Printing: Gerike, Berlin - Binding: Mikolai, Berlin - Cover Design: Hansbernd Lindemann, Berlin.

*Offprint*

# INNOVATIONS IN PROTEASES AND THEIR INHIBITORS

EDITOR

FRANCESC X. AVILÉS


WALTER DE GRUYTER · BERLIN · NEW YORK 1993

# CONTENTS





# COMPUTER - SUPPORTED NMR STRUCTURE DETERMINATION OF PROTEINS IN SOLUTION ILLUSTRATED WITH STUDIES OF PROTEIN PROTEINASE INHIBITORS


**Kurt Wüthrich*, Peter Güntert & Kurt D. Berndt**

*Institut für Molekularebiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich (Switzerland).*


## Introduction

The relationship between amino acid sequence, three-dimensional structure and biological function of proteins is one of the most intensely pursued areas of molecular biology and biochemistry. In this context, the three-dimensional structure has a pivotal role, as its knowledge is a prerequisite to understanding the physical, chemical, and biological properties of a protein (1). Until 1984, structural information at atomic resolution could only be determined by X-ray diffraction techniques with protein single crystals (2). The introduction of nuclear magnetic resonance (NMR) spectroscopy as a technique for protein structure determination has made it possible to obtain structures at comparable resolution in solution (3, 4). The establishment of the NMR method for biomolecular structure determination also means that there is now detailed structural information available from data reflecting two different molecular time scales, which promises to broaden our view of macromolecular conformation by supplementing crystal structure data with a picture that also includes dynamic properties of the molecules. In addition, proteins studied independently by X-ray diffraction and NMR provide the basis for meaningful comparisons of corresponding structures in single crystals and in noncrystalline states and provide insights into possible structural rearrangements caused by the local environment of the crystal lattice.

Protein inhibitors of proteinases, which are ubiquitous in animal and plant tissues as well as in microorganisms and viruses, have played an important role in the development of the NMR method for protein structure determination. For example, because of its ready avail-

ability, high stability in solution, and favorable NMR-spectral properties, the bovine pancreatic trypsin inhibitor (BPTI) (Fig. 1) has been used time and again to develop and optimize new NMR experiments for studies of proteins (*e.g.*, 5-9). BPTI was also the first protein for which complete sequence-specific NMR assignments and the NMR determination of the secondary structure were obtained (10-12). Subsequently, the bull seminal protease inhibitor IIA (BUSI IIA) was the first protein for which the determination of the three-dimensional structure by NMR in solution was completed (13-15).

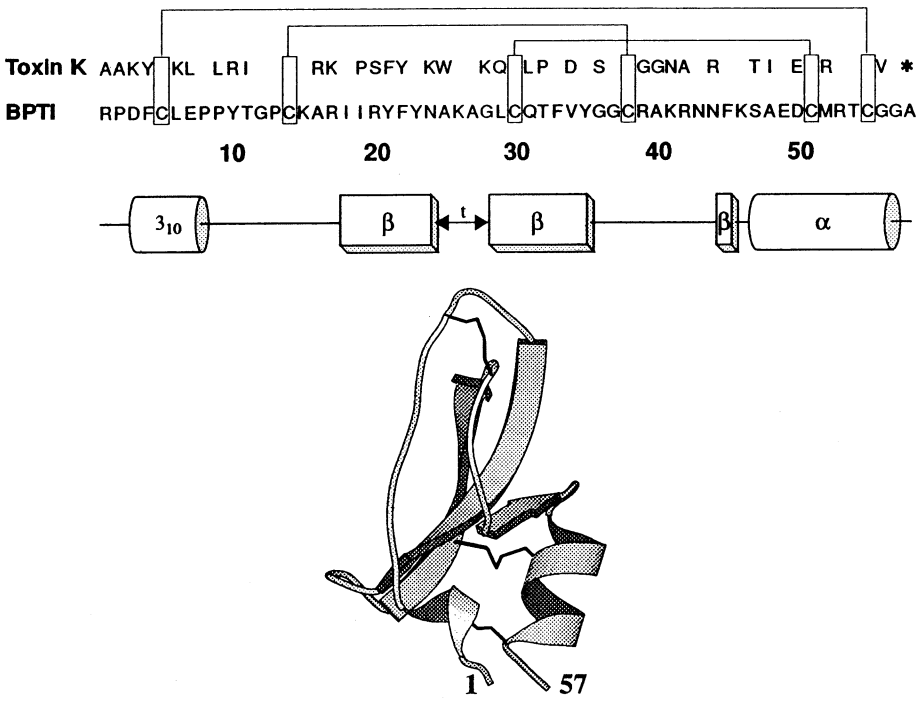


**Fig. 1.** Primary, secondary and tertiary structure of Toxin K. The amino acid sequence of Toxin K is shown in comparison to that of BPTI, with positional differences indicated with the appropriate amino acid in the sequence of Toxin K. The asterisk indicates that residue 58 is not present in Toxin K. The disulfide pairs are shown as boxes with connecting lines. The sequence positions of the regular secondary structures of these two homologous proteins as determined by NMR are identified below the sequence. At the bottom of the figure, the three-dimensional structure of Toxin K as determined by NMR is shown as a ribbon cartoon, with the three disulfide bonds displayed as solid lines tracing the covalently linked side chains.



This chapter presents a general overview of the structure determination of small proteins, such as Kunitz-type protein proteinase inhibitors, by homonuclear $^1$H NMR spectroscopy as practiced in our laboratory (16, 17), using the software package EASY (18) for spectral interpretation and the computer programs CALIBA (19, 20), HABAS (21), INFIT (22), DIANA (19, 23), ASNO (17), and GLOMSA (19, 20) for structure determination. As an illustration we use the recent structure determination of the Toxin K from the venom of the black mamba, *Dendroaspis polylepis polylepis* (24, 17). This protein, while having very similar three-dimensional structure but only 41% residue identity with BPTI (Fig. 1), has practically no tryptic or chymotryptic inhibitor activity. Its appears that its function is to facilitate the release of neurotransmitters from neuromuscular junctions (25).

## The NMR Method for Protein Structure Determination

The utility of the NMR method for the study of molecular structures depends to a large part on the sensitive variation of the resonance frequency of a nucleus with the chemical structure, the conformation of the molecule, and the solvent environment. These so-called chemical shifts ensure the necessary spectral resolution, although they do not provide direct structural information (3). They arise because a given nucleus is shielded from the externally applied magnetic field to a different extent depending on its local environment. Three of the four most abundant elements in biological materials, *i.e.*, hydrogen, carbon and nitrogen, have at least one naturally occurring isotope with nuclear spin $I = 1/2$, and are therefore suitable for high-resolution NMR experiments in solution (3). The proton ($^1$H) has the highest natural abundance (99.98%) and the highest sensitivity (due to its large gyromagnetic ratio) among the $I = 1/2$ isotopes of these elements, and hence plays a central role in most NMR experiments with biopolymers. The low natural abundance of $^{13}$C and $^{15}$N (1.11 and 0.37% respectively) and low relative sensitivity of these nuclei normally require isotope enrichment. Today, this is routinely achieved by overexpression of proteins in isotope-labeled media, which is indispensable for structure determinations with larger proteins (*e.g.*, 26, 27). Structures of small proteins with molecular weight ≤ 10,000 can be efficiently determined by $^1$H NMR alone. Structure determinations of proteins obtained from natural sources have to rely entirely on $^1$H NMR experiments - even for larger molecular weights.

The procedure used for the determination of small proteins by $^1$H NMR is outlined in Figure 2. It includes the preparation of the protein, the NMR experiments, the crucial problem of obtaining assignments of the NMR lines to individual atoms in the polypeptide chain, and finally the calculation and refinement of the 3D structure. In the following, these different steps are described in more detail and, where applicable, the software developed in our laboratory for support of individual processes is described.

## Sample Preparation

The sample preparation is largely dictated by the fact that the NMR method is a rather insensitive technique requiring millimolar concentrations for typical multidimensional (2D, 3D, 4D) experiments with proteins. Low concentration can, to some extent, be compensated for by increased measuring time, but this may deteriorate the quality of the spectra because of instrument instabilities. Therefore, NMR measurements are usually performed at the highest concentration obtainable. In practice, 5 mm o.d. sample tubes are used in most high-resolution NMR experiments with proteins, so that about 0.5 ml of the protein solution are needed for one sample. The protein concentration should be at least 1 mM, and ideally 5 mM or higher so that approximately 25 mg of a protein of molecular weight 10,000 should be available for a structure determination. The NMR experiments performed with Toxin K used a 10 mM solution, which corresponds to about 31 mg of the protein. Obviously, at these high protein concentrations, 10% impurity becomes a significant obstacle because of the additional lines

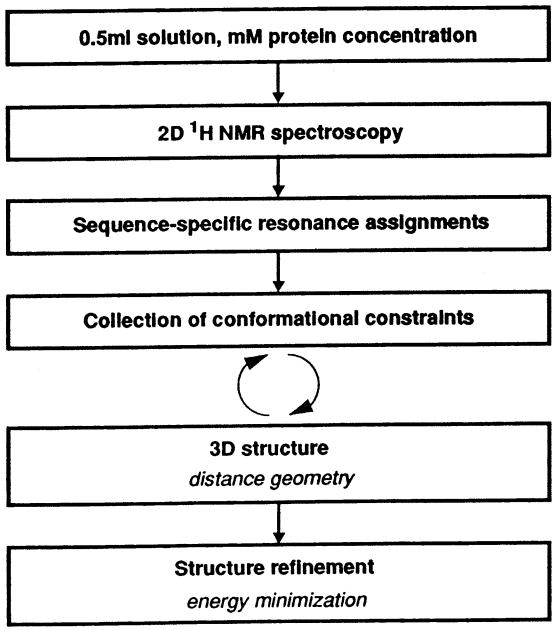


**Fig. 2.** Outline of the procedure described in this paper for NMR structure determination of small proteins (MW ≤ 10,000) in solution.



appearing in the spectra. Most proteins, especially larger ones, will not form ideal solutions at such high concentrations and therefore often require particularly long measuring times to achieve a workable spectral quality.

Considerable care must be taken in the preparation of samples for NMR experiments, which is normally based on thorough examination of suitable solution (pH and concentration) and temperature conditions. Protein stability under a variety of solution conditions can be efficiently ascertained in most cases by circular dichroism (CD), ultraviolet, or fluorescence spectroscopy (28). The small proteins of the Kunitz protease inhibitor family are highly soluble and very thermostable in pure water. If, with other proteins, a buffer is necessary to maintain the native conformation, caution must be exercised in the choice of the buffer so as not to introduce additional proton resonances which might interfere with the spectral interpretation. Phosphate is an ideal choice for buffering of most NMR samples, as it contains no NMR-observable protons and buffers adequately in the pH region 3-6 where most protein NMR measurements are performed. Deuterated organic buffers are available with minimal residual proton contamination and can be used if needed. Inclusion of additional salts, such as sodium or potassium chloride, to protein solutions should be avoided unless essential to the stability of the sample. It should then be kept to under 0.1 M to avoid undesirable artifacts during the NMR measurements.

## NMR experiments

In practice, the data for the structure determination of a small protein are collected by performing homonuclear $^1$H 2D NMR experiments. Typically, these spectra contain an array of diagonal peaks in the 2D frequency plane (Fig. 3), with $\omega_1 = \omega_2$ , which display the chemical shift positions of the resonance lines and can be directly compared to a conventional 1D spectrum. In addition there are a large number of off-diagonal peaks (cross peaks) with $\omega_1 \neq \omega_2$. Through simple geometric patterns, each cross peak establishes a correlation between two diagonal peaks. For example, in 2D nuclear Overhauser enhancement spectroscopy (NOESY) the cross peaks represent nuclear Overhauser effects (NOEs) and indicate that the protons corresponding to the two correlated diagonal peaks are separated by only a short distance in the three-dimensional molecular structure, say less than 5.0 Å. 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, where they are separated by at most three chemical bonds.

A large number of different 2D NMR experiments utilizing COSY-type through-bond $^1$H-$^1$H connectivities are currently available, for example, multiple-quantum-filtered correla-

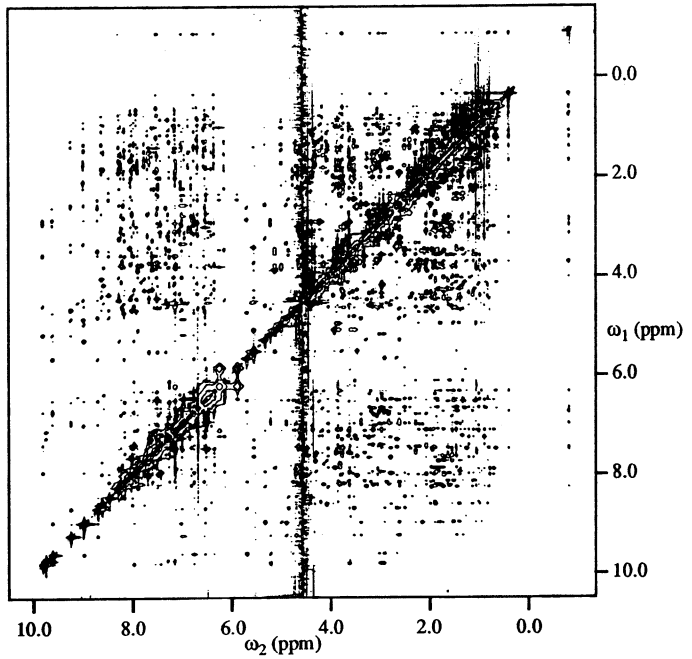


**Fig. 3.** Two-dimensional $^1$H NOESY spectrum of Toxin K (protein concentration ~10 mM, mixing time $\tau_m$ = 40 ms) in $H_2O$ at pH 4.6 and 36 °C. The noise band in the center of the spectrum is the result of residual perturbation from the suppression of the water resonance by selective irradiation.

tion spectroscopy (2QF-COSY (29) or 3QF-COSY (30)), multiple-quantum spectroscopy (31), or total correlation spectroscopy (TOCSY) (32, 33). Using a suitable selection of these experiments with $H_2O$ and/or $D_2O$ solutions of the protein in combination with NOESY for studies of through-space interactions, one can obtain a complete or nearly complete delineation of the $^1$H-$^1$H connectivities in a protein.

### Obtaining sequence-specific resonance assignments

In the first step of the spectral interpretation, sequence-specific resonance assignments must be obtained (Fig. 2). In other words, for each proton in the polypeptide chain (Toxin K has over 400 protons) the corresponding diagonal peak position must be identified. This is achieved by a combined analysis of through-bond COSY-type and through-space NOESY-type information using the sequential resonance assignment technique (3, 12, 34, 35). Briefly,



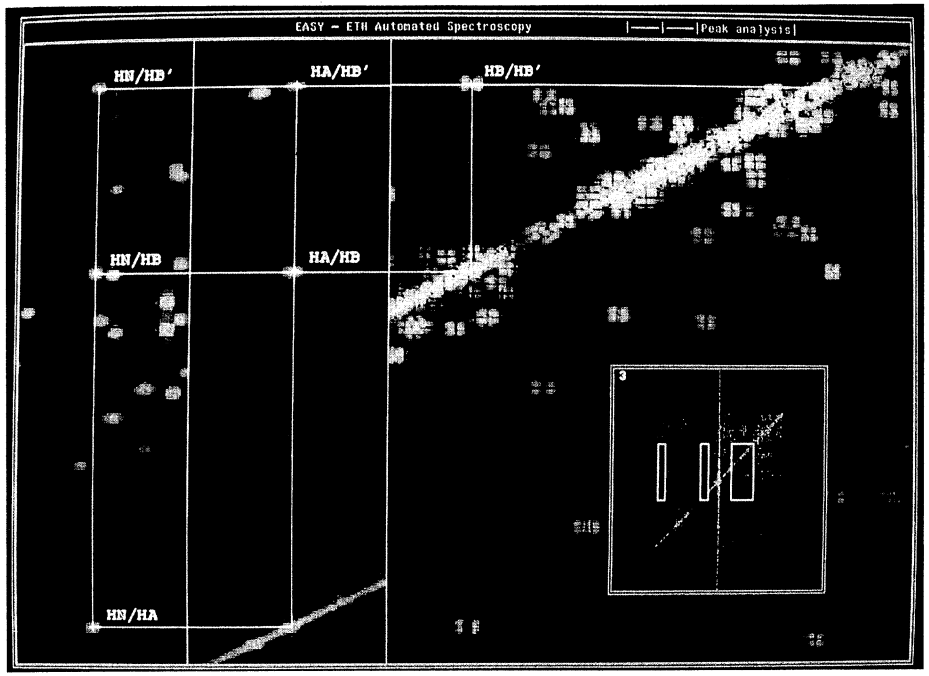


Figure 4. Combined display of TOCSY (left and center) and 2QF-COSY (right) spectral regions for interactive checks on spin system assignments for Toxin K. The insert in the lower right displays an entire spectrum, with the positions of the three zoomed regions indicated with boxes. Red and yellow colors represent positive and negative intensity in the intensity spectra, respectively. Yellow lines form the borders around the spectral regions, blue lines connect crosspeaks in the different spectral areas that belong to the same spin system. (see Appendix for a representation in colour)

for small proteins this involves first the identification of spin systems, *i.e.*, groups of protons (spins) connected by scalar (spin-spin) couplings, $J$, for each amino acid residue. Particular geometric patterns of cross peaks in the COSY-type spectra are characteristic for the different spin systems in the 20 naturally occurring amino acids. In all, the 20 amino acids give rise to 10 different COSY connectivity patterns for the aliphatic protons and four different patterns for the aromatic rings. By combining 2QF-COSY and TOCSY information, spin systems for all or nearly all of the amino acids can nowadays be identified. This points out one of the many advantages of using a computer to assist in the analysis of spectral data. In our laboratory, the EASY (ETH Automated Spectroscopy) program package was developed for this purpose (18). Patterns of cross peaks in the 2QF-COSY and TOCSY spectra can be searched with simple algorithms and possible spin system assignments automatically identified. Were it not for some inevitable chemical shift degeneracies leading to overlapping peaks in the 2D NMR spectra, spin system identification could be completely automated. However, in practice, a spectroscopist must be able to resolve interactively residual ambiguities. Figure 4 shows an AMX spin system from Toxin K as displayed with the EASY program software. By combining spectral regions of TOCSY and 2QF-COSY spectra simultaneously, possible spin system assignments suggested by the automated part of the program are displayed (as crosses denoting peak positions connected by blue lines) and are readily checked. In this way assignments for each proton in a spin system can be ascertained.

Obtaining sequence-specific assignments of the individual amino acid spin systems may be viewed as a two-step process, in which polypeptide segments of various lengths are first assembled based on the observation of characteristic sequential NOEs ($d_{\alpha N}$, $d_{NN}$, and $d_{\beta N}$ in Figure 5). Then the sequence-specific assignments are achieved by matching the sequence of these spectroscopically identified segments against the independently obtained

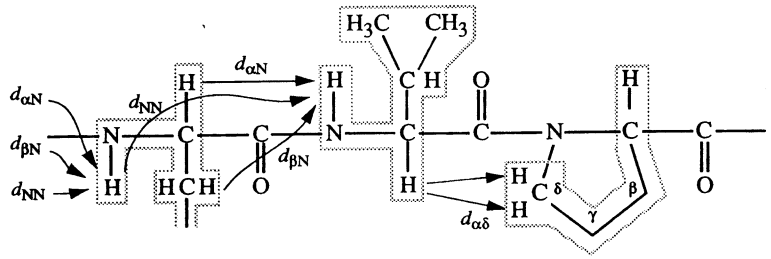


**Fig. 5.** Segment of a polypeptide chain with indication of the sequential NOE connectivities with amino acids and with proline (shown in the *trans* conformation). Stippled lines outline independent amino acid spin systems.



amino acid sequence. Special consideration must be given to proline, which will interrupt the conventional sequential segments as identified above due to its lack of an amide proton. Sequential connectivities for prolines can be established by observation of sequential $d_{\alpha\delta}$, $d_{N\delta}$, and $d_{\beta\delta}$ NOEs for the *trans* isomer, or sequential $d_{\alpha\alpha}$ and $d_{N\alpha}$ NOEs for the *cis* isomer (Fig. 5). The EASY program includes facilities for interactively establishing sequential assignment pathways, as well as a routine for automated sequential assignments which largely combines the two aforementioned steps by making reference to the amino acid sequence at each sequential addition of a spin system.

### Collection of conformational constraints

Once resonance assignments are available for all or nearly all polypeptide protons, one can start to collect the data needed for the determination of the 3D structure (Fig. 2). This data is of several types: NOEs which are converted into distance constraints, and spin-spin coupling constants which, when combined with intraresidual and sequential NOEs, restrict the dihedral angles $\phi$, $\psi$, and $\chi^1$. Thereby the most important input results from upper bounds on $^1$H-$^1$H distances derived from NOESY experiments like that shown in Figure 3. The intensity of a cross peak in a NOESY spectrum can be related to the distance, $r$, between the two correlated protons by an equation of the form

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

where the second term is a function of the correlation time, $\tau_c$, which accounts for the influence of motional averaging processes on the observed NOE. $\tau_c$ is governed by the combined effects of the overall rotational molecular motions, which depend on the size and shape of the protein, and on intramolecular mobility. Thus, $\tau_c$ may vary for different locations in a protein molecule, and therefore 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 sometimes difficult to exclude the possibility that individual NOEs might be partially quenched by competitive relaxation processes. In consideration of these potential pitfalls, only upper bounds on $^1$H-$^1$H distances are usually derived from NOE measurements, and the input for the structure calculations consists of an allowed distance range bounded by this NOE upper limit and a lower limit equal to the sum of the van der Waals radii of the two interacting protons, which is usually taken to be 2.0 Å.

Analysis of NOESY spectra acquired in both $H_2O$ and $D_2O$ solution of Toxin K resulted in a total of 1080 unambiguously assigned cross peaks. Removal of irrelevant distance

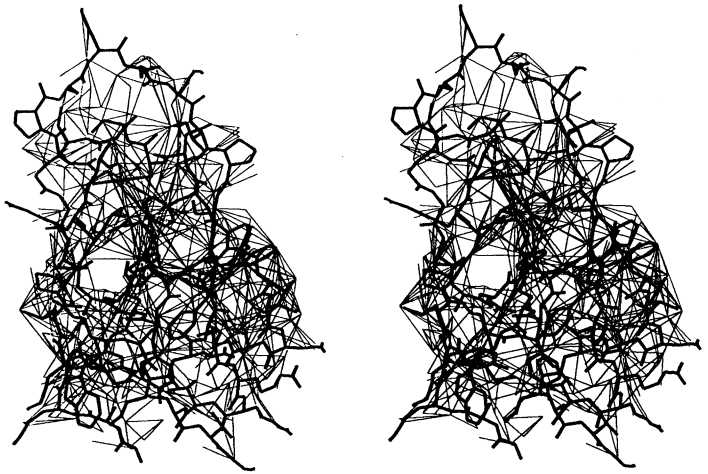

**Fig. 6.** Stereo view of an all-heavy-atom representation of Toxin K (bold lines). Each of the 809 upper distance constraints used in the calculation of this structure is shown by a thin line connecting the two protons (or pseudo-atoms) involved in the constraint. The lack of distance constraints in left-center of the figure coincides with the position of a cavity containing two internal water molecules in Toxin K (17).

constraints (*i.e.*, the distance in question is either independent of the conformation, or there exists no conformation that would violate the constraint) and the application of corrections for pairs of diastereotopic substituents without individual stereospecific assignments (36) reduced the original number of upper distance constraints to 809. The distribution of these constraints can be seen in Figure 6 which shows all distance constraints as thin lines connecting the respective protons in the solution structure of Toxin K.

Constraints on the dihedral angles $\phi$ and $\psi$ are important supplementary input for the structure determination. Measurements of vicinal spin-spin couplings, $^3J_{HN\alpha}$, present a valuable complementation of NOE distance data for studies of the polypeptide backbone conformation in proteins stemming from the relation between the magnitude of the coupling and the intervening dihedral angle $\phi$ (3, 37). Valuable information pertaining to the side chain conformation can be obtained from measuring the vicinal spin-spin couplings $^3J_{\alpha\beta}$, which place restrictions on the dihedral angle $\chi^1$. Using the HABAS routine in our software package, constraints for the dihedral angles $\phi$, $\psi$, and $\chi^1$ as well as stereospecific assignments for $\beta$-methylene protons can be obtained by a grid search of the allowed conformations, whereby these coupling constants are combined with the availablr local NOE distance constraints (21).



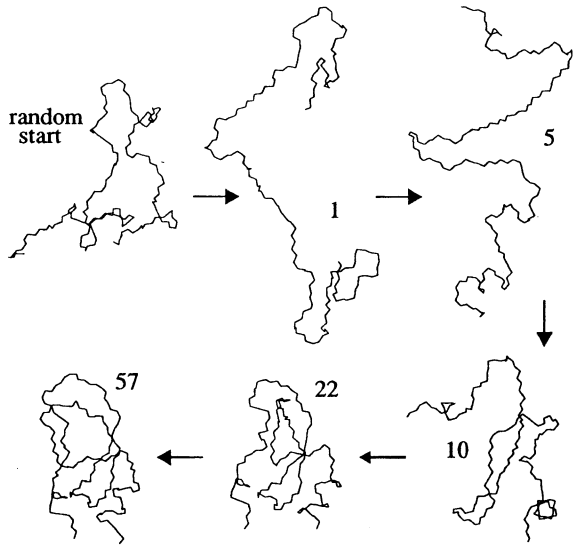


**Fig. 7.** Randomized starting structure, intermediate structures and final, converged structure of Toxin K during minimization with the distance geometry program DIANA. The backbone atoms N, $C^\alpha$, and C' for all 57 residues of the random start conformation, and of the conformations at the end of the minimization levels $l = 1, 5, 10, 22$, and 57 are shown (at level $l$, constraints between residues $i$ and $j$, with $|i\text{-}j| \leq l$ are included in the minimization).

## Calculation of the three-dimensional protein structure

A protein structure determination from NMR data using the distance geometry program DIANA (19, 23) begins with a random polypeptide conformation (upper left in Fig. 7). Under the influence of the experimental constraints, the variable target function method (38) used in DIANA 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 six selected stages in this optimization procedure with Toxin K are shown in Figure 7. At level 1, only intraresidual and sequential constraints are imposed on the random structure. Some turns that are subsequently found in the final structure may be recognized, but these are only strictly localized deviations from randomness. By level 5 (upper distance constraints from residues $i$ to $i + 4$), elements of helical secondary structure emerge. By level 10, the central $\beta$-hairpin is largely complete, and by level 22 the familiar global polypeptide fold is clearly visible. (The series of snapshots shown in Figure 7 is suggestive of how a small, single-domain protein might adopt its tertiary structure

(neglecting disulfide bonds), whereby local interactions responsible for the acquisition of local structure are important early in the folding process, followed by interaction of these preformed structures and finally coalescence into a compact tertiary fold.) The result of a single distance geometry calculation (Fig. 7) represents one conformer that is compatible with the NMR experiments. One cannot tell *a priori* if this conformation is unique, and therefore the structure calculation is repeated a number of times (typically 50 to 100 calculations) with different random starting conditions to ensure a good sampling of the conformation space defined by the NMR data set.

Assignment of interresidual cross peaks in the 2D NOESY spectra is an important and often tedious endeavor if chemical shift degeneracies result in excessive cross peak overlap. Were it not for these inevitable chemical shift degeneracies and the usually somewhat imprecise cross peak positional information, all assignments could of course be made from knowledge of the chemical shifts resulting from the sequence-specific resonance assignments. In practice, only a fraction of the NOESY cross peaks can be unambiguously identified in this direct way and used to generate a group of preliminary, "low resolution" conformers. In a subsequent step, these preliminary conformers are used to reduce the number of heretofore ambiguous assignments resulting from chemical shift degeneracies by eliminating possible assignments involving two protons which, on the basis of the preliminary structures, are beyond a maximum distance cutoff for the observation of NOEs. The program ASNO ("Assign NOEs") was developed to assist the assignment of NOE cross peaks based on both chemical shift values and reference to a set of preliminary three-dimensional conformers (17). Assignments made in this way can be examined interactively using the EASY program package for spectral display (Fig. 4).

As an illustration, the overall effect of increasing the number of NOE distance constraints for Toxin K is visualized in Figure 8, where four groups of 10 conformers each, selected at different stages of the structural refinement process, are displayed. A preliminary data set containing a total of 322 mostly interresidual and sequential upper distance constraints (approximately 40% of the final number) was obtained using only chemical shift information for the assignment of NOESY cross peaks. Using this data combined with explicit disulfide bond constraints and dihedral angle constraints from the program HABAS, a group of 100 conformers was calculated from randomized starting conformers with the program DIANA. Figure 8A shows the ten best conformers, identified on the basis of having the lowest residual target function. The RMSD value relative to the mean structure of 1.7 Å for the backbone heavy atoms in this group of 10 conformers is clearly indicative of a low-quality structure (3, 39). Although the conformers of Figure 8A do not form a tight bundle, the global



polypeptide fold is nonetheless clearly defined. In the first cycle of refinement, these conformers were used by the program ASNO to resolve ambiguities due to chemical shift degeneracies, which resulted in an approximately two-fold increase of the number of NOE distance constraints. The resulting data set now contained a total of 657 constraints from NOESY cross peaks. A new group of 100 conformers was again calculated from random starting conformers. The improvement in the precision of the structure after one cycle of refinement is visualized in Figure 8B. With the exception of some chain-terminal residues, the structure is now reasonably well defined. Further improvements of the structure by performing additional cycles of refinement, are mostly confined to the regions between the regular secondary structural elements (Fig. 8, C and D).

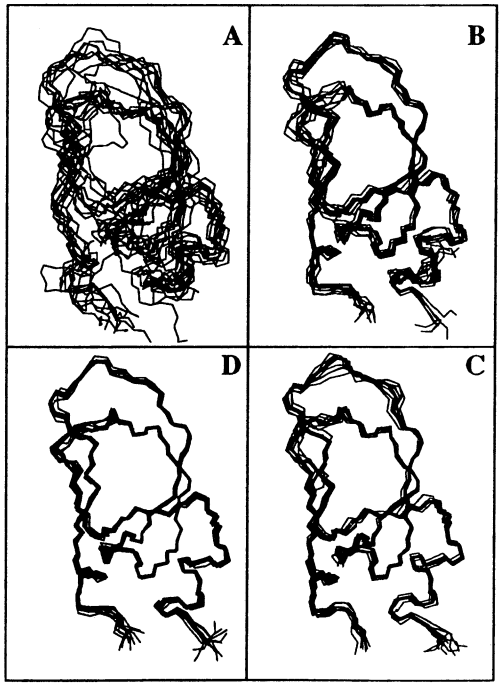


**Fig. 8.** View of the polypeptide backbone of the 10 best DIANA structures of Toxin K at four stages of structure refinement. (A) 322 NOE upper distance constraints (RMSD to the mean structure: $1.66 \pm 0.40$ Å). (B) 657 NOE upper distance constraints (RMSD to the mean structure: $0.65 \pm 0.10$ Å). (C) 747 NOE upper distance constraints (RMSD to the mean structure: $0.38 \pm 0.09$ Å). (D) 809 NOE upper distance constraints (RMSD to the mean structure: $0.32 \pm 0.07$ Å).

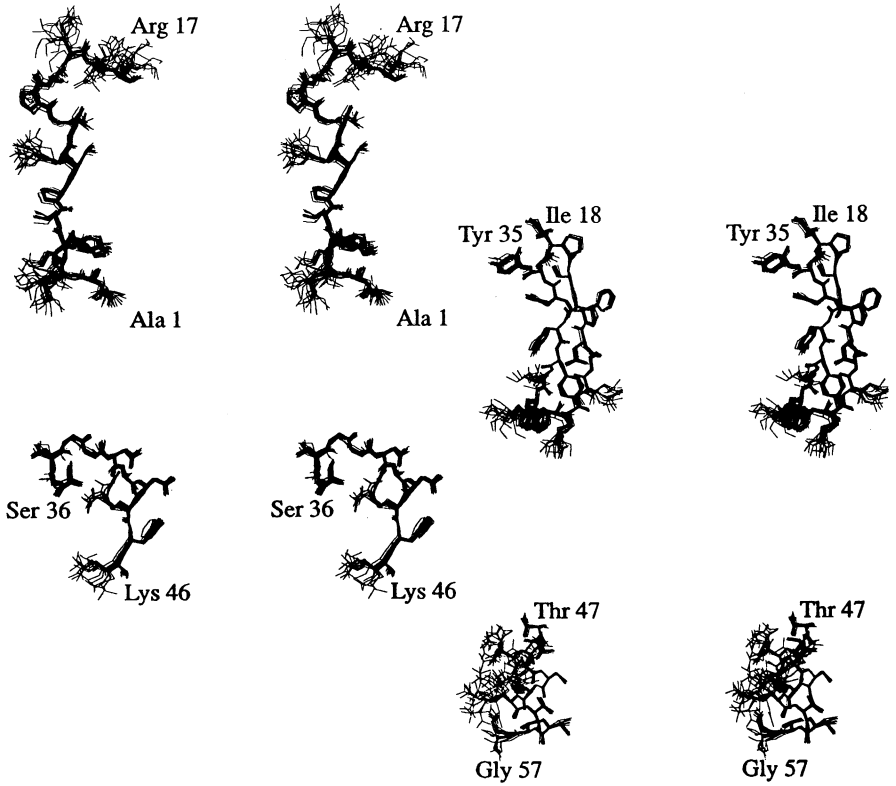


**Fig. 9.** Stereo views of an all-heavy-atom presentation of the solution structure of Toxin K. Local superpositions of the 20 conformers used to represent the solution structure was made for four polypeptide segments 1-17, 18-35, 36-46 and 47-57.

### Evaluation of the quality of protein structures determined by NMR in solution

The Figures 8-10 illustrate different facets of protein structures calculated from NMR data. A structure determination always uses the complete polypeptide chain with the amino acid side chains. A physically meaningful presentation of the solution structure consists of a superposition of a group of conformers calculated from different starting structures with the same NMR data set. For clarity only the polypeptide backbone atoms N, $C^{\alpha}$, and C' are usually drawn. The spread among the different individual distance geometry solutions (Figs. 8 and 9) occurs because the experimental input consists of allowed distance ranges rather than exact distances, and because there may be only a few constraints for some areas of the molecule.



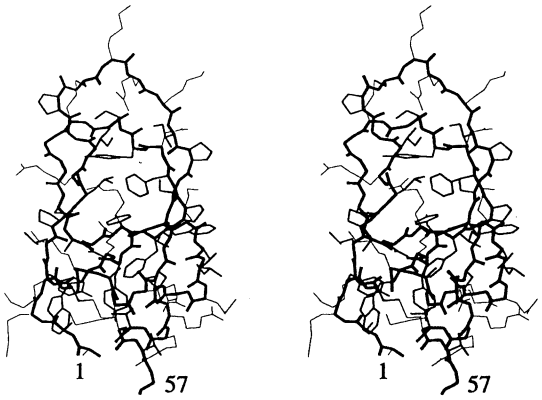


**Fig. 10.** Stereo view of the all-heavy atom representation of the best DIANA conformer (*i.e.*, the one with the smallest residual target function value) of Toxin K after energy refinement. The backbone atoms N, $C^{\alpha}$, C' and O' are connected with thick lines, best-defined side chains are drawn with lines of medium thickness and other side chains are drawn with thin lines.

A critical assessment of a NMR structure determination can be based on the facts that nearly complete sequence-specific resonance assignments are indispensable as a basis for a structure determination, that the quality of the structure determination is improved if stereospecific assignments are obtained for prochiral centers, that about 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 only small residual violations of the NMR and steric constraints. If the structure calculation is repeated a number of times, each with a different starting condition, a high-quality structure determination is reflected by a tight bundle of conformers, which corresponds to a small value for the average of the root mean square deviation (RMSD) values between the mean solution structure and each individual conformer. The variable precision in the definition of the side chain conformations can be appreciated from Figure 9, where all 20 final solution conformers are displayed in segments along the polypeptide chain of Toxin K. A number of side chain conformations, usually in the core of the molecule, are determined with a precision comparable to that of the polypeptide backbone (Fig. 8D). It is therefore possible to define a group of "best-defined" side chains for which inclusion of the side chain heavy atoms into the calculation of the RMSD after a global superposition does not significantly increase the RMSD value relative to that for the backbone atoms alone. A stereo view of an all-heavy-atom representation of a single DIANA conformer is shown in Figure 10 with the "best-defined" side chains highlighted.

In summary, for the presently used example, Toxin K, complete sequence-specific $^1H$ resonance assignments were obtained, with individual stereospecific assignments for 35 pairs of diastereotopic substituents. A total of 809 conformational constraints were obtained from the NOESY spectra, and HABAS produced 123 dihedral angle constraints (44 for $\phi$, 44 for $\psi$, and 35 for $\chi^1$). On average, each of the 20 final conformers had less than one residual NOE distance constraint violation larger than 0.1 Å, none of which exceeded 0.11 Å. The average of the pairwise RMSD values between each of the 20 final conformers and the mean solution structure is 0.36 Å for the backbone atoms N, $C^\alpha$, C', 0.36 Å for the backbone atoms plus the heavy atoms of the best-defined side chains, and 0.92 Å for the complete polypeptide structure. These numbers are indicative of a high-quality NMR structure determination, where the backbone and the best-defined side chains are defined with a similar precision as in a 2.0 Å X-ray crystal structure.

## Acknowledgments

We are indebted to many colleagues mentioned in the references, who have contributed to the development of the presently described approach to NMR structure determination. We thank R. Marani for help in preparation of the manuscript. Financial support by the Schweizerischer Nationalfonds (project 31.32033.91) is gratefully acknowledged.



## References

1. Schulz, G.E. & Schirmer, R.E. (1979) "Principles of Protein Structure" Springer-Verlag, New York
2. Blundell, T.L. & Johnson, L.N., (1976) "Protein Crystallography", Academic Press, New York
3. Wüthrich, K., "NMR of Proteins & Nucleic Acids" (1986) Wiley, New York
4. Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982) J. Mol. Biol. 155, 311– 319
5. A. Masson, A. & Wüthrich, K. (1973) FEBS Lett. 31, 114-118
6. Wüthrich, K. & Wagner, G. (1975) FEBS Lett. 50, 265-268.
7. Nagayama, K., Wüthrich, K., Bachmann, P. & Ernst, R.R. (1977) Biochem. Biophys. Res. Comm. 78, 99-105
8. Anil Kumar, Wagner, G., Ernst, R.R. & Wüthrich, K. (1981) J. Amer. Chem. Soc. 103, 3654-3658
9. Otting, G. & Wüthrich, K. (1989) J. Amer. Chem. Soc. 111, 1871-1875
10. Dubs, A., Wagner, G. & Wüthrich, K. (1979) Biochem. Biophys. Acta 577, 177-194
11. Wagner, G., Anil Kumar & Wüthrich, K. (1981) Eur. J. Biochem. 114, 375-384
12. Wagner, G. & Wüthrich, K. (1982) J. Mol. Biol. 155, 347-366
13. Štrop, P., Čechová, D. and Wüthrich, K. (1983) J. Mol. Biol. 166, 669-676
14. Williamson, M.P., Marion, D. & Wüthrich, K. (1984) J. Mol. Biol. 173, 341-359
15. Williamson, M.P., Havel, T.F. & Wüthrich, K. (1985) J. Mol. Biol. 182, 295-315
16. Berndt, K. D., Güntert, P., Orbons, L.P.M. & Wüthrich, K. (1992). J. Mol. Biol. (in press)
17. Berndt, K.D., Güntert, P. & Wüthrich, K. (manuscript in preparation)
18. Eccles, C., Güntert, P., Billeter, M. & Wüthrich, K. (1991) J. Biomol. NMR 1, 111-130
19. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1991) J. Mol. Biol. 217, 517-530
20. Güntert, P., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1991) J. Mol. Biol. 217, 531-540
21. Güntert, P. Braun, W., Billeter, M. & Wüthrich, K. (1989) J. Amer. Chem. Soc. 111, 3997-4004
22. Szyperski, T., Güntert, G., Otting, G. & Wüthrich, K. (1992) J. Magn. Reson. (in press)

23. Güntert, P. & Wüthrich, K. (1991) J. Biomol. NMR 1, 447-456

24. Strydom, D.J. (1973) Nature New Biol. 243, 88-89

25. Harvey, A.L. & Karlsson, E. (1982) Br. J. Pharmac. 77, 153-161

26. Bax, A. (1989) Ann. Rev. Biochem. 58, 223-256

27. Clore, G. M. & Gronenborn, A. M. (1991) Science 252, 1390-1399

28. Pace, C.N., Shirley, B.A. & Thompson, J.A. (1989) In: Protein Structure: A Practical Approach, (T.E. Creighton, ed.) IRL Press at Oxford University Press, Oxford. pp. 311-330

29. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. & Wüthrich, K. (1983) Biochem. Biophys. Res. Comm. 117, 479-485

30. Müller, N., Ernst, R.R. & Wüthrich, K. (1986) J. Amer. Chem. Soc. 108, 6482- 6492

31. Wagner, G. & Zuiderweg, E.P.R. (1983) Biochem. Biophys. Res. Commun. 113, 854-860

32. Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1987) "Principles of Nuclear Magnetic Resonance in One- and Two-dimensions" Clarendon Press, Oxford

33. Griesinger, C., Otting, G., Wüthrich K. & Ernst, R.R. (1988) J. Amer. Chem. Soc. 110, 7870-7872

34. Billeter, M., Braun, W. & Wüthrich, K. (1982) J. Mol. Biol. 155, 321-346

35. Wider, G., Lee K.H. & Wüthrich, K. (1982) J. Mol. Biol. 155, 367-388

36. Wüthrich, K., Billeter M. & Braun, W. (1983) J. Mol. Biol. 169, 949-961

37. Karplus, M. (1963) J. Amer. Chem. Soc. 85, 2870-2871

38. Braun, W. & Gō, N. (1985) J. Mol. Biol. 186, 611-626

39. Wüthrich, K. (1989) Science 243, 45-50