The Development of Nuclear Magnetic Resonance Spectroscopy as a Technique for Protein Structure Determination

KURT WÜTHRICH

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

Received June 13, 1988 (Revised Manuscript Received October 12, 1988)

In the past, three-dimensional protein structures at atomic resolution could be determined exclusively by diffraction techniques with protein single crystals.[1] Now a second method for protein structure determination is available, which makes use of data collection by nuclear magnetic resonance (NMR) spectroscopy of proteins in solution.[2] The development of this new approach during the period 1978–1983[3-6] and the subsequent determination of the complete three-dimensional structure of a globular protein in solution[7] (Figure 1) came at a time of rapidly increasing interest in protein structures, as indicated by a quotation from a recent *New York Times* article titled "Genetic engineers prepare to create brand new proteins":[8] "A big barrier, however, is that scientists have little experimental data on how proteins actually look.... Until recently, the only way to determine the structure of a protein has been through X-ray crystallography.... But it can take several years to determine the structure of a single protein by this method. For one thing, many proteins do not readily form crystals. Now a technique used in medical imaging, nuclear magnetic resonance spectroscopy, is being used to avoid crystallization altogether, because it can analyze proteins in solution."

The quoted text refers to the fact that genetic engineers designing proteins must rely on knowledge of structure and functions of natural proteins. This constraint is readily appreciated, since a polypeptide chain with n amino acid residues can have 20^n different sequences, which virtually rules out systematic screening as a strategy for the design of new or improved proteins. However, although the amino acid sequences of many thousand proteins are known, only approximately 400 three-dimensional structures have been described. At most, 10% of these structures are at high resolution, and until 1984 all information on three-dimensional structures came from protein crystals. The first two sections of the present article describe the kind of new, additional structural information that is now available from NMR studies of proteins and the current practical implementation of the NMR method.

It is important to note that the quotation erroneously relates protein structure determination by NMR to medical NMR imaging. In fact, imaging of macroscopic objects[9] and determination of protein structure[2] were developed independently during the last decade, and they rely on fundamentally different uses of the NMR phenomenon, which was first observed in 1946.[10,11] Sections III and IV, which are the main parts of this article, attempt to clarify this issue. They describe the foundations of the NMR method for protein structure determination and place them in perspective with other popular applications of NMR techniques.

I. Protein Structure Determination by NMR or by Diffraction Techniques Yields Complementary Information

X-ray diffraction and NMR can both be used independently to determine the complete three-dimensional structure of proteins. The complementarity of the two methods results from the facts that the time scales of the two measurements are widely different[1,2] and, in contrast to the need of single crystals for diffraction studies, the NMR measurements use proteins in solution.

(1) Protein structure determinations by NMR or by X-ray diffraction provide the basis for meaningful comparisons of corresponding structures in single crystals and in noncrystalline states. This is highly relevant, since the solution conditions for NMR studies may coincide closely with the natural, physiological environment of the protein, or they may be varied over a wide range for studies of structural transitions with pH, temperature, or ionic strength.[2] Extensive similarities between crystal and solution structures as well as major conformational rearrangements between the two states have already been documented. For example, the α-amylase inhibitor Tendamistat (Figure 2) is a protein that has the same global molecular architecture in crystals and in solution,[12,13] with localized dif-

Kurt Wüthrich was born in Switzerland in 1938. He received a diploma degree in chemistry and physics from the University of Bern in 1962 and a diploma degree in sports and a Ph.D. in chemistry from the University of Basel in 1964. Following postdoctoral work at the University of Basel and the University of California, Berkeley, he spent two years as a Member of Technical Staff at Bell Telephone Laboratories, Murray Hill, NJ. In 1969 he moved to the ETH in Zürich, where he is Professor of Biophysics and currently the head of the Institute of Molecular Biology and Biophysics. His research is concerned with the molecular biology, structure, and function of proteins, with special emphasis on the use of nuclear magnetic resonance in this field.

(1) Blundell, T. L.; Johnson, L. N. *Protein Crystallography*; Academic Press: New York, 1976.
(2) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(3) Dubs, A.; Wagner, G.; Wüthrich, K. *Biochim. Biophys. Acta* **1979**, *577*, 177–194.
(4) Braun, W.; Bösch, Ch.; Brown, L. R.; Gō, N.; Wüthrich, K. *Biochim. Biophys. Acta* **1981**, *667*, 377–396.
(5) Wüthrich, K.; Wider, G.; Wagner, G.; Braun, W. *J. Mol. Biol.* **1982**, *155*, 311–319.
(6) Braun, W.; Wider, G.; Lee, K. H.; Wüthrich, K. *J. Mol. Biol.* **1983**, *169*, 921–948.
(7) Williamson, M. P.; Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 295–315.
(8) Pollack, A. *New York Times* **1988**, *March 15*, C1 and C11.
(9) Morris, P. G. *Nuclear Magnetic Resonance Imaging in Medicine and Biology*; Oxford University Press: Oxford, 1986.
(10) Bloch, F.; Hansen, W. W.; Packard, M. E. *Phys. Rev.* **1946**, *69*, 127.
(11) Purcell, E. M.; Pound, R. V.; Bloembergen, N. *Phys. Rev.* **1946**, *70*, 986–987.

0001-4842/89/0122-0036$01.50/0

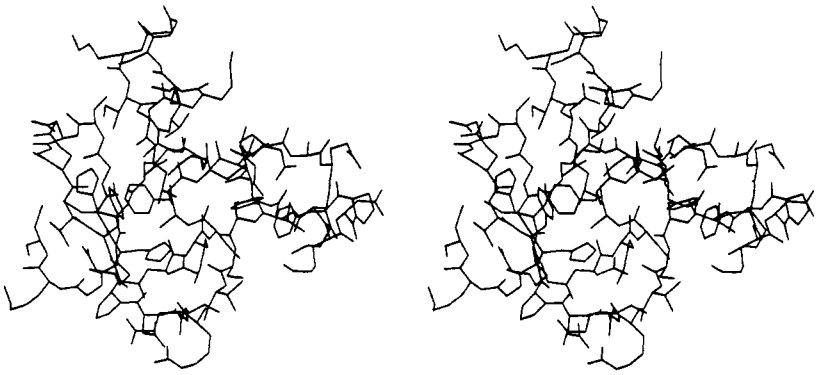

Figure 1. Stereoview of the three-dimensional structure of bull seminal proteinase inhibitor IIA (BUSI IIA) determined from NMR measurements in aqueous solution. All bonds connecting heavy atoms are shown for the entire polypeptide chain of 57 residues.[7]

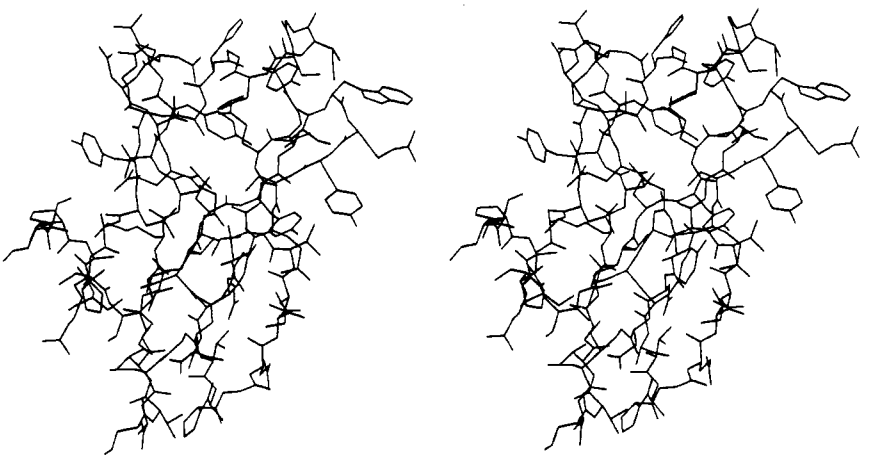

Figure 2. Stereoview of the three-dimensional structure of the α-amylase inhibitor Tendamistat determined from NMR measurements in aqueous solution. All bonds connecting heavy atoms are shown for the residues 5–73 (the complete molecule consists of residues 1–74). The orientation of the molecule was chosen so that the presumed active site consisting of the residues Trp18-Arg19-Tyr20 is clearly visible in the upper right.[12]

ferences seen primarily near the molecular surface, including the active site (Figure 2). Similar observations were reported for other globular proteins, such as the basic pancreatic trypsin inhibitor (BPTI)[14] and barley serine proteinase inhibitor-2.[15] Metallothionein-2 from rat liver (Figure 3) is a protein for which different global molecular architectures were reported in aqueous solution[16] and in crystals,[17] such that 21 of the total of 28 coordinative bonds linking the seven metal ions with the polypeptide chain are different.[18] Another well-documented example for the occurrence of different molecular architectures in different environments is the polypeptide hormone glucagon.[6,19,20]

(2) If no suitable single crystals can be grown, NMR is presently the only method available for the deter-

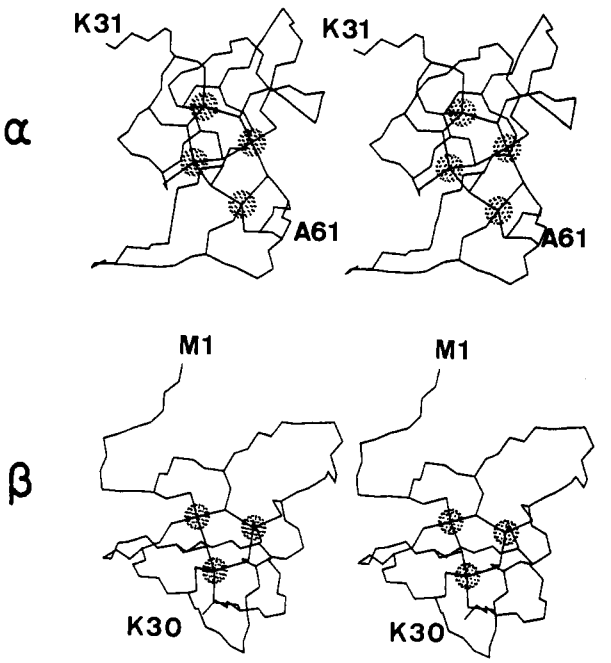

Figure 3. Stereoview of the three-dimensional structure of the α- and β-domains of rat $[Cd_7]$-metallothionein-2 determined from NMR measurements in aqueous solution. The bonds between the heavy atoms of the polypeptide backbone and the metal-bound cysteine side chains are drawn, and the Cd^{2+} ions are represented as dotted spheres. The lettering identifies the first and last residues in each domain. (Reproduced from ref 16. Copyright 1988 Academic.)

(12) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377–382; *Ibid.*, in press.
(13) Pflugrath, J. W.; Wiegand, G.; Huber, R.; Vértesy, L. *J. Mol. Biol.* **1986**, *189*, 383–386.
(14) Wagner, G.; Braun, W.; Havel, T. F.; Schaumann, Th.; Gō, N.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 611–639.
(15) Clore, G. M.; Gronenborn, A. M.; James, M. N. G.; Kjaer, M.; McPhalen, C. A.; Poulsen, F. M. *Protein Eng.* **1987**, *1*, 313–318.
(16) Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. *J. Mol. Biol.* **1988**, *203*, 251–268.
(17) Furey, W. F.; Robbins, A. H.; Clancy, L. L.; Winge, D. R.; Wang, B. C.; Stout, C. D. *Science* **1986**, *231*, 704–710.
(18) Vašák, M.; Wörgötter, E.; Wagner, G.; Kägi, J. H. R.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 711–719.
(19) Sasaki, K.; Dockevill, S.; Ackmiak, D. A.; Tickle, I. J.; Blundell, T. L. *Nature* **1975**, *257*, 751–757.
(20) Bösch, Ch.; Bundi, A.; Oppliger, M.; Wüthrich, K. *Eur. J. Biochem.* **1978**, *91*, 209–214.

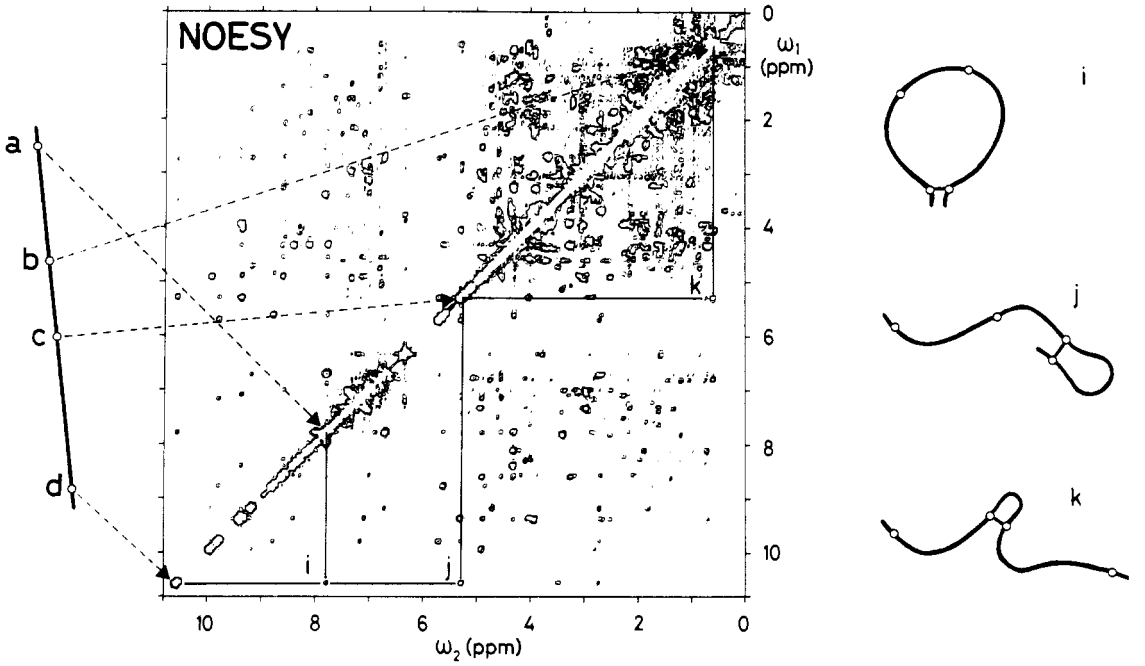

Figure 4. Illustration of the description of the NMR method for protein structure determination in solution. In the center, a contour plot of a 500-MHz ^{1}H NOESY spectrum of the protein basic pancreatic trypsin inhibitor (BPTI) is shown, with the two frequency axes ω_1 and ω_2. Three cross peaks are marked i–k and linked by horizontal and vertical lines with the diagonal positions of the protons connected by the corresponding NOEs. On the left, an extended polypeptide chain is represented by a straight line, and four protons in this chain are identified by circles and the letters a–d. The broken arrows connect these protons with their resonance positions on the diagonal of the NOESY spectrum. On the right, there is a schematic representation of three circular structures formed by the polypeptide chain, which are manifested by the cross peaks i–k.

mination of the three-dimensional protein structure. An example is bull seminal proteinase inhibitor IIA (BUSI IIA) (Figure 1).[7] Conversely, for large molecules, or generally if a stable concentrated (1 to several mM) solution cannot be prepared, X-ray diffraction remains the only avenue to the protein structure, always provided that suitable single crystals are available.

(3) For a characterization of the internal dynamics of proteins, NMR provides direct, quantitative measurements of the frequencies of certain high activation energy motional processes, and at least semiquantitative information on additional high-frequency processes.[21] The corresponding information from X-ray structure determinations commonly consists of an outline of the conformation space covered by the combination of static disorder and high-frequency structural fluctuations.[1] Furthermore, neutron diffraction in single crystals[22] and NMR in solution[2,21] can both be employed for studies of the exchange of labile protons, thus potentially enabling direct comparison of dynamic properties in the different states.

II. The Practice of Protein Structure Determination by NMR

Presently, the data for protein structure determination are collected by using two-dimensional (2D) NMR experiments[2,23] (Figure 4). The 2D NMR spectra which are of prime importance for work with proteins contain an array of *diagonal peaks* with frequencies $\omega_1 = \omega_2$, which correspond largely to the conventional one-dimensional (1D) spectrum and display the chemical-shift positions of the resonance lines. In addition, there are a large number of *cross peaks* with $\omega_1 \neq \omega_2$. Through simple geometric patterns, as indicated in Figure 4 for the peaks labeled i–k, each cross peak establishes a correlation between two diagonal peaks. In 2D nuclear Overhauser enhancement spectroscopy (NOESY), the cross peaks represent nuclear Overhauser effects (NOEs), which tell us that the protons corresponding to the two correlated diagonal peaks are separated only by a short distance, say less than 5.0 Å. In correlated spectroscopy (COSY) and many related experiments,[2,23] the cross peaks manifest scalar spin–spin couplings between the correlated protons. NOESY is the pivotal experiment in studies of protein structures, and its information content is used in the following way. First, sequence-specific resonance assignments must be obtained, i.e., for all the protons in the polypeptide chain, the corresponding diagonal peaks must be identified, as is indicated in Figure 4 by the dashed arrows linking the protons a to d with their diagonal positions in the NOESY spectrum. Each NOESY cross peak then tells us that two protons in known locations along the polypeptide chain are separated by a distance of less than 5.0 Å in the three-dimensional protein structure. Since the overall length of an extended polypeptide chain with n residues is $n \times 3.5$ Å, the NOEs may impose stringent constraints on the polypeptide conformation. On the right of Figure 4, this is schematically indicated by the formation of the three circular structures, which are closed by the near approach of the three pairs of protons correlated via the NOESY cross peaks i–k. Typically, the NOESY spectra of proteins contain many hundred cross peaks, which shows that the three-dimensional polypeptide chain forms a large number of circular structures of the type shown in Figure 4. These

(21) Wagner, G.; Wüthrich, K. In *Methods in Enzymology 131, Enzyme Structure Part L*; Hirs, C. H. W., Timasheff, S. N., Eds.; Academic Press: New York, 1986; pp 307–326.
(22) Kossiakoff, A. A. *Nature* **1982**, *296*, 713–721.
(23) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1987.

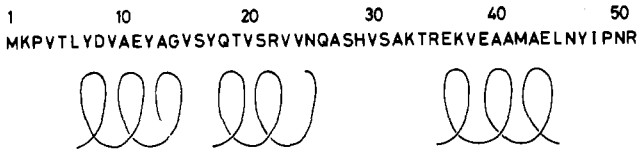

Figure 5. Secondary structure of the DNA-binding domain 1–51 of the *E. coli lac* repressor determined by NMR in aqueous solution. The locations of three α-helical segments are indicated below the amino acid sequence.[24]

data are interpreted either by an empirical search for distinct patterns of conformational constraints defining the secondary structure (Figure 5) or by computational techniques capable of identifying those three-dimensional arrangements of the polypeptide chain that satisfy all the experimental constraints, resulting in a complete description of the tertiary structure (Figures 1–3).[2]

III. Elements of the NMR Method for Protein Structure Determination

The development of the method outlined in section II involved three principal factors: (1) identification of NMR parameters that are experimentally accessible under the conditions of the spin physics prevailing in macromolecules in solution, and that can be related in a straightforward way to the molecular conformation; (2) an efficient technique for obtaining sequence-specific assignments of the many hundered to several thousand NMR lines in a protein; and (3) suitable interactive model building procedures and numerical techniques for the structural interpretation of the NMR data.

(1) NMR Parameters and Protein Conformation. To appreciate fully the contents of this section, one must consider that NMR studies of proteins evolved from NMR spectroscopy of low molecular weight chemical compounds, and that they were at all times critically dependent upon the available NMR instrumentation. Observations first made in the early 1950s have revealed that a given molecule can give rise to multiple NMR lines because of different chemical shifts for individual spins,[25] and because of scalar spin–spin couplings.[26] Using measurements of chemical shifts and spin–spin coupling constants for the elucidation of the covalent structures of synthetic compounds and natural products, NMR soon became an indispensable tool in chemical laboratories. Considering that protein structure determination by NMR is based primarily on measurements of NOEs,[2] the following text taken from p IX of ref 27, which was published in 1971, is an interesting historical document: The NOE "has found limited use over the past two decades in the study of chemical kinetics and, somewhat more recently, in the assignment of NMR spectra. Lately, interest in the NOE has grown enormously following the realization that detailed qualitative and quantitative information on molecular configuration and conformation can be obtained from it. The uniqueness of this approach to problems in molecular structure, together with the increasing availability of NMR spectrometers sufficiently sophisticated for NOE studies and sufficiently simple in operation to be used on a routine basis, have increased and will continue to increase the applications of the method manyfold. However, the numerous existing books on NMR written primarily for chemists barely mention the NOE and do not provide the background in nuclear relaxation theory necessary to understand it. Aside from original research papers, only a few references of a highly theoretical nature are extant. The difficulty and rigor of these references has been a source of frequent misunderstandings and has surely limited the growth of the field."

The foregoing quotation makes reference to the limitations on the accessible NMR parameters imposed by the NMR equipment. Standard proton resonance frequencies of commercial spectrometers were 40 MHz in the 1950s and 60 or 100 MHz in the 1960s. In 1966, superconducting magnets were first used for high-resolution NMR at 220 MHz, and the field strength was then further increased to 360 MHz in 1975, 500 MHz in 1980, and 600 MHz in 1987, resulting in improvements of the spectral resolution and the sensitivity. The sensitivity and the feasibility of complex experiments were independently greatly improved by novel experimental procedures. In the 1960s, computers of average transients were used for calculating the average of multiple scans recorded in the continuous wave (CW) mode; in the early 1970s, Fourier transform spectroscopy was introduced; and in the late 1970s, 2D NMR was added.[23]

The first report of a protein ^{1}H NMR spectrum appeared in 1957.[28] Considering the state of NMR spectroscopy at that time, it is not surprising that the early projects concentrated on chemical shift measurements, especially since spin–spin coupling fine structure could not be resolved in the spectra of proteins. It soon became apparent that the proton chemical shifts are highly sensitive to changes of the protein conformation and could be used, for example, in studies of protein denaturation,[29] or for pH titrations of ionizable groups.[30,31] In favorable cases, a qualitative explanation for unusual chemical shifts in proteins was obtained from correlations with the corresponding crystal structures, in particular for protons located near aromatic rings and thus subject to the influence of the local ring current fields.[29,32–34] The time scale defined by the proton chemical shifts further provided a basis for investigations on internal mobility in proteins.[33,34] With regard to the determination of otherwise unknown protein structures, however, the following message had to be accepted: Although the dispersion of the proton chemical shifts by the nonbonding interactions in folded proteins is sufficiently large to provide the resolution required for a complete spectral analysis,[5] the relations

(24) Zuiderweg, E. R. P.; Kaptein, R.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 5837–5841.
(25) Arnold, J. T.; Dharmatti, S. S.; Packard, M. E. *J. Chem. Phys.* **1951**, *19*, 507–509.
(26) Gutowsky, H. S.; McCall, D. W.; Slichter, C. P. *J. Chem. Phys.* **1953**, *21*, 279–292.
(27) Noggle, J. H.; Schirmer, R. E. *The Nuclear Overhauser Effect, Chemical Applications*; Academic Press: New York, 1971.
(28) Saunders, M.; Wishnia, A.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 3289–3290.
(29) McDonald, C. C.; Phillips, W. D. *J. Am. Chem. Soc.* **1967**, *89*, 6332–6341.
(30) Bradbury, J. H.; Scheraga, H. A. *J. Am. Chem. Soc.* **1966**, *88*, 4240–4246.
(31) Roberts, G. C. K.; Jardetzky, O. *Adv. Protein Chem.* **1970**, *24*, 447–545.
(32) Wüthrich, K. *Struct. Bonding* **1970**, *8*, 53–121.
(33) Wüthrich, K. *NMR in Biological Research, Peptides and Proteins*; North Holland: Amsterdam, 1976.
(34) Jardetzky, O.; Roberts, G. C. K. *NMR in Molecular Biology*; Academic Press: New York, 1981.

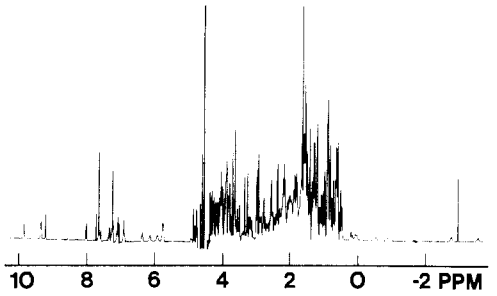

Figure 6. One-dimensional ^1H NMR spectrum at 360 MHz of ferrocytochrome c-551 from *Pseudomonas aeruginosa*. The well-separated lines between 9 and 10 ppm correspond to the four meso protons α–δ, and the five lines between 0 and −4 ppm correspond to the fragment βCH_2–γCH_2-S–ϵCH_3 of the axial methionine, which is coordinated to the heme iron through the sulfur atom (the molecular structure is shown in Figure 10).

between protein conformation and chemical shifts are too ambiguous to present a basis for structure determination.

NOEs in proteins were reported starting in 1970,[35-37] but with experiments devised for small molecules, the NOE distance information was masked by spin diffusion.[38,39] Around 1975, NOEs were more systematically applied in studies of cyclopeptide structures,[40,41] where the available conformation space is a priori drastically restrained by the ring-closure condition.[42] To refine the cyclic structures, quantitative NOE measurements of ^1H–^1H distances were attempted with the same experiments as for small organic molecules,[27,43] and these data were combined with the conformational information from spin–spin coupling constants[44] to distinguish between different proposed models for the peptide conformation.[40,41] Measurements of ^1H–^1H distances in proteins were introduced in 1978, when it was demonstrated that the NOE distance information in macromolecules could be recovered by recordings of the NOE buildup.[45,46] Subsequently, with further improved spectral resolution, supplementary conformational costraints resulted from measurements of spin–spin coupling constants,[2,47] and from observation of slowed exchange for hydrogen-bonded amide protons.[2]

This account would not be complete without reference to the important role played by hemoproteins. As a consequence of the large local magnetic field of the heme groups, which function as naturally built-in shift reagents, some ^1H NMR lines are well separated from

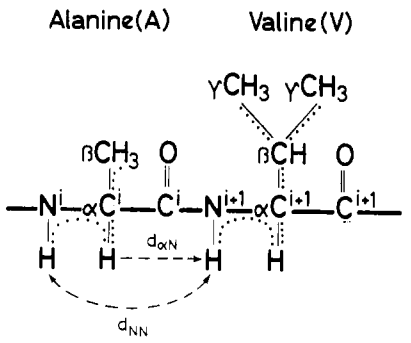

Figure 7. Illustration of the description of sequential resonance assignments. In the dipeptide segment -Ala-Val- the dotted lines indicate ^1H–^1H relations which can be established by the scalar spin–spin couplings observed in COSY. The broken arrows indicate relations between protons in sequentially neighboring residues, which can be established by NOESY cross peaks manifesting short sequential distances $d_{\alpha N}$ (between αCH and the amide proton of the following residue) and d_{NN} (between the amide protons of neighboring residues).

the bulk of the many hundred proton resonances in these molecules (Figure 6).[32,33] 1D NMR experiments with these well-resolved lines enabled detailed studies of the nuclear spin physics in macromolecules in solution long before such data became available for other proteins, where corresponding measurements are dependent on the use of 2D NMR. Both the initial observations of ^1H–^1H NOEs in a protein[35] and the initial NOE buildup studies[45,48] were made with cytochrome c. Hemoproteins were also important stepping stones in the development of the sequential resonance assignment procedure[48] and the structural interpretation of NMR data.[32,49,50]

(2) Sequential Resonance Assignments. Without resonance assignments, the size and the location in the amino acid sequence of the circular structures in Figure 4 would not be defined, and hence, there would be no way to determine a molecular structure from NOE distance information. The role of obtaining sequence-specific resonance assignments in protein structure determinations by NMR[5] is thus similar to that of solving the phase problem in protein crystallography.[1]

The elucidation of sequence-specific resonance assignments is not straightforward because proteins invariably contain more than one residue of each of the 20 common amino acids. As a consequence, the identification of the groups of scalar-coupled spins belonging to individual amino acid residues is in general not sufficient to define a unique sequence location. Prior to the introduction of the sequential assignment technique during the period 1979–82,[3,5,51-53] other techniques would in favorable cases yield assignments for a small number of protons, but could not provide the information needed as a basis for protein structure determination. These early attempts to assign protein ^1H

(35) Redfield, A. G.; Gupta, R. K. *Cold Spring Harbor Symp. Quant. Biol.* **1971**, *36*, 405–411.
(36) Balaram, P.; Bothner-By, A. A.; Dadok, J. *J. Am. Chem. Soc.* **1972**, *94*, 4015–4017.
(37) Campbell, I. D.; Dobson, C. M.; Williams, R. J. P. *Chem. Commun.* **1974**, 888–889.
(38) Hull, W. E.; Sykes, B. D. *J. Chem. Phys.* **1975**, *63*, 867–880.
(39) Kalk, A.; Berendsen, H. J. C. *J. Magn. Reson.* **1976**, *24*, 343–366.
(40) Gibbons, W. A.; Crepaux, D.; Delayre, J.; Dunand, J. J.; Hajdukovic, G.; Wyssbrod, H. R. In *Peptides: Chemistry, Structure, Biology*; Walter, R., Meienhofer, J., Eds.; Ann Arbor Science: Ann Arbor, 1975; pp 127–137.
(41) Glickson, J. D.; Gordon, S. L.; Pitner, T. P.; Agresti, D. G.; Walter, R. *Biochemistry* **1976**, *15*, 5721–5729.
(42) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178–187.
(43) Anet, F. A. L.; Bourn, A. J. R. *J. Am. Chem. Soc.* **1965**, *87*, 5250–5251.
(44) Bystrov, V. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, *10*, 41–82.
(45) Gordon, S. L.; Wüthrich, K. *J. Am. Chem. Soc.* **1978**, *100*, 7094–7096.
(46) Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1979**, *33*, 675–680.
(47) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741–751.
(48) Keller, R. M.; Wüthrich, K. *Biochim. Biophys. Acta* **1978**, *533*, 195–208.
(49) Keller, R. M.; Wüthrich, K. *Biochim. Biophys. Acta* **1980**, *621*, 204–217.
(50) Senn, H.; Billeter, M.; Wüthrich, K. *Eur. Biophys. J.* **1984**, *11*, 3–15.
(51) Wagner, G.; Anil-Kumar; Wüthrich, K. *Eur. J. Biochem.* **1981**, *114*, 375–384.
(52) Billeter, M.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1982**, *155*, 321–346.
(53) Wagner, G.; Wüthrich, K. *J. Mol. Biol.* **1982**, *155*, 347–366.

NMR spectra used either isotope labeling with ^2H, ^{13}C, or ^{15}N or comparisons with protein crystal structures, with oligopeptides corresponding to segments of the protein sequence, or with homologous proteins based on chemical shift measurements (for reviews, see ref 33 and 34 and Chapter 4 of ref 2).

The sequential assignment strategy[2,52,53] is illustrated by Figure 7. The dotted lines indicate the ^1H–^1H connectivities within the amino acid residues, which can be established via scalar spin–spin couplings, and which identify the amino acid types. Relations between protons in sequentially neighboring amino acid residues are established by NOEs manifesting close approach between αCH_i and NH_{i+1} ($d_{\alpha N}$), NH_i and NH_{i+1} (d_{NN}), or both. In Figure 7, we obtain the result that the protein studied contains a dipeptide segment Ala-Val. This dipeptide is then matched against the independently known amino acid sequence. If the latter contains Ala-Val only once, the assignment problem is solved. Otherwise, to distinguish between the different Ala-Val sites, tri- or tetrapeptide segments including Ala-Val must be identified by NMR and matched against the amino acid sequence.

Three principal advantages of the sequential assignment technique are as follows: (i) No prior knowledge of the polypeptide conformation is needed, since at least one of the two distances $d_{\alpha N}$ or d_{NN} (Figure 7) is always sufficiently short to be observed by NOEs.[52] (ii) In globular proteins, the probability is very small that identical tri- or tetrapeptide segments occur more than once, and thus the procedure is generally applicable.[2] (iii) Identification of the sequential NOEs is an integral part of the data collection for the protein structure determination. This ensures that the overall procedure is highly efficient. Limitations arise when a substantial proportion of the αCH and amide proton chemical shifts are degenerate. This will eventually be one of the factors that restrict the use of the sequential assignment technique to relatively small proteins, at present up to molecular weights of 10 000–20 000. Somewhat larger molecules may become accessible for structure determination in solution by combination of sequential ^1H NMR assignments with suitable isotope labeling of the proteins.[2,54,55]

(3) (a) Interactive Construction of Molecular Models for the Structural Interpretation of NMR Data. Identification of the well-known α-helix and other regular helical structures, antiparallel and parallel β-sheets, and tight turns can be achieved with a search for patterns of NOEs which coincide with those expected for these secondary structures.[56] This is best explained by way of an example: In an extended polypeptide chain, the distances from the hydrogen atoms in the first amino acid residue to those in the fourth residue are approximately 14 Å and, thus, far too long to be manifested by NOEs. In contrast, in an α-helix, the first and the fourth residue are only about 4.0 Å apart. Therefore, observation of ^1H–^1H NOEs between residues in relative sequence positions i and i + 3 over a certain polypeptide segment shows that this part of the chain forms an α-helix. Additional data may independently indicate the presence of an α-helix, including short sequential distances d_{NN} (Figure 7),[52] small vicinal spin–spin couplings $^3J_{HN\alpha}$,[47] and slowed exchange of the hydrogen-bonded amide protons.[2] Similarly, there are characteristic patterns of NOE data, spin–spin couplings, and amide proton exchange rates for the other common regular polypeptide secondary structures.[2,47,52,56]

The roots of this empirical approach for secondary structure identification can be traced back to investigations on correlations between local dipeptide conformation and intramolecular ^1H–^1H distances during the period 1975–79.[3,40,41,57–59] The presently used procedures for secondary structure identification in proteins[2,47,52,56] differ from this early work with amino acids and oligopeptides by the fact that they are based on qualitative observations rather than quantitative measurements of intraresidual and sequential ^1H–^1H distances and spin–spin coupling constants. With the exclusive use of qualitative experimental data, the procedure is efficient and robust in practical applications. It requires very little additional work besides the sequence-specific resonance assignments by the sequential method. The results obtained are nonetheless much more detailed and precise than those from other commonly used techniques. For example, in Figure 5, the length and the sequence location of the three α-helices were identified,[24] whereas circular dichroism experiments could only indicate an approximate value for the percentage of the residues in the protein that are involved in helical structures.

Interactive model building has also been used for the determination of the three-dimensional polypeptide fold at low resolution, or for more precise studies of local details in protein structures. For example, CONFOR[60] is a computer graphics program that displays molecular structures and visualizes violations of experimental NMR data in these structures, which can then be interactively modified by variation of individual torsion angles to minimize the residual constraint violations. Since the action of different individual constraints on the molecular structure can be highly cooperative, such interactive techniques are limited for use with low-dimensional problems. In the present practice of protein structure determination by NMR, the interactive approaches have been largely superseded by the numerical procedures described below. Early on, however, they provided important insights into the relations between NMR data and protein conformation, as is exemplified by the following three projects.

In addition to the data that resulted in the identification of the three helices in *E. coli lac* repressor 1–51 (Figure 5), approximately 20 constraints on distances between protons located in different helices were observed. Low dimensionality was achieved by representing the three helices as rigid cylinders, and the relative spatial orientation of these three cylinders was then determined by using CONFOR. The resolution of the resulting structure (Figure 8) was sufficient for the

(54) Senn, H.; Otting, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1987**, *109*, 1090–1092.
(55) LeMaster, D. M.; Richards, F. M. *Biochemistry* **1988**, *27*, 142–150.
(56) Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715–740.
(57) Khaled, M. C.; Urry, D. W. *Biochem. Biophys. Res. Commun.* **1976**, *70*, 485–491.
(58) Leach, S. J.; Némethy, G.; Scheraga, H. A. *Biochem. Biophys. Res. Commun.* **1977**, *75*, 207–215.
(59) Kuo, M.; Gibbons, W. A. In *Peptides: Structure and Biological Function*; Gross, E., Meienhofer, J., Eds.; 1979, pp 229–232.
(60) Billeter, M.; Engeli, M.; Wüthrich, K. *J. Mol. Graphics* **1985**, *3*, 79–83.

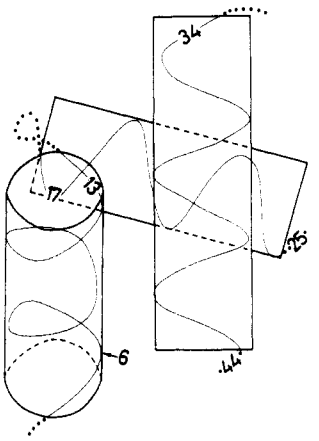

Figure 8. Three-dimensional arrangement of the three helices in the DNA-binding domain 1–51 of *E. coli lac* repressor (Figure 5) determined from NMR measurements in aqueous solution using interactive computer graphics with the program CONFOR.[60] The helices are represented by cylinders. Inside the cylinders a smooth line goes through the αC positions, and the sequence locations of the first and last helical residues are indicated. (Reproduced from ref 61. Copyright 1984 Elsevier.)

identification of the helix–turn–helix motif involved in specific DNA recognition.[61]

In cytochrome b_5, the orientation of the protoheme IX group was determined in single crystals as shown in Figure 9A. The figure shows that only two valine residues of the polypeptide chain, Val 45 and Val 61, are located near the heme, where they are in close proximity to the ring positions 1, δ, and 8. Using 1D ^{1}H NMR experiments, all resonance lines of the heme were individually assigned by a sequential approach which made no reference to the crystal structure,[48] and independently the spin system of one valine residue was identified. Subsequently, ^{1}H–^{1}H NOEs were observed between this valine and the heme substituents in the ring positions 4, β, and 5. For the molecular structure to be compatible with the NMR data, the heme group thus had to be rotated by 180° about an axis through the ring positions α and γ (Figure 9B).[49] Reexamination of the diffraction data subsequently confirmed that Figure 9B represents also the crystal structure.[62] This appears to be the first well-documented case where a precisely defined detail in a protein crystal structure was revised on the basis of structural studies in solution.

In cytochromes c, 1D ^{1}H NMR experiments (Figure 6) were used to obtain individual assignments for the heme ring substituents by a sequential approach without reference to the crystal structure,[48] and for the identification of the spin system of the axial methionine. Observation of NOEs between the methionine protons and those of the heme then enabled the determination of stereospecific resonance assignments[2,63] for the two methylene groups of methionine. Subsequently, a precise determination of the methionine conformation and its orientation relative to the heme plane was obtained, which revealed that the chirality at the sulfur atom of the axially bound methionine was different in homologous cytochromes c from different species (Figure 10).[50]

(61) Zuiderweg, E. R. P.; Billeter, M.; Boelens, R.; Scheek, R. M.; Wüthrich, K.; Kaptein, R. *FEBS Lett.* **1984**, *174*, 243–247.
(62) Mathews, F. S. *Biochim. Biophys. Acta* **1980**, *622*, 375–379.
(63) Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1983**, *169*, 949–961.

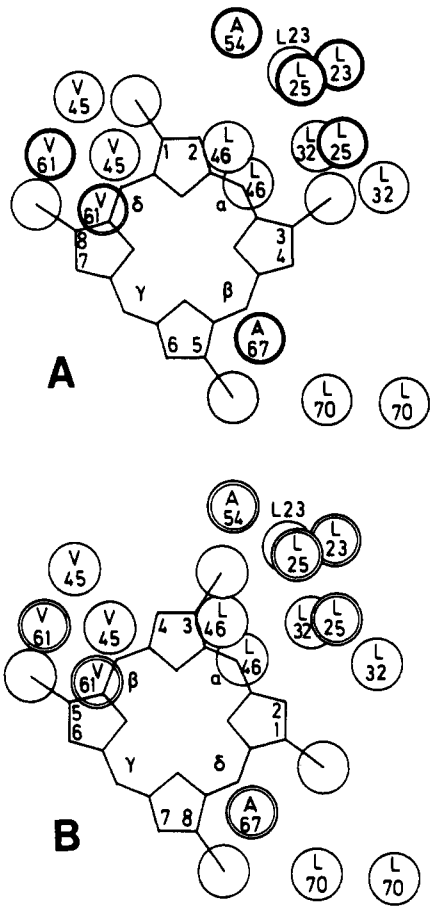

Figure 9. Heme crevice in cytochrome b_5. In the skeleton of the protoheme IX group, the four meso-proton positions are identified by α–δ and eight peripheral positions by 1–8, and the four ring methyls in positions 1, 3, 5, and 8 are represented by empty circles. Additional circles indicate the positions of amino acid methyl groups near the heme. Thick lines (in A) or double lines (in B) identify the methyls located above the heme plane, and the residues are identified by the one-letter symbol for amino acids and the sequence position. (A) Crystal structure. (B) Revised structure based on NMR data in solution[49] and reexamination of the X-ray data.[62] (Reproduced from ref 49. Copyright 1980 Elsevier.)

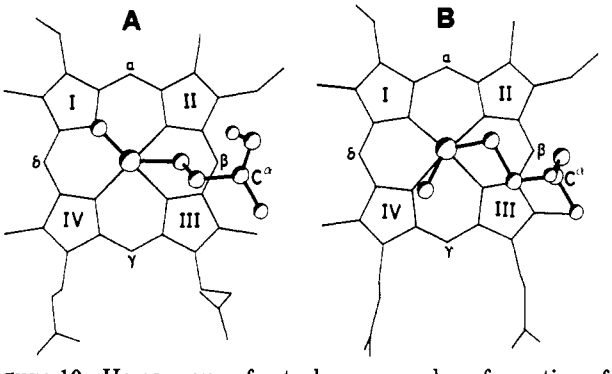

Figure 10. Heme group of cytochrome c and conformation of the iron-bound axial methionine determined by NMR in solution. Only the heavy atoms are drawn. The lettering identifies the four meso-proton positions α–δ and the four pyrrole rings I–IV of the heme and the α-carbon of the methionine. (A) Mammalian ferrocytochrome c. (B) Cytochrome c-551 from *P. aeruginosa*. Note the different chirality at the metal-bound sulfur atom in the two structures.[50]

(3) (b) Numerical Techniques for the Determination of Complete Three-Dimensional Protein Structures from NMR Data. Several fundamentally

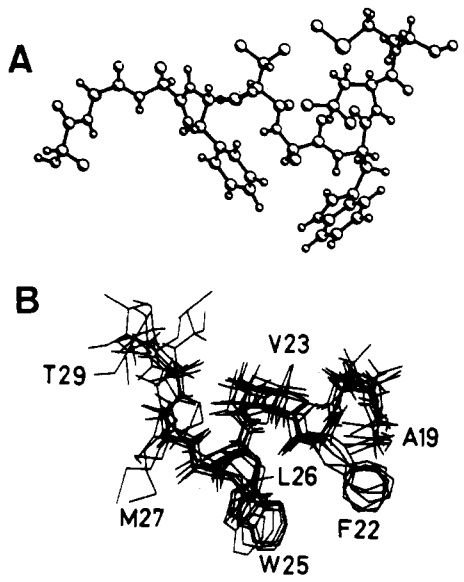

Figure 11. Three-dimensional structure of the polypeptide hormone glucagon in the lipid–water interphase on the surface of dodecylphosphocholine micelles. (A) Segment -Ala19-Glu-Asp-Phe-Val-Gln-Trp-Leu-Met27. A single conformer from a group of structures computed with a metric matrix distance geometry algorithm from a limited set of conformational constraints collected with 1D NMR experiments is shown, with Ala19 on the left (reproduced from ref 4; copyright 1981 Elsevier). (B) Segment 19–29. Six conformers computed from a more complete set of constraints collected by using 2D NMR measurements have been superimposed for minimal RMSD (reproduced from ref 6; copyright 1983 Academic). In both presentations, the side chains of the residues Gln20, Asp21, and Gln24, as well as that of Asn28 in B, are not shown, since they were not constrained by the experimental data.

different mathematical techniques have already been developed for the computation of protein structures from NMR data,[2,64] whereby the most promising approaches are based on the principles of distance geometry.[65] Work on improvements in the use of existing techniques and the development of new algorithms in this field has on its own already become a lively area of research.

Distance geometry is a long-established branch of mathematical research,[65] and starting in the mid-1970s, it was also applied for theoretical studies on protein folding and reconstruction of folded molecules from intramolecular distances.[66–68] For use in structure determinations from NMR data, new demands had to be met. Most importantly, NOE measurements are so far only used to obtain upper limits on intramolecular distances rather than precise distances,[2] and as a consequence, a complete, all-atom description of the covalent polypeptide structure, possibly supplemented with pseudoatoms,[63] must be used for the computations. This leads to quite formidable requirements for computing and storage capacity. A first program package for such calculations,[4] which made use of a metric matrix distance geometry algorithm,[64,66] had a size limitation for peptide fragments of up to ca. 10 amino acid residues, and it was applied for structure determinations with lipid-bound polypeptide hormones (Figure 11).[4,6] Another metric matrix program, DISGEO,[69] can handle globular proteins with up to ca. 100 amino acid residues. It was subjected to extensive tests, where a protein structure was reconstructed from different sets of distance contraints derived from its atomic coordinates in the X-ray structure,[70] and it enabled the first structure determination of a globular protein from experimental NMR data (Figure 1).[7] Alternative procedures, which were subjected to similar functional tests, include the program DISMAN[71] and restrained molecular dynamics calculations.[72] In practical applications, restrained molecular dynamics has so far been used primarily for refinement of structures obtained by model building, such as in Figure 8,[73] or by distance geometry calculations.[15]

Figure 11B shows a superposition for minimal root-mean-square deviation (RMSD) of a group of glucagon structures. Each of these structures is the result of a distance geometry calculation using the experimental NMR constraints.[6] The different structures were obtained by repeating the same calculation with different, randomly generated starting conditions. Since the experimental data represent only an incomplete set of intramolecular interatomic connections, and each constraint describes an allowed distance range rather than a precise value for the distance, the individual structures are similar but not identical. The result of a structure determination from NMR data is therefore commonly represented by a group of conformers, each of which represents a solution to the geometric problem of fitting the polypeptide chain to the ensemble of all experimental constraints. The spread among the different structures, usually expressed by the average of the pairwise RMSDs,[2,64,69,70] represents the precision of the structure determination. Figures 1–3 and 11A all correspond to individual conformers taken from groups of structures such as that shown in Figure 11B.

IV. Two-Dimensional NMR and Three-Dimensional Protein Structures

All three principal elements of the method for protein structure determination by NMR, i.e., measurements of intramolecular ^1H–^1H distances,[45,46] sequential resonance assignments,[3,48] and the structural interpretation of NMR data,[4,49,50] were realized with 1D NMR experiments before 2D NMR was ready for relevant applications with proteins. However, it was the combination of these principles with the use of 2D NMR experiments, as outlined in section II, that made protein structure determination by NMR a practical approach.

For experiments with macromolecules, 2D NMR[2,23] has three decisive advantages over 1D NMR. First, there is the improved resolution achieved by spreading the resonances over a two-dimensional frequency plane rather than a single frequency axis (Figure 4; compare the overlap of peaks on the diagonal and in the ω_1–ω_2 plane). Second, selective correlations between individual pairs of spins can be obtained by using nonse-

(64) Braun, W. *Quart. Rev. Biophys.* **1987**, *19*, 115–157.
(65) Blumenthal, L. M. *Theory and Application of Distance Geometry*; Chelsea: New York, 1970.
(66) Kuntz, I. D.; Crippen, G. M.; Kollmann, P. A.; Kimmelman, D. *J. Mol. Biol.* **1976**, *106*, 983–994.
(67) Crippen, G. M. *J. Comput. Phys.* **1977**, *24*, 96–107.
(68) Wako, H.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 961–969.
(69) Havel, T. F.; Wüthrich, K. *Bull. Math. Biol.* **1984**, *46*, 673–798.
(70) Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 281–294.
(71) Braun, W.; Gō, N. *J. Mol. Biol.* **1985**, *186*, 611–626.
(72) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801–3805.
(73) Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *J. Mol. Biol.* **1985**, *182*, 179–182.

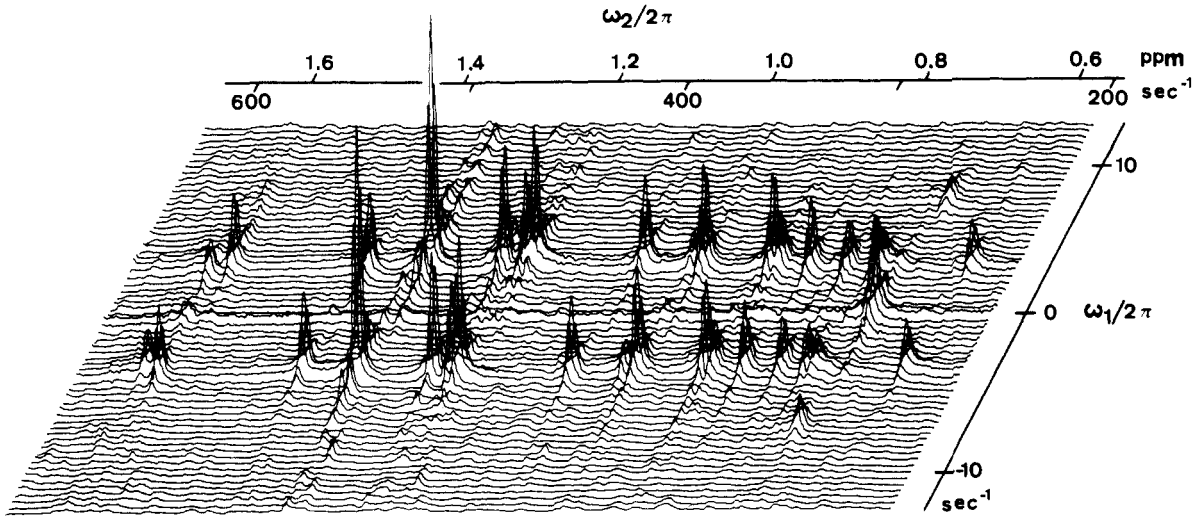

Figure 12. High-field region of a 2D *J*-resolved ^{1}H NMR spectrum at 360 MHz of the protein BPTI, which was recorded in 1977. (Reproduced from ref 76. Copyright 1977 Academic.)

lective irradiation pulses. Third, a single 2D NMR experiment can provide the data on all spin–spin interactions of a given type in the entire molecule, and the method is therefore highly efficient. It is mainly these three factors that make 2D NMR, and in the future perhaps also 3D NMR,[74] an indispensable tool for the practice of protein structure determination in solution.

The rapid advances in the practice of protein structure determination by NMR from approximately 1985 onward were largely a result of the fact that adaptation of existing 2D NMR experiments and the development of special new techniques for use with biomacromolecules had started immediately after the fundamental work on 2D NMR by Ernst and co-workers in 1975–6.[75] In 1977, the first 2D NMR spectra of a protein were recorded (Figure 12),[76] and by 1980, COSY[75,77,78] and NOESY[79] experiments with proteins could be performed in H_2O solution,[80] which provided the information needed for structure determinations.[2] Since approximately 1982, a constantly increasing number of research groups contribute to rapid further progress in 2D NMR. The resulting improvements of the classical experiments and the addition of a wide array of new experimental schemes[23] enable more rapid and thorough analyses of the spectra. As a consequence, bigger or otherwise more complex proteins can be tackled with the modern techniques.[2]

V. Conclusion and Outlook

Today the NMR method for protein structure determination can be used with proteins up to molecular weights around 20 000. It is a robust procedure, since it uses exclusively qualitative interpretations of the NMR spectra. As a consequence of the cooperative action of the conformational constraints, the resulting structures are much more precisely defined than the individual constraints. For Tendamistat (Figure 2), detailed comparisons with the crystal structure showed that for the bulk of the molecule the NMR structure is as precisely defined as the refined X-ray structure at 2.0 Å resolution, and the more pronounced disorder near the surface of the solution structure is probably functionally significant.[81] Quite generally, by providing data that are in many ways complementary to the protein structures obtained from X-ray crystallography, the NMR method promises to widen our view of protein molecules, providing a better grasp of the relations between structure and function. Considering that less than five years have gone by since the initial structure determination of a protein by NMR,[7] there is clearly much room for continued improvements. Increased efficiency from at least partial automation of the spectral analysis, and the combination of the presently described techniques with suitable isotope labeling for studies of bigger and more complex systems[54,55,82] are only two of many avenues where further progress seems imminent.

My research in the field of this review was supported by the Schweizerischer Nationalfonds and by special grants of the ETH Zürich. I thank the many colleagues mentioned in the references for the fruitful and enjoyable collaboration and R. Marani for the careful processing of the manuscript.

(74) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Magn. Reson.* **1987**, *73*, 574–579.
(75) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229–2246.
(76) Nagayama, K.; Wüthrich, K.; Bachmann, P.; Ernst, R. R. *Biochem. Biophys. Res. Commun.* **1977**, *78*, 99–105.
(77) Nagayama, K.; Anil-Kumar; Wüthrich, K.; Ernst, R. R. *J. Magn. Reson.* **1980**, *40*, 321–334.
(78) Bax, A.; Freeman, R. *J. Magn. Reson.* **1981**, *44*, 542–561.
(79) Anil-Kumar; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
(80) Anil-Kumar; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *96*, 1156–1163.
(81) Billeter, M.; Kline, A. D.; Braun, W.; Huber, R.; Wüthrich, K. *J. Mol. Biol.*, in press.
(82) Weiss, M. A.; Redfield, A. G.; Griffey, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1325–1329.