Background Information for the Keynote Lectures to be delivered at BRIDGES events in Thailand, December 13 – 17, 2004.

Structures of the Molecules of Life – Impact on Modern Biomedical Research

Kurt Wüthrich

Cecil H. and Ida M. Green Professor of Structural Biology

The Scripps Research Institute, La Jolla, CA, USA

and

Professor of Biophysics

Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

Synopsis

Our research area is structural biology and structural genomics. My research groups are specialized in the use of nuclear magnetic resonance (NMR) spectroscopy for studies of the molecular structures and functional interactions of proteins and nucleic acids. When compared to structure determination by X-ray crystallography, the NMR method is unique by the fact that atomic resolution structures of biological macromolecules can be determined in solution. The solution conditions can be adjusted such that they are very close to the physiological milieu in body fluids such as, for example, the blood, stomach fluid or saliva. The impact of NMR structure determination of proteins in solution on modern biological and biomedical research will be discussed, and I will further entertain some biographical data and issues of research organization in the natural sciences, based on my experience in organizing and supervising research teams in a European environment and in the USA.

Growing up with Natural Sciences – a Short Biography

My childhood was spent in the Kanton of Bern in Switzerland, a rural area of farmland, forests and rivers. I thus grew up in close contact with plants and animals, and natural science was then a special attraction for me in

high school. When I then moved on to the University of Bern in 1957, the number of faculty and the student classes were small, with 3 students majoring in physics and 7 students majoring in chemistry in 1957. I studied both chemistry and physics, and participated in an intense program of courses in mathematics. Only much later did I fully appreciate the extent to which this combination of undergraduate studies would provide an excellent foundation for my later scientific activities in an interdisciplinary area.

In the Spring of 1962 I moved from the University of Bern to the University of Basel, where I enrolled in the "Turn- und Sportlehrerkurs". In addition to about 25 weekly hours of intense physical exercise, these studies included premedical courses in human anatomy and physiology. Combined with experience gained from observations made on myself in the pursuit of competitive sports, this provided an additional dimension to my education. In the Fall of 1962 I also started to extend my training in chemistry at the University of Basel, with a Ph.D. thesis in inorganic chemistry with Prof. Silvio Fallab. The subject of my Ph.D. thesis was the catalytic activity of copper compounds in autoxidation reactions. This project led to my initial practical experience with magnetic resonance spectroscopy, using a state-of-the-art electron paramagnetic resonance (EPR) spectrometer available in the physics institute. Although the actual experiments were performed with low molecular weight metal complexes, the interest of the study was mainly focused on structure–function correlations in copper-containing metalloproteins. My formal University education was completed in March 1964, when I obtained both a Ph.D. degree in chemistry and the "Eidgenössisches Turn- und Sportlehrerdiplom". It is worth noting that the areas of our current research had not yet matured by 1964 and were not part of my University education: Nuclear magnetic resonance (NMR) spectroscopy was just being introduced as an analytical tool in chemistry, molecular biology was not yet established as an independent discipline, and the initial three-dimensional protein crystal structures at atomic resolution were just emerging.

After finishing my graduate studies I spent another year in Switzerland at the University of Basel, and in the spring of 1965 we moved to the USA for postdoctoral training with Prof. Robert E. Connick at the University of California, Berkeley. We used NMR spin relaxation measurements of ^{17}O, ^{2}H and ^{1}H in addition to EPR for studies of the hydration of metal ions and metal complexes. In 1967 I joined the Biophysics Department of Dr.

Robert G. Shulman at Bell Telephone Laboratories in Murray Hill, NJ, USA, which at the time was a world-leading place for research in physics and chemical physics. I was given responsibility for the maintenance of one of the first superconducting high resolution NMR spectrometers, which operated at a proton resonance frequency of 220 MHz, and started to use NMR for research on protein structure and function. In particular, using blood sampled from my arm in the first aid station, we prepared "hemoglobin (KW)", and within a few months we found entirely new avenues of deriving information on structure–function correlations from the NMR spectra of hemoglobin and other hemoproteins. Many years later, the special NMR-spectral features of hemoproteins were a great help in the development of the NMR method for three-dimensional protein structure determination. In 1969 I returned to Switzerland and joined the faculty of the Eidgenössische Technische Hochschule (ETH) in Zürich, where I still pursue my research and teaching as a Professor of Biophysics. Since 2001 I share my time between the ETH Zürich and The Scripps Research Institute (TSRI) in La Jolla, CA, USA.

Although scientific research has always been the main focus in my professional life, some other activities also represented major commitments of time and energy, in particular with functions in the International Union of Pure and Applied Biophysics (IUPAB) and the International Council of Scientific Unions (ICSU). From 1978–90 I served as Council Member, Secretary General and Vice President of IUPAB, and as a member of the "General Committee" and the "Standing Committee on the Free Circulation of Scientists" of ICSU. In connection with these positions I also got involved in the organization of numerous scientific meetings in developing countries. From today's perspective these extracurricular activities have been most valuable elements of my professional life, both for the scientific contacts and the personal friendships established with colleagues all around the world.

NMR in Structural Biology

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for three-dimensional structure determination of proteins and nucleic acids at atomic resolution, since the NMR data can be recorded in solution. Considering that body fluids such as blood, stomach liquid and saliva are protein solutions where these molecules perform their physiological functions, knowledge of the molecular structures in solution is

highly relevant. In the NMR experiments, solution conditions such as the temperature, pH and salt concentration can be adjusted so as to closely mimic a physiological fluid. Conversely, the solution conditions may also be changed to quite extreme non-physiological conditions, for example, for studies of protein denaturation. Furthermore, in addition to structure determination, NMR applications include investigations of dynamic features of the molecular structures, as well as studies of structural, thermodynamic and kinetic aspects of interactions between proteins or nucleic acids and other solution components, which may either be other macromolecules or low molecular weight ligands. Again, for these supplementary data it is of keen interest that they can be measured directly in solution.

An exciting feature of structural investigations in solution is that polypeptide chains can be studied even if they are only partially folded. Partially folded proteins are usually difficult to crystallize. If crystals are obtained, the chain segments that are disordered in solution will either be ordered by intermolecular contacts in the crystal lattice, or they will not be visible by diffraction methods. As a consequence, NMR has in many cases been the only method capable of providing structural information on partially folded polypeptides. Although a standard protocol for NMR structure determination provides only a static picture of unstructured chain segments, additional NMR experiments can provide information on the frequencies of the rate processes that mediate transitions between discrete states of the molecule in conformation space. Overall, the ability of the NMR technique to characterize macromolecular structures and their intermolecular interactions with high spatial as well as temporal resolution makes it a highly attractive tool for structural biology and structural genomics.

The Way to NMR Structures of Proteins

In the 1960s and 1970s, successful structural interpretations of NMR data invariably supplemented a previously known low resolution X-ray crystal structure of the same protein, which could rely primarily on measurements of so-called “chemical shifts”. In a *de novo* protein structure determination, however, there would be

inevitable ambiguities in the structural interpretation of chemical shifts. Different approaches were therefore called for, and eventually a NMR method for protein structure determination could be based on the following four principal elements:

1. Measurement of NOE upper distance limits as conformational constraints.

Nuclear Overhauser effects (NOE) are due to dipolar interactions between different nuclei. The intensity of the NOE is related to the product of the inverse sixth power of the internuclear distance and a correlation function, $f(\tau_c)$, which describes the modulation of the dipole–dipole coupling by stochastic rate processes, with an effective correlation time τ_c:

$$\text{NOE} \ \alpha \frac{1}{\langle r \rangle^6} \cdot f(\tau_c) \qquad [1]$$

Although the NOE is a common phenomenon for all combinations of closely spaced nuclear spins, NOEs between pairs of hydrogen atoms are of prime interest for structural studies of biological macromolecules. A ^{1}H–^{1}H NOE is related to the "through-space" distance between pairs of atoms that are either not at all linked by covalent bonds (intermolecular NOE), or that may be far apart in the amino acid sequence of a polypeptide chain.

Our studies showed that outstandingly favourable conditions for NOE distance measurements can be found in macromolecules, which have long effective correlation times for the modulation of dipole–dipole couplings, when compared to low-molecular-weight compounds, for which the condition of "extreme motional narrowing" applies. With proper selection of the duration of one of the experimental parameters, the so-called "mixing period", one can measure highly selective ^{1}H–^{1}H NOEs between distinct pairs of hydrogen atoms in proteins or nucleic acids in solution.

2. Sequence-specific resonance assignments

There are closely spaced pairs of hydrogen atoms in neighbouring residues of a polypeptide chain. These can be connected by the observation of "sequential NOEs". Suitable combinations of intraresidual $^{1}H–^{1}H$ connectivities established by scalar spin–spin couplings, and inter-residue connectivities established by sequential NOEs enable progressive resonance assignments while "walking along the polypeptide backbone". In other words, neighbouring amino acid residues can be connected by the intervening sequential NOE connectivities. Today, sequential resonance assignments for recombinant ^{13}C, ^{15}N-labeled proteins are mostly obtained with so-called "heteronuclear triple-resonance" experiments.

3. Two-dimensional (2D) NMR.

With the introduction of 2D NMR experiments, and subsequently 3D and 4D NMR experiments, NMR studies of biological macromolecules evolved from intellectually stimulating science to a practical approach for protein structure determination. When compared to "old-fashioned" one-dimensional (1D) NMR experiments, there are two crucial consequences of multi-dimensional NMR for studies of proteins. First, 2D ^{1}H NMR enables the recording of selective interactions between pairs of hydrogen atoms without selective irradiation of individual resonance lines. It thus enables a detailed analysis of the entire ^{1}H NMR spectrum of a protein, which contrasts with the 1D NMR situation of being limited to using only a small number of resolved lines. Second, the dispersion of the resonances along two or several frequency dimensions affords greatly improved separation of the individual peaks.

4. Structural interpretation of NOE distance constraints.

A polypeptide chain with 100 amino acid residues has a length of about 400 Å, whereas NOE-observable distances are shorter than about 5 Å. Observation of a NOE between a pair of hydrogen atoms with assigned chemical shift positions therefore enforces the formation of a ring-like structure. A successful structure determination generates three-dimensional arrangements of the polypeptide chain that simultaneously contain all the small and large circular structures imposed by the ensemble of all NOESY cross peaks.

For the calculation of complete three-dimensional protein structures from NMR data , it was quite clear from the outset that an input of quantitative NOE distance measurements would be difficult to obtain, since the observed NOEs depend on the proton–proton distance, *r*, as well as on the effective rotational correlation times, τ_c (Equation [1]). Since for each pair of hydrogen atoms the effective τ_c-value is governed not only by the overall rotational molecular tumbling (Brownian motions), which depends on the size and shape of the protein as well as on the viscosity of the solvent, but can also be affected by intramolecular motions, $f(\tau_c)$ may vary for different pairs of hydrogen atoms in a protein molecule. In view of this intrinsic limitation for efficient quantitative NOE distance measurements, we decided to use a constant value of the correlation function ($f(\tau_c)$ in Equation [1]) for all ^{1}H–^{1}H combinations in a protein, and to derive only upper limits on the ^{1}H–^{1}H distances from the NOE measurements. In practice, the input for a structure calculation then consists of allowed distance ranges, which are bounded by a NOE upper limit of 3.0 to 5.0 Å, depending on the intensity of the NOE, and a lower limit of 2.0 Å, which represents the sum of the van der Waals radii of the two NOE-connected hydrogen atoms. Although each individual entry in the input data thus has only limited precision, this procedure is robust and can conceptually account for the influence of intramolecular mobility in most of the situations that are commonly expected for the structured parts of globular proteins.

For the initial globular protein structure calculations from NMR data, we used a metric matrix distance geometry algorithm to search for molecular geometries that are consistent with the ensemble of all experimentally determined NOE distance constraints. Each such calculation ends with the minimization of an error function, and the residual error function value represents a straightforward measure for the success of having found a molecular geometry that satisfies the experimental input data. In view of the aforementioned distance-range format of the input, it is further of keen interest to evaluate the uniqueness of the calculated structure. To this end, the structure calculation is repeated with identical input data but different boundary conditions, and the uniqueness of the resulting NMR structure is judged from the tightness of the fit among the resulting ensemble of conformers. Typically, about 100 conformers are generated, and a sub-group of the 20 conformers with the smallest residual error function values is selected to represent the NMR structure of the protein. The average of the pairwise root-

mean-square distances (RMSD) calculated for this bundle of conformers is then taken as a measure for the precision of the structure determination. Today, metric matrix distance geometry algorithms have been replaced by molecular dynamics techniques for the calculation of protein structures from NMR data.

The NMR View of Protein Structures in Solution

A standard protocol for NMR structure determination of biological macromolecules that was first used successfully in 1984 includes the preparation of a homogeneous protein solution, the recording and handling of the NMR data sets, and the structural interpretation of the NMR data. In the meantime, this protocol has been used, with continuously improving experimental techniques, for over 3'000 NMR structure determinations of proteins and nucleic acids.

The static picture of a protein molecule obtained from the standard protocol for NMR structure determination typically shows variable precision of the structure determination along the polypeptide chain. Furthermore, when compared to crystal structures, increased disorder is observed toward the molecular surface. This pronounced surface disorder, which typically also involves the ends of the polypeptide chain, is in most instances the only significant difference between corresponding globular protein structures in single crystals and in solution. With the additional use of NMR spin relaxation measurements, one can distinguish between static disorder and dynamic disorder, with intramolecular motions on the nanosecond and sub-nanosecond time scale. Overall, quite independent of the dynamics issue, the observation of partially folded polypeptide chains in solution is important complementary information to the data that can be obtained by studies in crystals. It is also the main reason why the quality of a NMR structure determination is not usually characterized by a single, global parameter.

An important extension of the characterization of proteins in solution resulted from high resolution NMR studies of protein hydration. Thereby the location of hydration waters is determined by the observation of NOEs between water protons and hydrogen atoms of the polypeptide chain. Because of the dependence of the NOE on the inverse sixth power of the $^1H–^1H$ distance, only one layer of hydration water molecules is observed. For the hydration studies, the dependence of the NOE intensity on the correlation function describing the stochastic modulation of the dipole–dipole coupling between the interacting protons (Equation [1]) has a key role. The value of $f(\tau_c)$ may be governed either by the Brownian rotational tumbling of the hydrated protein molecule, or by interruption of the dipolar interaction through translational diffusion of the water molecules relative to the protein surface, whichever is faster. On this basis it could be established that surface hydration of peptides and proteins is characterized by very short residence times of the water molecules in the hydration sites, in the range from about 20 to 300 picoseconds at 10°C. This result presents an intuitive rationale for the generally observed dynamic disorder of the protein surface structure in solution, and indicates that the dehydration of the polypeptide surface will hardly ever be a rate-limiting step either in protein folding or in intermolecular interactions with proteins.

Another intriguing NMR observation bears on internal mobility of protein molecules. This is, for example, manifested by 180° flipping motions of the aromatic rings of phenylalanine and tyrosine. The first observation of these “ring flips” on the millisecond to microsecond time scale in 1975 was a genuine surprise, because the aromatic rings of phenylalanine and tyrosine are among the best-defined side chains in protein crystal structures. Possible functional roles of the “breathing motions” of proteins that are manifested by these ring flips will still need to be evaluated, and their continued analysis could have considerable impact in the newly emerging field of structural genomics.

Recent Progress and Outlook

With the availability of a rapidly increasing number of completely sequenced genomes, new challenges arise for the methods used for three-dimensional structure determination. On the one hand, “structural genomics” initiatives in several leading research centers focus on the development of technology for high-throughput structure determination to generate a comprehensive atlas of protein folds, so that remaining gaps could be filled by structure

prediction methods. There is clearly a lot of room to further enhance the efficiency of each step of protein structure determination by NMR. On the other hand, in structural genomics we face the situation that newly determined protein structures should enable prediction of novel functions, whereas in classical structural biology one encounters more typically the challenge of rationalizing known functions on the basis of the three-dimensional structure.

It has been widely recognized that supplementing the determination of new protein structures with data on intermolecular interactions may provide a key for the identification of unknown gene functions. Since efficient use of conventional NMR spectroscopy in solution had been limited to particle sizes with molecular weights up to about 30'000 Dalton, a new challenge for solution NMR techniques then arose from the fact that the supramolecular structures resulting from interactions of two or several proteins, or of other macromolecular components tend to have high molecular weights. A 30'000-Dalton size limit was thus not compatible with extensive use of NMR for studies of such supramolecular structures. For example, this size limit would severely narrow down the range of potential receptor systems accessible to NMR in drug discovery projects, and restrict studies of protein–nucleic acid complexes to a small number of systems with modest size. It would also prevent the use of solution NMR for studies of membrane proteins, since these have to be reconstituted and solubilized in large mixed micelles with detergents or lipids. A few years ago this limitation was successfully challenged, since the size-range for applications of solution NMR techniques could be significantly extended through the introduction of transverse relaxation-optimized spectroscopy (TROSY). With the use of the TROSY-principle, solution NMR can now be used for *de novo* membrane protein structure determination, and functional studies with receptor systems of molecular weights well beyond 100,000 Dalton have become possible. The "record" in my laboratory actually is that NMR spectra have been recorded for a structure with molecular weight 870'000 Dalton.

Intriguing possibilities for future use of these new NMR techniques include that NMR can now be employed in drug discovery projects with very large receptors. Combined with suitable isotope-labelling strategies, TROSY-based NMR techniques have also been shown to provide a powerful approach for investigations of intermolecular interactions in supramolecular structures with two or several macromolecular components. *De novo* NMR

determination of large structures appears to be particularly attractive for, but not limited to, nucleic acid–protein complexes and small membrane proteins reconstituted in soluble detergent or lipid micelles. Such structure determinations will in turn establish a novel basis for functional studies of higher-order structures with NMR, again with particularly exciting prospects in the area of drug discovery.