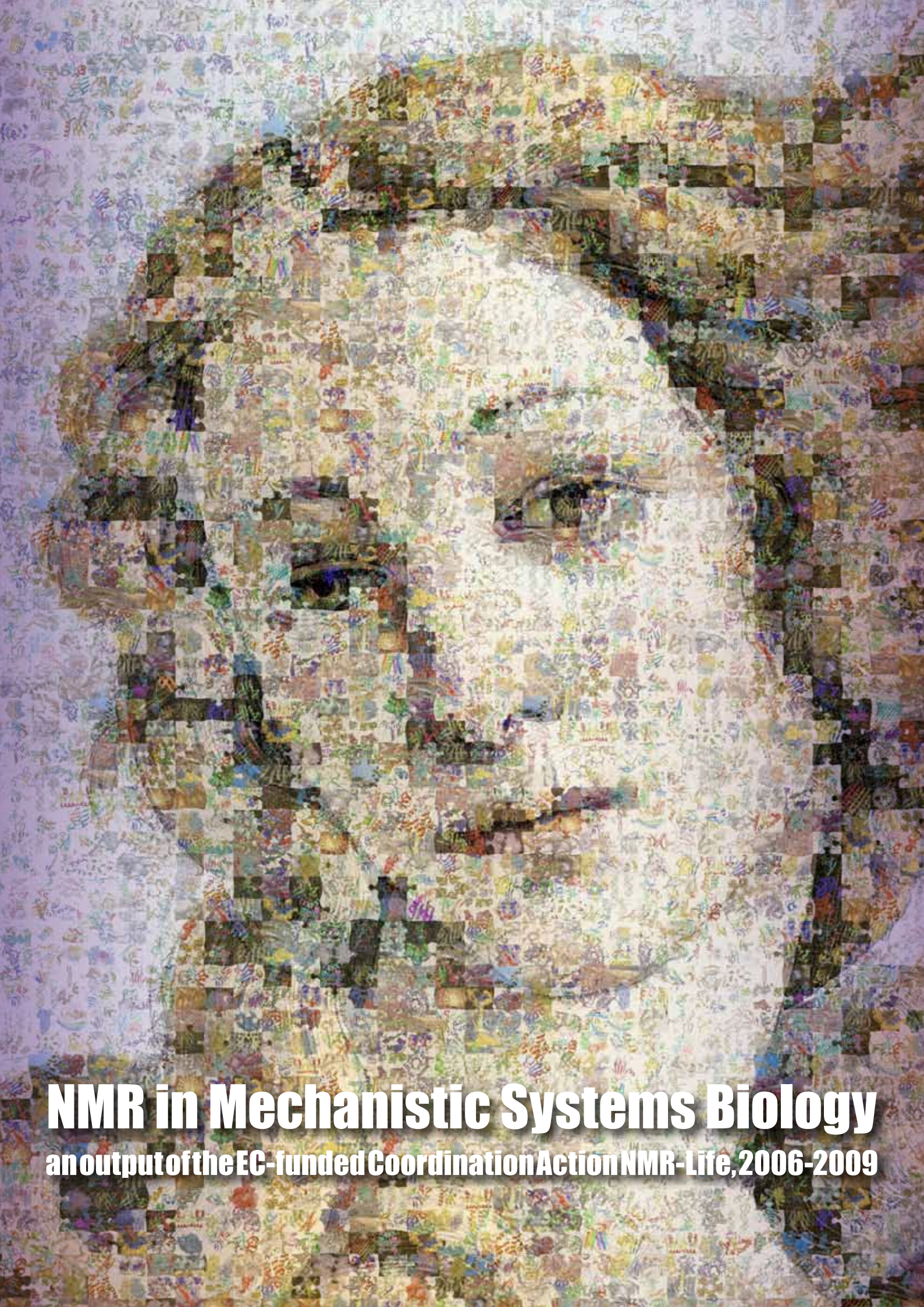


NMR in Mechanistic Systems Biology

an output of the EC-funded Coordination Action NMR-Life, 2006-2009

The Coordination Action NMR-Life is funded by the European Commission's Framework Programme 6, contract no. 018758.

Kathleen McGreevy (Magnetic Resonance Center, University of Florence) managed the volume's compilation and editing. Lorenzo Norfini (Studio Grafico Norfini, Florence, Italy) created the graphic design and lay-out. The cover art is adapted from Botticelli's The Birth of Venus.

Table Of Contents

Foreword

The Coordination Action (CA) "NMR-Life" is over. A large number of activities have taken place; this is normal, and is expected for every successful project. However, these activities have gained impact from the evolution of the international scientific scenario.

A target, if not "the" target, in the Life Sciences is that of understanding the mechanisms of Life at the molecular level and of modeling them in such a way to be able to predict outcomes when a given organism assimilates food, drugs, or other chemicals. Today, this is known as systems biology. NMR can play a strategic role in this scientific challenge. It can be used to study interactions, both from the thermodynamic and kinetic points of view, between and among biomolecules and between biomolecules and small ligands. Through metabonomic studies, it can monitor the whole metabolic process. On top of this, it can solve the structure of biomolecules and tell us how the biomolecules interact. We can refer to these achievements as the mechanistic contribution to systems biology, or **mechanistic systems biology**.

In order to evaluate the contribution of NMR to mechanistic systems biology, we can note that the percentage of structures in the PDB solved by NMR has been stable throughout the years at around 15%. Solid-state NMR has begun to contribute to the number of solved structures mainly of proteins in the form of microcrystals, but also immobilized in membranes or as fibrils.

The size of the biomolecules to be investigated by NMR in solution is steadily growing. In the solid-state we can possibly label a component of a large molecular machine and profitably study structure and function.

Progress in our technique is based on (i) increasingly intense magnetic fields to boost sensitivity and resolution; (ii) increasingly refined probes to enhance sensitivity; (iii) continuous reduction of the quantity of sample needed; (iv) development of software for reducing time and/or gaining resolution; (v) development and standardization of procedures for structure determination; (vi) integration with complementary techniques like X-ray diffractometry, SAXS, and cryo-electron microscopy.

All the above areas have been addressed during the duration of the CA, and documents have been issued on specific subjects (see www.postgenomicnmr.net). Some of them have been published in, or submitted to, widespread journals such as Nature Reviews Drug Discovery. I am proud of the results and thankful to all of the contributors.

Advantages have been gained from fruitful interactions with the work of other EC-funded projects such as SPINE2, which is devoted to protein-protein complexes, INSTRUCT, the European Infrastructure on Integrated Structural Biology, FESP, the Forum for European Structural Proteomics, and the NMR e-infrastructure, which aims to provide the scientific community with workflows of programs installed on a newly implemented grid.

During the final meeting held in Montecatini, a number of documents were discussed within the context of prestigious lectures and released on the CA web site. The final act was the presentation of such a collection of documents at the user meeting of the NMR Integrated Infrastructure Initiative (I3) EU-NMR (European Network of Research Infrastructures for Providing Access and Technological Advancements in NMR), January 26-29, 2009. Indeed, the community of NMR users is that which has contributed to and will profit from the EC-funded NMR-Life project.

The next objective is to interface other bio-NMR communities, starting with those of North America and Japan and continuing on to those of China, India, Russia, Australia, etc. We would like to share our analysis of scientific perspectives and to agree on possible platforms for the treatment of spectral data up to the description of structures and interactions. We would like to thank the EC for financing this CA project and for supporting the bio-NMR community through transnational access programs and through the development of an e-infrastructure.

Ivano Bertini

Chapter 1
Mechanistic Systems Biology

1.1 Foreword

Author: Lucia Banci[a]
[a] Magnetic Resonance Center, University of Florence, Italy

Over the years of activity of this Coordination Action, the role and impact of NMR in the Life Sciences has evolved dramatically, from being a mainly structural and dynamical tool to playing a strategic role in addressing mechanistic aspects at a comprehensive, systems level. It is becoming evident that NMR can have a major impact on Systems Biology, by providing critical functional and mechanistic information on processes at atomic resolution. The CA NMR-Life and the Forum for European Structural Proteomics (FESP) organized a meeting which involved about 30 prestigious scientists that were selected from among authorities in various aspects of systems biology and structural biology, as well as officers from major funding agencies worldwide and scientists from large structural genomic and proteomic initiatives. Drawing from the contributions of the participants at the meeting and the consequent discussion, Antonio Rosato, Joanne Kotz and Giulio Superti-Furga wrote a document that is the first to appear in the collection written by the CA NMR-Life. This document is strategic as it defines the frame of all the others, i.e. the frame within which the methodological and technical developments of NMR should be planned. The original mainly addressed the role of Structural Biology, of which NMR is a relevant aspect, in Systems Biology. The indication was that a "molecular" dimension needs to be attached to the overall modelling of systems; in this respect, structural biology can provide a striking contribution. The concept of "Molecular Systems Biology" was developed, which was defined as the ability to model systems to predict biological outcome at the molecular level.

It emerged that an understanding of molecular systems biology is unthinkable without an appreciation of the dynamic structure of proteins, the specificity of protein-protein interactions and the resulting properties of molecular machines, pathways and entire networks. It is believed that this molecular "protein" perspective is critical for systems biology to have a much-needed impact on medicine and pharmacology.

The document discussed a "Structural Biology" vision, pointing out at the critical opportunities for Structural Biology in a "systems" perspective and addressing the topics on which a molecular description has a major impact. The frontiers, in term of developments in computational modelling and experimental data acquisition tools, were also addressed. The role of Structural Biology within Systems Biology and its contribution to various aspects of the description of the biology of the system were analyzed and specific contributions were indicated. Finally, the impact of Molecular Systems Biology on molecular medicine was analyzed, with respect to both understanding and treating diseases.

As time passes, the scientific challenges and the goals and needs of the scientific community are evolving and changing, together with the concepts and perceptions of scientists.

We think that the document on "Molecular Systems Biology" is still valid and very authoritative.

However, during the general discussion at the CA closing meeting, the role of NMR in Systems Biology was addressed further and in a more in-depth way, and the evolution of the field was discussed.

It is evident that NMR, in terms of providing structural information in solution, i.e. in sample conditions close to physiological ones, has shown its power in characterizing biomolecular interactions, including transient and weak ones that usually escape other characterization approaches. However, its contribution to characterizing functional processes, including mechanistic, thermodynamic, and kinetic aspects is more striking. The concept of "Mechanistic Systems Biology" has therefore been developed, which can be defined as a description of functional pathways based on the 3D structural and dynamic interactions

of various actors.

Within the frame of this definition, the need to address the interactions among biomolecules at an atomic level, the thermodynamics and kinetics of their interactions, and how their 3D structure changes during interactions, all through the study of their structure and dynamics, emerges.

Towards these aspects NMR is the technique of choice. Therefore, the document on Molecular Systems Biology needs to be focused on the role of NMR in Systems Biology, stressing the mechanistic aspects which can be elucidated through NMR.

The various methodological and technological developments of NMR discussed within the CA and reported in the specific documents are instrumental for the perspectives of "Mechanistic Systems Biology". The bio-NMR community is committed to advancing the impact and role of NMR in this direction in order to provide quantitative data at an atomic level that can be used (and indeed, which are needed) to model functional processes and their mechanisms at a system level.

Within this frame, NMR, with its broad range of applications, i.e. from solution to solid state, from structural to dynamical characterization, from single molecules to large complexes, and integrating information of various natures, can contribute to providing data and models to be used to describe the mechanisms of the system.

1.2: New challenges in the Life Sciences - Prioritizing European research in Molecular Systems Biology

Authors: Antonio Rosato[a], Joanne Kotz[b], Giulio Superti-Furga[c]
[a] Magnetic Resonance Center, University of Florence, Italy
[b] Nature Chemical Biology
[c] Austrian Academy of Sciences

Motivation

Research in the post-genomic era is moving toward new horizons, which are largely embraced by the broad definition of *Molecular Systems Biology*. To encourage discovery in this epoch, we need to prioritize and focus on the future of such research. Where do we want to go, and how will we get there?

Systems biology addresses the properties of entire biological systems and subsystems as opposed to the isolated study of their individual components. Fundamental properties of biological systems rely on the spatial and temporal interactions of the macromolecules that compose the system and can only be understood by looking at the system as a whole. An understanding of *molecular* systems biology, i.e. the ability to *model* systems to predict biological outcome at the molecular level, is unthinkable without an appreciation of the dynamic structure of proteins, the specificity of protein-protein interactions and the resulting properties of molecular machines, pathways and entire networks. We believe that this molecular "protein" perspective is critical for systems biology to have a much-needed impact on medicine and pharmacology.

A symposium entitled, "New Challenges in the Life Sciences: Prioritizing European Research in Molecular Systems Biology", was held on October 18-19, 2007 in Florence, Italy, to provide a forum for the scientific community involved in on-going major European post-genomic projects to discuss the importance of molecular research in advancing systems biology. The meeting was sponsored by the European Commission (EC) through the *Coordination Action NMR-Life*, in conjunction with the *Forum for European Structural Proteomics (FESP)*. It is hoped that the newly established European Infrastructures in the Biomedical Sciences, as outlined by the *European Strategy Forum for Research Infrastructures (ESFRI)* and implemented by the European Commission, will find this discussion helpful in fine-tuning their funding objectives. The symposium provided an opportunity for the participants to share their work and make their voices heard as we look toward the future of molecular systems biology.

The meeting involved about 30 experts, representing the multidisciplinary nature of the field, and included scientists and policy-makers from the US, China, Japan, India and Europe. The meeting consisted of four sessions that included presentations from invited speakers followed by group discussions. The first session addressed the new challenges in protein chemistry associated with molecular systems biology, the second focused on the role of structural biology from a systems biology perspective, and the third on the impact of molecular systems biology on molecular medicine. The fourth session reviewed funding opportunities and strategies.

This document represents a synthesis of the contributions of the participating scientists and strives to define the importance of a molecular foundation for a systems-level understanding of biology, to identify the opportunities that will be afforded by a '3D' view of biological systems, and to recommend funding priorities for advancing molecular systems biology."

A vision for molecular systems biology

Individual protein and RNA structures, as well as protein-pro-

tein and protein nucleic acid complexes provide a knowledge base that is robust and detailed. This structural foundation, in turn, constitutes a basis for extrapolation from a given biological system to produce testable hypotheses about its response to perturbations. The information derived from one system can then be extrapolated to homologous processes. Structural data are essential to generate predictions of how genetic variation affects protein activity and, ultimately, impacts phenotype. Likewise, the resulting framework offers the ability to hypothesize how networks in humans may operate and respond by analogy with studies of the corresponding networks in model organisms.

The most critical opportunities:

- To understand a biological system it is critical to move beyond a focus on its individual components to the next level of complexity, i.e. an understanding of how the components assemble into functional units.
- Biomolecules for which a three-dimensional structure can be obtained represent particularly robust building blocks for starting systems-level integration.
- Experimentally-derived multi-scale structural information will enable the objective and reliable visualization of cellular organization from the molecular to the systems level.
- A structure-based scaffold will provide an ideal platform for integrating available biological data and will be synergistic with other systems biology efforts.
- Using this structure-based platform will be a particularly effective means for validating and interpreting genetic variation as it relates to disease and will guide more informed and precise therapeutic interventions.

State of the art

Many problems of key importance to human quality of life, health, economic development, agriculture and the environment depend critically on being able to predict the behavior of complex biological networks in the healthy organism as well as how they are perturbed in disease. Being able to predict the behavior of these networks is key both to tailoring available therapies to individual patients and to developing new therapies. The essence of systems biology lies in its ability to formulate these predictions. Proteins constitute a fundamental element of biological systems. The comprehensive analysis of their structure, location, function and mechanism of action, as well as of their interactions with each other and with other components of the cell, provides information crucial to the development of systems biology.

A long term goal of the Life Sciences is the complete comprehension, at the molecular level, of the processes at the basis of Life, which are essential for living organisms to survive and account for the function or malfunction of cells and tissues. Molecular Systems Biology is an integrative discipline that has evolved out of the need to describe the behavior of complex biological systems in terms of their molecular components and interactions; it provides a platform for data collection, data analysis, integration of data from various sources, and modeling of molecular and cellular phenomena.

Historically, molecular biology has progressed through the identification of individual genes and proteins and the study of their individual functions - components of biochemical pathways have been analyzed as though they were independent pieces of a larger puzzle. This approach imposes limitations, however; an organism is more than the sum of its individual functional processes, and each process is affected by all of the others. In the case of the human, we are left with an unclear picture of how the human body functions, and how we can best approach attempts to predict, prevent, and treat health problems. Efforts to cure complex diseases have met with limited success because only individual aspects of the organism have been studied at one time. Molecular Systems Biology approaches the study of an organism as an integrated and interacting network of genes, proteins, and biochemical reactions – and it is now well recognized that it is this *system* that gives rise to and maintains Life. Individual functional processes must therefore be studied in the context of an entire cell or organism, and not in isolation. It is the *interactions* that are ultimately responsible for the form and function of any organism.

Structural genomics efforts have provided a large number of structures that can be used for the next level of integration. In parallel, it has been possible to obtain a first coarse map of the cellular machinery of the model organism yeast by purification of complexes and mass spectrometry. Several additional technological platforms are available to systems biology, such as proteomics, metabolomics, etc. Taken together, these provide the basis for 3D systems biology.

The frontiers of Molecular Systems Biology

In order to achieve the goals outlined above, we must obtain a detailed description of the molecular components, their in-

teractions, and abundances under a wide range of cellular and *in vitro* conditions. To this end, we require a set of new methodologies and technologies, and the acquisition of appropriate large-scale datasets. Specifically:

- Systematic analysis of structures and interactions of proteins, protein-protein and protein-nucleic acid complexes.
- Innovative computational and experimental approaches to studying the effect of variations in the sequence and of post-translational modification on the structure and function of proteins and their complexes, e.g. somatic mutations in cancer, by experimental and computational approaches.
- Development of tools for measuring the often transient interactions between components of the complex systems involved (including dynamics and structural changes affecting function).
- Development of novel knowledge management systems capable of describing the components of such complex systems and the assurance of their reliability, for the purpose of being incorporated into models.
- Systematic and parallel *in vivo* and *in vitro* studies of protein interactions and biochemistry at intermediate levels of system complexity.
- Elucidation and modeling of the principles underlying signaling circuits.
- Analysis of host-pathogen interactions
- Understanding microbial communities and interpreting metagenomic data.
- Development of models incorporating dynamic, structural and mutational information able to predict interactions in biological systems.
- Identification of therapeutic opportunities for treating common human diseases.

The role of Structural Biology from a Systems Biology perspective

A cell may be compared to a car, for which, in order to understand its workings, we need to examine and understand the component subsystems, e.g. the gear train, transmission, fuel system, etc. Structural biology can make a crucial contribution to analyzing these subsystems, ultimately working towards understanding how proteins function in an *in vivo* context. A goal would be to obtain a mathematical description of networks and systems with the ultimate goal of achieving predictive understanding of a system or of a subsystem. This encompasses dynamic, often transient, interactions in a cellular context, and examining the nature of response to structural disorder in such interactions, moving away from solely looking at ordered domains. The goal is an understanding of the impact of these features on cellular systems and their role in disease mechanisms, e.g. in Alzheimer's, prion and other neurodegenerative diseases, in diabetes, cardiovascular disease, infectious diseases, both bacterial and viral, and in cancer.

We need to emphasize steps forward in key structural biology techniques to facilitate this. This requires continued development of individual technologies, but, importantly, also their key synergistic implementation. To combine and integrate the different techniques, we need advances in methods but also improved and new computational tools. A key objective will be to integrate the whole system as a function of time. Quantitative information is crucial for systems biology.

Important techniques where further improvements will be necessary span many areas, from structural to computational biology and from mass spectrometry to knowledge management.

Examples include: 1) X-ray crystallography (Remote access; automatic analysis of maps for ligand recognition; improvement in collection of data on microcrystals); 2) NMR (*in vivo* methods; transient interactions; disordered systems; solid state methods; improved computational methods); 3) Electron microscopy (Single particle cryo EM for large complexes; cryo-electron tomography of cellular structures); 4) SAXS; 5) Protein production

Impact of Molecular Systems Biology on molecular medicine

The following opportunities will be attainable through a Molecular Systems Biology approach:

Understanding Disease

- Systems level understanding of genetic perturbation reflecting disease. Efforts from medical genetics and mouse models of disease will become more easily interpretable.
- The nature of many disease mutations will become interpretable by mapping effects of the molecular defects on

3D protein complex structures, including inter-complex protein-protein interactions.

- Understanding the mechanisms of complex diseases will profit from integration of data at the level of the molecular machine, of the entire pathway and at the inter-pathway level.
- Extrapolation of mutant-protein-effect relationships from a validated system to other diseases becomes possible.

Treating Disease

- Entire disease pathways in three-dimensions will instruct the identification and use of research compounds, interfering peptides and drug leads.
- "Poly-pharmacology" (off-target effects) may be turned into "Systems Pharmacology" (where multiple targets are hit on purpose).
- The 3D systems biology understanding of disease may allow the informed treatment of multifactorial diseases and the use of combination therapy.
- More efficient target and scaffold hopping (the same drug class targeting a particular protein class can be used to hit an analogous target and the same target may be hit by unrelated chemical scaffolds).

The recommended research path to achieve the above goals should comprise the following steps:

Identification and structural characterization of individual elements → correlation of structure-function relations of these elements → experimental super-structural studies → 3D modeling → experimental perturbation of the system by mapping disease mutations → modeling and simulation to predict the molecular disease mechanism(s) and to propose drug targets / diagnostic markers.

These steps would have as their foundation and strength:

- Build on the continuation of the structural genomics effort to deliver the shapes and structures of components to enable 3D systems biology.
- Map human cellular machinery.
- Obtain hybrid structural information from patchwork analysis of suitable technologies.
- 3D computational modeling.
- Integrate experimental studies, disease genetics and pharmacological perturbation of the system of interest.
- Modeling/simulation of diseased states. Prediction of diseased network output and small molecule effect.

Appendix: Available funding instruments in the EU *Within the 7th Framework Programme*:

- Cooperation «: Health 3rd call 2008, 4th call 2009
  - ICT programme: (bioinformatics, databases)
- Capacities «: Research infrastructures: resources, facilities and related services aimed at top-level research
- Support to existing research infrastructures provided for Integrated Activities (bottom-up and targeted approaches) and e-Infrastructures
- Support to new research infrastructures provided for design studies and construction of new infrastructures
- Support for policy development and program implementation

ERA-NET schemes permit coordination between relevant national research programmes of Member States.

International dimension

EU-USA interaction

- Project participation level
  - FP7: US partners can participate in proposals, but, in most cases, cannot receive EU funding
  - NIH: EU and Canadian partners could be funded (like EBI) based on a decision from the appropriate governing board, or could be subcontractors of US partners
- Programme cooperation and coordination level
  - Project cooperation and coordination: example of PSI: international cooperation with Wellcome Trust (UK) established: 2000 Cambridge meeting (tackled: international coordination; policies on data-release, international coordination of target selection, facilities)
  - IKMC: international knock-out mouse consortium – running since spring 2007: EC-NIH-Genome Canada

- ICGC: International Cancer genome consortium – in preparation

EU -Third countries

Bilateral cooperation agreements: (EU-China, EU-India, EU-Russia): areas identified; implementation through » SICA « calls in FP7 (3rd call: Health, topics with China and Russia foreseen)

Mechanisms for identifying joint areas of cooperation at the international level

Agreements (top-down)

EC-US Task Force on biotechnology (bottom-up); workshops on yearly basis in selected areas, produce policy documents for joint actions (Workshop on infrastructures in Systems biology, May 2007)

Ad-hoc initiatives: Systems biology of cancer (EC-US) May 2008

EU: Identification of inputs to the creation and "fine tuning" of FP7

- Legal procedure for adoption of FP7 (codecision between Council and European Parliament)
- Inputs to annual work programme (scientific topics and instruments: how are the broad topics defined for FP7?)
  - Advisory groups for health research and research infrastructures
  - Input from on-going projects: SSA, CA (FESP)
  - Programme committees (Member States)
  - International dimension: implementation of political initiatives from agreements between EC and other countries
  - Addressing the needs of emerging international consortia (Mouse genome)
  - Top-down (FP7) vs. bottom-up (expression of interest; 2-stage procedure)

Choice of instruments: depends on the scale and ambition of the issue addressed

The situation in the USA (NIH)

The PSI is ongoing: large-scale initiatives started in late 90's. At present, the 2nd stage of the PSI is focused on selected targets in order to achieve coverage of large protein families. In the future a possible target will be the human gut microbiome, but this may depend on a decision with respect to funding of the 3rd stage of the PSI, which should be made in early 2008.

Systems biology: 6 National centers for systems biology active: interdisciplinary, based on collaborative research

PSI-systems biology: possibilities for interaction will be explored in the future (e.g. interactions with the Functional Glycomics Center, the National Center for Research Resources-NCRR, Synchrotrons, the National Cancer Institute-NCI)

US-Third countries: - policy initiatives

NIH -Wellcome trust; Japan (RIKEN); China – ISGO conferences

Official agreements with China, Japan, India. . .

Research Infrastructure financing to participants from third countries is possible, depending on evaluators.

Acknowledgements

The authors gratefully acknowledge the valuable presentations and discussions at the symposium on "New Challenges in the Life Sciences: Prioritizing European Research in Molecular Systems Biology" held in Florence, Italy on October 18-19, 2007 with special thanks to the following people: Lucia Banci[a], Wolfgang Baumeister[b], Ivano Bertini[c], Roslyn M. Bill[d], Steven E. Brenner[e], Bruno Canard[f], John Christodoulou[g], Bruno Domon[h], Peter Henderson[i], Sasa Jenko[j], Patrik Kolar[k], Andreas Kremer[l], Peter Kuhn[m], Hans Lehrach[n], Zhi-Jie Liu[o], Claudio Luchinat[p], John Moult[q], Pär Nordlund[r], John Norvell[s], Carlo Rizzuto[t], Harald Schwalbe[u], Israel Silman[v], Lalji Singh[w], David Stuart[x], Joel Sussman[y], Purificacion Tejedor Del Real[z], Georg Terstappen[aa], and Shigeyuki Yokoyama[bb].

[a] Magnetic Resonance Center (CERM), University of Florence and FESP
[b] Max Planck Institute for Biochemistry, Martinsried
[c] Magnetic Resonance Center (CERM), University of Florence
[d] Aston University, Birmingham
[e] University of California, Berkeley
[f] CNRS, Université de Provence Université de la Méditerranée
[g] University of Cambridge
[h] Institute of Molecular Systems Biology, ETH Zurich
[i] University of Leeds, Coordinator EU EDICT project
[j] European Commission, Directorate General for Research Genomics and Systems Biology
[k] European Commission, Directorate General for Research Genomics and Systems Biology
[l] SienaBiotech SpA
[m] The Scripps Research Institute, La Jolla
[n] Max Planck Institute for Molecular Genetics, Berlin
[o] Chinese Academy of Science, National Laboratory of Biomacromolecules
[p] Magnetic Resonance Center (CERM), University of Florence
[q] Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute
[r] Karolinska Institutet, Stockholm, SGC Consortium
[s] NIH, Director of the Protein Structure Initiative
[t] Chair European Strategy Forum on Research Infrastructures (ESFRI)
[u] Biological NMR Group, Johann Wolfgang Goethe University, Frankfurt
[v] Weizmann Institute of Science, Rehovot
[w] Director of the Centre for Cellular & Molecular Biology, India
[x] Oxford University, Coordinator of SPINE2
[y] Coordinator Forum for European Structural Proteomics (FESP)
[z] European Commission, Directorate General for Research, Research Infrastructures (Unit B3)
[aa] Siena Biotech SpA
[bb] RIKEN, Protein Research Group

Chapter 2

The strategic role of NMR in the study of protein-protein interactions

Authors: Lucia Banci[a], Rolf Boelens[b], Andres Ramos[c], Michael Sattler[d], Harald Schwalbe[e], Vladimir Sklenar[f], Marcellus Ubbink[g]
[a] Magnetic Resonance Center, Florence, Italy
[b] Bijvoet Center for Biomolecular Research, University of Utrecht, The Netherlands
[c] National Institute for Medical Research, Medical Research Council, London, UK
[d] Institute of Structural Biology, Helmholtz Zentrum München, Germany
[e] Biological NMR Group, Johann Wolfgang Goethe University, Frankfurt
[f] National Center for Biomolecular Research, Masaryk University, Brno, Czech Republic
[g] Institute of Chemistry, Leiden University, The Netherlands

Executive Summary

Proteins are the key players mediating biological function in living systems. At a molecular level biological function depends on precise interactions with other proteins, nucleic acids and small molecules/metabolites. These interactions are tightly regulated in the cell and are the molecular basis for biological activity. Defects in these processes can be the cause of severe diseases. Protein recognition inside a cell involves multi-component protein complexes, which are often dynamic in their assembly and constitution.

Many essential processes such as the regulation of gene expression at the level of DNA, chromatin and RNA or cellular signaling rely on multi-domain proteins, where protein-protein recognition modules are connected by flexible linkers. The structural flexibility of proteins is coupled to their regulatory functions and represents a primary feature essential to fulfilling their biological activity.

In order to gain molecular insight into the cellular functions it is crucial to describe not only the three-dimensional structures of protein complexes but also their temporal variations. Understanding the spatial and temporal dynamics of proteins is therefore a key requirement for a systems description of biology. NMR is a powerful and reliable tool to validate and quantify protein-protein interactions predicted by HTP methods. It is also the sole method that can localize binding sites on proteins in solution, which is an essential element for the description of complex protein interaction networks. Moreover, NMR is able to identify weak and transient interactions, which are ubiquitous in the crowded environment of a cell, but generally ignored by other approaches. Thus, NMR has important roles in providing molecular and structural details for Mechanistic Systems Biology.

State-of-the-art

Biomolecular NMR spectroscopy is the key technique for studying the molecular mechanisms underlying cellular function, since it can determine both the structure and dynamics of protein complexes under near physiological conditions. In particular, NMR plays a key role as it is an ideal method for studying weak and transient interactions, which are intimately linked to biological function. Transient interactions are quite common, as they represent the different steps of a functional process. In order for the process to proceed forward, the transient complexes must break apart and their components must rearrange. NMR is unique in its potential for studying the dynamical features of the interactions as well as the structures of the interacting molecules. NMR can characterize the oligomerization state of proteins, for example, in receptor oligomerization, which plays a critical role in cellular signaling. Furthermore, NMR is invaluable for characterizing those complexes involving partially or completely unstructured proteins, which cannot be studied at the atomic level by other techniques.

The importance of NMR in the study of protein/protein interactions has grown in parallel with the evolution of its designated role within the strategic planning of Structural Biology/Genomics projects in both the US and in Europe. NMR spectroscopy is a key player in the FP6 EC-funded SPINE2-COMPLEXES project. Indeed, NMR plays an increased role from the one it had in the previous FP5 SPINE project. The involvement of NMR in SPINE2-COMPLEXES builds on two competences: (i) methodological developments for determining the structure

of protein-protein complexes, and (ii) the structural characterization of selected protein-protein complexes of high biological relevance. As stated in the description of SPINE2-COMPLEXES, the project has:

"A strong and critical NMR component both for target-related structure determination (where it is suitable for complementary samples unsuitable for crystallography) and in methods development, as this is not funded elsewhere by FP6."

NMR is also contributing to the 3D Repertoire Project, which aims at describing the structures of protein complexes in the cell. Here, NMR is used i) to map protein-protein interfaces and ii) to determine the structures of dynamic protein complexes that are difficult to crystallize or may crystallize in a non-biological conformation.

A further step in defining the key role of NMR in the characterization of functional processes is its role in the European Research Infrastructure INSTRUCT. The INSTRUCT project is part of the ESFRI Roadmap, which strives to respond to the foreseen long-term research needs of the European scientific community. The INSTRUCT infrastructure is composed of seven Core Centers and a number of Associate Centers, and will provide a platform to link information obtained by the major structural biology methods with state-of-the-art cell biology techniques to provide a dynamic picture of key cellular processes at all scales.

Aspects of studying protein-protein interactions by NMR

The methodological aspects of the application of NMR spectroscopy to the study of protein-protein interactions have been extensively analyzed in the frame of the Coordination Action NMR-Life. Two large meetings (Utrecht, June 2006 and Murnau, October 2008), and two workshops on NMR sample preparation (September 2007 at EMBL Heidelberg and July 2008 at the Bavarian NMR Centre in Munich) were organized to evaluate the current state of biomolecular NMR with respect to protein-protein interactions and to identify areas where future methodological advances may further expand the applicability of NMR. The discussions and conclusions from these events are the basis for this document.

Studying transient/weak interactions in regulatory systems

The unique potential of NMR for studying weakly interacting, transient protein complexes has been demonstrated in recent years. Unprecedented new insight was gained for various multi-domain proteins that play crucial roles in intra- and extracellular signaling, the regulation of gene expression, chaperone-cochaperone complexes in kinase folding, and on metal-mediated transient protein-protein interactions, to name a few. In many cases, structural analysis by X-ray crystallography was not possible since crystallization failed or the crystallized protein constructs were strongly influenced by crystal packing. In others, although X-ray was successful in providing a structural view, NMR was necessary to relate it to the functional details of the post-translational regulatory mechanism.

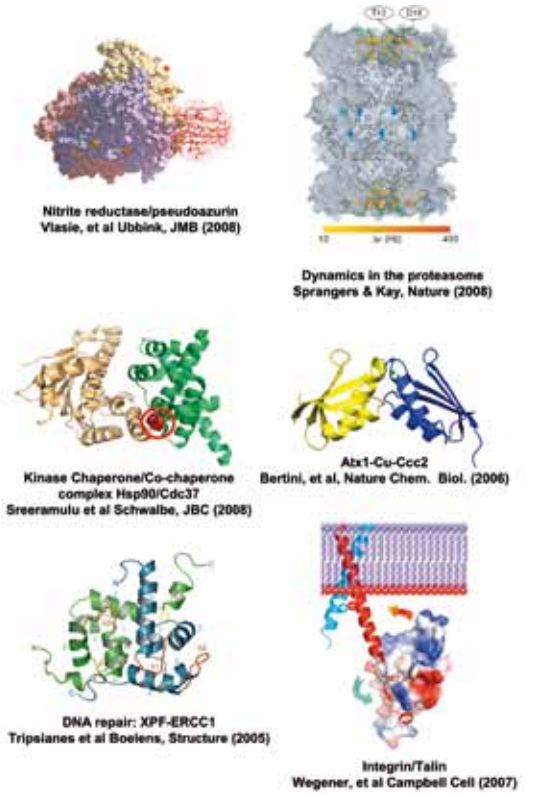


Figure 1: Examples of structure and dynamics in protein-protein complexes studied by NMR [1-5]

NMR methods

NMR methods for high-molecular weight protein-protein complexes

In the last decade, major breakthroughs have contributed to the potential for using NMR for studying very high-molecular weight protein-protein complexes such as the GroEL/GroES

co-chaperone [1] or the proteasome [2]. This potential is based on the use of optimized isotope-labeling schemes that leave only a few protons in an otherwise fully deuterated background [3]. Optimized NMR pulse sequences are available for these experiments [4]. Studies of such high-molecular weight systems are normally done in conjunction with an available crystal structure of the system and focus on elucidating ligand binding and protein dynamics.

NMR approaches to study protein-protein interactions mediated by cofactors/metals

In the last years it has become evident that a number of protein-protein interactions are not driven by direct molecular recognition but are mediated by the shared coordination of a cofactor such as a metal ion. NMR is the technique of choice for the characterization of the structural and dynamical properties of such interactions, as they are usually weak and transient [5]. For these systems strategies should be developed that integrate NMR with other approaches to characterize the protein-cofactor region. Optimized strategies have also been developed to study oxidation-linked protein complexes.

NMR tools for mapping protein interfaces

NMR is an excellent method and is routinely used for mapping the interaction surface of a protein with ligands, such as proteins, nucleic acids and small molecules [6]. Residue-specific information can be obtained in a fast and reliable manner and this often suggests key sites for mutational analysis. The simplest mapping technique is based on chemical shift perturbations (CSPs). This procedure typically involves the acquisition of NMR spectra of an isotope-labeled protein alone and when bound to an unlabeled ligand. The chemical shift perturbations provide a per residue characterization of the protein-protein interface, and give unique opportunities to study local effects and conformational dynamics that are associated with ligand binding. CSPs also allow quantification of the binding affinities. One drawback of CSPs is that they may also reflect indirect effects linked to structural changes induced by ligand binding.

An alternative approach to overcome this bottleneck is the use of through-space relaxation effects to map binding interfaces. This involves cross saturation transfer (CST) experiments and quantification of these effects [7]. CST can resolve ambiguities from CSP analysis but requires specific differential isotope labeling of both binding partners.

Structural analysis of protein complexes in solution

Paramagnetic relaxation enhancements (PRE) have been used as a tool for defining the structure and dynamics of protein-protein interactions [8-10]. Examples range from the use of PREs for the structure determination of proteins, including small membrane proteins and multi-domain proteins, up to the detection and structural characterization of transient, weakly populated encounter complexes. Solvent PREs can be obtained from using paramagnetic cosolvents that screen the surface of a protein and allow the identification of binding interfaces.

More generally, anisotropic effects derived from partial alignment in anisotropic solution or from anisotropic paramagnetic ions are very useful since they provide long-range distance and orientational information from PRE, pseudo contact shifts (PCS) and residual dipolar couplings (RDCs) [11]. These long-range effects augment the more local information (distance restraints) typically obtained from NOE-based experiments by one order of magnitude. To make these systems available for general studies of protein complexes, paramagnetic tags are employed. These can include spin labels, which provide only PRE, or lanthanide (metal) binding tags, which give structural information from PCS and RDC data as well. Lanthanide-binding tags are either cross-linked to cysteine residues in a protein or they can be genetically encoded for expression with the protein of interest. The use of different lanthanides provides an extendable ruler that can define interactions up to 70 Å, and is useful to define assembly in transiently assembled multi-component systems.

In cell NMR spectroscopy

The utility and potential of monitoring protein-protein interactions, for example to follow protein phosphorylation in living cells, has recently been demonstrated [12]. This emerging area of research is expected to have a major impact that will contribute to NMR's unique role in Mechanistic Systems Biology. Specific advantages of the technique are that protein-protein interactions can be monitored inside a living cell at residue-level. Moreover, the course of events such as phosphorylation can be followed in real time. This allows, in principle, the detection of the time course of phosphorylation cascades that play central roles in cellular signaling.

Sample preparation

Bacterial expression in *E.coli* is the most commonly used approach for NMR sample preparation. Well-established protocols and a wide range of expression vectors are available for this purpose. However, many highly interesting proteins cannot be expressed and isotope-labeled in *E.coli*. This will be a future challenge for NMR sample preparation.

Utility of non-bacterial hosts for sample preparation

Higher organisms, i.e. insect cells or mammalian cell lines, are not frequently used as expression hosts for NMR sample preparation. Examples have been reported in the production of isotope labeled proteins in SF9 cells and in HEK293 proteins [13,14]. Such expression systems are particularly important for the production of membrane proteins with native glycosylation status. The drawbacks of using insect cells or mammalian cell lines are the prohibitive costs of the incorporation of isotope labels in expression proteins in these hosts. Some recent progress and applications have been reported, for example, in the isotope labeling of protein kinases and G-protein coupled receptors (GPCRs), though further improvement of these techniques will be required for them to be more generally applicable.

Two CA workshops held at EMBL Heidelberg in September 2006 and at the TU Muenchen in July 2008 explored the potential of using non-bacterial expression hosts. The expression hosts *Dichtyostellim discoideum* and *C. elegans* can provide nice and economical means for the expression of some difficult target proteins. However, the expression protocols need to be further developed to be used in the preparation of isotopically labeled samples for NMR studies. One possibility is the use of waste cell pellets from isotopically labeled cultures to cultivate *C. elegans*. Efficient growth media should be developed for a more general use of the *Dichtyostellim discoideum* system in NMR sample preparation. Low cell density and batch-to-batch variations are limitations of using recombinant protein production in shake flasks. The Enbase technology allows a controlled growth of cells to high cell densities. It is also an efficient method for the preparation of ^{15}N labeled samples, but it still needs to be further developed for the preparation of other isotopically labeled samples.

In vitro expression

In vitro expression systems based on bacterial and wheat germ extracts are available for cell free expression and have been successfully used. Here, substantial progress has been made in the expression of membrane proteins [15]. *In vitro* expression can be well combined with SAIL isotope labeling [16], which potentially enables detailed NMR studies of protein complexes. However broad application of this technology still requires a cost-effective production of the SAIL amino acids.

Segmental labeling and posttranslational modifications

In addition to optimized NMR sample and isotope labeling strategies, additional protocols should be established for adding correct disulfide bonds, methylation, phosphorylation, glycosylation and other posttranslational modifications for NMR studies of proteins. This is particularly important as NMR is the method of choice in the study of the conformational rearrangement associated with the phosphorylation and methylation phenomena taking place in signaling cascades and during the regulation of gene expression.

Segmental isotope labeling is a powerful tool to reduce signal overlap in NMR spectroscopy of large proteins. Methods for protein ligation, such as native chemical ligation *in vitro* or intein mediated ligation *in vivo* are available [17;18]. More efficient protocols should be developed to make these techniques generally available for NMR studies of multi-domain proteins.

Computational tools

Docking driven by NMR data

The great utility of NMR for the efficient detection and mapping of molecular interactions of proteins is well recognized. Data from chemical shift perturbations can be readily employed to derive experimentally determined models for protein complexes, which provide already useful structural information and can be used for mutational analysis of protein function in biological assays. For such studies, computational tools are available; most prominently the HADDOCK program [19] is now being used by both the NMR community but also non-expert biologists, documenting the utility of these approaches.

Structures from chemical shifts

Chemical shifts contain a wealth of information about the structure and dynamics of biological macromolecules. Recently, initial reports have appeared which demonstrate that using ^{13}C (secondary) chemical shifts can be combined with structure prediction programs to obtain three-dimensional folds of small globular domains [20;21]. Software packages are available for these approaches, but further improvements are anticipated to optimize the speed, reliability and accuracy of the predictions. Nevertheless, this information can be readily exploited to aid studies of protein-protein interactions.

Multidisciplinary/complementary approaches

NMR data can be combined with data from complementary techniques to provide a comprehensive description of protein-protein interactions and to calculate structures of multi-domain proteins and protein complexes [23;24]. One such approach uses the combination of distance and orientational restraints from paramagnetic tagging and/or spin labeling with data from Small Angle X-ray and/or Neutron Scattering (SAXS/SANS) [22-25]. See Chapter 8 of this volume for more information.

Conclusions and future strategies

Discussions amongst colleagues in the NMR community during workshops and in the general frame of the CA NMR-Life have reached a consensus that NMR has an invaluable role in studying various aspects of mechanistic systems biology. The unique strengths of NMR for these studies that are not matched but are very complementary to other structural biology techniques are:

- Mapping binding interfaces for protein-protein interactions and characterizing the competition and cooperativity of interactions.
- Determining the structures and dynamics of (large) multi-component protein complexes using NMR solution data, also by employing multidisciplinary hybrid approaches. NMR is unique in that it provides structural information for individual components (or domains) in a multi-component system and in that it can follow the changes in dynamics upon complex formation.
- Studying transient and weak interactions, which are critical for the regulation of cellular processes.

In addition, the following topics have been identified for further developments to enhance NMR studies of protein complexes:

- Sample preparation, expression hosts and isotope labelling are critical issues, and frequently the bottleneck in the application of NMR for studying biomolecular complexes. Some issues may be solved by efficient high-throughput methods that could be economically and reliably run in expert facilities.
- Improvements in hardware (ultra-high magnetic fields, cryoprobes) should further enhance the application of NMR in studying protein-protein interactions.
- New computational methods for data analysis and structure calculations. It will become increasingly important to develop robust computational and analysis methods that produce reliable results with high efficiency. For this, and also for the potential further analysis of precious unique data, it must become possible to collect the raw NMR data and protocols in central databases.
- Applications of NMR to the study of protein-protein and protein-nucleic acid complexes of medical importance should be strengthened. This includes the analysis of misfolded, partially folded or intrinsically disordered proteins. Studies of large multi-component complexes involving multi-domain proteins can be best analyzed with high field NMR instruments. NMR has a unique role in characterizing dynamic interactions and transient complexes. It was suggested that NMR facilities should provide support for the biochemical community in all aspects, including data acquisition, sample preparation and structural analysis to enhance the biological impact of NMR studies on the structure and dynamics of protein-protein interactions.

Complementary methods, like SAXS, crystallography, electron microscopy and mass spectrometry are increasingly used together with NMR spectroscopy. This demonstrates the unique capability of NMR to analyze the dynamic nature of molecular assemblies. It is important to stimulate this integration of different methods, not only for reasons of better understanding complex biology but also for improved robustness and efficiency. Such studies will also emphasize the unique role of NMR in structural biology. Possibilities for integration can be created by stimulating conferences, scientific exchange between complementary research centres, collaborative projects and by integrating activities.

References

(1) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K. *Nature* **2002**, *418*, 207-211.
(2) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445*, 618-622.
(3) Tugarinov, V.; Kanelis, V.; Kay, L. E. *Nat Protc* **2006**, *1*, 749-754.
(4) Reik, R.; Pervushin, K.; Wüthrich, K. *Trends Biochem Sci* **2000**, *25*, 462-468.
(5) Banci, L.; Bertini, I.; Cantini, F.; Felli, I. C.; Gonnelli, L.; Hadjiliadis, N.; Pierattelli, R.; Rosato, A.; Voulgaris, P. *Nat.Chem.Biol.* **2006**, *2*, 367-368.
(6) Zuiderweg, E. R. *Biochemistry* **2002**, *41*, 1-7.
(7) Takahashi, H.; Nakanishi, T.; Kami, K.; Arata, Y.; Shimada, I. *Nat. Struct.Biol.* **2000**, *7*, 220-223.
(8) Mackereth, C.D.; Simon B.; Sattler M. *Chembiochem* **2005**, *6*, 1578-1584.
(9) Clore, G. M.; Tang, C.; Iwahara, J. *Curr Opin Struct Biol* **2007**, *17*, 603-616.
(10) Vlasie, M. D.; Fernández-Busnadiego, R.; Prudêncio, M.; Ubbink, M. *J.Mol.Biol.* **2008**, *375*, 1405-1415.
(11) Otting, G. *J Biomol NMR* **2008**, *42*, 1-9.
(12) Selenko, P.; Frueh, D. P.; Elsasser, S. J.; Hass, W.; Gygi, S. P.; Wagner, G. *Nat Struct Biol* **2008**, *15*, 321-329.
(13) Nisius, L.; Rogowski, M.; Vangelista, L.; Grzesiek, S. *Protein Expr. Purif.* **2008**, *61*, 155-162.
(14) Werner, K.; Richter, C.; Klein-Seetharaman, J.; Schwalbe, H. *J Biomol NMR* **2008**, *40*, 49-53.
(15) Schwarz, D.; Klammt, C.; Koglin, A.; Löhr, F.; Schneider, B.; Dötsch, V.; Bernhard, F. *Methods* **2007**, *41*, 355-69.
(16) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Guntert, P. *Nature* **2006**, *440*, 52-57.
(17) Cowburn, D.; Shektman, A.; Xu, R.; Ottesen, J. J.; Muir, T. W. *Methods Mol.Biol.* **2004**, *278*, 47-56.
(18) Zuger, S.; Iwai, H. *Nature Biotechnology* **2005**, *23*, 736-740.
(19) Dominguez, C.; Boelens, R.; Bonvin, A. M. *J.Am.Chem.Soc.* **2003**, *125*, 1731-1737.
(20) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl.Acad.Sci.U.S.A* **2007**, *104*, 9615-9620.
(21) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc.Natl.Acad.Sci.U.S.A* **2008**, *105*, 4685-4690.
(22) Mattinen, M. L.; Paakkonen, K.; Ikonen, T.; Craven, J.; Drakenberg, T.; Serimaa, R.; Waltho, J.; Annila, A. *Biophys.J.* **2002**, *83*, 1177-1183.
(23) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J.Am.Chem.Soc.* **2005**, *127*, 16621-16628.
(24) Gabel, F.; Simon, B.; Sattler, M. *Eur.Biophys.J.* **2006**, 1-15.
(25) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl.Acad.Sci.U.S.A* **2008**, *105*, 5762-5767.

Chapter 3

Advances in studies of protein-nucleic acid complexes

Authors: Rolf Boelens[a], Michael Sattler[b], Harald Schwalbe[c], Vladimir Sklenář[d], Richard Štefl[d], Lukáš Žídek[d]

[a] Bijvoet Center for Biomolecular Research, University of Utrecht, The Netherlands

[b] Institute of Structural Biology, Helmholtz Zentrum München, Germany

[c] Biological NMR Group, Johann Wolfgang Goethe University, Frankfurt

[d] National Center for Biomolecular Research, Masaryk University, Brno, Czech Republic

Executive Summary

It is increasingly recognized that nucleic acids, especially RNA, and their complexes with proteins, are of utmost importance in the description of human life [1]. Transcription, translation, and replication of nucleic acids are key dynamic cellular processes that require tight control by proteins. Mechanistic defects therein are the cause for severe diseases and reduced life span. During evolution, cells have developed a repertoire of repair processes, in which protein – nucleic complexes play a crucial function. Novel roles of nucleic acids are constantly being discovered; recently, for example the role of microRNAs and other non-coding RNAs [2]. Regulatory processes help to maintain functional states of cells in the complex environment of the biological system. Allosteric conformational transitions between alternate conformations represent key elements in regulation circuits, both in prokaryotes and eukaryotes. RNA mediated regulation in cells, in fact, entails a complex network of dynamic molecular interactions involving protein-RNA complexes. Their proper description represents a key element to gain understanding of the development of multi-cellular organisms and processes essential in the maintenance of Life.

Structures of protein-nucleic acid complexes are usually highly dynamic, characterized by intermolecular contacts mediated by very weak atomic interactions. NMR spectroscopy has an unparalleled capability to characterize not only the spatial structures of biological macromolecules at the atomic resolution but also to map their temporal variations across a time range spanning 14 orders of magnitude. NMR is unique in its capacity to provide information necessary for understanding regulatory functions. A review of accomplishments achieved during the past few years clearly shows that NMR is playing an ever increasing role in analyzing mechanisms studied by systems biology. European initiatives should focus on fostering advances in mechanistic systems biology using NMR as a major technology and a key asset.

State of the art

A vast amount of biological activity is triggered and regulated by mutual interactions between proteins, nucleic acids, and other biological macromolecules. Nucleic acids and proteins work together to maintain, replicate, transcribe, and repair the genetic code. It is now also understood that RNA can act alone, and various types of so-called non-coding functional RNAs have come into the focus of biomedical research. More recently, for example, it has become clear that microRNAs and other non-coding RNAs including Xist, Air, Rox, HAR and riboswitch-RNAs to name just a few play important roles in the regulation of gene expression and some even control the development of multi-cellular organisms. During the past decade, structural biology has reached a level of maturity which enables identification of individual atoms within large bio-macromolecular assemblies at very high resolution. Analysis of intermolecular interactions and their dynamics at atomic resolution reveals the mechanisms of biomolecular recognition and function, and represents the major approach to rationalize biological activity.

Currently, two major experimental techniques provide high resolution three-dimensional structures of biological macromolecules and their complexes; namely single crystal x-ray diffraction and NMR spectroscopy. As of November 2008, the majority of protein structures in the PDB represent free proteins. In fact, only 2206 structures out of 54,599 files deposited in the PDB, i.e. around 4%, concern protein - nucleic acid (protein-NA) complexes. Thus our mechanistic understanding of

most protein-NA complexes is highly limited and only partially based on precise observations. The small number of deposited structures clearly indicates that methodological obstacles accompanying the studies of molecular assemblies held together by weak intermolecular interactions are still severely restrictive. NMR spectroscopy has contributed 140 three-dimensional structures, 91 of protein-DNA and 49 of protein-RNA complexes.

Growth of the total number of all protein-nucleic acid structures determined using various experimental methods deposited in the PDB is documented in Figure 1. It is evident that technology developments of the structure determination of protein-nucleic acid complexes during the past few years have led to a significant increase in the amount of available data. To the total number of protein-NA structures deposited in the PDB, NMR spectroscopy has contributed 6.3%. As documented in Figures 2 and 3, yearly growth of deposited NMR structures varies significantly, with an average of 10 structures per year between 2000 and 2008. In contrast, X-ray crystallography has shown a steady increase in the number of yearly determined structures, growing from 86 to 267 in the same period. Considering the much higher relative success in determining the 3D structures of individual proteins and nucleic acids, with annual contributions by NMR reaching 15% and 44%, respectively, it is evident that the potential of NMR methodology for the structure determination of protein-NA complexes is still awaiting its full exploitation. As indicated in the remainder of this document there has recently been considerable progress in NMR studies of large biomolecular complexes that may enable this.

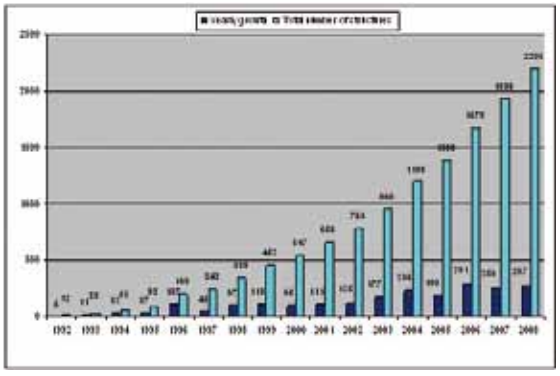


Figure 1: Number of structure of protein nucleic acid complexes determined using various experimental methods deposited in PDB

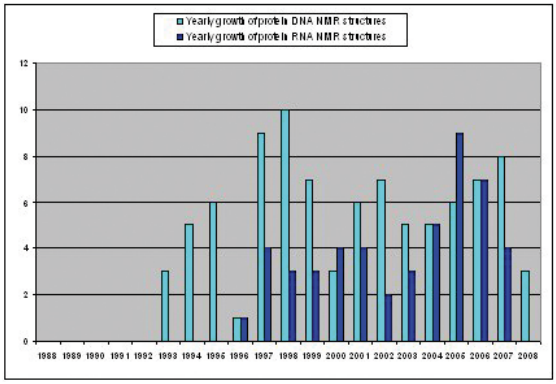


Figure 2: Yearly growth of protein nucleic acid complexes deposited in the PDB determined using NMR

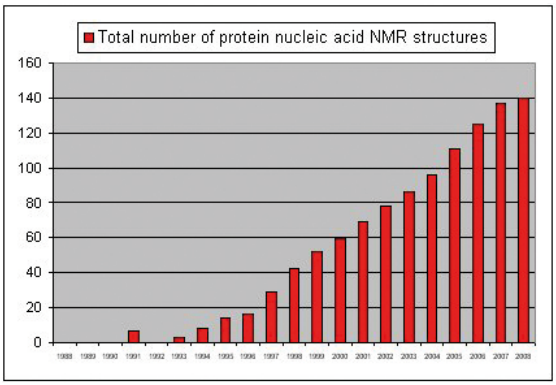


Figure 3: Total number of structures of protein-nucleic acid complexes deposited in the PDB determined using NMR

To illustrate the potential of NMR technology, two recently determined solution structures of protein-DNA and protein-RNA complexes are shown in Figures 4 and 5, respectively. They demonstrate that despite all the existing obstacles, NMR can provide unique structural information describing transient intermolecular interactions responsible for complex formation in solution in detail.

In functional states, nucleic acids often have complex and highly dynamic structures. During protein-NA complex formation, a selection chooses one or more of those states. This intrinsically highly dynamic process is hard to analyze; reduction to a single, static picture, as provided by both X-ray crystallography and NMR, thus reveals a wealth of valuable information. However, static pictures can lead to oversimplified views and generally only partially correlate with biochemical data. Even highly specific and tight complexes, as observed in repressor-operator and transcription factor-responsive element complexes, can show considerable structural plasticity and internal dynamics at key regulatory spots. For example, the high

on-rate of repressor-operator complex formation can only be explained by non-specific binding to DNA adjacent to the operator binding site and subsequent sliding and/or hopping to a final destination [3]. NMR studies are well suited to address such questions (Fig. 4). Ribosomal proteins are assumed to stabilize specific RNA structures and promote compact folding of the large rRNA. The conformational dynamics of the protein between the bound and unbound states play an important role in the binding process. A model for the ternary L11 – RNA - thiostrepton complex that is based on interaction data and a high resolution structure determination based on NOEs, J coupling and RDC of the L11 protein is shown in Fig. 5 [4]

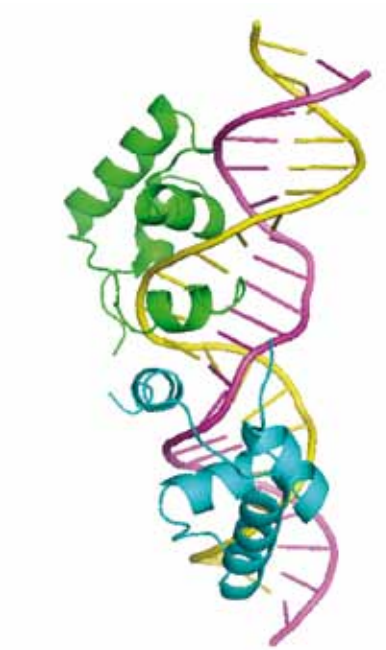


Figure 4: Protein-DNA complex of an altered-specificity mutant lac headpiece (PDB 2bjc)

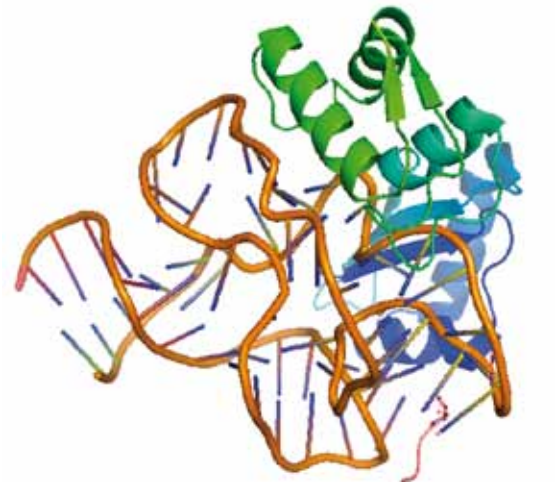


Figure 5: NMR structure of the tertiary complex between the GTPase region of 23S rRNA, ribosomal protein L11, and thiostrepton (PDB 2jq7)

The model provides an explanation for the role of the L11 N-terminal domain in elongation factor binding. A striking example of NMR capabilities to detect temporal variations of the spatial structure is shown in Figure 6, where temporal changes in the mutual orientations of two helical parts of HIV-TAR RNA, as mapped by measuring the RDCs, are shown [5]. NMR uncovered super-large amplitude helix motions that trace out a surprisingly structured and spatially correlated 3D dynamic trajectory. Notably, this trajectory samples conformations that are found in various ligand-bound states of the RNA.

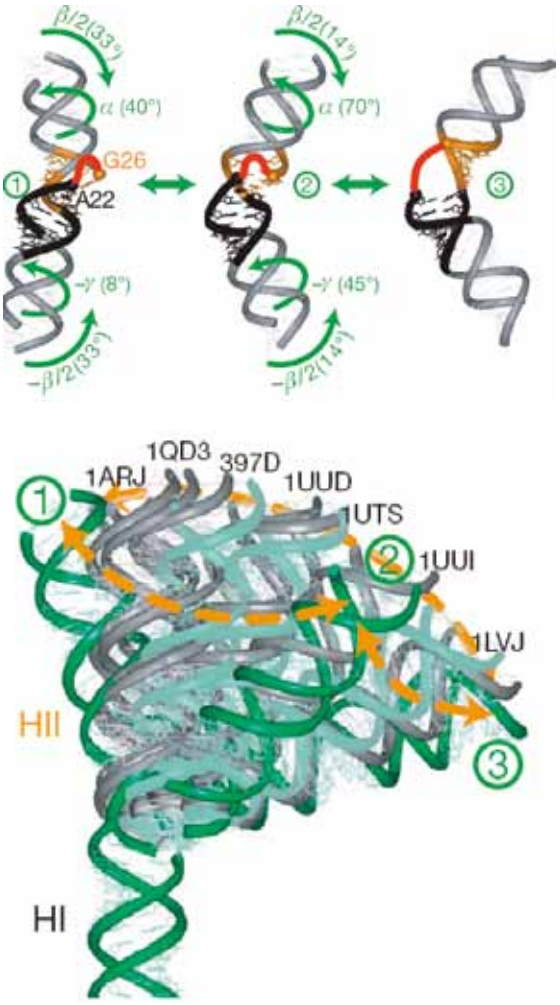


Figure 6: NMR mapping of the HIV-TAR interhelical motions from the measurments of RDCs.

Presently, NMR has proved to be a very powerful technique, providing information on subtle details of relatively complicated intermolecular complexes. However, such applications are still limited to studying a small number of carefully selected

biological systems exhibiting favorable physico-chemical behavior combined with efficient and well-optimized protocols of sample preparation. The great challenge is to convert such investigations of model systems into sample-targeted research, driven by physiological importance rather than by spectroscopic properties, which make the studied molecules exceptionally favorable for NMR investigations.

In contrast to X-ray crystallography, NMR can analyze (both structurally and kinetically) highly dynamic complexes and intrinsically disordered states participating in DNA transcription, replication, and repair. The challenge of NMR is to implement similarly efficient methodologies to those which make crystallography efficient. When possible, NMR should be combined with other techniques to complement the picture of highly complicated and dynamic structures of protein-NA assemblies.

The current document describes the capabilities of the present state-of-the-art technology used to study protein-DNA and protein-RNA complexes using NMR spectroscopy. Problems at the molecular biology level, as well as challenges and limitations in the presently available NMR methodology are addressed and possible remedies and future prospects briefly outlined.

Selected topics

Nuclear magnetic resonance has a unique advantage over X-ray crystallography in that it can determine the three-dimensional structure of biological macromolecules in solution. Sample pH and salt concentrations for the measurements can be adapted to match various physiological conditions as closely as possible. More than other techniques, NMR provides information about protein-nucleic acid complexes by detecting weak or transient intermolecular interactions. Investigations of protein-nucleic acid assemblies supply a wealth of information about interaction networks and interface recognition processes, yielding results having a potentially high impact on proteomics, system biology, and the entire biological community. In addition to structural information, NMR is unique in its ability to characterize both intramolecular and global motions on an extremely wide time scale, ranging from 10^{2} to 10^{-12} s. The intrinsic dynamics of proteins and nucleic acids has a distinct impact on a wide range of biological functions. Recent studies clearly document that protein-nucleic acid complexes are dynamic ensembles, in which both proteins and RNA or DNA molecules can change their conformational states dramatically. To describe those changes, the coupling of kinetic and structural studies becomes a necessity [6]. The study of the different functional states of macromolecular assemblies requires a combination of various experimental and theoretical approaches. Selected topics, which need a special attention in the near future, are briefly discussed in the subsequent paragraphs.

Sample preparation

Basic isotope labeling strategies for proteins and nucleic acids were developed during the early nineties. However, studies of protein-nucleic acids complexes with the size relevant to address important biological questions require sophisticated strategies beyond a simple introduction of ^{13}C and ^{15}N isotopes in a uniform manner. Sub-unit selective isotope labeling greatly facilitates analysis of NMR spectra by reducing their complexity. To succeed, the protein-NA complex must be reconstituted *in vitro* from its individual components, which very often represents a non-trivial task.

As the macromolecular ensemble becomes larger, the number of observable nuclei and the line widths of associated NMR signals increase. At the same time, the sensitivity and resolution decrease, making the assignment of signals prohibitively complicated. Amplification of the dipolar interactions between $^{13}C/^{15}N$ nuclei and neighboring protons in slowly tumbling, large molecular assemblies is primarily responsible for the line-broadening of NMR signals. The quality of the NMR spectra can be substantially improved, at the expense of their information content, by partial or complete deuteration, which by reducing the number of (the observable) ^{1}H nuclei attenuates the deteriorating effects of dipolar interactions. In the case of full deuteration, as a disadvantage, the number of observable protons is limited to the number of hydrogens in those amide groups, where deuterons can be effectively exchanged to protons in H_2O. When using fractional deuteration, resulting benefits generally break down for larger molecular assemblies. A site-specific deuteration, which reduces the number of observable nuclei and attenuates the dipolar interactions in a controlled fashion, represents a successful alternative. It has been documented that a combination of site-specific deuteration along with uniform ^{13}C and ^{15}N labeling allows NMR studies of macromolecules with the molecular weight up to 60 kDa. Labeling protocols for site-specific deuteration have been developed for both protein and nucleic acid parts. Further improvements can be achieved by a selective ^{13}C and/or ^{15}N labeling of proteins and nucleic acid constituents [7]. Nucleic acids can be very well produced by chemical methods, which allow

specific labeling and/or chemical modifications at key residues. In this way, dynamic studies can performed in ease and in high detail. In contrast to proteins, where such methodology is still in an early stage of development, the segmental labeling of nucleic acids has been well established. This allows attention to be on crucial parts of large complexes and intrinsic dynamic processes to be studied. When accompanied by fractional or site-specific deuteration and by a separate employment of ^{15}N and ^{13}C isotope labels, greatly improved quality of NMR spectra has been achieved in terms of both resolution and sensitivity [8]. In the case of proteins, the general application of the selective ^{13}C and/or ^{15}N labeling of individual amino acids using recombinant expression system is restricted by the scrambling of isotope label through metabolic pathways within the expression host organism. As a remedy, cell-free expression systems with a low level of metabolic conversion have been proposed. Additional gain in sensitivity and resolution can be achieved using stereo-array isotope labeling (SAIL), a technique that builds a complete stereospecific and regiospecific pattern of stable isotopes, optimal with regard to the quality and information content of the resulting NMR spectra. SAIL uses exclusively chemically and enzymatically synthesized amino acids for cell-free protein expression and offers sharpened lines, spectral simplification without loss of information, and the ability to rapidly collect the structural restraints required to obtain a high-quality solution structure.

In addition to spin labeling, molecular biology methods suitable to produce macromolecular assemblies and complexes in an economic and achievable fashion are crucial for successful studies of functionally relevant systems. In order to investigate a range of states representing various functions, new strategies for sample preparation are required, as a single component of large macromolecular machineries may not be stable. In many cases, development of specific co-expression and co-purification protocols for large functional assemblies is essential for their structural characterization.

NMR methodology

Structure determination of protein-NA complexes is challenging and involves many biochemical and biophysical aspects. The structure determination of protein- RNA complexes became possible by NMR spectroscopy only recently. Often, truncated forms of the studied protein-RNA complexes have to be generated in order to obtain a well-behaving system for NMR structure determination. The design of truncated constructs relies not only on computational alignment strategies but also on limited proteolysis and functional assays. Such a careful strategy should exclude the possibility of truncating functionally important elements out of the studied constructs. Many nucleic acids binding proteins have a modular domain organization, in which individual domains independently bind different sequences of nucleic acids or their independent structured elements. These multi-domain complexes can be simplified using a so-called "modular approach", in which an excision of thermodynamically stable secondary and tertiary structure elements out of their larger structural context, such as RNA stems or loops and individual protein domains, helps to achieve the resonance assignments and is valuable for the structure determination procedure.

Resonance assignment strategies resemble those used for the resonance assignment for unbound proteins and nucleic acids. Often, protein-NA complexes are prepared with only a single component isotopicaly labeled to reduce spectral overlap and line broadening. Single component labeled complexes in conjunction with isotope filter or editing experiments are very important to provide the intermolecular nuclear Overhauser enhancements (NOEs) that are essential to define protein-NA interaction interfaces. In these experiments, two data sets are recorded, which differ only by the phase of an editing 90º ^{13}C or ^{15}N pulse. Choice of addition or subtraction and their combinations in direct and indirect dimensions provides several possibilities to observe intra- and inter-molecular NOEs for single component labeled complexes. Other $^{13}C/^{15}N$-labeling schemes, producing selectively labeled domains or base/amino acid-specifically labeled molecules, represent additional tools for characterization of the protein-NA complex interface.

The conformation of large protein-NA assemblies can be probed by the use of TROSY methods [9] in combination with full or partial deuteration that significantly improves the relaxation properties of the observed NMR signals. Similarly, introduction of $^{1}H/^{13}C$-labeled methyl groups on per-deuterated background is very useful for assessing the conformational behavior of large systems. Their assignment is facilitated by COSY-based, multiple-quantum TOCSY and methyl-detected TROSY experiments specifically developed for this purpose. In the case of fully deuterated protein and nucleic acid constituents, it is possible to apply recently developed methodologies based on the direct detection of ^{13}C nuclei, which take advantage of cryogenically cooled probeheads with greatly enhanced sensitivity. Advantages of the ^{13}C direct detection NMR methodology are addressed in detail in chapter 6 of this volume.

Identification of the interaction surfaces

NMR spectroscopy allows relatively straightforward identification of binding interfaces between the constituents of protein - DNA/RNA complexes. Revelation and proper description of the interactions that are primarily responsible for the formation of protein-NA complexes are critical for understanding their function. Unfortunately, deciphering the structural details *de novo* is usually very time-consuming. However, in many cases, NMR can be applied to quickly map the binding interfaces of the complex constituents. In combination with selective labeling, this qualitative approach can be utilized to study very large molecular complexes with molecular weight exceeding 100 kDa.

Chemical shift mapping

The simplest method to obtain information about the molecular interfaces is based on the monitoring of chemical shift perturbations induced by complex formation. A comparison of the chemical shift data for ^{1}H, ^{13}C, ^{15}N, and ^{31}P nuclei obtained on free constituents and on the complex, using an appropriate combination of isotope labeling of the individual partners, directly identifies regions affected by the binding interaction. Resonances influenced by contacts display significant changes of the chemical shifts or residue specific line-broadening due to intermediate exchange [10]. Analysis of the spectral perturbations in a site-specific manner is called *chemical shift mapping***.** Although relatively simple, this approach cannot always unambiguously distinguish regions directly affected by the intermolecular contacts and regions distant from the binding interface, as additional conformational changes can be associated with the ligand binding. For tightly binding complexes ($K_D < 10^{-5}$ M), intermolecular assemblies must yield NMR spectra amenable for interpretation. For weakly associating systems, perturbations resulting from the intermolecular interactions are observed on one of the free complex constituents, removing the size of the complex as a limiting factor.

Cross-relaxation

Close spatial proximity of protons of residues close to the intermolecular interface of the protein – nucleic acid complex can lead to sufficiently large intermolecular dipole-dipole interactions. Resulting cross-relaxation and intermolecular NOEs provide direct evidence of their spatial proximity. If complete chemical shift assignments are known for all constituents, isotope-edited NMR experiments can be used to identify interfacial NOEs and to build up a high-resolution structure of the complex interface [11]. Although extremely useful, this approach can be very time-consuming and technically challenging, as well-resolved and unambiguously assigned spectra must be obtained for the entire complex. For the analysis, the strength and type of the intermolecular interactions, along with the number of intermolecular contacts that produce a sufficiently large number of intermolecular NOEs, represent the most limiting factors.

Saturation transfer

Cross-relaxation effects can also be utilized also in a less laborious way, which avoids the need for the full resonance signal assignments. Instead of relying on the precise identification of the interacting nuclei, cross-saturation or saturation transfer from one partner of the complex to the other constituent can be used. For example, protein aliphatic side-chain protons can be saturated and the polarization transfer monitored using intensity changes of the aromatic protons in the nucleic acid, thereby identifying the part in contact with the protein [12]. In general, the binding interface is determined from changes in the peak intensities of NMR spectra of one of the binding partners when signals of the other partner are saturated by radio-frequency pulses. For this purpose, usually one of the partners carries different isotope labels, enabling selective saturation, excitation, and detection.

Changes in the solvent accessibility can also provide a direct evidence for identification of the interfacial residues. This method utilizes the solvent saturation and polarization transfer between the solvent and the atoms of the complex. The signal intensity changes caused by solvent saturation are measured for a free and for a complexed molecule. The observed differences directly identify nuclei for which the intermolecular contacts modify solvent accessibility due to complex formation. In a related approach, water soluble relaxation agents can be added to the sample, leading to the relaxation increase of the atoms exposed to the solvent. Binding of the ligand protects residues at the interacting surface, thereby revealing atoms participating in the intermolecular contacts. Similarly, it has been shown that monitoring the hydrogen/deuterium exchange rates may serve the same purpose.

Spin labels and paramagnetic tagging

Paramagnetic spin labels and metal binding tags provide information about the molecular interfaces by supplying data

on long-range distances by carrying paramagnetic centers enhancing relaxation in distances much larger than covered by proton-proton dipolar interactions (NOEs). The spin label is a small, free-electron carrying molecule, which is covalently attached to the protein or modified RNA/DNA molecules. The paramagnetic species increase the relaxation rate of NMR resonances in the vicinity of the unpaired electrons (PRE – Paramagnetic Relaxation Enhancement) in proportion to the inverse sixth power of the distance between the label and the reporter nucleus and generate distance dependent line-broadening providing quantitative restraints up to 25 Å compared to the 5-6 Å range covered by NOEs. Typically, the differences in the line broadening are monitored by measuring spectra with an active, free-electron stabilized spin label and spectra acquired after reducing the spin label by adding ascorbic acid or another reducing agent. To achieve specific localization of the spin label on the protein chain, a nitroxide group, usually incorporated into a heterocyclic ring, is introduced into the protein by site-directed mutagenesis. Only one accessible cysteine residue is left unmodified, while other are mutated to alanine or serine. Instead of nitroxide labels, metal chelating tags, capable of paramagnetic metal binding, can be utilized. Paramagnetic metal labels carry additional information by inducing pseudo-contact chemical shifts and partial alignment in the magnetic field. Analysis of the induced shifts and measurements of residual dipolar couplings provide supplementary distance restraints. In the case of nucleic acid spin labeling, introduction of several labels along the extended helical RNA/DNA sequence might be an advantage. For RNA, a 4-thiouracil base is used to attach a 3-(2-Iodoacetamido)-proxyl residue with a free, unpaired electron [13]. It has been demonstrated that such a modification does not disturb the base stacking within a double-stranded RNA. Paramagnetic tags can be used also to detect weakly populated, transient states, which encounter complexes involved in protein-protein and protein-nucleic acid recognition. When using spin labeling and paramagnetic tags, several precautions must be observed. First, the spin label should not physically perturb the binding interface, yet its distance must ensure sufficiently large relaxation enhancements for interfacial residues. In addition, the spin label should not be placed on mobile segments of the complex, which would lead to motional averaging of the measured distances. Nevertheless, this area of research is flourishing and significant progress has been made recently, demonstrating the utility of these experiments especially for the structural analysis of higher molecular weight molecular complexes.

Structure calculation

Structure determination of small proteins using nuclear magnetic resonance (NMR) spectroscopy is becoming a routine procedure due to significant advances in automated resonance and NOE assignment methods. These automated procedures are less suitable for nucleic acids due to their intrinsic properties (high resonance overlap and low proton density compared to proteins), and entirely ineffective for protein-NA complexes. While working with protein-NA complexes, NOEs in the nucleic acid and intermolecular NOEs between the nucleic acid and the protein are usually assigned manually, directed by human expertise, and the protein part is subject to the automated NOE assignment technique. This hybrid approach provides the most effective results and has been used for a number of protein-NA structure determinations carried out in the past.

The major bottleneck in the determination of protein-NA complexes resides in defining the interaction interface between the protein and nucleic acid. Due to intrinsic dynamics and chemical nature the interface is usually characterized by an extremely low number of intermolecular NOEs (varying from dozens to a few hundreds) compared to a protein region that represents a similar volume, and frequently contains ambiguous data. Residual dipolar couplings (RDCs) are often of critical importance as they help to define the mutual orientation of the protein and the nucleic acid. Nonetheless, the majority of protein-NA interfaces are relatively "under-determined" systems. To avoid pitfalls that may lead to the incorrect definition of protein-NA interaction interfaces, careful analysis and interpretation of experimental data is mandatory. To precisely define mutual orientation between the protein and the nucleic acid, additional structural information can be extracted from SAXS and PRE data. The intermolecular interface frequently includes electrostatic interactions in the sugar-phosphate backbone of the nucleic acid (mediated by arginine and lysine residues). These interactions are observed only indirectly through the neighboring proton-containing groups. Precise and "chemically correct" charge-charge interactions can then only be achieved by using a proper force field in the structure calculation. The choice of a force field also has a large effect on the conformation of structurally underdetermined regions and therefore the best existing, state-of-the-art force fields should be utilized.

Chemical shift mapping in combination with docking calculations has developed as an efficient methodology to study protein-protein complexes and to understand the effects of mutations. Such modeling calculations tend to be more powerful in docking rigid proteins, with limited conformational changes

upon complex formation. Since RNA and DNA are intrinsically dynamic molecules, this initially slowed down the application of docking calculations for protein-NA complexes. More recently, the protocols of docking calculations have considerably improved and developed technology allows sampling of larger conformational space. Docking calculations now become feasible also for dynamic complexes [14].

Combination of NMR and SAXS (and other) data

The development of new techniques to tackle larger molecules and molecular complexes as well as the increased coverage of the 3D fold space of protein domains in the Protein Data Bank opens the way for new and efficient methods for studying multi-domain proteins and molecular complexes by NMR. Frequently, the individual structural domains are known and the aim is to rapidly determine their quaternary arrangement. The relative orientation of the independent units can be determined from residual dipolar couplings (RDCs). Translational restraints can be derived from i) paramagnetic relaxation enhancements (PRE) using site-specific spin labels, which yield long-range inter-domain distances or ii) small angle scattering data (SAXS/SANS), which provide information regarding the overall shape.

To calculate the quaternary arrangement of the molecules, a robust and flexible protocol for calculating such complexes has been developed and is embedded in a standard NMR structure calculation setup (e.g. ARIA/CNS). The relative orientation of the known substructures is achieved in two steps: (i) the domains are locally refined against the experimental data, (ii) the relative domain orientation is determined with random starting orientation of the substructures. Both steps use the same simulated annealing protocols as in a standard NOE-based NMR protein structure calculation. The major difference in the latter is the inclusion of additional forces to constrain all atoms harmonically to their starting position. The force constants can be chosen by the user and can vary for different regions, such that tighter constraints can be applied to parts of the starting structures, where more reliable information or fewer experimental NMR restraints are available.

The translation packing of the complexes is achieved by introducing distance restraints from paramagnetic relaxation enhancements (PRE). The distances are entered in a way similar to NOE-derived distance restraints. The nitroxide spin labels attached to specifically engineered cysteine residues induce signal bleaching of amide protons in $^{1}H,^{15}N$ - HSQC spectra up to a radius of approximately 25 Å from the paramagnetic center. The calculation allows an ensemble refinement of the experimental distances to compensate for multiple conformations of the spin labeled side chain. An additional force was introduced into CNS to directly incorporate small angle scattering data (SAXS/SANS) in the structure calculation. Advances in coupling of SAXS/SANS and other related methodologies with NMR are addressed in detail in Chapter 8 of this volume.

Strategies

Three large workshops, namely in Utrecht, June 14-16, 2006, in Brno, February, 1-2, 2008, and in Murnau, October 16-18, 2008 were organized within the Coordination Action NMR-Life to discuss problems associated with the structure determination of protein-nucleic acid complexes using nuclear magnetic resonance in detail. The following strategies have been proposed to address the existing limits.

Isotope labeling

To successfully solve the structures of larger protein-NA complexes, a combination of various labeling strategies and sample preparation protocols both at the protein and the nucleic acid levels is required. It was concluded that a successful combination of various approached generates a need for wide and highly specialized expertise, which is generally beyond the capacity of a single lab. The participants expressed their interest in establishing a "coordinated activity" focused on automated sample preparation. As stated above, the methodology of designing samples of large complexes suitable for structural NMR studies is at hand, but it remains technically extremely demanding, very often expensive, and not available to the broader structural biology community. Especially in the case of labeled DNA and RNA samples, a public-access sample production facility, ideally with financial support, would be greatly appreciated. Segmental labeled RNA structures will shift the size of systems amenable to NMR studies by one order of magnitude. The same applies to the production of site-specifically and segmental labeled proteins. In addition, it was stressed that the complex structures studied so far represent kinds of "model" structures. These systems were intentionally selected to provide high quality NMR spectra and were optimized for years to match the requirements of NMR structure determination. Therefore, the current technology is not yet a "plug-and-play" tool applicable to any biologically interesting system. The efforts to move in the direction of general applicability were

mentioned. For example, automated methods of screening the sample stability based on optical scattering have been found to be very helpful. It has been also stressed that NMR can serve as a screening tool for X-ray crystallography since well folded samples providing characteristic NMR spectra usually crystallize with a much higher probability.

Data collection

Acquisition of NMR spectra is still a rather demanding activity requiring expensive equipment and advanced know-how. Although standard NMR experiments have been implemented in a ready-to-use format in the software controlling data collection, large molecules and molecular complexes often require a customized approach. It is desirable to modify and optimize standard NMR parameters to achieve optimal performance when working with samples of limited concentration, stability or unfavorable relaxation properties. In addition, large molecular complexes often require high magnetic fields in order to suppress relaxation effects. European large scale NMR facilities have been established in order to provide the possibility to use high field spectrometers to external users. Less experienced users can also ask for advice, technical help, or for complete spectra acquisition. The operation of such facilities was discussed at the 2008 meeting in Brno. Participants appreciated the effective procedure of evaluating proposals for such measurements, the high technical quality of the equipment offered by the large-scale facilities, and the friendly environment. While the technical level of the large scale facilities is very high, many potential users would welcome simplified and standardized protocols for data acquisition and introduction of auto set-up features. At the same time, flexibility should be maintained for advanced users and for samples requiring specific and highly optimized experimental setups.

Structure calculation

A broad variety of software packages designed to calculate biomacromolecular structures based on NMR data exists. There is a general agreement that unification of the available tools would make their use much more efficient. The X-ray crystallography community may serve as an example of more advanced developments in that direction. Initiatives such as e-NMR, CCPN, and Extend-NMR already started to create the necessary pressure and received financial support. The general idea is not to define "the only recommended software" but to establish the compatibility of the existing tools and to facilitate conversions between data formats. The users would like to see new structure calculation tools that are more reliable, user friendly, and accessible to non-specialists.

Structure validation

The whole process of structure determination must be accompanied by a careful validation of the data used for calculation and obtained structures. It has to become mandatory that validation tests are done and data are scrutinized prior to structure deposition in public repositories. Therefore, it is necessary to establish standardized validation procedures. Even more important is the need to deposit not only structures, but also restraints extracted from the data and used for structure calculation, and ideally also raw experimental data. A policy requiring data deposition upon publication would not only allow anybody to check the reliability of solved structures, but it would greatly enhance developments of structure modeling tools. In many cases such protocols could be tuned in blind tests using large amounts of experimental data made available prior to experimental structure determination. On the other hand, some participants of the CA NMR workshops expressed concern that supplying data to the modeling tests might create an undesirable time pressure leading to deposition of "premature" structures. Therefore, deposition and release of structures prior to publication, which would allow blind testing, should remain voluntary. Whereas deposition prior to publication, as well as strict rules on handling and/or presenting data before deposition is clearly not compatible with all – in particular long-term – studies, it is crucial that upon publication all data (raw, restraints, and coordinates) are well documented and become accessible via public depositories. It is important to keep high standard of the databases content in general. Not just structures but also the deposited NMR data should be reliable. Formal mistakes and inconsistencies, such as incorrect chemical shift referencing and atom labeling, should be avoided as much as possible. There is a general agreement that editors of scientific journals and reviewers must be the force that makes submission of experimental data mandatory.

Conclusions and outlook

- A "coordinated activity" focused on automated sample preparation is highly desired. The methodology of designing samples of large complexes suitable for structural NMR studies has been developed, but it remains techni-

cally extremely demanding, expensive, and not available to the broader structural biology community. Especially in the case of labeled DNA and RNA samples, a public-access sample production facility, ideally with financial support, would be highly appreciated. Sample preparation is frequently the bottleneck in the application of NMR to biomolecular complexes. Efficient high-throughput methods can only be economically and reliably run in expert, highly specialized facilities.

- The deposition of experimental data, both raw data and structural restraints, should be a mandatory part of structure publication. The small number of nucleic acid structures available in the databases presents the major obstacle to the structure determination methods based on database searches, which have been so successfully applied in protein studies.
- So far, only proton chemical shifts can be productively used for the prediction of the secondary structure of RNA and DNA molecules. The lack of ^{13}C chemical shift data for RNA, DNA, and protein-NA complexes in the BMRB database prevents reliable correlation between the structure and carbon chemical shift values.
- The sequences of target molecules for studies of large macromolecular assemblies before their structures are solved should be shared in a similar initiative as in the CASP (Critical Assessment of Structure Prediction) network to stimulate the development of structure prediction methods.
- The communication between the structural genomics centers, where a number of large macromolecular complexes are studied, and the broad scientific community needs improvement. Workshops and courses, as well as other activities organized within the framework of genomics projects should be better advertised.
- Modeling methods represent an important complement extending our knowledge beyond the current experimental limits. Synergy between evaluations of experimental data and computer simulations should be used to achieve improvements in the force fields and structure calculation protocols.
- It is clear that biomolecular NMR spectroscopy in solution can contribute significantly in the context of systems biology, as molecular interactions along a given biological pathway or engaged in a specific mechanism can be analyzed both qualitatively and quantitatively. For example, the regulation of gene expression often involves the binding of a multi-domain protein with a single-stranded RNA and therefore involves the recognition of degenerate RNA sequences. The combined binding of multiple domains, which are connected by flexible linkers, allows the assembly of specific protein-RNA complexes by providing an additive input to generate a particular functional readout, such as one involved in the regulation of splicing and alternative splicing.

References

(1) Puglisi, J. D.; Doudna, J. A. *Curr.Opin.Struct.Biol.* **2008**, *18*, 279-281.
(2) Eddy, S. R. *Nature Reviews Genetics* **2001**, *2*, 919-929.
(3) Salinas, R. K.; Folkers, G. E.; Bonvin, A.; Das, D.; Boelen, R.; Kaptein, R. *ChemBioChem* **2005**, *6*, 1628-1637.
(4) Jonker, H. R. A.; Llin, S.; Grimm, S. K.; Wohnert, J.; Schwalbe, H. *Nucleic Acids Res* **2007**, *35*, 441-454.
(5) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. *Nature* **2007**, *450*, 1263-U1214.
(6) Cowieson, N. P.; Kobe, B.; Martin, J. L. *Curr.Opin.Struct.Biol.* **2008**, *18*, 617-622.
(7) Tugarinov, V.; Kanelis, V.; Kay, L. E. *Nat Protc* **2006**, *1*, 749-754.
(8) Tolbert, T. J.; Williamson, J. R. *J Am Chem Soc.* **1997**, *119*, 12100-12108.
(9) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc.Natl.Acad. Sci.U.S.A* **1997**, *94*, 12366-12371.
(10) Matsuo, H.; Walters, K. J.; Teruya, K.; Tanaka, T.; Gassner, G. T.; Lippard, S. J.; Kyogoku, Y.; Wagner, G. *J Am Chem Soc.* **1999**, *121*, 9903-9904.
(11) Sattler, M.; Schleucher, J.; Griesinger, C. *Progr.NMR Spectrosc.* **1999**, *34*, 93-158.
(12) Takahashi, H.; Nakanishi, T.; Kami, K.; Arata, Y.; Shimada, I. *Nat. Struct.Biol.* **2000**, *7*, 220-223.
(13) Ramos, A.; Varani, G. *J.Am.Chem.Soc.* **1998**, *120*, 10992-10993.
(14) van Dijk, M.; van Dijk, A. D. J.; Hsu, V.; Boelens, R.; Bonvin, A. *Nucl.Acids Res.* **2006**, *34*, 3317-3325.

Chapter 4

Perspectives of NMR in drug discovery

Authors: Maurizio Pellecchia[a], Ivano Bertini[b], David Cowburn[c], Claudio Dalvit[d], Ernest Giralt[e], Wolfgang Jahnke[f], Thomas L. James[g], Steve W. Homans[h], Horst Kessler[i], Claudio Luchinat[b], Bernd Meyer[j], Hartmut Oschkinat[k], Jeff Peng[l], Harald Schwalbe[m], Gregg Siegal[n].

[a] Burnham Institute for Medical Research, La Jolla, California, USA
[b] Magnetic Resonance Center, University of Florence, Italy
[c] New York Structural Biology Center, New York, USA
[d] Nerviano Medical Sciences, Milan, Italy
[e] Institute for Research in Biomedicine, Barcelona Science Park and the Organic Chemistry Department, University of Barcelona, Spain
[f] Novartis Institutes for BioMedical Research, Basel, Switzerland
[g] School of Pharmacy, University of California at San Francisco, California, USA
[h] Institute of Molecular and Cellular Biology, University of Leeds, UK
[i] Department of Chemistry, Technische Universität, Munich, Germany
[j] Institute for Organic Chemistry University of Hamburg, Germany
[k] Forschungsinstitut für Molekulare Pharmakologie, Berlin, Germany
[l] University of Notre Dame, Indiana, USA
[m] Center for Biomolecular Magnetic Resonance, Frankfurt, Germany
[n] Institute of Chemistry, Leiden University, The Netherlands

Foreword

In the past decade, scientists from both industry and academia realized that Nuclear Magnetic Resonance (NMR) spectroscopy could be very useful and versatile in monitoring inter-molecular interactions, making it a potentially powerful and general tool for drug discovery. Here several common views on the usefulness of NMR spectroscopy in the drug discovery process, both in industry and in academia, are presented. Then a glimpse at the future is attempted, and at the challenges to the current drug discovery paradigm that are posed by the emerging systems biology approaches.

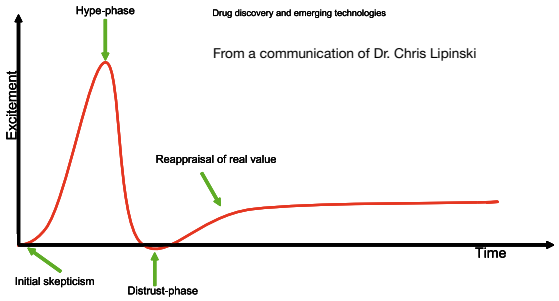
Introduction

While NMR is often valued for its ability to shed light on structure, its real value in drug discovery probably lies in evaluating molecular interactions at the atomic level. NMR is a multifaceted phenomenon that enables an impact in various aspects of the small molecule drug discovery process [1-18] Every organic chemist is familiar with the chemical shift. This simple parameter is highly sensitive to the exact environment of the atom, and therefore yields information about whether a small molecule binds to a target protein or nucleic acid, what parts of the small molecule are interacting, and to which part of the macromolecular target the ligand is bound. Yet other NMR experiments are sensitive to molecular motions. The variety of readily measurable parameters permits NMR to contribute efficiently to the ligand discovery process by assessing target *drugability*, initial hit identification via fragment-based drug design (FBDD) [19-34] with screening of very small molecule (<300 dalton) libraries, pharmacophore identification, hit validation, hit optimization and potentially structure-based drug design. NMR can also be used to determine low resolution structures of target-ligand complexes for natively unstructured proteins or membrane proteins that are not amenable to crystallographic approaches. The combination of advances in instrumentation, the use of orientational restraints [35-39] in partially oriented media , the use of selective $^{13}C,^{1}H$-methyl labeling in otherwise deuterated samples [40-42], the simultaneous acquisition of multi-dimensional experiments, [43] the use of segmental labeling techniques [44-47] and the use of TROSY-type experiments [48;49] have made NMR spectroscopy a very powerful and efficient tool for structural biology initiatives [50]. Combined with automated assignment strategies and projection spectroscopy, these approaches promise to significantly reduce the time needed for data collection and analysis [51-57]. Finally, solid-state NMR spectroscopy has also advanced significantly in recent years, making it possible to study proteins such as neurologically relevant GPCRs, transporters or ion pumps [58-66]. Taking into account such developments, it may be envisaged that new areas of biology may be made accessible for structural research, focusing more on 'systems' rather than on purified material via solid-state NMR .

As mentioned, NMR could also be very useful in monitoring inter-molecular interactions involving macromolecules (proteins or nucleic acids) or a small molecule ligand, making NMR spectroscopy a potentially powerful and general tool for drug discovery.

Drug Discovery and Emerging Technologies

When it comes to developing and introducing novel drug discovery technologies, the response of the scientific and business development communities is often similar (from a communication of Dr. Chris Lipinski at a recent symposium at the Burnham Institute for Medical Research): after a first phase of skeptical resistance whose duration is somehow proportional to the current development stage of other competing techniques, an overwhelming and often exaggerated enthusiasm (hype-phase) is generated around the approach; as time goes by, the overly optimistic predictions turn into distrustful set backs (distrust-phase); eventually, the real value of the new approach emerges and its realistic impact to the drug discovery process is reappraised. This process takes on average about five to ten years, or more. How does NMR fit into this general trend? Even taking into account that NMR spectroscopy is a multifaceted technique that can be used at many different stages and levels of the drug discovery process, one can notice that each and every general NMR approach has followed, to a certain extent, a similar trend. Two general considerations can be further made: first, only valuable techniques usually survive the hype phase, and, second, when forthright and open discussions can take place evaluating the merits and pitfalls of a given approach, the technique can be appropriately and effectively justified. We believe that many aspects of NMR spectroscopy as applied to the drug discovery and development processes are now mature enough for such critical assessments. Our collective evaluation on past and future of NMR spectroscopy in the hit identification, validation, characterization and optimization processes are summarized in the next paragraphs, hopefully unbiased by innate skepticisms, overrated enthusiasms or over-trusting disappointments.



The response of the scientific and business development communities to the development and introduction of novel drug discovery technologies.

NMR in Fragment-Based Drug Design: Puzzling Approaches to Drug Discovery

It has been estimated that the number of potential drug molecules is of the order of 10^{10}–10^{50} [67] However, for a given target system it is difficult to imagine high-throughput screening (HTS) performed with much more than 10^6 compounds, especially considering that such endeavors would be very expensive and subject to a sizeable number of false positives and false negatives. The traditional approach of testing variations of known drugs is certainly not going to dive very deeply into this potential pool either, but at least it has the advantage of exploring compound space based on knowledge, so the search will be made more effectively. Of course, our chances of encountering cross-resistance are enhanced if we limit ourselves to compounds similar to those currently in clinical use. Consequently, it would be most useful to find molecules that might lead to development of drugs with novel chemical scaffolds. These statements represent the basic premise of the so-called fragment-based drug discovery approaches (FBDD) [19-24;26-33].

In principle, there are in principle several ways to *construct* novel ligands designed for a particular target that could subsequently become lead candidates. These entail a modest exploration of "*drugable*" molecular space, but the efficacy of this approach could be largely enhanced by using knowledge of the target. Sometimes that knowledge may be functional, but more often it is structural. That is, we presume the target protein (or nucleic acid) assumes a structure, and some aspect of that structure is used to search for small molecule ligands that might bind and be used to develop a drug candidate.

One approach is to build up a drug candidate (with typical molecular weight of 500 daltons) from screening a database of typically 1,000-15,000 compounds composed of smaller molecules (fragments) with molecular weight < 300 daltons and good aqueous solubility, using tethering [19-24;26-34], X-ray diffraction [29;68-72] or NMR approaches [22;32;33]. These techniques enable one to identify the location of binding of any fragment. Often, a second screen is carried out to find a second fragment that will bind in close proximity to the first fragment. Use of a second screen has an advantage in that binding of the first fragment may select for a particular conformation of the target that can enable binding of a particular second fragment that would not have been found in the absence of the first. Individually, the fragments may bind to the target with K_D values of only 10^{-4} to 10^{-3} M. However, by covalently linking the fragments, the additivity of the binding enthalpy and the favorable entropic contribution may pay off in a big way. In fact, sub-micromolar bi-dentate compounds can be found that are generally novel in structure. These can serve as starting points for further structure-activity studies by synthesizing focused libraries of related compounds or by determining the structure

of the linked compound bound to its target and using that knowledge to suggest structural modifications. Several recent manuscripts and review articles report on critical technical aspects of the use of NMR spectroscopy in FBDD [1-4;6-18;73].The fragment-based approach for primary screening has proved to be viable for the identification of lead molecules. The probability of detecting the binding of a low complexity fragment with high sensitivity exceeds that of full-sized ligands with lower screening sensitivity. The functional groups of fragment-based libraries should already include synthetically accessible starting points for chemical linkage. In a follow-up screen, chemical building block fragments with masked linker groups can be utilized, an optimization step in library design called the *'fragment pair concept'*. Key to the success of such a strategy is the quality of the fragment database. Quality, of course, refers to the purity of the compounds, but also to the diversity and chemical nature of the fragments chosen. Several different laboratories have developed their own fragment databases, some of which are emerging as commercially available libraries (Maybridge Corp., Chembridge Corp., Asinex Corp., Life Chemicals, ActiveSight Corp., Pyxis Discovery, to mention a few).

Several NMR strategies, which follow the initial screening trials, have been proposed (Tables 1 and 2), ranging from the more traditional chemical shift mapping to ligand-based techniques monitoring changes in ligand nuclear spin-relaxation properties upon binding, to measurements of diffusion, etc. Some of these approaches are better suited as screening methods and/or to validate hits coming from HTS campaigns (Table 1), whereas others are better suited to guide their optimization into more potent, selective and *drug-like* compounds (Table 2). It should be also possible to extend some of these approaches to *in-cell* NMR experiments to provide, for example, mapping information from chemical shift perturbations for serially expressed protein systems [74-76]. Other applications could include possibly deriving novel compounds with reduced serum albumin binding and/or cytochrome P450 enzymes interactions by designing out unacceptable properties during the iterative optimization process [77]. The detailed description of these methods can be found in the reported citations within Tables 1 and 2 as well as in recent review articles cited throughout this manuscript.

NMR in structure-based drug design

The identification of new possible targets or of possible "*druggable*" sites on known targets can also begin with structural studies. However, many multi-domain proteins show considerable flexibility in the organization of their components during interactions with multiple ligands, and allosteric modulation of activities is of considerable significance in their activity. In contrast to a structure determined in a crystal, where the inter-domain interactions accommodate the need for the lowest crystal packing energy, structures in solution reflect a more physiological milieu and can characterize the dynamic inter-conversions available [78]. NMR methods to characterize these interactions, using relaxation properties and special isotopic labeling can be applied to complex systems like protein tyrosine kinases, widely identified as significant targets, but where plasticity of interaction with ligands (or known drugs) is evident [79]. Magic-angle-spinning solid-state NMR entered the scene in the past five years as an additional alternative method for protein structure determination, and offers new perspectives for structural investigations on samples that could not easily be analyzed before, such as native membranes, fibrils and cytoskeletal complexes. Recently, models of a potassium channel-toxin complex, of various fibrils, and of receptor-agonist complexes were published as a result of the constant advance of the field [58-66]. Projects aiming at well-determined structures of membrane proteins are underway in several laboratories.

However, how does this structural information materialize into new potential drug candidates? An obvious strategy is to employ *in silico* approaches for the initial screening of a compound database to predict those that should bind to a (usually static) target protein structure. This has the apparent advantages that the search should be less expensive and faster (if time to develop an experimental assay is considered) than HTS. Computational "hits" from this virtual screening will need to be tested experimentally, but a relatively small number of top-ranking binders can be selected, e.g., 1%, for NMR screening; while certainly not high-throughput, a pleasant feature of NMR is that a system-specific assay needs not be developed. Currently, it is feasible to screen $>10^6$ compounds computationally; for the money spent on a single HTS screen of a pharmaceutical company's in-house database of compounds, a computer cluster could be constructed to screen all of the pure compounds extant in the world (assuming this number is *ca.* 10^7).

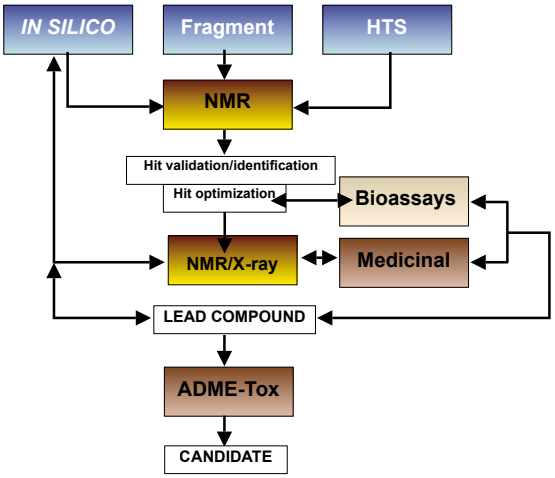
Numerous programs have followed the initial DOCK [80] algorithm and have been used on many systems [81]. For successful prediction of hits that are subsequently verified experimentally, the actual search algorithm used at this point in history seems to matter little. However, challenges for the field largely center around two general aspects: (a) most molecules adapt their

conformation upon binding to another molecule, i.e., both the ligand and the target are malleable such that an induced fit occurs; and (b) the scoring functions used to rank the ligands for binding affinity are highly imperfect. There is work in the field on both points, promising an improved future, but computational predictions of molecular recognition are today at a state where predictions of protein folding were about fifteen years ago.

With either of the two approaches mentioned above, one must be cognizant that any promising binders are only that: promising. The vast majority of compounds that look promising at this stage will fail in further development, chiefly from lack of bioavailability, i.e., not getting to the site of action, and from toxic side effects from insufficient selectivity. It is a challenge to increase the probability of any promising preclinical compound making it through the drug development pipeline to become a clinically viable drug. With the realization that ADME-Tox problems are an expensive limitation in drug development, scientists are working to develop tools to identify compounds or classes of compounds early that may engender these problems. One school of thought in recent years has been to filter the databases to be screened such that criteria largely comply with Chris Lipinski's "Rule of Five", an experiential list based on clinically successful drugs with good bioavailability [82;83]. However, as Lipinski reports, there are numerous examples of drugs that do not abide by these rules, so it is not really a rule to be followed blindly, e.g., the rules of Veber et al. are suited as well and provide an alternative way to look at potential drugs [84]. Likewise, methods to predict toxicity are now being developed. While computational toxicology is still in its infancy, it will undoubtedly improve with time. There is a fairly new initiative (DSSTox) aiming at creating a common format for chemical structures and searchable data files for toxicity databases that will be available to the public. This worthwhile venture is described at http://www.epa.gov/nheerl/dsstox/.

NMR spectroscopy in the drug discovery process: a critical assessment

Penetration of NMR into drug discovery remains rather limited when compared to HTS and X-ray crystallography, despite repeated large scale investments, and, as reported above, the fact that various NMR techniques can be considered essential tools in a vast array of academic and industrial research. One problem is that the utility of NMR as a structural biology tool in the hit to lead stage has fallen far short of original expectations. Clearly X-ray crystallography can provide higher resolution structures much faster and, as a result, it is far more widely used. The only exceptions are companies the have been setup based on NMR expertise, certain large pharmaceutical companies where NMR has proven itself over the long term, or in academia where individual research groups may focus on NMR. The question becomes: are there other areas of drug discovery where NMR information is clearly superior or for which there are no alternatives? In our admittedly biased opinion, the answer to the question posed above is a resounding yes! One very important application that we foresee is extending the current principles of fragment based drug discovery (FBDD) to membrane proteins. Although membrane proteins represent something like a third of our genome, more than 2/3 of all marketed drugs target them. Further, there is great underutilized potential in membrane proteins as pharmaceutical targets. Advantages of targeting small molecules to cellular membrane proteins include the fact that the compound needs not traverse the outer cell membrane to reach its target. However, a major hurdle to overcome is that membrane proteins are, by and large, very challenging in terms of biochemical manipulation. At present, most drug discovery efforts targeting membrane proteins (mostly GPCRs and ion channels) utilize cell-based assays and high throughput screening of large corporate compound collections. FBDD is having tremendous success in developing orally bio-available drugs to soluble targets via NMR [33] and we anticipate making a similar impact in the future in targeting membrane proteins.



The NMR structure-based drug design process

However, FBDD has had as much success also by using other techniques. Therefore, one has to ask the question: what unique insights does NMR actually bring to drug discovery? Currently, NMR techniques provide some information about the binding epitope on the ligand and map the binding site on

the target quickly. For the future, increasing numbers of membrane proteins can be recombinantly expressed and solubilised. These protein samples are often not suitable for crystallography yet can be used to perform NMR-based ligand screening. Based on the SAR of hits, follow-up focussed libraries could be synthesized to jump start the drug discovery process. NMR can also be used to determine structures of the ligands in the bound state and, using data from paramagnetic labels or traditional nuclear Overhauser effects (NOE), low-resolution structures of target-ligand complexes can be determined. No other biophysical technique can provide this sort of information for membrane proteins. On the strengths side, NMR can very quickly deliver information about ligand binding properties even if the receptor cannot be characterized at high resolution: it can provide a detail picture of the bound ligand, even if the receptor cannot be characterized; as mentioned, ligand binding to membrane-integrated proteins can be analyzed; if the receptor can be characterized, a limited set of NMR data can provide information on the location and orientation of ligands [85;86]; and last but not least, as mentioned in this article, NMR is well suited for fragment-based screening because it allows characterization of weak binding (see another recent example in reference [87]).

Clearly, however, drug discovery is an immensely complex venture, requiring a multi-disciplinary effort. One issue that we recognize is that each of the intervening disciplines (e.g., HTS, crystallography, cell biology, protein NMR, medicinal chemistry) is usually poorly integrated. Fragment-based approaches such as the SAR by NMR strategy require excellent integration with medicinal chemistry and possibly biology. Effective use then implies some degree of centralized organization, and specialization of labor. In an academic setting, this must come from collaboration. We envision that a possible solution would be to engage in collaborative programs that would pull together the state-of-the-art design of new drugs using NMR and other technologies to optimize the speed and quality of lead optimization. Large NMR-based infrastructures worldwide could play a role in these programs. Another major need is for laboratories of mainly synthetic chemistry groups that would be willing and able to collaborate in such an effort. Research groups are also needed that would be willing to perform biological and functional testing of intermediately generated new compounds in order to combine more efficiently and rapidly binding studies with functional assays. In the United States, there are several screening centers that may provide such support (Examples are the Molecular Libraries Screening Centers Network initiative, http://nihroadmap.nih.gov/molecularlibraries; the NIAID's Antimicrobial Acquisition and Coordinating Facility, http://niaid-aacf.org; the NCI's Developmental Therapeutics Program, http://dtp.nci.nih.gov).

Another issue that is being discussed at length is the use of macromolecular NMR as a structural tool to complement X-ray crystallography. That paradigm is not fully deployed in industry for a variety of reasons, but especially because it is not rapid. It may still be viable in academic research, where the choice of protein targets is less constrained by immediate therapeutic relevance. It is possible that a structural focus can re-emerge with the growing interest in integral membrane proteins, particularly with emerging solid-state NMR techniques, but this is not yet certain. NMR-based structure determination of protein-ligand complexes should also gain more weight in the future. But it must become fast and should take advantage of synergies with X-ray crystallography and computational tools [86]. To be viable in a fast-paced industrial setting, structure refinement via NMR must be stream-lined. Another option that may pay off with more research is computational modeling using a limited amount of NMR structural data. One issue we recognize is that there is a lack of suitable NMR structure determination software for such industrial purposes; a suggestion is to establish a consortium to develop new compatible, easy-to-use software, much as it was done by and for the X-ray community in the past. NMR screening, on the other hand, *via* several of the briefly mentioned methods for lead generation, optimization, and 'rescue' is a newer paradigm. It provides quicker turn-around for information of more immediate impact on medicinal chemistry. This application seems to have reached a plateau now, and it is unclear as to what further room for expansion remains. Again, possibly applications to 'non-traditional' targets is the best venue, e.g., RNA [88-90] or membrane proteins. A renewed interest in NMR as a metabolomic tool for predictive toxicology [91] seems to be a powerful contributor to drug development, permitting forays into systems biology [76]. The general impression is that this field is not yet saturated.

In addition to monitoring ligand binding and to determining macromolecular structures, NMR is among the most powerful methods for profiling biomolecular motion [92-97]. There is potentially great value in learning how to ascertain and ultimately exploit intrinsic motion to guide drug discovery and delivery. A huge challenge, however, is that dynamics studies are of similar or slower pace than structural studies. Moreover, there is not yet a consensus on how dynamics information can be best used to advance ligand design. A promising area is in estimat-

ing configurational entropy changes associated with binding, and coupling to calorimetry [98].

Finally, it is important to put NMR-based drug discovery in the context of the emerging trends in systems biology. Indeed, the very paradigm of the *target-directed* drug may change in the near future to the one of *pathway-directed* drug. Although it is true that ultimately a drug will interact with a target, the latter should be more and more seen in the context of biochemical pathways and fluxes. Targeting protein-protein or protein-nucleic acid interactions (see chapter 3) rather than single proteins will become increasingly important, and it is envisaged that NMR will be specially suitable for this task. Indeed, many of the physiopathological biomolecular interactions are weak and/or transient, and therefore more amenable to NMR characterization than, for instance, X-ray characterization. Furthermore, one should keep in mind that NMR may still undergo another "revolution" if the promises of DNP techniques to boost sensitivity (see chapter 7) are maintained. Increases in sensitivity by two orders of magnitude or more will make feasible approaches that today are hardly imagined like, for instance, differential spectroscopy experiments on intact cells grown in enriched media in the presence and absence of a candidate drug, to assess binding to the proper target, selectivity, and metabolism of the drug in a *in vivo* situation.

Therefore, at least in theory, NMR spectroscopy should find a significant role in each step in the drug discovery process (Figure 2), both of today and beyond. However, it seems that NMR suffers from a sort of Leibniz syndrome, in that it is often the "second best" at so many things. The diversity of problems it can attack is extremely alluring to its practitioners, leading to a state of mind: "All I have is a hammer, so everything is a nail". This attitude is, of course, unhealthy for NMR in drug discovery. Macromolecular NMR works best in drug discovery when its data can be quickly integrated with those from other analytical techniques. It has to be comprehensible, portable, and available on a time scale compatible with medicinal chemistry. No matter how unique NMR information may be, it will not be used unless it meets these criteria. Some years ago, the "word around the campfire" was that researchers in modeling and bio-informatics eschewed solution NMR structures (as opposed to crystal structures) because one gets a series of structures, thus calling into question the accuracy and meaning of the data. This may be myth; nevertheless, if true, it underscores the importance of integrating NMR data with the world-views of other disciplines that are the stronger driving forces of drug discovery.

One important issue is that training researchers able to translate basic discoveries to new drugs is not at all established in academia. In the US, medicinal chemistry is predominantly taught at the employer, with no focus on professional degrees in the specific areas. Only recently, several Schools of Pharmacy have either instituted or increased their investment in educating research scientists relative to professional pharmacists. Within these educational programs, there is controversy attending the issue of how much spectroscopy is required for its effective use, but with few exceptions, there is little educational effort in NMR relative to other means of lead generation/optimization. Defining the necessary steps for effective training is a valuable exercise. For example, for the use in screening and FBDD, medicinal chemists rather than physicists, biologists or even organic chemists would be preferred 'users' of the techniques. For NMR, the current educational focus still has a strong structure-determination bent, which appeals more to biologists or biophysicists than medicinal chemists. Yet, many of these students are interested in entering drug-discovery in industry. What do we tell them? What kinds of jobs await them? An increasing number of students want "turn-key" biophysical methods. There is less interest in mastering the underlying theory of a given technique, and much more on fast downloads to rapidly summarize the results of multiple techniques. This is natural, given our relentless emphasis on the urgency and competition in drug discovery. That is all fine, but what does this imply in terms of designing an appropriate curriculum for such students? Maybe a curriculum which includes more detailed studies whereby NMR is used to decompose the overall thermodynamics of binding for a given ligand-protein interaction into enthalpic and entropic contributions from the ligand, protein and solvent, could be a good compromise between basic and applied research in this area.

Concluding remarks

When all facts are considered, NMR remains a multifaceted and unique technique that is sensitive to both structure and dynamics and that can monitor the binding of low molecular weight ligands to biological macromolecules in the *early stages* of drug discovery due to its ability to detect even very weak binders. One common pitfall of the implementation of NMR in the industrial drug discovery pipe lines is that it is often brought in too late. On the other hand, while many examples do exist of successful drug discovery projects that are entirely jump-started by NMR-based approaches, it is clear that when applied in isolation, these methodologies, much like any other

technique, cannot be fully effective. The successful implementation of NMR in the drug discovery process is often based on the early and effective integration of medicinal chemistry, computational approaches, biology, and now systems biology. Training the scientists of the future based on these observations may be the long-term solution of these problems; establishing large collaborative efforts in the academic setting or the coordination of technologies in the industrial setting may represent the short term solutions.

Acknowledgements

We thank Dr. Marco Fragai[a] and Dr. Christopher Lipinski[b] for suggestions and comments. We also thank Mr. Lloyd Slivka[c] for careful assistance in the preparation of the manuscript.

[a] Magnetic Resonance Center, University of Florence, Italy
[b] Melior Discovery, Exton, Pennsylvania, USA
[c] Burnham Institute for Medical Research, La Jolla, California, USA

Table 1: Most used NMR methods for compound screening and hit validation

Approach (original references)	Observation	Used for	Description (references to recent pplications)
Chemical Shift Perturbation (Reviewed for example in reference [6])	Target (protein or nucleic acid) resonances	Primary screening/ Hit validation/ Site of binding	Identifies binders by means of chemical shift perturbation of resonances of target ([90;99-103])
Saturation Transfer Difference (STD) NMR ([88])	Ligand	Primary screening/ Hit validation	Identifies weak binders, build-up curve identifies interacting functional groups ([89;90;104-107])
WaterLOGSY ([108])	Ligand	Primary screening	Identifies binders by water mediated NOEs [109;110])
SLAPSTICK Using Spin-labeled protein ([111])	Ligand	Primary screening	Highly sensitive detection of fragments ([32;111])
TINS ([112])	Ligand	Primary Screening/Hit validation	Identifies small molecule compounds by screening libraries against immobilized protein targets ([113])
$T_{1\rho}$ and T_2 relaxation; Line broadening ([114])	Ligand	Primary screening/ Hit validation	Binding enhances relaxation, affinity estimate, build-up curve identifies interacting functional groups ([115])
Transferred NOEs ([116])	Ligand	Hit validation/ Conformation of flexible ligands	Interaction of binders with the target ([8;117]) Also helpful to determine bioactive conformation of flexible ligands such as peptides ([115;118])
FABS ([119])	Substrate of cofactor	Primary screening/ Hit validation	Utilizes reference substrates or cofactors to monitor enzymatic reactions ([3;120-124])
FAXS ([125-127])	Reference Ligand	Primary screening/ Hit validation	Utilizes reference substrates or cofactors to monitor enzymatic reactions ([3;120-124;128-130])
Diffusion measurements ([131;132])	Ligand	Primary screening/ Hit validation	Measures the difference is diffusion rates for ligands in the bound versus free state ([133])

Table 2: Representative NMR methods for hit/lead optimization

Approach (original references)	Observation	Used for	Description (references to recent applications)
SAR by NMR ([22;134])	Ligand/Target	Structural information FBDD Screening/Compound optimization	Design bi-dentate compounds ([135;136]).
SLAPSTICK with first-site spin- labelled compound ([137])	Ligand	FBDD Screening/ Compound optimization	Highly sensitive detection of fragments and weakly interacting second site compounds ([138])
SAR by ILOEs ([139;140])	Ligand-to-Ligand	FBDD Screening/ Compound optimization	Detects protein mediated ligand-ligand interactions (compounds occupying adjacent sites) ([140])
Pharmacophore by ILOEs ([141])	Ligand-to-Ligand	FBDD Screening/ Compound optimization	Detects protein-mediated ligand-ligand interactions and uses information for pharmacophore-based search of bi-dentate compounds ([141])
H_2O/D_2O exchange rate measurements	Target	Compound characterisation	Identifies binding epitope ([142])
INPHARMA ([143])	Ligand-to-Ligand	Compound characterization	Detects protein mediated ligand-ligand interactions (competition for the same binding site)

References

(1) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W:; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. *Nat.Rev.Drug. Discov.* **2008**, *7*, 738-745.

(2) Coles, M.; Heller, M.; Kessler, H. *Drug Discov.Today* **2003**, *8*, 803-810.

(3) Pellecchia, M.; Becattini, B.; Crowell, K. J.; Fattorusso, R.; Forino, M.; Fragai, M.; Jung, D.; Mustelin, T.; Tautz L. *Exp.Opin.Ther. Targets* **2004**, *6*, 597-611.

(4) Salvatella, X.; Giralt, E. *Chem.Soc.Rev.* **2003**, *32*, 365-372.

(5) Betz, M.; Saxena, K.; Schwalbe, H. *Curr.Opin.Cell Biol.* **2006**, *10*, 219-225.

(6) Pellecchia M.; Sem, D. S.; Wuthrich, K. *Nat Rev Drug Discov* **2002**, *1*, 211-219.

(7) Pellecchia M. *Chemistry & Biology* **2005**.

(8) Lepre, C. A.; et al *Combinatorial chemistry & high throughput screening* **2002**, *5*, 583-590.

(9) Wyss, D. F.; McCoy, M. A.; Senior, M. M. *Current opinion in drug discovery & development* **2002**, *5*, 630-647.

(10) Jahnke, W.; Widmer, H. *Cell Mol Life Sci* **2004**, *61*, 580-599.

(11) Huth, J. R.; Sun, C.; Sauer, D. R.; Hajduk, P. J. *Methods Enzymol* **2005**, *394*, 549-571.

(12) Klages, J.; Coles, M.; Kessler, H. In *NMR-Based Screening in Exploiting in Chemical Diversity for Drug Discovery*; P.A.Bartlett & M.Etzeroth, ed. RCS Publishing: 2006; pp 263-290.

(13) Kessler, H.; Klages, J. In *Comprehensive Medicinal Chemistry II*; Elsevier, Oxford: 2006; pp 901-920.

(14) Luy, B.; Frank, A.; Kessler, H. In *Molecular Drug Properties: Measurement and Prediction*; R.Mannhold, ed. Wiley-VCH, Weinheim: 2008; pp 207-254.

(15) Takeuchi, K.; Wagner, G. *Curr.Opin.Struct.Biol.* **2006**, *16*, 109-117.

(16) Zartler, E. R.; Shapiro, M. J. *Current Pharmaceutical Design* **2006**, *12*, 3963-3972.

(17) Fry, D. C.; Emerson, S. D. *Drug design and discovery* **2000**, *17*, 33.

(18) Klages, J.; Coles, M.; Kessler, H. In *NMR-based Screening*; RCS publishing: 2006.

(19) Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. *Drug Discov. Today* **2005**, *10*, 987-992.

(20) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 199-223.

(21) Zartler, E. R.; Shapiro, M. J. *Current Opinion in Chemical Biology* **2005**, *9*, 366-370.

(22) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531-1534.

(23) Congreve, M.; Carr, R.; Murray, C. W.; Jhoti, H. *Drug Discov Today* **2003**, *8*, 876-877.

(24) Erlanson, D. A.; Hansen, S. K. *Current Opinion in Chemical Biology* **2004**, *8*, 399-406.

(25) Erlanson, D. A.; McDowell R.S.; O'Brien, T. *J Med Chem* **2004**, *47*, 3463-3482.

(26) Gill, A.; Cleasby, A.; Jhoti, H. *ChemBioChem* **2005**, *6*, 506-512.

(27) Hartshorn, M. J.; et al *J Med Chem* **2005**, *48*, 403-413.

(28) Moore, W. R. *Current opinion in drug discovery & development* **2005**, *8*, 355-364.

(29) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat Rev Drug Discov* **2004**, *3*, 660-672.

(30) Schade, M.; Oschkinat H *Current opinion in drug discovery & development* **2005**, *8*, 365-373.

(31) Villar, H. O.; Yan, J.; Hansen, M. R. *Current Opinion in Chemical Biology* **2004**, *8*, 387-391.

(32) Jahnke, W.; Erlanson, D. A. In *Fragment-based Approaches in Drug Discovery*; Wiley-VCH: 2006.

(33) Hajduk, P. J.; Greer, J. *Nat Rev Drug Discov* **2007**, *6*, 211-219.

(34) Langer, T.; et al *ChemBioChem* **2004**, *5*, 1508-1516.

(35) Bax, A. *Protein Sci* **2003**, *12*, 1-16.

(36) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. *Chem.Rev.* **2004**, *104*, 3519-3540.

(37) Luy, B.; Kessler, H. W.G.& D.Craik, ed. Kluwer Academic Publishers: 2007; pp 1261-1267.

(38) Luy, B.; et al *J Am Chem Soc.* **2005**, *127*, 6459-6465.

(39) Klages, J.; Neubauer, C.; Coles, M.; Kessler, H.; Luy, B. *ChemBioChem* **2005**, *6*, 1672-1678.

(40) Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. *J Biomol NMR* **1999**, *13*, 369-374.

(41) Kay, L. E.; Gardner, K. H. *Curr.Opin.Struct.Biol.* **1997**, *7*, 722-731.

(42) Kay, L. E. *J.Magn.Reson.* **2005**, *173*, 193-207.

(43) Szyperski, T.; Yeh, D. C.; Sukumaran, D. K.; Moseley, H. N.; Montelione, G. T. *Proc.Natl.Acad.Sci.U.S.A* **2002**, *99*, 8009-8014.

(44) Cowburn, D.; Muir, T. W. *Methods In Enzymology* **2001**, *339*, 41-54.

(45) Yamazaki, T.; et al *J Am Chem Soc.* **1998**, *120*, 5591-5592.

(46) Romanelli, A.; Shektman, A.; Cowburn, D.; Muir, T. W. *Proc.Natl. Acad.Sci.U.S.A* **2004**, *101*, 6397-6402.

(47) Zuger, S.; Iwai, H. *Nature Biotechnology* **2005**, *23*, 736-740.

(48) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc.Natl.Acad.Sci. USA* **1997**, *94*, 12366-12371.

(49) Pervushin, K. *Q.Rev.Biophys.* **2000**, *33*, 161-197.

(50) Hol, W. G. *Nat Struct Biol* **2000**, *7 Suppl.*, 964-966.

(51) Kupce, E.; Freeman, R. *J.Am.Chem.Soc.* **2003**, *125*, 13958-13959.

(52) Kupce, E.; Freeman, R. *J.Am.Chem.Soc.* **2004**, *126*, 6429-6440.

(53) Hiller, S.; Fiorito, F.; Wüthrich, K.; Wider, G. *Proc.Natl.Acad. Sci.U.S.A* **2005**, *102*, 10876-10881.

(54) Liu, G. H.; et al *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 10487-10492.

(55) Kupce, E.; Freeman, R. *J Biomol NMR* **2004**, *28*, 391-395.

(56) Kupce, E.; Freeman, R. *J Magn Reson* **2003**, *163*, 56-63.

(57) Kupce, E.; Freeman, R. *J Biomol NMR* **2003**, *27*, 383-387.

(58) Luca, S.; Heise, H.; Baldus, M. *Acc.Chem.Res.* **2003**, *36*, 858-865.

(59) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98-102.

(60) Castellani, F.; van Rossum, B. J.; Diehl, A.; Rehbein, K.; Oschkinat, H. *Biochemistry* **2003**, *42*, 11476-11483.

(61) Fernandez, C.; Wuthrich, K. *FEBS Lett* **2003**, *5 55*, 144-150.
(62) Hong, M. *Structure* **2006**, *14*, 1731-1740.
(63) Sanders, C. R.; Sonnichsen, F. *Magn.Reson.Chem.* **2006**, *44 Spec No*, 24-40.
(64) Luca, S.; Heise, H.; Lange, A.; Baldus, M. *Arch.Pharm.(Weinheim)* **2005**, *338*, 217-228.
(65) Zech, S. G.; Olejniczak, E.; Hajduk, P.; Mack, J.; McDermott, A. E. *J Am Chem Soc.* **2004**, *126*, 13948-13953.
(66) Werner, K.; et al *J Biomol NMR* **2007**, *37*, 303-312.
(67) Valler, M. J.; Green, D. *Drug Discov Today* **2000**, *5*, 286-293.
(68) Allen, K. N.; Lavie, A.; Petsko, G. A.; Ringe, D. *Biochemistry* **1995**, *34*, 3742-3749.
(69) Rutenber, E. E.; et al *The Journal of Biological Chemistry* **1993**, *268*, 15343-15346.
(70) Rutenber, E. E.; et al *Bioorganic & Medical Chemistry* **1996**, *4*, 1545-1558.
(71) Shoichet, B. K.; Stroud, R. M.; Santi, D. V.; Kuntz, I. D.; Perry, K. M. *Science* **1993**, *259*, 1445-1450.
(72) Verlinde, C. L. M. J. In *Perspectives in Medicinal Chemistry*; F.E., W. &. G. R., ed. Verlag Helvetica Chimica Acta, Basel: 1993; pp 135-148.
(73) Betz, M.; Saxena, K.; Schwalbe, H. *Current Opinion in Chemical Biology* **2006**, *10*, 219-225.
(74) Koglin, A.; et al *Magn.Reson.Chem.* **2006**, *44 Spec No*, S17-23.
(75) Selenko, P.; Serber, Z.; Gadea, B.; Ruderman, J.; Wagner, G. *Proc. Natl.Acad.Sci.U.S.A* **2006**, *103*, 11904-11909.
(76) Reckel, S.; Lohr, F.; Dotsch, V. *ChemBioChem* **2005**, *6*, 1601-1606.
(77) Sun, C.; Huth, J. R.; Hajduk, P. J. *ChemBioChem* **2005**, *6*, 1592-1600.
(78) Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Mangani, S.; Terni, B.; Turano, P. *Proc.Natl.Acad.Sci. USA* **2005**, *102*, 5334-5339.
(79) Vogtherr, M.; et al *Angew.Chem.Int.Ed* **2006**, *20*, 993-997.
(80) Moustakas, D. T.; et al *J Comput Aided Mol Des* **2006**, *20*, 601-619.
(81) Shoichet, B. K. *Nature* **2004**, *432*, 862-865.
(82) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv.Drug Deliv.Rev.* **2001**, *46*, 3-26.
(83) Lipinski, C. A. *Drug Discov Today* **2003**, *8*, 12-16.
(84) Veber, D. F.; et al *J Med Chem* **2002**, *45*, 2615-2623.
(85) Schieborr, U.; et al *ChemBioChem* **2005**, *6*, 1891-1898.
(86) Bertini, I.; Fragai, M.; Giachetti, A.; Luchinat, C.; Maletta, M.; Parigi, G.; Yeo, K. J. *J.Med.Chem.* **2005**, *48*, 7544-7559.
(87) Arendt, Y.; Bhaumik, A.; Del Conte, R.; Luchinat, C.; Mori, M.; Porcu, M. *ChemMedChem* **2007**, *2*, 1648-1654.
(88) Mayer, M.; Meyer, B. *Angew.Chem.Int.Ed.* **1999**, *38*, 1784-1788.
(89) Mayer, M.; James, T. L. *J Am Chem Soc.* **2004**, *126*, 4453-4460.
(90) Salvatella, X.; et al *Angew.Chem.Int.Ed* **2004**, *43*, 196-198.
(91) Claudino, W. M.; Quattrone, A.; Biganzoli, L.; Pestrin, M.; Bertini, I.; Di Leo, A. *J.Clin.Oncol.* **2007**, *25*, 2840-2846.
(92) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc.Natl.Acad.Sci. USA* **2004**, *101*, 6841-6846.
(93) Gadner, K. H.; Kay, L. E. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 357-406.
(94) Kay, L. E. *Biochem Cell Biol* **1997**, *75*, 1-15.
(95) Schwalbe, H.; et al *Methods In Enzymology* **2001**, *338*, 35-81.
(96) Wuthrich, K. *Biochem Soc Symp* **1981**, 17-37.
(97) Kern, D.; Zuiderweg, E. R. *Curr.Opin.Struct.Biol.* **2003**, *13*, 748-757.
(98) Bertini, I.; Calderone, V.; Fragai, M.; Giachetti, A.; Loconte, M.; Luchinat, C.; Maletta, M.; Nativi, C.; Yeo, K. J. *J.Am.Chem.Soc.* **2007**, *129*, 2466-2475.
(99) Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D. S. *J.Biomol.NMR* **2002**, *22*, 165-173.
(100) Medek, A.; Olejniczak, E. T.; Meadows, R. P.; Fesik, S. W. *J Biomol NMR* **2000**, *18*, 229-238.
(101) Hajduk, P. J.; Gerfin, T.; Boehlen, J.-M.; Haberli, M.; Marek, D.; Fesik, S. W. *J.Med.Chem.* **1999**, *42*, 2315-2317.
(102) Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Q Rev Biophys* **1999**, *88*, 211-240.
(103) Frutos, S.; et al *Biopolymers* **2007**, *88*, 164-173.
(104) Mayer, M.; et al *Chem Biol* **2006**, *13*, 993-1000.
(105) Mayer, M.; James, T. L. *J Am Chem Soc.* **2002**, *124*, 13376-13377.
(106) Mayer, B.; et al *Ernst Schering Res Found Workshop* **2004**, 149-167.
(107) Mayer, M.; Meyer, B. *J.Am.Chem.Soc.* **2001**, *123*, 6108-6117.
(108) Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. *J.Biomol.NMR* **2000**, *18*, 65-68.
(109) Dalvit, C.; Caronni, D.; Mongelli, N.; Veronesi, M.; Vulpetti, A. *Current drug discovery technologies* **2006**, *3*, 115-124.
(110) Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. *J.Biomol.NMR* **2001**, *21*, 349-359.
(111) Jahnke, W.; Rudisser, S.; Zurini, M. *J.Am.Chem.Soc.* **2001**, *123*, 3149-3150.
(112) Vanwetswinkel, S.; Heetebrij, R. J.; van Duynhoven, J.; Hollander, J. G.; Filippov, D. V.; Hajduk, P. J.; Siegal, G. *Chem.Biol.* **2005**, *12*, 207-216.
(113) Vanwetswinkel, S.; Heetebrij, R. J.; van, D. J.; Hollander, J. G.; Filippov, D. V.; Hajduk, P. J.; Siegal, G. *Chem.Biol.* **2005**, *12*, 207-216.
(114) Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W. *J.Am.Chem.Soc.* **1997**, *119*, 12257-12261.
(115) Stebbins, J. L.; Jung, D.; Leone, M.; Zhang, X. K.; Pellecchia, M. *The Journal of Biological Chemistry* **2006**, *281*, 16643-16648.
(116) Fejzo, J.; et al *Chem Biol* **1999**, *6*, 755-769.
(117) Johnson, E. C.; Feher, V. A.; Peng, J. W.; Moore, J. M.; Williamson, J. R. *J Am Chem Soc.* **2003**, *125*, 15724-15725.
(118) Leone, M.; Freeze, H. H.; Chan, C. S.; Pellecchia, M. *Current drug discovery technologies* **2006**, *3*, 91-100.
(119) Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J. *J.Am.Chem.Soc.* **2003**, *125*, 7696-7703.
(120) Tarrago, T.; Kichik, N.; Segui, J.; Giralt, E. *ChemMedChem* **2007**, *2*, 354-359.
(121) Tarrago, T.; Frutos, S.; Rodriguez-Mias, R. A.; Giralt, E.

ChemBioChem **2006**, *7*, 827-833.
(122)Frutos, S.; Tarrago, T.; Giralt, E. *Bioorganic & Medicinal Chemistry Letters* **2006**, *16*, 2677-2681.
(123)Forino, M.; Johnson, S.; Wong, T. Y.; Rozanov, D. V.; Savinov, A. Y.; Li, W.; Fattorusso, R.; Becattini, B.; Orry, A. J.; Jung, D. W.; Abagyan, R. A.; Smith, J. W.; Alibek, K.; Liddington, R. C.; Strongin, A. Y.; Pellecchia, M. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 9499-9504.
(124)Fattorusso, R.; Frutos, S.; Sun, X.; Sucher, N. J.; Pellecchia, M. *Phytomedicine* **2006**, *13*, 16-22.
(125)Dalvit, C.; et al *J Am Chem Soc.* **2005**, *127*, 13380-13385.
(126)Dalvit, C.; Ardini, E.; Fogliatto, G. P.; Mongelli, N.; Veronesi, M. *Drug Discovery Today* **2004**, *9*, 595-602.
(127)Dalvit, C.; Flocco, M.; Stockman, B. J.; Veronesi, M. *Combinatorial chemistry & high throughput screening* **2002**, *5*, 645-650.
(128)Fattorusso, R.; Jung, D.; Crowell, K. J.; Forino, M.; Pellecchia, M. *J Med Chem* **2005**, *48*, 1649-1656.
(129)Papeo, G.; et al *J Am Chem Soc.* **2007**, *129*, 5665-5672.
(130)Dalvit, C.; Hadden, D. T.; Sarver, R. W.; Ho, A. M.; Stockman, B. J. *Combinatorial chemistry & high throughput screening* **2003**, *6*, 445-453.
(131)Price, S. W. *Concepts Magn.Reson.* **1997**, *9*, 299-336.
(132)Price, S. W. *Concepts Magn.Reson.* **1998**, *10*, 197-237.
(133)Dehner, A.; Kessler, H. *ChemBioChem* **2005**, *6*, 1550-1565.
(134)Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. *J.Am.Chem.Soc.* **1997**, *119*, 5818-5827.
(135)Petros, A. M.; et al *J Med Chem* **2006**, *49*, 656-663.
(136)Oltersdorf, T.; et al *Nature* **2005**, *435*, 677-681.
(137)Jahnke, W.; Perez, L. B.; Paris, C. G.; Strauss, A.; Fendrich, G.; Nalin, C. M. *J.Am.Chem.Soc.* **2000**, *122*, 7394-7395.
(138)Vazquez, J.; Tautz, L.; Ryan, J. J.; Vuori, K.; Mustelin, T.; Pellecchia, M. *J.Med.Chem.* **2007**, *50*, 2137-2143.
(139)Becattini, B.; et al *Chem Biol* **2004**, *11*, 1107-1117.
(140)Becattini, B.; Pellecchia, M. *Chemistry* **2006**, *12*, 2658-2662.
(141)Chen, J.; et al *ACS chemical biology* **2007**, *2*, 329-336.
(142)Carulla, N.; et al *Nature* **2005**, *436*, 554-558.
(143)Sanchez-Pedregal, V. M.; et al *Angew.Chem.Int.Ed* **2005**, *44*, 4172-4175.

Chapter 5

Protein structure determination by magic angle spinning solid-state NMR today and in the future: toward systems biology

Authors: Hartmut Oschkinat[a], Mark Baldus[b], Anja Böckmann[c], Frank Engelke[d], Gil Goobes[e], Claudio Luchinat[f], Beat H. Meier[g], Bernd Reif[a], Barth van Rossum[a]

[a] Forschungsinstitut für Molekulare Pharmakologie, Berlin, Germany
[b] Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[c] University of Lyon, France
[d] Bruker Biospin, Rheinstetten, Germany
[e] Bar Ilan University, Rimat, Israel
[f] Magnetic Resonance Center, University of Florence, Italy
[g] ETH Hoenggerberg, Zurich, Switzerland

The function of a living cell relies on its heterogeneous composition, involving membranes separating specialized compartments such as organelles and the cytosol, or cytoskeletal proteins for maintaining its shape, localizing individual protein complexes and enabling active transport. Both membranes and cytoskeletal features may be considered as quasi-solid objects due to their large size and the concomitant anisotropic and slow motions in solution. Solid-state NMR is a method that may be applied to non-crystalline and non-purified samples of such quasi-solid cellular features, independent of size, allowing the tackling of so far unsolved problems.

An outlook on the development of solid-state NMR will naturally require some balance between what we would like to do and what we can do at the moment, even though the latter might have increased a lot in the last years. Starting originally with model systems, informative structures of protofibrils and a structure of αB-crystallin as a highly dynamical, polydisperse complex are now available, and structural investigations on membrane proteins aiming for a de-novo structure are under way. The current experience allows us to extrapolate that a wide area of yet unconquered ground in structural biology will be amenable to solid-state NMR investigations, including a number of 'old' structural problems representing heterogeneous, multi-component systems, providing that some experimental obstacles can be overcome and that technologies and concepts will be developed to speed up the process of structure determination considerably. Solid-state NMR thereby offers a special potential for studying structural rearrangements on a variety of time scales, even enabling correlations with thermodynamical parameters. Along these lines, membrane protein systems are an obvious target of solid-state NMR, and in particular the structural rearrangements and their dynamics in native lipid environments. Looking at the X-ray structures of the ß-adrenergic receptor and the many mutations applied to generate a crystallizing construct, subsequent investigations by solid-state NMR seem mandatory, including the characterization of the plasticity of its membrane-integrated features. Furthermore, a wide variety of biological functions is associated with the appearance of heterogeneous, "dynamical" complexes that are very difficult to prepare in pure states by in-vitro methods. Examples are actin networks which bear monomers with nucleotides in different phosphorylation states. It would be of great interest, for example, to see the structure of nebulin in complex with f-actin, being largely unstructured in an aqueous environment. Other examples are the tubulin-associated proteins that are highly flexible in aqueous solutions but presumably structured when bound to tubulin fibres. Complexes and molecular machineries attached to the cytoskeleton are often large and polydisperse, like, for example, focal adhesions, making them a good target for solid-state NMR. The same holds for small heat shock proteins, which form large homo- or hetero-oligomeric assemblies. Structural investigations on such systems are difficult, since crystallization is hampered by their polydisperse nature, being also too large for the application of solution NMR. Many of these systems are of high medical relevance such as the small Heat Shock Protein αB-crystalline which is involved in cardiomyopathies, multiple sclerosis and Alzheimer's Disease; mutations also affect its function in the eye-lens, through the formation of cataracts.

While structure determination on such systems would currently represent the next step in solid-state NMR research, it is worthwhile to judge its potential while looking at a wider

context. Coherent with a systems-oriented view employed in current biological research, a new type of structural biology that integrates information obtained by high- and low-resolution techniques on the same objects is highly desirable: native-close preparations which maintain a high degree of their physiological functionality. However, a gap exists in current structural biology between the global information delivered by laser scanning microscopy or other optical techniques using 'systems-like' samples, and the techniques that yield high-resolution structural information such as X-ray crystallography, NMR and in part electron microscopy. EM-tomography was suggested recently as one technique that is able to close the gap, yielding sufficiently resolved displays of molecular machineries inside cells, at membranes and around the cytoskeleton. Potentially, solid-state NMR can be applied to similar samples, providing portions can be labelled with magnetically active nuclei such as ^{13}C and ^{15}N. This may be accomplished by reconstituting functional complexes on templates prepared from natural sources such as membranes, microtubules and actin fibres. Another prerequisite is that signal-to-noise can be increased by a factor of 50-100 through the application of dynamic nuclear polarisation (DNP) (see Chapter 7 in this volume). A combination of LSM, EM-Tomography and solid-state MAS/DNP NMR has the attractive potential to deliver this structural information from the micrometer to the Angstrom scale, also making use of X-ray structures deposited in the PDB, while maintaining a native-close state of the samples. Applying these technologies, 'systems biology' studies of whole drosophila eyes, for example, can be envisaged, focusing on the environment of the isotope-labelled retinal in the various contained rhodopsins, and aiming at obtaining a picture of the isomerisation processes taking place upon irradiation of light with a monochromatic laser beam.

The current stage of solid-state NMR, however, is 'doing home work' as a preparation for these studies, through aiming at a first membrane protein structure, by investigating cytoskeleton-attached complexes, and the like. The X-ray and solution NMR communities neeed to take a series of small steps, and these must be taken in turn by the solid-state NMR spectroscopists.

The various brands of solid-state NMR: magic-angle-spinning, oriented samples, and related techniques

The most commonly used way to record NMR spectra of quasi-solid samples involves the application of magic-angle spinning. Rotating a sample about an axis at 54.7 degrees with respect to the external magnetic field leads to an averaging of dipolar interactions or chemical shift anisotropy. If the rotation is about three times the size of the interactions, the latter are usually averaged to zero. This method produces spectra with narrow lines of quasi-solid biological samples, and allows for the application of an arsenal of pulse techniques for achieving chemical shift assignments and for collecting structural constraints to be used in structure calculation. Due its experimental variability and the richness of the data, de-novo structure determination independent of additional knowledge is possible [1;2].

There is a light version of magic-angle spinning NMR, HR-MAS, which involves hardware developed initially to validate reaction intermediates in resin-based syntheses. HR MAS hardware is characterized by a probe head design that enables magic-angle spinning combined with pulse field gradients but using solution-like coils and circuits. High power proton decoupling is prohibitive. Today, the same kind of hardware can in principle be used in the acquisition of solid-state NMR spectra of samples for which no high power ^{1}H decoupling is required, i.e. perdeuterated peptides and proteins.

Another successful method specific to membrane proteins involves the use of static, oriented membrane samples. For proteins that are embedded in mechanically or magnetically oriented lipid bilayers, it is possible to retain the information-rich dipolar couplings and chemical shift anisotropies while maintaining reasonable spectral resolution [3-5]. In this case, the partial mobility and the well-defined orientation relative to the external magnetic field B_0 produces good spectral resolution, even without MAS. Extensive chemical shift assignments have been made for proteins oriented in bilayers, such as gramicidin A [6;7] the M2 transmembrane helix of the acetylcholine receptor [8] and the coat protein of fd bacteriophage [9]. These experiments show the orientation of protein structure elements such as α-helices [10] or β-barrels [11] with respect to the membrane, and provide useful information about the orientation of the cytoplasmic domain of phospholamban [12] in membrane bilayers. The chemical shifts measured in this type of experiment have isotropic and anisotropic contributions, where the anisotropic part yields information about the orientation of the chemical shift tensor relative to B_0. Because peptide bonds are nearly planar, measuring chemical shift anisotropies of the carbonyl ^{13}C and amide ^{15}N provides a direct measurement of the orientation of each peptide plane relative to the magnetic field. The amide ^{1}H-^{15}N dipolar couplings are related to the internuclear distance and the orientation of the internuclear vector with respect to B0. This principle has been used to generate

multidimensional correlation spectra with 15N chemical shift in one dimension and 1H-15N or 1H-13C dipolar coupling in the other [13-15].

NMR spectra with isotropic-anisotropic correlations can also be measured by varying the angle of the sample spinning with respect to B_0, either in a series of discrete one-dimensional spectra as in variable angle spinning (VAS) or in two correlated Fourier dimensions with a fast hop during each scan, as in switched-angle spinning (SAS). VAS is more readily implemented experimentally, because it does not require rapid reorientation of the sample. Correlations between isotropic and anisotropic chemical shifts can be obtained after data processing and reconstruction using many one-dimensional spectra, yielding information about structure and dynamics in solids [16]. Variable angle spinning experiments have also been used to measure scaled dipolar couplings in liquid crystals.

Recent developments in solid-state MAS NMR protein structure determination

In the past years, solid-state NMR has accumulated a considerable track record with respect to *de novo*-structure determination of proteins (Table 1). The deposited structures were obtained mostly on nano- or microcrystalline preparations of soluble proteins and on protofibrils. The work on soluble proteins such as the SH3 domain of β-spectrin, ubiquitin, the β1 immunoglobulin binding domain of protein G (GB1), Kaliotoxin, and Crh were mostly methodological exercises, demonstrating the advancement of the field [1;17-24]. Looking at the structure statistics, it appears that there are now first and second generation structures, characterized by an increasing number of constraints, and different types of constraints. The original structure of the SH3 domain was determined with 286 and 6 carbon-carbon and nitrogen-nitrogen inter-residue constraints, respectively, whereas the recently published structure of the domain-swapped dimer of the CRH domain, for example, was determined by using 643 distance constraints, combining carbon-carbon and proton-proton distances. The GB1 structure determined by Franks et al., including also orientational constraints, and in total 7826 distance constraints is another example of the advancement of technology.

The structural investigations on protofibrils also reflect the development in methodology. Initial investigations by the group of Rob Tycko yielded a model of $A\beta_{1-40}$ at low resolution [25] and investigations by Marc Baldus on α-synuclein fibrils revealed the quaternary structure arrangement [26]. Coordinates describing a well determined structure of the monomeric subunit of fibrils formed by a short peptide derived from transthyretin were deposited by the group of Griffin [27] and structures of the monomeric subunits of peptides with a length of approximately 35 amino acids were published for the protofilaments formed by a β-2-microglobulin fragment and of the WW domain of CA 150 [28;29]. In 2008, the group of Beat Meier published the structure of fibrils formed by the protein Het-s, again reaching a new level of technology (Figure 1)[30].

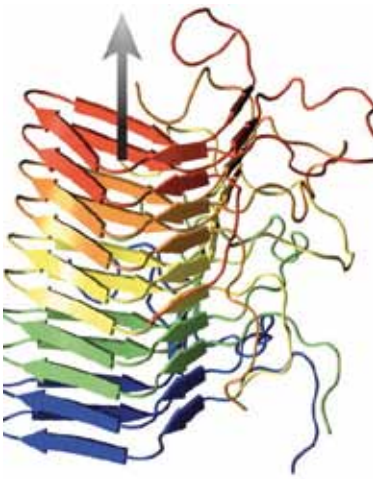


Figure 1: Structure of the HET-s(218-289) fibrils as achieved by solid-state NMR (Wasmer C, Lange A, Van Meickebeke H, Siemer AB, Riek R, Meier BH, Science, 319:1523-1526, 2008).

Along with these investigations that yielded independently determined 3-dimensional structures, investigations on membrane protein systems are on the way. As an intermediate approach, models are generated by chemical shift analysis and subsequent molecular modelling. These include, for example, structural models of two peptide hormones bound to G-protein-coupled receptors derived by the groups of Baldus and Glaubitz, respectively, which are of interest to pharmacologists and researchers in the drug design area [31]. The group of Baldus has published a number of very interesting membrane protein models derived by a combination of solid-state NMR, utilizing chemical shift changes, and molecular modelling, that enabled a deeper understanding of, for example, potassium channel activation and inhibition (Figure 2) [32;33], or of the plasticity of the PAS domain when interacting with the histidine kinase DcuS [34]. Furthermore, the group of Chad Rienstra has published extensive assignments of the membrane protein DsbB [35]. Partial assignments of the outer membrane protein G (OmpG) of *E.coli* were published by the group of Oschkinat [36].

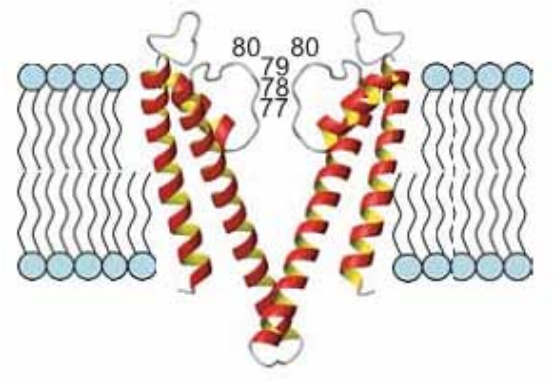


Figure 2: Solid-state NMR structure of a membrane-bound potassium channel (Lange A, Giller K, Hornig S, Martin-Eauclaire MF, Pongs O, Becker S, Baldus M, Nature, 440:959-962, 2006).

tions might also be useful, as e.g. in the case of αB-crystalline[39]. Protofibrils may be investigated as such after precipitation according to recipes derived by the individual biologists.

With regard to the most suitable preparations of membrane proteins, various options exist. It is possible, of course, to investigate frozen solutions of detergent-solubilised proteins[40;41]. Preferred are, however, liposomes prepared from various kinds of lipids, such as bacterial or soja bean extracts[32;42], or two-dimensional crystals using native lipids such as *E.coli* proteins embedded in *E.coli* membranes. With regard to the various possibilities of preparing membrane proteins, it seems beneficial to investigate their structures in relatively native environments which speaks for liposomes or two-dimensional crystals prepared using native lipids. Another possibility is the use

Table 1

Protein	PDB	#S	#AA	All Dist	Eff. Con	Dist str.	T H	rmsd	Experiments	Reference
CRYSTALLINE PROTEINS				Con or CP	H-H	C-C, N-N, C-N		Ensemble / **X-ray**		
Spectrin SH3 domain	1M8M	4	62	292	--	292	--	1.6*/**2.6***	PDSD	17
Spectrin SH3 domain		4	62	889	--	883	T	0.7*/**1.2***	3D NCC	23
Kaliotoxin	1XSW	1	38	28	28	--	T	0.8/(**1.9**)	CHHC	1
Ubiquitin		2	76	336	--	336	T	0.7*/ **?**	PDSD, DARR	24
GB1	2JU6	1	56	517	517	--	T	0.8 /**1.9**	3D NCC,2H-Protein	18
GB1	2JSV	5	56	7826	190	607	T	0.3 /**1.4**	3D NCC, CHHC	19
Ubiquitin	2JZZ	1	76	996	--	981	T	0.6 /**1.4**	PDSD/Candid/Atnos	20
Crh domain, dimer	2RLZ	1	85	1002	643	--	T	0.8 /**1.6**	CHHC, NHHC, Aria	21
Kaliotoxin	2KTX	2	38	290	290	--	T	0.8/(**0.6**)	CHHC, NHHC	22
PROTEINS FORMING FIBRILS										
Transthyretin in amyloid fibrils	1RVS	3	10	35	--	35	T	0.7 / --	C-N, 2D Tedor	27
Protofibrils of beta2-microglobulin fragm.	2E8D	2	22	27	--	27	TH	1.4 / --	DARR	28
Protofibrils of the CA150 WW domain	2NNT	3	40	37	--	37	H	1.3 / --	PDSD, DARR	29
Amyloid fibrils of the HET-s(218-289) prion	2RNM	3	71	134	44	90	TH	0.4 / --	CHHC, PDSD, NHHC	30

In general, research in this field is still concentrated on achieving resonance assignments and improving concepts for structure determination. Of considerable interest is the logarithmic curve with which the number of structures is increasing, and the improvement in methodology.

Sample preparation

With regard to sample preparations, a large arsenal of possibilities is applied, with a clear tendency to investigate functionally relevant states. For small proteins, the preparation of nano- or micro-crystals is still an option for achieving well resolved spectra[37;38], and depending on the situation, PEG-precipita-

of nano discs as investigated by the group of Rienstra which yielded spectra of high quality[43].

The current structure determination paradigm in MAS solid-state NMR

The structures determined so far concerned small proteins (< 100 AA), which were either crystallized in their native state or investigated as protofibrils. For such applications, well-defined solid-state NMR assignment strategies exist, as well as concepts for obtaining distance constraints, utilizing mainly ^{13}C and ^{15}N chemical shifts. Stepping further towards membrane proteins or cytoskeleton-attached complexes, assignment

concepts need to be adapted and tuned to the particular problem, e.g. by including specially labelled preparations, and making use of proton chemical shifts at a larger scale.

The structure determination concept for small soluble proteins involves a minimum number of samples which are either uniformly labelled with ^{13}C and ^{15}N or generated by using 1,3- and 2-^{13}C-glycerol as carbon sources during the growth of the bacteria, apart from ^{15}N. All three types of samples are used for both, to achieve resonance assignments and extract distance constraints. Special samples are additionally prepared when intermolecular contacts play a large role. In nano- or microcrystalline proteins, crystal packing may lead to the observation of large effects in carbon-carbon correlation spectra. In such cases, inter- and intra-molecular constraints may be distinguished on the basis of samples made from a mixture of labelled and unlabelled protein (e.g. 1:4). The direct determination of intermolecular constraints is achieved best by producing samples containing mixtures of ^{15}N and ^{13}C-labelled protein, with the option of diluting naturally abundant ^{13}C in the ^{15}N-sample, and vice versa.

With respect to resonance assignments, three-dimensional correlation techniques such as NCACX and NCOCX (or CB instead of CX) yield the bulk of assignments. The pulse sequences usually contain PDSD, DARR or PAR-type carbon-carbon mixing units, while carbon-nitrogen transfer is achieved by selective CP. The CB-versions of the experiments employ double-quantum mixing, such as, for example, DREAM. The combined evaluation of the 3D-experiments allows for so-called sequential walks along like nitrogen and like-C', Cα or Cβ frequencies when superposing NCO- and NCA-type spectra. The labelling pattern of the glycerol-made samples may also be taken into account in the interpretation to remove ambiguities and to resolve overlap. Furthermore, cross peaks involving correlations between α-carbons have proven to be of high value for the sequential assignment procedure when the ^{15}N lines are too broad. Using such procedures, resonance assignments to a high degree of completeness have been obtained for a domain of 100 amino acids from a 170 residue protein [39]. With regard to the measurement of distance constraints, there are various options. First, one may decide for proton-proton or carbon-carbon distances, or both. Secondly, these may be determined at different levels of accuracy. Applying three-dimensional techniques, only generous upper bounds may be obtained due to the many factors determining individual signal intensities in solid-state NMR, no matter which type of constraints are measured. Extracting them from 2D-spectra allows for rough estimations of distance ranges, especially when some spin dilution is applied. In the papers describing the determination of those solid-state structures that are deposited in the PDB, carbon-carbon constraints were extracted from PDSD, DARR and PAR spectra, and proton-proton constraints from CHHC or NHHC techniques. The NHHC techniques play a particular role in the determination of intermolecular constraints on samples with mixed labelling, i.e. one sort of molecule is ^{15}N-labelled and the other is ^{13}C-labelled.

A problem occurs when increasing the number of amino acids into a range which covers small membrane proteins or interesting protein complexes. In such cases amino acid specific labelling – or unlabelling – is required for resolving spectral overlap. This has been particularly fruitful in the assignment process of sensory rhodopsin [44] or DsbB [35]. Specially labelled amino acids have also been used for this purpose [36] or combinations of precursor-driven labelling with either labelled or unlabelled amino acids. In those cases where sufficient assignments could be achieved for extracting the required distance or angular constraints for structure calculation, the bulk of distance constraints were obtained from carbon-carbon correlations or CHHC-type spectra recorded with very short CP and proton-proton exchange times. Recently, the DARR and PAR techniques are used in conjunction with PDSD. Investigations by Marc Baldus and Anja Bockmann have demonstrated the utility of proton-proton constraints [1;21]. Further constraints e.g. between carbons and nitrogens may be useful to define hydrogen bonds or salt bridges and other features important for the integrity of protein structure, measured by PAIN, for example.

As to membrane proteins and cytoskeletal complexes, data sets will be evidently more sparse, and there will always be a fight to achieve a sufficient amount of distant constraints. In particular, CHHC techniques might not work in all cases for sensitivity reasons, and C-N constraints may be difficult to extract from crowded correlations. The basis for structure determination will probably be carbon-carbon constraints from biosynthetically directed dilution of carbon spins by using 1,3- or 2-^{13}C-glycerol as precursors for making the samples, or other types of carbon dilutions. In addition, protons for dispersing the chemical shifts will be necessary to exploit more carbon nitrogen constraints.

Protein dynamics by magic-angle-spinning NMR

At present, the characterization of dynamic processes in biological macromolecules tends to be a field of solution-state

NMR spectroscopy. However, solid-state NMR spectroscopy also holds a great potential for the study of these processes. So far, it is not very well understood why the spectra of a few amyloidogenic peptides or membrane proteins are characterized by very broad resonance lines [25;45] whereas others display a very favorable spectral quality [26;46;47]. The assumption that dynamic processes may be the origin of this line broadening is corroborated by studies which demonstrate that molecules in an amyloid fibril can undergo chemical exchange between an amyloid fibril-type conformation and a soluble molecule [48;49]. Analogous to solution-state NMR spectroscopy, dynamic processes result in averaging of the H^{α}/H^{β} spin states in standard ^{1}H decoupled solid-state NMR experiments, with all the detrimental consequences to the spectra that are observed in solution-state NMR for very large proteins. On the other hand, interference between different relaxation mechanisms can be employed to manipulate the relaxation properties of a certain spin. These effects were explored with great success by Wüthrich and co-workers and resulted in the development of the TROSY (Transverse Relaxation Optimized Spectroscopy) technique [50]. In the solid-state, dynamics occurr even in crystalline proteins [51;52]. Motional effects in the protein backbone are quantified using ^{15}N-T_1 [53;54], ^{15}N-CSA, ^{1}H-^{15}N dipole cross correlated relaxation [55] and direct assessment of order parameters via measurement of dipolar couplings [56;57]. In addition, side chain motional properties are accessible by exploiting the spin-1 properties of the deuterons in a perdeuterated protein [58;59]. Interestingly, side chain methyl relaxation rates turn out to match very well the relaxation rates found in the solution-state [60;61]. This opens up the possibility for a combined analysis of solution-state and solid-state NMR data [62]. Furthermore, deuteration allows the characterization of water molecules which are bound to a crystalline protein [63-65].

Signal-to-noise, dynamic nuclear polarisation and other experimental advances

In comparison to conventional samples used so far in liquid-state NMR, the rotors commonly used for solid-state NMR experiments (1.3-4.0 mm diameter) provide less productive volume which, in the case of membrane proteins, needs to be shared with water, detergents and lipids. Furthermore, the cryoprobe technology now well-established in solution NMR has not yet found a clear equivalent in solid-state NMR, leaving a gap in sensitivity. In addition, the most interesting applications of solid-state NMR involve larger proteins, or complexes with large components like actin or tubulin fibres. This leads to a constant battle for better signal-to-noise, and demands for further instrumental developments. One recent development potentially affecting signal-to-noise is concerned with the reduction of sample heating through proton decoupling by reducing the effect of the electrical component of the radiofrequency field and thus allowing for higher decoupling fields. Initial experiments are promising, but in practice, the so-called e-free or low-e probes are not yet allowing the application of sufficient power for these purposes. An equivalent increase in sample rotation as an alternative is hampered by the required reduction in volume. An approach equivalent to solution NMR, using so-called cryoprobes, is very interesting but not free of experimental pitfalls. In principle, and as a special advantage of solid-state NMR, a gain of sensitivity through cooling the sample down to liquid helium or nitrogen temperatures by exploiting the Boltzmann factor should be easily achievable. However, in most cases an increase in line width was observed so far, compensating the gain. This demands cryoprobes employing cooling of the detection circuit and the preamp, hoping for an improvement of signal-to-noise by a factor of 2-4. At the same time, biological experiments could be done at native temperatures.

Considering the vision of investigating samples that are closer to a native situation, a substantial increase in signal-to-noise is required, for which a generally applicable solution was proposed by the pioneering work of Bob Griffin: dynamic nuclear polarization. Keeping all factors constant, enhancements in the range of 20-100 were observed on biological samples, opening new perspectives for solid-state NMR in biological basic research [66-68]. The current state-of-the-art involves the exploitation of the 'three-spin solid effect' via application of biradicals [69] and measurements at liquid nitrogen temperatures (see Chapter 7 in this volume). Samples need to be prepared such that a 'glass' is formed by protein, radicals and solvent, and mixtures of DMSO or glycerol with water form such glasses, ensuring in this way slow relaxation of the electron spin resonance. Currently, variants of Tempo, and in particular Totapol [70] are used as biradicals, and deuteration of the solvent is required, again for ensuring slow relaxation of the electron spin resonance as one of the prerequisites for optimal polarisation transfer. The most interesting demonstration of the technology were investigations on the membrane protein bacteriorhodopsin which yielded new insights into intermediate states occurring during the photo cycle, and where enhancements of a factor of 40 were observed for resonances of the retinal which is in the interior of the protein [68]. From this investigation it can be concluded that enhancements may be achieved over a longer distance from the radical-containing area, into a membrane-embedded protein.

Deuteration versus higher spinning frequencies: Protons in biological solid-state NMR

Recently, various attempts have been made to include proton chemical shifts and distances between protons into solid-state NMR assignment and structure determination concepts. There are in principle two different ways to allow for the measurement of sufficiently well-resolved proton spectra of proteins, applying either extensive deuteration and allowing for a few protons at selected sites [51] or using very high spinning frequencies, up to 60 kHz, while still applying extensive deuteration at non-exchangeable sites, but not at exchangeable sites, if experimentally achievable [18;71-73].

Extensive deuteration is applied by Bernd Reif and coworkers who demonstrated that line widths of around 20 Hz for amide proton signals can be obtained when applying 100% deuteration of non-exchangeable sites, and reinstalling protons to 10% at exchangeable sites [74]. These line widths are obtained at a spinning frequency of 32 kHz, using 3.2 mm rotors. It was demonstrated that solution-like correlation techniques employing INEPT transfer steps such as HNCA, HN(CO)CA, etc. may be applied to record a data set suitable for resonance assignment, while carbon-carbon correlation steps require some attention due to the absence of protons. The other option for achieving sufficient resolution in proton spectra is to apply ultra-high spinning, e.g. 40-60 kHz, using 1.3-1.6 mm rotors, for example, and fully back-exchanged protons in an otherwise perdeuterated protein. The group of Chad Rienstra has demonstrated that at these spinning frequencies reasonably resolved correlations can be obtained [18;71-73]. The proton line width of NH moieties was observed to be 140 $\pm$ 30 Hz, and for the respective nitrogens 37 $\pm$ 5 Hz. Surprisingly, the overall transverse relaxation time of the proton lines was estimated to be only 45 Hz, as determined by direct T_2 measurements. The line widths of the proton signals depended in a linear manner on the spinning frequency, leading to the expectation that it is doubled when spinning only at 20 kHz. Pulse sequences applied to such samples would contain cross polarisation units rather than INEPT steps in comparison to the concept published by Reif. However, signal intensities and the apparent line widths are suitable for achieving assignments and in the end also extracting structural constraints.

Both approaches have their special merits; the application of solution-like sequences certainly requires sparse protonation in the 10-20% range since the critical step of correlating through the CN bond via an INEPT mixing step to achieve sequential correlations is otherwise not effective.

A special issue occurring in the solid-state NMR of proteins is water suppression, since water signals are broad, and in case of 100% protonation of exchangeable sites, also large. First attempts by Kurt Zilm, employing double CP with a z-filter, demonstrate efficient water suppression. However, it is evident that the development of solid-state NMR probes needs to be continued such that gradients are available on triple resonance probes to exploit the full potential of proton NMR. Using gradients, the group of Rienstra published water suppression schemes based on Mississippi, while others employed WATERGATE sequences.

More restraints: Exploiting the angular dependency of parameters

Similar to solution NMR, where information on the angular dependence of parameters is now included in structure determination concepts, solid-state NMR should take advantage of these parameters in a richer manner. The group of Rienstra has demonstrated that relative orientations of individual moieties may be obtained from the dipolar line shapes of small proteins, improving the position of the structural ensemble by a factor of 2^{19}. The group presented 3D experiments for the measurement of the dipolar line shapes. It can be envisaged that in principle 4D experiments would be necessary for larger proteins which however, should be feasible due to the few numbers of points that are required to obtain a reasonable fit.

Another very promising approach is based on the clever use of paramagnetic centres (typically transition metals or lanthanides), either present in the molecules or added to the sample. The characteristics of these paramagnetic centres are however opposite to those required for DNP. The latter needs paramagnets with the longest possible electron relaxation rates (typically organic radicals) which are achieved because these systems have very little if any magnetic anisotropy. On the contrary, high magnetic anisotropy denotes the presence of strong electron spin-orbit coupling and therefore fast electron relaxation rates. This high magnetic anisotropy, however, causes pseudocontact shifts (pcs) on surrounding nuclei. Pcs are the part of the dipolar interaction between electrons and nuclei that is not averaged to zero by molecular tumbling (in solution) or by MAS (in the solid state). Pcs have a very precise and quantitative dependence on structural parameters, namely the polar coordinates of the nucleus in the reference axis system of the magnetic anisotropy tensor of the paramagnetic centre [75].

Pcs have long been known to constitute accurate structural restraints in solution [76], and theory predicts that pcs should be

equally well observable in the solid state under MAS conditions. Recently, an experimental verification of the underlying theory for solid state samples has been obtained by measuring a number of pcs in a paramagnetic metalloprotein [77]. Moreover, it turns out that the measurement of pcs in the solid state is as easy as in solution, i.e. by the difference between the paramagnetic sample and a diamagnetic reference [78]. On the test protein of 17 kDa in a microcrystalline state, more than 300 pcs could be measured straightforwardly, and used for structural purposes [79].

Dipolar couplings between the unpaired electron(s) and the surrounding nuclei also cause nuclear relaxation, in the form of both longitudinal relaxation enhancement and line broadening. The former can be measured in the solid state with the same sequences adopted in solution, and contain quantitative distance information from the metal centre. Relaxation-based restraints are also being introduced in today's solid state restraints weaponry [80].

Perspectives

Magic-angle spinning solid-state NMR has progressed tremendously towards structure determination of larger systems. Recent progress has demonstrated that in comparison to solution NMR, a higher number of restraints may be collected, including new restraint types, offering the potential for determining higher resolution structures. Obstacles are on the experimental side, regarding sensitivity of probes and inclusion of protons into structure determination concepts, and on the software side. Considering that solid-state NMR should be particularly useful to investigate large protein complexes and membrane proteins, the exploitation of its full potential will largely depend in the future largely on the availability of suitable software aiding in the assignment process and facilitating the extraction of structural parameters.

Applying existing and future technologies, in particular DNP technologies, studies on 'systems biology samples' can be envisaged, consisting mainly of endogeneous material with a few components labelled, either through biosynthetic pathways or through reconstitution steps. Along these lines, investigations on whole drosophila eyes will for instance be feasible, focusing on the environment of isotope-labelled retinal in the various rhodopsins, and aiming at obtaining a picture of the isomerisation processes taking place in the eyes upon irradiation of light with a monochromatic laser beam.

Synaptosomes, either investigated directly after isolation, or after steps of reconstitution of membrane-associated macromolecular complexes are also potential objects for structural systems biology. Synaptosomes are hollow membrane preparations from nerve tissue comprising the pre- and postsynaptic membranes, connected by protein complexes in the synaptic cleft. They contain mainly receptors, cytoskeletal, matrix and adhesion proteins, all associated with the membrane, and are competent for uptake, release and storage of neurotransmitters. The function of synapses depends critically on the well-tuned interplay of processes involving cytoskeletal and adhesion proteins, and receptors. A typical experiment will thus be the study of the homer/shank/PSD95-involving protein complexes, and their association with receptors and cytoskeletal proteins at the post-synaptic side of the synaptosomes. One first aim of the solid-state NMR study would be the unravelling of the native interaction partners of the individual PSD95 PDZ domains, by reconstituting the complexes into the isolated synaptosomes.

It is evident that these very challenging 'dream experiments' are only feasible if the methodological advancements remains intense. The introduction of DNP is one example, and current results suggest a major role for this technique. Another new technique potentially important in the context of systems biology is magnetic resonance force microscopy [81] which allows the investigation of topological features of biological solids. It will allow for new types of contrast without using contrast agents or fluorescent groups such as GFP.

References

(1) Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M. *Angew.Chem.Int.Ed.* **2005**, *44*, 2089-2092.

(2) Bockmann, A. *Angew.Chem Int.Ed Engl.* **2008**, *47*, 6110-6113.

(3) Nielsen, N. C.; Malmendal, A.; Vosegaard, T. *Molecular Membrane Biology* **2004**, *21*, 129-141.

(4) Opella, S. J.; Marassi, F. M. *Chemical Reviews* **2004**, *104*, 3587-3606.

(5) De Angelis, A. A.; Jones, D. H.; Grant, C. V.; Park, S. H.; Mesleh, M. F.; Opella, S. J. *In Nuclear Magnetic Resonance of Biological Macromolecules, Part C* **2005**, *394*, 350-382.

(6) Cross, T. A. *In Solid-Phase-Peptide Synthesis* **1997**, *289*, 672-696.

(7) Kovacs, F.; Quine, J.; Cross, T. A. *Proc.Natl.Acad.Sci.U.S.A* **1999**, *96*, 7910-7915.

(8) Opella, S. J.; Marassi, F. M.; Gesell, J.; Valente, A. P.; Kim, Y.; Oblatt-Montal, M.; Montal, M. *Nature Structural Biology* **1999**, *6*, 374-379.

(9) Zeri, A. C.; Mesleh, M. F.; Nevzorov, A. A.; Opella, S. J. *Proc.Natl. Acad.Sci.U.S.A* **2003**, *100*, 6458-6463.

(10) Marassi, F. M.; Opella, S. J. *J Magn Reson* **2000**, *144*, 150-155.

(11) Marassi, F. M. *Biophysical Journal* **2001**, *80*, 994-1003.

(12) Abu-Baker, S.; Lu, J. X.; Chu, S.; Shetty, K. K.; Gor'kov, P. L.; Lorigan, G. A. *Protein Sci.* **2007**, *16*, 2345-2349.

(13) Ramamoorthy, A.; Wu, C. H.; Opella, S. J. *J Magn Reson B* **1995**, *107*, 88-90.

(14) Ramamoorthy, A.; Wu, C. H.; Opella, S. J. *J.Am.Chem.Soc.* **1997**, *119*, 10479-10486.

(15) Gu, Z. T.; Opella, S. J. *J Magn Reson* **1999**, *138*, 198.

(16) Frydman, L.; Chingas, G. C.; Lee, Y. K.; Grandinetti, P.; Eastman, M. A.; Barral, G. A.; Pines, A. *Israel J.Chem.* **1992**, *32*, 161-164.

(17) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98-102.

(18) Zhou, D. H.; Shea, J. J.; Nieuwkoop, A. J.; Franks, W. T.; Wylie, B. J.; Mullen, C.; Sandoz, D.; Rienstra, C. M. *Angew.Chem Int.Ed Engl.* **2007**, *46*, 8380-8383.

(19) Franks, W. T.; Wylie, B. J.; Schmid, H. L.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M. *Proc.Natl.Acad.Sci.U.S.A* **2008**, *105*, 4621-4626.

(20) Manolikas, T.; Herrmann, T.; Meier, B. H. *J Am Chem Soc.* **2008**, *130*, 3959-3966.

(21) Loquet, A.; Bardiaux, B.; Gardiennet, C.; Blanchet, C.; Baldus, M.; Nilges, M.; Malliavin, T.; Bockmann, A. *J Am Chem Soc.* **2008**, *130*, 3579-3589.

(22) Korukottu, J.; Schneider, R.; Vijayan, V.; Lange, A.; Pongs, O.; Beckers, S.; Baldus, M.; Zweckstetter, M. *Plos ONE* **2008**, *3*, e2359.

(23) Castellani, F.; van Rossum, B. J.; Diehl, A.; Rehbein, K.; Oschkinat, H. *Biochemistry* **2003**, *42*, 11476-11483.

(24) Zech, S. G.; Wand, A. J.; McDermott, A. E. *J Am Chem Soc.* **2005**, *127*, 8618-8626.

(25) Petkova, A.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc.Natl.Acad.Sci.U.S.A* **2002**, *99*, 16742-16747.

(26) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. *Proc.Natl.Acad.Sci.U.S.A* **2005**, *102*, 15871-15876.

(27) Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G. *Proc.Natl.Acad.Sci.U.S.A* **2004**, *101*, 711-716.

(28) Iwata, K.; Fujiwara, T.; Matsuki, Y.; Akutsu, H.; Takahashi, S.; Naiki, H.; Goto, Y. *Proc.Natl.Acad.Sci.U.S.A* **2006**, *103*, 18119-18124.

(29) Ferguson, N.; Becker, J.; Tidow, H.; Tremmel, S.; Sharpe, T. D.; Krause, G.; Flinders, J.; Petrovich, M.; Berriman, J.; Oschkinat, H.; Fersht, A. R. *Proc.Natl.Acad.Sci.U.S.A* **2006**, *103*, 16248-16253.

(30) Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. *Science* **2008**, *319*, 1523-1526.

(31) Luca, S.; Heise, H.; Baldus, M. *Acc.Chem.Res.* **2003**, *36*, 858-865.

(32) Lange, A.; Giller, K.; Hornig, S.; Martin-Eauclaire, M. F.; Pongs, O.; Becker, S.; Baldus, M. *Nature* **2006**, *440*, 959-962.

(33) Ader, C.; Schneider, R.; Hornig, S.; Velisetty, P.; Wilson, E. M.; Lange, A.; Giller, K.; Ohmert, I.; Martin-Eauclaire, M. F.; Trauner, D.; Becker, S.; Pongs, O.; Baldus, M. *Nature Structural Biology* **2008**, *15*, 605-612.

(34) Etzkorn, M.; Kneuper, H.; Dunnwald, P.; Vijayan, V.; Kramer, J.; Griesinger, C.; Becker, S.; Unden, G.; Baldus, M. *Nature Structural Biology* **2008**, *15*, 1031-1039.

(35) Li, Y.; Berthold, D. A.; Frericks, H. L.; Gennis, R. B.; Rienstra, C. M. *ChemBioChem* **2007**, *8*, 434-442.

(36) Hiller, M.; Higman, V. A.; Jehle, S.; van Rossum, B. J.; Kuhlbrandt, W.; Oschkinat, H. *J Am Chem Soc.* **2008**, *130*, 408-409.

(37) Pauli, J.; van Rossum, B.; Forster, H.; de Groot, H. J.; Oschkinat, H. *J.Magn.Reson.* **2000**, *143*, 411-416.

(38) Martin, R. W.; Zilm, K. W. *J Magn Reson* **2003**, *165*, 162-174.

(39) Jehle, S., van Rossum, B., Stout, J. R., Noguchi, S. M., Falber, K., Rehbein, K., Oschkinat, H., Klevit, R. E., and Rajagopal, P. alphaB-Crystallin: A Hybrid Solid/Solution-State NMR Investigation Reveals Structural Aspects of Heterogeneous Oligomer. J Mol Biol . 2008.

Ref Type: In Press

(40) Lopez, J. J.; Shukla, A. K.; Reinhart, C.; Schwalbe, H.; Michel, H.; Glaubitz, C. *Angew.Chem Int.Ed Engl.* **2008**, *47*, 1668-1671.

(41) Luca, S.; Heise, H.; Baldus, M. *Acc Chem Res* **2003**, *36*, 858-865.

(42) Hiller, M.; Krabben, L.; Vinothkumar, K. R.; Castellani, F.; van Rossum, B.; Kuhlbrandt, W.; Oschkinat, H. *ChemBioChem* **2005**, *6*, 1679-1684.

(43) Kijac, A. Z.; Li, Y.; Sligar, S. G.; Rienstra, C. M. *Biochemistry* **2007**, *46*, 13696-13703.

(44) Etzkorn, M.; Martell, S.; Andronesi, O. C.; Seidel, K.; Engelhard,

M.; Baldus, M. *Angew.Chem Int.Ed Engl.* **2007**, *46*, 459-462.
(45) Ventura, S.; Zurdo, J.; Narayanan, S.; Parreno, M.; Mangues, R.; Reif, B.; Chiti, F.; Giannoni, E.; Dobson, C. M.; Aviles, F. X.; Serrano, L. *Proc.Natl.Acad.Sci.U.S.A* **2004**, *101*, 7258-7263.
(46) Jaroniec, C. P.; MacPhee, C. E.; Astrof, N. S.; Dobson, C. M.; Griffin, R. G. *Proc.Natl.Acad.Sci.U.S.A* **2002**, *99*, 16748-16753.
(47) Siemer, A. B.; Ritter, C.; Ernst, M.; Riek, R.; Meier, B. H. *Angew. Chem.Int.Ed* **2005**, *44*, 2441-2444.
(48) Carulla, N.; Caddy, G.; Hall, D. R.; Zurdo, J.; Gairi, M.; Feliz, M.; Giralt, E.; Robinson, C.; Dobson, C. M. *Nature* **2005**, *436*, 554-558.
(49) Narayanan, S.; Reif, B. *Biochemistry* **2005**, *44*, 1444-1452.
(50) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc.Natl.Acad. Sci.U.S.A* **1997**, *94*, 12366-12371.
(51) Chevelkov, V.; Reif, B. *Concepts Magn.Reson.* **2008**, *32A*, 143-156.
(52) Chevelkov, V.; Faelber, K.; Schrey, A.; Rehbein, K.; Diehl, A.; Reif, B. *J Am Chem Soc.* **2007**, *129*, 10195-10200.
(53) Giraud, N.; Blackledge, M.; Goldman, M.; Bockmann, A.; Lesage, A.; Penin, F.; Emsley, L. *J Am Chem Soc.* **2005**, *127*, 18190-18201.
(54) Chevelkov, V.; Diehl, A.; Reif, B. *J Chem Phys* **2008**, *128*, 052316.
(55) Chevelkov, V.; Diehl, A.; Reif, B. *Magn.Reson.Chem.* **2007**, *45*, S156-S160.
(56) Lorieau, J. L.; McDermott, A. E. *J Am Chem Soc.* **2006**, *128*, 11505-11512.
(57) Lorieau, J. L.; Day, L. A.; McDermott, A. E. *Proc.Natl.Acad. Sci.U.S.A* **2008**, *105*, 10366-10371.
(58) Hologne, M.; Faelber, K.; Diehl, A.; Reif, B. *J Am Chem Soc.* **2005**, *127*, 11208-11209.
(59) Hologne, M.; Chevelkov, V.; Reif, B. *Prog.NMR Spectrosc.* **2006**, *48*, 211-232.
(60) Reif, B.; Xue, Y.; Agarwal, V.; Pavlova, M. S.; Hologne, M.; Diehl, A.; Ryabov, Y. E.; Skynnikov, N. R. *J Am Chem Soc.* **2006**, *128*, 12354-12355.
(61) Agarwal, V.; Xue, Y.; Skrynnikov, N. R.; Reif, B. *J Am Chem Soc.* **2008**, *130*, 16611-16621.
(62) Chevelkov, V.; Zhuravleva, A. V.; Xue, Y.; Reif, B.; Skrynnikov, N. R. *J Am Chem Soc.* **2007**, *129*, 12594-12595.
(63) Lesage, A.; Bockmann, A. *J Am Chem Soc.* **2003**, *125*, 13336-13337.
(64) Chevelkov, V.; Faelber, K.; Diehl, A.; Heinemann, U.; Oschkinat, H.; Reif, B. *J Biomol NMR* **2005**, *31*, 295-310.
(65) Lesage, A.; Gardiennet, C.; Loquet, A.; Verel, R.; Pintacuda, G.; Emsley, L.; Meier, B. H.; Bockmann, A. *Angew.Chem.Int.Ed* **2008**, *47*, 5851-5854.
(66) van der Wel, P. C.; Hu, K.-N.; Lewandowski, J. R.; Griffin, R. G. *J Am Chem Soc.* **2006**, *128*, 10840-10846.
(67) van der Wel, P. C.; Lewandowski, J. R.; Griffin, R. G. *J Am Chem Soc.* **2007**, *129*, 5117-5130.
(68) Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J. *Proc.Natl.Acad.Sci.U.S.A* **2008**, *105*, 883-888.
(69) Hu, K.-N.; Yu, H.-H.; Swager, T. M.; Griffin, R. G. *J Am Chem Soc.* **2004**, *126*, 10844-10845.
(70) Song, C.; Hu, K.-N.; Joo, C.-G.; Swager, T. M.; Griffin, R. G. *J Am Chem Soc.* **2006**, *128*, 11385-11390.
(71) Chen, L.; Kaiser, J. M.; Polenova, T.; Yang, J.; Rienstra, C. M.; Mueller, L. J. *J Am Chem Soc.* **2007**, *129*, 10650-10651.
(72) Zhou, D. H.; Shah, G.; Cormos, M.; Mullen, C.; Sandoz, D.; Rienstra, C. M. *J Am Chem Soc.* **2007**, *129*, 11791-11801.
(73) Zhou, D. H.; Rienstra, C. M. *J Magn Reson* **2008**, *192*, 167-172.
(74) Chevelkov, V.; Rehbein, K.; Diehl, A.; Reif, B. *Angew.Chem Int. Ed Engl.* **2006**, *45*, 3878-3881.
(75) Bertini, I.; Luchinat, C.; Parigi, G. *Progr.NMR Spectrosc.* **2002**, *40*, 249-273.
(76) Bertini, I.; Luchinat, C.; Piccioli, M. *Methods Enzymol.* **2001**, *339*, 314-340.
(77) Balayssac, S.; Bertini, I.; Lelli, M.; Luchinat, C.; Maletta, M.; Yeo, K. J. *J.Am.Chem.Soc.* **2007**, *129*, 2218-2219.
(78) Balayssac, S.; Bertini, I.; Falber, K.; Fragai, M.; Jehle, S.; Lelli, M.; Luchinat, C.; Oschkinat, H.; Yeo, K. J. *ChemBioChem* **2007**, *8*, 486-489.
(79) Balayssac, S.; Bertini, I.; Bhaumik, A.; Lelli, M.; Luchinat, C. *Proc.Natl.Acad.Sci.USA* **2008**, *105*, 17284-17289.
(80) Kervern, G.; Steuernagel, S.; Engelke, F.; Pintacuda, G.; Emsley, L. *J.Am.Chem.Soc.* **2007**, *129*, 14118-14119.
(81) Lin, Q.; Degen, C. L.; Tomaselli, M.; Hunkeler, A.; Meier, U.; Meier, B. H. *Phys.Rev.Lett.* **2006**, *96*, 137604.

Chapter 6

Recent advancements in solution NMR: fast methods and heteronuclear direct detection

Authors: Isabella Felli[a], Bernhard Brutscher[b]
[a] Magnetic Resonance Center, University of Florence, Italy
[b] Institut de Biologie Structurale, Grenoble, France

Introduction

Nuclear Magnetic Resonance Spectroscopy (NMR) and Imaging (MRI), since the first observations by Bloch and Purcell in 1944,[1-3] have found an increasingly broad range of applications in fields of research as diverse as physical and material sciences, chemistry, biology, and medicine. Because it merely manipulates nuclear spins with very weak electromagnetc fields, NMR is virtually the only technique that provides atomic-level information without disturbing the chemical properties of the molecules and materials under investigation. This enormous versatility has been possible because of the development of a large panoply of NMR tools through the years. Among the major achievements one should cite Fourier-transform NMR that had a dramatic effect on the experimental sensitivity of NMR[4], and the introduction of multidimensional NMR spectroscopy by Jeener[5] and Ernst[6] in the early 70s. This has laid the basis for complete assignments and high-resolution investigations of biological macromolecules with the first protein structure presented in 1985[7]. Since then, NMR spectroscopy has been established as the principal technique with atomic resolution for structural and dynamic investigations of biological macromolecules and their complexes in solution.

In the context of mechanistic systems biology, NMR spectroscopy faces a number of new challenges, such as the investigation of the structure and dynamics of biological molecules of increasing size and complexity, the characterization of protein-protein and protein-nucleic-acid complexes (see Chapter 3), as well as the study of kinetic features of biochemical processes in the cell. This requires further technical and methodological improvements in terms of experimental sensitivity, spectral and temporal resolution, and the development of reliable and standardized methods for automated data analysis. One may expect that the availability of higher magnetic fields (see Chapter 9), improved probe technology, and new hyperpolarization instrumentation such as DNP (see Chapter 7) will lead to further substantial improvements in sensitivity in the near future. In addition, recent progress in magic-angle-spinning solid-state NMR makes large and highly immobilized molecular systems, such as membrane proteins or amyloid fibrils, amenable to NMR investigations (see Chapter 5). New advanced NMR pulse sequences and acquisition schemes are required that make optimal use of the improved instrumental performance, and are best adapted to the scientific problems in mechanistic systems biology. Last but not least, there are also an increasing number of computational tools available for automated data processing, data analysis, sequential resonance assignment, and structure determination.

In view of the extremely wide variety of experiments that are available today for biomolecular NMR studies, it may seem that there is little space left for new developments. However, recent hardware improvements, and the challenge of studying more complex biomolecular systems have triggered the development of many new exciting NMR methods. Here we focus on two examples, chosen among a wide variety of methods developed recently: (i) the reduction of experimental time requirements for multidimensional NMR by the development of new innovative data acquisition schemes, and (ii) the direct detection of low-γ heteronuclei. These examples allow us to highlight recent progress and anticipate future developments in biomolecular NMR. Both research topics benefit enormously from the improved sensitivity brought about by high magnetic field strengths (currently up to 22 T), cryogenically cooled detection coils and preamplifiers, and improved digital electronics of modern NMR instruments.

Selected topics

The power of NMR spectroscopy lies in its ability to obtain in-

formation about structure and dynamics of individual atoms in large macromolecules. In order to deal with overlap problems, multidimensional NMR techniques have been introduced that spread and correlate the signals of individual nuclear spins along different frequency dimensions. A major drawback of multidimensional (nD) NMR is the long experimental time required to record hundreds or thousands of scans (repetitions of the basic NMR pulse sequence) required for the sampling of the indirect time domain grid. While acquisition times for 1D and 2D spectra are on the order of seconds and minutes, respectively, for 3D NMR typical experimental times can be as long as several days. For four- and higher-dimensional experiments that may be of interest to the study of large molecular systems, as well as partially or completely unstructured proteins, the acquisition times become unreasonably long, unless one is willing to compromise on the spectral resolution in the indirectly detected dimensions. As a large number of different spectra must be recorded during the NMR investigation of a macromolecule such as a protein, this imposes a high time stability on the protein samples. Long experimental times represent a major limitation for high-throughput protein studies in the context of structural genomics and proteomics. By contrast, reduced acquisition times allow real-time investigations of kinetic processes, such as biochemical reactions, protein/RNA folding, and molecular assembly by multidimensional NMR. For all these reasons, there is an enormous interest in the development of new acquisition schemes that speed up the acquisition of multidimensional NMR data.

The impact of heteronuclear studies on biomolecular NMR is evident from the widespread use of ^{13}C and ^{15}N isotope enrichment, and from the large variety of heteronuclear correlation experiments that have been designed for many different purposes. Heteronuclei can be detected with very high sensitivity through so-called inverse detection methods, based on proton direct detection [8;9]. The growing interest in direct detection of heteronuclei, initially proposed by Markley and coworkers [10], stems from the intrinsically different properties of heteronuclear spins compared to protons (different gyromagnetic ratio, chemical shift ranges, …) that can provide alternative solutions when protons find limitations, such as for example, in the case of paramagnetic, unfolded or very large proteins. In this frame, a set of 'protonless' NMR experiments based on ^{13}C direct detection was recently developed [11] to perform complete sequence-specific resonance assignment of heteronuclei, and to determine NMR observables providing structural and dynamic information. This was paralleled by a tremendous improvement in instrumentation that contributed in recent years to an increase of the ^{13}C sensitivity by about one order of magnitude [12] equivalent to a gain of almost two orders of magnitude in experimental NMR time. Therefore, experiments that could only be imagined in the past are becoming possible today, bringing ^{13}C direct detection in the sensitivity range suitable for biomolecular NMR applications.

In this document we will give a short overview of the strategies used, and achievements made recently in these fields of current research.

Fast multidimensional NMR spectroscopy

Strategies

Multidimensional (nD) NMR experiments are crucial for the study of biomolecular structure and dynamics as they provide the required resolution to extract spectral parameters for individual nuclear sites in the molecule. While in 1D NMR the time evolution of nuclear spin magnetization is detected directly via the electric current induced in a receiver coil, to monitor the spin evolution in a so-called indirect time domain, it is necessary to repeat the experiment by stepwise incrementation of a delay in the pulse sequence. As a consequence of this time incrementation procedure, the experimental time required for the acquisition of an nD NMR spectrum increases by roughly 2 orders of magnitude per additional dimension. Therefore, even if the inherent sensitivity is sufficient, complete sampling of the indirect time domain grid imposes *lower limits* on the experimental times: several minutes for 2D, several hours for 3D, several days for 4D, and several months for 5D. New acquisition schemes are therefore required that allow recording data more rapidly, i.e., not slower than required to obtain a sufficient signal-to-noise ratio. In order to speed up multidimensional NMR data acquisition, the sampling problem can be resolved either by limiting the number of data points (*sparse or non-uniform sampling techniques*), or by reducing the duration of each repetition of the experiment (*fast pulsing techniques*). In recent years, motivated by the increased sensitivity of modern NMR spectrometers and the apparent drawbacks of long acquisition times for biomolecular NMR studies, many research groups around the world have contributed to the development of new spectroscopic and computational tools for fast NMR spectroscopy (figure 1). Many of the ideas have been borrowed from imaging techniques such as echo planar MRI where short acquisition times have always been a crucial issue, notably for clinical applications. The principal ideas and concepts will be shortly reviewed in the following sections, while the reader is referred to the original literature for specific examples.

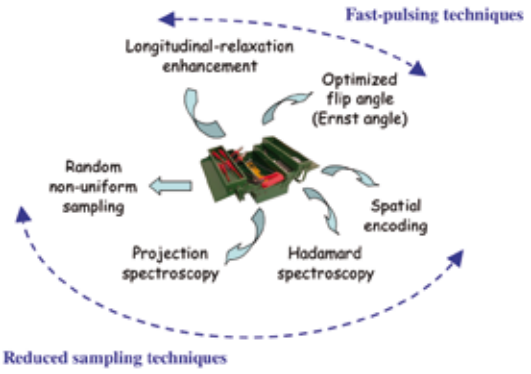


Figure 1: The toolbox of the "fast" NMR spectroscopist.

To simplify the discussion we will focus in the following on 3D experiments that require sampling of a two-dimensional grid formed by the time variables t_1 and t_2 (figure 2). For conventional linear sampling in time domain, the spacing of two subsequent time points is determined by the inverse of the chemical shift range (spectral width) of the frequency-encoded nuclei, and the number of sampled points is given by the desired spectral resolution. The same concepts, described below for the case of 3D NMR also apply to higher dimensional NMR spectroscopy.

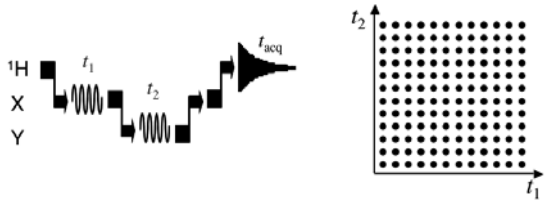


Figure 2: Schematic drawing of a three-dimensional (3D) H-X-Y correlation experiment (left), and corresponding regular time-domain sampling grid (right). Each point on the grid corresponds to one repetition of the basic pulse sequence.

Sparse non-uniform random data sampling. The simplest approach for reducing the total acquisition time of an NMR experiment is to record only a reduced number of FIDs by randomly selecting (t_1, t_2) pairs from the sampling grid (figure 3). Typically, only about 10-30% of the data points are recorded. Non-uniformly sampled data sets can no longer be processed by a regular fast Fourier transformation, but other processing methods are required to reconstruct the multidimensional NMR spectrum. Examples of possible spectral reconstruction methods proposed for non-linear data sampling are maximum entropy reconstruction (MER) [13;14] the filter diagonalisation method (FDM) [15;16], multi-dimensional decomposition (MDD) [17;18], and numerical Fourier integration (NFI) [19;20]. In special cases a statistical covariance-based analysis [21;22] may also be used for spectral reconstruction. Such methods provide increased spectral resolution with respect to standard experiments based on linear data sampling, recorded in the same time. All these methods benefit from advanced algorithms and computer technology that drastically increase the speed of such calculations. A drawback of these non-linear processing methods is that most of them rely on some kind of fitting process that is not always completely reliable, and requires some parameter adjustment either from prior knowledge about the system, or using a trial-and error approach.

Radial data sampling (projection NMR, GFT NMR). Another interesting approach to overcome the sampling problem consists in the joint incrementation of several indirect evolution times (e.g., t_1 and t_2 for the 3D case of figure 2). This translates into radial sampling of the indirect time-domain sampling grid (figure 3). Radially sampled NMR data can be processed, after appropriate recasting of the data, using regular fast Fourier transformation. The resulting spectra of "reduced dimensionality" are conceptually identical to 2D projections of the full 3D spectrum [23-26] The projection angle depends on the relative scaling of the time increments chosen for the t_1 and t_2 evolutions. Generally several 2D spectra recorded with different projection angles are needed to retrieve the full 3D correlation information using either back-projection techniques [27;28] or algebraic reconstruction methods [24]. Projection NMR yields an impressive reduction in experimental time by several orders of magnitude, in particular for higher dimensions ($n \geq 4$). Because the same number of peaks is detected in a spectral space of reduced dimensionality, projection NMR may appear especially attractive when only a limited number of correlation peaks need to be characterized, i.e., when there is a great deal of of "empty" spectral space. However, this limitation of projection NMR may be overcome when examining several projections simultaneously [29]. The presence of many peaks, forming sets with identical line shapes often allows resolving ambiguities caused by overlap in crowded NMR spectra, e.g., such as NOESY spectra. So far this technique has been mainly applied for recording series of triple-resonance correlation spectra required for resonance assignments of small to medium-sized proteins. Recently, projection NMR spectroscopy has also become popular for resonance assignment of intrinsically unfolded proteins based on projected high-dimensional (5D-7D) correlation experiments [30;31].

Optimized undersampling (spectral aliasing). When the time increments between successive data points along an indirect dimension are increased, the number of data sampling points and thus the data acquisition time are reduced accordingly. Undersampling of the time domain grid (figure 3) translates into reduced apparent spectral widths. If the spectral width is smaller than the chemical shift range of the observed nuclear spins, this results in spectral folding (aliasing). Intensive spectral aliasing is impractical, as it generally results in accidental peak overlap, and thus in a loss of spectral resolution. However, if the chemical shift of one or several nuclei are already known from previously recorded NMR data, this information can be exploited to minimize spectral widths without creating additional peak overlap [32;33]. It has been demonstrated that spectral compression factors (equivalent to the final time savings) of up to a factor of 4 can be obtained by applying optimized undersampling to the ^{15}N dimension of ^{1}H-^{15}N correlation experiments of proteins. For compression of 2D spectral space, e.g. ^{15}N-^{13}CO correlation spectra, this factor can reach values of 50 and higher. The simplicity of this method makes it attractive for routine applications provided prior chemical shift information is available.

Discrete frequency-domain sampling (Hadamard spectroscopy). Another way of exploiting existing information about chemical shifts is to replace the time-domain frequency encoding by multiple-selective excitation sequences using appropriately shaped radio-frequency pulses. Let us take the practical example of 16 different resonance frequencies that shall be edited in a multidimensional NMR experiment. An array of selective rf pulses simultaneously applied to the 16 resonances of interest, created based on the knowledge of the chemical shifts, replaces the usual non-selective rf pulse for spin excitation. The spins resonating at one of the 16 selected frequencies can be manipulated independently, in order to realize a 180° phase shift (sign change) of the corresponding line in the NMR spectrum. In order to separate the different frequencies in the final spectrum a Hadamard encoding scheme is used, instead of the usual chemical shift evolution delay t_1. In Hadamard encoding, the NMR signals from different sites in the molecule are sign modulated according to a Hadamard matrix of size $n \times n$, with n being equal or higher than the number of resonance frequencies that have to be distinguished by setting up the array of rf pulses (16 in the present case). The experiment is then repeated n (16) times, while changing the sign modulation scheme according to a particular row of the Hadamard matrix. This sign encoding ensures that each frequency contributes fully to the detected NMR signal in each of the experimental repetitions, thus providing the same multiplex advantage as standard time domain NMR spectroscopy [34;35]. The Hadamard matrix chosen for frequency encoding is then also used at the processing stage to disentangle the individual resonance frequencies. Hadamard spectroscopy presents a speed advantage whenever n is smaller than the number of data points to be sampled along an indirect time-period using standard time-domain frequency encoding followed by Fourier transformation. In the context of protein NMR spectroscopy, Hadamard encoding allows focusing on a small number of nuclear sites, e.g., in the active sites of proteins that are of particular interest [34;36]. In this context it is noteworthy that the resolution of individual peaks – unlike the situation in FT NMR – does not depend on the number of increments that must be acquired, but only on the excitation profile of the encoding pulses, which allows one to focus on a small set of cross-peaks using only a limited number of experiments. Hadamard spectroscopy has also been successfully applied to amino-acid-type and base-type editing in NMR spectra of proteins and nucleic acids [37;38]

Spatial frequency encoding (ultrafast NMR). The ultimate solution to the NMR sampling problem is obtained when encoding the chemical shift information along a spatial coordinate or k-axis (figure 3), instead of an indirect time domain, and replacing the standard signal detection at the end of the pulse sequence by an imaging-type readout sequence, e.g., as in echo-planar-imaging (EPI). This so-called "ultrafast" NMR technique in principle allows on to record any multidimensional experiment within a single scan, thus yielding a dramatic reduction in acquisition time [39;40]. However, ultrafast NMR currently still suffers from a relatively low sensitivity, which is approximately proportional to $\sqrt{N}$, where N is the number of indirect time points that one would sample in a conventional experiment. In addition to this drawback, further instrumental and methodological improvements are needed to make ultrafast NMR a standard and widely used tool for protein applications.

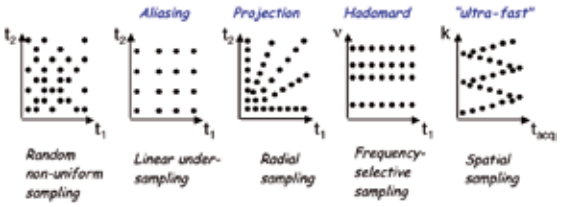


Figure 3: Alternative sampling schemes used for different fast NMR methods.

Fast-pulsing NMR techniques present an alternative way of reducing acquisition times. The main idea is to shorten the time delay between successive scans (recycle delay) to achieve higher repetition rates and thus collect the same number of scans in less time. Of course, the number of data points to be recorded can also be reduced as discussed above, which makes fast-pulsing techniques fully compatible with sparse sampling approaches. A recycle delay is required to allow relaxation of the excited spins (usually ^{1}H) towards their thermodynamic equilibrium, and to build up sufficient ^{1}H polarization to be used for the next scan. In order to keep the experimental sensitivity high enough while using fast repetition rates, some spectroscopic tricks are required. A first approach has become known as longitudinal relaxation enhancement (LRE) [41] LRE exploits the fact that the efficiency of ^{1}H spin-lattice relaxation is increased if nearby ^{1}H are unperturbed by the pulse sequence, so that they can take up some of the energy put into the system via dipole-dipole interactions (nOe effect), or via hydrogen exchange. In practice, LRE is realized by selectively manipulating a subset of the ^{1}H spins of interest, e.g., those of amide protons, throughout the pulse sequence, thus ensuring that the spin states of all other protons that are not directly involved in the coherence transfer pathways of a particular experiment remain unperturbed. This yields reductions in effective longitudinal ^{1}H relaxation times from a few seconds to a few hundred milliseconds. In some circumstances, e.g., in HMQC experiments, the sensitivity of fast-pulsing experiments can be even further enhanced by adjusting the excitation flip angle to the so-called Ernst angle [42;43]. Both effects have been combined in the SOFAST experiment [44;45] that allows one to record 2D ^{1}H-^{15}N or ^{1}H-^{13}C correlation spectra of proteins in only a few seconds, thus opening new avenues for real-time investigations of protein kinetics at atomic resolution.

Time-shared techniques. Although they do not fit into the general frame of sparse sampling or fast pulsing techniques discussed so far, time-sharing methods provide additional tools for reducing NMR data acquisition times by performing different experiments at the same time. Time-sharing has been proposed for simultaneous ^{15}N and ^{13}C editing in HMQC/HSQC [46] NOESY-HSQC-type experiments [47], for simultaneous H^{α} and H^{N} detection in triple-resonance experiments [48], and simultaneous methyl and amide TROSY experiments of large deuterated proteins [49]. The use of multiple receivers also enables the simultaneous detection of ^{1}H ,^{15}N and ^{13}C in heteronuclear correlation experiments, as recently demonstrated by Kupce et al. [50;51].

Interest for NMR studies of proteins

In the following, we shall briefly outline a few selected areas where fast multidimensional NMR data acquisition schemes can be used to increase the efficiency, or extend the applicability range of NMR studies of protein structure and dynamics.

Quality control of protein samples. NMR is a powerful tool for fast screening of protein samples to check whether they are suitabile for a detailed structural investigation by either NMR spectroscopy or X-ray crystallography. In particular, information about the presence of stable structures along the polypeptide sequence is obtained from the chemical shift dispersion and distribution observed in the NMR spectra. Often when dealing with complex molecular systems comprising independently folding subunits, or in the context of structural proteomics projects, a large number of protein constructs need to be prepared and tested. Therefore, sensitive multidimensional NMR methods are required that allow fast identification of promising protein constructs and sample conditions for further detailed structural and dynamic investigation. The ^{1}H-^{15}N correlation spectrum provides a fingerprint of the protein that is particular indicative of the protein's purity, and its structural and dynamic properties. The SOFAST pulse scheme [44;45], based on fast pulsing techniques, provides the most efficient way for recording such a fingerprint spectrum in the shortest amount of time. Furthermore a variant of SOFAST, the so-called HET-SOFAST NMR experiment [52], allows quantification of local structure along the polypeptide chain. In other words, HET-SOFAST NMR measures the structural heterogeneity, i.e., the fractions of structured and unstructured regions, within a protein construct. The high sensitivity of the experiment allows data recording in short experimental time even for protein samples with low concentration (µM range).

Small proteins – reduced experimental time. The majority of NMR studies reported in the literature concern relatively small (< 200 residues) globular proteins or small fragments of larger proteins. Because of their (relatively) high chemical shift dispersion and favourable spin relaxation properties, such proteins yield often high-quality NMR spectra. Therefore, a 3D spectral space is generally sufficient to resolve most of the observed correlation peaks. In addition, owing to the use of high magnetic fields and cryogenically cooled probes, the sensitivity is sufficient, making fast NMR approaches very attractive. Fast acquisition methods are particular useful for triple-reso-

nance H-N-C correlation experiments required for backbone and side-chain resonance assignments. Such spectra are characterized by a relatively low peak density. Recently it has been demonstrated that for small proteins the complete backbone resonance assignments can be obtained from NMR data recorded in only a few hours, using various optimized data acquisition schemes and analysis tools [31;53;54]. Furthermore, taking advantage of fast multidimensional NMR data acquisition and of advanced structure calculation protocols, high-resolution structures of small proteins can be obtained from a limited set of NMR data within a few days. Short overall experimental times not only save expensive NMR instrument time; they also make protein samples of limited stability that previously would have been discarded amenable to NMR investigation. Another advantage of fast acquisition methods consists in the possibility of recording a larger number of experimental data in order to increase the robustness of (automated) data analysis.

Large and intrinsically unstructured proteins – increased dimensionality. The situation is quite different for proteins of increasing molecular size, or proteins possessing large unstructured regions. While experimental sensitivity is usually not a limiting factor for unstructured proteins, or can be enhanced by appropriate isotope labelling and/or by relaxation-optimized pulse schemes, the major bottleneck for NMR studies on such systems is a the extensive NMR signal overlap (high frequency degeneracy.) Experiments of high dimensionality ($n \geq 4D$) can provide a solution to this problem. It has been demonstrated that highly resolved 4D spectra, recorded using either projection NMR or non-uniform sampling techniques combined with spectral reconstruction, are useful for the study of large molecular systems [17;55;56], while even higher dimensional experiments (6D, 7D) have been proposed for sequential resonance assignment of unstructured proteins [30;57]. The long transverse spin relaxation times observed in these unfolded systems allow multiple coherence transfer steps within a single correlation experiment. Experimental times are kept short by projecting these hyper-dimensional spectra onto 2D planes. The coordinates of the 6D or 7D correlation peaks are then obtained from a set of projection spectra using advanced spectral reconstruction methods.

Real-time studies of protein/RNA kinetics. NMR spectroscopy allows studies with atomic resolution of molecular dynamics over a wide range of time scales (figure 4). While steady-state NMR methods are well suited to characterize equilibrium dynamics occurring on a sub-seconds time scale, kinetic molecular processes taking place on longer time scales can be followed by real-time multidimensional NMR methods [58;59] where spectral changes are monitored during a conformational transition, e.g., during protein folding. In the past, the long acquisition times associated with 2D NMR have limited the application of real-time 2D NMR to slow kinetic processes with characteristic time constants of minutes to hours. The introduction of fast 2D data acquisition schemes such as the SOFAST [44;45] and ultraSOFAST [60] experiments has extended the time window accessible to real-time 2D NMR to the range of seconds to minutes. SOFAST real-time 2D NMR provides a powerful new tool for investigations of protein folding at atomic resolution, conformational transitions, ligand binding, and hydrogen-deuterium exchange reactions that can be completed within a few seconds [61]

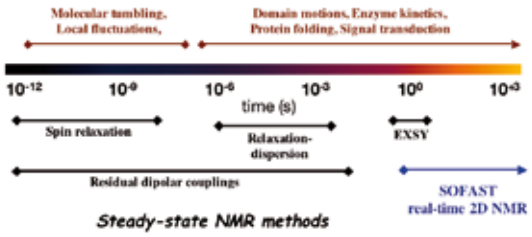


Figure 4: Time scales of various molecular dynamic processes, and multidimensional NMR methods available to characterize these motional modes at atomic resolution.

Perspectives

As outlined above, numerous tools are nowadays available (see Figure 1) for speeding up multidimensional NMR data acquisition. The achievable time saving for a given experiment depends on the technique, on the inherent sensitivity, and on the dimensionality of the experiment. Generally there is no hierarchy among the proposed fast NMR methods, but they should be considered as complementary with individual tools performing best for different applications.

With the availability of all these new spectroscopic tools, there is now an urgent need for making them widely accessible to the scientific community. Integrated software packages are required that help the experimentalist to set up NMR experiments in a time-optimized way, including all the processing tools required for data analysis. This will then allow target-driven NMR data acquisition whenever NMR data are recorded in small »pieces« that are analyzed while the next piece of data

is recorded, and data acquisition is stopped once all required information has been obtained.

The importance of fast NMR data acquisition schemes will further increase with the availability of higher magnetic field magnets, improved sensitivity obtained through new NMR probe technology or polarization enhancement techniques. All these developments combined will make NMR spectroscopy applicable to molecular systems of increasing complexity, addressing questions of biological relevance.

Direct detection of heteronuclei

Strategies

The motivation. The main motivation for the development of ^{13}C direct detection NMR experiments to study biological macromolecules comes from the intrinsically different properties of ^{13}C, compared to ^{1}H, in terms of chemical structure and magnetic properties (e.g., paramagnetic interactions, chemical shifts, relaxation rates). The large gyromagnetic ratio that is responsible for the high sensitivity of protons also causes large dipole-dipole interactions that lead to rapid relaxation rates. These in turn may broaden lines beyond detection with increasing molecular mass, or in the presence of a paramagnetic center. In this context, heteronuclei, which are characterized by smaller gyromagnetic ratios, may provide additional spectroscopic probes with a better performance. The conformational heterogeneity characteristic of specific protein regions, often linked to molecular recognition and/or to catalysis, causes exchange processes that may also broaden signals beyond detection. Since these processes affect different nuclear spins to different extents, NMR methods based exclusively on heteronuclei may reveal unique information. The intrinsically small range of proton chemical shifts may cause severe resonance overlap, in particular in unfolded systems. By contrast heteronuclear chemical shifts, which have a much larger chemical shift dispersion even in the absence of any stable 3D structure, provide an ideal tool to characterize these systems. Proline residues, often very abundant in intrinsically unfolded proteins, have very characteristic ^{15}N shifts and become the starting point of assignment protocols based on direct detection of carbonyl carbons. Heteronuclear direct detection NMR experiments can also provide additional input for automatic assignment protocols, where the reliability of the results strongly depends on the completeness and redundancy of the input data. Other technical aspects in favour of heteronuclear direct detection include the absence, in general, of intense solvent signals that need to be suppressed and a reduced sensitivity to high ionic strength (high salt concentrations) that may cause problems for proton detection. Therefore, heteronuclear NMR experiments based on ^{13}C direct detection provide new tools complementing or replacing the standard ^{1}H-detected experiments. Some general aspects of heteronuclear direct detection experiments, as well as research areas where these methods can be applied will be described hereafter.

Homonuclear ^{13}C decoupling. The presence of large homonuclear carbon-carbon scalar couplings constitutes a first key difference between ^{1}H and ^{13}C direct detection experiments. In ^{1}H NMR the homonuclear ^{1}H-^{1}H scalar couplings are generally limited to the relatively small ^{2}J and ^{3}J, which in many multidimensional NMR experiments are not even resolved. ^{13}C direct detection experiments reveal large homonuclear one-bond couplings, that are generally well-resolved, and that yield complicated multiplets in the direct acquisition dimension causing signal overlap and reduced sensitivity. The ^{13}C nuclei of amino acid side chains have specific topologies and coupling patterns. The backbone nuclei C′ and C^{α} are the most suitable to be treated in a general way. Many solutions to the problem of homonuclear decoupling were proposed in the literature for different kinds of applications, including band-selective homodecoupling [62], spin-state selection methods [63-68], and post acquisition processing protocols [69]. When tested for ^{13}C direct detection in solution [70-72], as in solid state applications [67;68], the approaches that gave the best results were spin-state selective methods, such as IPAP or S^3E. Several variants were adapted to C′ and C^{α} direct detection allowing simplification of the spectra by removing large one-bond homonuclear splittings [11;71-73]. User-friendy protocols are available to process the data.

A suite of heteronuclear-detected NMR experiments. After solving the problem of homonuclear decoupling in the direct dimension of heteronuclear NMR experiments, a set of dedicated ^{13}C direct detection experiments was designed for the structural and dynamic characterization of biological molecules. These experiments exploit the large homonuclear and heteronuclear scalar couplings for coherence transfer, similar to the widely used ^{1}H-detected triple resonance experiments [74]. As an example, a set of $^{13}C'$ direct detection experiments designed for sequence-specific resonance assignment [75-77] is illustrated in Figure 5.

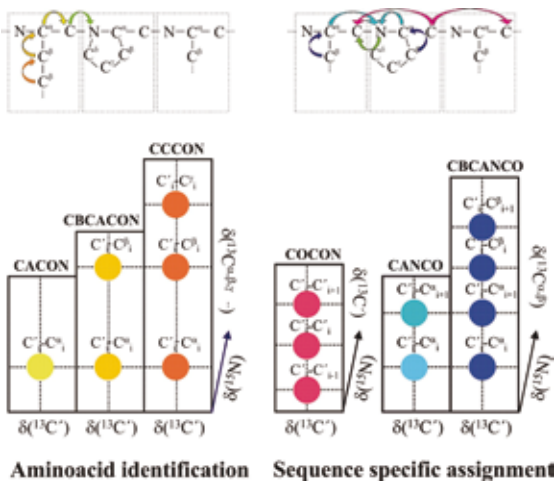


Figure 5: Set of heteronuclear 3D NMR experiments designed to obtain sequence-specific assignments in proteins, based on direct detection of carbonyl carbons, shown schematically highlighting the coherence transfer pathways as well as the kind of correlations expected in each experiment. The green arrow represents the C′-N correlation used to expand the experiments into a third dimension by including the ^{15}N chemical shifts.

These experiments can be generally applied to the study of proteins as long as the molecular size does not lead to excessive broadening of the carbonyl resonances (small/medium size proteins). When compared to the analogous methods based on amide proton detection, generally used for sequence-specific assignment of backbone signals, these experiments also give information on proline residues and are less prone to exchange processes that may broaden amide protons beyond detection. Therefore they may provide additional information for sequence-specific resonance assignment. This can be particularly useful in the process of automation, even at the expense of longer acquisition times compared to analogous amide proton detected experiments. The experiments can obviously benefit from implementation of the features discussed in the previous paragraph to reduce experimental time and speed up data acquisition, such as folding, sparse data sampling and longitudinal relaxation enhancement.

Finally, although this may sound somewhat counterintuitive, protons may also play an important role in ^{13}C direct detection experiments. As long as fast transverse ^{1}H relaxation does not constitute a limiting factor, this mixed approach can still provide valuable solutions; several experiments based on ^{13}C detection, but using ^{1}H chemical shift labelling in one of the indirect dimensions, have been proposed both for proteins [78-83] and nucleic acids [84;85]. The latter are very useful for the low proton density of nucleic acids [84;85]. Alternatively, the ^{1}H polarization can simply be used as a starting point of otherwise exclusively heteronuclear NMR experiments (only heteronuclei are frequency labelled in all dimensions), thus increasing the sensitivity of the experiments, while exploiting the favourable chemical shift dispersion of heteronuclei [86].

For larger molecules, the large CSA of carbonyl carbons leads to rapid transverse relaxation rates and thus it is advisable to focus on direct detection of aliphatic carbons that, provided isotopic enrichment with ^{2}H is available, retain relatively narrow lines also in very large systems (hundreds of kilodaltons). In these cases, if fast transverse relaxation becomes a limiting factor for experiments based on *J*-couplings, efficient magnetization transfer may still be achieved by exploiting ^{13}C-^{13}C dipolar interactions. Because homonuclear cross-relaxation rates (nOe effects) are larger for slower-tumbling molecules, nOe-based magnetization transfer between adjacent carbons within the polypeptide chain becomes very efficient for large molecules [87;88]. Therefore, ^{13}C-^{13}C NOESY spectra can be used to identify spin systems [70;88;89] when fast transverse relaxation becomes a limiting factor for standard *J*-coupling based correlation experiments.

Application areas

Unfolded systems – improved chemical shift dispersion. The higher intrinsic chemical shift dispersion of heteronuclei compared to protons constitutes a great advantage, particularly for the study of completely or partially unfolded systems, where the lack a stable 3D structure causes a dramatic reduction of the ^{1}H chemical shift dispersion and severe overlap problems [90;91]. Heteronuclear-detected NMR experiments are particularly well suited for the study of these unfolded systems [77;92]. As an example, the richer information content of experiments based on direct detection of carbonyl signals, in comparison with the analogous ones based on detection of amide protons, can be appreciated by inspection of Figure 6 that schematically shows the expected correlations of the amide nitrogens with the attached carbonyls (C'_i-N_{i+1}) or with the attached amide protons (H^N_i-N_i).

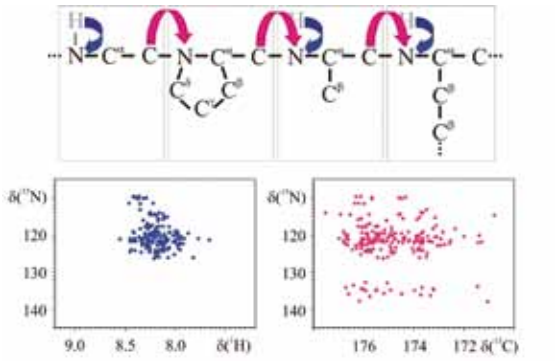


Figure 6: The graphs represent expected correlations involving backbone nitrogens with the directly bound amide proton (N^H_i-H^N) or carbonyl carbon (N^H_i-C'_{i-1}), calculated using random coil chemical shifts[111] appropriately corrected for the contributions from the primary sequence[112], taking human securin as an example [92]. For a fair comparison, the same spectral widths (in units of Hz) are shown in the two graphs (2 ppm for ^1H correspond to 8 ppm for ^{13}C).

It is evident that CON experiments also reveal correlations involving prolines, which are often very abundant in intrinsically unfolded systems. Moreover the chemical shifts characteristic of unfolded systems are generally very close to random coil values, with a narrow range for each type of aminoacid. Hence, the inter-residue (C'_i-N_{i+1}) correlations detected in the CON experiment mayimprove the dispersion of the cross peaks. Therefore all experiments based on direct detection of carbonyl reasonances (see Figure 5) provide more complete information than the analogous ones based on amide proton detection. These experiments, tested on standard samples, were employed for sequence specific assignment of several unfolded proteins [77;92].

The excellent performance of this kind of experiments for the study of unfolded systems stimulates additional improvements. In particular, since fast transverse proton relaxation does not constitute a limiting factor in unfolded systems, the ^1H polarization can still be used as a starting point to increase sensitivity, while keeping heteronuclear detection in the NMR experiments. This feature, which can be easily implemented in many of the ^{13}C direct detection NMR experiments proposed, stimulates the design of additional variants exploiting the additional chemical shift dispersion of heteronuclei in all dimensions of an nD NMR experiment, including the directly detected dimension [86]. Variants of the CON experiment can be designed for the quantification of various spectral observables, such as residual dipolar couplings (N-H^N, C^α-H^α, C^α-C', etc), auto- and cross-correlated relaxation rates, to characterize residual structures and dynamics of unfolded proteins.

Paramagnetic systems – recover lost information. Another advantage of low-γ nuclei is the reduced magnitude of the dipole-dipole interactions that affect spin relaxation. In this context, a typical example is given by a paramagnetic center that provides additional contributions to relaxation rates through electron-nucleus dipolar interactions [93;94]. The dipolar contributions depend on the internuclear electron-nucleus distance ($1/r^6$) and on the gyromagnetic ratio of the observed nucleus (γ^2). Therefore, at equal distance from the paramagnetic center, protons are much more affected by paramagnetic broadening than carbons (γ^2_H/γ^2_C). The use of ^{13}C direct detection NMR experiments allows us to observe ^{13}C signals for residues whose proton resonances are broadened beyond detection, and to obtain spectral information for residues much closer to the paramagnetic center. On the other hand, the measurement of ^1H and ^{13}C paramagnetic relaxation rate enhancements can provide complementary distance information for nuclei at different locations with respect to the paramagnetic center. These ideas have recently been exploited to study several paramagnetic molecules that contain a paramagnetic center in their native state [95-100] or after metal substitution [101-103]. As an example the case of Cu(II),Zn(II) monomeric superoxide dismutase is shown in Figure 7.

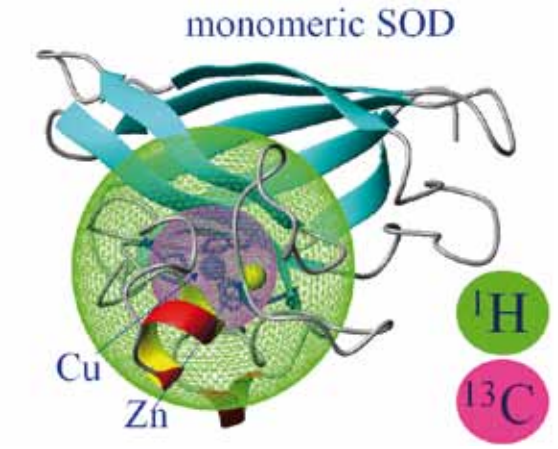


Figure 7: The green sphere (with a radius of 11 Å) shows the region of the monomeric mutant of Cu(II)Zn(II)-superoxide dismutase (PDB ID 1BA9) where the ^1H signals are broadened beyond detection by the contributions to transverse relaxation due to the presence of a paramagnetic type-II Cu(II). The use of protonless NMR experiments, based on ^{13}C direct detection, permits the identification of all residues except for the Cu(II) ligands (pink sphere) [96].

^{13}C direct detection experiments may also complement ^1H-based experiments for the study of molecules containing

paramagnetic tags to obtain additional long-range distance information. Paramagnetic tags are often employed to study intermolecular interactions by observing changes induced in the spectra of one partner by the presence of a tag attached to the other partner. Paramagnetic tags can provide information on possible long-range contacts present in the various conformers characteristic of unfolded protein states. Protonless NMR experiments can thus provide unique structural information. Finally, borrowing an idea from the early NMR literature on paramagnetic systems, it has recently been proposed to exploit the effect of heavily relaxing agents, such as the contrast agents generally used for imaging, to characterize molecular surfaces [104-106]. These substances, when added to the solution, induce paramagnetic line broadening that depends, among other factors, on the distance of a nucleus from the surface as well as on the square of the gyromagnetic ratio of the observed nucleus. The complementary use of ^{1}H and ^{13}C detected NMR experiments in samples containing a paramagnetic center provides a powerful tool to characterize protein surfaces and unfolded protein states.

Another interesting application of protonless ^{13}C direct detection NMR experiments consists in the possibility to determine "residual chemical shifts", resulting from incomplete averaging of the CSA tensors in solution. These effects can in principle also be determined when the linewidths are too large to determine residual dipolar couplings.

Large systems – reduced spin relaxation. In large, slowly-tumbling macromolecules, dipole-dipole interactions may broaden lines beyond detection. Isotopic replacement of ^{1}H with ^{2}H can be used to reduce the dipole-dipole interactions. Furthermore, the constructive use of cross-correlated relaxation effects [107] allows one to detect ^{1}H resonances for specific kinds of protons (amide [107;108], aromatic [108], methyl [109]), but this approach generally requires specifically designed isotope enrichment schemes. Therefore, ^{13}C direct detection NMR experiments can provide additional solutions for NMR spectroscopy of large molecules or molecular complexes. While carbonyls are broadened beyond detection at high magnetic fields by the large chemical shift anisotropy (CSA), the aliphatic nuclear spins are still characterized by relatively narrow lines, and can be exploited to obtain site-specific information in large macromolecular assemblies, for which little information is available from ^{1}H detection experiments. Particularly interesting in this context is the ^{13}C-^{13}C NOESY experiment [70;88;89]. The short internuclear distances between directly bound carbons and the large rotational correlation times typical of large molecules are responsible for detectable ^{13}C-^{13}C nOe effects and for very efficient spin diffusion within a network of directly bound carbon spins [70;88]. For these reasons the ^{13}C-^{13}C NOESY experiment can be used to identify the ^{13}C spin systems of aminoacids (intra-residue correlations) in large molecules with hundreds of kilodaltons [89]. This was demonstrated for a very large multimeric protein (Figure 8) [89].

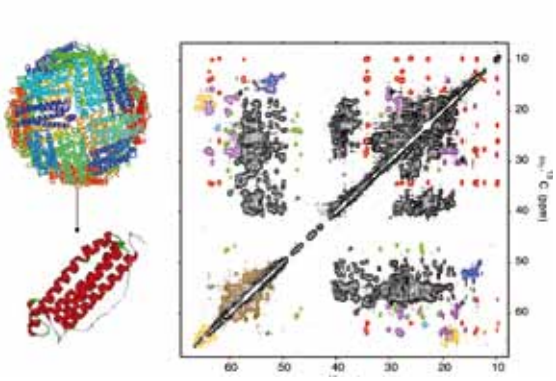


Figure 8: The 2D ^{13}C-^{13}C NOESY spectrum of a sample of 0.2 mM bullfrog ferritin uniformly labelled with ^{13}C, ^{15}N, and >90% ^{2}H [89]. Only the aliphatic region of the spectrum is shown. Cross peaks corresponding to specific types of amino acids are colour coded. The 3D structure of the protein (PDB ID: 1MFR) as well as the monomeric unit are shown on the left.

Still, the problem of sequence-specific resonance assignment of large proteins is not a trivial task, given that fast relaxation, insufficient polarization, and poor coherence transfer efficiency prevent the detection of the correlations that are essential to obtain sequence-specific assignments. For systems where the increase in molecular mass is not accompanied by an increase of the number of signals, such as, for example, for multimeric proteins or isotopically labelled subunits of larger biomolecular assemblies, one solution to the problem of sequence-specific assignment may consist in combining solution and solid state NMR data. Carbonyls and nitrogens, which in solution are broadened beyond detection due to their large CSA's combined with slow molecular tumbling, can be exploited in the solid state to obtain sequence-specific assignments. Assignments obtained in the solid state can be transferred to solution NMR data by comparison of the 2D carbon-carbon correlation maps that can be acquired in solution (^{13}C-^{13}C NOESY) and in the solid state (PDSD, DARR). Therefore ^{13}C-^{13}C NOESY experiments can become a powerful tool to access site-specific information, in solution, on systems as large as a few hundreds of kilodaltons, provided that the number of signals is reduced to a few hundred of aminoacids, using segmental labelling

approaches [110]. For example, ^{13}C-^{13}C NOESY experiments in solution can be used to follow perturbations that occur upon changing the sample conditions, much like what is often done with ^{1}H-^{15}N HSQC experiments.

Perspectives

Sets of ^{13}C-detected NMR experiments offer new tools that can be applied to study biological macromolecules, also thanks to recent improvements in the instrumentation. They can be generally used in parallel to NMR experiments based on direct proton detection, and can provide unique additional information. The sets of experiments can be further improved and expanded. For example, many of the proposed approaches to reduce experimental time, in particular those to reduce the number of experiments in the indirect dimensions, can also be implemented in heteronuclear-detected NMR experiments to take advantage of the favourable heteronuclear chemical shift dispersion. Additional experiments may be designed to determine a variety of different NMR observables required for the structural and dynamic characterization of biomolecules. The development of user-friendly experimental tools, paralleled by the impressive increase in the sensitivity of the instrumentation and of the experiments themselves is now bringing ^{13}C-detected NMR in a suitable range of sensitivity to provide important additional information about biological systems.

Summary and outlook

The development of advanced NMR methods in parallel with tremendous improvements in NMR instrumentation is continuously expanding the scope of NMR spectroscopy. Here, we have focused on two examples among the many exciting recent developments in the field of biomolecular high-resolution NMR: i) experimental schemes yielding reduced acquisition times for multi-dimensional NMR spectra, and ii) dedicated NMR experiments for direct detection of low-γ nuclei. These new methods provide new tools for the characterization of biomolecular systems. They pave the way for NMR studies at atomic resolution of molecular systems of increasing complexity and of limited lifetime, of proteins lacking a stable tertiary structure, and of fast kinetic molecular processes that would have been impossible, or at least very difficult without these new tools.

The methods described here, together with other ongoing developments in biomolecular NMR, demonstrate that NMR is quickly adapting to the new challenges that Life Sciences will be facing in the future, and which are shifting from the characterization of single biomolecules to an integrated view of interacting molecular networks observed at varying levels of biological organization. After the genome and the proteome, the *interactome* is becoming an increasingly important target of biomolecular research. This in turn means that we must develop new tools for identifying the reaction pathways in which a specific molecule is involved, either under normal conditions or when subjected to different stimuli from the environment. Many different techniques, with different time- and space-resolutions, should probably be combined to approach this ambitious goal. NMR has the unique ability of providing information on structural, dynamic, thermodynamic, and kinetic aspects of biological macromolecules, and about their interactions at atomic resolution. Thus NMR can address many fundamental aspects of Mechanistic Systems Biology. It can be foreseen that future progress in NMR instrumentation and methodology will further enhance the potential of NMR spectroscopy to make significant contributions to these ambitious goals.

Acknowledgements

We would like to thank the following people for their valuable discussions and comments:

Wolfgang Bermel[a], Ivano Bertini[b], Martin Billeter[c], Geoffrey Bodenhausen[d], Lucio Frydman[e], Christian Griesinger[f], Eriks Kupce[g], Tobias Madl[h], Roberta Pierattelli[b], Paul Schanda[i], Paola Turano[b], Gerhard Wider[j]

[a] Bruker BioSpin, Rheinstetten, Germany
[b] Magnetic Resonance Center, University of Florence, Italy
[c] Department of Chemistry, University of Gothenburg, Sweden
[d] École Normale Supérieure, Paris, France and Institute of Molecular and Biological Chemistry, Swiss Federal Institute of Technology, Lausanne, Switzerland
[e] Weizmann Institute, Rehovot, Israel
[f] Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[g] Varian, Darmstadt, Germany
[h] Institute of Chemistry, University of Graz, Austria
[i] Institut de Biologie Structurale, Grenoble, France
[j] Institute of Molecular Biology and Biophysics, ETH Zurich, Switzerland

References

(1) Bloch, F. *Phys.Rev.* **1946**, *70*, 460-474.
(2) Bloch, F.; Hansen, W. W.; Packard, M. *Physical Review* **1946**, *69*, 127.
(3) Purcell, E. M.; Torrey, H. C.; Pound, R. V. *Phys.Rev.* **1946**, *69*, 37.
(4) Ernst, R. R.; Anderson, W. A. *Rev.Sci.Instr.* **1966**, *37*, 93.
(5) Jeener, J. AMPERE INTERNATIONAL SUMMER SCHOOL . 1971. Ref Type: Abstract
(6) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J.Chem.Phys.* **1976**, *64*, 2229-2235.
(7) Williamson, M. P.; Havel, T. F.; Wüthrich, K. *J.Mol.Biol.* **1985**, *182*, 295-315.
(8) Bodenhausen, G.; Ruben, D. J. *Chem.Phys.Lett.* **1980**, *69*, 185-188.
(9) Bax, A.; Griffey, R. H.; Hawkins, B. L. *J.Magn.Reson.* **1983**, *55*, 301-315.
(10) Oh, B.-H.; Westler, W. M.; Darba, P.; Markley, J. L. *Science* **1988**, *240*, 908-911.
(11) Bermel, W.; Bertini, I.; Felli, I. C.; Piccioli, M.; Pierattelli, R. *Progr. NMR Spectrosc.* **2006**, *48*, 25-45.
(12) Kovacs, H.; Moskau, D.; Spraul, M. *Prog.NMR Spectrosc.* **2005**, *46*, 131-155.
(13) Hoch, J. C.; Stern, A. S. In *Nuclear Magnetic Resonance of Biological Macromolecules*; 2001; pp 159-178.
(14) Rovnyak, D.; Filip, C.; Itin, B.; Stern, A. S.; Wagner, G.; Griffin, R. G.; Hoch, J. C. *J.Magn.Reson.* **2003**, *161*, 43-55.
(15) Chen, J. H.; Nietlispach, D.; Shaka, A. J.; Mandelshtam, V. A. *J.Magn.Reson.* **2004**, *169*, 215-224.
(16) Hu, H.; De Angelis, A. A.; Mandelshtam, V. A.; Shaka, A. J. *J.Magn.Reson.* **2000**, *144*, 357-366.
(17) Luan, T.; Jaravine, V.; Yee, A.; Arrowsmith, C.; Orekhov, V. Y. *J.Biomol.NMR* **2005**, *33*, 1-14.
(18) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J.Biomol.NMR* **2003**, *27*, 165-173.
(19) Kazimierczuk, K.; Kozminski, W.; Zhukov, I. *J.Magn.Reson.* **2006**, *179*, 323-328.
(20) Marion, D. *J.Magn.Reson.* **2005**, *32*, 141-150.
(21) Brüschweiler, R. *J.Chem.Phys.* **2004**, *121*, 409-414.
(22) Zhang, F.; Brüschweiler, R. *J.Chem.Phys.* **2004**, *120*, 5253-5260.
(23) Brutscher, B.; Morelle, N.; Cordier, F.; Marion, D. *J.Magn Reson.B* **1995**, *109*, 238-242.
(24) Kim, S.; Szyperski, T. *J.Am.Chem.Soc.* **2003**, *125*, 1385-1393.
(25) Kupce, E.; Freeman, R. *J.Am.Chem.Soc.* **2003**, *125*, 13958-13959.
(26) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wüthrich, K. *J.Am. Chem.Soc.* **1993**, *115*, 9307-9308.
(27) Kupce, E.; Freeman, R. *J.Am.Chem.Soc.* **2004**, *126*, 6429-6440.
(28) Coggins, B. E.; Venters, R. A.; Zhou, P. *J.Am.Chem.Soc.* **2004**, *126*, 1000-1001.
(29) Malmodin, D.; Billeter, M. *J.Am.Chem.Soc.* **2005**, *127*, 13486-13487.
(30) Hiller, S.; Wasmer, C.; Wider, G.; Wüthrich, K. *J.Am.Chem.Soc.* **2007**, *129*, 10823-10828.
(31) Staykova, D. K.; Fredriksson, J.; Bermel, W.; Billeter, M. *J.Biomol. NMR* **2008**, *42*, 87-97.
(32) Jeannerat, D. *J.Magn.Reson.* **2007**, *186*, 112-122.
(33) Lescop, E.; Schanda, P.; Rasia, R.; Brutscher, B. *J.Am.Chem.Soc.* **2007**, *129*, 2756-2757.
(34) Kupce, E.; Nishida, T.; Freeman, R. *Progress in NMR Spectroscopy* **2003**, *42*, 95-122.
(35) Kupce, E.; Freeman, R. *J.Magn.Reson.* **2003**, *162*, 300-310.
(36) Bougault, C.; Feng, L.; Glushka, J.; Kupce, E.; Prestegard, J. H. *J.Biomol.NMR* **2004**, *28*, 385-390.
(37) Lescop, E.; Rasia, R.; Brutscher, B. *J.Am.Chem.Soc.* **2008**, *130*, 5014-5015.
(38) Van Melckebeke, H.; Simorre, J. P.; Brutscher, B. *J.Am.Chem.Soc.* **2004**, *126*, 9584-9591.
(39) Frydman, L.; Scherf, T.; Lupulescu, A. *Proc.Natl.Acad.Sci.USA* **2002**, *99*, 15858-15862.
(40) Shrot, Y.; Frydman, L. *J.Am.Chem.Soc.* **2003**, *125*, 11385-11396.
(41) Pervushin, K.; Vogeli, B.; Eletsky, A. *J.Am.Chem.Soc.* **2002**, *124*, 12898-12902.
(42) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in one and two dimensions*; Oxford University Press: London, 1987.
(43) Ross, A.; Salzmann, M.; Senn, H. *J.Biomol.NMR* **1997**, *10*, 389-396.
(44) Schanda, P.; Brutscher, B. *J.Am.Chem.Soc.* **2005**, *127*, 8014-8015.
(45) Schanda, P.; Kupce, E.; Brutscher, B. *J.Am.Chem.Soc.* **2005**, *33*, 199-211.
(46) R. Boelens, M. Burgering, R. H. Fogh, R. Kaptein, *J.Biomol.NMR* **1994,** *4,* 201-214, B. T. Farmer II, *J.Magn Reson.* **1991,** *93,* 635-641, M. Sattler, M. Maurer, J. Schleucher, C. Griesinger, *J.Biomol.NMR* **1995,** *5,* 97-102.
(47) H. Vis, R. Boelens, M. Mariani, R. Stroop, C. E. Vorgias, K. S. Wilson, R. Kaptein, *Biochemistry* **1994,** *33,* 14858-14870, B. T. Farmer II, L. Mueller, *J.Biomol.NMR* **1994,** *4,* 673-687, S. M. Pascal, D. R. Muhandiram, T. Yamazaki, J. Forman-Kay, L. E. Kay, *J.Magn Reson.B* **1994,** *103,* 197-201, D. P. Frueh, D. A. Volsburg, C. T. Walsh, G. Wagner, *J.Biomol.NMR* **2006,** *34,* 31-40.
(48) Frueh, D. P.; Arthanari, H.; Wagner, G. *J.Biomol.NMR* **2005**, *33*, 187-196.
(49) Guo, C.; Zhang, D.; Tugarinov, V. *J.Am.Chem.Soc.* **2008**, *130*, 10872-10873.
(50) Kupce, E.; Cheatham, S.; Freeman, R. *Magn.Reson.Chem.* **2007**, *45*, 378-380.
(51) Kupce, E.; Freeman, R. *J.Am.Chem.Soc.* **2008**, *130*, 10788-10792.
(52) Schanda, P.; Forge, V.; Brutscher, B. *Magn.Reson.Chem.* **2006**,

44, S177-S184.
(53) Jaravine, V. A.; Zhuravleva, A. V.; Permi, P.; Ibraghimov, I.; Orekhov, V. Y. *J.Am.Chem.Soc.* **2008**, *130*, 3927-3936.
(54) Lescop, E.; Brutscher, B. *J.Am.Chem.Soc.* **2007**, *129*, 11916-11917.
(55) Hiller, S.; Garces, R. G.; Malia, T. J.; Orekhov, V. Y.; Colombini, M.; Wagner, G. *Scienc* **2008**, *321*, 1206-1210.
(56) Tugarinov, V.; Kay, L. E.; Ibraghimov, I.; Orekhov, V. Y. *J.Am. Chem.Soc.* **2005**, *127*, 2767-2775.
(57) Fiorito, F.; Hiller, S.; Wider, G.; Wüthrich, K. *J.Biomol.NMR* **2006**, *35*, 27-37.
(58) Balbach, J.; Forge, V.; Van Nuland, N. A. J.; Winder, S. L.; Hore, P. J.; Dobson, C. M. *Nature Struct.Biol.* **1995**, *2*, 865-870, N. A. J. Van Nuland, V. Forge, J. Balbach, C. M. Dobson, *Acc.Chem.Res.* **1998,** *31*, 773-780.
(59) B. Fuertig, J. Buck, V. Manoharan, W. Bermel, A. Jaeschke, P. Wenter, S. Pitsch, H. Schwalbe, *Biopolymers* **2007,** *86,* 360-383, J. Buck, B. Fuertig, J. Noeske, H. Schwalbe, *Proc.Natl.Acad. Sci.USA* **2007,** *40,* 15699-15704, K. H. Mok, L. T. Kuhn, M. Goez, I. J. Day, C. L. Jasper, N. H. Andersen, P. J. Hore, *Nature* **2007,** *447,* 106-109
(60) Gal, M.; Schanda, P.; Brutscher, B.; Frydman, L. *J.Am.Chem.Soc.* **2007**, *129*, 1372-1377.
(61) Schanda, P.; Forge, V.; Brutscher, B. *Proc.Natl.Acad.Sci.U.S.A* **2007**, *104*, 11257-11262.
(62) Hammarström, A.; Otting, G. *J.Am.Chem.Soc.* **1994**, *116*, 8847-8848.
(63) Andersson, P.; Weigelt, J.; Otting, G. *J.Biomol.NMR* **1998**, *12*, 435-441.
(64) Meissner, A.; Duus, J. O.; Sørensen, O. W. *J.Magn.Reson.* **1997**, *128*, 92-97.
(65) Sørensen, M. D.; Meissner, A.; Sørensen, O. W. *J.Biomol.NMR* **1997**, *10*, 181-186.
(66) Ottiger, M.; Delaglio, F.; Bax, A. *J.Magn.Reson.* **1998**, *131*, 373-378.
(67) Duma, L.; Hediger, S.; Lesage, A.; Emsley, L. *J.Magn.Reson.* **2003**, *164*, 187-195.
(68) Duma, L.; Hediger, S.; Brutscher, B.; Böckmann, A.; Emsley, L. *J.Am.Chem.Soc.* **2003**, *125*, 11816-11817.
(69) Shimba, N.; Stern, A. S.; Craik, C. S.; Hoch, J. C.; Dötsch, V. *J.Am. Chem.Soc.* **2003**, *125*, 2382-2383.
(70) Bertini, I.; Felli, I. C.; Kümmerle, R.; Luchinat, C.; Pierattelli, R. *J.Biomol.NMR* **2004**, *30*, 245-251.
(71) Bermel, W.; Bertini, I.; Duma, L.; Emsley, L.; Felli, I. C.; Pierattelli, R.; Vasos, P. R. *Angew.Chem.Int.Ed.* **2005**, *44*, 3089-3092.
(72) Bermel, W.; Bertini, I.; Felli, I. C.; Kümmerle, R.; Pierattelli, R. *J.Magn.Reson.* **2006**, *178*, 56-64.
(73) Bermel, W.; Felli, I. C.; Matzapetakis, M.; Pierattelli, R.; Theil, E. C.; Turano, P. *J.Magn.Reson.* **2007**, *188*, 301-310.
(74) Sattler, M.; Schleucher, J.; Griesinger, C. *Progr.NMR Spectrosc.* **1999**, *34*, 93-158.
(75) Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R. *Angew.Chem.Int.Ed.* **2004**, *43*, 2257-2259.
(76) Bermel, W.; Bertini, I.; Felli, I. C.; Pierattelli, R.; Vasos, P. R. *J.Magn.Reson.* **2005**, *172*, 324-328.
(77) Bermel, W.; Bertini, I.; Felli, I. C.; Lee, Y.-M.; Luchinat, C.; Pierattelli, R. *J.Am.Chem.Soc.* **2006**, *128*, 3918-3919.
(78) Pervushin, K.; Eletsky, A. *J.Biomol.NMR* **2003**, *25*, 147-152.
(79) Eletsky, A.; Moreira, O.; Kovacs, H.; Pervushin, K. *J.Biomol.NMR* **2003**, *26*, 167-179.
(80) Lee, D.; Vögeli, B.; Pervushin, K. *J.Biomol.NMR* **2005**, *31*, 273-278.
(81) Hu, K.; Vögeli, B.; Pervushin, K. *J.Magn.Reson.* **2005**, *174*, 200-208.
(82) Hu, K.; Vögeli, B.; Clore, G. M. *J.Am.Chem.Soc.* **2007**, *129*, 5484-5491.
(83) Pasat, G.; Zintsmaster, J. S.; Peng, J. *J.Magn.Reson.* **2008**, *193*, 226-232.
(84) Farès, C.; Amata, I.; Carlomagno, T. *J.Am.Chem.Soc.* **2007**, *129*, 15814-15823.
(85) Fiala, R.; Sklenar, V. *J.Biomol.NMR* **2007**, *39*, 152-163.
(86) Bermel, W., Bertini, I., Csizmok, V., Felli, I. C., Pierattelli, R., and Tompa, P. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins, *J.Magn.Reson.*, in press.
(87) Fischer, M. W. F.; Zeng, L.; Zuiderweg, E. R. P. *J.Am.Chem.Soc.* **1996**, *118*, 12457-12458.
(88) Bertini, I.; Felli, I. C.; Kümmerle, R.; Moskau, D.; Pierattelli, R. *J.Am.Chem.Soc.* **2004**, *126*, 464-465.
(89) Matzapetakis, M.; Turano, P.; Theil, E. C.; Bertini, I. *J.Biomol. NMR* **2007**, *38*, 237-242.
(90) Dyson, H. J.; Wright, P. E. *Chem.Rev.* **2004**, *104*, 3607-3622.
(91) Mittag, T.; Forman-Kay, J. *Curr.Opin.Struct.Biol.* **2007**, *17*, 3-14.
(92) Csizmok, V., Felli, I. C., Tompa, P., Banci, L., and Bertini, I. *J.Am. Chem.Soc.* **2008**, *130*, 16873-16879.
(93) Bertini, I.; Luchinat, C.; Parigi, G. *Solution NMR of paramagnetic molecules*; Elsevier: Amsterdam, 2001.
(94) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536-1549.
(95) Arnesano, F.; Banci, L.; Bertini, I.; Felli, I. C.; Luchinat, C.; Thompsett, A. R. *J.Am.Chem.Soc.* **2003**, *125*, 7200-7208.
(96) Bermel, W.; Bertini, I.; Felli, I. C.; Kümmerle, R.; Pierattelli, R. *J.Am.Chem.Soc.* **2003**, *125*, 16423-16429.
(97) Caillet-Saguy, C.; Delepierre, M.; Lecroisey, A.; Bertini, I.; Piccioli, M.; Turano, P. *J.Am.Chem.Soc.* **2006**, *128*, 150-158.
(98) Kostic, M.; Pochapsky, S. S.; Pochapsky, T. C. *J.Am.Chem.Soc.* **2002**, *124*, 9054-9055.
(99) Machonkin, T. E.; Westler, W. M.; Markley, J. L. *J.Am.Chem.Soc.* **2002**, *124*, 3204-3205.
(100) Pochapsky, S. S.; Sunshine, J.; Pochapsky, T. C. *J.Am.Chem.Soc.* **2008**, *130*, 2156-2157.
(101) Babini, E.; Bertini, I.; Capozzi, F.; Felli, I. C.; Lelli, M.; Luchinat,

C. *J.Am.Chem.Soc.* **2004**, *126*, 10496-10497.
(102) Bertini, I.; Jiménez, B.; Piccioli, M. *J.Magn.Reson.* **2005**, *174*, 125-132.
(103) Balayssac, S.; Jiménez, B.; Piccioli, M. *J.Biomol.NMR* **2006**, *34*, 63-73.
(104) Respondek, M.; Madl, T.; Göbl, C.; Golser, R.; Zangger, K. *J.Am. Chem.Soc.* **2007**, *129*, 5228-5234.
(105) Hilty, C.; Wider, G.; Fernandez, C.; Wüthrich, K. *ChemBioChem* **2004**, *5*, 467-473.
(106) Pintacuda, G.; Otting, G. *J.Am.Chem.Soc.* **2002**, *124*, 372-373.
(107) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc.Natl.Acad. Sci.USA* **1997**, *94*, 12366-12371.
(108) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *J.Am.Chem.Soc.* **1998**, *120*, 6394-6400.
(109) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J.Am. Chem.Soc.* **2003**, *125*, 10420-10428.
(110) Skrisovska, L.; Allain, F. H. T. *J.Mol.Biol.* **2008**, *375*, 151-164.
(111) Zhang, H.; Neal, S.; Wishart, D. S. *J.Biomol.NMR* **2003**, *25*, 173-195.
(112) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J.Am.Chem.Soc.* **2001**, *123*, 2970-2978.

Chapter 7

Perspectives of high-field dynamic nuclear polarization (DNP)

Authors: Thomas F. Prisner[a], Robert G. Griffin[b]
[a] Institute of Physical and Theoretical Chemistry and Center of Biomoleculear Magnetic Resonance, Goethe University, Frankfurt am Main, Germany
[b] Francis Bitter Magnet Laboratory and Dept. of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachussets, USA

Introduction

Sensitivity is a key issue in NMR spectroscopy, microscopy and imaging and often the factor that limits the success of various applications. The origin of low sensitivity in NMR is well known and is due to the small magnetic moment of nuclear spins. Correspondingly, this yields small nuclear spin polarizations and weak absorption signals. Historically, each advance in technology and methodology that has increased the signal-to-noise in NMR has shifted the boundaries of what is achievable, often opening new areas of application and directions of research. Recent technological developments of note include the continuing development of higher field superconducting magnets, and cryoprobes in which the excitation/detection coil is maintained at 20 K. In addition, innovations in NMR methodology have improved sensitivity, classic examples being Hartmann-Hahn cross polarization and INEPT transfer methods and the introduction of ^{1}H detection of ^{13}C/^{15}N resonances. Furthermore, techniques for non-inductive detection of resonance such as the AFM based technique of magnetic resonance force microscopy (MRFM) have recently allowed observation of a single electron spin and NMR signals from ~100 nuclear spins.

Coupling the nuclear spin states to systems having much higher polarizations, such as electron spins, is another approach to enhance the sensitivity in NMR experiments. For example, laser-polarized noble gases, chemical-induced dynamic nuclear polarization (CIDNP), para-hydrogen induced polarization (PHIP) as well as microwave-driven dynamic nuclear polarization (DNP) all rely on this principle. In the cases of CIDNP and PHIP, polarized states are generated by spin sensitive chemical reactions, and, while they are very successful, they are generally system specific. In contrast, microwave-driven DNP experiments are evolving as a broadly applicable approach to sensitivity enhancement in solid state and solution NMR. Thus, the discussion here focuses on advances in the millimeter wave technology required for DNP experiments at high magnetic fields as well as methods for polarization transfer.

As we will see below DNP currently improves the sensitivity in NMR spectra by ~10^2 and/or in principle reduces the acquisition time in multidimensional experiments by ~10^4 thereby permitting studies of larger molecules, dynamics of reactions, or high throughput screening. Concurrently, it can improve the information content by providing selectivity and contrast. For example, specific sections of a protein can be enhanced, metabolic cycles examined and contrast in MRI spectra increased. In structural studies of proteins additional constraints are in principle available from electron-nuclear dipolar or scalar coupling and paramagnetic shifts of sites in close proximity to spin labels or metal centers.

State-of-the-art

DNP is based on the transfer of the large electron spin polarization to nuclear spins (γ_e/γ_n>657). This concept, originally proposed by Overhauser in 1953 [1], was at first experimentally demonstrated in metals [2] and later also observed in liquids [3-5], two classes of samples with mobile electrons. In addition, analysis of low temperature DNP experiments designed to produce highly polarized solid targets for nuclear scattering revealed that other polarization transfer mechanisms are present. In particular when the paramagnetic centers are localized the so-called solid-state effect [6], cross-effect [7] and thermal mixing [8] dominate the polarization transfer, and involve the dipolar coupling of the nuclear spin to one, two or more electron spins, respectively. The theory for all three of these mechanisms predicts reduced transfer efficiencies at higher magnetic field values. This in combination with the paucity of high frequency microwave technology to effectively excite electron spins at magnetic field values above 1 T relegated

DNP to an interesting scientific curiosity. Concurrently, both solution and solid-state NMR were briskly moving towards higher magnetic fields and higher spectral resolution [9-11]. This situation persisted until the early 1990's when high field, solid state MAS DNP experiments directed at structural biology and utilizing gyrotron microwave sources were reported by Griffin and coworkers [12;13]. Subsequently, in 2003 the Amersham group explored the possibility of polarizing samples at very low temperatures followed by fast dissolution and heating of the samples [13;14] and observation of the liquid state spectrum. These two experiments received a good deal of attention in the magnetic resonance community and stimulated additional worldwide activities and initiatives in the field of solid- and liquid state DNP and high-frequency microwave technology.

New Instrumental Strategies

High Frequency Microwave Sources

Currently, semiconductor microwave technology (Gunn and IMPATT diodes) reaches its limit at frequencies of ~100 GHz, corresponding to a magnetic field of 3.5 T (150 MHz ^{1}H NMR). Higher frequencies can be attained most conveniently by generating higher harmonics, but with significant losses in power. Alternatives are vacuum electron devices, where an accelerated electron beam is modulated by suitable slow wave structures or magnetic fields. A number of different designs exist for continuous wave or pulsed operation, variable or fixed frequency. Devices such as backward wave oscillators, orotrons, and carcinotrons are used at high frequencies. Because of the presence of a slow wave structure, which has a size comparable to the microwave wavelength, the electron beam density close to this structure is limited, and leads to maximum deliverable microwave powers in the 0.1-1 W range.

Gyrotrons, referred to as fast wave devices, circumvent this problem by replacing the slow wave structure with a cavity immersed in a magnetic field. In this configuration CW output powers in the watt range are achieved in devices designed specifically for DNP at MIT and more recently at Fukui University [15-18]. In a gyrotron, an electron beam is launched from an annular cathode and accelerated through the field of a superconducting magnet. The field profile is designed to compress the beam as it moves through the vacuum tube to a resonant cavity that converts the transverse kinetic energy from the helical motion of the electrons into microwaves. A quasi-optical mode converter couples the radiation to the output window of the device and into at transmission line to the sample. Details of the physics and engineering of gyrotrons are described elsewhere [15-18] and their great virtue for DNP/NMR experiments is the fact that they are scalable to experiments in high magnetic fields. Specifically, because they are fast wave devices, they can therefore generate 10's of watts of microwave power at frequencies up to 800 GHz, corresponding to ^{1}H frequencies of 1.2 GHz. Furthermore, they operate in true CW mode for periods of days, enabling the multidimensional experiments that are commonplace in magnetic resonance. To date the gyrotrons used for DNP experiments are fixed frequency oscillators, but recently tunable sources were described, and gyroamplifiers for time domain experiments are under development.

Transmission lines

Transmitting the microwaves to the sample in the probe with minimal loss, and monitoring the microwave power output is important experimentally. Fundamental mode waveguides have unacceptable insertion losses, and do not couple to a free-space propagation of a Gaussian beam, which is typically used for quasi-optical transmission outside of the probe. Corrugated overmoded or metallodielectric waveguides can be used inside the DNP probe for transmission [19;20]. These differ from classical fundamental waveguides in that the losses in such systems are less then 1-2 dB. Detection of the EPR signal requires quasioptical duplexing devices to prohibit the strong excitation power from the microwave detector.

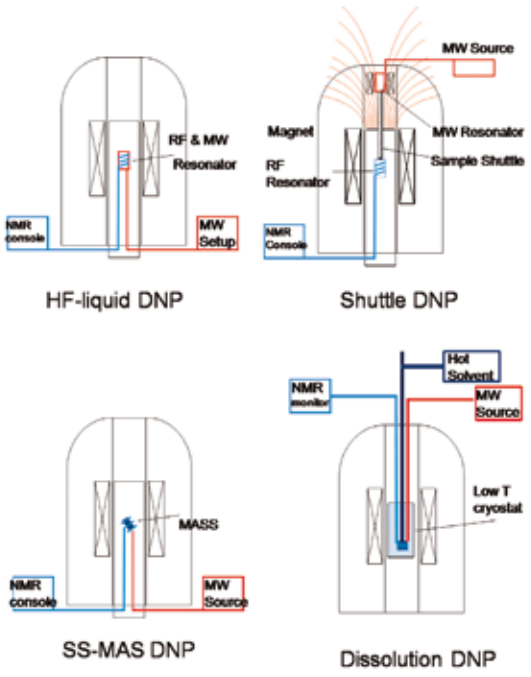


Figure 1: Typical experimental approaches for dynamic nuclear polarizaton spectrometers.

Solid-State MAS DNP

Figure 1 illustrates some typical DNP/NMR spectrometers. As mentioned above low sensitivity is a significant limit to the application of MAS dipolar recoupling experiments to biological systems. Accordingly this stimulated the development of DNP spectrometers opearating at 211 and 380 MHz proton frequencies [16;21;22]. Performing modern SSNMR experiments while irradiating with the high frequency microwaves required the development of a new generation of quadruple resonance MAS probes that support triple resonance NMR experiments as well as microwave irradiation, all at 90 K or lower. Generation of strong B_1 fields at the nuclear Larmor frequencies is usually accomplished with solenoid RF coils. The coil and MAS apparatus complicate the design of a μw resonator cavity and hence limit the quality factor, Q. This contrasts to Q values in EPR resonators which are ~10^3, whereas the Q for the μw circuit in the MAS probe is closer to unity. The inside of the stator cavity is typically coated with a thin layer of silver, which reflects the microwaves and increases the Q slightly. Initially, the μw radiation was introduced parallel to the MAS axis, but more recently probes employ irradiation perpendicular to the rotor axis. As long as the spacing between the turns of the coil is >l/2 λ the microwaves will penetrate the sample. Using a gyrotron microwave source concurrently permits the experimenter to satisfy the constraints of MAS at kilohertz frequencies at cryogenic temperatures, high NMR sensitivity, and efficient microwave pumping of the electrons. The MAS-DNP probe typically functions routinely down to 85 K and spinning frequencies of 10-15 kHz

Enhancements of 45-400 in the proton NMR signals have been obtained using the thermal mixing effect with mM concentrations of radicals at 90 and 20 K respectively, with static samples [12]. In MAS experiments enhancements up to ~300 are observed with biradical polarizing agents and the cross-effect [23]. The proton polarization is transferred to ^{13}C by classical solid state NMR CP and was further used for 1D- and 2D SS-NMR applications [22].

Low-T Dissolution Polarizer

This method utilizes DNP in the solid state at very low temperatures (1.2 K) at magnetic fields of 3-5 T. The polarization step is followed by rapid dissolution with a suitable solvent, and finally transfer of the sample to either a higher magnetic field, high resolution NMR spectrometer or MR imager. It has been reported that very high polarizations for ^{13}C can be retained within the dissolution and transfer process [14]. However, when comparing the reported enhancements it is important to keep in mind that they are the product of the DNP enhancement (up to 2600 for ^{13}C) multiplied by the large Boltzmann enhancement (about 250) due to the 1.2 K temperature where the DNP process occurs.

Typically, DNP polarization is obtained at a magnetic field of 3.5 T or 5 T, where all the necessary microwave components for the excitation of the electron spin system are commercially available [24;25]. Microwave excitation at 95 GHz (W-band, 3.5 T) is realized by solid state or vacuum tube devices with typical output power of more than 200 mW. Recently, experiments at 140 GHz (5 T) were also reported. The solid sample containing a high concentration of a suitable radical (usually trityl, but recently TEMPO) is excited by continuous microwave radiation for a period of hours and the progress of the polarization transfer can be monitored by ^{13}C NMR detection [26]. After the saturating polarization is obtained, the sample is rapidly dissolved in a warm solvent and transferred to another magnet for NMR or MRI measurements [27]. Dissolution and transfer times are in the order of a few seconds.

In-Situ TJ-DNP (laser melting)

Another approach to enhancing spectra of liquids is to polarize the sample in the solid-state as usual followed by rapid laser melting and subsequently to record a liquid state NMR spectrum. The melting is accomplished with an IR laser and an optic fiber [28] and is highly reproducible. The advantage to this approach for analytical NMR is that the sample is not diluted by a factor of ~100 as is the case with the dissolution experiment, and it is not necessary to physically move the sample between two different magnetic fields. Further, the existing repertoire of multidimensional solution state experiments can be inserted as mixing periods into this scheme. If the temperature for the polarization is reduced, then enhancements similar to those obtained with dissolution should be achieved. Finally since the scheme includes a ^{1}H-^{13}C cross polarization step, the long polarization times associated with directly polarizing ^{13}C or ^{15}N are shortened to that of the ^{1}H T_1. A similar approach would dramatically shorten the polarization times in the dissolution experiments.

HF-Liquid-DNP Spectrometers

Experiments performed in liquid solutions with radicals show

a significant attenuation of the polarization transfer efficiency at magnetic field larger than 1 T [3-5]. This effect could be rationalized as a reduced Overhauser effect efficiency related to the spectral density function of the rotational and translational motion of the radicals in solution. Thus, modeling of the spectral density function based on a simple spherical model predicts a vanishing polarization transfer at magnetic field values above 1 T. However, this calculations is at odds with experimental observations, which showed polarization enhancements in the range of 10 for such magnetic field strengths. More elaborate simulations of the rotational and translational motion of the radicals in solution support the experiments [29;30].

Based on these findings and predictions and on experimental DNP enhancements obtained in liquid solution at 5 T [31] a dedicated high-field liquid DNP spectrometers was designed at Frankfurt University, operating at 9.5 T magnetic field, 260 GHz EPR and 400 MHz proton NMR frequency [20]. The key problem is to avoid excessive microwave heating of the liquid water solution samples, mandatory for high-resolution NMR studies of proteins. The microwave penetration at such high microwave frequencies in liquid water is below µm excluding a design similar to low temperature solid state DNP applications, where the microwave excites the sample as a freely propagating wave. Instead, a microwave resonance structure, allowing spatial separation of the necessary magnetic fields from the futile electrical fields of the microwave, must be used. This design relaxes the restrictions on sample size – that the sample has to be much smaller than the microwave wavelength - to a single dimension. For the first prototype liquid DNP spectrometer a helical structure developed for ENDOR (Electron nuclear double resonance) experiments was used, where a cylindrical microwave resonator functions simultaneously as a multi-resonance NMR coil [32]. With such a structure it is possible to simultaneously apply microwaves and radiofrequency NMR pulses to the sample, allowing coherent manipulations of the coupled electron-nuclear spin system [33]. Additionally, the intense microwave magnetic field strength at the sample permits saturation of the EPR transitions of typical radicals in solution with low microwave power levels. For the cylindrical TE_{011} mode resonator currently used in Frankfurt, a microwave power of 45 mW, generated by a solid-state source is sufficient to saturate nitroxide radicals. The disadvantage of this fundamental mode design is the sample size, which is only 3 nl and the rather poor NMR coil filling factor. Enhancements of 7 and 11 were observed for water protons in a 40 mM solution with TEMPONE and Fremy' Salt nitroxide radicals respectively at room temperature [34]. Similar results were obtained at lower field strengths of 3.4 T (95 GHz) with a cylindrical ENDOR resonator [28].

More elaborate resonance structures are under development, which could potentially increase the sample volume up to few µl and achieve a higher NMR filling factor. Such extended structures would need microwave power levels of about 1 W to saturate the electronic transitions of nitroxide radicals.

Shuttle DNP

This approach exploits the fact that polarization transfer processes in liquids can be more efficient and technically less demanding at magnetic field values below 1 T [3-5]. In this case the microwave irradiation is typically performed at X-band frequencies (9 GHz / 0.3 T) in an electromagnet [35;36] or a permanent Halbach magnet [37;38]. The liquid sample resides inside a microwave cavity to achieve high microwave magnetic field strengths at the sample. This is required for saturation of the electron spin system of the radicals, which have very short relaxation times in liquid solution at room temperature. Following the polarization process, which only takes few seconds, the sample is shuttled to a high magnetic field for NMR detection. This experimental approach is used in flow systems [35], fast shuttling of the sample [38] or by a rapid transfer of the whole DNP probe [39]. Typical transfer times are in the order of several 100 ms. Because the polarization transfer is achieved at very low fields, a 'Boltzmann penalty' is present, which is the ratio of the polarization magnetic field (0.3 T) over the detection magnetic field (2-14 T). Depending on the application, ranging from MRI, analytical chemistry or high-resolution spectroscopy this factor might range from 1/5 to 1/60. A second problem related to this approach concerns the magnetic field profile during the transfer from the polarization field to the detection field. Care has to be taken as the polarization enhancement of the nuclear spins does not get blurred by passage through low magnetic fields [39]. Experimentally, enhancement factors for small molecules in water have been achieved of -3.5 for proton and -7 for carbon for a pump field of 0.3 T and a detection field of 14T [38]. This is close to the theoretical limit with the present apparatus. Repetition of the experiment is possible thus standard approaches to signal averaging and multidimensional NMR are possible.

Application areas

Some applications of DNP that highlight the potential of the method in different areas of research are as follows:

- *Biomolecular Structure Studies on Membrane Proteins:* MAS DNP was used to investigate the intermediate states in the photocycle of bacteriorhodopsin. The enhanced sensitivity of DNP permitted for the first time the characterization of the retinal conformation in the K, L and M states [15;40]. In addition, DNP enhanced MAS spectra of amloid nanocrystals illustrate the potential of polarizing structures with dimensions of 100-200 nm via ^{1}H spin diffusion from the solvent.
- *Ultrafast 2D-NMR spectroscopy*: Combination of gradient encoded ultrafast 2D-NMR methods [41] with dissolution DNP allow to obtain more dimensional spectra of polarized small molecules with very high sensitivity [42].
- *MRI with hyperpolarized metabolites*: ^{13}C-pyruvate was polarized with a dissolution spectrometer and used for cardiac MR imaging [43;44]
- *Time resolved DNP spectroscopy*: Reactions such as enzyme catalysis can be followed in real time by hyperpolarized NMR spectroscopy by following the kinetics of the substrate and product resonances [45]
- *Hydration and local water dynamics in lipids*: Nitroxide spin labels covalently attached along the hydrophobic tail of stearic acid molecules allowed to probe the interaction with solvent water molecules via DNP in micelles and vesicle assemblies [46].

Perspectives

All of the approaches to DNP discussed above are currently undergoing extensive instrumental development and simultaneously refinement of the experimental methods. We now briefly summarize the status of each area and suggest directions that will facilitate applications to problems in systems biology:

- MAS DNP experiments are currently operating at 140, 250, 263, and 395 GHz and plans are underway to move the experiments to 460 and 527 GHz. A series of stable biradicals are known to give much larger signal enhancements and are being optimize to further improve the results. At present the enhancements achieved in the experiment range from 40-400 depending on the details with 200-300 being routine in standard samples at 90 K. Details of the polarization transfer mechanism as well as applications of the method are subject of current research. The microwave excitation field used for this method can be delivered by fixed frequency gyrotrons adapted for DNP requirements. Higher frequency tunable CW sources are under development and will interface to existing solid state NMR spectrometters. The approach has successfully been applied to enhance MAS spectra of membrane and amyloid proteins as well as soluble proteins, and small molecules. Time domain experiments are expected to address the field dependence of the CW experiments. The instrumentation for MAS DNP experiments -- including gyrotrons, transmission lines, and low temperature probes -- is now commercially available and should accelerate propagation of the technique to many laboratories. In the past this has been the major limitation for a broader spread of this approach. A final topic that now requires attention is preparation of protein samples for the experiments.
- Dissolution-DNP experiments are currently operating at 95 and 140 GHz using Gunn and Impatt diodes to generate the microwaves for the polarization process. Limitations of the approach are the fact that the polarization times are hours, that the sample is diluted by a factor of ~70 in the dissolution process, that the dissolution is irreversibility, and that the sample must be shuttled between magnets. Some of these problems are partially circumvented by integrating single scan 2D-NMR methods into the technique. The DNP enhancement obtained in the experiment is roughly identical to that observed in gyrotron based experiment. Note, that the enhancements reported in many applications are relative to room temperature spectra, and therefore are a product of the DNP enhancement and the Boltzmann factor of ~250. The experiment has been used to boost sensitivity, time resolution and contrast for applications in spectroscopy and imaging, and will undoubtedly be developed further in the future.
- Laser-melting Liquid DNP uses approaches developed for MAS DNP experiments and melts the sample *in-situ* for observation of a solution spectrum. The polarization times are short, the sample is not diluted with solvent, the experiment can be recycled, and it is not necessary to shuttle between magnetic fields. The experiment will be most useful for small molecules which can be repeatedly frozen and thawed. It presently has a larger bandwidth than single scan techniques and when combined with multiple receivers, sparse sampling and other improvements could be quite useful for analytical NMR. The protocol could be extended to lower temperatures (10 K) and would benefit from the larger Boltzmann factor as does dissolution DNP.
- High-Field Liquid DNP allows polarizing the liquid solution

in-situ without altering the sample parameters within the polarization-detection cycle. Thus, if technical problems related to sample size, electric field heating and field homogeneity can be solved, this method could be very useful for extended NMR measurements typically encountered in structural studies on biological macromolecules. The simple Overhauser polarization transfer approach will not be applicable to very high magnetic fields and also the necessary resonant microwave structures will probably become impractical at field values above 10 T. On the other hand, the possibility of exciting electron and nuclear spins simultaneously without any time delays is unique for this approach and offers versatile prospects for coherent spin manipulations which might lead to improved polarization transfer pathways and new type of experiments.

- Shuttle-DNP combines well-established low-field EPR and high-field NMR knowledge in a straightforward manner. The shuttle process is the most critical technological part and must be optimized with respect to shuttle speed and field profile. The reduced polarization enhancement because of the polarization/detection field ratio might be compensated by more sophisticated polarization transfer pathways, as for example pulsed microwave excitations, which can be implemented at low microwave frequencies. Additionally, it might be possible to use higher polarized electron spin systems (as optical excited triplet states or radical pairs) in such systems.

All of these approaches are potentially applicable to NMR spectroscopy in biology, chemistry and physics and their successful development will have an enormous impact on the field. Recently, a number of academic and industrial research groups have initiated research efforts to overcome the current limitations of the techniques. Technical developments of high-frequency microwave sources and components and of the various DNP spectrometers will be of vital importance for the further development of this method. Other areas such as the optimization of polarizing agents, the development of new type of polarization transfer methods, and the design of new experiments concentrating on selectivity, contrast and additional structural restrains are research area ripe for investigation. Thus, collaborative efforts between scientists from chemistry, physics and biology will be required to optimize DNP for applications to high-field NMR and MRI.

Acknowledgments

The authors acknowledge many helpful comments and suggestions from several participants at the CA meeting *Magnetic Resonance in the Life Sciences: What's New?*, especially Claudio Luchinat, Christian Griesinger, and Lucio Frydman.

References

(1) Overhauser, A. W. *Phys.Rev.* **1953**, *92*, 411.
(2) Carver, T. R.; Slichter, C. P. *Phys.Rev.* **1953**, *92*, 212.
(3) Hausser, K. H.; Stehlik, D. *Adv.Magn.Reson.* **1968**, *3*, 79.
(4) Dwek, R. A.; Richards, R. E.; Taylor, D. *Ann.Rev.Nmr Spectroscopy* **1969**, *2*, 293.
(5) Muller-Warmuth, W.; Meise-Gresch, K. *Adv.Magn.Reson.* **1983**, *11*, 1.
(6) Abragam, A. *Phys.Rev.* **1955**, *98*, 1729.
(7) Hwang, C. F.; Hill, D. A. *Phys.Rev.* **1967**, *18*, 110.
(8) Goldman, M. *Spin temperature and magnetic resonance in solids*; Oxford, 1970.
(9) Wind, R. A.; Duijvestijn, M. J.; Vanderlugt, C.; Manaenscijn, A.; Vriend, J. *Prog.NMR Spectrosc.* **1985**, *17*, 33.
(10) Singel, D. J.; Seidel, H.; Kendrick, R. D.; Yannoni, C. S. *J Magn Reson* **1989**, *81*, 145.
(11) Afeworki, M.; Schaefer, J. *Macromolecules* **1992**, *25*, 4092-4096.
(12) Becerra, L. R.; Gerfen, G. J.; Temkin, R. J.; Singel, D. J.; Griffin, R. G. *J Chem Phys* **1995**, *102*, 9494-9497.
(13) Maly, T.; Debelouchina, G. T.; Bajaj, V. S.; Hu, K.-N.; Joo, C.-G.; Mak-Jurkauskas M.L.; Sirigiri, J. R.; van der Wel P.C.A.; Herzfeld, J.; Temkin, R. J.; Griffey, R. G. *J Chem Phys* **2008**, *128*, 052211.
(14) Ardenkjaer-Larsen, J.-H.; Fridlund, B.; Gram, A.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K. *Proc.Natl.Acad. Sci.U.S.A* **2003**, *100*, 10158.
(15) Bajaj, V. S.; Hornstein, M. K.; Kreischer, K. E.; Sirigiri, J. R.; Woskov, P. P.; Mak-Jurkauskas, M. L.; Herzfeld, J.; Temkin, R. J.; Griffin, R. G. *J Magn Reson* **2007**, *189*, 251.
(16) Bajaj, V. S.; Farrar, C. T.; Hornstein, M. K.; Mastovsky, I.; Vieregg, J.; Bryan, J.; Elena, B.; Kreischer, K. E.; Temkin, R. J.; Griffey, R. G. *J Magn Reson* **2003**, *160*, 85.
(17) Saito, T.; Nakano, T.; Hoshizuki, H.; Sakai, K.; Tatematsu, Y.; Mitsudo, S.; Ogawa, I.; Idehara, T.; Zapevalov, V. E. *Intl J Infrared and Millimeter wawes* **2007**, *28*, 1063-1078.
(18) Idehara, T.; Saito, T.; Ogawa, I.; Mitsudo, S.; Tatematsu, Y.; Agusu, L.; Mori, H.; Kobayashi, S. *Appl.Magn.Reson.* **2008**, *34*, 265.
(19) Woskov, P. W.; Bajaj, V. S.; Hornstein, M. K.; Temkin, R. J.; Griffin, R. G. *IEEE Trans.Microwawe Theory Techn.* **2005**, *53*, 1863.
(20) Denysenkov, V. P.; Prandolini, M. J.; Krahn, A.; Gafurov, M.; Endeward, B.; Prisner, T. F. *Appl.Magn.Reson.* **2008**, *34*, 289.
(21) Becerra, L. R.; Gerfen, G. J.; Bellew, B. F.; Bryant, J. A.; Hall, D. A.; Inati, S. J.; Weber, R. T.; Un, S.; Prisner, T. F.; McDermott, A. E.; Fishbein, K. W.; Kreischer, K. E.; Temkin, R. J.; Singel, D. J.; Griffin, R. G. *J Magn Reson* **1995**, *A117*, 28.
(22) Rosay, M.; Weis, V.; Kreischer, K. E.; Temkin, R. J.; Griffin, R. G. *J Am Chem Soc.* **2002**, *124*, 3214.
(23) Hu, K.-N.; Yu, H.-H.; Swager, T. M.; Griffin, R. G. *J Am Chem Soc.* **2004**, *126*, 10844.
(24) Wolber, J.; Ellner, F.; Fridlund, B.; Gram, A.; Johannesson, H.; Hansson, G.; Hansson, L.; Lerche, M. H.; Mansson, S.; Servin, R.; Thaning, M.; Golman, K.; Ardenkjaer-Larsen, J.-H. *Nucl.Instr. Meth.Phys.Res.A* **2004**, *526*, 173.
(25) Granwehr, J.; Leggett, J.; Kockenberger, W. *J Magn Reson* **2007**, *187*, 266.
(26) Reynolds, S.; Patel, H. *Appl.Magn.Reson.* **2008**, *34*, 495.
(27) Comment, A.; van den Brant, B.; Uffmann, K.; Kurdzesau, F.; Jannin, S.; Konter, J. A.; Hautle, P.; Wenckebach, W. Th.; Gruetter, R.; van der Klink, J. J. *Concepts Magn.Reson.* **2007**, *31B*, 255.
(28) Joo, C.-G.; Hu, K.-N.; Bryant, J. A.; Griffin, R. G. *J Am Chem Soc.* **2006**, *128*, 9428.
(29) Hoefer, P.; Parigi, G.; Luchinat, C.; Carl, P.; Guthausen, G.; Reese, M.; Carlomagno, T.; Griesinger, C.; Bennati, M. *J Am Chem Soc.* **2008**, *130*, 3254.
(30) Luchinat, C.; Parigi, G. *Appl.Magn.Reson.- A special volume on DNP* **2008**, *34*, 379-392.
(31) Loening, N. M.; Rosay, M.; Weis, V.; Griffin, R. G. *J Am Chem Soc.* **2002**, *124*, 8808.
(32) Bennati, M.; Rosay, M.; Bryant, J. A.; Griffin, R. G. *J Magn Reson* **1999**, *140*, 293.
(33) Weis, V.; Griffin, R. G. *Solid State Nucl.Magn.Reson.* **2006**, *29*, 66.
(34) Prandolini, M. J.; Denysenkov, V. P.; Gafurov, M.; Lybenova, S.; Endeward, B.; Bennati, M.; Prisner, T. F. *Appl.Magn.Reson.* **2008**, *34*, 399.
(35) Dorn, H. C.; Gitti, R.; Tsai, K. H.; Glass, T. E. *Chem.Phys.Lett.* **1989**, *155*, 227.
(36) Reese, M.; Lennartz, D.; Marquardsen, T.; Hofer, P.; Tvernier, A.; Carl, P.; Schippmann, T.; Bennati, M.; Carlomagno, T.; Engelke, F.; Griesinger, C. *Appl.Magn.Reson.* **2008**, *34*, 301.
(37) Armstrong, B. D.; Lingwood, M. D.; McCarney, E. R.; Brown, E. R.; Bluemler, P.; Han, S. *J Magn Reson* **2008**, *191*, 273.
(38) Muennemann, K.; Bauer, C.; Schmiedeskamp, J.; Spiess, H. W.; Schreiber, W. G.; Hinderberger, D. *Appl.Magn.Reson.* **2008**, *34*, 321.
(39) Miesel, K.; Ivanov, K. L.; Koechling, T.; Yurkovskaya, A. V.; Vieth, H.-M. *Appl.Magn.Reson.* **2008**, *34*, 423.
(40) Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J. *Proc.Natl.Acad.Sci.U.S.A* **2008**, *105*, 883.
(41) Frydman, L.; Blazina, D. *Nature Physics* **2007**, *3*, 415.
(42) Mishkovsky, M.; Frydman, L. *ChemPhysChem* **2008**, *9*, 2340.
(43) Golman, K.; Zandt, R. I.; Lerche, M.; Pehrson, R.; Ardenkjaer-Larsen, J.-H. *Cancer Res* **2006**, *66*, 10855.
(44) Kohler, S. J.; Yen, Y.; Wolber, J.; Chen, A. P.; Albers, M. J.; Bok, B.; Zhang, V.; Tropp, J.; Nelson, S.; Vigneron, D. B.; Kurhanewicz, J.; Hurd, R. E. *Magn Reson Med* **2007**, *58*, 65.
(45) McCarney, E. R.; Armstrong, B. D.; Kausik, R.; Han, S. *Langmuir* **2008**, *24*, 10062.
(46) Bowen, S.; Hilty, C. *Angew.Chem.Int.Ed* **2008**, *47*, 47.

Chapter 8

Complementary techniques to NMR for structure determination of biological macromolecular complexes: from atoms to the cell – from NMR to cryo-EM

Authors: Lucia Banci[a], Wolfgang Baumeister[b], Ivano Bertini[a], Alex Grishaev[c], Lee Makowski[d], Antonio Rosato[a], Michael Sattler[e], Dmitri Svergun[f], Peter A. Timmins[g], Matthias Wilmanns[f]

[a] Magnetic Resonance Center, University of Florence, Italy
[b] Francis Bitter Magnet Laboratory and Dept. of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachussets, USA
[c] National Institutes of Health, Bethesda, USA
[d] Argonne National Laboratory, Chicago, USA
[e] European Molecular Biology Laboratory, Heidelberg, Germany
[f] European Molecular Biology Laboratory, Hamburg, Germany
[g] Institut Laue-Langevin, Grenoble, France

Future challenges in structural genomics and structural biology

The genomic revolution prompted the need for characterizing the products of the human genome and of an ever increasing number of organisms. The knowledge of their structure is a fundamental and necessary step for the understanding of their molecular function and the mechanisms of life of any living organism at the molecular level. In the meantime, the investigation of the whole network of interacting proteins produced by a living organism (the 'interactome') and of how these interactions are being modulated by changing concentrations of metabolites (the 'metabolome') have become emerging field of research[1-2]. Interactome studies provide information on the biochemical processes which occur through protein-protein and protein-DNA/RNA interactions. Such interactions are often weak presumably because they have to be "reversible" in order to repeat the operation whenever and as many times as needed by the organism. Sometimes protein-protein interactions are strong ($K_D < 10^{-6}$ M), but must be triggered by other events e.g. as in the case of EF-hand proteins where the triggering event is calcium binding. Protein-protein and protein-nucleic acid interactions are relevant to several aspects of the physiology of living cells, such as transcription, DNA repair, RNA processing, regulation of gene expression, cell surface recognition and adhesion, and signal transduction, to name a few. Malfunctioning of these processes is often linked with, or can lead to, pathological states. The understanding of the molecular bases of diseases also requires an atomic-level knowledge of the mechanisms of interaction between different biological macromolecules. Therefore, one of the next major challenges for structural biology will be the integration of structural knowledge at the cellular level in the context of systems biology. A major, further step ahead in the comprehension of the processes of Life will be deciphering the human interactome, which will be completed soon. While this knowledge would contribute to advance the comprehension of functional processes, it will also open a large range of new questions and unresolved problems based on the description of the interactions at the atomic level. These will constitute new challenges for scientists in the next years. Within this frame, NMR can play an essential and unique role.

Usefulness of NMR

NMR, together with X-ray diffraction, is one of two techniques that are at the heart of structural biology. Both are used nowadays to provide molecular structures at the atomic level. Despite structure determination of biological macromolecules by NMR began 30 years later than X-ray diffraction, NMR is now a well established, core technique that contributes about 15% of the structures annually deposited in the Protein Data Bank. Although NMR has some limitations with respect to the size of the molecules and the resolution of the structures, it provides considerably more than mere structural information, i.e. it goes beyond a static picture of the three-dimensional structure of genome products. In particular, it can give functionally relevant information on molecular dynamics and can characterize weak and strong interactions with other biomacromolecules as well as small ligands. Transient intermolecular interactions are crucial for phenomena such as regulation of protein expression, enzyme activation/repression, signalling, etc. NMR is particularly well suited to study weak and transient interactions, as it allows researchers to investigate the systems

of interest in solution, which is often the physiologically relevant state, or, for membrane systems, in bi-layers or micelles, which simulate the membrane environment. Protein-protein interactions, when weak, cannot be studied with any other technique at the atomic level. In this respect the results of interaction studies by NMR can provide unique information for structural analysis of the interactome maps.

From a more technical point of view, if NMR chemical shift assignments are available and the structure has been determined by either NMR, X-ray or modeling, then simple NMR chemical shift variations provide information on the protein surfaces which are interacting. Programs are already available for proposing models of the complexes on this basis, making NMR an excellent tool for the characterization of protein-protein as well as protein-DNA/RNA adducts. NMR can be used to determine the structure of the complex in solution, both in the case of strong interactions, when therefore a stable complex is formed, and of weak interactions, when the complex is exchanging with the isolated proteins, thus obtaining an average structure. NMR can characterize internal motions as well as the overall tumbling of macromolecules. Therefore, it can also provide insight into the dynamics of protein-protein adducts, as well as of other intermolecular adducts. The dynamics of adducts can be important for molecular recognition processes (e.g. selectivity or induced fit binding). Changes in the internal motion regime occurring in the partners upon complex formation are also important to understand functional features, and can be fully characterized only by means of NMR techniques.

NMR is also particularly suitable for the study of intrinsically unstructured proteins, as required by their function, and of large protein aggregates, which are the results of pathogenic misfolded states. The role of NMR in drug design is well established. A frontier is represented by the screening of molecules interacting with membrane protein receptors. Finally, the impact of NMR in metabolomics, i.e. the analysis of living organism small molecules, is growing steadily.

The outcomes of Structural Genomics ultimately need to be integrated with cell and systems biology, to be able to describe how a cell functions. This would have a great impact and constitute the background knowledge for understanding pathological states originating from alterations of the necessary cell functional processes and/or the presence of SNPs and mutations. So the outcomes of these studies would be beneficial for human health as they are functional for developing therapeutic treatments and new, more efficient drugs.

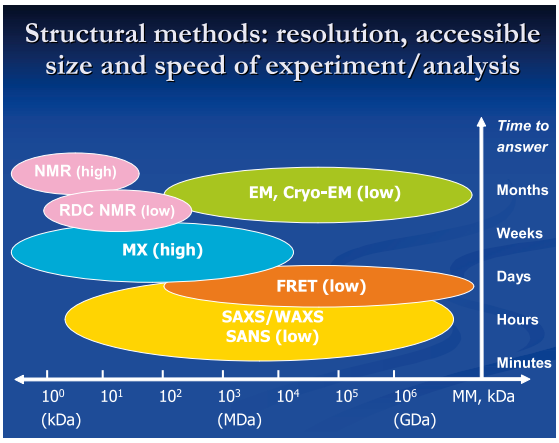
For the above reasons, we can anticipate that the role of NMR will become increasingly important. In parallel, the evolution of NMR technical and methodological aspects will contribute to widening the range of systems amenable to NMR studies, in particular by relieving the current limitations in terms of molecular size. NMR can be applied to investigate protein structures and protein-protein interactions, as well as nucleic acid structures and protein-DNA/RNA interactions. These scientific aspects have significant direct implications for the understanding of the information derived from genome sequences, as they directly address not only the question of the 3D structure of gene products but also of their mechanism of function, in particular of how the dynamic interactions between proteins, RNA and DNA work together to define biological pathways.

Discussion rounds on complementary techniques

Two major events took place during the NMR-Life Coordination Action related to the joint use of NMR with complementary methods. First, a workshop was organized in Florence, on May, 5-6, 2006 by Lucia Banci and Michael Sattler, with the aim of discussing the potentials of combining NMR spectroscopy with other techniques relevant for samples in the liquid state, in order to advance the scope and applicability of biophysical methods for structure determination. Second, a special discussion session devoted to complementary techniques was chaired by Dmitri Svergun at the closing meeting of NMR-Life. Below a summary of the major results of the two meetings is presented.

NMR is now a well established technique for structure determination in solution, i.e. in conditions as close as possible to the physiological ones and in the most recent years it is developing into a relatively HTP method for structure determination. One of the next steps in SG is the structure determination of protein-protein or macromolecular complexes. While efforts in this direction are underway (see for example: http://www.3drepertoire.org/ and http://www.3dem-noe.org), they focus on relatively sturdy complexes involving high-affinity interactions. On the other hand, weak and transient protein-protein interactions are key to most of the functional processes in the cell, but these complexes in most of the cases escape crystallization. Under these conditions NMR can be quite powerful as it can characterize weak or transient molecular interactions, where a protein binds in fast exchange with the free form. In this respect, NMR will have a relevant impact in the EC-funded integrated project (IP) SPINE2-COMPLEXES (http://www.spineurope.org/). However, despite the huge advancements seen in the last years, NMR is still limited in the molecular size of

the systems that can be studied. Presently, macromolecules and complexes larger than 100 kDa can be hardly characterized by NMR. Therefore, there is the need of integration with other techniques.



Complementarity of structural methods

A schematic complementarity chart of different methods to analyze the structure of biological macromolecules is presented in the Figure, presenting estimated resolution ranges and the time required for the experiment and analysis of the data. All the methods shown with the exception of macromolecular crystallography are applicable in solution. NMR together with crystallography can provide high resolution structures of individual proteins and protein domains. Complementary scattering techniques, described in more detail in the next sections, allow the study of the overall structure of individual macromolecules and complexes, while cryo-electron microscopy can be employed to provide additional low resolution information on large macromolecular assemblies. Cryo-electron tomography can locate multi-protein complexes within the cell, at a resolution superior to light microscopy. Additional long-range distance information can be obtained from FRET data. Thus, while interactome studies provide coarse and somewhat abstract information about the overall genome, NMR, x-ray and neutron scattering, and cryo-electron microscopy give atomic to molecular level structural information describing cellular processes.

During the two discussion rounds we have analysed the above-mentioned techniques in the perspective of their integration with NMR spectroscopy. We considered their potentialities and limits in this respect, as well as their requirements in terms of experimental conditions and characteristics of the systems investigated in order to define the corresponding ranges of applicability. Finally, we analysed the information that can be obtained from their application in particular to protein-protein complexes. Focus has been made on possible methods for the assessment and the validation of the results obtained through the above-mentioned techniques.

Description of Complementary Techniques

SAXS

Small-angle X-ray scattering (SAXS)[3] allows one to study native biological macromolecules, from individual proteins to large complexes (1 to 200 nm) in solution. SAXS is not limited by the molecular size and it is well suited to determine at least the shape of molecules in a broad range of molecular masses including those of a few hundred kDa (too large for NMR and too small for cryo-electron microscopy (EM)) in their native conditions. SAXS does not only provide low resolution three-dimensional models of particle shapes, but kinetic SAXS experiments allow one to analyze structural changes in response to variations in external conditions, protein-protein and protein-ligand interactions, and to study kinetics of assembly/dissociation or folding/unfolding. Nowadays, high flux SAXS beamlines at various synchrotrons allow for time-resolved studies of fast structural transitions in the sub-millisecond time region of solutions and partly ordered systems yielding resolution in the range from 100 to 1-2 nm. The time limit for these measurements is the mixing of the solutions, which is set by the mixing technique used (typically stopped-flow experiments). SAXS requires relatively small amounts of sample (20-50 µl of a solution containing about 1 mg/ml of protein). It should however be taken into account that SAXS may often destroy the sample because of radiation damage. In terms of sample conditions, up to e.g. 4.0 M NaCl is an acceptable ionic strength. However, very high salt concentrations can promote radiation damage and will decrease the electron density contrast between the macromolecule and the solvent, leading to the loss of the scattering signal intensity. Use of denaturants with high electron density such as guanidinum chloride are not recommended because of absorption problems, while urea is generally tolerated.

In the presence of equilibria between species in fast exchange, SAXS cannot discriminate between time averaging and ensemble averaging. Some information can nevertheless be obtained by performing a series of experiments at varying concentrations, e.g. during a titration allowing one to shift the equilibrium and to change the species distribution. SAXS can be performed on the same sample as for NMR. However, the two techniques are affected by sample aggregation at a different extent. Indeed, SAXS is much more sensitive than

NMR to this problem, and even a small percentage (5-10%) of aggregated species can be deleterious to data analysis, while it would not be observed by NMR. On the other hand, local small-magnitude structural changes such as those corresponding to conformational exchange can have a significant adverse effect on the NMR spectra, while only marginally affecting SAXS curves.

Initial strategies for the application of SAXS in conjunction to NMR spectroscopy have been already developed. For example, methods for the combination of NMR and rigid body mofdeling in SAXS and SANS have been proposed and implemented (4, 5, 6). Here, the utility of NMR residual dipolar couplings (RDCs) is applied to define the relative orientation of structural domains or complex subunits, while scattering data are used to obtain long-range translational restraints. SAXS data thus allow one to reduce the orientational ambiguities of RDC data or to exploit different implementations of restraining potentials. These potentials permit the simultaneous use of SAS and RDC data in structure calculations based on molecular dynamics/simulated annealing methods. Future challenges will be to implement and further improve such protocols for routine structural analysis of macromolecular complexes in solution. For example, it is likely that additional data will be useful and may be required to obtain unambiguous structural models. Other issues are the combined use of SAXS and SANS data, for example, to study protein-lipid or protein-nucleic acid complexes, where contrast matching can be employed to obtain additional information (see below). Very important is also the joint use of SAXS and NMR for the analysis of flexible systems, including intrinsically unfolded proteins (Bernado et al, 2007).

SAXS experiments are best carried out at dedicated synchrotron beam lines, taking advantage of their high brilliance. These lines are typically equipped with a thermostated flow-through cell to minimize radiation damage, an area detector and a setup allowing for a variable sample-to-detector distance distance and variable energy of the indent radiation. In Europe, Grenoble (ESRF), Hamburg (DESY), Paris (Soleil), Oxford (Diamond), and Trieste (Elettra) provide access to SAXS users, with funds awarded by the EC. The number of users performing SAXS measurements in order to combine their results with NMR data is steadily and rapidly growing, and their share is currently, over 20% at the X33 beamilne of the EMBL in Hamburg optimized for biological solution scattering. As the typical SAXS experiments and their analysis require much less time than NMR studies, and the overall structural information from SAXS is highly complementary to more local NMR information, application of SAXS should become a standard supplementary step in the structural NMR studies of proteins and complexes.

An important issue in SAXS method development is validation of the scattering data and the models resulting from its interpretation. In the absence of any prior structural information that could be compared against parameters directly extractable from the data such as approximate radius of gyration or maximum molecular dimension, one has to rely on the interpretation of the zero-angle extrapolated scattering intensity to confirm both the expected oligomerization state and absence of aggregation. For such measurement, acquisition of the scattering data on the samples of standards such as lysozyme or BSA, coupled with an accurate measurement of the macromolecule's concentration, are required. Validation of the structural models produced by the analysis of scattering data is more challenging and ultimately depends on the availability and amount of the prior structural or biochemical information about the molecule. In cases where SAXS data is combined with NMR data, *ab initio* SAXS-only low-resolution density reconstructions could be compared with NMR/SAXS generated high-resolution models using a measure like normalized spatial discrepancy to assess their similarity. An additional way to validate SAXS-only density reconstruction is by performing such reconstructions with several different software packages such as GASBOR, DAMMIN, SASHA, SAXS3D, etc.

A crucial issue for the incorporation of SAXS data into the main-stream structural biology is standardization of the data formats and deposition of the SAXS data into a publicly accessible database. One possible way to approach this is by requiring data deposition upon publication of any work that involves structural interpretation of SAXS data into a database such as RSCB PDB (this issue was discussed at a recent IUCr Congress (Osaka, August, 2008), and appropriate measures are planned).

SANS

Small angle neutron scattering (SANS) is conceptually very similar to SAXS but uses neutron radiation and exploits important differences between their scattering properties and those of X-rays. The most important factor is the scattering length density, the neutron analogue of electron density for X-rays. Neutrons are scattered by the nucleus of an atom rather than the electronic cloud as is the case for X-rays. The scattering process for neutrons involves the formation of a complex nucleus with subsequent liberation of a neutron and is a complex process which is not related in any simple way to

the mass or atomic number of the atom contrary to the X-ray case. Thus the scattering length densities of heavy atoms are not necessarily any greater than those for light atoms and can indeed be different for different isotopes of the same element which, by definition, contain a different number of neutrons and therefore have a different nuclear structure. In practise the scattering length densities for neutrons of the atoms common found in biological molecules vary only from 0.54 for carbon to 0.94 for nitrogen, all other atoms, with one exception, being between these extremes. The striking exception is hydrogen which has a negative scattering amplitude and is hence easily distinguished from all other atoms. The other important factor is the scattering power of deuterium, an isotope of hydrogen, which is positive and has a scattering amplitude very similar to that of carbon (0.66). Substitution of hydrogen by deuterium is therefore a powerful labelling technique in all domains of neutron scattering, from crystallography, through SANS to inelastic scattering. Concentrating on SANS, the difference between hydrogen and deuterium scattering leads to the contrast variation method whereby mixtures of H_2O and D_2O have very different scattering density as do proteins, nucleic acids or lipids. Similarly if biological macromolecules can be perdeuterated then their scattering is very different from their hydrogenated counterparts. Via the contrast variation method it is possible to perform SANS experiments on a macromolecular complex in aqueous solution where one component of the complex can be rendered invisible by choosing an appropriate H_2O/D_2O mixture such that solvent and that component have the same scattering density and hence zero contrast. Hence in this way the structure of different parts of a complex can be determined individually without dissociating the complex. There are numerous examples of this such as the study of the troponin complex (King *et al*, 2005) where samples having any one or two or all three sub-units could be deuterated and by carrying out SANS experiments in solutions containing 42% D_2O/58% H_2O it was possible to visualize, and hence model, the deuterated sub-units alone. Another potentially important class of molecules which can be investigated using this method is membrane proteins solubilised in lipids or detergent micelles. Lipids or detergents are invisible in water containing no or very little Deuterium and thus a neutron scattering experiment in these condition leads to a structure of the membrane protein alone without its solubilising layer of lipid or detergent.

Sample requirements for SANS are similar to SAXS requiring a slightly greater sample volume, 150 – 300µl per sample multiplied by the number of contrasts. There is however, no radiation damage with neutrons which avoids the necessity for multiple or circulating samples. Typical concentrations are 1 – 10 mg/ml. Small molecule additives such as salts or denaturing agents such as urea have no effect on the contrast so long as they do not form nanometre sized aggregates such as micelles. Experiments are usually performed in standard quartz cuvettes which may also be used for UV measurements to determine sample concentration. Exposure times are very variable depending on sample concentration, macromolecular dimensions as for X-rays but also on contrast and signal to noise factors. The highest contrast and hence strongest signals are obtained from deuterated proteins in H_2O or hydrogenated samples in D_2O. The latter conditions provide the lowest background which comes mainly from the incoherent scattering of hydrogen atoms.

Data analysis is in most respects similar to that used for X-rays, particularly in model fitting as exemplified in the methods described above for SAXS. The particularity is the increased information content available due to contrast variation leading to several different scattering curves being obtained from the same molecular complex. Interpretation at a simpler level, that of radius of gyration, can also give important information as the contrast variation method allows distances between subunits of a complex to be determined in a model independent way. Data may also be placed rather easily on an absolute scale allowing molecular weight measurements to be made.

The complementarity with NMR lies, as with SAXS, in the possibility of defining an overall molecular envelope delimiting the overall structure of the molecule or molecular complex. Unlike a crystallographic structure it is measured in solution, as with NMR, and the envelope obtained is a space and time average over all solution conformations.

SANS experiments can of course only be carried out at a neutron source which is without exception a central facility – no laboratory neutron sources suitable for scattering exist. Indeed the major difference between neutron and X-ray facilities is the much smaller flux available at even the most powerful sources compared with synchrotron radiation. This is however compensated in part by bigger beams and the flexibilities introduced by the contrast variation method as described above. The most powerful SANS facilities currently available in the world are the D11 and D22 SANS instruments at the Institut Laue-Langevin (ILL) in Grenoble, France. The ILL has the world's most powerful research reactor and is a facility available to all its member countries and even to a limited extent to non member countries. Set up in 1967, the reactor started operation in 1972 and is owned and run by France, Germany and

the UK. It has also as scientific members, currently (October 2008) Austria, Belgium, the Czech Republic, Hungary, Italy, Poland, Spain, Sweden and Switzerland. Beam time is allocated by peer review committees twice a year and experimenters from member countries who have their experiments selected also have the travel and subsistence paid to carry out the experiments. More details can be obtained from http://www.ill.fr. The ILL also has a laboratory specialising in the production of deuterated macromolecules for neutron scattering experiments. This laboratory is also open to outside users in a way similar to the neutron instruments except that treatment of proposals is entirely by e-mail and application may be made at any time. Further details are available at http://www.ill.fr/deuteration.

SANS facilities are operated also at several other neutron sources in Europe, the USA, Japan and Australia. There are both reactor sources such as the Laboratoire Leon Brillouin at Saclay in France or the FRM-2 reactor in Munich Germany and spallation sources such as ISIS at the Rutherford Laboratory, UK or the SNS in Oak Ridge, USA. All these facilities have external user programs and the list is by no means exhaustive.

WAXS

Wide-angle x-ray scattering (WAXS) from proteins in solution is an emerging technique which is a direct extension of SAXS derived through the collection and analysis of data to significantly higher scattering angles. Whereas most SAXS studies of proteins in solution are limited to $q < 0.2$ Å^{-1} ($q = 4\pi \sin(\theta)/\theta$ and 2θ is the scattering angle), WAXS studies have extended that limit to collect the weak data that extend to $q \sim 3.0$ Å^{-1}. The amount of information in solution scattering patterns is proportional to the resolution limit, so that, in principle, WAXS patterns may contain up to 15 times as much structural information as a SAXS pattern. As outlined above, SAXS data can be used to generate information about the size and shape of a protein. WAXS data provide information about finer details of the structure since scattering at progressively wider angles corresponds to progressively smaller structural features within the protein. X-ray scattering data from proteins in solution correspond roughly to the spherical average of data collected by x-ray crystallography. Thus they preserve information about the relative abundances of interatomic vectors of different lengths, but contain no information about their relative orientations. At very small angles of scattering (i.e. low q), long inter-atomic vectors dominate, limiting the use of small-angle x-ray scattering (SAXS) to determination of protein size and shape. At wider angles (i.e. higher q), shorter inter-molecular vectors contribute proportionately more to the intensity, embedding information about protein secondary and tertiary structure in the weak wide-angle scattering. As such, a WAXS pattern represents a weighted mapping of all the interatomic vector lengths in a protein. Since secondary structural elements such as α-helices or β-sheets have characteristic patterns of interatomic vector length, WAXS data can provide information about the secondary, and potentially, tertiary structure of a protein.

A major advantage of WAXS is that WAXS data can be predicted from atomic coordinates. A number of software packages that carry out this calculation are available, with CRYSOL being the most widely used. This capability makes it possible to test structural models based on atomic coordinates against experimental WAXS data. For instance, if the structure of two proteins are known but the way in which they interact is unknown, WAXS data from a solution of the complex can be used to test models for their interaction by comparing experimental data to that predicted from the atomic coordinates of the individual proteins and their relative positions and orientations. Consequently, WAXS data from complexes that cannot be studied by NMR or x-ray crystallography may provide adequate information for construction of a molecular model of the complex. WAXS is also well suited to the study of conformational changes in macromolecular complexes. Changes in WAXS data reflect changes in quaternary, tertiary or secondary structure, providing them with the potential for being a sensitive, global method for detecting structural changes in proteins induced by ligand binding, changes in environment or interactions with other macromolecules. WAXS is a sensitive tool for detection of the structural changes in proteins that accompany functional changes, thereby providing an indirect assay for functional interactions with small molecules. Recently, time-resolved WAXS experiments have been used to characterize structural transitions in proteins over time intervals as short as 10 ns, opening a new opportunity for the study of protein dynamics[2].

The use of WAXS has, in the past, been limited by the relative weakness of scattering from proteins at relatively wide angles, and by the presence of strong scattering from buffer solutions in this range. More recently, however, third-generation synchrotron sources, such as the European Synchrotron Radiation Facility (ESRF; Grenoble) and the Advanced Photon Source (APS; Chicago) have made possible the collection of wide-angle x-ray scattering (WAXS) of macromolecules in solution

to unprecedented resolutions. Although wide-angle x-ray scattering data are significantly weaker than the small angle data collected using SAXS, they can generally be collected using less than 100 µl of a solution with protein concentrations of 5–10 mg/ml in less than 30 s with the high flux available at a third generation synchrotron X-ray source. Higher concentrations result in data with higher signal to noise ratio. In general, WAXS data is less sensitive to the effects of aggregation than SAXS data. That, plus the need to measure much weaker data than in SAXS experiments, makes the use of more concentrated solutions far more favorable than in SAXS experiments. Care must be taken to avoid radiation damage to proteins studied by WAXS. Radiation damage and/or radiation-induced aggregation can be averted by a number of experimental parameters, including flowing the protein sample through the beam during data collection.

WAXS provides data complementary to NMR data in a manner similar to that provided by SAXS data. WAXS data contains information about the relative abundance of interatomic vectors that vary in length from ~ 2Å to the diameter of the scattering particles. As such, it provides information about the global arrangement of material in the protein or protein complex. It is not, in general, possible to obtain information about a specific inter-atomic vector from WAXS data. But the data provide an accurate overall view of the distribution of interatomic vector lengths in the protein which can provide important constraints on the form of the molecular model constructed from NMR data.

FRET

Fluorescence Resonance Energy Transfer (FRET) imaging is a powerful microscopy technique that overcomes some of the usual limitations of light microscopy to allow researchers to visualize and quantify protein associations under physiological conditions in individual cells. Conventional widefield fluorescence microscopy enables localization of fluorescently labelled molecules within the optical spatial resolution limits defined by the Rayleigh criterion, approximately 200 nanometers (0.2 micrometer). FRET, when applied to optical microscopy, permits determination of the approach between two molecules within several nanometers, a distance sufficiently close for molecular interactions to occur. The introduction of the green fluorescent protein (GFP) to FRET-based imaging microscopy allowed the use of this technique as a sensitive probe of protein–protein interactions and protein conformational changes *in vivo*. This was the beginning of real-time *in vivo* imaging of dynamic molecular events, providing researchers with crucial insight into the biological mechanisms as well as the physiological functions of a cell. This technique allows studying interactions that occur in the pico and nanosecond time scale. However, it can also be as a powerful tool to characterize conformational changes, both in vitro and in vivo, thereby providing a unique link to correlate atomic resolution structural information with biological function in a living cell. In many cases qualitative information derived from FRET changes can suffice to provide an understanding of conformational changes.

While examples are available of FRET measurements used in combination with X-ray structures, few examples are available for the combined use of FRET and NMR data applied to structural analysis of macromolecular complexes. NMR is being used to characterize and study the molecular basis of a newly designed FRET probe that exploits a phosphorylation-induced conformational switch.

FRET data provide local information on protein-protein distances (at longer distances than NMR, typically tens to hundreds of Å). As a consequence, these data are best suited for the validation of a structural model of an adduct or to suggest possible ways of interaction. The application of FRET requires that the protein partner can be chemically modified to introduce on specific residues (generally engineered cysteine residues for proteins) a pair of donor and acceptor fluorophores. This is needed in order to be able to measure a residue-residue distance within the complex in solution. While in some instances this kind of information can be of invaluable importance, it is unlikely that FRET can become a standard technique for the high-throughput structural characterization of complexes. It is also to be noted, that the information provided by FRET could be more readily obtained by pulsed electron-electron double resonance method, which requires spin labelling instead of fluorescent probes.

Cryo EM

Single particle reconstruction methods from EM of frozen-hydrated samples allow the structural analysis of macromolecular objects of molecular weights > 150 kDa. While resolutions between 10 and 15 Å can be achieved routinely, higher (< 10 Å) resolution can only be attained with advanced instrumentation and extensive image classification and averaging. Higher (atomic) resolution structures of components (subunits, domains) derived from crystallography and/or NMR can be fitted

into the medium resolution structures of complexes obtained by cryo EM. Image reconstruction relies on averaging of large numbers (up to 100,000) of individual particles. When there are sample inhomogeneities or when different conformational states coexist, image classification must precede averaging. EM of frozen-hydrated samples ensures a close-to-life preservation at the expense of a low signal-to-noise ratio. Negative stain provides high contrast and is therefore less demanding in terms of data processing but it is prone to artefacts and yields only low-resolution structural information (20-30 Å).

Cryoelectron tomography is the only technique that enables the study of large pleomorphic structures such as organelles or whole cells embedded in vitrified ice and, therefore, in a close-to-life state, with a resolution of a few nanometers. With the advent of computer-controlled transmission electron microscopes equipped with eucentric goniometers, and the availability of highly sensitive CCD cameras, it became possible to develop and implement automated image-acquisition procedures, which operate in low-dose conditions, thereby keeping the dose applied to the sample as low as possible. A major limitation in electron tomography is specimen thickness and the limited tilt-range, which allows the recording of only 70% of the necessary data in order to obtain a complete reconstruction. Nevertheless the information provided in the cryo-electron tomograms is sufficient to recognize distinct macromolecular complexes. The macromolecules are embedded in their natural environment and their density distribution is not affected by staining reactions, which tend to produce aggregations, and therefore compromise their molecular interpretation.

X-ray crystallography

X-ray crystallography provides the most powerful technique for the determination of structures of biological macromolecules. There is virtually no size limit for targets to be investigated, as long as they fold into a defined three-dimensional structure. The method is equally applicable for routine applications, for instance in the context of structural genomics projects, and for challenging multi component complexes. Advanced protocols have been developed for routine structure determination of integral membrane proteins as well as for complexes with partially unfolded protein components.

X-ray crystallography has enormously advanced during recent years because of major efforts to provide state-of-the-art data acquisition facilities at 3rd generation synchrotron radiation facilities, accompanied by developments and implementations of highly automated pipelines in terms of hardware and software provision. These efforts need to continue to ensure competitiveness. Because of the advances in automation and user-friendliness, translating into remote forms of execution of experiments, we are observing a decrease of a profound knowledge about the underlying theoretical foundations of this technique. It will be essential to keep a critical mass of experts in the field with the ability for further methods and technology developments, as well as to provide advanced methods-oriented training opportunities for young researchers.

At a time where an emerging key objective by the scientific community is to unravel functional/structural relations within entire biological systems, there is an increasing and urgent need to confirm and to validate high resolution molecular data by complementary *in vitro* and *in vivo* imaging methods as well as by functional assays. While the first category (presented by electron microscopy, electron tomography, small angle X-ray scattering, for instance) generally does not require fundamentally different sample preparation facilities, the application of *in vivo* validation methods generally requires access to cell culture facilities and specific know-how on respective applications.

Probably the most serious caveat of X-ray crystallography is its limitation to static structures, which is natural because of the requirement for fixed, diffracting crystal lattices. Therefore, in order to capture the dynamics of a given structure as a function of external parameters of a given biological system, it appears to be essential to combine X-ray crystallography either with experimental or computational structural biology methods to explore dynamic processes. At the molecular level, NMR spectroscopy provides a leading complement, as long as the method is feasible for a given biological system. At the cellular and/or whole-organism level, state-of-the-art complements are provided by life-video imaging methods.

Concluding Remarks

From the analysis of the applications of the techniques addressed during the workshop, and in particular of their possible integration with high-resolution NMR spectroscopy for the structural characterization of macromolecular complexes, a few conclusions can be drawn:

- SAXS and SANS can be integrated with NMR data with relative ease. Indeed, some pioneering protocols and programs have been developed. These techniques have the features and potentialities to attain high-throughput application in the structural characterization of complexes;

- WAXS is still in its infancy as far as application to biological macromolecules (and their adducts) is concerned. In principle, it can be well integrated with NMR spectroscopy, and can have a precious role thanks to its ability to provide information on conformational changes and aggregation modes of proteins and nucleic acids. Further methodological developments are needed to bring WAXS to a high-throughput level of application.
- X-ray crystallography and NMR are already employed as complementary methods, where structural data from crystallography are complemented by information concerning intermolecular interactions and dynamics obtained from NMR. Some examples of joint structural refinement against crystallographic and NMR data have been reported. Structures obtained from crystallography, but also from NMR or homology modelling, can be used to obtain structural models of protein complexes when combined with NMR data and computational approaches.
- Cryo EM and cryo electron tomography are complementary to NMR analysis. For example, structures and protein interfaces determined by NMR can be fitted into electron densities derived from EM. However, a direct combined application of the two methods is more difficult given the different size preferences.
- A direct combination of FRET data with NMR is less practical, since very different experimental conditions and, for FRET, extensive sample modifications are required. FRET is of potential interest because it provides long-range residue-residue distance information. Indeed, similar to crystallographic structures, NMR derived structural data may be validated by FRET experiments in vivo. Moreover, NMR can be used to characterize the intermolecular interactions of FRET probes, e.g. also in terms of structural perturbations induced. FRET however will probably remain a system-specific technique for at least some time.
- During the discussion at the CA closing meeting, other techniques were also considered and discussed with respect to their potentialities to provide complementary information. Focus was put on the possible exploitation of ELDOR (Electron DOuble Resonance) which can provide distance information for pairs of paramagnetic centers from 2 to 6 Å apart. ELDOR can only be applied to proteins with at least two native paramagnetic centers or to specifically tagged proteins. The relevance and the potential impact of single molecule microscopy were also discussed.

In summary, the currently available techniques for the structural investigation of macromolecular complexes have been experiencing tremendous developments in the last few years. These advances allow scientists to tackle the study of more and more challenging systems, such as weak and/or transient protein interactions. These adducts are often hardly addressed with a single method but they can benefit of the integration of a variety of methods able to characterize different properties, scales and aspects of macromolecular complexes.

References

(1) Bernado, P.; Mylonas, E.; Petoukhov, M.V.; Blackledge, M.; Svergun, D.I. *J.Am.Chem.Soc.* **2007**, 129, 5656.

(2) Cammarata, M.; Levantino, M.; Schotte, F.; Anfinrud, P.A.; Ewald, F.; Choi, J.; Cupane, A.; Wulff, M.; Ihee, H. *Nat.Methods* **2008**, 5, 881.

(3) King, W.A.; Stone, D.B.; Timmins, P.A.; Narayanan, T.; von Brasch, A.A.; Mendelson, R.A.; Curmi, P.M. J.Mol.Biol **2005**, 345, 797.

Chapter 9

Perspectives and roles in the development of high magnetic fields in biomolecular NMR

Authors: Ivano Bertini[a], Claudio Luchinat[a], Roberta Pierattelli[a]
[a] Magnetic Resonance Center, University of Florence, Italy
Appendix 1 by: Gerhard Roth[b]
[b] Bruker Biospin, Rheinstetten, Germany
Appendix 2 by: Andrew J. Varney[c]
[c] Varian, Inc., Oxford, United Kingdom

Motivation for this document

This document constitutes an attempt to identify the scientific challenges that can be addressed using high magnetic fields, and the importance of further developing experimental capabilities, instrumentations and techniques in this sector of NMR. Advancement requires research efforts in areas such as magnet technology and probe head design as well as developments in software, computing and data analysis. However, it is widely recognized that the mechanical and electromagnetic properties of current superconducting materials pose a major impediment in the design and development of highly homogeneous magnets with a static magnetic field beyond 23 T.

The present document is not the first to address these issues. In the year 2000, a *Baseline Study* by the committee for "High Field NMR: *A New Millennium Resource*" (http://nmr.magnet.fsu.edu/resources/NMRstudy.pdf) was prepared to help the United States Office of Science and Technology Policy assess the potential use of high field NMR technology and to provide recommendations for future efforts of US research funding agencies in addressing these challenges. Subsequently, in 2003, the National Research Council's Committee on "Opportunities in High Magnetic Field Science" (COHMAG) was established to determine the current state and future prospects of high magnetic field science and technology in the USA and to identify promising areas for research and development (http://books.nap.edu/catalog/11211.html). Both working groups have produced and made available position papers reviewing the current state of the art in magnet technology and high magnetic field applications. These documents are strongly supportive of research and massive investments in the field.

In Japan, no initiatives of this kind have been promoted by researchers, however, in the frame of the project "Protein 3000", the importance of developing a 1.3 GHz NMR spectrometer was stressed.

NMR-Life, the EC Coordination Action for NMR in life sciences, has therefore decided to organize a working group to pinpoint the European view on this subject.

The state of the art in 2006

NMR has established itself as a leading technique for the determination of the structure of biological macromolecules (proteins, nucleic acids) in solution. For those systems that are readily crystallized, X-ray diffraction offers highly efficient methods for structure determination. Indeed, to date 84% of the structures deposited in the PDB are obtained via X-ray crystallography. However, for non-crystallisable systems, (of limited dimensions, see below), NMR is the only technique available for structural studies. Furthermore, NMR is the method of choice for experimental characterization of the dynamic properties of proteins and macromolecular complexes, and the study of how these correlate with protein folding, molecular recognition, protein-protein interactions, enzymatic catalysis and other biological processes.

Challenges and opportunities with high fields

The contribution of NMR to structural biology is presently limited by both sensitivity and spectral resolution. A limiting factor is magnetic field strength. As of Summer 2006, commercial instrumentation suitable for high-resolution studies feature magnetic fields of at most 22.2 T (corresponding to 950 MHz proton Larmor frequency). If high-resolution NMR spectrometers with substantially higher fields become available, their impact on biomedical sciences will be dramatic.

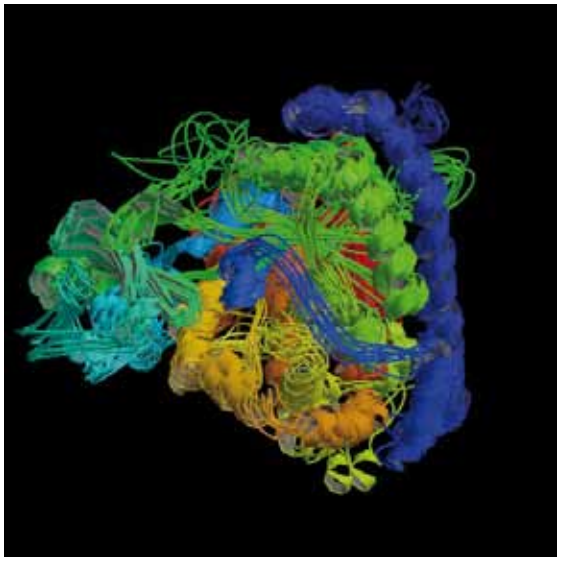


Figure 1: Solution global fold of the monomeric 723-residue (82-kDa) enzyme malate synthase G from Escherichia coli (Tugarinov V, Choy WY, Orekhov VY, Kay LE., Proc Natl. Acad. Sci U.S.A., 18, 102:622-7, 2005).

Higher magnetic fields lead to increased signal strength in proportion to the square of the magnetic induction, B_0^2, and a decrease in signal averaging time in proportion to B_0^3. Moreover, spectral resolution improves linearly with B_0 and this effect is multiplied by the dimensionality of the NMR spectrum, becoming even larger for 2D and 3D experiments. The increased sensitivity and resolution of such instruments would probably allow the use of NMR in the structural determination of proteins two-three times larger than the proteins whose structures can be determined today, as well as of larger molecular aggregates.

Until now, the largest protein whose (low resolution) structure has been resolved by NMR is malate synthase G from *Escherichia coli* (723 amino acids, with a molecular weight around 82 kDa), reported in Figure 1. However, current methods are limited to smaller systems, setting the generally accepted threshold for NMR to a molecular weight in the range of 30-35 kDa, or 250-300 amino acids. This limitation is severe, although not as much as commonly believed. Indeed, taking into account the human genome and neglecting protein-protein complexes, more than 50% of the approximately 40,000 proteins coded have less than 300 amino acids (Figure 2); if NMR is able to successfully afford the structure of proteins of around 600 amino acids, this technique can be used to study up to 80% of the total proteins of the human genome, opening avenues for the investigation of all the dynamic processes that are key for understanding biological functions.

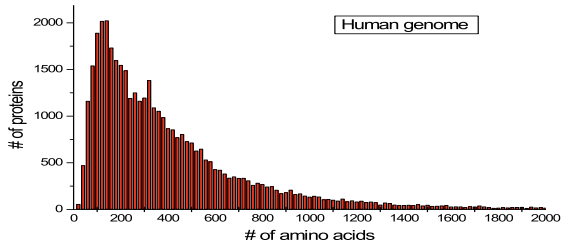


Figure 2: Number of proteins per number of amino acids.

Higher dimensionality experiments (4D or higher) are being developed in conjunction with computational tools for their automated analysis to expand the capabilities of NMR with proteins of larger sizes. The application of these methods would also significantly benefit from high magnetic fields, particularly due to improved resolution. Improved resolution would be contributive even when reduced dimensionality or projection reconstruction methods are adopted, because of decreased signal overlap and a resultant enhanced performance of the algorithms for analysis.

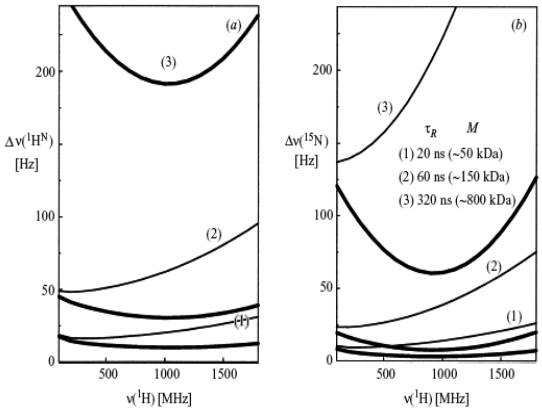


Figure 3: Frequency dependence from 100-1800 MHz of the full resonance line width at half height for amide groups in [15N, 1H]-SQ-TROSY (bold lines) and in [15N, 1H]-HSQC (thin lines) experiments calculated for three correlation times of τc = 20, 60 and 320 ns, which represent spherical proteins with molecular weights of 50000, 150000 and 800000 Mχ (Mr = molecular mass). (a) 1HN linewidth. (b) 15N linewidth. The following parameters were used: rHN = 1·04 Å, ΔσN = -155 p.p.m., ΔσH = 15 p.p.m., θN = 15º, θH = 10º, remote protons r(1HiN-Hia) = 0·22 nm, r(1HiN-Hi+1a) = 0·28 nm, r(1HiN-1Hi+1N) = 0·4 nm, r(1HiN-1Hi-1N) = 0·4 nm, r(1HiN-Hiβ2) = 0·3 nm, r(1HiN – Hiβ3) = 0·3 nm. Figure taken from K. Pervushin, Quarterly Reviews of Biophysics 33 , 2 (2000), pp. 161–197

So-called transverse relaxation-optimized spectroscopy (TROSY) has already offered an increase in spectral resolution at high field. Figure 3 shows that TROSY provides its maximum effect, for an amide moiety, at proton frequencies of around 1100 MHz, with significant attenuation of the transverse relaxation achieved simultaneously for the TROSY component of the ^{1}H and ^{15}N multiplets. For comparison purposes the linewidth behaviour for the same components in a standard ^{1}H-^{15}N correlation experiment are reported. Identical principles have been exploited on aromatic moieties and, more recently, on methyl and methylene groups, providing benefits similar to those for NHs signals in terms of resolution. It is expected that a similar trend will be observed for other nuclei with increasing magnetic field strength.

Another breakthrough of the past decade is the re-discovery and exploitation of the phenomenon of residual dipolar couplings (RDCs) experienced by dipole-dipole coupled nuclei (e.g. in a ^{1}H-^{15}N backbone amide moiety) in the presence of orienting media (bicelles, phages, stretched gels, etc.) in a magnetic field. RDCs provide long range restraints, which are particularly precious because they complement the short range information provided by NOEs, and permit a much more accurate determination, for instance, of the reciprocal orientation of secondary structure elements or subdomains even when they are very far apart. RDCs scale with the square of the magnetic field. The availability of, e.g. a 1.3 GHz instrument would more than double the magnitude of the RDC values that are currently obtained at 900 MHz, thereby sizably increasing their usefulness.

Finally, increasingly high fields, together with increased sensitivity, will allow a better exploitation of the *protonless* experiments that are already becoming useful at current fields. In fact, line broadening at very high field is mainly due to chemical shift anisotropy (CSA). Because line broadening scales with the square of the magnetogyric ratio, line broadening on carbons and nitrogens would be 1/16 and 1/100, respectively, of that of the protons. Especially for carbons with modest CSA, such as those in CH and CH_2 groups, *protonless* experiments may benefit dramatically by higher fields. Even the TROSY effect on CH coupling should improve, although the theoretical optimum is still beyond reach of any developments that may occur in the near future.

The impact on solid-state NMR of proteins may be even more dramatic. As described above, an increase of the magnetic field induction improves sensitivity and resolution in solution NMR. On the contrary, it can lead also to line broadening due to chemical shift anisotropy (CSA) effects, which increase quadratically with the field. CSA broadening depends on the correlation time of molecular diffusion. Its contribution therefore increases linearly with the molecular weight and becomes a problem for large biological systems. Indeed, the TROSY and related experiments are based on clever ways of reducing or cancelling out CSA broadening. In solid-state NMR experiments, which are not influenced by molecular diffusion, the linewidths of protein signals do not depend on the molecular weight. With the continuous progress in probe design, and higher and higher magic angle speeds achieved, direct proton observation may become routinely feasible. When very high magnetic fields are available, solid state NMR can become a method of choice for structure determination of large proteins.

Finally, NMR has proved to be one of the most powerful techniques for the study of biofluids and the only one capable of studying intact tissues, producing comprehensive sets of spectra that can be interpreted in terms of metabolites of sane or pathological organisms. With current high field spectrometer technology, standard analysis of NMR spectra does not provide an efficient means of interpreting biofluid spectra, because at current resolution the number of metabolite signals is too high. Therefore, automatic data reduction and chemometric approaches have been developed to enable efficient mining and extraction of information from large spectral databases. Unfortunately, information may be fatally lost in the process. The availability of higher field would improve significantly our ability to interpret these spectra and progress in the use of NMR for metabolomics. Similarly, all the uses of NMR for quality control and analysis will greatly benefit from the gain in resolution obtainable with the use of high field. It should be stressed that once the highest fields are employed to expand the database of components in complex mixtures, routine experiment at lower fields will also be more informative. This notion is strategically important from the market point of view, as new NMR-based analytical techniques will spread only if the costs associated with the instrumentation are affordable.

Technological challenges for developing higher fields

High field magnets for high resolution NMR spectroscopy must comply with the most stringent demands on field-strength homogeneity and stability. Their design is strongly dependent on the availability of an appropriate superconducting wire. A key property of such a wire is its critical current density J_C, defined as critical current divided by cross sectional area, and a function of the temperature T and the magnetic field B expe-

rienced by the wire. Similarly, one can define a critical temperature T_C and a critical field B_C. If any of these parameters (J, T or B) exceeds its critical value there is a transition from the superconducting to the resistive state. The high current flowing in a resistive wire generates heat able to "quench" the superconductive state, causing the conversion of the entire energy stored in the magnetic field into heat and the boiling off of the cryogenic liquids (He and N_2).

In addition to the technical aspects of superconduction, advances in other aspects of magnet engineering are essential. A primary characteristic of high field magnets is the high mechanical stress developed in the windings, constraining the wire to be capable of withstanding high forces. Furthermore, the wire, as well as the dewar containing the coil, should comply with all the requirements of safety in the event of quench and other unexpected occurrences that may cause a partial to total loss of superconducting capabilities.

The growth of the field strength of NMR in the past 40 years has been almost linear, with each step requiring approximately 3-4 years to be accomplished. However, the transition from NbTi superconducting wire technology, that permits the construction of coils up to 9.4 T at 4.2 K, to Nb_3Sn technology, that allows 21.1 T at 2.2 K, required almost 10 years (Figure 4).

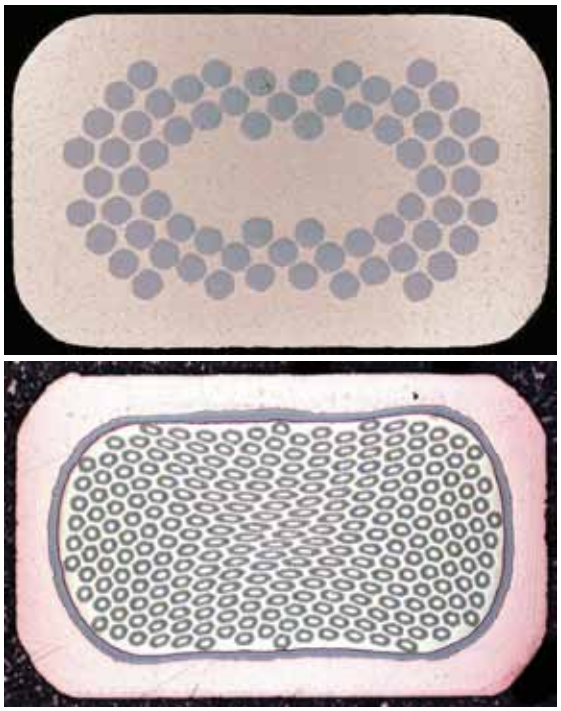


Figure 4: Filaments of (above) NbTi and (below) (NbTaTi)3Sn. Courtesy of Bruker Biospin

The development of multifilamentary wires, high temperature heat treatment, and persistent joints required great efforts and eventually lead to stable 22.2 T (950 MHz) magnet systems. In Figure 5 the dependence of the current density on the magnetic field strength is reported. It can seen that, generally, the current density decreases with increasing field, with a steep drop-off around 23 T. In the internal Nb_3Sn category, several processes can be used to produce wires. One such method uses Nb rod extrusions (Restacked Rod Process, RRP) while a second approach, the Modified Jelly Roll (MJR) method, employs coiled expanded Nb mesh. Both methods can be used to produce conductors having very high J_C values at a given field (Figure 6). The proprietary RRP treatment of the Nb_3Sn composite resulted in a 3-fold increase in critical current density at 20 T, making it possible to stabilize it at 23 T. However, above 23 T the current carrying capability degrades so severely that this field strength represents an upper limit for the Nb_3Sn composite.

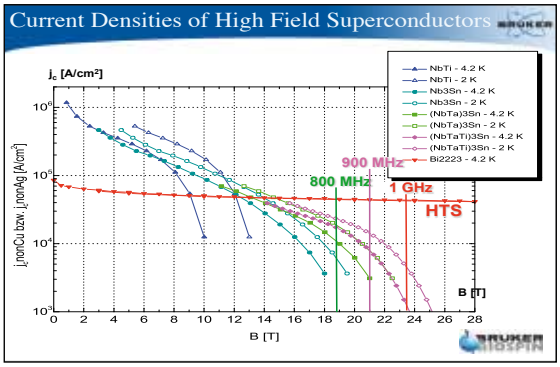


Figure 5: Dependence of current density on magnetic field strength. Blue: NbTi, dark green: Nb3Sn, light green: (NbTa)3Sn and purple (NbTaTi)3Sn. Open symbols give the current density at 2 K (Courtesy of Bruker Biospin).

Other promising composites, such as Nb_3Al, represent incremental improvements over Nb_3Sn and should have the potential to permit the development of 25 T magnet systems. However, the projected investment for a truly operative product is extremely high for a relatively small gain.

Indeed, beyond 23 T, new technologies are required. The current J_C of many high temperature superconductors (HTS) does not decrease significantly even at field strengths as high as 30 T and, therefore, magnet development groups are looking to this technology for the development of higher fields. The technology of HTS superconductors is developing rapidly, but a number of difficulties, such as heat resistance, the anisotropy of Jc, and others must still be overcome before a usable high field NMR magnet is constructed. These issues are being addressed in a

number of candidate HTS conductors including $Bi_2Sr_2Ca_2Cu_3O_x$ multifilamentary tape (Bi-2223), multifilamentary $Bi_2Sr_2CaCu_2O_x$ tape (Bi-2212) and YBa_2Cu_3O tape (YBCO). These are the most promising materials, according to magnets manufacturers.

At the National High Magnetic Field laboratory of Tallahassee (Florida, USA), the magnet R&D group is following a different strategy to achieve stable, very high magnetic fields. The approach is based on the development of a hybrid composed of an outer superconducting magnet and an inner resistive magnet. The unique feature to this design is that the resistive and superconducting components will be run in series so that the large inductance of the superconducting magnet will dampen the field fluctuations of the resistive component, increasing considerably, with respect to previous attempts, performance in terms of stability and homogeneity. The availability of HTS would also be highly beneficial for this hybrid approach.

To achieve the full potential of >22 T magnets, ancillary equipment needs to be developed. In particular, NMR probes circuitries that permit double- and triple-resonance experiments at ^{1}H NMR frequencies of 1.0 GHz and above will have to be designed and incorporated into the probes, including those with high-speed MAS capabilities. Already the construction of probes working at 21.1 T poses several problems in terms of interaction of the resonating coil with the stationary short radiowaves. Even more challenging is the design of cryogenically cooled probes, where, in addition to the probe, the components of the preamplifiers must also work at extreme temperature conditions.

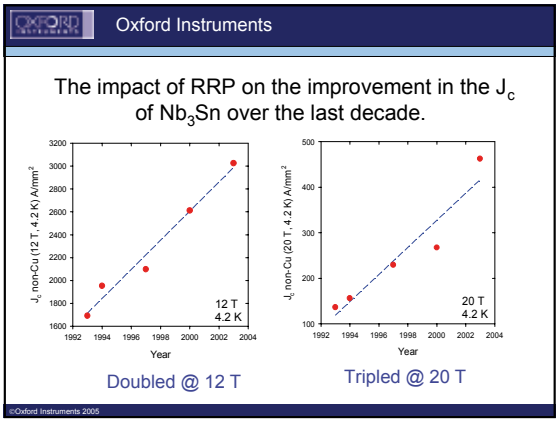


Figure 6: Impact of the Restacked Rod Process on the improvement in the Jc of Nb3Sn over the last decade (courtesy of Oxford Instruments).

Finally, radio-frequency power amplifiers that produce 1000 W pulses at proton NMR frequencies of 1 GHz and more must be developed, as the currently available technology is not capable of providing performance at the requested level.

Although outside the scope of this document, it is worth mentioning that any significant technological advancement in high-resolution magnets is expected to have an impact also on the design of magnets for MRI, either by increasing potency or by decreasing the stray fields, which currently represent non-negligible practical problems. From an economical point of view, the MRI market is much larger than the market for high resolution, very high field NMR spectrometers.

The role of infrastructures

The operation of very high field magnets with continued efficiency and quality of results for an extended period is not trivial. It requires dedicated personnel with adequate technical skills bolstered by a range of support services. Complementary to these technical aspects, it should be stressed that the development of a new technology requires extensive testing by a range of (potential) users in order to both identify possible problems and to propose new solutions or further improvements. Such testing should be accomplished either before or immediately upon release of the new technology on the market to achieve maximal benefits.

Research Infrastructures are optimal locations to drive the development of innovative high field equipment such as that described in this document, because they constitute centres capable of combining top-level technical and scientific expertise within individual countries or even at the continental level. Research Infrastructures in the field of bio-NMR are typically equipped with an array of spectrometers in the 10-20 T range, and thus already employ the technical staff and have access to the support services required to operate very high field magnets.

From a scientific point of view, the combination of infrastructure staff and the pool of its users provides not only a large and variegated ensemble of expertises to challenge any new spectrometer's performance but is also well suited to the identification of innovative scientific problems and approaches for maximal exploitation of the newly available technology.

The role of Research Infrastructures in the development of high field magnets suitable for applications in bio-NMR clearly cannot be that of addressing the technological problems involved in the construction of the magnet itself. Research infrastructures constitute an opportunity to provide solid scientific background for the innovative applications accompanying the release of new instrumentation as well as to provide access to a wide audience of potential users for the testing of the instrumentation.

More importantly, the Infrastructures must contribute significantly to the development of advanced experiments made possible by the high magnetic fields, from design to validation on a variety of experimental systems and conditions. In addition, Research Infrastructures should be involved in identifying limits and challenges of the newest technological developments, concurrently driving optimisation.

A final consideration regards the evolution of NMR manufacturing companies over the last decades. Pushing innovation may not always be rewarding from the budgetary point of view. It is to be expected, for instance, that the market for the > 1 GHz instruments will never be very large. In this light, the role of the Infrastructures in Europe, and of the national governments behind them, should also grow from that of simple customers of NMR manufacturers to that of "stakeholders" of NMR as a technique that continues to have the potential to produce breakthroughs in our understanding of the mechanisms of Life. In this sense, investments on the side of the EC and of national European governments are needed, not only to provide money to purchase the new NMR instruments in the companies' catalogues, but also to co-finance the joint development of new instruments between companies and infrastructures.

Conclusions and recommendations

The potential of NMR spectroscopy in the field of Life Sciences is still not fully realized. At present, NMR is a thoroughly validated tool for structural determination of soluble proteins with small to medium-range molecular size. Furthermore, NMR is uniquely suited for the study of protein dynamics and of intermolecular interactions (protein-protein, protein-nucleic acid, protein-ligand). NMR is rapidly progressing towards becoming the tool of choice for the structural investigation of membrane-bound proteins and other immobilized systems. Each of these applications would benefit significantly from a leap forward in magnet field strength, for different technical reasons. Among the main advances that would be achieved are the inclusions of proteins up to 600 amino acids long in solution NMR structural studies (80% coverage of human proteome), an improvement in solution structure determination throughput and accuracy, and a widened range of applications for the investigation of membrane proteins.

The technical challenges posed by the construction of the magnets required for very high field spectrometers devoted to bio-NMR are so great that manufacturers are unlikely to undertake their construction without at least some direct support from customers. Thus, it is likely necessary that the largest Research Infrastructures and manufactures join their efforts to obtain support from funding agencies. This partnership will permit focused efforts towards real breakthroughs for bio-NMR, with follow-ups in all fields where high magnetic fields are beneficial.

Acknowledgements

Discussions with R. Boelens[a], J. Breslin[b], T.A. Cross[c], N. Dajada[d], J. Gilbert[e], C. Griesinger[f], S. Hayashi[g], T. Keller[h], S. Pittard[d], G. Roth[h], B. Scanlon[e], H. Schwalbe[i], V. Sklenar[j], A. Twin[e], and A. Varney[d] are gratefully acknowledged.

[a] Bijvoet Center for Biomolecular Research, University of Utrecht, The Netherlands
[b] Varian Ltd., Oxford, United Kingdom
[c] National High Magnetic Field Laboratory, Tallahassee, Florida
[d] Magnex Scientific Ltd., Yarnton, United Kingdom
[e] Oxford Instruments, Abingdon, United Kingdom
[f] Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[g] JASTEC Inc., Tokyo, Japan
[h] Bruker Biospin, Rheinstetten, Germany
[i] Biological NMR Group, Johann Wolfgang Goethe University, Frankfurt
[j] National Center for Biomolecular Research, Masaryk University, Brno, Czech Republic

Appendix 1
New developments in magnet technology, 2006-2008

Over the past two years considerable new developments have taken place in both superconducting wire technology and magnet technology itself. For the classical bronze route Nb_3Sn wires, which exhibit well-proven NMR capability, significant progress towards higher current densities was achieved. With these new wires it is possible to improve magnet design with respect to the highest possible field and, in addition, it is possible to reach new levels of compactness with respect to the physical size of the magnet as well as to the size of the stray field.

With these new wires it was possible to build existing highest field magnets, such as the 900 MHz magnet (21.1 T) in an actively shielded version, reducing the stray field in volume by more than a factor of 10 and reducing the size of the 5 G footprint to approx. one quarter of the size of a non-shielded magnet (Fig. 1). The physical size of the 900 MHz magnet remained unchanged.

From this magnet also a new 850 MHz US^2 WB magnet with 89 mm room temperature bore could be derived. This is the highest field commercially available WB magnet, which in addition is actively shielded and which allows to place further instrumentation relatively close to the magnet. In the meantime,

four of these highest field WB magnets have been delivered to leading research labs in Europe.

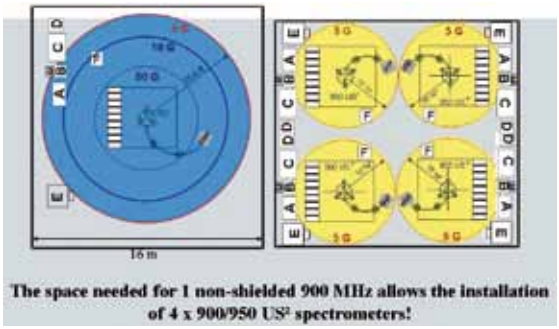


Fig. 1: Space requirements for 900 / 950 MHz magnets

With the new technology it was even possible to expand the maximum field strength to 950 MHz (22.3 T), again in a shielded version and with no increase in stray field and physical size of the magnet compared to 900 MHz.

The new high current density wires also opened the road to minimizing the size of existing magnets, especially for the already existing actively shielded 800 MHz magnet, which was as big as the new 900/950 MHz magnets. The new actively shielded 800 MHz magnet is a dramatically smaller version of an 800 with absolutely no compromise in performance, but with a radial 5 G line now at 1.5 m (instead of 6.1 m non-shielded), having essentially the same footprint as a former non-shielded 300 MHz magnet.

For this magnet, which fits into a single story lab, the stray field does not define the space requirement anymore, but the space requirement is dominated by the additional components, which come with the complete spectrometer.

Fig. 2: New 800 UltraShield Plus Magnet

Fig. 2 shows the new 800 USPlus magnet and Fig. 3 shows the reduction in space given with such a system, which now allows installation of 8 800 USPlus magnets in the same area as the former non-shielded 800, with all necessary equipment, including a cryoprobe.

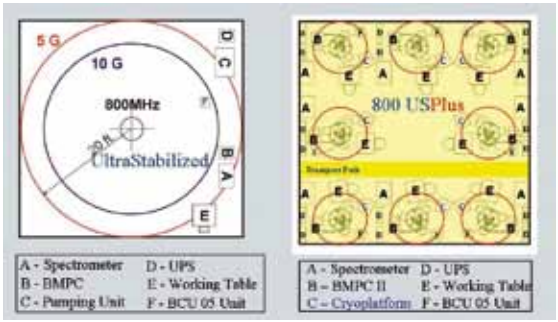


Fig. 3: 800 USPlus – minimum space requirements

The compactness of the magnet facilitates the use of the spectrometer within the lab dramatically and it especially allows to develop new experiments, where additional equipment needs to be positioned close to the magnet and which cannot tolerate the magnet stray field such as for example in Metabonomics.

The most important aspect however with these new improved Nb_3Sn wires is the fact that the properties of these wires could be stabilized in the manufacturing process in a way that allows achievement of highest current densities, even at 23.5 T. This field is very close to the region of rapid decline of current densities and in the past small production variations used to lead to unacceptable losses in current density, not allowing to meet the necessary wire specifications for 23.5 T any more.

This aspect could not be controlled very well in the past, but today properties are stable enough at this high field, so that the design of a 1 GHz magnet, based on classical Nb_3Sn wires, became feasible. This magnet has been constructed over the past two years and is now in the stage of final testing (Fig. 4).

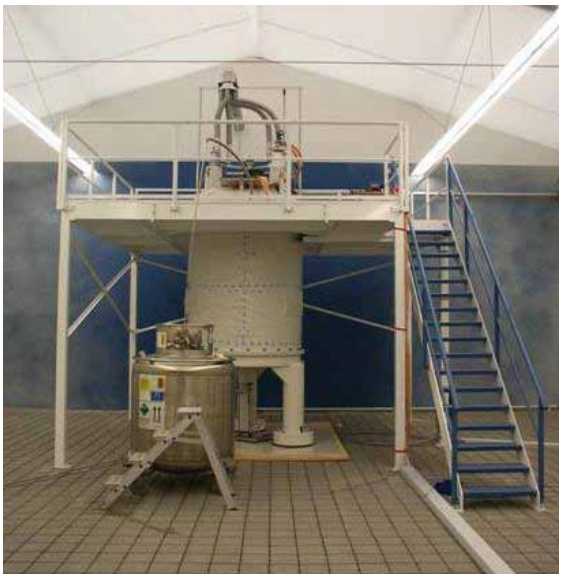


Fig. 4: First 1 GHz Test Site

This magnet reaches a number of upper limits in magnet design and in manufacturing capabilities of wire manufacturers. The mere size is so big that it reaches the limit of the maximum length of Nb_3Sn conductors that can be manufactured in one piece. For this reason, this magnet is non-shielded, as for a shielded version the field generating coil must be even larger, so that a shielded 1 GHz magnet is currently out of reach.

The next step towards a shielded 1 GHz magnet and to frequencies even higher than 1 GHz will only be possible with new superconducting materials, which have significantly higher current densities than Nb_3Sn at very high fields. Such most promising conductors are the high temperature superconductors (HTS) of which two types are in development:

BSCCO (Bismuth, Strontium, Calcium, Copper Oxide), the so-called "first-generation-conductor", has been under development for more than 10 years and is in use today for several low field applications. However, the development of this conductor type did not proceed as expected and as needed for the use in high field magnets. Main disadvantages are mechanical weakness and an insufficient current density.

The second conductor **YBCO** (Yttrium, Barium, Copper Oxide) or "second-generation-conductor" from today's view is the most promising conductor for highest field magnets. However, this conductor currently is at a too early stage of development to be actually used in an NMR magnet. It has a completely different wire architecture (Fig. 5).

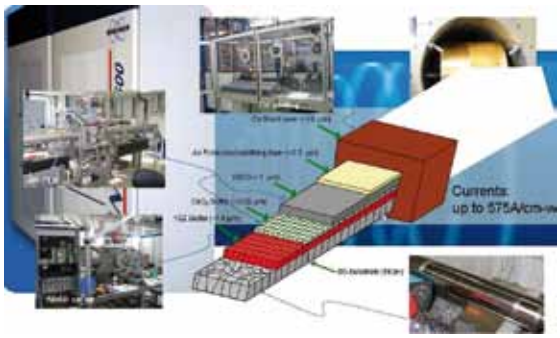


Fig. 5: YBCO Coated Conductors Production Route

The superconductor consists of a very thin superconducting layer (typically 1 µm), which is evaporated on the surface of a buffer layer and the buffer layers are deposited on a stainless-steel substrate. With this architecture the conductor simultaneously fulfils the most important requirements for high field magnets: very high current density in even a very thin layer and mechanical strength because of the stainless steel substrate. To protect the conductor, it is covered with a thin gold layer and – for quench protection – it is surrounded with a thick Copper layer.

While all preconditions to reach a technically feasible high field superconductor are promising, the technical challenges to develop a reliable production method, which can generate this conductor in sufficiently long lengths (~ 2,000 m), are high.

EHTS most recently was successful in manufacturing a maximum length of 100 m (Fig. 6), which shows a quite consistent high current of 250 A at 77 K in a 4 mm wide tape over the whole length.

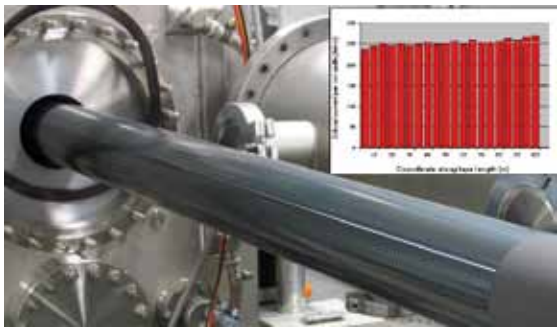


Fig. 6: 100 m tape coated with YBCO

A closer look however shows, that at the beginning of the tape and at the end of the tape the current deviates from its average value and it is a big challenge to maintain constant production conditions over a long time for a tape length of 2,000 m. However, constant tape quality over long lengths is not the only challenge faced in the design of future very high field magnets.

As this new conductor is only available as a tape, a new solenoid winding technology for tape conductors, needs to be developed as well.

The superconductor itself exhibits strong anisotropic properties, which need to be overcome, given the three-dimensional distribution of the magnetic field within the coil windings. The long-term stability (years) of this new conductor material needs to be proven and the method for reliable quench protection for this conductor needs to be developed (see Fig. 7).

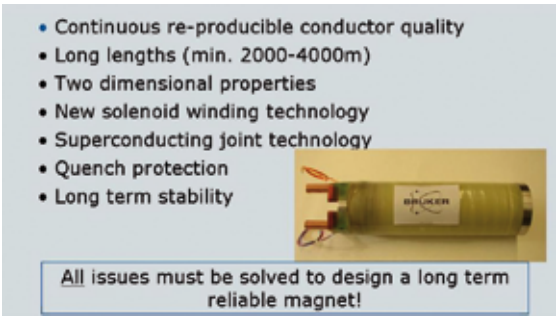


Fig. 7: HTS Challenges

Most challenging will be the development of a truly superconducting jointing technology, which combines the classical metallic superconductors (not oxygen tolerant) with the new high temperature superconductors (which require oxygen). Such a jointing technology is a precondition to achieve a truly persistent high resolution NMR magnet.

All the challenges need to be met successfully in order to design a long-term stable and reliable magnet.

The insert in Fig. 7 shows a solenoid test coil, which has successfully been wound and tested with "first-generation" taped conductor. Coil technology development will continue as soon as sufficient lengths of "second-generation" conductor will become available.

In summary, wire technology and magnet technology have taken a big step towards higher fields over the past two years. This allows the design and construction of a 1 GHz magnet based on Nb_3Sn conductors. A further step towards even higher fields will require that the promising high temperature superconductors can be manufactured in long lengths and that the numerous technical challenges associated with these conductors can be resolved successfully.

Appendix 2
The latest developments in high-field magnet technology

For a period of almost forty years the magnetic field strength of NMR magnets has increased approximately linearly with time up to 920 MHz in 2001. Above 500 MHz, the incremental increases in field up to that point have been enabled by improvements in bronze route Nb_3Sn wire (including higher Sn content and the use of additives such as Ti and Ta) and associated magnet technology developments. As the ultimate performance limits of bronze route Nb_3Sn wire are being approached, the rate of increase of field has been slowing.

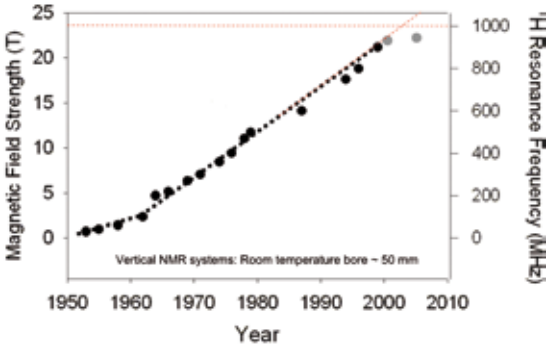


Figure 1: Highest magnetic field as a function of time for persistent narrow-bore NMR magnets. (Graph follows figures from Rooney et al. and Freeman [1] with additional points at 920 MHz [2] and 950 MHz [3]).

Recently, however, Nb_3Sn wire constructed via a different route has become suitable for high field NMR magnets. This internal tin technology, known as RRP™ (Restacked Rod Process) was used on the first successful 950 MHz NMR magnet installed in late 2005 at the University of Oxford [3]. The magnet in question was otherwise identical in geometry to previous 900 MHz magnets using older-style Nb_3Sn conductors. Since then, this step up in wire technology has enabled the introduction of new PremiumCOMPACT™ magnets at 600 MHz and 800 MHz.

Figure 2 shows the engineering critical current density, J_E – which is a parameter expressing the superconducting performance of real wire for its entire cross-section – for a variety of state-of-the-art Nb_3Sn wire types. The graph clearly shows that the RRP internal tin wire performance is at least double that of bronze route Nb_3Sn wire at magnetic fields of 800 MHz and above. A similar improvement is seen at lower fields. This means that conductor of a smaller cross-section can be used to achieve the same performance, thereby resulting in smaller Nb_3Sn coils and allowing the whole magnet size to be reduced.

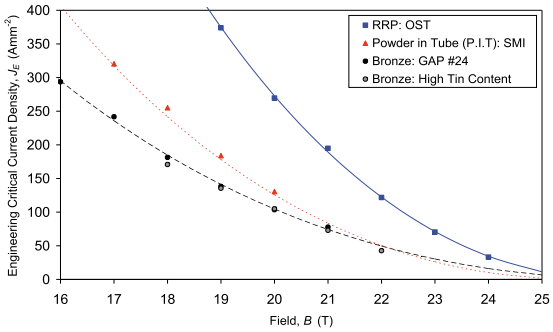


Figure 2: Performance of state-of-the-art Nb_3Sn wire at 4.2 K. (Data is from a 2008 paper by Flükiger et al.[4], except high tin content bronze[5])

While the new RRP conductor is the enabler, the details of the magnet design are also critical. Incorporation of the conductor has required optimisation of the superconducting joints, detailed consideration and testing of the mechanical behaviour of the wire in real magnets in order to use it to its full potential, analysis and testing of the quench protection, and other such details. These are all issues which have had to be addressed in the last few years in the development of the new PremiumCOMPACT magnet range.

The result is an effective step down in size for a magnet built to the same design criteria in other respects as the previous generation of magnets. For instance, the new 600 MHz PremiumCOMPACT magnet is the same size and weight and has the same fringe field footprint as the standard Premium Shielded 500 MHz magnet. Figure 3 shows the significant reduction in footprint of the 600 MHz PremiumCOMPACT magnet compared to the 600 MHz Premium Shielded magnet.

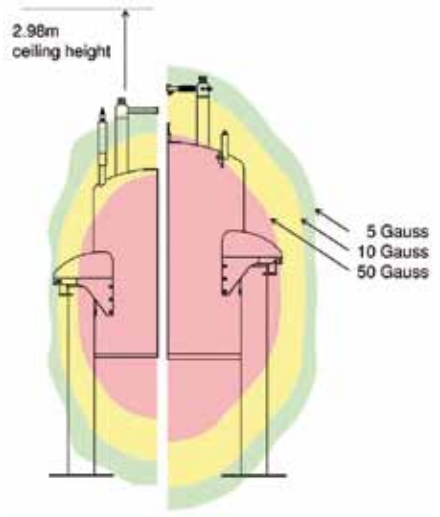


Figure 3: The improvement in footprint for the PremiumCOMPACT 600 MHz magnet (left) compared to the Premium Shielded 600 MHz magnet (right)

The gains at higher field are perhaps even more dramatic. The RRP wire is incorporated into the 4.2K PremiumCOMPACT 800 MHz magnet (see Figure 4). The technology enables a reduction in size and fringe field – to the point where the radial 5 gauss position (at 1.45 m) is within that of an old unshielded 200 MHz magnet. The reduction in physical size means that the magnet can be installed in a little over 3.5 metres ceiling height, with the 5 gauss fringe field not impinging into the storey above. In addition, this has been achieved while operating the magnet at 4.2 K rather than using a reduced temperature helium bath with the associated requirement for a pumping station.

Figure 4: New 4.2K PremiumCOMPACT 800 MHz magnet

In Figure 5 wire performance data for different types of Nb_3Sn wire is shown at a temperature of 1.8 K rather than 4.2 K. As can be seen, by reducing the operating temperature of the magnet to around 2 K, the performance of the wire can be further enhanced and hence it can be used to generate higher magnetic fields. This is just what has been done for most NMR magnets operating at 800 MHz and above using traditional bronze route Nb_3Sn wire. The RRP, however, achieves an engineering critical current density, J_E, at 1 GHz (23.5 T) comparable to the best published data for bronze route Nb_3Sn at 900 MHz (21.1 T). A more detailed analysis shows that a 1 GHz magnet of a similar size to the previous generations of 900 MHz magnet can be constructed within the design parameters successfully demonstrated to work on the new 4.2K PremiumCOMPACT magnets at lower fields (i.e. those up to 800 MHz). In fact, the wire performance data at 24.5 T look encouraging enough to suggest that NMR magnets approaching 1.05 GHz (24.65 T)

may be a possibility in the future without requiring any fundamentally new wire technology.

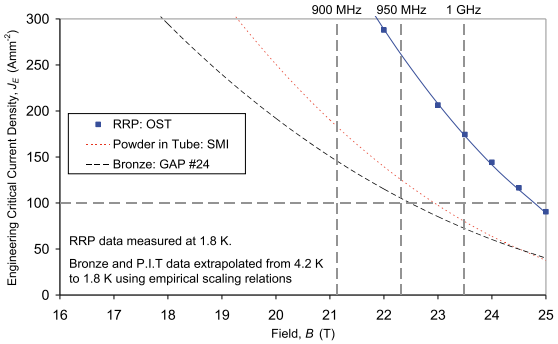


Figure 5: Performance of state-of-the-art Nb_3Sn wire at 1.8 K. (Data for RRP measured at 1.8 K[3]; other data scaled from 4.2 K measurements[4])

Despite the developments over recent years, it appears clear that to go *significantly* beyond 1 GHz in magnetic field will still require completely new wire types. Possible conductors for these higher fields include Nb_3Al, BSCCO (either $Bi_2Sr_2Ca_1Cu_2O_x$ referred to as Bi-2212 or $Bi_2Sr_2Ca_2Cu_3O_y$ referred to as Bi-2223) and YBCO ($YBa_2Cu_3O_{7-\delta}$). In this list the first is a conventional low temperature superconductor (LTS) whereas the latter materials are high temperature superconductors (HTS). For any of these conductors there are significant challenges which still need to be resolved before they can be of use in NMR magnets above 1 GHz.

One issue for incorporating solenoids wound from these conductors into persistent magnets is how they can be jointed to the other conductors in the magnet in a way which does not introduce too high an electrical resistance to make the magnet unsuitable for high resolution NMR work. From this perspective Nb_3Al is more hopeful, since joints with a resistance at the femto-ohm level between this material and other superconductors have been demonstrated [6]. Although low resistance joints at a level adequate for an NMR magnet have been demonstrated in BSCCO [7], jointing between the ceramic-based HTS and the metallic LTS remains an open issue. Another possibility is to minimize the resistance of the joints without striving to make them truly superconducting and instead to compensate for the drift by driving the field in another way. Studies in Japan in this context have suggested that this can be done at a level which could make a magnet above 1 GHz with a slightly resistive HTS coil usable for NMR [8].

Another issue in developing the highest field NMR magnets with HTS or Nb_3Al innermost coils relates to the technology for the rest of the magnet. Developing a narrow-bore 1 GHz magnet is a start, but to build a magnet suitable for incorporating a 5 Tesla inner coil to take the magnetic field up to 1.2 GHz (for example) requires the outer part of the magnet to increase in size significantly and this introduces additional challenges. This is one of the reasons why the increase in magnetic field for magnets at larger bore sizes lags behind those for narrow bore magnets. For instance, in 2005, while the record for a successful narrow-bore NMR magnet was 950 MHz, at a room temperature bore size of 260 mm, a field of 12 T (just over 500 MHz) was the state of the art. The last few years have seen significant progress in this area. By using Nb_3Sn coils rather than just NbTi and operating them in reduced temperature helium, examples of these large magnets first at 14.1 T (600 MHz) and then 16.4 T (700 MHz), as shown in Figure 6, have been successfully built, tested and installed.

Figure 6: 700 MHz (16.4 T) magnet with 260mm bore

Although Nb_3Al in principle has a higher upper critical field limit than Nb_3Sn, the conductor produced by the jelly-roll method has been unable to deliver the performance that might have been expected at high fields [9]. Instead an alternative 'rapid heating, quenching and transformation' (RHQT) method has

been developed which results in much better critical current performance, especially at high field. To date, however, the performance of this wire does not match that of RRP at fields of around 1 GHz and at 1.2 GHz it is unusable.

The quality and length demands for HTS material required for high field NMR magnet applications is higher than for other applications. Most of the wire produced for low field and high current applications has a residual resistance which makes it unusable for high field NMR applications. Producing long lengths of wire with sufficiently low resistance is in itself a challenge. In addition, most HTS wire is made in tape form. This presents other difficulties, such as an inherent critical current anisotropy with background magnetic field variations and the need to employ different coil construction techniques. In the case of silver-sheathed BSCCO it also has the tendency to develop voids which can break the wire as it is thermally cycled in the presence of liquid helium. Although this latter difficulty can be solved by sealing the tape, this has the disadvantage of reducing J_E.

Of all the materials mentioned, YBCO is the one with the highest critical current density at high fields. It is a type of superconductor referred to as 2G ('second generation') HTS. The construction of such wire is fundamentally different from that of the other superconductors, requiring complex processes to lay down thin layers on a substrate. The superconductor itself is typically a layer around 1 µm thick. The construction of the wire has the advantage that the substrate provides it with exceptional stress-handling capabilities. The disadvantages include the small fraction of superconductor, which dilutes the superconducting current carrying capacity (i.e. J_E), the difficulties in making long enough lengths with the properties required for this application and the other issues with tape already mentioned. While magnet coils have been tested in magnetic fields of up to nearly 27 T [10], these are a double-pancake construction which is unsuitable for high resolution NMR magnets for a number of reasons.

A conductor which avoids these concerns associated with HTS tapes is the Bi-2212 wire produced by OST [11]. This is true multi-filamentary wire rather than tape, designed to be braid-insulated so that it can be wound as multiple layers of a solenoid and reacted in very similar processes to those used for Nb_3Sn already in NMR magnets. The similarities of the processes as well as the avoidance of many of the issues with tape make it a natural choice in the development of the next generation of high field NMR magnets. In addition the J_E for this wire has been measured to be significantly higher than in a range of Bi-2223 tapes across a wide range of magnetic field strengths, attaining a value in excess of 250 A/mm^2 even at 45 T. Recently solenoids constructed from such wire have been tested as part of a complete magnet assembly generating a total of 22 T at 4.2 K [12]. The project is aimed at understanding the quench behaviour of the HTS solenoids incorporated with an LTS magnet in order to develop the quench protection system required for a future 25 to 30 T magnet. While there are still significant obstacles to be overcome, this work illustrates the promise that such an approach may hold for future high field magnets suitable for NMR.

References

(1) Rooney. W.D.; Johnson, G.; Li, X.; Cohen, E.R.; Kim, S-G.; Ugurbil, K.; Springer, C.S. *Magn. Reson. Med.* **2007**, *57*, 308-318. Freeman, R.; *Concepts in Magn. Reson.* **1999**, *11*, 61-70.

(2) Kiyoshi, T.; Sato, A.; Takeuchi, T.; Itoh, K.; Matsumoto, S.; Ozaki, O.; Fukushima, K.; Wada, H.; Yoshikawa, M.; Kamikado, T.; Ito, S.; Miki, T.; Hase, T.; Hamada, M.; Hayashi, S.; Kawate, Y.; Hirose, R. *IEEE Trans. Appl. Supercond.* **2002**, *12*, 711-714.

(3) Hong, S.; Field, M.B.; Parrel, J.A.; Zhang, Y. *IEEE Trans. Appl. Supercond.* **2006**, 16, 1146-1151.

(4) Flükiger, R.; Uglietti, D.; Senatore, C.; Buta, F. *Cryogenics* **2008**, *48*, 293-307.

(5) Miyazaki, T.; Miyatake, T.; Kato, H.; Zaitsu, K.; Murakami, Y.; Hase, T.; Hamada, M. *Cryogenics* **2008**, *48*, 341-346. Private communication.

(6) Fukuzaki, T.; Maeda, H.; Matsumoto, S.; Yokyama, S.; Kiyoshi, T. *IEEE Trans. Appl. Supercond.* **2007**, *17*, 1435-1437.

(7) Hase, T.; Shibutani, K.; Hayashi, S.; Shimada, M.; Ogawa, R. *Cryogenics* **1996**, *36*, 21-25. Hase, T., Shibutani, K.; Hayashi, S.; Shimada, M.; Ogawa, R.; Kawate, Y.; Kiyoshi, T.; Inoue, K. *Cryogenics* **1997**, *37*, 201-206.

(8) Takahashi, M.; Saito, Y.; Yanagisawa, Y.; Hobo, F.; Takao, T.; Nakagome, H.; Hosono, M.; Hamada, M.; Kiyoshi, T.; Yamazaki, T.; Maeda, H. poster presented at ENC 2008.

(9) Takeuchi, T.; Kikuchi, A.; Banno, N.; Kitaguchi, H.; Iijima, Y.; Tagawa, T.; Nakagawa, N.; Tsuchiya, K.; Mitsuda, C.; Koizumi N.; Okuno, K. *Cryogenics* **2008**, *48*, 371-380.

(10) Hazelton, D.W.; Selvamanickam, V.; Duval, D.M.; Larbaleister, D.C.; Markiewicz, W.D.; Weijers, H.W.; Holtz, R.L. *IEEE Trans. Appl. Supercond.* [submitted].

(11) Miao, H.; Marken, K.R.; Meinesz, M.; Czabaj, B.; Hong, S. *IEEE Trans. Appl. Supercond.* **2005**, *15*, 2554-2557.

(12) Melhem, Z.; Hong, S. Press release on Oxford Instruments web site, **2008**.

Chapter 10

Protein structure determination using chemical shifts

Authors: Michele Vendruscolo[a], Ad Bax[b], Rolf Boelens[c], Gijs van der Schot[c] and Roberta Pierattelli[d]

[a] Department of Chemistry, University of Cambridge, United Kingdom
[b] National Institutes of Health, Bethesda, USA
[c] Bijvoet Center for Biomolecular Research, University of Utrecht, The Netherlands
[d] Magnetic Resonance Center, University of Florence, Italy

Executive Summary

The extraordinary success of the application of NMR spectroscopy to protein structure determination is largely based on the exploitation of Nuclear Overhauser Effects (NOEs). The detection of NOE signals provides estimates of inter-proton distances from which the complete structures of proteins can be reconstructed [1]. NOE-based protein structure determination is usually, however, a rather time-consuming process since the collection and assignment of the NOE spectra is a lengthy procedure. This aspect is particularly relevant in the context of structural genomic initiatives, which have enormously increased the number of target proteins for structure determination. Moreover, structure elucidation via the current methods becomes increasingly complex as proteins become larger. In addition, NOE intensities are weak which limits NMR to the structural analysis of proteins at high concentrations. There is therefore a need for alternative methods.

The aim of this document is to discuss an emerging paradigm, according to which protein structures can be determined from the information provided by chemical shifts alone. Through recent advances it has now become possible to generate three-dimensional structures of proteins using backbone chemical shifts with an accuracy comparable to that achieved by more standard NMR approaches. The three computer packages that are currently available for obtaining these results are Cheshire [2], CS-Rosetta [3] and CS23D [4]. These methods in their present implementation are currently still restricted by protein size, as they have been proven to accurately solve the structures of proteins only up to about 130 amino acids. However, there are good prospects that the current chemical shift-based methods may become applicable to larger systems and can be used to investigate large molecular machines as well as especially with the continuous improvement of methods, either *ab initio* [5] or empirical [6;7], for predicting chemical shifts from structures.

In order to fully establish the current approaches it is necessary to increase our understanding of the fundamental relationship between chemical shifts and protein structures, as well as to further develop computational methods to reconstruct structures from chemical shifts. To promote advances in this direction, our primary recommendation is to systematically compare new structures determined from NOE measurements with those determined from chemical shifts using Cheshire, CS-Rosetta and CS23D, which are all publicly available and fully supported. Furthermore, exploration is needed as to whether chemical shift refinement can be combined with complimentary data from methods such as SAXS and can define molecular rearrangements as occur in regulatory systems.

Opportunities and Limitations of Current Chemical Shift Methods

General properties of chemical shifts

Among all the observables that can be measured by NMR spectroscopy, chemical shifts are those that can be determined most readily and with the greatest accuracy. The chemical shift of a nucleus is dependent on the electron density around the nucleus. The electron density is directly dependent on the type of nucleus and on the number, nature, and location of neighboring atoms. Therefore, chemical shifts offer an important source of local structural information. In addition, chemical shifts are dependent on a variety of other factors, including in particular ring currents and hydrogen bonds. Advances in computational methods and the appearance of large repositories of structural (PDB) and NMR (BMRB) information have made it possible to generate accurate structures of proteins, using solely chemical shifts as input [2-4].

Chemical shifts are primary parameters which have several advantages over other mostly secondary NMR observables: *i*) They are measured with large accuracy, high sensitivity and within full coverage, in the early stages of the NMR process; *ii*) They are often the only NMR parameters accessible for large molecular complexes or molecules in a cellular environment; *iii*) They can be measured for regions where other parameters cannot be measured (for example, NOEs cannot usually be measured in flexible regions); *iv*) Statistical methods for error analysis or for the detection of incorrect assignments are possible; *v*) Misassignments only give rise to small deviations since chemical shifts are primarily of a local nature. These advances make chemical shifts an inviting alternative to the NOE-based structure determination methods, and could, at least in principle, be applied to much larger macromolecules and complexes.

Chemical shifts and secondary structure determination

The first successes in using chemical shift to obtain structural information have been the discovery of a correlation between chemical shifts and backbone torsion angles [8;9], and the prediction of secondary structures by using the Chemical Shift Index (CSI) [10] or the PSSI methods [11]. With the rapid increase in the number of proteins in structural repositories such as the Protein Data Bank (PDB), more quantitative methods have been developed. For example, the TALOS method [12] enables the accurate prediction of torsion angles from chemical shifts. This program compares the chemical shift and the residue type of a three-residue fragment from the query protein with the chemical shift and residue type of all three-residue fragments in the protein database. The assumption made is that fragments with similar chemical shifts and residue type have similar torsion angles. TALOS predicts torsion angles for an average of 67% of the residues; 3% of the predictions are poor [12]. The predicted torsion angles can be used to cross-validate independently determined structures, or directly as restraints in structure refinement protocols.

Chemical shifts and tertiary structure determination in solution

Within the last two years three methods have been presented that enable the determination of native states of proteins using chemical shifts: Cheshire [2], CS-Rosetta [3] and CS23D [4]. These methods are based on similar strategies, which in essence involve three steps: (1) Generation of a library of fragments that are selected from structural databases on the basis of sequence and chemical shift similarity, (2) Assembly of these fragments into an ensemble of putative structures, (3) Refinement of these structures and choice of the optimal one. The last two steps in all cases involve the use of force fields and they can also include the secondary and tertiary structure information contained in the chemical shifts. For example in the Cheshire method, the SHIFTX method [6] is used to translate chemical shifts intro structural restraints that are used in the assembly of the fragments and in the subsequent refinement.

After the initial demonstration that chemical shifts can be used for determining the structures of proteins [2], Bax, Baker and coworkers have convincingly shown that this approach is capable of solving previously unknown structures and is useful in structural genomics initiatives [3]. Furthermore, Wishart and coworkers have recently introduced a method that can achieve these types of results with a very limited computational cost, at least when homologous structures are available [5].

Chemical shifts and tertiary structure determination in the solid state

Chemical shifts can also be used in solid state NMR. This technique has the considerable advantage over solution NMR and X-ray crystallography of not requiring soluble or crystalline samples, and thus can be used on powders, fibrils, frozen solutions, microcrystals, gels or proteoliposomes [13]. Furthermore, the samples used in solid state NMR are not limited by protein size or structural order [13]. A deepened understanding of chemical shift will therefore aid studies on for example amyloid proteins, molecular folding, ligand-membrane interactions and membrane proteins [13]. An initial demonstration of the possibility of determining native states of proteins in the solid state using chemical shifts has been recently provided, by the demonstration that the native structures of protein G, ubiquitin and SH3 can be determined at a resolution comparable to that of other methods [14;15].

Chemical shifts and protein-protein complex structure determination

For several years it has been recognized that chemical shifts can provide useful information about the structures of protein-protein complexes. In the HADDOCK method, chemical shift perturbations upon complex formation are translated into am-

biguous distance restraints (AIRs) [16]. This method has proven very effective and has become widely used. More recently, it has been proposed that chemical shifts can be used more directly for the determination of the structures of protein-protein complexes. In the CamDock method [17] the secondary and tertiary information provided by chemical shifts is translated into structural restraints, in a strategy that is very similar to that used for the structures of native states of globular proteins. In an initial application of this method, it has been shown that it is possible to determine in this way the structure of the E9-Im9 complex (PDB code 2K5X, Fig. 1). It was previously impossible to determine the structure of this complex by NMR methods because of its significant dynamics, which makes it extremely difficult to detect NOE signals, and of the considerable rearrangements upon complex formation, which extend up to 5 Å at the interface.

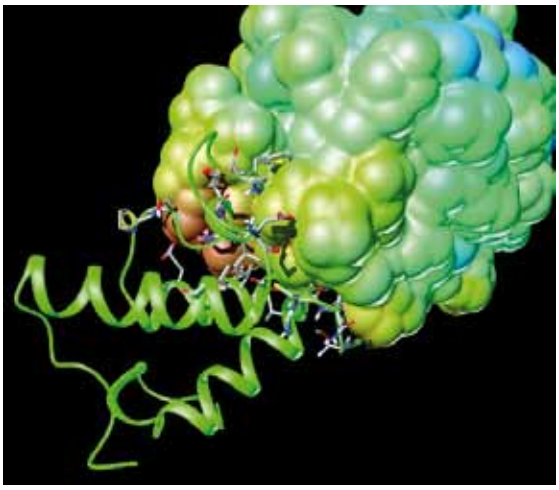


Figure 1: Structure of the E9-Im9 complex (PDB code 2K5X) determined using chemical shifts with the CamDock method [17].

Conclusions and Recommendations

There are great opportunities in the use of chemical shift for structure determination of proteins. Chemical shifts can already be used for the calculation of secondary structure [10-12], and in favourable cases also of tertiary [2-4] and quaternary [16;17] structures, both in solution and in the solid state. It will be possible to improve the accuracy and widen the scope of these methods by increasing our understanding of the relationship between chemical shifts and the structure and dynamics of proteins. NMR chemical shifts are the perfect parameters to be used in the structure elucidation of large molecular machines, since they can be measured at high accuracy, high sensitivity (being a primary parameter) and with considerable coverage. We therefore conclude this document with recommendations aimed at facilitating these strongly needed advances. Since the assignment of backbone chemical shifts is routine in essentially any NMR strategy for protein structure determination, we invite researchers to test the use of current publicly available methods (Cheshire, CS-Rosetta, CS23D) together with more standard procedures. These systematic efforts will generate a solid background upon which further developments will be built. Furthermore, researchers are invited to explore the possibility of using NMR chemical shifts in combination with other NMR parameters such as NOEs and RDCs and with complementary information such as that obtained from SAXS to define molecular rearrangements as occur in regulatory systems, to explore inclusion of the unique dynamical aspects and to extend the methodology to protein-protein and protein-nucleic acid complexes. A future in which NMR can also be used to elucidate larger molecular machinery is dawning.

Software

Cheshire: http://www-vendruscolo.ch.cam.ac.uk/software.html

CS Rosetta: http://spin.niddk.nih.gov/bax/software/CSROSETTA/

CS23D: http://www.cs23d.ca/

References

(1) Wuthrich, K. *NMR in Structural Biology*; World Scientific, Singapore: 1995.

(2) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl.Acad.Sci.U.S.A* **2007**, *104*, 9615-9620.

(3) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc.Natl.Acad.Sci.U.S.A* **2008**, *105*, 4685-4690.

(4) Wishart, D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. *Nucl.Acid Res.* **2008**, *36*, w496-w502.

(5) Vila, A. J.; Aramini, J. M.; Rossi, P.; Kuzin, A.; Su, M.; Seetharaman, J.; Xiao, R.; Tong, L.; Montelione, G. T.; Sheraga, H. A. *Proc. Natl.Acad.Sci.U.S.A* **2008**, *105*, 14389-14394.

(6) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J Biomol NMR* **2003**, *26*, 215-240.

(7) Shen, Y.; Bax, A. *J.Biomol.NMR* **2007**, *38*, 289-302.

(8) Spera, S.; Bax, A. *J.Am.Chem.Soc.* **1991**, *113*, 5490-5492.

(9) Kuszewski, J.; Qin, J.; Gronenborn, A. M.; Clore, G. M. *J.Magn. Reson.Ser.B* **1995**, *106*, 92-96.

(10) Wishart, D. S.; Sykes, B. D. *J.Biomol.NMR* **1994**, *4*, 171-180.

(11) Wang, Y. J.; O Jardetzky *Protein Sci* **2002**, *11*, 852-861.

(12) Cornilescu, G.; Delaglio, F.; Bax, A. *J Biomol NMR* **1999**, *13*, 289-302.

(13) Baldus, M. *J Biomol NMR* **2007**, *39*, 73-86.

(14) Robustelli, P.; Cavalli, A.; Vendruscolo, M. *Structure* **2008**, *16*, 1764-1769.

(15) Shen, Y., Vernon, R., Baker, D., and Bax, A. J Biomol NMR . 2009. (In Press)

(16) Dominguez, C.; Boelens, R.; Bonvin, A. M. *J.Am.Chem.Soc.* **2003**, *125*, 1731-1737.

(17) Montalvao, R. W.; Cavalli, A.; Salvatella, X.; Blundell, T. L.; Vendruscolo, M. *J Am Chem Soc.* **2008**, *130*, 15990-15996.