (Charifson et al. 1999; Wang and Wang 2001).

Experimentally, the binding constant or its inverse, the dissociation constant, is determined (or approximate values such as IC50s) to characterize binding of putative lead candidates. Assuming equilibrium thermodynamics, this entity corresponds to a Gibbs free energy, composed by an enthalpic and entropic contribution. Whereas the former relates to energetic features, the latter regards configurational and ordering phenomena. It estimates how the energy content is distributed over internal and external molecular degrees of freedom.

Up to now, three approaches have been followed to predict binding affinities on the basis of a given protein-ligand binding geometry. The most rigorous and theoretically most solid approaches are first principle methods (Gohlke and Klebe 2002). Using quantum mechanics or computationally less demanding, however approximate force-fields, the partition function of a system is computed and free energy differences between bound and unbound state are determined. With increasing speed of computers such methods obtain growing relevance for scoring (Raha and Merz 2005). However, screening large samples of docked solutions is still far beyond tractability.

Nevertheless, first principle methods do not need any calibration or training at experimentally determined affinity data, thus they will not suffer from inherent experimental short-comings or accuracy limits. This is quite different for the other two approaches based on empirical concepts, the regression- and knowledge-based scoring functions (Gohlke and Klebe 2002).

Regression-based approaches assume additivity of individual terms considered in a master equation to describe the total Gibbs free energy of binding. In this context a "term" can reflect any physicochemical property of relevance for the protein-ligand binding process, e.g., number of charged or uncharged H-bonds, size of polar or nonpolar surface portion, number of rotatable bonds or desolvation enthalpy, just to mention some. They are assumed to be independent from each other and their individual contribution in reproducing known affinities of some training set ligands is extracted by regression analysis, PLS, or neural networks (Gohlke and Klebe 2002). Independence of the terms is hardly likely and fair to strong correlations among terms is probable. This can result in the fact that in the analysis another, on the first glance surprising property is selected that is perhaps not the explaining one in a physical sense, however it is highly correlated with the one that actually matters. We recently parameterized a new regression-based scoring function on the basis of a large sample of crystal structures with known affinities for the bound ligands. Depending on the composition of the training set, different terms turn-up as relevant in the analysis. This clearly shows that such empirical scoring functions reflect a best fit with respect to the training set used, however they hardly achieve generality.

An alternative to regression-based approaches are knowledge-based scoring functions (Gohlke and Klebe 2002). They evaluate the occurrence frequencies of some properties of interest, e.g., the mutual distance between particular atom types found across the protein-ligand interface. The sample distributions are compared to a statistical mean reference situation and deviations from the average are translated by some functional form into occurrence probabilities. The derived propensities describe statistical preferences that can be used to score the geometry of protein-ligand complexes. To estimate binding constants, either the regression- and knowledge-based scoring functions require experimental affinity data for internal calibration. Thus, they can never achieve prediction accuracy better than the precision by which binding data can be measured. It is interesting to note that the estimated standard deviations of such empirical scoring functions are reduced if data for a selected number of targets, determined in one lab based on assay data recorded under strictly conserved conditions are used or if broad-range data covering many targets are considered based on assay data determined in various labs. The determination of relative differences in binding data within a series of compounds can be achieved quite precise usually much better

than the accuracy across the data for various systems where the comparison has to be performed on absolute scale. In consequence, the standard deviations of presently available functions range between 0.7 and 1.5 logarithmic units in binding affinity. This range matches with the experimental accuracy achieved for binding data across training sets of growing heterogeneity covering a broad spectrum of targets. In our experience, knowledge-based scoring functions are better able to extract binding poses closely approximating the experimentally confirmed binding mode from a sample set of decoy placements, whereas regression-based scoring functions are better in the actual affinity prediction, provided a fairly accurate binding geometry is given.

In contrast to the professed opinion that in docking the geometry problem has been resolved to a sufficient extent and the scoring problem remains as open question (Dixon 1997; Owens 2003), it appears that both are intimately related. We recently developed a knowledge-based scoring function based on more accurate contact data from small molecule crystal data (Velec et al. 2005). The latter function reliably recognizes the binding mode found in a crystal structure among a set of decoy poses. This suggests that the scoring problem can only be increasingly alleviated if more relevant, near-native binding poses are produced by docking programs. Accordingly, it appears most advisable to drive docking solutions as close as possible to the native geometries, e.g., as they would show-up in a corresponding crystal structure by minimizing them with respect to the function used for scoring. As a disadvantage such doing will be computationally rather demanding.

Discussing scoring and ranking, the question must be allowed to how well do we understand the target value Gibbs free energy? Is it advisable to focus scoring on free energy or better on enthalpy and entropy separately (Luque and Freire 2002)? A mutual compensation of enthalpy and entropy retains simply due to the fact that both entities scatter over much larger ranges than the free energy itself. Affinity differences coincide for ligand binding into a rather small window. The mutual compensation of enthalpy and entropy can be effective already across closely related molecules (Gerlach et al. 2005). Furthermore, many physicochemical phenomena of relevance for the binding process are not yet fully understood and accordingly not correctly incorporated into scoring functions, e.g., the role of water, change in protonation states, or an appropriate consideration of entropy. Interestingly, microcalorimetry indicates that with increasing temperature the protein-ligand binding process becomes more exothermic and entropically less favourable (Gohlke and Klebe 2002; Reyda et al. 2003). This observation points to some general phenomena not yet understood on a molecular level, as it applies in general to all targets. Despite of our present deficiencies in understanding of the physics of the binding process, scoring

still works satisfactorily, most likely because we consider the binding of ligands to a protein on a relative scale to each other. Accordingly, any unappreciated phenomena, similar across all complexes in the analysis, will simply cancel out. As mentioned, one concept to consider protein flexibility applies parallel docking into several binding-competent conformers of the protein. Disadvantage of this strategy is the fact that additional degrees of scoring are created: what discriminates the scoring against different protein conformers? Cancellation of unreflected internal protein energy contributions is no longer given. Studies have shown that a special scoring is required (Wei et al. 2004; Barril and Morley 2005). Since dramatic energy differences between low-energy conformers are unlikely, a modulated pocket size experienced by the different protein conformers possibly requires individual scoring of the altered desolvation properties of the binding site.

VS is used to enrich putative actives from a large sample set of test ligands. Accordingly, the desired accuracy of scoring depends whether at first glance prioritizing of the sample set for testing is anticipated or whether putative actives are expected among the top 10 or 100 of a hit list. The latter requires very powerful discrimination of actives over inactive decoy binders. Present scoring functions have been optimized to discriminate for a particular ligand decoy binding mode from near-native ones. The discrimination of binders from decoy non-binders still remains as major challenge for present VS protocols (Graves et al. 2005). Perhaps the consideration of similarity criteria in the search (Grüneberg et al. 2002; Hindle et al. 2002; Fradera et al. 2002; Sotriffer et al. 2002; Gohlke and Klebe 2002; Radestock et al. 2005) allows one to alleviate the problem to some degree since it drives the search toward closer related ligands for which some of the disregarded effects in scoring possibly cancel out.

Only in very rare cases the crystal structure of detected VS hits has been determined subsequently. Such case studies have recently been reviewed by Shoichet (Shoichet 2004). Cases have been described for which VS has correctly predicted the later found binding mode (Brenk et al. 2004). Other examples point to deficiencies arising from the superficially understood phenomena described above. Finally, also cases have been reported where VS suggested an active compound, however for the wrong reasons because binding actually occurs in a totally different region of the protein surface.

6. CONCLUSIONS

VS has been established as a powerful alternative and complement to HTS. Thoroughly performed, quite impressive hit rates have been reported, significantly higher than those for HTS. Comparative studies of HTS and VS indicate that the methods can capture alternative and complementary ligands.

This review tried to analyze the limitations of the VS in a frank and unvarnished way and sought for explanations why we still face them. Undoubtedly, VS is not yet a fully mature technology following a well-established process line. Few of the foundations of protein-ligand recognition are understood well enough to be deployed in a large scale, multi-compound effort such as commonly undertaken in virtual screening. This calls upon further in-depth research. Particularly, protein flexibility and induced-fit adaptations, the role of water in solvation/desolvation and ligand binding and the electrostatics including changes in protonation states are the major short-comings in understanding. Frequently, experimental work performed in parallel or as follow-up to a VS campaign provides a whole bunch of unexpected results pointing toward manifold deficiencies of the concepts applied. Nevertheless, VS has proven successful and as a valuable alternative, in particular if it is used as a tool to support and complement hit discovery. For example, it is interesting to note that an experienced modeller or medicinal chemist can often figure out whether a particular binding pocket appears druggable or a certain molecule obeys the rules of drug-likeness, but putting such knowledge into computer algorithms makes the multifactorial nature of these rules apparent and this complicates to generalize them. The same holds for docking and scoring. It also explains why at the present level of understanding it is highly advisable to refrain from fully automated strategies in VS. Experience and human intervention is of utmost importance to keep control over the various filter steps in an VS run. Such applied, VS can be very successful; in particular if information about molecular similarity is considered in terms of generic physicochemical properties and not simply as chemical formulae. Considering similarity concepts takes the risk that highly diverse molecules remain undetected, however, most likely it makes the searches simpler because many parameters that actually matter in VS and which are not properly considered simply cancel out in a relative comparison. Nevertheless, it is still a long way to go until it is an established tool to routinely suggest innovative leads.

REFERENCES

Abagyan, R., Totrov, M. (2001) High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**: 375-382.

Barril, X., Morley, S.D. (2005) Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **48**: 4432-4443.

Barril, X., Hubbard, R.E., Morley, S.D. (2004) Virtual Screening in Structure-Based Drug Design. *Mini-Rev. Med. Chem.* **4**: 779-791.

Baurin, N., Baker, R., Richardson, C., Chen, I., Foloppe, N., Potter, A., Jordan, A., Roughley, S., Parratt, M., Greaney, P., Morley, D., Hubbard, R.E. (2004) Drug-like annotation and

duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J. Chem. Inf. Comput. Sci.* **44**: 643-651.

Bissantz, C., Bernard, P., Hibert, M., Rognan, D. (2003) Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **50**: 5-25.

Bohacek, R.S., McMartin, C., Guida, W.C. (1996) The art and practice of structure-based drug design: a molecular modelling perspective. *Med. Res. Rev*. **16**: 3-50.

Bolten, B.M., DeGregorio, T. (2002) Trends in development cycles. *Nat. Rev. Drug Discov.* **1**: 335-336.

Brenk, R., Meyer, E., Reuter, K., Stubbs, M.T., Garcia, G.A., Diederich, F., Klebe, G. (2004) Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening *J. Mol. Biol.* **338**: 55-75.

Brenk, R., Naerum, L., Gradler, U., Gerber, H.D., Garcia, G.A., Reuter, K., Stubbs, M.T., Klebe, G. (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **46**: 1133-1143.

Brooijmans, N., Kuntz, I.D. (2003) Molecular recognition and docking algorithms. Annu. *Rev. Biophys. Biomol. Struct.* **32**: 335-373.

Carlson, H.A., McCammon, J.A. (2000) Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **57**: 213-218.

Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P. (1999) Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **42**: 5100-5109.

Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P. Morley, S.D., Afshar, M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock. *J. Comput.-Aided Mol. Des.* **18**: 189-208.

Claussen, H., Buning, C., Rarey, M., Lengauer, T. (2001) FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308**: 377-395.

Convreve, M. et al. (2003) A rule-of-three for fragment-based lead discovery. *Drug Discov. Today* **8**: 876-877.

Cummings, M.D., DesJarlais R.L., Gibbs, A.C., Mohan, V., Jaeger E.P. (2005) Comparison of Automated Docking Programs as Virtual Screening Tools, *J. Med. Chem.* **48**: 962-976.

Czodrowski, P., Dramburg, I., Sotriffer, C.A., Klebe, G. (2006) Development of a generally applicable charge model for the calculation of protonation states in proteins and protein-ligand complexes. *Proteins, Structure, Functions and Bioinformatics* **65**: 424-437.

Czodrowski, P., Sotriffer, C.A., Klebe, G. (2006) Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments. *J. Mol. Biol. in press.*

DesJarlais, R.L., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., Ortiz de Montellano, P.R., DeCamp, D.L., Babé, L.M., Craik, C.S. (1990) Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA* **87**: 6644-6648.

DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. (1986) Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* **29**: 2149-2153.

Dixon, J.S. (1997) Evaluation of the CASP2 Docking Section. *Proteins*, *Suppl.* 198-204.

Erlanson, D.A. et al. (2004) Fragment-based drug discovery. *J. Med. Chem*. **47**: 3465-3482.

Evers, A., Gohlke, H., Klebe, G. (2003) Ligand-supported homology modelling of protein binding sites using knowledge-based potentials. *J. Mol. Biol.* **334**: 327-345.

Evers, A., Klabunde T. (2005) Structure-based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the Alpha1A Adrenergic Receptor. *J. Med. Chem.* **48***:* 1088-1097.

Evers, A., Klebe, G. (2004**)** Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J. Med. Chem.* **47**: 5381-5392.

Ferrara, P., Gohlke, H., Price, D.J., Klebe, G., Brooks, C.L., III (2004) Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **47**: 3032-3047.

Filikov, A.V., Mohan, V., Vickers, T.A., Griffey, R.H., Cook, P.D., Abagyan, R.A., James T.L. (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J Comput.-Aided Mol. Design* **14**: 593-610.

Fink, T., Bruggesser, H., Reymond, J.L. (2005) Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons, *Angew. Chem. Int. Ed.* **44**: 1504-1508.

Fiser, A., Sali, A. (2003) Comparative protein structure modelling. In: Chasman D., ed., *Protein Structure*, Marcel Dekker, Inc. New York, pp. 167-206.

Fishman, M.C., Porter, J.A. (2005) A new grammar for drug discovery. *Nature* **437**: 491-493.

Fradera, X., Knegtel, R.M., Mestres, J. (2002) Similarity-driven flexible ligand docking. *Proteins* **40**: 623-636.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J. et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**: 1739-1749.

Gerlach, C., Velec, H., Smolinski, M., Hangauer, D., Heine A., Klebe, G. (2005) Library design based on privileged scaffolds through docking and direct design in the protein binding pocket. Abstracts of Papers, 230th ACS National Meeting, Washington DC, United States, Aug. 28- Sept. 1, COMP-185.

Gohlke, H., Hendlich, M., Klebe, G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295**: 337-356.

Gohlke, H., Hendlich, M., Klebe, G. (2000) Predicting binding modes, binding affinities and "hot spots" for protein-ligand complexes using a knowledge-based scoring function. *Perspect. Drug Discovery Des.* **20**: 115-144.

Gohlke, H., Klebe G. (2002) Approaches to the Description and Prediction of Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chemie, Int. Ed. Engl.* **41**: 2644-2676.

Gohlke, H., Klebe, G. (2002) DrugScore meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* **45**: 4153-4170.

Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**: 849-857.

Graves, A.P., Brenk, R., Shoichet, B.K. (2005) Decoys for Docking. *J. Med. Chem.* **48**: 3714-3728.

Grüneberg, S., Stubbs, M.T., Klebe, G. (2002) Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* **45**: 3588-3602.

Günther, J., Bergner, A., Hendlich, M., Klebe, G. (2003) Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.* **326**: 621-636.

Hajduk, P.J., Huth, J.R., Fesik, S.W. (2005) Druggability Indices for Protein Targets Derived from NMR-Based Screening Data, *J. Med. Chem.* **48***:* 2518-2525.

Halperin, I., Ma, B., Wolfson, H., Nussinov, R. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* **47**: 409-443.

Hendlich, M., Bergner, A., Günther, J., Klebe, G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol***, 326**: 607-620.

Hillisch, A., Pineda, L.F., Hilgenfeld, R. (2004) Utility of homology models in the drug discovery process. *Drug Discovery Today* **9**: 659-669.

Hindle, S.A., Rarey, M., Buning, C., Lengauer, T. (2002) Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.* **16**: 129-149.

Hopkins, A.L., Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* **1**: 727-730.

Irwin, J.J. and Shoichet, B.K. (2005) ZINC - a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**: 177-182.

Jain, A.N. (2004) Virtual screening in lead discovery and optimization**.** *Curr. Opin. Drug Discov. Devel.* **7**: 396-403.

Jelesarov, I., Bosshard, H.R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **2**: 3-18.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**: 727-748.

Joseph-McCarthy, D., Thomas, B.E.T., Belmarsh, M., Moustakas, D., Alvarez, J.C. (2003) Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* **51**: 172-188.

Kellenberger, E., Rodrigo, J., Muller, P., Rognan, D. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**: 225-242.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev. Drug Discov.* **3**: 935-949.

Knegtel, R.M., Kuntz, I.D., Oshiro, C.M. (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **266**: 424-440.

Kubinyi, H. (2006) Success Stories of Computer-Aided Design. In: Ekins, S., ed., *Computer Applications in Pharmaceutical Research and Development*, in press.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. (1982) A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**: 269-288.

Lahana, R. (1999) How many leads from HTS? *Drug Discovery Today* **4**: 447-448.

Lam, P.Y., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y. et al. (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**: 380-384.

Langer, T., Krovat, E.M. (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discov. Devel.* **6**: 370-376.

Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility (1994) *J. Mol. Biol.* **235**: 345-356.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**: 3-25.

Luque, I., Freire, E. (2002) Structural parameterization of the binding enthalpy of small ligands. *Proteins* **49**: 181-190.

Lyne, P.D. (2002) Structure-based virtual screening: an overview. *Drug Discovery Today* **7**: 1047-1055.

Martin, Y.C., (2005**)** A Bioavailability Score, *J. Med. Chem.* **48***:* 3164-3170.

McGovern, S.L., Shoichet, B.K. (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **46**: 2895-2907.

Moebius, F.F., Bermoser, K., Reiter, R.J., Hanner, M., Glossmann, Kurogi, Y., Guner, O.F. (2001) Pharmacophore modeling and threedimensional database searching for drug design using Catalyst. *Curr. Med. Chem.* **8**: 1035-1055.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **19**: 1639-1662.

Morris, G.M., Goosell, D.S., Huey, R., Hart, W.E., Belew, R., Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., et al. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**: 1739-1749.

Muegge, I., Rarey, M. Small Molecule Docking and Scoring. (2001) *Reviews in Computational Chemistry*, Wiley-VCH, New York, pp. 1-60.

Oprea, T.I. (2002) Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput.-Aided Mol. Design.* **16**: 325-334.

Österberg, F., Morris, G.M., Sanner, M.F., Olson, A.J., Goodsell, D.S. (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **46**: 34-40.

Owens, J. (2003) Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discov. Today* **8**: 12-16.

Perola, E., Walters, W.P., Charifson, P.S. (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56**: 235-249.

Radestock, S., Böhm, M., Gohlke H. (2005) Improving Binding Mode Predictions by Docking into Protein-Specifically Adapted Potential Fields, *J. Med. Chem.* **48***:* 5466-5479.

Raha, K., Merz K.M., Jr. (2005) Large-Scale Validation of a Quantum Mechanics Based Scoring Function: Predicting the Binding Affinity and the Binding Mode of a Diverse Set of Protein-Ligand Complexes, *J. Med. Chem.* **48***:* 4558-4575.

Ramesha, C.S. (2000) Comment: How many leads from HTS? *Drug Discovery Today* **5**: 43-44.

Rarey, M., Dixon, J.S. (1998) *J. Comput. Aided Mol. Des*. **12**: 471-90.

Rarey, M., Kramer, B., Lengauer, T. (1999) The particle concept: placing discrete water molecules during protein–ligand docking predictions. Proteins: *Struct. Funct. Genet.* **34**: 17-28.

Rarey, M., Kramer, B., Lengauer, T., Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**: 470-489.

Rauh D., Klebe G., Stubbs M.T. (2003) Understanding protein-ligand interactions: The price of protein flexibility *J. Mol. Biol.* **335**: 1325-1341.

Rauh, D., Klebe G., Stürzebecher J., Stubbs M.T. (2003) ZZ Made EZ: Influence of Inhibitor Configuration on Enzyme Selectivity. *J. Mol. Biol.* **330**: 761-770.

Rees, C.D., Congreve, M., Murray, C.W., Carr, R. (2004) Fragment-based lead discovery, *Nat. Reviews Drug Discov.* **3**: 660-672.

Reyda, S., Sohn, C., Klebe, G., Rall, K., Ullmann, D., Jakubke, H.D., Stubbs, M.T. (2003) Reconstructing the binding site of factor Xa in trypsin reveals ligand-induced structural plasticity. *J. Mol. Biol.* **325**: 963-977.

Rutenber, E, Fauman, E.B., Keenan, R.J., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., DeCamp, D.L., Salto, R., Rosé, J.R., Craik, C.S., Stroud, R.M. (1993) Structure of a nonpeptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *J. Biol. Chem.* **268**: 15343-15346.

Schafferhans, A., Klebe, G. (2001) Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **307**: 407-427.

Schnecke, V., Kuhn, L.A. (1999) Virtual screening with solvation and ligand-induced complementarity. *Persp. Drug Discov. Des.* **20**: 171-190.

Schulz-Gasch, T., Stahl, M. (2003) Binding site characteristics in structure-based vistual screening: evaluation of current docking tools. *J. Mol. Model.* **9**: 47-57.

Shoichet, B.K., McGovern, S.L., Wei, B., Irwin, J.J. (2002) Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **6**: 439-446.

Shoichet, B.K. (2004) Virtual screening of chemical libraries. *Nature* **432**: 862-865.

Sotriffer, C.A., Krämer, O., Klebe, G. (2004) Probing Flexibility and "Induced-Fit" Phenomena in Aldose Reductase by Comparative Crystal Structure Analysis and Molecular Dynamics Simulations, *Proteins* **56***:*52–66.

Sotriffer, C.A., Gohlke, H., Klebe, G. (2002) Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J Med Chem.* **45**: 1967-1970.

Sotriffer, C.A., Winger, R.H., Liedl, K.R., Rode, B.M., Varga, J.M. (1996) Comparative docking studies on ligand binding to the multispecific antibodies IgE-La2 and IgE-Lb4. *J Comput.- Aided Mol. Des.* **10**: 305-320.

Stahl, M., Rarey, M. (2001) Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.,* **44**: 1035-1042.

Steuber, H., Czodrowski, P., Sotriffer, C.A., Klebe, G. (2006a) Assignment of Superimposed Proton Steps as a Prerequisite to Factorize Thermodynamic Inhibitor Binding Data to Aldose Reductase, *J. Mol. Biol. submitted.*

Steuber, H., Zentgraf M., Gerlach C., Sotriffer C.A., Heine A., Klebe G. (2006b) Expect the Unexpected or Caveat for Drug Designers: Multiple Structure Determinations Using Aldose Reductase Crystals Treated under varying Conditions. *J. Mol. Biol. submitted.*

Stubbs, M.T., Reyda, S., Dullweber, F., Moller, M., Klebe, G., Dorsch, D., Mederski, W.W., Wurziger, H. (2002) pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *Chembiochem* **3**: 246-249.

Teague, S.J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2**: 527-541.

Velec, H.F.G, Gohlke, H., Klebe, G. (2005) DrugScoreCSDsKnowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction, *J. Med. Chem.,* **48***:* 6296-6303.

Verdonk, M.L., Berdini, V., Hartshorn, M.J., Mooij, W.T., Murray, C.W. et al. (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **44**: 793-806.

Verdonk, M.L., Cole, J.C., Taylor, R. (1999) SuperStar: A knowledge based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **289**: 1093-1108.

Verdonk, M.L., Chessari, G., Cole J.C., Hartshorn ,M.J., Murray, C.W., Nissink, J.W.M., Taylor, R.D., Taylor, R. (2005) Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem*. **48**: 6504-6515.

Wallner, B., Elofsson, A. (2005) All are not equal: a benchmark of different homology modelling programs. *Protein Sci.* **14**: 1315-1327.

Wang, R., Lu, Y., Wang, S. (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **46**: 2287-2303.

Wang, R., Wang, S. (2001) How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **41**: 1422-1426.

Wei, B.Q., Weaver, L.H., Ferrari, A.M., Matthews, B.W., Shoichet, B.K. (2004) Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **337**: 1161-1182.

From Solid State to Bio-Complexity: On the Emerging Science of Emergence

DAVOR PAVUNA
Institute of Physics of Complex Matter, Faculty of Basic Sciences, Ecole Polytechnique Federale de Lausanne (EPFL), Station 3, CH-1015 Lausanne, Switzerland,
Phone: +41 21 6933301, Fax: +41 21 693 4666, E-mail: davor.pavuna@epfl.ch

Abstract

The challenge of understanding bio-complexity can hardly be addressed within the 'usual' reductionism school of thought. We approach contemporary bio-challenges by adopting the view 'more is different', hence the *science of emergence*. Emergence was introduced by biologists, as it has a natural place within bio-complexity and yet is an integral part of the contemporary physics. Emergence is the process of complex pattern formation from simpler rules. Emergent structures are patterns not created by a single event or rule. For a phenomenon to be termed emergent it should generally be unpredictable from a lower level description. There is nothing that commands the system to form a pattern, but instead the interactions of each part to its immediate surroundings causes a complex process which leads to order. Most of our present science is emergent and many of the 'archetype' models and techniques of physics are successfully adopted to the bio-context. We also give a summary of some of our ongoing projects on scanning near-field optical microscopy applied to the fluorescent imaging of neuronal cells and the expression of tagged glutamate receptors in organotypic slice cultures from rat brain.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 273–285

Keywords: Physics, Biology, Emergence, Reductionism, SNOM, Bio-imaging.

1. INTRODUCTION

My approach to the main themes of this summer school could be best described as an attitude of a pragmatic, experimental physicist. First of all, the discipline nowadays called physics has emerged from the Greek notion of *Physis* = Nature, hence one can treat all bio-related topics as a part of what used to be quite correctly called *natural philosophy*, i.e. physics. Actually, bio-complexity and biophysics are very loosely used notions and concern disciplines that welcome anybody who can constructively contribute (Carry 2003). Physicists have certainly done their fair share in that respect. Moreover, since the beginning of the 3rd Millennium, we are actively involved in staging a regular, albeit somewhat iconoclast, conference series 'From Solid State to Biophysics' (Forró and Pavuna 2006). These meetings bring together various distinguished scientists and cover the 'full range' of topics, from basic condensed matter physics to biology or genomics and even actual medical treatments. And for all those who may still have some doubts that physics encompasses biology and life, I refer to recent articles that clearly show that even femto-second quantum phenomena are relevant to biology and life (Chergui 2006)!

Like this school, our regular conferences prove in practice that the contemporary 'biophysics' indeed accommodates *every* constructive contribution. In these conferences we learned that, as thoroughly trained physicists, we can quickly figure out most of the essential arguments in contemporary bio-challenges. Namely, so far the human mind has produced a finite number of archetype math-based models, so many of the key models already exist in physics. Therefore, once you learn the language of the actual problem at hand you are most of the time well placed to figure out the most sensible route to the qualitative understanding. Very rarely do we encounter some stunning scientific breakthroughs, equivalent to Newton's insights into the Physical Law (as that is also a rare event in contemporary physics too). Consequently, in western Switzerland we have a joint sub-committee for condensed matter *and* biophysics and we all have often an easier task to understand the contemporary lectures in (say) cell motility than the advanced theories of correlated electron systems. The former use mainly relatively simple language of classical statistical mechanics, albeit applied to horrendously complex bio-samples, while the latter use some 'wild' quantum physics of correlated electrons in 'simpler' forms (samples) of matter. Hence, it seems evident (at least to me) that the optimal training for bio-complexity

might be a thorough undergraduate training in physics with some advanced postgraduate studies in one of the stimulating bio-topics, as discussed at length by other authors in this volume.

To align my contribution with a spirit of our age, in the next section I argue that the emerging *science of emergence* is the appropriate approach useful for us all (biologists and physicists) dealing with complexity, living at the very end of the era of reductionism. I then proceed to argue that many physicists actually do have a different view of life and nature than most of the contemporary biologists. To me, and many physicists all Universe reflects Life and it undoubtedly contains plenty of (presently undiscovered) forms of matter and life. The usual separation to inanimate and living matter seems unnecessary and rather artificial. Lastly, in section 5, I discuss two of our ongoing biophysics experiments and conclude with some brief remarks on the bright future of our science.

2. THE CHALLENGE OF UNDERSTANDING THE BIO-COMPLEXITY: 'STRUCTURE IS FUNCTION'

If in the early 16th century you were able to watch the night sky from this lovely island (Crveni Otok, Rovinj) you would have seen essentially the same sky as today. It would seem rather complex, even if we could establish some relationships concerning the movements of Planets or even galaxies. It was Isaac Newton who has formulated a well known Physics Law of Gravity that in a way represents the 'old science dream': it's a reductionist's triumph in that a simple formula accurately describes 'the clock of the Universe'. His formula was rigorously verified experimentally and only when it was confirmed by numerous, very thorough *experiments* (with a precision appropriate to the epoch of the experiment), did it become a generally accepted Physical Law. Moreover, it worked so well that from the 18th century onwards all experiments were judged by the degree of compatibility with the Law. Today we know most of the limitations of Newton's laws of mechanics, yet nevertheless it illustrates the power of such a simple, yet powerful mathematical statement and how it is useful as an archetype model of general physics and all of pre-quantum, pre-emergence science.

Obviously, nowadays we would all rejoice if bio-chemists' mantra, 'structure is function', could be summarized by some contemporary 'Newton' into an equivalent BioPhysics Law and, for example, protein folding be simply described by some appropriate, however complex, formula (see also

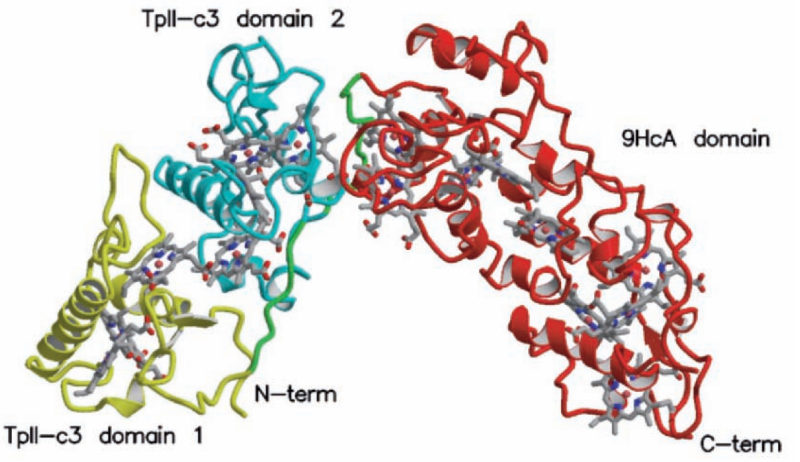


Figure 2.1. While Newton's Law describes remarkably well apparently complex movements of Planets in the night sky, it is rather unlikely that just one, even complex formula, will describe the function of relatively simple macromolecules, such as illustrated here.

Fig. 2.1). However, given effectively an infinite complexity of relevant, dynamic bio-structures, and their associated functions, such a possibility at present seems highly unlikely. Therefore, in what follows, we shall assume that unlike 'simple classical physics', the actual challenge of understanding the bio-complexity can hardly be addressed within reductionism school of thought ('*In science there is only physics; all the rest is stamp collecting*.') (Kelvin 1900) or with a view that there is the end in sight to theoretical physics (Hawking 1988). Rather, we will full-heartedly adopt the emerging science of *emergence*, a promising subject that has evolved from the discussions at the Santa Fe Institute in the later part of the 20th century (Laughlin and Pines 2000). Actually, the emergence was introduced by bio-logists as it has a natural place within bio-complexity, yet is very important in contemporary physics and chemistry as well (Laughlin 2005).

3. THE END OF REDUCTIONISM: THE EMERGING *SCIENCE OF EMERGENCE*

Horgan (1997) argued that our civilization is facing barriers to the acquisition of knowledge so fundamental that the Golden Age of Science must be thought of as over. It is an instructive and humbling experience to attempt explaining this idea to any child. And I did try the test with my own children: The outcome is always the same. The child eventually stops

listening, smiles politely, and then runs off *to play* and to explore the countless infinities of new possibilities in his or her world. Hence Horgan's insights may more properly be considered the End of Reductionism, for it is actually a warning to all those concerned with the health of physical science to face the truth that in most respects the reductionist ideal has reached its limits as a guiding principle. Rather than a Theory of Everything, we seem to face *a hierarchy of Theories of Things*, each emerging from its parent and evolving into its children as the energy scale is lowered (Laughlin and Pines 2000) or as our experimental resolution is enhanced by some new 'nanoscope'.

Therefore, our contemporary (r)evolution in bio-sciences is best approached by adopting the view '*more is different*' (Anderson 1972), hence the *science of emergence* (Laughlin 2005) and many of the models and techniques developed in physics can be adopted to the new context (Forró and Pavuna 2006). Emergence is the process of complex pattern formation from simpler rules. Emergent structures are patterns not created by a single event or rule. For a phenomenon to be termed emergent it should generally be unpredictable from a lower level description. At the very lowest level, the phenomenon usually does not exist at all or exists only in trace amounts: it is irreducible. There is nothing that commands the system to form a pattern, but instead the interactions of each part to its immediate surroundings causes a complex process, which leads to order. One might conclude that emergent structures are more than the sum of their parts because the emergent order will not arise if the various parts are simply coexisting; the interaction of these parts is central (Stanford Encyclopedia 2006). And physics methods can often help in developing the understanding of reliable bio-technologies (Forró and Pavuna 2006).

As for the biologist, although there are many reductionist tendencies as regards the definition of life (see section 4), the evolution and emergence are part of our daily life. For many physicists, on the other hand, the transition from a reductionist approach may not be easy, yet is, in the long run, much more satisfying. (Actually, nowadays it is often geneticists that think and behave like reductionists!). Living with emergence means, among other things, focusing on what experiment tells us about candidate scenarios for the way a given system might behave before attempting to explore the consequences of any specific model. This contrasts sharply with the imperative of ultimate reductionism, which requires us never to use experiment, as its objective is to construct a deductive path from the ultimate equations to the experiment without cheating. But this is unreasonable when the behavior in question is emergent, for the higher organizing principles, the core physical ideas on which the model is based, would have to be deduced from the underlying equations, and this is, in general, impossible. Repudiation of this

physically unreasonable constraint is the first step down the road to some fundamental discovery (Laughlin 2005). And in many ways this is *the lesson* in contemporary bio-complexity or rather in complexity, as examples are numerous even in solid state physics.

Let me give one example of relevant (non-biological) complexity where the emergence does play a role: it is my familiar ground of macroscopic quantum phenomena (Pavuna 2007). Namely, no problem in physics in our time has received more attention than high-Tc superconductivity in cuprate oxides (discovered in 1986), whose striking properties continue to surprise us all. As the high-Tc community has learned by now, deduction from (any) microscopic theory has not explained, and probably cannot explain as a matter of principle, the wealth of crossover behavior discovered in the normal state of the underdoped compounds, much less the remarkably high superconducting transition temperatures measured at optimal doping of these layered materials. Paradoxically, high-Tc continues to be the most important problem in solid-state physics, and perhaps physics in general, because this very richness of behavior strongly suggests the presence of a new and unprecedented kind of quantum emergence (Laughlin and Pines 2000).

The aforementioned end of reductionism is, however, not the end of science, or even the end of theoretical physics, as claimed by some (Hawking 1988). The scientific questions with non-satisfactory or no answer abound: How do proteins work their wonders? (Fig. 2.1.) Why do magnetic insulators superconduct or why is 3He a superfluid? Why does black hole formation so resemble a quantum phase transition? How is Life distributed across the universe, or even in our galaxy? The list is endless, and it does not include the most important questions of all, namely those raised by (mostly unexpected) discoveries yet to come. And they will come! So, according to Laughlin (2005) the central task of physics in our time is no longer to write down the ultimate equations but rather to catalogue and understand emergent behavior in its many guises, including potentially life itself. Although there are several scientifically different approaches to this problem (Thorpe and Phillips 2001), that's indeed the state-of-the-art physics (or rather *Physis*) of the present century i.e. the fascinating manifesttations of dynamic energy landscapes (Bishop and Conradson 2004, 2007) and the systematic study of complex adaptive matter (Pines et al. 2006). For better or worse we are now witnessing a transition from the science of the past, so intimately linked to reductionism, to the study of complex adaptive matter and dynamic energy patterns. It is all firmly based on experiments, with its hope for providing a jumping-off point for new discoveries, new concepts, new wisdom and new vision. For example, we note that the maximum entropy production (MEP) is an organizational principle common to non-equilibrium systems in physics and biology (Juretic et al. 2006). And for all those who do understand that

we actually *think* our own Universe (it's our *mind-stuff*!), there is at least one certainty: there will be many unexpected new 'discoveries', actually, new mind-stuff manifestations!

4. THE UNIVERSE, CONSCIOUSNESS AND LIFE CANNOT BE AN ACCIDENT

As we are considered 'objective' scientists, most of us are lead to believe that there is some universal *definition of life*. One can easily find web pages where various definitions of life are discussed in scientific terms, yet even a rapid scan leads to the conclusion that actually we do *not* have an accepted definition of life. For example, cell biologists have worked out the basic functioning of organisms, while molecular biologists are trying to understand the molecular mechanisms of these basic functions. Geneticists have been involved in the comprehension of the replication of life, another property related to evolution and which is considered an important property of life, while ecologists have centered their attention on the basic rules that govern the complex relationship between different organisms sharing the same habitat.

However, the fact is that in spite of these, and many other different approaches, the hard reality is that we are still lacking an adequate definition of life. This difficulty in defining our object of study adds an important factor of complexity and uncertainty to the discussions of its existence, and especially to the search for life in other parts of the universe. And that is certainly one of the facts we have to keep in mind here while discussing the bio-challenges and opportunities of our scientific age. Moreover, the evolution is used as a dogma of biosciences to the extent that we are told that in biology everybody now accepts the evolution. Most scientists simply *believe* that the biological life somehow 'naturally' emerged from the Big Bang scheme from some 'pra-soup' that evolved into cells, primitive organisms, plants, animals and eventually us humans. And evolution, while very useful as a concept, is often non-critically used in present day biosciences, so that it often appears just like some 'magic' random permutation machine that enables only the 'fittest' products to survive within the dynamic space-time web.

However, *Physis* i.e. physics encompasses all nature, and therefore the whole universe with it's continuing changes: all subtle, dynamic energy patterns within the space-time. Therefore, by definition, *Physis* includes all physics, biology and all life as well and possibly even the consciousness. So, to an open minded physicist, experienced in dealing with quantum correlated systems, chemical complexities, macroscopic quantum phenomena and dynamic, functional molecular and glassy networks, the whole world (including

life) seems noble and positive i.e. resembles the poetic saying in Hindu-philosophy: '*God sleeps in minerals, dreams in flowers, wakes up in animals and lives in Men*'. All Universe (including life) is a complex emergent phenomenon, yet there is a fine underlying unity to it all, although we cannot use just one 'simple' reductionist formula to describe it. And to some of us even a phenomenon like high-Tc superconductivity (Cyrot and Pavuna 1992; Pines 1998), where some hundred thousand billion billion electrons coherently 'dance' ensemble, also belongs to a precursor bio-complexity problem, an *intermediate phase* phenomenon (Thorpe and Phillips 2001) even though most biologists at this (st)age wouldn't even dream of any connections (Pines 2006). Yet, everything in nature is highly correlated and the quantum nature of all phenomena, life included, is fundamental and has already been noticed in the previous century (London and Bohr 1935).

Moreover, the Universe, the consciousness and life cannot be an accident as it is a mathematical impossibility that chance processes produced the first living matter (Korthof 2001). Namely, the usual self-organization scenarios do not distinguish between order and complexity and fail to give a plausible method of generating sufficient information content. Even brilliant agnostic thinkers like Feynman (1967) were amazed by the fact that the smaller the space-time object physicists study, the greater intelligence is required to decode what exactly is going on. Unlike lone Newton thinking about the gravity of the Universe, in CERN thousands of physicists try to figure out properties of quarks and their ever smaller 'particle-cousins'. As the information is actually a physical quantity, the space-time-web indeed seems to be full of dynamic information; that implies that the Universe is somehow intrinsically intelligent. Finally even a simple common logic can state the undeniable fact that 'the Universe is' and 'I am'. And, as my own intelligence and consciousness are clearly a part of The Whole, consequently The Whole has to be at least equally or more intelligent than I am.

Such simple considerations are still not fully understood or accepted in the mainstream science (and especially not in biology), although bookstores seem to be flooded with books, often by distinguished writers, that discuss some form of post-Darwinistic view of Nature. Nevertheless, with physicists' positive view of the whole of life, the whole Universe appears rather optimistic. What a stark contrast to most geneticists' "bluesy" views of nature! That is primarily due to horrors that biologists daily observe in the survival-of-species, like in the survival struggle of bacteria. In general, unlike the Platonic beauty of physics, present biology does appear cruel and unfair. The aforementioned Hindu metaphor and positive physicists' views of the Universe make one certainly more relaxed about any genomic experiments that claim to be crucial to the understanding of all life (it is also often just a form of scientific 'marketing'). And it also helps to be deeply aware of the

emergence principle within all that complexity: every event in nature counts, and every experiment has it's share of importance!

5. ON OUR ONGOING EXPERIMENTS

There are several biophysics experiments that are going on in our laboratory. Namely, over the years we have used the X-rays in imaging the the 'dead' and 'live' forms of matter (Margaritondo 2006). Incidently, X-rays are an obvious example of application of physics methods to bio-challenges. A well known fact is that without X-ray data on DNA, systematically measured by Rosalind Franklin (who caught deadly cancer in the process), Watson and Crick most likely wouldn't have been able to propose a double helix model, or certainly not as the very first acknowledged authors. A direct extension of these early X-ray methods is a present day use of advanced synchrotron radiation sources. In the past several years we were also involved in the use of synchrotron radiation for biomedical imaging. Examples include the state-of-the-art spectromicroscopy or the use of coherent X-rays in 3D radiology and Scanning Near-field Optical Microscopy (SNOM; Fig. 5.2).

Using SNOM we have recently performed fluorescent imaging of neuronal cells. The SNOM is relatively recent experimental observational tool, still in development, and it aims for subwavelength optical resolution images, thereby surpassing the possibilities of the classical microscopy. This study was focused on rats' neurons whose glutamate receptors are labeled with fluorophores. Glutamate receptors are the primary mediators of excitatory synaptic transmission in mammalian central nervous system. The changes in their number underlie aspects of synaptic plasticity and information storage in the brain.

The comprehension of their trafficking through the neuron body is thus essential to understand the different functionalities and illnesses of the brain. Using SNOM setup that is able to provide the topography, the fluorescence emission map and the optical transmission properties of the neuron samples, we have obtained high resolution transmission images and determined their resolution with wavelet analysis. These images offer new information about the inner constituents of the cells. The fluorescence images are treated and interpreted in conjunction with the topography and the optical transmission (Fig. 5.3). Without giving single receptor resolution, data show a non uniform surface distribution of the receptors on the cell membrane with a higher concentration near the nucleus, thus confirming the biological models. We note the intrinsic new information given by the fluorescence measurements, independent of topographical or transmission artifacts (Lanz 2007).

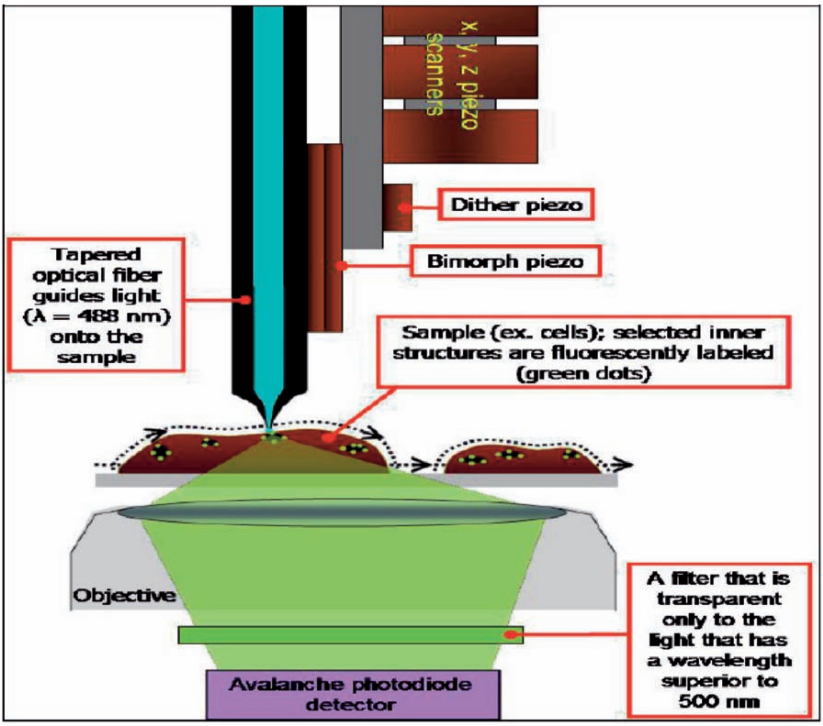


Figure 5.2. Fluorescence SNOM Experimental Scheme (Vobornik 2005).

Another ongoing experiment is the work on expression of tagged glutamate receptors in organotypic slice cultures from rat brain. Trafficking of -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors into and out of the postsynaptic membrane of neurons plays a central role in the regulation of synaptic strength, a phenomenon implicated in learning and memory. There are two main ways of trafficking for AMPA receptors: the endosomal way, i.e. exchange between surface and intracellular receptors through exocytosis/endocytosis, and the diffusive way, i.e. lateral diffusion of surface receptors in the plasma membrane. We have established a novel technique to study the trafficking of AMPA receptors. The O6-alkylguanine-DNA alkyltransferase (AGT) system combined with lentiviral gene delivery enables to label specifically the surface AMPA receptors. We have established this expression system in organotypic slice cultures from rat hippocampus instead of the so far used dissociated cultured hippocampal

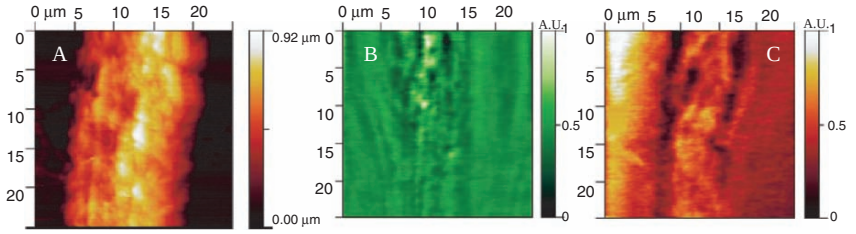


Figure 5.3. a) 25µm square topography image of a hippocampal neuronal axon, with 250 pixels per line. b) Fluorescence recording. The intensity is in arbitrary unit. c) Transmission signal. The intensity is in arbitrary unit. The borders of the axon give rise to a strong contrast due to myelin absorption.

neurons. The surface trafficking of these receptors in neurons is observed by confocal microscopy and spinning-disk confocal microscopy. In future we intend to observe larger number of receptors and statistically analyze their diffusion coefficient. As it has been shown that the trafficking of AMPA receptors depends on their subunit composition, we intend to investigate the role of the different AMPA receptor subunits in their diffusion. The open question remains whether GluR1/GluR2 and GluR2/GluR3 receptors behave differentially (Rey 2007).

6. CONCLUDING REMARKS

While physics by definition (*Physis* = Nature) encompasses all nature, we should not fall into the reductionist trap and insist to explain everything with some simple, fundamental 'formulae'. Namely, our age represents the end of reductionism and no, there is no end in sight to science, or to (theoretical) physics. It is the emerging science of *emergence* that should dominate our approaches to challenges of bio-complexity. It was introduced by biologists, as it has a natural place within bio-complexity, and yet is extremely useful in all of physics and chemistry as well. However, that should in no way reduce our critical view of (far too many) inconclusive or even irreproducible experiments that often dominate highly financed biosciences, genomics especially. The very best experiments are time consuming and consequently very expensive, but it's the experiments that are the basis of all science, so it is our collective duty to make sure that we all perform careful, meaningful and reproducible, albeit expensive, experiments. Last but not least, let us again emphasize that we are only in the very *beginning* of science. We have no idea on the exact origin of dark matter and dark energy i.e. more than 90% of the stuff of our Universe (NASA 2006), and we have as many

unknowns in the contemporary bio-challenges. So let's enjoy our respective roles in this brilliant human endeavor and do our best, so, that our ever better understanding and improved scientific methods be truly useful to all life and to the future of our young civilization.

ACKNOWLEDGEMENTS

I wish to thank Greta Pifat-Mrzljak for a marvellous organization of the summer school, her stamina and patience. I acknowledge my numerous colleagues at the EPFL and friends worldwide, all of whom encouraged me to have a bold look at biocomplexity with an open eye and audacity of a *true Physis-researcher: an eternally playing, curious child.*

REFERENCES

Anderson, P.W. (1972) More is Different. *Science* **177:** 393-396.

Bishop, A. and Conradson, S. (2004) Dynamic Energy Landscapes and Functional Systems, see: http://www.lanl.gov/mst/delfs2004/program.shtml

Cyrot, M. and Pavuna, D. (1992) *Introduction to Superconductivity and High-Tc Materials*, World Scientific, Singapore, New York.

Craig, C. (2003) Origins of Biocomplexity: Colonization and Succession of Microbial Communities in a Dynamic Geochemical Environment, NSF award #0083837, see: http://www.ocean.udel.edu/extreme2003/mission/biocomplexity/index.html

Chergui, M. (2006) Controlling Biological Functions, *Science* **313**: 1246.

Feynman, P.R. (1967) *The Character of Physical Law*, MIT Press and Penguin Press Science, ISBN-10: 0679601279.

Forró, L. and Pavuna, D. (2006) From Solid State to BioPhysics I, II, III, see: http://dubrovnik2006.epfl.ch, see also : http://ipmc.epfl.ch

Hawking, S. (1988) Interview in Newsweek, see: http://www.hawking.org.uk/home/hindex.html

Horgan, J. (1997) *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age,* Addison-Wesley, Reading, MA.

Juretic, D., Zupanovic, P. et al. (2006) 4th international meeting on maximum entropy production in physics and biology, 6-7 July 2006, Split, Croatia, see: http://www.pmfst.hr/razno/entropy/

Kelvin Lord aka Sir William Thompson (1900) Quotations, see: http://zapatopi.net/kelvin/quotes

Korthof, G. (2001) Is life an accident?, see: http://home.planet.nl/~gkorthof/kortho50.htm

Lanz, B. (2007) High resolution imaging of neuronal receptors Scanning Near-field Optical Microscopy, EPFL Master Disertation.

Laughlin, R.B. and Pines, D. (2000) The Theory of Everything, *Proc. Nat. Acad. Ssi. USA* **97**: 28-31.

Laughlin, R. (2005) *A Different Universe,* Basic Books, ISBN 0-465-03828-X.

London, F., Bohr, N. (1935) Quantum bio-physics in living organisms, see: http://twm.co.nz/quant_biology.htm and http://www.psrast.org/defknquant.htm

Margaritondo, G. (2006) Research activities, see: http://sb3.epfl.ch/gmpage.html

NASA (2006) What is the Universe Made Of?, see: http://map.gsfc.nasa.gov/m_uni/uni_101matter.html

Pavuna, D. (2007) Laboratory for Physics of Functional Matter, see: http://ipmc.epfl.ch/page31440.html

Pines, D. (1998) In: Bok, J., Deutscher G., Pavuna D., Wolf S., eds., *The Gap Symmetry and Fluctuations in High-Tc Superconductors*, Plenum, New York, pp. 111-142.

Pines, D. et al. (2006) International Institute of Complex Adaptive Matter, see: http://www.i2cam.org

Rey, G. (2007) Expression of tagged glutamatereceptors in organotypic slice cultures from rat brain, EPFL Master Disertation.

Stanford Encyclopedia (2006) Emergent Properties, see: http://plato.stanford.edu/entries/properties-emergent/

Thorpe, M.F. and Phillips, J.C. eds., (2001) *Phase Transitions and Self-Organization in Electronic and Molecular Networks*, Springer USA.

Vobornik, D. (2005) Scanning near-field infrared microscopy and spectromicroscopy applied to nano-systems and cells, EPFL Doctoral Thesis 3296.

This page intentionally blank

MR Spectroscopy and the Early Detection of Cancer in Human Subjects

IAN C.P. SMITH AND RACQUEL BAERT
Institute for Biodiagnostics, National Research Council, Winnipeg, Canada, R3B 1Y6, Phone: +1204 983 7526, Fax: +1 204 984 6978, E-mail: ian.smith@nrc-cnrc.gc.ca

Abstract

Nuclear magnetic resonance (NMR) is used to produce metabolic profiles of cancer tissue. The large amounts of data acquired necessitate the use of sophisticated statistical methods for analysis. We apply a Statistical Classification Strategy, developed in our institute, to data for fine needle aspirate biopsies of breast tumours, tissue biopsies from the prostate gland, and head and neck tumours. In all cases the diagnostic accuracy for cancer is very high (>95%). For the breast biopsies we can also diagnose if the cancer has spread to the lymphatic system, and if the tumour has established a vascular system. For the head and neck tumours we can predict which will respond to radiation therapy. We propose that high field (>/ = 3 Tesla) *in vivo* magnetic resonance spectroscopy will play a significant role in the non-invasive diagnosis of cancer.

Keywords: *in vivo* magnetic resonance spectroscopy, cancer (prostate, breast, head and neck)

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 287–298.

1. INTRODUCTION

Cancer kills too many people! In many cases a cancer detected early can be treated with minimally traumatic methods and good prognosis. So why is cancer not treated early? The main reason is that there are not many good methods for early detection. Another reason is the reluctance of affected patients to go to the doctor with minor symptoms, confident that the symptoms will go away. Our goal therefore is to overcome these deterrents and develop new methods for early detection of cancer in human subjects. The methods must be accurate, simple, non-invasive, and patient-friendly.

Our principal tools are magnetic resonance and infrared imaging and spectroscopy. Magnetic resonance has a superb spectral dispersion but relatively low detection sensitivity, whereas infrared spectroscopy has low chemical dispersion but high detection sensitivity. Thus, we choose our tools according to the needs of the problem at hand.

Our early studies concentrated on ^{1}H NMR of tissue *ex vivo*. These have been well described in a series of reviews (Smith and Baert 2004; Smith and Baert 2005; Smith and Stewart 2002; Mountford et al. 2006). We have expanded our repertoire to include stool samples, cellular biopsies, and *in vivo* studies. As we went along it became clear that the analysis methods of the day were quite inadequate to develop a robust analytical method. With our colleagues Ray Somorjai and Carolyn Mountford we therefore developed a Statistical Classification Strategy (Lean et al. 2002). We are happy to report that the combined methods are remarkably accurate, and that some of them are well enough developed to be well on their way to be introduced into clinical process.

2. THE STATISTICAL CLASSIFICATION STRATEGY

In a trial of a new method, one measures many properties of a number of subjects or specimens, represented by Table 2.1. This demonstrates an immediate problem. We have many observations (P), of which only a few may be usefully discriminatory, and relatively few subjects (N). This is the opposite of what a statistician prefers – a few critical discriminators and a large number of subjects! Our first task must thus be to reduce the number of observables dramatically such that X, the number of subjects, is much greater (x 10) than the number of observables we use. We must find the most discriminatory properties and use only a few of them. Of course we must normalize all the data and if they are spectra, register them against a reference point.

Table 2.1. A Typical medical data se of N subjects and P properties.

	PROPERTIES (P)										
	1	**2**	**3**	**4**	**:**	**:**	**:**	**:**	**:**	**:**	**Y**
Subjects N											
1	P11	P12	P13	P14	**:**	**:**	**:**	**:**	**:**	**:**	P1Y
2	P21	P22	P23	P24	**:**	**:**	**:**	**:**	**:**	**:**	P2Y
3	P31	P32	P33	P34	**:**	**:**	**:**	**:**	**:**	**:**	P3Y
:	P41	P42	P43	P44	**:**	**:**	**:**	**:**	**:**	**:**	P4Y
:	:	:	:	:	**:**	**:**	**:**	**:**	**:**	**:**	:
X	PX1	PX2	PX3	PX4	**:**	**:**	**:**	**:**	**:**	**:**	PXY

To determine the minimum number of discriminatory points in the MR spectra of biopsies, (Nikulin et al. 1998) developed a method employing the Genetic Algorithm. Surprisingly, very few properties were required, of the order of five.

This achieved, we turn next to develop a classifier. The data set is divided arbitrarily into two sets, a training set and a test set. The training set develops a classifier by methods such as linear discriminant analysis, and the classifier is assessed by analyzing the test set. We then randomly assign our data set into another training and test set, leading to a classifier that may differ from the first classifier. We then repeat this process 500–1000 times to generate a large number of classifiers. Finally, we combine all the 500–1000 classifiers to produce an overall classifier. This process is known as bootstrapping – it ensures that the classifier has been developed with access to all members of the data set, and thus has sampled as much data space as possible.

Should the accuracy at this point not be acceptable, we have yet another arrow in our quiver. We generate a new data set by performing some operation on the original data set, such as first or second derivative, or rank ordering. This new data set undergoes the above bootstrapping process to produce another overall classifier. Finally, these new classifiers are combined to form a consensus classifier. This last process, classifier fusion, usually results in increased accuracy.

Diagnostic success is expressed in various ways. Table 2.2 demonstrates the most common descriptors.

3. PROSTATE CANCER

Prostate cancer is widely prevalent in males over the age of 60. It can be slow growing, and rarely the cause of death, or fast growing and often lethal. Our early studies in this area looked at biopsies of prostate from patients

Table 2.2. Commonly used measures of classification results.

Result		Cancer	Normal	
Clinical	Cancer	TP	FN	TP = true positive FP = false positive
	Normal	FP	TN	TN = true negative FN = false negative

Sensitivity is the ability of the method to detect disease

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity is the ability to detect that disease as cancer

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{Accuracy} = \frac{TP+TN}{FP+FN}$$

with benign prostate hypertrophy (overgrown gland) and prostate cancer. Representative spectra are shown in Figure 3.1. (Hahn et al. 1997). Applying an early form of SCS to the data yielded sensitivity and specificity of 100% and 96% respectively, with an accuracy of 97%. The most discriminatory resonances were those of taurine, glutamine, glutamate anion, choline, and citrate anion.

Amongst the spectra from the patients with benign hypertrophy of the prostate (BPH), differences were seen between tissue that was predominantly stromal, and that which was mainly glandular (Figure 3.2). The resonance due to the citrate anion is much more intense in the glandular specimen. The presence of citrate in the spectra of the biopsies is a clear indication of non-malignant tissue.

A subsequent study by Swindle et al. (2003) confirmed the usefulness of ^{1}H MRS to detect cancer in biopsies of the prostate gland. The authors studied histologically benign tissue as well as that with hypertrophy, and confirmed the differences between stromal and glandular hypertropic tissue. They stressed that when comparing the MRS data with the histological reports, it was essential that the entire tissue specimen be analyzed by the histologist. This has been the case in our studies as well – whenever the two methods disagreed, step sectioning of the entire tissue sample for histological analysis confirmed the MRS result.

A therapeutic application of this technology was reported by Menard et al. (2001). They found that ^{1}H MRS could identify whether an irradiated prostate still contained viable tumour, with sensitivity and specificity of 89 and 92%, respectively. Choline was once again amongst the discriminatory resonances. Interestingly, neither type of irradiated tissue, benign or malignant,

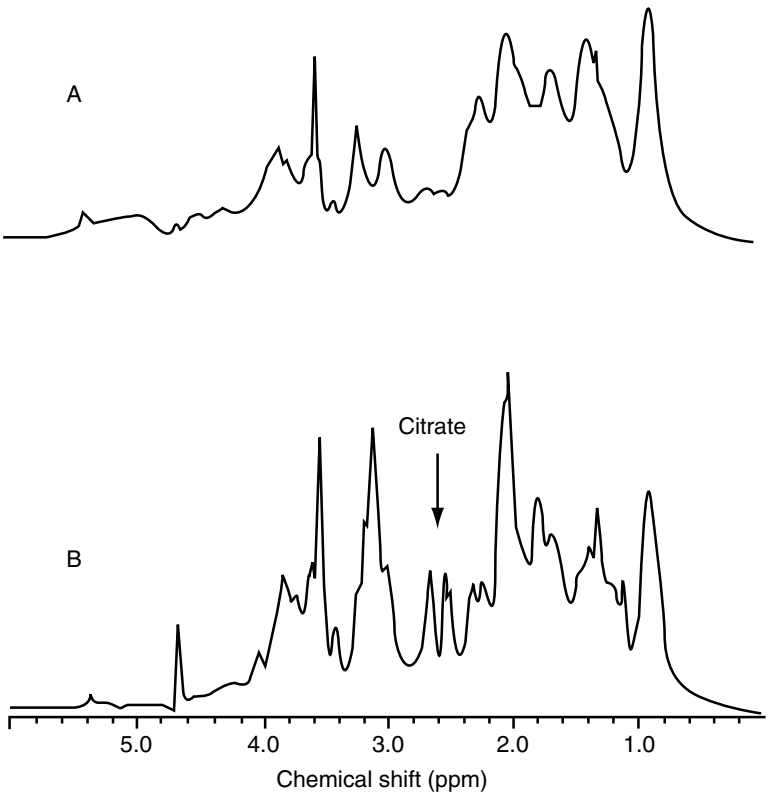


Figure 3.1. 360 MHz MR spectra, 37°C, of prostate tissue specimens A. cancer, Gleason grade 3 + 3. B. benign prostatic hypertrophy (BPH). Assignments of the resonances are: Tau, taurine; Chos, choline-containing compounds; Crs, creatine-containing compounds; Lac, lactic acid; Glu, glutamic acid; Gln, glutamine; Lys, lysine; Leu, leucine; Val, valine. From Hahn et al., 1997.

manifest the citrate resonance so characteristic of healthy prostate. Thus, the citrate resonance could not in this case be used as a monitor of benignancy.

In vivo MRS studies of prostate are of course possible, although the signal strength is much lower for a variety of reasons. The magnetic field is lower, 1.5 or 3 Tesla compared to 8 or 9 Tesla for the biopsy studies. Shimming the magnetic field is also more challenging. Volume selection can be effected by means of the gradients, and volumes of the order of millilitres can be examined. A map of the entire gland can be constructed in terms of citrate, choline, creatine, lipid, etc.

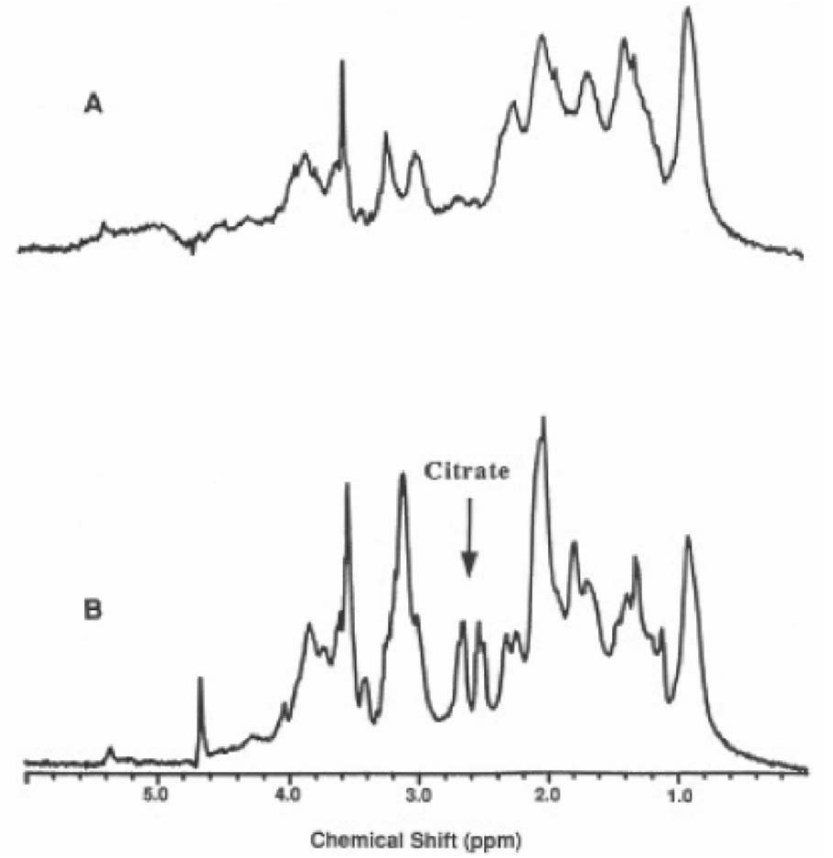


Figure 3.2. 360 MHz in ^{1}H MR spectra of BPH tissue of different compositions. A. 95% stromal, 5% glandular. B. 90% glandular, 10% stromal. From Hahn et al., 1997.

Pioneering *in vivo* studies of prostate were performed at the University of California (Kurhanewicz et al. 1992; Roach et al. 2001) where technological improvements have brought the technique to a level useful in clinical practice (Chen et al. 2006, Jung et al. 2004). Better and better spectra were obtained as the methods were improved, culminating in excellent spectra *in vivo* at 3 Tesla (Lange et al. 2006; Chen et al. 2006). Analytical methodology also improved with the inclusion of available clinical data (Dhingsa et al. 2004) and a standardized evaluation system (Jung et al. 2004). Excellent data have also been obtained by the groups in Nijmegen (Sheenen et al. 2005) and New York (Pucar et al. 2005). A typical data set is shown in Figure 3.3.

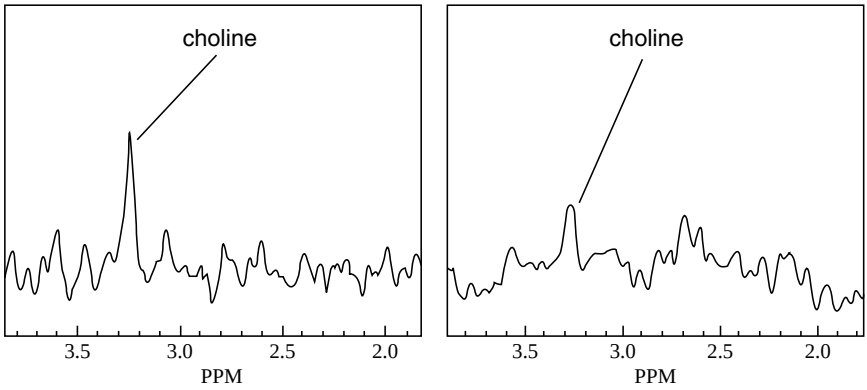


Figure 3.3. A. 130 MHz (3 Tesla) and B. 65 MHz (1.5 Tesla) ^{1}H MR spectra of human prostate *in vivo* from voxels (0.16 and 0.34 cc, respectively) placed over the zone of malignancy. Note the absence of the citrate resonance. From Chen et al., 2006.

This figure is a comparison of the spectral quality of the choline region obtained from the same patient, *in vivo*, at magnetic fields of 1.5 and 3.0 Tesla. The quality (signal to noise ratio) at 3 Tesla is clearly better – roughly a factor of two. As expected choline levels were high, and citrate levels were low in the cancerous region. It appears that the 3 Tesla instruments will soon be preferred for *in vivo* spectroscopic studies. They have become relatively affordable over the past several years, comparable to what 1.5 Tesla instruments cost five years ago.

4. BREAST CANCER

Breast cancer is the second deadliest killer of women. If contained in the breast it can be treated with considerable success. If it has spread into the lymphatic system, treatment is lengthy and debilitating, and frequently unsuccessful.

Working with scientists at the Institute for Magnetic Resonance Research, Sydney, Australia we have analyzed the MR spectra of a large number of fine needle aspirate (FNA) biopsies from breast tumours and normal breasts. FNA is the least invasive method to obtain a breast specimen. A fine needle bearing syringe is used to penetrate the tumour, and action of the plunger is used to liberate and aspirate cells for biopsy. NMR spectra at 360 MHz and clinical data were obtained in Sydney. Figure 4.4 shows results of a simple

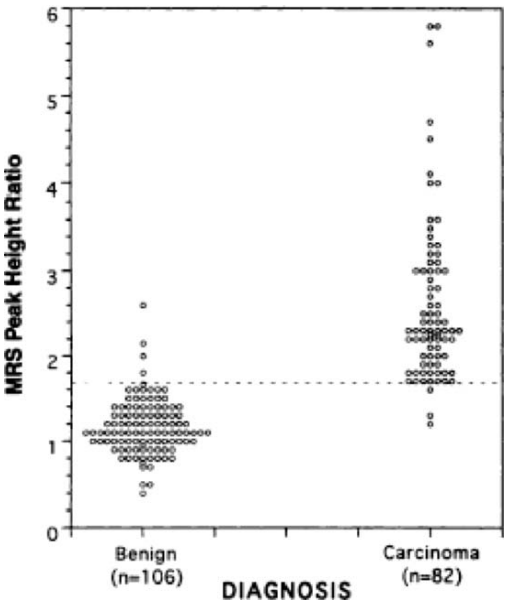


Figure 4.4. Ratio of resonances at 3.25 and 3.05 ppm in the 360 MHz MR spectra of fine needle aspirate biopsies from benign and malignant breast tissue. Data are grouped according to histopathologic findings. From Mackinnon et al., 1997.

analysis of the data for normal and malignanttissue (Mackinnon et al. 1997). The presence of a strong peak at 3.25 ppm, plus other resonances, typify the spectra of the cancer specimens. The ratios of resonances at 3.25 and 3.05 ppm show clustering for the normal and malignant specimens, with an apparent accuracy of 95%. This is not a robust analysis, and a more rigorous statistical approach was taken in a subsequent study using SCS in Winnipeg (Mountford et al. 2001). These yielded sensitivity, specificity, and accuracy of 98, 94, and 96%, respectively.

A surprising result is that analysis of the spectra of the cancer cohort led to high accuracy for metastasis to distant lymph nodes. This is incredibly valuable information since accurate diagnosis of spread to lymph nodes is critical for appropriate treatment and the life of the patient. Further interrogation of the cancer cohort data indicated another useful diagnosis of the establishment of a vascular system by the tumour, with sensitivity, specificity, and accuracy of 84, 100, and 94%, respectively. This is useful knowledge in planning the treatment of the disease with antiangiogenesis compounds.

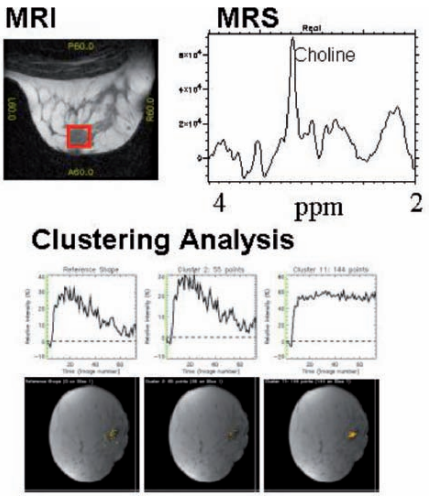


Figure 4.5. MR data (1.5 Tesla) for the breast of a subject with invasive ductal carcinoma. The tumour was located by analysis of the time-dependent uptake of gadolinium-based contrast reagent. The uptake curves (centre and right) show different behaviour. The rapid uptake to a high steady level (cluster 11) is characteristic of malignancy. That region is outlined with a square in the top left image. Its MR spectrum is shown in the top right, and contains the characteristic choline resonance at 3.2 ppm (unpublished data from Lawrence Ryner, NRC, Canada). Clustering analysis was performed with the NRC software EvIdent.

Further studies on these data led to diagnosis of the presence of receptors for estrogen and progesterone, also very valuable information for treatment planning (Lean et al. 2004).

While these discoveries promise a rapid, inexpensive, and accurate diagnosis of breast cancer, and lymph node and vascular involvement, a biopsy must be taken. Given the power of MRI to image anatomy and discern abnormalities, it would be preferable to use MRI to locate suspicious lesions and then *in vivo* MRS to seek the same discriminations that were observed in the biopsies. Patient compliance would surely be higher. We have therefore begun a programme to implement this. Figure 4.5 shows a 1.5 Tesla MRI, and the malignnant voxel detected by means of a time-dependence analysis of the uptake of paramagnetic contrast reagent by the breast. This analysis is shown in the lower part of the figure. The MRS of the malignant voxel clearly shows the choline peak so frequently detected in malignancies. Spectral quality in terms of signal to noise and dispersion of resonances will be much improved at higher fields, so we are currently conducting our studies at 3 Tesla.

5. HEAD AND NECK CANCER

The heterogeneity of neoplastic processes in the upper aerodigestive tract and the difficulty in interpreting histopathological data have led to a search for new, non-subjective, and preferably non-invasive, diagnostic techniques for head and neck cancer. The region is difficult to study by *in vivo* MRS due to the heterogeneity of materials (tissue, teeth, air), and the consequent problems with shimming the magnetic field. Therefore a study of biopsies was first performed to determine the diagnostic and prognostic value before attempting the more difficult MRS *in vivo* experiment.

We have enjoyed considerable success with diagnosis of this cancer, using the methods described above (El-Sayed et al. 2002). Sensitivity, specificity, and accuracy of 90, 83, and 87%, respectively, were obtained.

A prognostic use for the ^{1}H MRS method was demonstrated by Bezabeh et al. (2005). They explored the utility of MRS to indicate the aggressiveness of head and neck tumours and their response to treatment. By means of a simple analysis of resonance ratios (choline/creatine 3.2/3.0; lipid and/or lactic acid 1.3/0.9) and a paired t-test they were able to distinguish responders from non-responders with sensitivity and specificity of 83 and 82%, respectively. Some spectra are shown in Figure 5.6.

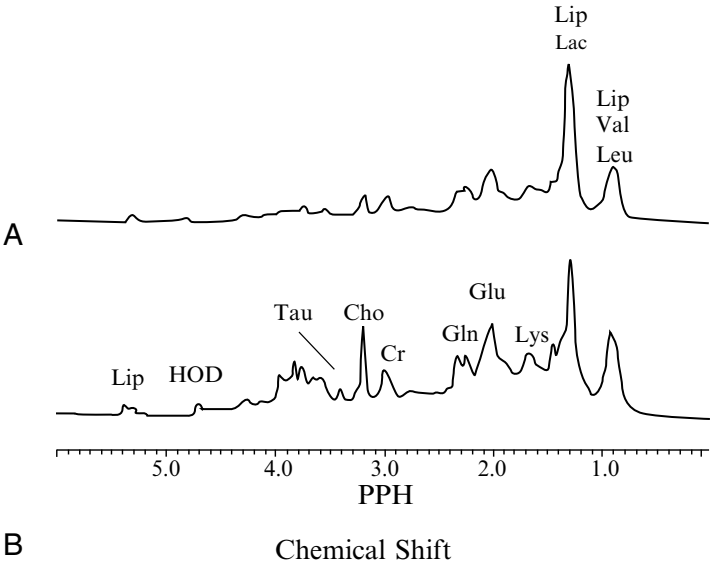


Figure 5.6. 360 MHz ^{1}H MR spectra of head and neck tumours that (a) responded and (b) failed to respond to radiation treatment: Assignments of resonances as in Figure 1. From Bezabeh et al. 2005.

6. CONCLUSION

In this brief overview I hope that the power of magnetic resonance spectroscopy for the early diagnosis of human malignancy is clear. The biopsy studies, combined with vigorous and robust statistics, lead to very high diagnostic accuracies. The *in vivo* MRI/MRS studies show potential for the same kind of accuracy to be obtained non-invasively. The advent of higher field MRI systems (3-4 Tesla) should facilitate the validation of the *in vivo* method, and its entry into routine clinical use. We have much confidence that a great many lives will be enhanced, or saved, using this high technology approach.

REFERENCES

Bezabeh, T., Odlum, O., Nason, R., Kerr, P., Sutherland, D., Patel, R., Smith, I.C.P. (2005). *Am. J. Neuroradiology* **26**: 2108-2113.

Chen, A.P., Cunningham, C.H., Kurhanewicz, J., Xu, D., Hurd, R.E., Pauly, J.M., Carvajal, L., Karpodinis, K., Vigneron, D.B. (2006). *Magn. Res. Imaging* **24**: 825-832.

Dhingsa, R., Qayyum, A., Coakley, F.V., Lu, Y., Jones, K.D., Swanson, M.G., Carroll, P.R., Hricak, H., Kurhanewicz, J. (2004). *Radiology* **230**: 215-220.

El-Sayed, S., Bezabeh, T., Odlum, O., Patel, R., Ahing, S., MacDonald, K., Somorjai, R.L., Smith, I.C.P. (2002). *Head and Neck* **24**: 766-772.

Hahn, P., Smith, I.C.P., Leboldus, L., Littman, C., Somorjai, R.L., Bezabeh, T., (1997). *Cancer Research* **57**: 3398-3401.

Jung, J.A., Coakley, F.V., Vigneron, D.B., Swanson, M.G., Qayyum, A., Weinberg, V., Jones, K.D., Carroll, P.R., Kurhanewicz, J. (2004). *Radiology* **233**: 701-708.

Kurhanewicz, J., Thomas, A., Jajodia, P., Weener, M.W., James, T.J., Vigneron, D.B., Narayan, P. (1991). *Magn. Res. Medicine* **22**: 404-413.

Lange, T., Trabesinger, A.H., Schulte, R.F., Dydak, V., Boesiger, P. (2006) *Mag. Res. Medicine* **56**: 1220-1228.

Lean, C., Doran, S., Somorjai, R.L., Malycha, P., Clarke, D., Himmelreich, U., Bourne, R., Dolenko, B., Nikulin, A.E., Mountford, C.E. (2004). *Tech. in Cancer Research and Treatment* **3**: 551-556.

Lean, C.L., Somorjai, R.L, Smith, I.C.P., Russell, P., Mountford, C.E. (2002). *Ann. Reports NMR Spectroscopy* **48**: 71-111.

Mackinnon, W.B., Barry, P.A., Malycha, P.L., Gillett, D.J., Russell, P. (1997). *Radiology* **204**: 661-666.

Menard, C., Smith, I.C.P., Somorjai, R.L., Leboldus, L., Patel, R., Littman, C., Robertson, S.J., Bezabeh, T. (2001). *Int. J. Radiation Oncology Biol. Phys.* **50**: 317-323.

Mountford, C., Smith, I.C.P. and Bourne, R. (2006) In: Webb, G., ed., *Modern Magnetic Resonance*, Vol.II, Springer, London (in press).

Mountford, C., Somorjai, R.L., Malycha, P., Gluch, L., Lean, C., Russell, P., Barraclough, B., Gillett, D.J., Himmelreich, U., Dolenko, B., Nikulin, A.E., Smith, I.C.P. (2001). *Br. J. Surgery* **88**: 1234-1240.

Nikulin, A.E., Dolenko, B., Bezabeh, T., Somorjai, R.L. (1998). *NMR Biomed.* **11**: 209-216.

Pucar, D., Shukla, A., Horicak, H., Moskowitz, C.S., Kuriowa, K., Olgac, S., Ebora, L.E., Scardino, P.T., Koutcher, J.A., Zakian, K.L. (2005). *Radiology* **236**: 545-553.

Roach, M., Kurhanewicz, J., Carroll, P. (2001). *Oncology* **15**: 1399-1410.

Scheenen, T.W.J., Gambarota, G., Weiland, E., Klamp, D.W.J., Fűtterer, J.J., Barentz, J.O., Herschap (2005). *Mag. Res. Medicine* **53**: 1268-1274.

Smith, I.C.P. and Baert, R. (2004). *Proc. Indian Natl. Sci. Acad.* **A70**: 521-531.

Smith, I.C.P. and Baert, R. (2005) In: Tosi, R. and Tugnoli, V., eds., *Nuclear Magnetic Resonance Spectroscopy in the Study of Neoplastic Tissue*, Nova Publishers, New York, pp. 29-58.

Smith, I.C.P. and Stewart, L.C. (2002). *Prog. Nuc. Magn. Spectr.* **40**: 1-34.

Microarray and Single Cell Analysis Techniques in Bio-medical Fields

WILHELM J. ANSORGE
IAET Institute of Anvanced and Emerging Technology, CH-1277 Lausanne – Borex, Switzerland, Phone: +41 79 5900 517, Fax: E-mail: Wilhelm.ansorge@epfl.ch

Abstract

Molecular diagnostic technologies will play a significant role in practice of medicine, public health and pharmaceutical industry. Together with genomics and proteomics techniques, the field of pharmacogenomics will develop, with the goal of personalization of diagnosis and therapy, for optimal efficiency and reduced toxicity, contributing to integrated healthcare. After the production of the first complete Human Genome Array at the EMBL in 2001, the array was validated in applications in several biomedical fields and the results published during the few following years. By the end of 2003 and early 2004, several well known commercial companies have announced the availability of similar comprehensive DNA arrays with oligonucleotides, short (22-70 bases long) DNA fragments. Among them Affymetrix, Agilent, Applied Biosystems. In the meantime the start-up company Nimblegen Systems offers the whole human genome on an array it sells as a service, with clients sending in the test sample mRNA, the company synthesises the oligonucleotide array designed by the customer, performs the tests and analysis, then delivers the results to the customers. Illumina company offers a bead based platform for gene expression profiling and genotyping.

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 299–306.

Keywords: Microarray techniques, protein and antibody arrays, single cell analysis, RNA interference, cell arrays, stem cells.

1. INTRODUCTION

Among the significant technologies in the hands of biological and biomedical researchers for assessing gene expression are the DNA microarrays (Microarrays collection (March 2004) Nature Reviews, 2004, www.nature.com/reviews/focus/microarrays). On glass slides, thousands of known DNA sequences are spotted or synthesized. Test mRNA is isolated from samples, eventually converted to cDNA or amplified, labeled and then hybridized on the slides. The array on the slide is scanned to measure the amount of sample bound to a particular spot. Subsequent computer analysis generates gene expression profiles of samples, the test results leading to hypotheses and conclusions. Microarrays are providing insights into such areas as study of tumor classification, vaccine development, graft rejection, aging process, circadian gene expression and stress response. Studies are carried out to verify and improve the reproducibility and accuracy of the microarray techniques, as well as to introduce standards in microarray data formats submitted to data bases. (Brazma et al. 2004).

With the completion of the Human genome sequence, the interest turned to the production of a complete Human genome DNA microarray, as one of the most powerful and desirable tools for molecular diagnostics. The Human Genome Array, completed and reported in October 2002 (Human Genome on a Chip, EMBL press release, October 2002, (www.embl.de); Wirkner et al. 2004; Wirkner et al. 2005**)** at European Molecular Biology Laboratory (EMBL) by the group of W. Ansorge in collaboration with the German Resource Centre (RZPD) team led by Uwe Radelof, was the first large whole human genome array reported worldwide. It contained the best characterized and largest cDNA clone selection at the time, with what is believed to be the almost entire Human Genome with about 52000 clones of known and predicted human genes. In contrast to "topic specific" arrays that contain only a limited number of known genes, this is a tool enabling to perform a genome-wide differential gene expression profile from any investigation on human samples, and to identify new genes which may play an important role in the investigated test characterized by the samples. During the years 2003 and 2004, this complete Human Genome microarray was validated in collaborations in several biology and clinical projects (Angiogenesis, Stem cells, Cholesterol traffic, Testicular cancer diagnosis, Gene deletions, Study of cellular receptors). The array was also used in comparative genome hybridisation for evolution studies comparing humans with chimpanzees

and mice (Khaitovich et al. 2004). The DNA microarrays, in combination with the proteomics techniques like protein and antibody arrays (www.functionalgenomics.org.uk, Maercker et al. 2004), enable the analysis and diagnostics of cellular signalling pathways, similarly to the work on endostatin pathway in angiogenesis (Abdollahi et al. 2004). A promising product of the genome array techniques will be the finding of biomarkers for reliable diagnostics of diseases, as was demonstrated already in many applications, e.g. the study of testicular cancer (Almstrup et al. 2004; Almstrup et al. 2005). The complete Human Genome microarray is a tool that can be used to compare genome-wide gene expression analysis in different types of cells, with an enormous experimental potential for biologists and clinicians. It will be possible to select for each biomedical application a suitable, most complete smaller subset of genes relevant in the concrete application, and to produce smaller dedicated chips for fast and less expensive analysis (Muckenthaler et al. 2003; Cortes-Canteli et al. 2004; Huber et al. 2004; Boengler et al. 2003; Pellicano et al. 2006; Arnold et al. 2006). Performance and reliability of various probe labeling methods were studied and characterized (Richter et al. 2002).

By the end of 2003 and early 2004, several well known commercial companies have announced the availability of similar comprehensive DNA arrays with oligonucleotides, short (22-70 bases long) DNA fragments. Among them Affymetrix, Agilent, Applied Biosystems. In the meantime the start-up company Nimblegen Systems offers the whole human genome on an array it sells as a service, with clients sending in the test sample mRNA, the company synthesises the oligonucleotide array designed by the customer, performs the tests and analysis, then delivers the results to the customers. Illumina Company offers a bead based platform for gene expression profiling and genotyping.

1.1. Microarray database, data management

There is an urgent need to set up a central database for microarray data, and a central storage of the image files. The system should allow easy access to large datasets produced over several years, as well as easy data exchange between collaborating groups. The ArrayExpress (Data base ArrayExpress and Expression profiler at EBI institute, www.EBI.ac.uk); universal database for storage of any microarray platform, organism, and experimental conditions was developed at the European Molecular Biology Laboratory (EMBL) and its European Bioinformatics Institute (EBI); (www.EBI.ac.uk., www.ncbi.nlm.nih.gov/Uni-Gene; ChipSkipper software available from Dr. Christian Schwager, (schwager@embl.de, Christian.schwager@dkfz.de).

1.2. Human genome array

The complete Human Genome microarray was validated during one year, in several collaborative biology and clinical projects (Wirkner et al. 2004; Wirkner et al. 2005) (study of angiogenesis, search of marker genes relevant in clinical cancer research studies, testicular cancer diagnosis, gene deletions, stem cells, transcriptome evolution studies in human vs. chimpanzee vs. mouse). The result is a tool that can be used to compare gene activity in different types of cells, with an enormous experimental potential for biologists and clinicians. It will be possible to select for each biomedical application a suitable most complete smaller subset of genes relevant in the concrete application, and to produce smaller dedicated chips for fast and less expensive analysis.

1.3. Oligonuclotide-based DNA arrays, alternative splicing

Oligonucleotide-based DNA microarrays are becoming increasingly useful for the analysis of gene expression and single nucleotide polymorphisms. We carried out a systematic study of the sensitivity, specificity and dynamic range of microarray signals and their dependence on the labeling and hybridization conditions, as well as on the length, concentration, attachment moiety and purity of the oligonucleotides (Relogio et al. 2002). Both a controlled set of *in vitro* synthesized transcripts and RNAs from biological samples were used in these experiments. An algorithm was developed that allows efficient selection of oligonucleotides capable to discriminate a single nucleotide mismatch. One of the conclusions about the longer oligonucleotides (60 nucleotide in length), is that they can provide better sensitivity than 25 or 30 mers, but their specificity in complex mixtures of RNA is lower than that obtained for 25 mers (Relogio et al. 2002). These data will facilitate the design and standardization of custom-made microarrays applicable to gene expression profiling, expanding analysis of alternative splicing and its increasing significance in diagnostics, as well as in sequence analyses.

As mentioned above, several well known commercial companies are producing and distributing oligonuclotided-based DNA microarrays (Affymetrix, Applied Biosystems, Nimblegen, Qiagen), which are being applied successfully in numerous biomedical analysis (Microarrays collection, see www.nature.com/reviews/focus/microarrays).

1.4. Molecular tools for array based genomic analysis, protein and antibody arrays

Techniques for both protein and antibody microarrays have been established, and applied in routine projects to verify the functionality and in diagnostics (overview information at www.functionalgenomics.org.uk; Maercker et al. 2004). A key application, particularly in medical field, requiring highly specific protein ligands or antibodies, will be the microarrays for sensitive detection of presence of certain proteins, for example bacterial antigens. They will become valuable tools in the proteomic research, in applications for monitoring of protein expression levels, protein-protein, and protein-DNA interactions. The work is in progress to develop organ and disease specific protein arrays for diagnostics, and array based analysis of protein expression in tumor cells, defining cancer subtypes by protein expression profile signature (www.functionalgenomics.org.uk).

An indispensable part of these techniques will be the wide availability of specific protein ligands, e.g. antibodies, enabling the specific recognition and detection (www.functionalgenomics.org.uk), even of low abundance proteins.

We have established a novel method for high throughput screening of monoclonal antibody producing hybridomas (De Masi et al. 2005). This method, based on protein microarrays technology, allows for the simultaneous parallel analysis of up to potentially 40000 hybridomas against one antigen, on a single chip. The novel method for monoclonal antibody screening, two orders of magnitude less expensive and about ten times faster than the standard techniques, will find applications in diagnostics projects, clinical biology, as well as in large-scale proteomics analysis (www.functionalgenomics.org.uk).

2. CELL ARRAYS, MICROINJECTION, RNA INTERFERENCE

The cell array technique will be undoubtedly another technique contributing greatly to the field of molecular diagnostics, with its promise to allow studying and determining function of genes in living cells, from various tissues and organs.

The development of DNA microarrays has allowed systematic analysis of gene expression of different organisms. But genes are only one step in the response of an organism to stimulation; the proteins play an important role in

this response. The presence or absence of proteins, the protein-protein and DNA-protein interactions are vital for the understanding of behaviour and response of cells.

The cell array, or transfected-cell microarray technique, is a development of the last few years (Ziauddin and Sabatini 2001). In a cell array, one slide contains thousands of cell clusters transfected each with a defined cDNA, resulting in the cellular expression of a specific protein. This system thus allows studying the presence and localization of proteins inside the cells. Different cDNAs, coding for the various proteins, are spotted on glass slides. Cells grow over the slide and are transfected with the cDNAs, tested proteins with the GFP-fusion tag reporter are expressed by the cells, and allow localization of the proteins in the cell by fluorescence microscopy. Each cluster of transfected cells growing above one DNA spot on the slides, expresses the corresponding protein. The technique may be applied to silencing of genes by the RNA interference. Gene silencing is analysed by digital image analysis at a single cell level, in a high throughput cell based analysis.

3. SUMMARY AND PERSPECTIVES

Initial applications of molecular diagnostics were in the field of infections, but are now expanding in the areas of genetic disorders, genetic screening, cancer, tropical infectious diseases and others (see www.nature.com/reviews/focus/microarrays). In the near future, molecular diagnostics technologies will be involved in development of personalized medicine, based on pharmacogenetics and pharmacogenomics, analyzing DNA markers for predicting patient responses to many common drugs. New inputs will be provided by very fast novel DNA sequencing gel-less techniques (e.g. Ansorge 1991, Roche and Solexa companies) applied to gene expression profiling by the quantitative SAGE method, and to important DNA fragments. Re-sequencing of gene regions by dedicated microarrays, will provide a fast technique for the detection of single nucleotide polymorphism markers (SNPs), or mutations. Examples of special interest are techniques for defining molecular fingerprint of tumors and identification of their subtypes, identification of number of gene copies in individual genomes, acquired genetic changes and diagnosing congenital genetic defects. Whole genome analysis by the microarray technique has been used as a tool in studying and characterizing Stem cells (Wagner et al. 2004; Wagner et al. 2005a; Wagner et al. 2005b).

The power, speed, sensitivity and decreasing cost will be multiplied in combination with the novel techniques of labelling, miniaturization, nanotechnology and computing.

REFERENCES

Abdollahi, A., Hahnfeldt, P., Maercker, C., Gröne, H.J., Debus, J., Ansorge, W., Folkman, J., Hlatky, L., and Huber, P.E. (2004) Endostatin's Antiangiogenic Signaling Network. *Mol. Cell.* **13**(5): 649-63.

Almstrup, K., Hoei-Hansen, C.E., Wirkner, U., Blake, J., Schwager, C., Ansorge, W., Nielsen, J.E., Skakkebæk, N.E., Rajpert-De Meyts, E., and Leffers, H. (2004) Embryonic stem cell-like features of testicular carcinoma in situ revealed by genome-wide gene expression profiling. *Cancer Research* **64**(14): 4736-43.

Almstrup, K., Hoei-Hansen, C.E., Nielsen, J.E., Wirkner, U., Ansorge, W., Skakkebaek, N.E., Rajpert-De Meyts, E., Leffers, H. (2005) Genome-wide gene expression profiling of testicular carcinoma in situ progression into overt tumours. *Br. J. Cancer.* **92**(10): 1934-41.

Ansorge, W. (1991) Process for Sequencing of Nucleic Acids Without Gel Support, German Patent Application, Date of Application: 13.12.1991, Nr. DE 41 41 178 A1.

Arnold, I., Wagner-Ecker, M., Ansorge, W., Langer, T. (2006) Evidence for a novel mitochondria-to-nucleus signalling pathway in respiring cells lacking i-AAA protease and the ABC-transporter Mdl1. *Gene* **367**: 74-88.

Boengler, K., Pipp, F., Fernandez, B., Richter, A., Schaper, W., Deindl, E. (2003) The ankyrin repeat containing SOCS box protein 5: a novel protein associated with arteriogenesis. *Biochem. Biophys. Res. Commun.* **302**(1): 17-22.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo, J., Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**(4): 365-71.

ChipSkipper software available from Dr. Christian Schwager, (schwager@embl.de, Christian.schwageer@dkfz.de).

Cortes-Canteli, M., Wagner, M., Ansorge, W., Perez-Castillo, A. (2004) Microarray analysis supports a role for CCAAT/enhancer binding protein beta in brain injury. *J. Biol. Chem*. **279**(14): 14409-17.

Data base ArrayExpress and Expression profiler at EBI institute, www.EBI.ac.uk,

De Masi, F., Chiarella, P., Wilhelm, H., Massimi, M., Bullard, B., Ansorge, W., Sawyer, A. (2005) High-throughput mouse monoclonal antibodies using antigen microarrays. *Proteomics* **16**: 4070-81.

Functional genomics, www.functionalgenomics.org.uk

Huber, P.E., Abdollahi, A., Weber, K.J., Rastert, R., Trinh, T., Krempien, R,, Ansorge, W., Wannenmacher, M., Debus, J. (2002) Expression Profiling of irradiated human lung endothelial cells using a large DNA Chip. *Inter. J. of Rad. Oncol. Biol. and Phys.* **54**: 24-25.

Human Genome on a Chip**,** EMBL press release, October 2002, (www.embl.de).

Khaitovich, P., Weiss, G., Lachmann, Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. and Pääbo., S. (2004) A Neutral Model of Transcriptome Evolution. PLoS Biol. **5**: E132.

Maercker, Ch., Abdollahi, A., Rutenberg, Ch., Ridinger, H., Paces, O., Wang, J., Ansorge, W.J., Korn, B. and Huber, P.E. (2004) Antibody chips as a versatile tool for functional analysis of genes involved in tumor angiogenesis. In *Proceedings of the HUGO meeting, Berlin*, p. 59.

Microarrays collection (2004) *Nature Reviews*, see: www.nature.com/reviews/focus/microarrays

Muckenthaler, M., Richter, A., Gunkel, N., Riedel, D., Polycarpou-Schwarz, M., Hentze, S., Falkenhahn, M., Stremmel, W., Ansorge, W., Hentze, M.W. (2003) Relationships and distinctions in iron-regulatory networks responding to interrelated signals. *Blood* **101**(9): 3690-8.

Pellicano, F., Inglis-Broadgate, S.L., Pante, G., Ansorge, W., Iwata, T. (2006) Expression of coiled-coil protein 1, a novel gene downstream of FGF2, in the developing brain. *Gene Expr Patterns* **6**(3): 285-93.

Relogio, A., Schwager, C., Richter, A., Ansorge, W., Valcarcel, J. (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res.* **30**(11): e51.

Richter, A., Schwager, C., Hentze, S., Ansorge, W., Hentze, M.W., Muckenthaler, M. (2002) Comparison of fluorescent tag DNA labeling methods used for expression analysis by DNA microarrays. *Biotechniques* **33**(3): 620-630.

Wagner, W., Ansorge, A., Wirkner, U., Eckstein, V., Schwager, C., Blake, J., Miesala, K., Selig, J., Saffrich, R., Ansorge, W., Ho, A.D. (2004) Molecular evidence for stem cell function of the slow dividing fraction among human hematopoietic progenitor cells by genome wide analysis. *Blood* **104**(3): 675-86.

Wagner, W., Wein, F., Seckinger, A., Frankhauser, M., Wirkner, U., Krause, U., Blake, J., Schwager, C., Eckstein, V., Ansorge, W., Ho, A.D. (2005a) Comparative characteristics of mesenchymal stem cells from human bone marrow, adipose tissue, and umbilical cord blood. *Exp. Hematol.* **33**(11): 1402-16.

Wagner, W., Saffrich, R., Wirkner, U., Eckstein, V., Blake, J., Ansorge, A., Schwager, C., Wein, F., Miesala, K., Ansorge, W., Ho, A.D. (2005b) Hematopoietic progenitor cells and cellular microenvironment: behavioral and molecular changes upon interaction. *Stem Cells.* **23**(8): 1180-91.

Wirkner, U., Maercker, Ch., Abdollahi, A., Wagner, M., Selig, J., Drzonek, H., Kellermann, A., Ansorge, A., Maurer, J., Schwager, Ch., Blake, J., Korn, B., Wagner, W., Ho, A.D., Huber, P., Radelof, U., and Ansorge, W.J. (2004) Human Genome on a Chip. In: *Proceedings of the HUGO meeting*, Berlin, p. 61.

Wirkner, U., Maercker, C., Abdollahi, A., Wagner, M., Selig, J., Drzonek, H., Radelof, U., Kellermann, A., Ansorge, A., Maurer, J., Schwager, C., Korn, B., Wagner, W., Ho, A.D., Huber, P. and Ansorge, W. (2005) In: Patrinos, G.P. and Ansorge, W.J. eds., Human genome on a chip in bio-medical research. *Molecular Diagnostics,* Elsevier, Amsterdam.

Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cell expressing defined cDNAs. *Nature* **411**: 107-110.

SOS Hypothesis and the Emergence of Integrative Biology

MIROSLAV RADMAN
Faculté de Médecine René Descartes, Université Paris-5, Inserm U571, 156, rue de Vaugirard, 75015 Paris, France, and Mediterranean Institute for Life Sciences, Split, Croatia, Phone: +33 1 40 61 53 21, Fax: +33 1 40 61 53 22, E-mail: radman@necker.fr

A major discovery can often be recognized by the need to coin a new word (e.g., atom, radioactivity, antibiotic, apoptosis, etc). If the discovered entity or phenomenon does not have its cognate word in our vocabulary, it is surely a breakthrough. Coining the word SOS response (SOS system, SOS replication, SOS repair) to describe coordinated multiple inducible cellular responses to DNA damage resulting in induced mutagenesis and cell survival, was such a privilege (Radman 1974). This is a personal historical account of the intellectual circumstances that led to the birth of SOS hypothesis. I shall suggest that the synthetic thinking about diverse and disconnected experimental observations that eventually led to the concept of cellular SOS response could, in retrospect, be considered as the emergence of integrative biology. It was not an easy birth, for, back in 1970/71, I could not be understood by the "best and brightest" molecular biologists residing in the Biological Laboratories of the Harvard University where I was a postdoctoral fellow with Matthew Meselson. I lived with the certainty that my failure was due only to my lack of talent and know how in clear presentation. But, now I think that it was also due to a clash of scientific cultures: the almighty rigorous, unidirectional, analytical thinking versus a fragile, creative, "lateral", synthetic thinking. My memo on the "SOS hypothesis" sent out by the end of 1970 to a dozen of top experts in DNA

G. Pifat-Mrzljak (ed.), Supramolecular Structure and Function 9, 307–313.

repair and mutagenesis inspired no answer. Fortunately, I started to work on Meselson's project of mismatch repair in genetic recombination.

Two elements of SOS hypothesis were without precedent. One was flatly iconoclast: mutation was no more seen as the unavoidable rare stochastic chemical accident (Monod's *le hazard*) but as an active inducible cellular process. The second was best expressed by Evelyn Witkin: SOS was "the first case of a coordinated cellular regulation of metabolically unrelated genes" reminiscent of cellular differentiation in eukaryotes, but transient, i.e., reversible.

Because the intellectual history of the SOS repair hypothesis was extensively reviewed (Witkin 1976; Friedberg 1997; Bridges 2005), I shall recall here only some key personal events that inspired my thinking and experiments. My passion to understand mutagenesis was triggered in 1969 at the end of my PhD work at the University of Brussels when René Thomas told me: "Miro, try to understand why we must irradiate the *E. coli* host cell in order to obtain mutagenesis of UV irradiated phage lambda? Read the 1953 paper by Jean Weigle!" Weigle showed that UV irradiated lambda survives much better when infecting an irradiated *E. coli* host than an un-irradiated one (Weigle's "UV reactivation") and that is also the necessary condition for lambda mutagenesis (Weigle 1953). More induced phage lambda mutations were obtained by irradiating the host cell alone than phage alone! That did not make sense within the chemical paradigm of mutagenesis. Weigle thought that the mutagenesis was caused by some kind of recombination between the damaged phage and host cell genomes. A later observation that the recombination-deficient *recA* mutant is also deficient in phage mutagenesis seemed to support Weigle's hypothesis.

Raymond Devoret was fond of Weigle's hypothesis and tried in 1969/70 with Manuel Blanco to obtain evidence in its favour. Coincidently, I joined Devoret's lab in Gif-sur-Yvette as a Curie post-doctoral fellow and spent the whole year thinking about correlations between physiological conditions and genetic requirements for repair, mutagenesis and lysogenic induction of phage lambda, and for induced mutagenesis of *E. coli* as well. Conceptually, when I started considering phage genome just as an externally introduced probe for cellular repair and mutagenic activities, an important step forward was made. Obvious? Not when phage T4 studies dominated mutation research and T4 does not use host replication, repair and mutagenesis functions. While I seduced my colleagues Martine Defais and Perrine Fauquet into doing in Brussels the comparative studies of the three phage-related phenomena (Defais et al. 1971), I was fortunate that R. Devoret was working, with his associate late Jacqueline George, on Ryan-Borek phenomenon of "indirect prophage induction" (Ryan-Borek et al. 1974). Indirect induction of prophage lambda is triggered in unirradiated lysogenic "female" host cell

following its mating with an irradiated "male" partner devoid of prophage. The mere reception of UV damaged F' episome DNA from another cell was sufficient to induce the intact endogenous prophage. That cleared our minds from dirtiness of radiation effects: damaged replicon can signal lysogenic induction! Can such mating also trigger DNA repair and mutagensis of infecting irradiated phage? I convinced Jacqueline George to do this new experiment and – it worked! Jacqueline went on doing endlessly all kinds of control experiments and we published this important observation only in 1974 under the unattractive title "Indirect UV reactivation of bacteriophage lambda" (George et al. 1974). A bad "maketing": the paper was forgotten from the very beginning!

Yet, when Walter Gilbert tried to make life simple by proposing that the damage within lambda operator could prevent the repressor binding – hence induction - I reminded him that for lysogenic induction lambda DNA does not need to be irradiated! Wally said: Come on Miro, how can all kinds of damage anywhere in the genome cause induction of specific undamaged host and prophage genes? The alternative was that there is a cellular "SOS repressor" that responds very much like lambda repressor (same optimal doses are required and both require functional host *recA* and *lexA* genes) (Defais et al. 1971). Now, because the first observed phenotype of *recA* mutants was the defect in recombination, when its defect in induced mutagenesis was discovered, nearly everybody thought that recombination causes mutation. If mutation defect were first to be observed, the name of the gene would have been mutA and the thinking would probably go "How can mutation cause recombination?"

Before leaving Gif-sur-Yvette for Meselson's laboratory, I gave a seminar in September 1970 in R. Thomas' lab in Brussels and speculated that, if prophage induction and induction of mutagenesis and repair were regulated by the same pathway, then the mutants of *E. coli* temperature-inducible for prophage (isolated and called *tif* by Goldthwaite and Jacob; (Goldthwait and Jacob 1964)), should be temperature inducible also for repair and mutagenesis. Knowing that I will be working on Meselson's projects, I proposed the experiment to J. George who with Marc Castellazzi showed that my prediction was correct (Castellazi et al. 1972).

At that time, thinking about mutagenesis and repair of phage was in a separate mental compartment from that of prophage induction, and the mutagenesis of bacterial chromosome was in a yet another compartment. While settling down in Cambridge Mass. and getting ready to start working on my new project, one night the fatiguing complexity collapsed in my head into a simple picture: all induced mutagenesis, viral and bacterial, is due to an inducible cellular process! The cell can "decide" to mutate and survive, or not to mutate and die! Furthermore, the process is not recombination but

erroneous replication which saves damaged DNA's life (replication) by copying damaged sites and making mutations – it is a life-saving "SOS replication"! And I started immediately writing a memo entitled "Phenomenology of an inducible mutagenic SOS replication" that was finished at the sunrise. I was excited like a child. Yes, let's call it SOS, like the distress signal on the sea where I have spent my childhood fishing with my father. This was the night of November 21, 1970 but the memo was sent out some weeks later.

Overexcited, I kept on arguing about SOS with Harvard geniuses, an exercise that ended always in their desperate looks for escaping from me. For the giants of molecular biology all I said was too complicated, and unnecessarily so. Fortunately, a year later, as I was giving up, the supreme authority in mutation research Evelyn Witkin called me from Rutgers University saying "Miro, I think you may well be right about the inducibility of the mutagenic process in *E. coli*. I can understand my experiments with the *polA* mutant only if I invoke inducibility of the process". When she visited Cambridge, I told her also about the *E. coli* mutants (*tif* and *dnaB*) temperature-sensitive for prophage induction, phage mutagenesis and repair. Within a year, alone, she obtained the expected quite sophisticated results: without SOS induction by heat, mutations rise with square of the UV dose, but with previous induction they rise linearly at higher frequencies (Witkin 2005).

However, most important for me was Evelyn's drawing my attention to her 1967 paper (Witkin 2005) showing parallels between prophage induction and filamentous growth (longitudinal growth without cell division) and there was the cell cycle control by the SOS system! The list of inducible phenotypes grew quickly to a dozen involving co-regulated genes that are clearly metabolically unrelated. In 1973, Witkin presented the SOS hypothesis on the International Genetics Congress in Berkeley and I did the same (thanks to her and Jan Drake's insistence) on an Environmental Mutagenesis congress in Rochester. SOS became the public story. I published my text, a modification of the 1970 memo, in the difficultly quotable proceedings of the Rochester meeting (Radman 1974).

Thanks to some unpublished experiments of Maury Fox at MIT that he shared with me in our long discussions, my young group in Brussels showed later the kinetics of SOS induction of the mutagenic activity on infecting phage and the requirement for *de novo* protein synthesis after irradiation (Defais et al. 1976). SOS showed the standard enzyme induction pattern! But of which enzyme(s)? A DNA polymerase, of course! Witkin and Bridges dreamed about "error-prone" polymerases quite a while ago. There must be an inducible SOS polymerase! We used in new experiments the irradiated single-stranded phage phiX174, which is also subject to SOS repair and

mutagenesis and can be rescued only by trans-lesion synthesis across damaged sites. Analyzing in CsCl density gradients and sucrose velocity gradients the extent of synthesis on damaged single stranded templates *in vivo*, we could observe further extension of synthesis past blocks caused by the damage, but only under SOS conditions. The TLS, trans-lesion synthesis, was demonstrated and published in 1977 (Caillet-Fauquet et al. 1977).

No event made way for SOS as much as did a meeting organized by E. Witkin and E. Pollard in 1975 in Gainsville Florida. All actors and key future actors of SOS research were there. Jeff Roberts came to the meeting with the revelation that the lysogenic induction occurred by inactivation of lambda repressor via a proteolytic cleavage. The RecA protein was identified, by L. Gudas and A. Pardee, as an inducible "protein X". Oxidative metabolism appeared also under SOS control. I made the point that the inducible mutagenesis can lead to adaptive evolution on the population level and Hatch Echols called SOS the "inducible evolution". Walter Troll and Bernie Weinstein from Columbia told us about the importance of proteolysis in cancer promotion. Small natural protease inhibitors prevented chemically induced cancers in mice. With no clues what happens in mammalian cells, we dreamed that prevention of SOS response in mammalian cells could lead to cancer prevention. The atmosphere was one of a scientific revolution that is about to happen. Evelyn Witkin took a sabbatical to write the first comprehensive review on SOS that became the citation classic (Witkin 1976).

Later on, Little and colleagues identified the SOS repressor as the product of the well known *lexA* gene identified by Witkin in 1967 as a non-inducible mutant. LexA was shown to be built, and proteolytically self-cleaved with the assistance of RecA, just like lambda repressor (Little et al. 1980). But the "SOS polymerase" remained elusive until 1999 when three labs showed almost coincidently that there are three SOS inducible polymerases in *E. coli* (PolII, PolIV and PolV) all involved in some aspects of mutagenesis (Friedberg et al. 2002). The principal bacterial mutagenic polymerase is PolV encoded by the SOS inducible *umuCD* genes whose activity is doubly controlled by the RecA-assisted proteolytic cleavage inactivating the LexA repressor and activating the induced UmuD protein. But the work with yeast and human cells produced a surprising story: the SOS polymerases, now called TLS polymerases (because not all are inducible in eukaryotes) have evolved to copy the cognate DNA damage in an error-free fashion (without Watson-Crick base pairing), but the non-cognate lesions and healthy DNA in an error-prone fashion! They are specialized polymerases and there are many of them (Friedberg et al. 2002).

Thinking now back to the period of 1970-75 – what an extraordinary intellectual adventure it was! The SOS gene network also created a network of scientists and many precious friendships. In late 1980-ies the SOS system

became the textbook knowledge as the best understood complex multi-gene control system. Several cancer syndromes happen to be due to SOS-like processes in human cells. The discoveries of conceptual descendants of SOS response in human cells kept lining up: numerous cell cycle genes were discovered, p53, ATM, ATR, Rb etc. with their crucial involvement in human malignancies and in cancer predisposition syndromes (Friedberg et al. 2006), earned at least five Nobel prizes. The concept of the SOS story became one of the milestones of molecular and cell biology and the probable forerunner of today's integrative or systems biology. But how painful was its birth!

REFERENCES

Bridges, B.A. (2005) Error-prone DNA repair and translesion DNA synthesis. II: The inducible SOS hypothesis. *DNA Repair* (Amst) **4**: 725-6, 739.

Caillet-Fauquet, P., Defais, M. and Radman, M. (1977) Molecular mechanims of induced mutagenesis. Replication *in vivo* of bacteriophage ox174 single-stranded, ultraviolet light-irradiated DNA in intact and irradiated host cells. *J. Mol. Biol.* **177**: 95-112.

Castellazzi, M., George, J., and Buttin, G. (1972) Prophage induction and cell division in *E. coli*. I. Further characterization of the thermosensitive mutation tif-1 whose expression mimics the effect of UV irradiation. *Mol. Gen. Genet.* **119**: 139-152.

Defais, M., Caillet-Fauquet, P., Fox, M. S., and Radman, M. (1976) Induction kinetics of mutagenic DNA repair activity in *E. coli* following ultraviolet irradiation. *Mol. Gen. Genet.* **148**: 125-30.

Defais, M., Fauquet, P., Radman, M., and Errera, M. (1971) Ultraviolet reactivation and ultraviolet mutagenesis of lambda in different genetic systems. *Virology* **43**: 495-503.

Friedberg, E.C., Wagner, R. and Radman, M. (2002) Specialized DNA polymerases, cellular survival and the genesis of mutations. *Science* **296**: 1627-1630.

Friedberg, E.C. (1997) *Correcting the Blueprint of Life – An Historical Account of the Discovery of DNA Repair Mechanisms*. Cold Spring Harbor Laboratory Press.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R. D., Schultz, R.A. and Ellenberger, T. (Eds), 2006 *DNA Repair and Mutagenesis, 2nd Edition*. ASM Press, Washington, D.C.

George, J., Devoret, R. and Radman, M. (1974) Indirect Ultraviolet Reactivation Of Phage Lambda. *Proc. Natl. Acad. Sci. USA* **71**: 144-147.

Goldthwait, D., Jacob, F. (1964) On The Mechanism Of Induction Of Prophage Development In Lysogenic Bacteria. *C R Hebd Seances Acad. Sci.* **259**: 661-664. In French.

Little, J.W., Edmiston, S.H., Pacelli, L.Z., and Mount, D.W. (1980) Cleavage of the *Escherichia coli* lexA protein by the recA protease. *Proc. Natl. Acad. Sci. USA* **77**: 3225-3229.

Radman, M. (1974) Phenomenology of an inducible mutagenic DNA repair pathway in E. coli: SOS repair hypothesis. In: Prakash, L., Sherman, F., Miller, M.W., Lawrence, C.W. and Taber, H.W., eds. *Molecular and environmental aspects of mutagenesis*, (6th Rochester Conf. on Environemental Toxicity, 1973) C.C. Thomas Publ., Springfield, Illinois (1974) 128-142.

Ryan-Borek I., George, J., Devoret, R. and Radman, M. (1974) Indirect Ultraviolet Reactivation Of Phage Lambda. *Proc. Natl. Acad. Sci. USA* **71**: 144-147.

Weigle, J.J., (1953) Induction of mutation in a bacterial virus. *Proc. Natl. Acad. Sci. USA* **39**: 628-636.

Witkin, E.M. (1974) Thermal enhancement of ultraviolet mutability in a tif-1 uvrA derivative of *Escherichia coli* B-r: evidence that ultraviolet mutagenesis depends upon an inducible function. *Proc. Natl. Acad. Sci. USA* **71**: 1930-1934.

Witkin, E.M. (2005) The radiation sensitivity of *Escherichia coli* B: a hypothesis relating filament formation and prophage induction. 1967. *DNA Repair* (Amst) **4**: 727-31.

Witkin, E.M. (1976) Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*. *Bacteriol. Rev*. **40**: 869-907.

This page intentionally blank

Subject Index

Actin, 176, 213
Agonist ligand, 59, 61
Amloid, 24, 111, 118, 120,
  Deposite, 24
  Fibril, 21, 103, 125
  Formation, 116
  Forming peptide, 113
AMPA receptor, 282
Amylin, 112
Antibiotic, 209, 211, 214
Antioxydant, 153, 155, 160
Atomic force microscopy (AFM), 118, 154, 168
ATPase, 61, 63

Bacteriorhodopsin (bR), 55, 57
Benzoguanine, 234, 235, 237, 241
Biocomplexity, 284
Blocking ligand, 59
Bootstrapping, 289

Ca^{++}-ATPase, 46
Cancer, 83, 84, 91, 132, 149, 153, 156, 186, 281, 300, 302, 311
  Breast, 293, 295
  Head and neck, 296, 297
  Prostate, 290
Capsid, 136, 142, 144, 148
Carbohydrate, 92, 93
Chromatin, 81

Conformational plasticity, 7
Conformational space annealing (CSA), 15
Complementary measures, 14
Core
  Assembly, 136
  Immature, 141
  Mature, 144
  Protein, 136, 139, 146
  Viral, 136
Crystallography, 2, 34, 42, 104, 109, 120, 142, 209
Critical Assessment of Predicted Interactions (CAPRI), 1, 2, 9, 14, 17
Critical concentration, 108
Cryo electron microscopy, 2, 149, 167

Docking , 1, 6, 13, 251, 252, 255, 260, 265, 267
  Algorithm, 14
  Methods, 1, 9, 14
  Procedures, 2, 6
DNA, 132, 133, 153, 214
  Circular, 134
  Double-stranded, 133, 139, 147
  Minus-strand, 137, 146
  Plus-strand, 137, 146
  Viral, 131

Drug target, 46
Druggability, 254, 259
Dynamic nuclear polarization (DNP), 182

Electron microscopy, 59, 104, 109
Electron Paramagnetic Resonance (EPR), 24, 184

Imaging, 181, 201
In-vivo, 181
Longitudinally-detected (LOFEPR), 182
Oxymetry, 182, 184, 201
pH-imaging, 193, 201
Pulsed Fourier transform, 186
Spin-labelling, 206
Electron Spin Resonance (ESR), *see* Electron Paramagnetic Resonance (EPR)
Emergence, 273, 281, 283, 307, 313
Envelope protein, 134, 148

Fibril, 21
Formation, 21, 25, 27,
Fluorescence, 105, 117, 157, 182
Imaging, 273, 281
Folicle-associated epithelia (FAE), 212
Fullerene, 154, 158
Fullerol, 154, 160, 164, 166, 168, 173, 175, 176

Glycan, 92, 93, 99
Glutamate receptor, 273
Glutathione (GSH), 187, 192
Glycomics, 89
Glycoproteins, 92, 99
Glycosilation, 92, 96
Gramicin A, 53, 54
GTP, 104, 105. 107

Hepadna Virus, 132
Hepatitis B virus (HBV), 132
Hepatocyte, 134, 148
Heteronuclear
Coupling, 33, 50, 53, 62
Experiment, 36
Spectroscopy, 36
High-Throughput Screening (HTS), 251
Homology modelling, 17
Homonuclear,
Coupling, 50, 51, 54
Human Proteome Organization (HUPO), 91

Inferential structure determination (ISD), 42
Infrared spectroscopy, 21, 30, 111
Imaging, 288
Insulin, 21, 25, 26, 27
Interactome, 2
Ion channel, 45, 46, 54, 59
Isothermal titration calorimetry (ITC), 256

Jabłoński-Kasha diagram, 158

Magic angle oriented sample spinning (MAOSS), 48
Magic angle spinning (MAS), 48, 49, 50
Magnetic resonance imaging (MRI), 159
Mass spectrometry (MS), 89
Electrospray (ESI), 89, 99,
Fourier Transform Ion Cyclotron Resonance (FT-ICR), 89, 90
Fast atom bombardment (FAB), 89
Linear and threedimensional ion traps (IT), 90
Matrix-assisted laser desorption ionization (MALDI), 89, 99
Quadrupole time-of-flight (QTOF), 90
Microarray techniques, 299, 300
Antibody, 303
Cell, 303, 304
Human genome DNA, 300, 301
Oligonucleotide-based DNA, 302
Protein, 303
Microcalorimetry, 265
Microtubule, 103, 104, 108, 125
Dynamic instability, 103, 108
Molecular diagnostic, 299, 304
Molecular interaction field (MIF), 15
Mutagenesis, 307, 308

Neuron, 281, 282
Neurotensin, 64
Neutron scattering, 24

Nicotinic acetilcholine receptor, 53, 54, 59
Nitroxide, 181, 182, 184, 187, 194
  Nitronyl (NNR), 199, 201
  NO-sensitive, 198
  Thiol-sensitive, 187
  Thiol-specific, 190
Nuclear-cytoplasmic transport, 75
Nuclear Magnetic Resonance (NMR), 2, 16, 25, 33, 45, 92, 98, 112, 119, 181, 200, 252, 254, 287, 288, 293
  Biomolecular, 33, 38
  Correlation spectroscopy, 50
  Cross-polarization (CP), 51
  High power proton decoupling, 50
  In-vivo, 287, 291, 295
  Multiple-quantum (MQ) coherence excitation, 53
  Polarization inversion spin exchange at the magic angle (PISEMA), 53, 55

  Rotationl Resonance (R^2), 51
  Rotational echo double resonance (REDOR), 52, 55, 62
  Solid state, 45, 115
  Solution state, 46
  Two dimensional (2D) exchange, 53
Nuclear Overhauser effect (NOE), 33, 34, 40, 254

Oxidative stress (OS), 153, 155, 187

Peroxisome proliferator-activated receptor-γ (PPAR-γ), 81
Phage, 308, 310
Pharmacophore, 217
Phospholamban, 46
Photodynamic therapy, 153, 156
Photo-oxidative stress, 154, 174
Photosensitizer (PS), 154, 156, 158, 174
Plasmid, 211
Polarization Index Slant Angle (PISA), 53
Polymerase, 131, 135, 147, 310, 311
Polymorphonuclear leucocyte (PMN), 212
Porin, 55
Prion
  Desease, 123, 125, 126
  Protein oligomers, 122
  Protein (PRp), 119, 121, 122, 123
Promyelocytic leukemia protein (PML), 82
Prophage, 308, 309
Protein
  Complex, 1, 2, 4, 5, 9, 37
  Data bank (PDB), 4, 54
  Folding, 22
  Membrane, 45, 48
  Polytopic, 55
  Structure determination, 33, 38, 42
Protein kinase (HprK), 12
Protofilament, 104, 107, 109, 118, 121, 126
Proton electron double-resonance imaging (PEDRI), 182, 195
Pteridine, 222
Pyridazindione, 227

Reactive oxygen species (ROS), 153, 160, 165, 168
Reptin, 81
Residual Dipolar Coupling (RDC), 33, 38
Retrovirus, 132, 137
Rhodopsin, 46, 55, 64
RNA, 213
  Cellular, 134
  Pri-genomic, 132, 134, 135, 138, 146

Quinazolinone, 219, 223, 224, 227, 228
Queuine, 215, 216

Scanning near-field optical microscopy (SNOM), 273, 281
Scoring, 14, 251, 253, 255, 261, 263, 264
Shigellosis, 209, 210, 246
Single cell analysis, 299

Singlet oxygen, 154, 155, 156, 157, 158, 159, 164, 165, 166, 168, 176
Side chain
  Comformation, 15
  Flexibility, 15, 16, 17
  Optimization, 15
Small Ubiquited Modifier (SUMO)
SOS response, 307, 312
  System, 310
  Replication, 307, 310
  Repair, 310, 312
Statistical classification strategy (SCS), 287, 288, 290, 294
Structure base design, 215
Substituted imidazole pyridines (SIPS), 62
SUMO *see*: Small Ubiquited Modifier, 75, 76, 83
  Binding proteins, 79
  De-conjugation, 79
  Isopeptidase, 80
Sumoylation, 78, 80, 81, 82

TEMPO, 186
TEMPOL, 162, 163, 165
TEMPONE, 184
Thiol, 187
Total internal reflectance fluorescence microscopy (TIRF), 117
Transmembrane domain peptide (TMD), 54
Transverse relaxation- optimized spectroscopy (TROSY), 35, 46
tRNA-guanine transglycosylase (TGT), 209, 215, 216, 218, 219, 220, 223, 232, 233, 235, 241, 245, 246
Tubulin, 103, 104, 105, 109, 126

Ubiquitin, 75, 77, 82
  Ligase, 83
Uoabain, 62

Vaccine, 211
Virion, 131, 132, 140, 144
Virtual ligand screening, 222, 251, 252, 257, 259, 260, 261, 267

Young model, 169, 170, 171, 173