Protein Structure Determination in Solution by NMR Spectroscopy

Kurt Wüthrich
From the Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

The introduction of nuclear magnetic resonance (NMR) spectroscopy as a second method for protein structure determination at atomic resolution, in addition to x-ray diffraction in single crystals, has already led to a significant increase in the number of known protein structures. The NMR method provides data that are in many ways complementary to those obtained from x-ray crystallography and thus promise to widen our view of protein molecules, giving a clearer insight into the relation between structure and function.

Biological Macromolecules and NMR Spectroscopy

The first nuclear magnetic resonance (NMR) experiments with biological macromolecules were reported more than 30 years ago (1), and with the advent of modern NMR techniques in the late 1960s and early 1970s, which included superconducting magnets, Fourier transform spectroscopy, and computer control of the instrumentation, NMR spectroscopy yielded an ever widening array of insights into the behavior of such molecules. Examples are studies of protein conformation changes, denaturation and internal mobility, pH titration of individual ionizable amino acid side chains in enzyme active sites, observation of paramagnetic centers in metalloproteins, etc. (for surveys see Refs. 2 and 3). In addition, NMR in solution has become a technique for protein three-dimensional structure determination at atomic resolution (4), which is the subject of this review.

Survey of Protein Structure Determination by NMR

Fig. 1 presents an outline of the method (4, 5) that covers the preparation of the protein for the NMR experiments, the NMR measurements, the crucial problem of obtaining assignments of the NMR lines to individual atoms in the polypeptide chain, and two separate avenues for the structural interpretation of the NMR data.

Sample Preparation—The protein is usually dissolved in 0.5 ml of water, and the ionic strength, pH, and temperature may be adjusted so as to ensure near-physiological conditions (it is advantageous to work in the slightly acid pH range from 3 to 5 (41)). The protein concentration should be at least 1 mM, ideally 3-6 mM, so that 15-30 mg of a protein with molecular weight 10,000 should be available for a structure determination. Although this concentration is high relative to that of most proteins in their physiological milieu, it is not far from the total protein concentration in many body fluids. So far, structure determinations by NMR have been reported for proteins with molecular weights up to approximately 15,000. This upper size limit may perhaps be raised to about 30,000. For molecular sizes above approximately 12,000 the NMR study will have to include the preparation of protein enriched with 15N and/or 13C, which is best achieved with biosynthetic techniques.

NMR Measurements—Because of the large number of hydrogen atoms in a protein, a one dimensional 1H NMR spectrum is crowded with mutually overlapping lines. Therefore, two-dimensional (2D)1 and three-dimensional (3D) NMR experiments are used. Fig. 2 shows a small region of a homonuclear 2D 1H NMR spectrum. The NMR peaks are spread out along the two frequency axes $\omega_1$ and $\omega_2$, and they are therefore quite well separated. In the 3D spectrum of Fig. 3 the NMR peaks have been further spread out along a third frequency axis, which corresponds to the NMR frequencies of the 15N spins in the 15N-labeled protein. As a result, the NMR peaks of the 2D 1H-1H spectrum (Fig. 2) are distributed among several 1H-1H planes, typically 64 or 128. The ensuing improved separation of the peaks is indispensable for work with larger proteins.

For an intuitive understanding of the information contained in the spectral region of Fig. 2 it may be helpful to imagine that the region from 3.75 to 4.15 ppm of the 1D 1H NMR spectrum, which contains the resonance lines of $\alpha$ protons, is along $\omega_1$, and the 1D 1H NMR spectrum from 7.4 to 8.0 ppm, which contains resonance lines from amide protons and aromatic protons, is along $\omega_2$. All the peaks seen in Fig. 2 are "cross-peaks" manifesting an interaction between a resonance line in the 1D spectrum along $\omega_1$ and a line in the 1D spectrum along $\omega_2$. For example, the peak Y11a-Q12N at the bottom of the figure correlates the $\alpha$ proton resonance of Tyr-11 along $\omega_1$ with the backbone amide proton resonance of Gln-12 along $\omega_2$. Depending on the experiment used, the cross-peaks manifest different types of interactions between the spins. The most important information needed for a de novo 3D structure determination can be obtained from nuclear Overhauser enhancement (NOE) spectroscopy, or NOESY. In a properly executed (see below) NOESY experiment (Fig. 2) a cross-peak between two hydrogen atoms is observed only if those two protons are separated by a shorter distance than approximately 5.0 Å. Since the NOE depends on the through-space distance, the locations of the two interacting protons in the primary structure may be far apart, as much as 100 residues or more (Fig. 4). In different 2D NMR experiments, which are primarily used to support obtaining the 1H NMR assignments, the cross-peaks manifest through-bond relations between protons that are separated by not more than three covalent bonds, i.e., up to those of the same amino acid residue. Frequently used experiments are correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) (4, 6).

In contrast to all other NMR parameters, $\text{H}^\text{1}-\text{H}$ distance measurements by NOE experiments can be directly related to the protein conformation. Recognizing that NOE buildup experiments (7), which enable one to eliminate derogatory effects from spin diffusion (8–10), yield reliable distance measurements in macromolecules was therefore one of three fundamental elements that constitute the foundations of the NMR method for protein structure determination. NOE

1. The abbreviations used are: 1D(2D,3D), one-dimensional (...); NOE, nuclear Overhauser enhancement; NOESY, 2D NOE spectroscopy; COSY, 2D correlation spectroscopy; TOCSY, 2D total correlation spectroscopy; RMSD, root mean square deviation.
The protein was uniformly labeled with $^1$H to the extent of 295%. The NMR data with a variable target function algorithm supplemented by molecular mechanics energy minimization (e.g., Ref. 41) or an initial analysis with metric matrix distance geometry followed by molecular dynamics calculations (e.g., Ref. 42).

Each successful calculation with one of the aforementioned computational procedures yields a molecular structure that represents a good fit of the experimental data. To check that

---

**Fig. 2.** Three-dimensional $^1$H-$^1$H NOESY spectrum of the N-terminal domain comprising residues 1-76 of the P22 repressor. The protein was uniformly labeled with $^1$H to the extent of 295%.

The NOE buildup for a polypeptide chain (represented by the horizontal line in the center) with and without sequence-specific resonance assignments. Open circles represent hydrogen atoms of the polypeptide, and dotted lines the short $^1$H-$^1$H distances of less than 9 A manifested by the NOEs (see text) (reproduced from Ref. 4).

**Fig. 3.** Three-dimensional $^1$H-$^1$H NOESY spectrum of Antennapedia homeodomain. To indicate the pairs of hydrogen atoms for which close proximity is evidenced by these data, the cross-peaks are identified either by the one-letter amino acid symbols of 2 residues, their sequence positions, and the proton types, or for intraresidual NOEs by the identification of 1 residue and two proton types. This spectral region was taken by taking a part of one of the $^1$H-$^1$H planes of the three-dimensional spectrum of Fig. 3 (reproduced from Ref. 22).

**Fig. 4.** Scheme illustrating the information content of $^1$H-$^1$H NOEs in a polypeptide chain (represented by the horizontal line in the center) with and without sequence-specific resonance assignments. Open circles represent hydrogen atoms of the polypeptide, and dotted lines the short $^1$H-$^1$H distances of less than 9 A manifested by the NOEs (see text) (reproduced from Ref. 4).
the NMR data determine a unique three-dimensional structure, a group of conformers is compared that was obtained from a series of calculations using the same input but different, randomly chosen starting conditions (see below).

The availability of two methods for protein three-dimensional structure determination may have a useful mutual control function. For example, a crystal structure of rat metallothionein-2 (62) that was different from the NMR structure in solution (63, 64; Fig. 6) was subsequently found to need revision, and a new crystal structure of this metallothionein is nearly identical with the solution structure (65).

For several globular proteins a close similarity was observed between the molecular architectures in single crystals and in solution (e.g. Refs. 5, 38, 46, 61, 66, 67), including hydrogen-bonded secondary structures and the spatial arrangement of the interior amino acid side chains (61, 66). Protein surface areas, however, have been found to have, as a rule, significantly different structure and dynamic properties in crystals and in solution. The complementary information on the molecular surface obtained with the two methods is of special interest, since protein functions depend largely on the nature of molecular surface areas in direct contact with the substrates.

For polypeptides that do not form globular structures one may quite generally expect to find different conformations in crystals and in noncrystalline milieux. The polypeptide hormone glucagon is a typical example (34, 35).

**Protein Structures in Solution and in Single Crystals**

There is already an impressive list of three-dimensional NMR structures of proteins that have never been crystallized, including, for example, the Antennapedia homeodomain from Drosophila (54), several zinc finger proteins (55, 56), epidermal growth factors (57, 58), and interleukin 8 (59).

Even when protein crystals are available it is often difficult to obtain isomorphous heavy atom derivatives suitable for solving the crystallographic phase problem (23, 24). A Patterson rotation search (24) using an independently solved NMR structure of the same or a closely related protein in solution may then be a viable alternative route to the phase determination (60, 61).

A critical assessment of a NMR structure determination can be based on the facts that nearly complete sequencespecific resonance assignments are indispensable as a basis for a structure determination (21), that the quality of a structure determination is improved if stereospecific assignments are obtained for the prochiral centers (68, 69), that at least 10 conformational constraints per residue should have been measured, and that each individual structure calculation must represent an acceptable fit of the experimental data, with small residual violations of the NMR and steric constraints. If the structure calculation is repeated with different starting conditions, a high quality structure determination generates a tight bundle of conformers, which corresponds to
a small value for the average of the pairwise root mean square deviations (RMSDs). As an illustration, for the Antiparamecia (Anpt) homoeodomain complete sequence-specific resonance assignments were obtained, with stereospecific assignments for 33 out of a total of 88 pairs of diastereotopic substituents. For the 53 residues shown in Fig. 7, 533 conformational assignments were obtained, with stereospecific assignments for the backbone atoms, 0.9 Å for the backbone dihedral angle, and 0.25 Å for the complete polypeptide backbone spectrum. These numbers are representative of the NMR method for studies of intermolecular interaction, in particular with the use of stable isotopes, to extend NMR structure determination to larger proteins, maybe in the size range 15,000-30,000 (48-50, 52). The potential of the NMR method for studies of intermolecular interactions, for example, with the use of iso- tope-edited 1H NMR spectroscopy (51, 70) and for structural and kinetic studies relating to the protein folding problem (e.g. Refs. 71-73) may be even more attractive with regard to obtaining new fundamental insights to be used as a platform for the design of functionally improved proteins.

Acknowledgments—I owe a great debt to a large number of colleagues who participated in this work and whose names appear in the references. I thank Dr. K. V. Chary and Dr. G. Otting for Fig. 3, and R. Marani for the careful processing of the manuscript.

REFERENCES