

# Protein Structure Determination in Solution by Nuclear Magnetic Resonance Spectroscopy

KURT WÜTHRICH

Knowledge of three-dimensional protein structures is one of the foundations of protein design and protein engineering. Nuclear magnetic resonance spectroscopy was recently introduced as a second method for protein structure determination, in addition to the well-established diffraction techniques with protein single crystals. This new approach enables one to carry out detailed structural studies of proteins in solution and other noncrystalline states, which may be similar or identical to the physiological environment, and promises new insights into the dynamics of protein molecules and the protein-folding problem.

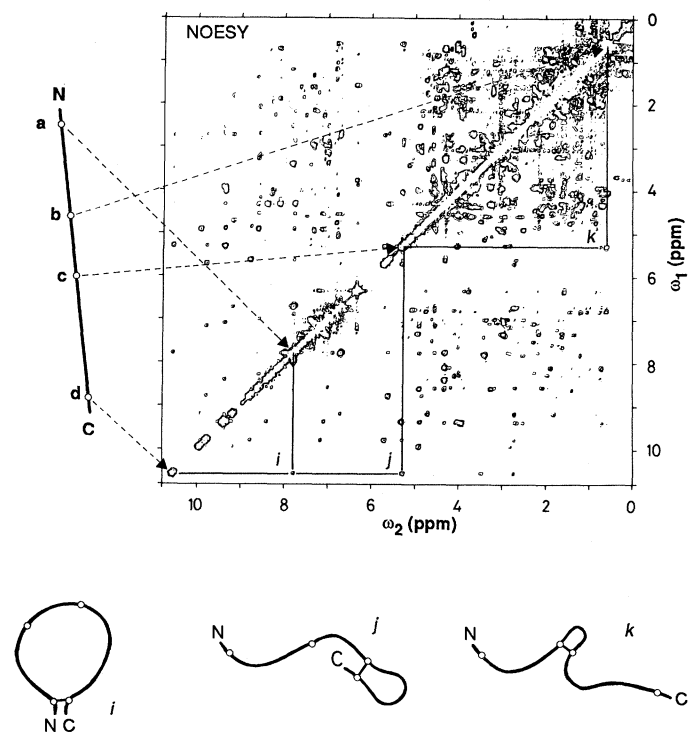
GENETIC ENGINEERS WHO ARE DESIGNING PROTEINS WITH new or improved functional properties must rely primarily on knowledge of relations between three-dimensional (3D) structure and functions of natural proteins. This constraint is readily appreciated if one considers that there are  $20^n$  different sequences for a polypeptide chain with  $n$  amino acid residues, which virtually rules out systematic screening of polypeptides produced with recombinant techniques or chemical synthesis as a strategy for the design of new or improved proteins. However, the scarcity of data on protein conformations is a major bottleneck in the progress of protein engineering. Although the amino acid sequences are known for many thousand different proteins, only  $\sim 400$  3D protein structures are available, and at most 10% of these are at high resolution. Furthermore, until recently all detailed structural information was obtained from protein single crystals with diffraction techniques (1, 2). In this article I survey a new method for protein structure determination that uses nuclear magnetic resonance (NMR) spectroscopy for data collection in solution and distance geometry for the structural analysis of the NMR data (3, 4).

The introduction of NMR as a second method for the determination of protein structures will help to increase the number of known protein structures. However, it is also of more fundamental interest, because it provides data that are in many ways complementary to those obtained from x-ray crystallography. Its use thus promises to give a better grasp of the relations between structure and function in protein molecules. The complementarity of the two methods results from the facts that the time scales of the two types of measurements are widely different (1, 3) and that, in contrast to the need to grow single crystals for diffraction studies, the NMR measurements use proteins in solution or other noncrystalline states.

The author is professor of biophysics at the Eidgenössische Technische Hochschule, Hönggerberg, CH-8093 Zürich, Switzerland.

## Survey of the Method

In present practice, the data for protein structure determination are collected by performing 2D NMR experiments (3, 5). The spectra of prime importance for work with proteins contain an array of diagonal peaks in the 2D frequency plane (Fig. 1), with  $\omega_1 = \omega_2$ , which display the chemical shift positions of the resonance lines and resemble the conventional 1D spectrum. In addition, there are a large number of cross peaks with  $\omega_1 \neq \omega_2$ . Through simple geometric patterns, each cross peak establishes a correlation between two diagonal peaks, as indicated in Fig. 1 for the cross peaks labeled  $i$ ,  $j$ , and  $k$ . For example, in 2D nuclear Overhauser enhancement spectroscopy (NOESY), the cross peaks represent nuclear Overhauser



**Fig. 1.** Proton NOESY spectrum of the protein basic pancreatic trypsin inhibitor. The spectrum was recorded at a Larmor frequency of 500 MHz. A contour plot is shown, with the two frequency axes  $\omega_1$  and  $\omega_2$ . Three cross peaks are marked  $i$ ,  $j$ , and  $k$  and are connected by horizontal and vertical lines with the diagonal positions of the protons related by the corresponding NOEs. The straight line on the left of the spectrum represents an extended polypeptide chain. The chain termini are identified by N and C, and four protons are identified by circles and the letters a to d. The broken arrows connect these protons with their diagonal peaks. Below the spectrum three circular structures formed by the polypeptide chain, which are manifested by the NOESY cross peaks,  $i$ ,  $j$ , and  $k$  (see text), are schematically represented.

**Fig. 2.** Survey of a protein structure determination by NMR.

**Protein conformation by NMR:**

1. Sequence-specific resonance assignments
2. Collect constraints (NOE  $\approx d \leq 5.0 \text{ \AA}$ )  
→ Secondary structure by pattern recognition  
→ Tertiary structure by distance geometry
3. Structure refinements

1 10 20 30 40 50  
MKPVTLYDVAEYAGVSYQTVSRV V NQASHVSAKTREKVEAAMAE L N Y I P N R



**Fig. 3.** Secondary structure of the DNA-binding domain 1 to 51 of the *Escherichia coli lac* repressor determined by NMR spectroscopy in aqueous solution. The one-letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The locations of three  $\alpha$ -helical segments are indicated below the amino acid sequence (37).

effects (NOE) and indicate that the protons corresponding to the two correlated diagonal peaks are separated by only a short distance, say, less than  $5.0 \text{ \AA}$ . In 2D correlated spectroscopy (COSY) the cross peaks represent scalar spin-spin couplings. For a protein, a COSY cross peak thus indicates that the protons corresponding to the two correlated diagonal peaks are in the same amino acid residue, where they are separated by at most three chemical bonds.

A large number of different 2D NMR experiments yielding COSY-type through-bond  $^1\text{H}$ - $^1\text{H}$  connectivities are available, for example, relayed coherence transfer spectroscopy, multiple-quantum filtered COSY, multiple-quantum spectroscopy, or total correlation spectroscopy (TOCSY) (3, 5). Using a suitable selection of three to six of these experiments in combination with NOESY or a related experiment for studies of through-space interactions, one can usually obtain a complete or nearly complete delineation of the  $^1\text{H}$ - $^1\text{H}$  connectivities in a protein. With these experimental data the structure determination is pursued as outlined in Fig. 2. In practice one needs the liberty of going back and forth between the three distinct steps of Fig. 2 in several cycles of improvement and completion of the structure determination.

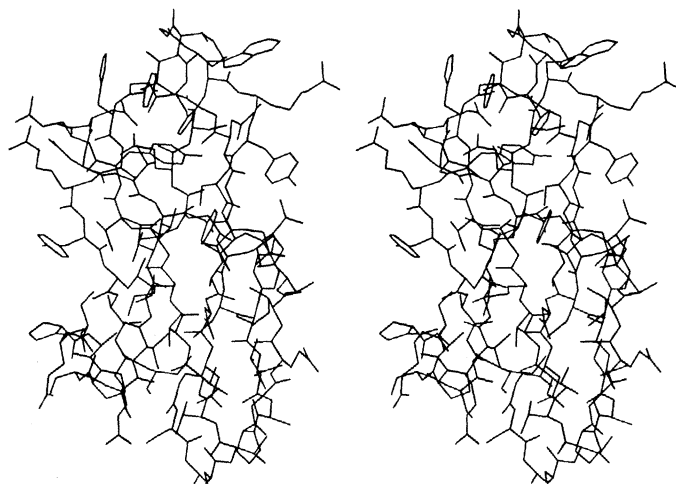
At the outset, sequence-specific resonance assignments must be obtained (6); that is, for all the protons in the polypeptide chain the corresponding diagonal peaks must be identified, as illustrated in Fig. 1 for the four protons a, b, c, and d. One achieves this result by combined analysis of COSY-type and NOESY-type experiments, using the sequential resonance assignment technique (3, 7). Once resonance assignments are available for all or nearly all of the protons, one starts to collect the data needed as input for the determination of the 3D structure. Each NOESY cross peak then indicates that two protons in known locations along the polypeptide chain are separated by a distance of less than approximately  $5.0 \text{ \AA}$  in the protein. Because the overall length of the extended polypeptide chain of a protein is several hundred angstroms, the NOEs may thus impose stringent constraints on the polypeptide conformation. Such constraints are illustrated at the bottom of Fig. 1 by the information contained in the three NOESY cross peaks *i*, *j*, and *k*; the presence of cross peak *i* in the spectrum shows that the distance between protons a and d must be short, and hence the chain must form a large circular structure. Similarly, cross peaks *j* and *k* manifest the formation of smaller circular structures. Interpretations of these data can be based on empirical recognition of patterns of such circular structures corresponding to those expected for the common secondary structure elements in proteins (3, 8), which results in the

identification of the secondary structure (Fig. 3). For the determination of the complete tertiary protein structure (Fig. 4) mathematical techniques have been developed, in particular, distance geometry algorithms (9-13). These methods consider the observed NOE distance constraints (Fig. 1) and supplementary conformational constraints from other NMR experiments in a computer search for those protein conformations that are compatible with all of the experimental measurements and the steric constraints imposed by the covalent polypeptide structure (3, 14).

## Historical Background

The NMR method for protein structure determination is of recent origin, and the first complete structure of a globular protein in solution was reported only in 1985 (15). In general, the development of the method was closely linked with that of NMR instrumentation and NMR methodology, in particular the introduction of superconducting high-field magnets, the development of 1D and 2D Fourier transform NMR spectroscopy (5), and the use of ever more powerful computers for data accumulation, storage, and processing. More specifically, the introduction of NOE experiments making possible the measurement of  $^1\text{H}$ - $^1\text{H}$  distances in macromolecules (16) and the development of an efficient technique for obtaining sequence-specific assignments of the many hundred to several thousand NMR lines in a protein (6, 7, 17) were the decisive steps leading to the method for structure determination outlined in the preceding section. The potential power of NMR for elucidating biopolymer conformations had long before been anticipated (18), but prior to the introduction of the sequential resonance assignment technique (7, 17) the situation was similar to that of protein crystallography before the successful application of the heavy-atom derivative technique for solving the phase problem (1).

Because the information obtained from NMR experiments is fundamentally different from x-ray crystal diffraction data, for which techniques for protein structure determination and refinement have long been available (1), new procedures for the structural interpretation of NMR data with proteins had to be developed. These techniques are based on relations between intramolecular proton-proton distances and the molecular conformation, and include empirical model building with molecular models or computer graphics (8, 19), distance geometry algorithms (11-13) and, more



**Fig. 4.** Stereoview of the 3D structure of the protein Tendamistat determined from NMR measurements in aqueous solution (one of the conformers in Fig. 8A is shown). All bonds connecting heavy atoms are drawn for residues 5 to 73 (32). The complete molecule contains 74 amino acid residues and has a molecular weight of 8000.

**Fig. 5.** Example of a constraint from  $^1\text{H}$  NOESY measurements used as input for protein structure calculations. The upper limit or the distance between proteins  $i$  and  $j$ ,  $d_{ij}$ , derived from  $^1\text{H}$ - $^1\text{H}$  NOEs may have different values, depending on the quality of the experimental data (see text). The lower limit corresponds to the sum of the atomic core radii (11–13) and may differ slightly from 2.0 Å, depending on the atoms to which the hydrogens are bound.

**Distance constraints:**  
**NOESY**  $d_{ij} < 4.0 \text{ \AA}$   
**van der Waals**  $d_{ij} > 2.0 \text{ \AA}$

**Allowed distance range used as input for distance geometry:**  
 $2.0 \text{ \AA} < d_{ij} < 4.0 \text{ \AA}$

recently, computational techniques such as molecular dynamics (20) and an ellipsoid algorithm (21). Distance geometry, which is a long-established branch of mathematical research (9), is at present most important among these different procedures. Starting around 1975 it was already being applied in model studies on the reconstruction of folded protein molecules from intramolecular distances (10), but for use in structure determinations from NMR data new demands had to be met with quite formidable requirements for data storage and computing capacity (11–14).

NMR experiments for proton-proton distance measurements in proteins (16), the sequential assignment technique (17), and 3D polypeptide structure determination by NMR and distance geometry (11) had all been developed and applied before 2D NMR or 3D NMR experiments were ready for the relevant measurements with proteins (22). However, full realization of the method was achieved only with the greatly improved resolution and efficiency afforded by 2D NMR spectroscopy (3, 5). It was of great practical importance that immediately after the fundamental work on 2D NMR by Ernst and co-workers in 1975 (23), work was begun to adapt existing 2D NMR experiments and to develop special new techniques for use with biomacromolecules (24).

## Information from NMR with Proteins

With the use of the NMR method of Figs. 1 and 2, complete 3D structures (Fig. 4) have so far been determined for proteins with molecular weights up to 10,000 (3). This size limit will soon reach 20,000 and with the additional use of isotope labeling (3, 25) it may reach molecular weights in the range of 30,000 to 40,000. A large fraction of the naturally occurring polypeptides and proteins are thus amenable to structure determination by NMR. The remaining sections of this article are devoted to discussions on the general characteristics and the precision of 3D protein structures determined by NMR.

In addition to the determination of the 3D structure by NMR, further information may be gained from the data, in that high-resolution NMR spectra are a highly sensitive test of the purity of a protein preparation (26). The sequential resonance assignment procedure affords a check on the amino acid composition and the amino acid sequence determined by chemical methods (3, 7), and for several proteins the published sequences had to be modified in light of the NMR data (27). With little additional work the sequential  $^1\text{H}$  NMR assignments lead to a description of the polypeptide secondary structure (3, 8), which is more detailed and precise than secondary structure determinations obtained by prediction algorithms or by optical spectroscopic techniques (Fig. 3). Because all of these data can be obtained efficiently, they might conceivably serve also for product control in the production of proteins by recombinant techniques or chemical synthesis.

It had long been recognized that conformational transitions, denaturation, and internal mobility in proteins are manifested in the NMR spectra and the NMR is also a powerful tool for studies of intermolecular interactions with proteins (18). With the availability

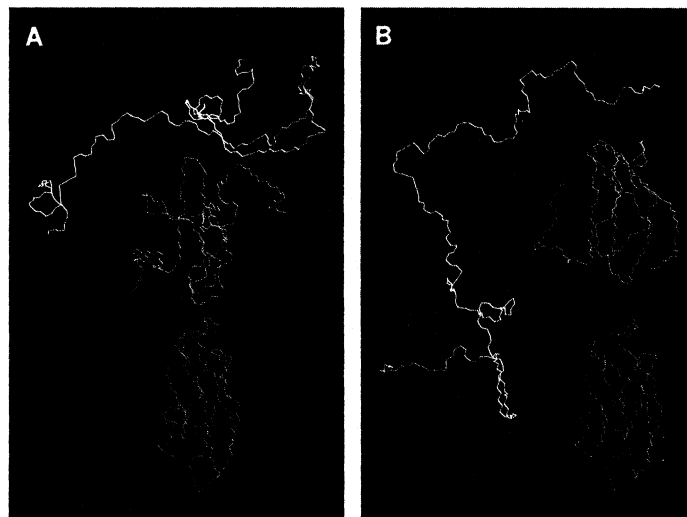
of sequence-specific resonance assignments and the 3D solution structure, NMR studies of these dynamic processes can now be analyzed in much greater detail (3), which promises to add new dimensions to the overall characterization of protein molecules.

## NMR Measurements, Distance Geometry, and Protein Solution Structures

The conformational constraints used as input for distance geometry calculations of a protein structure result primarily from NOE experiments, usually NOESY (Fig. 1). To appreciate the format of the data used (Fig. 5), one must understand that quantitative distance measurements are intrinsically difficult, because the observed NOEs depend not only on the proton-proton distances  $r$  but also on the effective rotational correlation times  $\tau_c$  (3, 5)

$$\text{NOE} \propto \frac{1}{\langle r^6 \rangle} \cdot f(\tau_c) \quad (1)$$

The correlation function  $f(\tau_c)$  accounts for the influence of motional averaging processes on the NOE. These are governed by the combined effects of the overall rotational molecular motions, which depend on the size and shape of the protein and on intramolecular motions. Thus  $\tau_c$  may vary for different locations in a protein molecule (3), and much care must be exercised in quantitative assessments of the relative distances between different pairs of protons from the corresponding NOE intensities. In addition, there are experimental limitations on the accuracy of NOE measurements, and it is difficult to exclude the possibility that individual NOEs might be quenched by competitive relaxation processes, such as the influence of traces of paramagnetic metal ions. Overall, it follows that reliable estimates of the lower bounds  $^1\text{H}$ - $^1\text{H}$  distances corresponding to given, measured NOE intensities are difficult. In contrast, upper bounds on  $^1\text{H}$ - $^1\text{H}$  distances can safely be established from NOE measurements, and for dynamic molecular structures one can also take into account that the apparent distances manifested in the NOEs correspond to a time average (Eq. 1) (11). On the basis of these considerations, only upper limits on the  $^1\text{H}$ - $^1\text{H}$  distances are



**Fig. 6.** (A and B) Snapshots from two distance geometry calculations of the protein Tendamistat with the program DISMAN (13). Both calculations used the same NMR input but different randomly generated starting structures. Yellow-green: starting structures; blue: intermediate state of the structure calculations [target size 20; see (13)]; red: final results of the calculations; only the bonds linking the backbone atoms N, C $\alpha$ , and C' have been drawn (32).

derived from the NMR experiments, and the input for the structure calculations consist of an allowed distance range bounded by this upper limit and a lower limit equal to the sum of the van der Waals radii of the two protons (Fig. 5).

For short-range connectivities between protons separated by not more than three single bonds in the chemical structure, that is, intraresidual and sequential distances (3), the NOE intensities are translated into corresponding upper bounds, typically in steps of 2.5, 3.0, and 4.0 Å. For longer range connectivities, a predetermined upper limit, usually 4.0 or 5.0 Å, depending on the protein and the experimental conditions used, is applied. Unless stereospecific assignments were obtained, in particular for the  $\beta$  methylene groups in amino acid side chains and the isopropyl groups of valine and leucine, correction factors for the use of pseudoatoms must be added to the upper limits on distances to prochiral groups of protons (28). The resulting allowed distance ranges may then be as large as from 2.0 to 9.8 Å. Supplementary conformational constraints, for example, from spin-spin coupling constants, are represented in the input by similar allowed ranges, which account for the limited accuracy of the measurements (3).

Distance geometry algorithms can be used to obtain the Cartesian coordinates of spatial molecular structures that are consistent with a predetermined set of intramolecular, interatomic distances (9, 10). To this end such algorithms identify the conformations

$$\{r_i = (x_i, y_i, z_i); i = 1, 2, \dots, N\} \quad (2)$$

which are consistent with the inequality

$$L_{ij} \leq |r_i - r_j| \leq U_{ij} \quad (3)$$

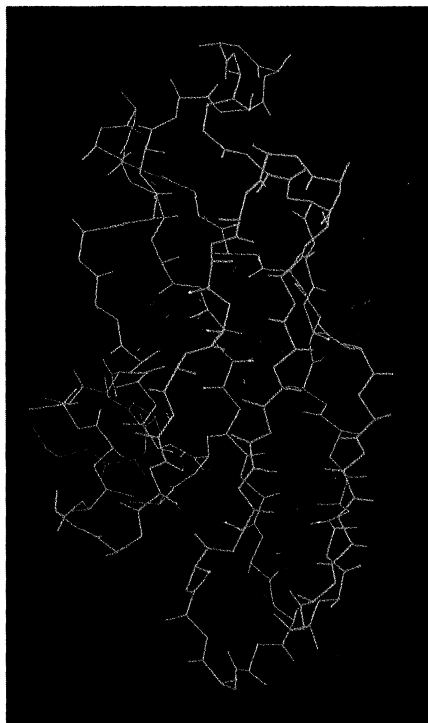
where  $L_{ij}$  and  $U_{ij}$  are lower and upper bounds, respectively, on the Euclidian distance  $|r_i - r_j|$  between the points  $r_i$  and  $r_j$ , and  $N$  is the number of points (or atoms) in the system studied. The problem described by inequality 3 may alternatively be expressed in the form of an error function, which can then be minimized. In the process of this conversion of distance information into Cartesian coordinates, one inevitably encounters the problem of local minima (29). Metric matrix methods (10–12) and a variable target function method in dihedral angle space (13, 14) have been implemented in the search for protein conformations obeying inequality 3, and the energy functions in molecular dynamics programs have been supplemented with terms representing the constraints expressed by inequality 3 (20). These three methods have been independently tested and the results compared (12, 13, 20). For this comparison,  $^1\text{H}$ – $^1\text{H}$  distances were measured in protein crystal structures and translated into NMR-type distance constraints (Fig. 5). The protein structures computed with the different algorithms from these simulated input data were then compared with the crystal structure from which the input had been derived and to each other. The result was, overall, that a crystal structure could be faithfully reproduced with the use of any of these methods, demonstrating that protein structures can be uniquely solved by NMR data (Fig. 5). Furthermore, it followed that the structures obtained are largely independent of the mathematical techniques used for the structural interpretation of the NMR data. Similar tests were subsequently repeated with experimental NMR data used as the input (30, 31). Two facets of the results obtained are particularly noteworthy. First, the overall dimensions of the protein structures are well defined, although the NMR measurements are limited to short distances compared with the diameter of a protein. Second, as a result of the concerted influence of a large number of constraints, the determination of the atom coordinates in the computed protein structures is much more precise than the individual constraints (Fig. 5). As an illustration, Figs. 6 to 8 [which were prepared with the program CONFOR (19)] exemplify some salient points of a protein structure computa-

tion from NMR data. [The distance geometry program DISMAN (13) was used with an experimental input of almost 1000 conformational constraints for Tendamistat (Fig. 4) (4, 32).]

*A protein structure determination from NMR data uses randomly selected starting conditions.* In a DISMAN calculation, for example, the starting polypeptide conformation (yellow-green structures in Fig. 6) is obtained by randomizing the dihedral angles about the single bonds. Under the influence of the experimental constraints, the variable target function method used in DISMAN then adjusts this conformation through variation of the dihedral angles, whereby it first satisfies the local constraints and gradually considers also constraints acting on longer segments of the polypeptide chain. Snapshots of three stages in this optimization procedure are displayed in Fig. 6A. In Fig. 6B a second calculation is shown that used the same NMR input but a different, randomly generated starting structure.

*Convergence of a structure calculation is judged from the residual constraint violations.* At any target size of a DISMAN calculation, the fit of the calculated structure to the experimental constraints is evaluated by an error function representing the residual constraint violations. When the target size reaches the entire polypeptide chain, the structure calculation is stopped at a predetermined value of the error function. In Fig. 7 are shown the residual violations of NOE distance constraints (Fig. 5) in the final Tendamistat structure of Fig. 6A. The red lines are in the direction of the constraints, and their lengths represent the extent of the violations. All of the residual violations in Fig. 7 are short compared to a C–C bond length. Thus Fig. 6A represents a successful structure calculation, and the red drawing represents a good solution to the geometric problem of finding a 3D arrangement of the linear polypeptide chain that satisfies all of the experimental constraints.

*The precision of the structure determination is evaluated from comparison of a group of conformers that are calculated from the same experimental data but that use different starting conditions.* The result of a single distance geometry calculation (Fig. 6A) represents one structure that is compatible with the NMR experiments, but it cannot tell whether this solution is unique. Thus the calculation is repeated with



**Fig. 7.** The Tendamistat structure obtained as the solution of the distance geometry calculation in Fig. 6A is shown with blue lines, where only the polypeptide backbone and the  $\text{C}'=\text{O}$  and  $\text{C}^\alpha-\text{C}^\beta$  bonds are drawn. The red lines indicate the direction of NOE distance constraints, and their lengths represent the residual violations of the experimental constraints at the time when the calculation was stopped. Some red lines appear to lie outside of the structure because the side chains are not shown.

different starting conditions (Fig. 6B). For each calculation, convergence is judged by the residual constraint violations (Fig. 7), and all satisfactory solutions are included in a group of conformers representing the structure of the protein. The final result of a structure determination is commonly presented as a superposition of the individual distance geometry solutions for pairwise minimum root mean square deviations (RMSDs) relative to a predetermined conformer. In Fig. 8 a superposition is shown of nine DISMAN solutions for Tendamistat. The entire polypeptide backbone is shown (Fig. 8A), as well as two short polypeptide segments complete with all amino acid side chains (Fig. 8, B and C). All nine structures are similar, have nearly the same overall dimensions, and contain the same secondary structure. The precision of the structure determination, as reflected by the dispersion among the nine structures, varies along the polypeptide chain, both for the backbone and for the amino acid side chains. The same type of structure presentation is obtained if in place of DISMAN a metric matrix distance geometry algorithm, such as DISGEO (4), is used (3, 11, 12, 15).

The individual conformers obtained as the solutions of the distance geometry calculations generally contain some close interatomic contacts and correspondingly high conformational energies. Evidence obtained so far indicates that by subsequent application of a suitable energy minimization routine low-energy structures are obtained that have nearly the same molecular geometry as the distance geometry solutions (3, 15, 20, 33).

### Protein Dynamics and Precision of Structure Determinations by NMR

A measure for the precision of a structure determination by NMR is the average of the pairwise RMSDs among a group of distance geometry solutions. For Tendamistat (Fig. 8), these numbers are 0.85, 1.04, and 1.53 Å for the polypeptide backbone, the backbone and the interior amino acid side chains, and all heavy atoms, respectively. These numbers, and the results in Fig. 8, indicate that different molecular regions have significantly different local RMSDs. In particular, the interior side chains are nearly as well determined as the backbone, whereas increased disorder is indicated for the surface side chains. For the polypeptide backbone, increased disorder is seen at the amino terminus on the extreme left, and at the top of the structure in Fig. 8A. Comparison of Fig. 8B (an interior loop) and Fig. 8C (the loop at the top of the molecule in Fig. 8A) illustrates different degrees of precision in side chain structure determination.

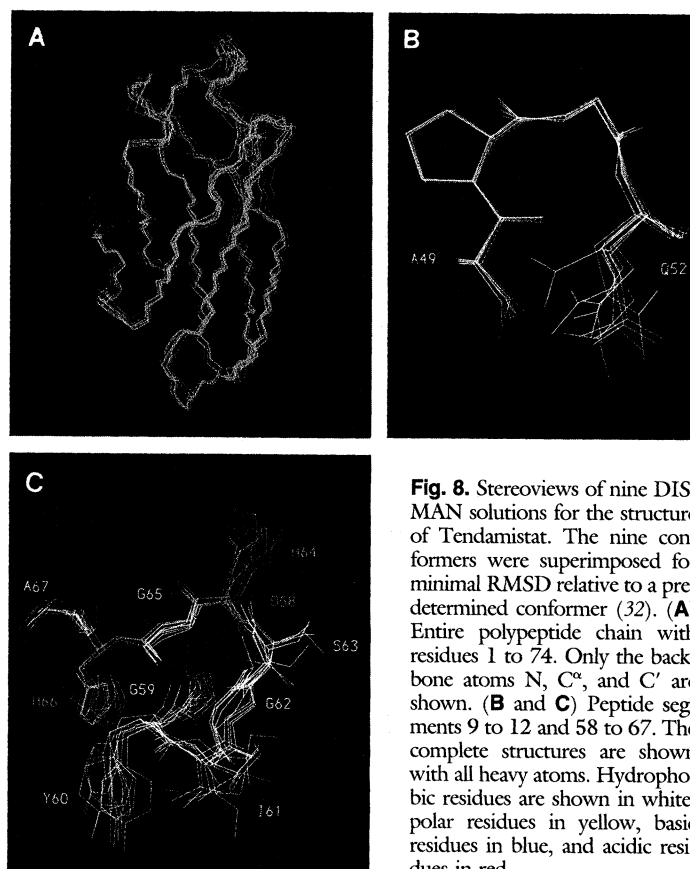
Technically speaking, the spread among the different individual distance geometry solutions (Fig. 8) occurs because the experimental input consists of allowed distance ranges (Fig. 5) rather than exact distances, and because there may be only a few constraints for some areas of the molecule. [Test calculations showed that groups of nearly identical solutions would result from similar calculations with complete sets of more accurate distance constraints (12, 13).] Nonetheless, the structure determinations are more precise than the individual input constraints (Fig. 5). To better rationalize this result, one should consider that the lower limit given in Fig. 5 is largely hypothetical, because the steric constraints imposed by the heavy atoms to which the protons are bound generally prevent a very close approach of the two protons.

Knowledge about internal mobility of proteins at ambient temperature is one of the reasons for the conservative interpretation of NOEs in terms of upper bounds on distances (Fig. 5) (3, 11). Although possible effects of internal dynamics have thus been allowed for, there are no straightforward relations between the

variations of the local RMSDs across the protein molecule (Fig. 8) and intramolecular motions. For example, on the protein surface the number of NOE constraints per residue is on average smaller than in the interior of the molecule, which also affects the results of the calculations. The relations between the local RMSDs, static structural disorder, and dynamic disorder are an interesting theme for continued research.

### Conclusions and Outlook

Tendamistat is presently the most precisely determined solution structure of a protein (32), and its crystal structure was independently determined (34). For the polypeptide backbone and the interior side chains, the Tendamistat structures in solution and in the crystals are nearly the same; for these parts of the molecule the precision of the NMR structure in solution is comparable to the refined crystal structure at 2.0 Å resolution (35). The structural disorder reflected by the local RMSDs along the polypeptide backbone (Fig. 8A) mimics faithfully the corresponding information from the temperature factors in the crystal structure. Significant structural differences occur near the protein surface, and generally the increase of structural disorder near the surface relative to the core of the molecule is more pronounced in solution than in the crystal structure. Similar observations were made with other globular proteins, but there are also polypeptide chains where the structures in solution and in the crystals were found to be different, for example, the polypeptide hormone glucagon and the metal-rich protein metallothionein (36). The NMR method for protein structure determination is a valuable complement to crystal structure determination. The efficiency of the method should be considerably improved by at least partial automation. Practical use seems ensured not only in academic institutions



**Fig. 8.** Stereoviews of nine DISMAN solutions for the structure of Tendamistat. The nine conformers were superimposed for minimal RMSD relative to a predetermined conformer (32). (A) Entire polypeptide chain with residues 1 to 74. Only the backbone atoms N, C $\alpha$ , and C' are shown. (B and C) Peptide segments 9 to 12 and 58 to 67. The complete structures are shown with all heavy atoms. Hydrophobic residues are shown in white, polar residues in yellow, basic residues in blue, and acidic residues in red.

but also by NMR groups established in conjunction with the protein engineering projects of a number of industrial enterprises.

Besides the practical applications for protein structure determination, the NMR methodology outlined in this article promises to provide new insights into different fundamental aspects of the physical and structural chemistry of biopolymers. Attractive topics are more detailed characterization of protein surfaces in solution and the internal dynamics of proteins, investigations relating to the protein-folding problem, and studies of intermolecular interactions with proteins, in particular in protein-DNA complexes.

#### REFERENCES AND NOTES

1. T. L. Blundell and L. N. Johnson, *Protein Crystallography* (Academic Press, New York, 1976).
2. G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, Berlin, 1979); J. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
3. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
4. Abbreviations: Tendamistat, an  $\alpha$ -amylase inhibitor from *Streptomyces tendae*; DISGEO and DISMAN, distance geometry programs for the calculation of protein structures from NMR data.
5. R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1987).
6. K. Wüthrich, G. Wider, G. Wagner, W. Braun, *J. Mol. Biol.* **155**, 311 (1982).
7. G. Wagner, Anil-Kumar, W. Wüthrich, *Eur. J. Biochem.* **114**, 375 (1981); M. Billeter, W. Braun, K. Wüthrich, *J. Mol. Biol.* **155**, 321 (1982); G. Wagner and K. Wüthrich, *ibid.*, p. 347; G. Wider, K. H. Lee, K. Wüthrich, *ibid.*, p. 367.
8. K. Wüthrich, M. Billeter, W. Braun, *J. Mol. Biol.* **180**, 715 (1984).
9. L. M. Blumenthal, *Theory and Application of Distance Geometry* (Chelsea, New York, 1970).
10. G. M. Crippen, *J. Comput. Phys.* **24**, 96 (1977); T. F. Havel, I. D. Kuntz, G. M. Crippen, *Bull. Math. Biol.* **45**, 665 (1983).
11. W. Braun, Ch. Bösch, L. R. Brown, N. Gö, K. Wüthrich, *Biochim. Biophys. Acta* **667**, 377 (1981).
12. T. F. Havel and K. Wüthrich, *Bull. Math. Biol.* **46**, 673 (1984); *J. Mol. Biol.* **182**, 281 (1985).
13. W. Braun and N. Gö, *J. Mol. Biol.* **186**, 611 (1985).
14. W. Braun, *Q. Rev. Biophys.* **19**, 115 (1987).
15. M. P. Williamson, T. F. Havel, K. Wüthrich, *J. Mol. Biol.* **182**, 295 (1985).
16. S. L. Gordon and K. Wüthrich, *J. Am. Chem. Soc.* **100**, 7094 (1978); G. Wagner and K. Wüthrich *J. Magn. Reson.* **33**, 675 (1979).
17. A. Dubs, G. Wagner, K. Wüthrich, *Biochim. Biophys. Acta* **577**, 177 (1979).
18. C. C. McDonald and W. D. Phillips, *J. Am. Chem. Soc.* **89**, 6332 (1967); G. C. K. Roberts and O. Jardetzky, *Adv. Protein Chem.* **24**, 447 (1970); K. Wüthrich, *Struct. Bonding (Berlin)* **8**, 53 (1970); K. Wüthrich, *NMR in Biological Research, Peptides and Proteins* (North-Holland, Amsterdam, 1976); R. A. Dwek, I. D. Campbell, R. E. Richards, R. J. P. Williams, Eds., *NMR in Biology* (Academic Press, New York, 1977); O. Jardetzky and G. C. K. Roberts, *NMR in Molecular Biology* (Academic Press, New York, 1981).
19. M. Billeter, M. Engeli, K. Wüthrich, *J. Mol. Graphics* **3**, 79 (1985).
20. R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, W. F. van Gunsteren, *J. Mol. Biol.* **182**, 179 (1985); A. T. Brünger, G. M. Clore, A. M. Gronenborn, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3801 (1986).
21. M. Billeter, T. F. Havel, K. Wüthrich, *J. Comput. Chem.* **8**, 132 (1987).
22. Anil-Kumar, R. R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **95**, 1156 (1980); Anil-Kumar, G. Wagner, R. R. Ernst, K. Wüthrich, *ibid.* **96**, 1156 (1980); *J. Am. Chem. Soc.* **103**, 3654 (1981).
23. W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **64**, 2229 (1976).
24. K. Nagayama, K. Wüthrich, P. Bachmann, R. R. Ernst, *Biochem. Biophys. Res. Commun.* **78**, 99 (1977).
25. M. A. Weiss, A. G. Redfield, R. H. Griffey, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1325 (1986); H. Senn, G. Otting, K. Wüthrich, *J. Am. Chem. Soc.* **109**, 1090 (1987); D. M. LeMaster and F. R. Richards, *Biochemistry* **27**, 142 (1988).
26. G. Otting, P. Marchot, P. E. Bougis, H. Rochat, K. Wüthrich, *Eur. J. Biochem.* **168**, 603 (1987).
27. P. Štrop, G. Wider, K. Wüthrich, *J. Mol. Biol.* **166**, 641 (1983); A. S. Arseniev, V. I. Kondakov, V. N. Maiorov, V. F. Bystrov, *FEBS Lett.* **165** (57) (1984); G. Wagner, D. Neuhaus, E. Wörgötter, M. Vašák, J. H. R. Kägi, K. Wüthrich, *Eur. J. Biochem.* **157**, 275 (1986); D. E. Wemmer, N. V. Kumar, R. M. Metrione, M. Lazdunski, G. Drobny, N. R. Kallenbach, *Biochemistry* **25**, 6842 (1986); P. A. Kosen, J. Finer-Moore, M. P. McCarthy, V. J. Basus, *ibid.* **27** 2775 (1988).
28. K. Wüthrich, M. Billeter, W. Braun, *J. Mol. Biol.* **169**, 949 (1983).
29. G. Némethy and H. A. Scheraga, *Q. Rev. Biophys.* **10**, 239 (1977).
30. G. Wagner, W. Braun, T. F. Avel, Th. Schaumann, N. Gö, K. Wüthrich, *J. Mol. Biol.* **196**, 611 (1987).
31. G. M. Clore, M. Nilges, A. T. Brünger, M. Karplus, A. M. Gronenborn, *FEBS Lett.* **213**, 269 (1987).
32. A. D. Kline, W. Traun, K. Wüthrich, *J. Mol. Biol.* **189**, 377 (1986); *ibid.*, in press.
33. M. Billeter, Th. Schaumann, W. Braun, K. Wüthrich, unpublished results.
34. J. Pflugrath, E. Wiegand, R. Huber, L. Vértessy, *J. Mol. Biol.* **189**, 383 (1986).
35. M. Billeter, A. D. Kline, W. Braun, R. Huber, K. Wüthrich, *ibid.*, in press; W. Braun, O. Epp, K. Wüthrich, R. Huber, *ibid.*, in press.
36. W. Braun, G. Wider, K. H. Lee, K. Wüthrich, *ibid.* **169**, 921 (1983); P. Schultze et al., *ibid.* **203**, 251 (1988).
37. E. R. P. Zuiderweg, R. Kaptein, K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5837 (1983).
38. The author's laboratory is supported by the Schweizerischer Nationalfonds (project 3.198.85) and by special grants of the Eidgenössische Technische Hochschule Zürich. I thank M. Billeter, W. Braun, and P. Schultze for help in the preparation of Figs. 6 through 8 and R. Marani for careful processing of the manuscript.

## AAAS–Newcomb Cleveland Prize

### To Be Awarded for an Article or a Report Published in *Science*

The AAAS–Newcomb Cleveland Prize is awarded to the author of an outstanding paper published in *Science*. The value of the prize is \$5000; the winner also receives a bronze medal. The current competition period began with the 3 June 1988 issue and ends with the issue of 26 May 1989.

Reports and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration of the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the competition period, readers are invited to nominate papers appearing in the Reports or Articles sections. Nominations must be typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to the AAAS–Newcomb Cleveland Prize, AAAS, Room 924, 1333 H Street, NW, Washington, DC 20005, and **must be received on or before 30 June 1989**. Final selection will rest with a panel of distinguished scientists appointed by the editor of *Science*.

The award will be presented at the 1990 AAAS annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.