5

Calculating Protein Structures from NMR Data

Peter Güntert

1. Introduction

Today many, if not most, NMR measurements with proteins are performed with the ultimate aim of determining their three-dimensional (3D) structure *(1)*. However, NMR is not a "microscope with atomic resolution" that would directly produce an image of a protein. Rather, it is able to yield a wealth of indirect structural information from which the 3D structure can be revealed only by extensive calculations. The pioneering first structure determinations of proteins in solution (e.g., *2–6*) were year-long struggles, both fascinating and tedious because of the lack of established NMR techniques and numerical methods for structure calculation, and hampered by limitations of the spectrometers and computers of the time. Recent experimental, theoretical, and technological advances —and the dissemination of the methodological knowledge—have changed this situation completely: Given a sufficient amount of a purified, water-soluble protein with less than approx 200 amino acid residues, its 3D structure in solution can be determined routinely by the NMR method. Protein structures with up to about 100 residues can be solved by [^{1}H]-NMR alone, whereas for larger proteins labeling with ^{13}C and ^{15}N is required.

This chapter reviews the computational methods for solution structure determination of proteins from a practical point of view. This chapter will also summarize the key NMR data carrying structural information that can be exploited readily in a structure calculation; treat the conversion from this NMR data to geometric conformational restraints; explain the preliminaries of a structure calculation, such as the systematic analysis of local conformation and stereospecific assignments; concentrate on the currently used structure

From: *Methods in Molecular Biology, Vol. 60: Protein NMR Techniques*
Edited by: D. G. Reid Humana Press Inc., Totowa, NJ

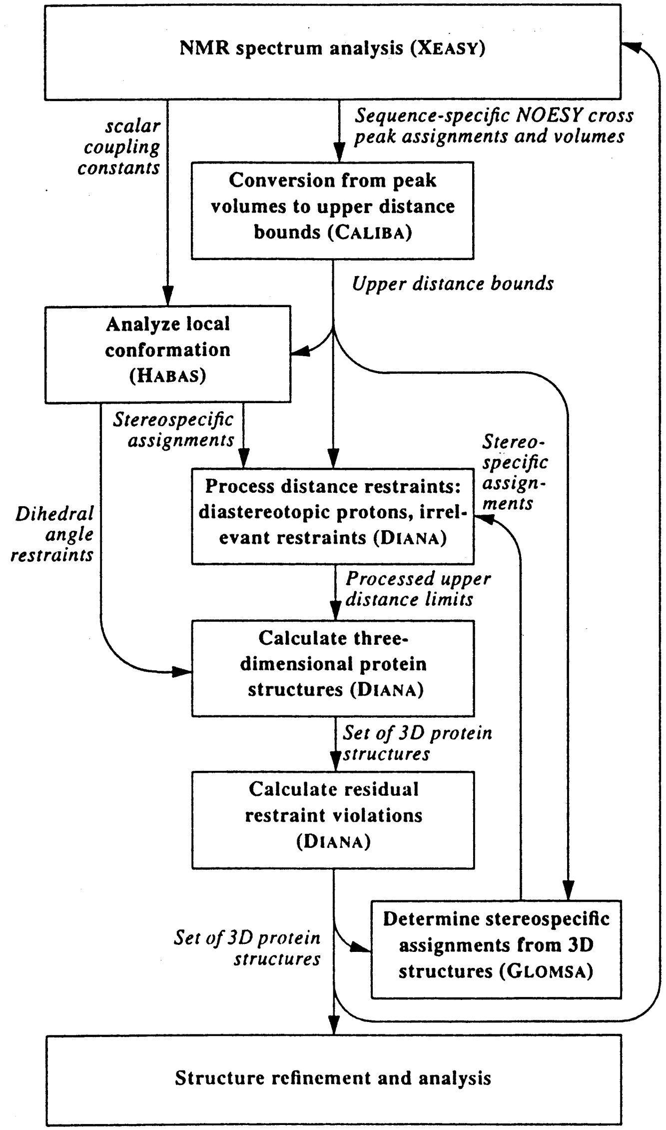

Fig. 1. Flowchart of a protein structure calculation. Various steps of a structure calculation are represented by boxes, and arrows indicate the flow of data. In parentheses, the names of computer programs in the DIANA and XEASY program packages *(51,98)* that perform the corresponding step are given.

calculation algorithms; discuss general ways of analyzing NMR solution structures of proteins; and, finally, give a brief overview on structure refinement methods (Fig. 1).

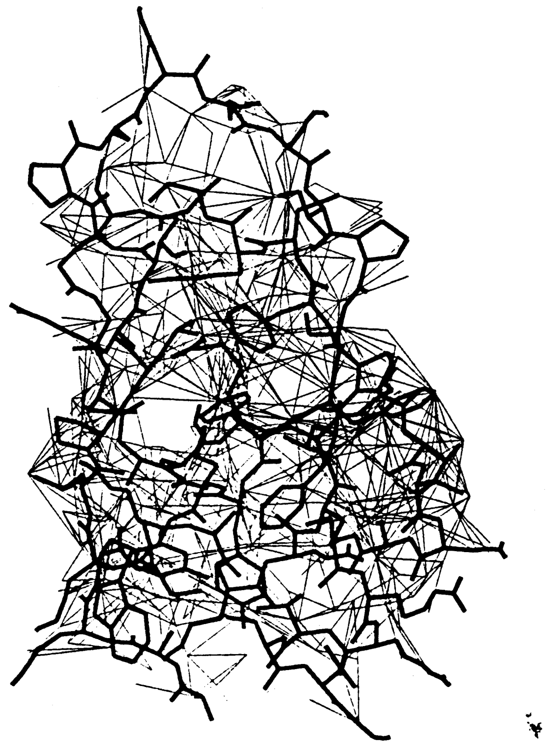

Fig. 2. Representation of the network of NOE distance restraints used for the structure calculation of the protein toxin K from the black mamba, *Dendroaspis polylepis polylepis (97)*. Covalent bonds between nonhydrogen atoms are shown as bold lines. Each of the 809 upper distance bounds is indicated by a thin line connecting the two atoms (hydrogens or pseudoatoms) involved in the restraint.

2. NMR Data for Protein Structure Calculation

2.1. Nuclear Overhauser Effects

The NMR method for protein structure determination relies on a dense network of distance restraints derived from nuclear Overhauser effects (NOEs) between nearby hydrogen atoms in the protein (*1,7–9;* Fig. 2). NOEs are the essential NMR data to define the secondary and tertiary structure of a protein because they connect hydrogen atoms separated by less than about 5 Å (1 Å = 0.1 nm) in amino acid residues that may be far away along the protein sequence but close together in space. The NOE reflects the transfer of magnetization between spins coupled by the dipole–dipole interaction in a molecule that undergoes Brownian motion in a liquid *(10–14)*. The intensity of a NOE, that is the volume, *V*, of the corresponding crosspeak in a NOESY

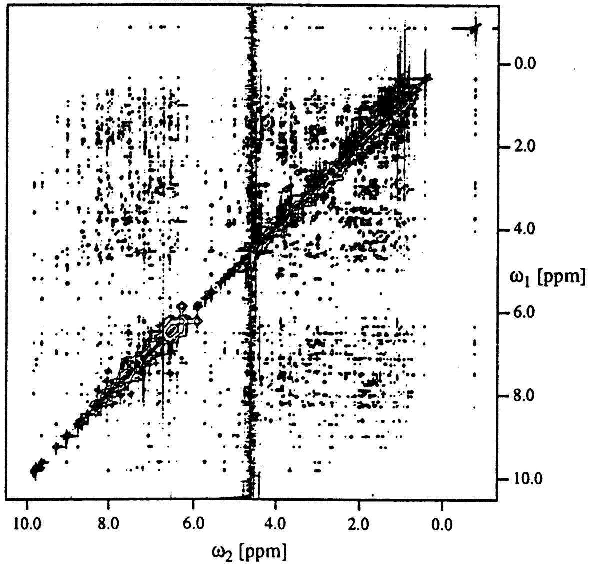

Fig. 3. Two-dimensional ^{1}H NOESY spectrum of toxin K in H_2O at pH 4.6 and 36°C (protein concentration 10 m*M*, mixing time τ_m = 40 ms; *97*).

spectrum (*15;* Fig. 3), is related to the distance, *r*, between the two interacting spins by

$$V \propto \langle r^{-6} \rangle f(\tau_c) \tag{1}$$

The averaging indicates that in molecules with inherent flexibility the distance *r* may vary and thus has to be averaged appropriately. The remaining dependence of the magnetization transfer on the motion enters through the function $f(\tau_c)$ that includes the effects of global and internal motions of the molecule in a nontrivial way. Since globular proteins are relatively rigid—with the exceptions of the protein surface and disordered segments of the polypeptide chain—often it is assumed that there exists a single rigid conformation that is compatible with all NOE data simultaneously, provided that the NOE data are interpreted in a conservative, semiquantitative manner *(1)*. More sophisticated treatments that take into account that the result of a NOESY experiment represents an average over time and space usually are deferred to the structure refinement stage *(16,17)*.

In principle, all hydrogen atoms of a protein form a single network of spins, coupled by the dipole–dipole interaction. Magnetization can be transferred from one spin to another not only directly but also via other spins in the vicinity—an effect called spin diffusion *(13,18,19)*. The approximation of isolated spin pairs is only valid for short mixing times in the NOESY experiment. However, the mixing time cannot be made arbitrarily short because (in the limit of short mixing times) the intensity of a NOE is proportional to the mixing time *(20)*. In practice, a compromise has to be made between the suppression of spin diffusion and sufficient crosspeak intensities, usually with mixing times in the range of 40–80 ms. Spin diffusion effects can also be included in the structure calculation by complete relaxation matrix refinement *(21–24)*. However, since assumptions about internal and overall motions that hardly can be confirmed experimentally also enter into the relaxation matrix refinement, care has to be taken not to bias the structure determination by overinterpretation of the data.

The quantification of a NOE is equivalent to determining the volume of the corresponding crosspeak in the NOESY spectrum *(19)*. Since the linewidths can vary appreciably for different resonances, crosspeak volumes should be determined by integration over the peak area rather than by measuring peak heights, for example, by counting contour lines. For isolated crosspeaks integration is straightforward, and for clusters of overlapping crosspeaks deconvolution methods have been proposed to distribute the total volume among the individual signals *(25)*. Although the reliable quantification of NOEs is important to obtain a high-quality protein structure, one should also keep in mind that, according to Eq. (1), the relative error of the distance estimate is only one-sixth of the relative error of the volume determination.

2.2. Scalar Coupling Constants

A second source of structural information are vicinal scalar coupling constants between atoms separated by three covalent bonds from each other *(26)*. These scalar coupling constants, ^3J, are related to the enclosed dihedral angle, θ, by Karplus relations *(27)*. For the structure determination of proteins the most important Karplus relations are *(26,28–30)*:

$$\begin{aligned} {}^3J_{HN\alpha}(\theta) &= 6.4\cos^2\theta - 1.4\cos\theta + 1.9 \\ {}^3J_{\alpha\beta}(\theta) &= 9.5\cos^2\theta - 1.6\cos\theta + 1.8 \\ {}^3J_{N\beta}(\theta) &= -4.4\cos^2\theta + 1.2\cos\theta + 0.1 \\ {}^3J_{C'\beta}(\theta) &= 8.0\cos^2\theta - 2.0\cos\theta \end{aligned} \tag{2}$$

$^3J_{HN\alpha}$ denotes the scalar coupling constant between a backbone amide proton and an α-proton, $^3J_{\alpha\beta}$ between an α- and a β-proton, $^3J_{N\beta}$ between a backbone nitrogen and a β-proton, and $^3J_{C'\beta}$ between a backbone carbonyl and a β-proton. All coupling constants are given in Hertz. In contrast to distance restraints derived from NOESY spectra, scalar coupling constants give information only on the local conformation of a polypeptide chain. They are, nevertheless, important to define accurately the local conformation, to obtain stereospecific assignments for diastereotopic protons (usually for the β-protons), and to detect dihedral angles (usually χ^1) that occur in multiple states.

Scalar couplings are manifested in the crosspeak fine structures of most NMR spectra *(19)*. Many NMR experiments have been proposed for the measurement of scalar coupling constants *(31)*. Scalar coupling constants conventionally are measured from the separation of fine-structure components in antiphase spectra. One has to be aware, however, of the cancellation effects between positive and negative fine-structure elements that lead both to an overestimation of the coupling constant and to a decrease of the overall crosspeak intensity *(32)*. These effects inhibit the determination of coupling constant values that are much smaller than the line-width from antiphase crosspeaks. The cancellation effects can be reduced in E. COSY type spectra where the crosspeak fine-structure is simplified by suppression of certain components of the fine-structure *(33)*. Other methods to determine coupling constants rely on a series of spectra with crosspeak volumes modulated by the coupling constant *(34)* or on in-phase spectra *(35)*. In general, scalar coupling constants can be determined in proteins with an accuracy of up to about ±1 Hz, but one has to keep in mind when interpreting them with the use of Eq. (2) that there is averaging because of internal mobility and that both the functional form and the parameters of the Karplus curves are approximations. This second source of error usually limits the applicable accuracy to approximately ±2 Hz.

2.3. Other NMR Data

NOEs and scalar coupling constants are the NMR data that most directly provide structural information. Additional NMR parameters that are sometimes used in structure determinations include hydrogen exchange data and chemical shifts, in particular for $^{13}C^{\alpha}$. Slow hydrogen exchange indicates that an amide proton is involved in a hydrogen bond *(9)*. Unfortunately, the acceptor oxygen or nitrogen atom cannot be identified directly by NMR, and one has to rely on NOEs in the vicinity of the postulated hydrogen bond or on assumptions about regular secondary structure to define the acceptor. Hydrogen bond restraints are thus either largely redundant with the NOE network or involve structural assumptions, and they should only be used in special situations, for instance, if not enough NOE data are available in preliminary structure calculations of larger proteins.

It was recognized that the deviations of $^{13}C^{\alpha}$ (and, to some extent, $^{13}C^{\beta}$) chemical shifts from their random coil values are correlated with the local backbone conformation *(36,37)*: $^{13}C^{\alpha}$ chemical shifts larger than the random coil values tend to occur for amino acid residues in α-helical conformation, whereas deviations toward smaller values are observed for residues in β-sheet conformation. Such information can be included in a structure calculation by restricting the local conformation of a residue to the α-helical or β-sheet region of the Ramachandran plot, although care should be applied because the correlation between chemical shift deviation and structure is not perfect. Similar to hydrogen bond restraints, conformational restraints based on $^{13}C^{\alpha}$ chemical shifts therefore should be used only as auxiliary data in special situations.

3. Conformational Restraints

For use in a structure calculation, geometric conformational restraints have to be derived from the NMR parameters. These geometric restraints should, on the one hand, convey to the structure calculation as much as possible of the structural information inherent in the NMR data, and, on the other hand, be simple enough to be used efficiently by the structure calculation algorithms. Therefore, predominantly distance and dihedral angle restraints are used in practice.

3.1. Distance Restraints

On the basis of Eq. (1), NOEs are usually treated as upper bounds on interatomic distances rather than precise distance restraints because the presence of internal motions and, possibly, chemical exchange may diminish the strength of a NOE *(19)*. In fact, much of the robustness of the NMR structure determination method is owing to the use of upper distance bounds instead of exact distance restraints in conjunction with the observation that internal motions and exchange effects usually reduce rather than increase the NOEs *(1)*. For the same reason, the absence of a NOE should not be interpreted as a lower bound on the distance between the two interacting spins. Certain NOEs, however, may also be enhanced by internal motions or chemical exchange and then be incompatible with the assumption of a rigid structure that fulfills all NMR data simultaneously *(17,38)*.

The upper bounds, u, are derived from the corresponding NOESY crosspeak volumes, V, according to calibration curves, $V = f(u)$, for example, assuming a rigid molecule,

$$V = k/u^6 \tag{3}$$

Here, k denotes a constant that depends on the arbitrary scaling of the NOESY spectrum. This constant is determined on the basis of known distances, for example, $d_{\alpha N}$ and d_{NN} in regular secondary structure elements *(39)*, or by reference to a preliminary structure.

In practice, it has been observed that flatter calibration curves, for example, of the type

$$V = k/u^n \tag{4}$$

with $n = 4$ or 5, may often give a better representation of the volume-to-distance relationship, in particular for NOEs that involve peripheral side chain protons *(40)*. The uniform average model *(2)* provides another, very conservative, calibration curve by making the assumption that, because of internal motions, the

Table 1
Repulsive Core Radii Used by the Program DIANA[a]

Atom type	Radius, Å
Amide hydrogen	0.95
Other hydrogen	1.00
Aromatic carbon	1.35
Other carbon	1.40
Nitrogen	1.30
Oxygen	1.20
Sulfur	1.60

[a]From Braun and Gō *(61)* and Güntert et al. *(51)*.

interatomic distance, r, assumes all values between the steric lower limit, l, and an upper limit, u, with equal probability:

$$V = k/(u-l)\int_l^u dr/r^6 = k'\,[1/l^5 - 1/u^5]/(u-l) \tag{5}$$

In practice, either the upper distance bounds obtained from Eqs. (3–5) are directly used as distance restraints, or they are classified into the three classes of strong, medium, and weak crosspeaks *(6,41)*, with corresponding upper limits of, typically, 2.7, 3.3, and 5.0 Å. NOEs that involve groups of protons with degenerate chemical shifts, in particular methyl groups, commonly are referred to pseudoatoms located in the center of the protons that they represent, and the upper bound is increased by a pseudoatom correction equal to the proton–pseudoatom distance *(42)*.

Hydrogen bonds also can be introduced into the structure calculation as distance restraints, typically by restraining the acceptor-hydrogen distance to 1.8–2.0 Å and the distance between the acceptor and the atom to which the hydrogen atom is covalently bound to 2.7–3.0 Å. The second distance restraint restricts the angle of the hydrogen bond.

Usually, a simple geometric force field is used for the structure calculation that retains only the most dominant part of the nonbonded interaction, namely, the steric repulsion in the form of lower bounds for all interatomic distances between pairs of atoms separated by three or more covalent bonds from each other. These steric lower bounds are generated internally by the structure calculation programs by assigning a repulsive core radius to each atom type (Table 1), and imposing lower distance bounds given by the sum of the two corresponding repulsive core radii. To allow the formation of hydrogen bonds, potential hydrogen bond contacts are treated specially with lower bounds that are smaller than the sum of the corresponding repulsive core radii.

Depending on the structure calculation program used, special covalent bonds, such as disulfide bridges or cyclic peptide bonds have to be enforced by distance restraints. For example, in the program DIANA disulfide bridges are fixed by restraining the distance between the two sulfur atoms to 2.0–2.1 Å and the two distances between the C^{β} and the sulfur atoms of different residues to 3.0–3.1 Å *(6)*.

3.2. Dihedral Angle Restraints

Dihedral angle restraints in the form of an allowed interval are used to incorporate scalar coupling information into the structure calculation. Using Eqs. (2), a given scalar coupling constant value gives in general rise to several (up to four) allowed intervals for the enclosed dihedral angle. However, most structure calculation programs allow only for a single allowed range for a dihedral angle. Using the smallest interval that encloses all dihedral angle values compatible with the scalar coupling constant often results in a loss of structural information because the dihedral angle restraint may encompass large regions that are forbidden by the measured coupling constant. Therefore it is often advantageous to combine local data—for example, all distance restraints and scalar coupling constants within the molecular fragment defined by the dihedral angles ϕ, ψ, and χ^1—in a systematic analysis of the local conformation and to derive dihedral angle restraints from the results of this grid search rather than from the individual NMR parameters *(43)*.

In addition, dihedral angle restraints may be used to restrict the conformation of the polypeptide chain, for example, to certain regions of the Ramachandran plot, on the basis of assumptions about regular secondary structure or $^{13}C^{\alpha}$ chemical shifts.

4. Preliminaries

4.1. Systematic Analysis of the Local Conformation

Before starting a structure calculation for the complete protein it is advisable to perform a systematic analysis of the local conformation in order to detect inconsistencies among the local conformational restraints, to derive dihedral angle restraints from the scalar coupling constant and local NOE data, and to obtain stereospecific assignments for diastereotopic protons and methyl groups. A systematic analysis of the local conformation is performed conveniently in dihedral angle space as a grid search over all sterically allowed combinations of dihedral angle values in a molecular fragment *(43,44)*. In practice, most of the available coupling constant and local NOE data involve the polypeptide backbone and the β-protons. They can be analyzed in a grid search

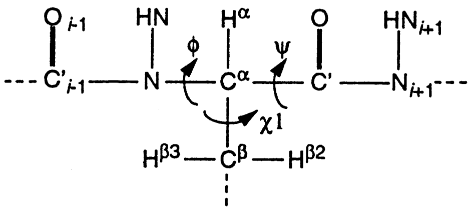

Fig. 4. Polypeptide fragment whose conformation can be analyzed systematically by a grid search over the dihedral angles ϕ, ψ, and χ^1 with the program HABAS *(43)*.

over the dihedral angles ϕ, ψ, and χ^1 of a given residue as it is, for instance, implemented in the program HABAS (Fig. 4).

4.2. Stereospecific Assignments

The standard method for obtaining resonance assignments in proteins *(45,46)* cannot provide stereospecific assignments, i.e., individual assignments for the two diastereotopic substituents of a prochiral center, for example, in methylene groups and in the isopropyl groups of valine and leucine. In the absence of stereospecific assignments, restraints involving diastereotopic substituents have to be referred to pseudoatoms *(42)*, or otherwise treated such that they are invariant under exchange of the two diastereotopic substituents, which inevitably results in a loss of information and less well-defined structures (*43;* Fig. 5). It is therefore essential for obtaining a high-quality structure that as many stereospecific assignments as possible are determined. Stereospecific assignments of valine and leucine isopropyl groups can be determined experimentally by biosynthetical fractional ^{13}C-labeling *(47,48)*. Stereospecific assignments for methylene protons have to be determined in the course of the structure calculation, either manually *(49)*, by systematic analysis of the local conformation around a methylene group, or by reference to preliminary 3D structures.

The local method, implemented, for example, in the program HABAS *(43)*, consists of two separate grid searches, one for each of the two assignment possibilities. An unambiguous stereospecific assignment results if allowed conformations occur only for one of the two possible assignments. This local method exclusively relies on scalar coupling constants and local distance restraints, for the stereospecific assignment of β-methylene protons with the program HABAS, for instance, on distance restraints and scalar coupling constants within the molecular fragment of Fig. 4. Assuming realistic error ranges for experimental data, generally it will not be possible to obtain unambiguous stereospecific assignments by the local method in all cases. Using complete simulated sets of local distance restraints and homonuclear coupling

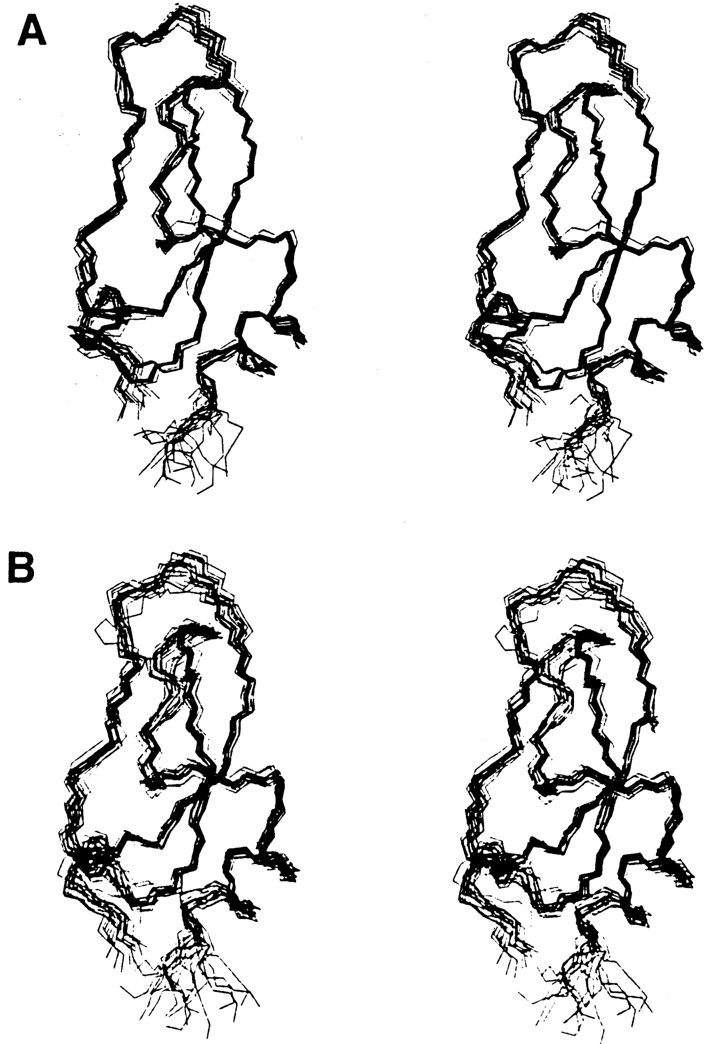

Fig. 5. The influence of stereospecific assignments on the precision of solution structures determined by NMR and distance geometry calculations, illustrated by two groups of 20 conformers of the protein BPTI, one calculated with the use of the 32 experimentally determined stereospecific assignments **(A)**, the other without **(B)**. In the stereo views, lines indicate covalent bonds between the polypeptide backbone atoms N, C^{α}, and C'.

constants with an accuracy of ±2 Hz, it was estimated that the program HABAS can yield unambiguous stereospecific assignments for about 50% of the β-methylene protons.

In contrast to the local method, global methods aim at the determination of stereospecific assignments either during the calculation of a 3D structure or by reference to preliminary 3D structures. They have the potential advantage over the local method that all conformational restraints, not only local ones, can be exploited, but, on the other hand, a systematic search of allowed conformations is no longer feasible, and the stereospecific assignments have to be based on a statistical analysis of a limited number of conformers. In conjunction with structure calculation programs working in Cartesian coordinate space, the so-called method of "floating stereospecific assignments" *(50)* can be used: At the beginning of a structure calculation a strong reduction of the corresponding potential energy terms allows the two diastereotopic substituents to interchange freely under the influence of the restraints before they later become fixed when the potential energy terms are restored slowly to their normal values (which inhibit an interchange of the diastereotopic substituents). A stereospecific assignment is considered to be unambiguous if it is consistently found in all conformers that were calculated. Another simple method for obtaining stereospecific assignments is implemented in the program GLOMSA of the DIANA package *(51)* and consists of the analysis of preliminary 3D structures: If there are two NOEs of significantly different strength from a given proton to both diastereotopic substituents of a prochiral center *and* if the distances from the given proton to the two diastereotopic substituents differ consistently in the structures, the stronger NOE can be identified with the diastereotopic substituent that is closer to the given proton.

4.3. Treatment of Distance Restraints to Diastereotopic Protons

Distance restraints involving diastereotopic substituents that could not be assigned stereospecifically have to be modified such that they are invariant under exchange of the two diastereotopic substituents. Traditionally, this is achieved by referring the restraints to a pseudoatom located centrally with respect to the two diastereotopic substituents and a concomitant increase of the upper distance bound, b_Q, by a pseudoatom correction, c_Q, equal to the distance from the pseudoatom to the individual protons, i.e., $b_Q = \min(b_1, b_2) + c_Q$ *(42)*. This approach, however, completely discards the weaker of the two possible NOEs from a given proton to the two diastereotopic substituents. In this case, an improved treatment implemented in the program DIANA *(51)* makes use of the information from both upper bounds, b_1 and b_2, by assigning a more restrictive upper limit, b_Q to the restraint to the pseudoatom,

$$b_Q = [(b_1^2 + b_2^2)/2 - c_Q^2]^{1/2} \tag{6}$$

and simultaneously imposing the weaker of the two upper bounds, $\max(b_1, b_2)$, on both distances to the individual diastereotopic substituents. Another

approach, used, for example, in the program XPLOR *(52)*, does not explicitly introduce a pseudoatom but imposes a distance restraint on the average distance, $\langle d \rangle$, rather than on the distances to the two diastereotopic substituents, d_1 and d_2. The average distance is calculated, for example, according to

$$\langle d \rangle = (d_1^{-6} + d_2^{-6})^{-1/6} \tag{7}$$

4.4. Removal of Irrelevant Distance Restraints

The number of experimental distance restraints used in a structure calculation is an important parameter that determines the accuracy of the resulting structure. To allow for meaningful comparisons it is therefore important to report the number of relevant distance restraints, i.e., of those actually that restrict the allowed conformation space, rather than the total number of NOESY crosspeaks that have been assigned. In addition, the removal of irrelevant distance restraints slightly increases the efficiency of the structure calculation. In practice, often more than half of the intraresidual and many sequential restraints are irrelevant. Those include restraints for fixed distances, for example, between geminal protons among the protons attached to an aromatic ring, and distance bounds that cannot be reached by any conformation, for example, an upper bound of 3.5 Å for the intraresidual distance between the amide- and the α-proton. Assuming rigid bond lengths and bond angles, the latter condition can be checked readily for distances that depend on one or two dihedral angles *(51)*. The distance d_{ij} between two atoms, i and j, that are separated by a single dihedral angle, a, is confined to the range

$$(A - B)^{1/2} \le d_{ij} \le (A + B)^{1/2} \tag{8}$$

where

$$A = \vec{d}_i^2 + \vec{d}_j^2 - 2(\vec{e}_a \cdot \vec{d}_i)(\vec{e}_a \cdot \vec{d}_j),\ B = 2\{[\vec{d}_i^2 - (\vec{e}_a \cdot \vec{d}_i)^2][\vec{d}_j^2 - (\vec{e}_a \cdot \vec{d}_j)^2]\}^{1/2} \tag{9}$$

In Eq. (9), $\vec{d}_i = \vec{r}_i - \vec{r}_a$, $\vec{d}_j = \vec{r}_j - \vec{r}_a$, $\vec{r}_i$ and $\vec{r}_j$ denote the position of the atoms i and j, and $\vec{r}_a$ and $\vec{e}_a$ denote the position of the start point and a unit vector along the rotatable bond a, respectively. Similar, albeit more complicated formulas can be derived for distances that depend on two dihedral angles.

5. Structure Calculation

The calculation of the 3D structure forms a cornerstone of the NMR method for protein structure determination. Because of the complexity of the problem—a protein consists typically of more than one thousand atoms that are restrained by a similar number of experimentally determined restraints in conjunction with stereochemical and steric constraints—in general it is neither possible to do an exhaustive search of allowed conformations nor to find solutions by interactive model building. In practice, the calculation of the 3D structure therefore is formulated usually as a minimization problem for a target function that measures the agreement between a conformation and the given set of restraints. In the following, the three most widely used types of algorithms are discussed.

5.1. Metric Matrix Distance Geometry

Distance geometry based on the metric matrix was the first approach used for the structure calculation of proteins on the basis of NMR data *(2,53,54)*. It relies on the fact that the NOE data and most of the stereochemical data can be represented as distance restraints. Metric matrix distance geometry is based on the theorem *(55,56)* that, given exact values for all distances among a set of points in 3D Euclidean space, it is possible to determine Cartesian coordinates for these points, which are unique except for a global inversion, translation, and rotation.

To see this, assume that we are given n points in 3D Euclidean space with coordinates, $\vec{r}_i, \ldots, \vec{r}_n$, chosen such that $\Sigma_i \vec{r}_i = 0$. The distance matrix, D, and the metric matrix, G, are the $n \times n$ matrices with elements

$$D_{ij} = |\vec{r}_i - \vec{r}_j| \quad \text{and} \quad G_{ij} = \vec{r}_i \cdot \vec{r}_j \tag{10}$$

The metric matrix can be obtained from the distance matrix according to the relations

$$G_{ii} = 1/n \sum_i D_{ij}^2 - 1/2n^2 \sum_{j,k} D_{jk}^2 \quad \text{and} \quad G_{ij} = (G_{ii} + G_{jj} - D_{ij}^2)/2 \tag{11}$$

It is positive semidefinite because for any $x_1, \ldots, x_n$ we have

$$\sum_{ij} x_i G_{ij} x_j = \sum_{\alpha=1}^{3} (\Sigma_i r_i^\alpha x_i)^2 \geq 0 \tag{12}$$

and of (maximal) rank 3 because for any $x_1, \ldots, x_n$ that fulfill the three linear conditions $\Sigma_i r_i^\alpha x_i = 0$ $(\alpha = 1, 2, 3)$ it follows that $\Sigma_j G_{ij} x_j = 0$. Therefore, the metric matrix has at most three positive eigenvalues, λ^α, with corresponding eigenvectors, e^α, that are related to the Cartesian coordinates, $\vec{r}_1 \ldots, \vec{r}_n$, of the n points by

$$r_i^\alpha = (\lambda^\alpha)^{1/2} e_i^\alpha \ (\alpha = 1, 2, 3) \tag{13}$$

Eqs. (10–13) provide a straightforward way to embed a distance matrix in 3D space, i.e., to obtain Cartesian coordinates for a set of points if all distances are known exactly. To use this idea in a structure calculation one has to account for the fact that in practice neither complete nor exact distance information is available: Only for a small subset of all distances, d_{ij}, restraints in the form of lower and upper bounds, $l_{ij} < d_{ij} < u_{ij}$, can be determined. Upper bounds result from NOEs, lower bounds from the steric repulsion, and there are some exact distance constraints from known bond lengths and bond angles of the covalent structure. To apply Eqs. (10–13), unknown upper bounds first are initialized to a large value, and unknown lower bounds to zero. Subsequently they are reduced by "bounds smoothing" *(53)*, i.e., repeated application of the triangle inequality until all lower and upper bounds are consistent with the triangle inequality. Then a complete set of distances is produced by "randomly" selecting a value between the corresponding lower and upper bounds, and the embedding procedure, Eqs. (10–13), is used to obtain Cartesian coordinates. Because the assumptions of the embedding theorem are not met exactly, the resulting structure will in general have the correct 3D fold (or its mirror image) but will be very distorted. It has to be regularized extensively, typically by conjugate gradient minimization of an appropriate target function in Cartesian coordinate space *(53)*, before it can be used as start structure for molecular dynamics simulated annealing *(52,57)*. Starting from the smoothed distance bound matrix, the calculation is repeated with different "random" selections of distances, in order to obtain a group of conformers whose spread should give an indication of the allowed conformation space.

Note that, since all conformational data have to be encoded into the distance matrix, it is not possible to introduce any handedness or chirality, and that for metric matrix distance geometry, a structure and its mirror image are always equivalent. The correct handedness is only enforced during regularization. For the same reason, dihedral angle restraints cannot be used directly in the embedding; they have to be represented by distance bounds too, thereby losing part of the information.

The sampling of conformation space by a group of conformers resulting from metric matrix distance geometry is determined decisively by the "random" selection of distance values between corresponding lower and upper bounds. The most straightforward method, namely, selecting the distances as independent, uniformly distributed random variables between the two limits, leads, because meaningful upper bounds exist only for a small subset of all distances, on the average to an overestimation of the true distances with the consequence of artificially expanded structures *(58,59)*. This effect is most pronounced for regions of the polypeptide chain for which only few restraints are available. For example, chain ends that are unstructured in solution tend to be in an extended conformation. A method to reduce—at the expense of considerably increased computation time—such biased sampling of the allowed con-

Table 2
Computation Times for the Calculation of a Group of 50 Conformers of the 40-Residue Protein Er-2 with the Program DIANA[a]

Computer type	CPU time, min
NEC SX-3	8.1
Cray Y-MP	8.5
IBM RS-6000-590	19.7
HP 9000/735	22.8
Convex C3820	39.6
Sun Sparcstation 5	132.5
Silicon Graphics Indigo	132.7

[a]On the basis of an experimental NMR data set consisting of 604 distance and 89 dihedral angle restraints *(93)* 50 conformers of the pheromone Er-2 from *Euplotes raikovi* were calculated with the program DIANA, using the REDAC strategy for convergence improvement. The average number of target function evaluations per conformer was 7200.

formation space is metrization *(53)*: Instead of selecting the individual distances independently from each other, the bounds smoothing is repeated after a next distance value has been chosen, thereby resulting in a more consistent set of distances for the embedding. This, however, introduces a strong dependence of the sampling properties on the order in which the distances are chosen *(59)*. The efficiency of metrization can be enhanced by partial metrication, i.e., by repeating the bounds smoothing only after the selection of the first few percent of the distances, which are chosen in random order *(60)*.

The first metric matrix distance geometry program commonly used for protein structure calculation was the program DISGEO *(53)*. Other widely used implementations include its successor, DG-II *(59)*, and the program XPLOR *(52)*.

5.2. Variable Target Function Method

The basic idea of the variable target function algorithm *(61)* is to fit gradually an initially randomized starting structure to the conformational restraints collected with the use of NMR experiments, starting with intraresidual restraints only, and increasing the "target size" stepwise up to the length of the complete polypeptide chain. Advantages of the method are its conceptual simplicity, its computational efficiency (Table 2), and the fact that it works in dihedral angle space, so that the covalent geometry is preserved during the entire calculation *(62)*. The variable target function algorithm

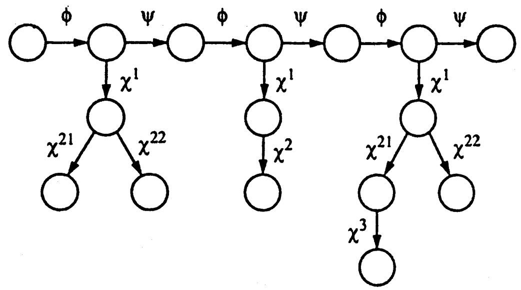

Fig. 6. Tree structure of dihedral angles as it is used in the program DIANA for the tripeptide Val–Ser–Ile. Numbered circles represent rigid units. Rotatable bonds are indicated by arrows that point toward the part of the structure that is rotated if the corresponding dihedral angle is changed.

was first implemented in the program DISMAN *(61)*. Today, the program DIANA *(51)* is the most commonly used implementation to which the following description refers.

The algorithm is based on the minimization of a variable target function, $T(\phi_1, \ldots, \phi_n)$, where the n degrees of freedom are the dihedral angles, $\phi_1, \ldots, \phi_n$, about rotatable bonds of the polypeptide chain. During the calculation bond lengths, bond angles and chiralities of the covalent structure are kept fixed at the standard values of the ECEPP force field *(63,64)*. The molecule thus is represented as a tree structure made up of rigid units that consist of one or several atoms and that are connected by rotatable bonds (Fig. 6). The target function, T, with $T \geq 0$, is defined such that $T = 0$ if and only if all experimental distance and dihedral angle restraints are fulfilled and all nonbonded atom pairs satisfy a check for the absence of steric overlap. $T(\phi_1, \ldots, \phi_n) \leq T(\phi'_1, \ldots, \phi'_n)$ if the conformation $(\phi_1, \ldots, \phi_n)$ satisfies the restraints "better" than the conformation $(\phi'_1, \ldots, \phi'_n)$. The problem to be solved is to find values $(\phi_1, \ldots, \phi_n)$ that yield low values of the target function.

The exact definition of the DIANA target function is as follows: Upper and lower bounds $c_{\alpha\beta}$ on distances between two atoms α and β, and restraints on individual dihedral angles ϕ_a in the form of allowed intervals $[\phi_a^{min}, \phi_a^{max}]$ with $\phi_a^{min} < \phi_a^{max} < \phi_a^{min} + 2\pi$ are considered. Let $\Gamma_a = \pi - (\phi_a^{max} - \phi_a^{min})/2$ be the half-width of the forbidden interval of dihedral angle values, and Δ_a the size of the dihedral angle restraint violation. I_u, I_l, and I_v denote the sets of atom pairs $\alpha\beta$, for which upper, lower, or van der Waals distance bounds $u_{\alpha\beta}$, $l_{\alpha\beta}$, or $v_{\alpha\beta}$ exist, $I'_c \subseteq I_c$ $(c = u, l, v)$ subsets thereof, I_d the set of restrained dihedral angles, w_u, w_l, w_v, and w_d weighting factors for the different types of restraints, $d_{\alpha\beta}$ the distance between the atoms α und β, $\Theta_u(t) = \max(0, t)$ and $\Theta_l(t) = \Theta_v(t) = \min(0, t)$. Then the target function, T, is defined by *(51)*:

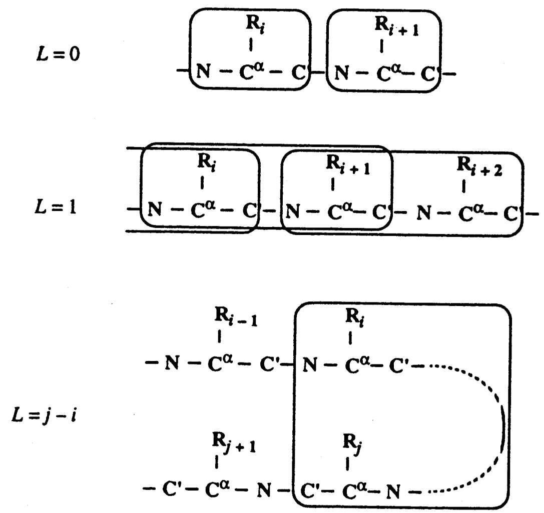

Fig. 7. Illustration of the variable target function algorithm as implemented in the program DIANA *(51)*. Boxes indicate examples of allowed ranges of active restraints at various minimization levels L.

$$T=\sum_{c=u,l,v} w_c \sum_{\alpha\beta\in\Gamma_c} [\Theta_c(d^2_{\alpha\beta}-c^2_{\alpha\beta})/2c_{\alpha\beta}]^2 + w_d\sum_{\alpha\in Id}[1-1/2\,(\Delta_a/\Gamma_a)^2]\,\Delta^2_\alpha \qquad (14)$$

The target function of Eq. (14) is differentiable continuously over the entire conformation space, and is chosen such that the contribution of a single small violation, δ_c, is given by $w_c\delta_c^2$ for all types of restraints. Because only squared interatomic distances and no square roots have to be computed, the target function can be calculated rapidly. The subsets $I'_c \subseteq I_c$ $(c = u,l,v)$ for which the program DIANA can calculate the target function consist of all distance restraints between residues whose sequence numbers differ by not more than a given minimization level, L (Fig. 7).

To reduce the danger of becoming trapped in a local minimum with a function value much higher than the global minimum, the target function is varied during a structure calculation. At the outset, only local restraints with respect to the polypeptide sequence are considered. Subsequently, restraints between atoms further apart with respect to the primary structure are included in a stepwise fashion (Fig. 7). Consequently, in the first stages of a structure calculation the local features of the conformation will be established, and the global fold of the protein will be obtained only toward the end of the calculation (Fig. 8).

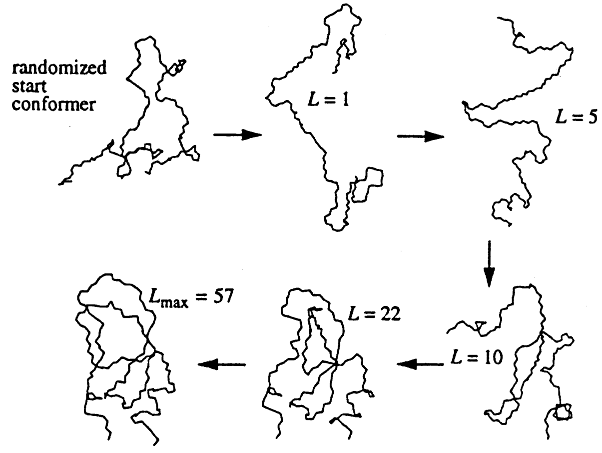

Fig. 8. Randomized starting conformer, intermediate structures, and final, converged structure of the protein toxin K during variable target function minimization with the program DIANA. Covalent bonds among the backbone atoms N, C^α, and C' of the random start conformation, and of the conformations at the end of the minimization levels L = 1, 5, 10, 22, and L_{max} = 57 are shown *(97)*.

The minimization algorithm used in the program DIANA is the well-known method of conjugate gradients *(65)*. At each minimization step conjugate gradient iterations are done until either the norm of the gradient is smaller than some preset value, or the maximal number of iterations at this minimization step is exceeded. The gradient of the target function can be calculated with a fast algorithm because the target function can be written as a sum of functions of individual interatomic distances and dihedral angles *(66,67)*. The partial derivative of the function T of Eq. (14) with respect to a dihedral angle a' is given by

$$\partial T/\partial\phi_{a'} = -\vec{e}_{a'}\cdot\vec{f}_{a'} - (\vec{e}_{a'}\wedge\vec{r}_{a'})\cdot\vec{g}_{a'} + 2w_d\sum_{\alpha\in Id}[1-(\Delta_a/\Gamma_a)^2]\,\Delta_a\delta_{aa'} \qquad (15)$$

where

$$\vec{f}_{a'} = \sum_{c=u,l,v} w_c \sum_{\substack{\alpha\beta\in\Gamma_c\\ \alpha\in M_{a'}}} [\Theta_c(d^2_{\alpha\beta}-c^2_{\alpha\beta})/c^2_{\alpha\beta}]\,(\vec{r}_\alpha\wedge\vec{r}_\beta),$$
$$\vec{g}_{a'} = \sum_{c=u,l,v} w_c \sum_{\substack{\alpha\beta\in\Gamma_c\\ \alpha\in M_{a'}}} [\Theta_c(d^2_{\alpha\beta}-c^2_{\alpha\beta})/c^2_{\alpha\beta}]\,(\vec{r}_\alpha - \vec{r}_\beta) \qquad (16)$$

$\vec{r}_\alpha$ and $\vec{r}_\beta$ denote the position vectors of the atoms α and β, respectively, $\vec{e}_{a'}$ denotes the unit vector along the rotatable bond a', $\vec{r}_{a'}$ the start point of it, and $M_{a'}$ the set of all atoms for which the positions are affected by a change of the dihedral angle a' if the N-terminal part of the protein molecule is kept fixed.

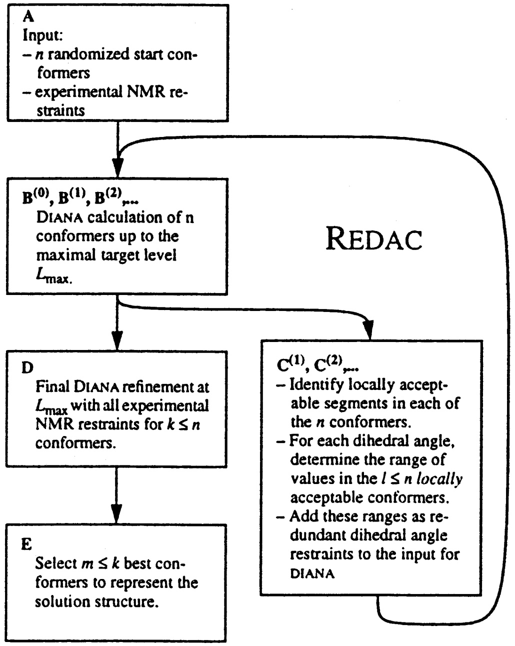

Fig. 9. Flowchart outlining the course of a protein structure calculation with the program DIANA using either the "direct" way (A-$B^{(0)}$-D-E) or redundant dihedral angle constraints (REDAC; *69*) (A-$B^{(0)}$-[$C^{(1)}$-$B^{(1)}$-...]-D-E). Typically, the number of REDAC-cycles is 1 or 2.

A drawback of the straightforward implementation of the variable target function algorithm is that for all but the most simple molecular topologies only a small percentage of the calculations converge with small residual restraint violations, which is a typical local minimum problem. Because of the low yield of acceptable conformers, calculations typically have been started with a large number of randomized starting conformers in order to obtain a group of good solutions, and sometimes a compromise had to be made between the requirements of small residual violations and the available computing time *(68)*. With the introduction of the highly optimized program DIANA, which significantly reduced the computation time needed for the calculation of a single conformer, a workable situation was achieved for α-proteins *(40)*, but for β-proteins with more complex topology the situation remained unsatisfactory. However, with the use of redundant dihedral angle restraints (REDAC), a greatly improved yield of converged conformers is obtained also for β-proteins.

In Fig. 9 the use of DIANA with redundant dihedral angle constraints *(69)* is outlined and placed in perspective with the "direct" variable target function

method as proposed originally by Braun and Gō *(61)*. In the direct approach, n start conformers with randomized dihedral angles are subjected to DIANA minimization against the experimental NMR constraints in step $B^{(0)}$. Experience has shown that the target function can be reduced further by repeating the DIANA refinement with all constraints and variable weights for the van der Waals constraints for a limited number k of well-converged solutions ($m \leq k \leq n$) in step D. Among the resulting solutions, m conformers with the smallest final target function values are selected to represent the solution structure. In practice, n is adjusted so as to obtain, typically, $m = 20$ acceptable conformers.

To use REDAC, one or several cycles $C^{(i)}$–$B^{(i)}$ are added to the calculation, providing a partial feedback of structural information from all conformers that were calculated up to the maximal minimization level L_{max} in the step $B^{(i-1)}$. In the step $C^{(i)}$, a particular amino acid residue is considered to have an acceptably well-defined conformation if the target function value owing to constraint violations that involve atoms or dihedral angles of this residue is less than a predefined value, and if the same condition holds for the two sequentially neighboring residues. Redundant dihedral angle constraints are generated and added to the input for the DIANA structure calculation in step $B^{(i)}$ for all those residues that were found to be acceptable in at least a predefined minimal number of conformers, typically 10, by taking the two extreme dihedral angle values in the group of acceptable conformers as upper and lower bounds.

The empirically found higher yield of good conformers with the use of REDAC—that does not lead to a reduction in the sampling of conformation space *(69)*—can be rationalized as follows: In many regions of a protein structure, in particular in β-strands, the local conformation is determined not only by the local conformational constraints derived from intraresidual, sequential, and medium-range NOEs *(1)*, but also by longer-range constraints, e.g., interstrand distance constraints in β-sheets. Therefore, the local constraints alone may allow for multiple different local conformations at low target levels in a DIANA calculation, of which some may be incompatible with the longer-range constraints taken into account at higher minimization levels. Obviously, incorrect local conformations that satisfy the experimentally available local constraints are potential local minima, which could only be ruled out from the beginning if the information contained in the long-range constraints were available already at low levels of the minimization. The use of REDAC achieves this: Information contained in the complete data set is translated into (by definition intraresidual) dihedral angle constraints. Further, it makes clear why the yield of good solutions with the direct strategy was in general higher for α-proteins than for β-proteins, since the conformation of an α-helix its particularly well-determined by sequential and medium-range constraints *(1)*. The success of DIANA structure calculations using REDAC primarily is owing to the feedback of useful structural information derived from conformers calculated up to the maximal level L_{max} into a subsequent round of structure calculations, which starts with local constraints only. In this way information gathered from the entire structure calculation is used in obtaining the final result, whereas most of this information is discarded in the direct approach.

5.3. Molecular Dynamics Simulated Annealing

This third major method for NMR structure calculation is based on numerically solving Newton's equation of motion in order to obtain a trajectory for the molecular system *(70,71)*. The degrees of freedom are the Cartesian coordinates of the atoms. In contrast to "standard" molecular dynamics simulations that try to simulate the behavior of the real physical system as closely as possible (and do not include restraints derived from NMR), the purpose of a molecular dynamics calculation in a NMR structure determination is simply to search the allowed conformation space of the protein. Therefore, the simulated annealing *(57,72)* is performed at high temperature using a simplified force field that treats the atoms as soft spheres without attractive or long-range (i.e., electrostatic) nonbonded interactions, and that does not include explicit consideration of the solvent. An important feature of molecular dynamics when compared to the straightforward minimization of a target function is that the presence of kinetic energy allows to cross barriers of the potential surface, thereby reducing the problem of becoming trapped in local minima. Since molecular dynamics cannot generate conformations from scratch, a start structure is needed, which is generated typically either by metric matrix distance geometry or by the variable target function method, but—at the expense of substantially increased computation time—it is possible also to start from a random structure *(73)* or even from a set of atoms randomly distributed in space *(74)*. Any general molecular dynamics program, such as GROMOS *(75)*, AMBER *(76)*, or CHARMM *(77)*, can be used for the simulated annealing of NMR structures, provided that pseudoenergy terms for distance and dihedral angle restraints have been incorporated. In practice, the program best adapted and most widely used for this purpose is XPLOR *(52)*.

The forces, $\vec{F}_i$, acting on n particles with masses, m_i, and positions, $\vec{r}_i$, in Newton's equation of motion,

$$m_i\, d^2\vec{r}_i/dt^2 = \vec{F}_i \quad (i = 1,\ldots, n) \tag{17}$$

are given by the negative gradient of the potential energy function, E, with respect to the Cartesian coordinates, i.e., $\vec{F}_i = -\vec{\nabla}_i E$. For simulated annealing, a simplified potential energy function is used that includes terms to maintain the covalent geometry of the structure by means of harmonic bond length and bond angle potentials, dihedral angle potentials, terms to enforce the proper chiralities

and planarities, a simple repulsive potential instead of the Lennard-Jones and electrostatic nonbonded interactions, as well as terms for distance and dihedral angle restraints. For example, in the program XPLOR *(52)*,

$$\begin{aligned} E = &\sum_{\text{bonds}} k_b (r-r_0)^2 + \sum_{\text{angles}} k_\theta (\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1+\cos(n_\phi+\delta)] \\ &+ \sum_{\text{impropers}} k_\phi (\phi-\delta)^2 + \sum_{\substack{\text{nonbonded}\\ \text{pairs}}} k_{\text{repel}} \{\max[0,(sR_{\min})^2 - R^2]\}^2 \\ &+ \sum_{\substack{\text{distance}\\ \text{restraints}}} k_d \Delta_d^2 + \sum_{\substack{\text{angle}\\ \text{restraints}}} k_a \Delta_a^2 \end{aligned} \tag{18}$$

In Eq. (18), k_b, k_θ, k_ϕ, k_{repel}, k_d, and k_a denote the various force constants, r and r_0 the actual and the correct bond length, respectively, θ and θ_0 the actual and correct bond angle, ϕ the actual dihedral angle or improper angle value, n the number of minima of the dihedral angle potential, δ an offset of the dihedral angle and improper potentials, $R_{\min}$ the distance where the van der Waals potential has its minimum, R the actual distance between a nonbonded atom pair, s a scaling factor, and Δ_d and Δ_a the size of the distance or dihedral angle restraint violation. To obtain a trajectory, the equation of motion is numerically integrated using a small, finite time step, δt, for example by advancing the coordinates, $\vec{r}_i$, of the particles according to *(70)*.

$$\vec{r}_i(t+\delta t) = 2\vec{r}_i(t) - \vec{r}_i(t-\delta t) + (\delta t^2/m_i)\, \vec{F}_i(t) \tag{19}$$

The velocities, $\vec{v}_i$, are given by

$$\vec{v}_{i(t)} = [\vec{v}_i(t+\delta t) - \vec{v}_i(t-\delta t)]/2\delta t \tag{20}$$

The time step, δt, must be small enough to still sample adequately the fastest motions, i.e., of the order of 10^{-15} s. In general, the highest frequency motions are bond length oscillations. Therefore, the time step can be increased if the bond lengths are constrained to their correct values by the SHAKE method *(78)*. To control the temperature of the system, it is loosely coupled to a temperature bath *(79)*.

For the simulated annealing of a (possibly distorted) start structure, certain measures have to be taken in order to achieve sampling of the conformation space within reasonable time *(57)*. In a typical simulated annealing protocol *(52)*, the simulated annealing is performed for a few picoseconds at high temperature, say $T = 2000$ K, starting with a very small weight for the steric repulsion that allows atoms to penetrate each other, and gradually increasing the strength of the steric repulsion during the calculation. Subsequently, the system is cooled down slowly for another few picoseconds and finally energy-minimized. This process is repeated individually for each of the start conformers. In general, simulated annealing by molecular dynamics requires substantially more computation time

Table 3
Computation Times for the Calculation of Groups of 10 Conformers with the Program XPLOR on a Convex C210 Computer[a]

	CPU time, min	
Calculation	Protein G[b], 855 atoms	Interleukin-8[c], 2362 atoms
Metrix matrix distance geometry		
Substructure embedding[d]	16	346
Full embedding[e]	405	11,520
Simulated annealing		
Starting from embedded structures	254	691
Starting from random structures	599	1680

[a] From Brünger *(52)*, p. 328.
[b] Gronenborn et al. *(95)*.
[c] Clore et al. *(96)*.
[d] Embedding is performed for a substructure consisting of all C^α, H^α, N, HN, C', C^β, and C^γ atoms; the remaining atoms will be added from a template structure during regularization.
[e] Embedding is performed for all atoms.

than the other methods (Table 3). This lower efficiency partially is compensated by the high success rate of 40–100% of the start structures that is caused by the ability of the algorithm to "escape" from local minima.

6. Structure Analysis

The result of a NMR structure calculation is a group of conformers that represents the solution structure of the protein. This section summarizes some of the commonly used formal measures for the analysis of a solution structure. A computer program, PROCHECK, that incorporates many different analyses of groups of NMR conformers is available *(80)*. Of course, in order to obtain a biologically meaningful assessment of the structure one will also analyze it visually with the help of a molecular graphics system and compare it with structures of related proteins.

6.1. Residual Restraint Violations

At the end of a structure calculation, the immediate question arises whether the structure calculation was successful, i.e., whether the algorithm was able to find structures that fulfill the given restraints, and, if not, which are the restraints that could not be satisfied. Therefore, an analysis of the residual restraint violations seen in the final conformers is performed, which is usually summarized in a table (Table 4). In addition, a list of residual restraint violations that indicates, for each violation separately, the individual conformers

Table 4
Analysis of the Residual Restraint Violations in the 20 Conformers of the Protein Toxin K with the Smallest DIANA Target Function Values[a]

Quantity	Average value ± standard deviation	Range
DIANA target function[b]	0.50 ± 0.14 Å^2	0.19–0.74 Å^2
Residual NOE distance restraint violations		
Number >0.1 Å	3.8 ± 2.1	0–7
Sum	3.3 ± 0.4 Å	2.0–3.8 Å
Maximum	0.15 ± 0.05 Å	0.07–0.31 Å
Residual dihedral angle restraint violations		
Number >5°	0	
Sum	6.1 ± 2.4°	2.4–10.6°
Maximum	1.8 ± 0.9°	0.6–4.2°

[a] From Berndt et al. *(97)*. A total of 50 conformers were calculated with the program DIANA using the REDAC strategy for convergence improvement but only the 20 structures with the smallest final target function values are included in the analysis.
[b] The weighting factor in Eq. (14) for NOE upper distance bounds was $w_u = 1$, for van der Waals lower distance bounds $w_v = 2$, and for dihedral angle restraints $w_d = 5$ Å^2.

where the violation occurs, can reveal consistent violations and distinguish them from casual violations resulting from the occurrence of different local minima in different conformers. Consistent violations most likely point to an inconsistency of the input data rather than to a convergence problem of the structure calculation algorithm.

6.2. Atomic Root-Mean-Square Deviations and Displacements

The standard measure used to quantify differences between 3D structures is the root-mean-square deviation (RMSD) for a given set of corresponding atoms *(81)*. For two sets of n atoms each, $\vec{r}_1, \ldots, \vec{r}_n$ and $\vec{q}_1, \ldots, \vec{q}_n$ with $\sum_i \vec{r}_i = \sum_i \vec{q}_i = 0$, the RMSD is defined by

$$\mathrm{RMSD} = \min_{R \in SO(3)} [1/n \sum_{i=1}^{n} \vec{r}_i - R\vec{q}_i^{\,2}]^{1/2} \tag{21}$$

R denotes a rotation matrix, and $SO(3)$ the rotation group. The optimal superposition according to Eq. (21) is used also for the simultaneous display of several conformers on a molecular graphics system. In practice, RMSD values usually are calculated for the backbone atoms N, C^α, and C' or for all heavy (i.e., nonhydrogen) atoms of those residues for which the conformation is well-defined, excluding, for instance, chain termini and loops that are unstructured in solution. For a group of m conformers, either the average of the m RMSD

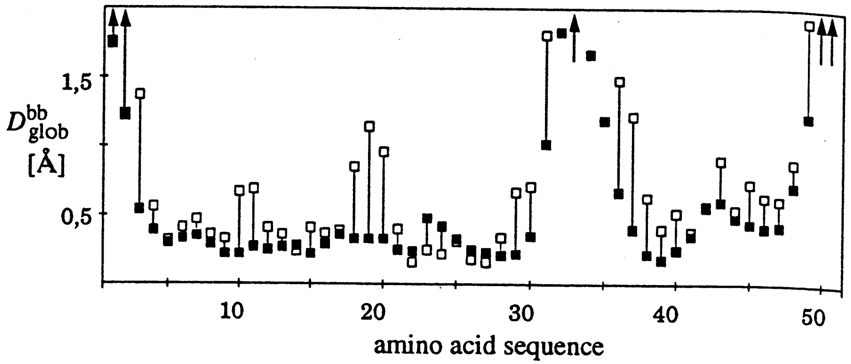

Fig. 10. Plot vs the amino acid sequence of the global backbone displacements, D^{bb}_{glob}, relative to the mean NMR structure of hirudin(1–51) for the 20 conformers of the solution structure (*99*; filled squares), and for hirudin in the X-ray crystal structure of the hirudin-thrombin complex (*100*; open squares).

values between the individual conformers and their mean structure or the average of the $m(m-1)/2$ pairwise RMSD values among the individual conformers are reported. The latter, pairwise RMSD is about 1.4 times larger than the corresponding RMSD to the mean structure. To define the mean structure, the conformers are superimposed according to Eq. (21) onto the first conformer, and the arithmetic mean of the corresponding Cartesian coordinates is taken.

It cannot be overemphasized that a small RMSD value *per se* is not indicative of a high-quality structure: It neither gives any information about the consistency of the experimental data nor does it correspond necessarily to the conformation space that is really allowed by the conformational restraints because the sampling of conformation space by the structure calculation algorithm may have been biased, i.e., there exist structures that would be in agreement with the data but significantly different from those resulting from the structure calculation. This may be a result of the limited statistics—typically, only about 20 conformers are analyzed—or of an inherent deficiency of the structure calculation algorithm *(58)*.

Displacements *(82)* are a generalization of the RMSD values, since the set of atoms used for the superposition of the conformers, M_{sup}, differs from the set of atoms for which the RMSD of the positions actually is calculated, M_{RMSD}. For example, for the evaluation of the backbone displacement of a given residue *i* after global superposition, D^{bb}_{glob}, M_{sup} consists of the backbone atoms N, C$^{\alpha}$, and C' of the residues used for global superposition, and M_{RMSD} of the backbone atoms N, C$^{\alpha}$, and C' of residue *i*. To evaluate local backbone displacements for a residue *i*, D^{bb}_{loc}, M_{sup} consists of the backbone atoms N, C$^{\alpha}$, and C' of the residues $i-1$, i, and $i+1$, and M_{RMSD} consists of the backbone atoms of residue *i*. Displacements are conveniently displayed in a plot vs the amino acid sequence of the protein (Fig. 10).

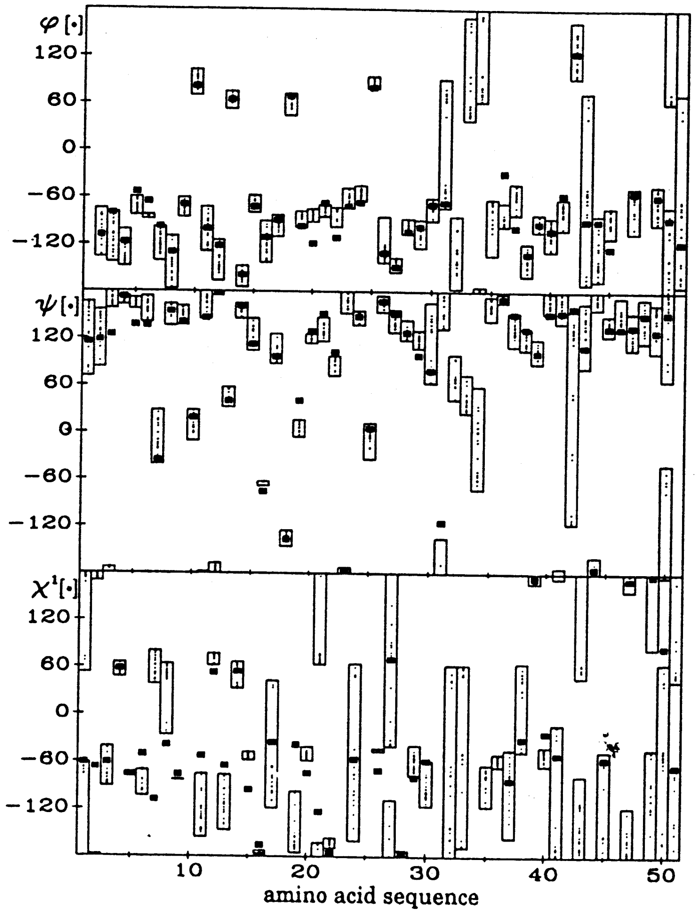

Fig. 11. Values of the dihedral angles ϕ, ψ, and χ^1 for hirudin(1–51) observed in the 20 conformers that represent the solution structure (*99*; dots enclosed by bars) and for hirudin in the X-ray crystal structure of the hirudin–thrombin complex (*100*; black squares).

6.3. Dihedral Angle Distributions

To analyze the conformation of the polypeptide chain on a local level, plots of the distributions of the individual values of the dihedral angles ϕ, ψ, and χ^1 vs the amino acid sequence of the protein (Fig. 11) and Ramachandran plots are convenient. They allow, for example, the identification of secondary structure elements, the classification of tight turns, and an assessment of the local precision of the structure determination.

To obtain the average value, $\overline{\phi}$, and the standard deviation, σ, from the dihedral angle values, $\phi_1, \ldots, \phi_n$, of the individual confornners, one has to take into account the periodicity of the dihedral angles, for example, by using

$$\bar{\phi} = \arg \sum_k e^{i\phi_k}$$

and

$$\sigma = [-2 \log|1/n \sum_k e^{i\phi_k}|]^{1/2} \tag{22}$$

The values defined by Eqs. (22) have a clear meaning only when the dihedral angle is well-defined. For example, the common situation that there are two groups of conformers, each with a well-defined value of the dihedral angle, but with a large difference between the two groups, cannot be distinguished from the situation of a truly disordered dihedral angle by means of Eqs. (22).

6.4. Hydrogen Bonds

Another important feature of protein structures are hydrogen bonds. They can be identified readily in the structure, for example, by the criterion that the hydrogen-acceptor distance must be shorter than 2.4 Å and that the angle between the hydrogen, the atom to which the hydrogen is covalently bound, and the acceptor must be smaller than 35° *(82)*. The second condition ensures that the hydrogen bond roughly is linear.

7. Structure Refinement

There exist many possibilities for the refinement of NMR structures of proteins. The following brief overview can mention only a few often used refinement methods.

Fig. 12. *(opposite)* Illustration of the use of a 3D protein structure to resolve amiguities in the NOESY crosspeak assignment. The program ASNO *(83)* is used in conjunction with the program package XEASY *(98,100)* for interactive spectral analysis. **(A)** Contour plot of a spectral region containing two crosspeaks in the NOESY spectrum of toxin K *(97)* are displayed by XEASY together with the corresponding assignment matrices. In each matrix the rows and columns are identified with the type of hydrogen atom and the sequence position of the amino acid residue, the hatched squares denote chemical shift-based assignments, and black squares denote assignments corresponding to distances shorter than 5 Å in the structure that were determined by ASNO. The figure shows that ASNO yielded unique assignments for both peaks, whereas the chemical shift information alone lead to six- and eightfold degenerate assignments, respectively. **(B)** View of the toxin K conformer with the lowest final DIANA target function value *(97)*. All backbone atoms N, C^{α}, and C', the backbone amide protons of residues 10, 38, and 50, and all side chain heavy atoms of the residues Cys 14 and Tyr 21 have been drawn. All possible chemical shift-based assignments of the upper NOESY crosspeak in (A) are identified by dashed lines, and the averages of the corresponding distances in the 20 final DIANA conformers are indicated. The arrow points to the distance corresponding to the unique assignment by ASNO.

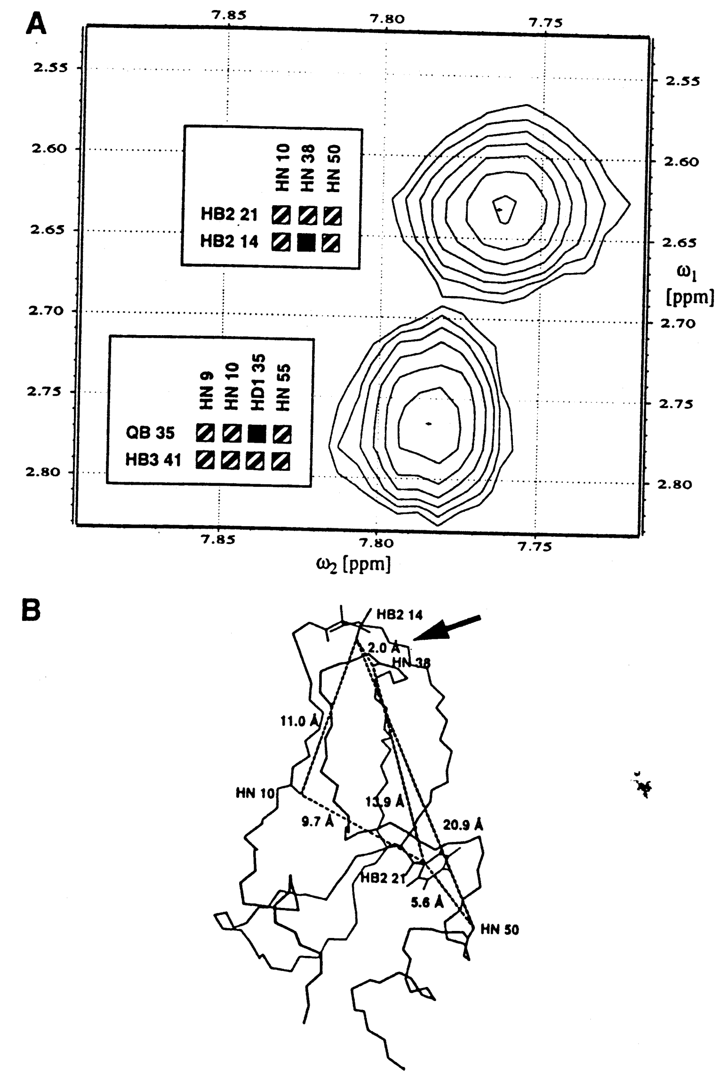

Fig. 12

7.1. Resolving Assignment Ambiguities

The most obvious, straightforward, and effective refinement of a NMR structure is to make use of preliminary 3D structures as a reference to resolve ambiguities in NOE assignments that are otherwise based on the chemical shifts available from the sequence-specific resonance assignments. In general, this process is repeated iteratively in several rounds of NOESY crosspeak assignments and structure calculations. It can be automated readily *(83;* Fig. 12). Typically, the number of unambiguously assigned NOESY crosspeaks, and

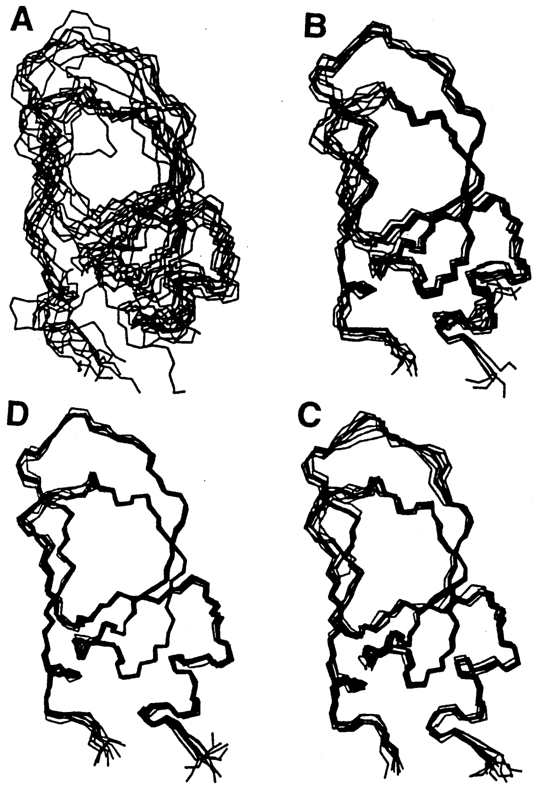

Fig. 13. View of the polypeptide backbone of the 10 best DIANA conformers of toxin K at four stages of structure refinement using the program ASNO (*83*; Fig. 12). **(A)** Structure calculated on the basis of 322 NOE upper distance bounds assigned from chemical shift information only. **(B)** Structure calculated from an input of 657 NOE upper distance bounds assigned using ASNO with the structure from (A). **(C)** Structure calculated from an input of 747 relevant NOE upper distance bounds assigned using ASNO with the structure from (B). **(D)** "Final" structure calculated from 809 NOE upper distance bounds *(97)*.

hence the number of distance restraints, can be increased at least twofold over the situation when only assignments based on the chemical shifts are included. The most pronounced improvement is a result of the increased number of long-range distance restraints (Fig. 13).

7.2. Restrained Energy Minimization

The structure calculation algorithms for NMR structures usually use a simplified force field that contains only the most dominant parts of the conformational energy. Therefore, the resulting structures may be unfavorable with respect to a

full, "physical" energy function *(63,75,84)* that includes, in addition to the terms used by the distance geometry structure calculation algorithms of Section 5., also a Lennard-Jones potential and electrostatic interactions for nonbonded atom pairs, and possibly other terms. The conformational energy of a distance geometry structure can be reduced significantly by restrained energy minimization, i.e., by locating a local minimum of the conformational energy function in the near vicinity of the distance geometry structure *(82)*. Visually, restrained energy minimization results in only small changes of the conformation.

7.3. Molecular Dynamics Simulation

An unrestrained or restrained molecular dynamics simulation under physiological conditions using the full physical force field and explicit water to represent the solvent can often give new insights into a protein structure, in particular for the generally disordered protein surface *(85,86)*. Such "realistic" molecular dynamics simulations that try to represent the molecular system as faithfully as possible are different fundamentally from simulated annealing, where artificial conditions, such as high temperature, are chosen in order to enhance the sampling of conformation space. A limiting factor in molecular dynamics simulations are the relatively short simulation times of up to approximately 1 ns that are feasible with present computers, because many motions in proteins occur on longer time scales.

7.4. Time- or Ensemble-Averaged Conformational Restraints

The commonly used structure calculation algorithms try to find rigid conformations that fulfill all distance and dihedral angle restraints simultaneously, and the effects of internal mobility of the polypeptide chain are taken into account implicitly by the interpretation of the NOE data as conservative upper distance bounds instead of exact distance constraints *(1)*. In reality, the NOEs and scalar coupling constants measured by NMR constitute an average over time and space. Methods have been proposed to include distance and dihedral angle restraints as time-averaged rather than instantaneous restraints into a molecular dynamics simulation *(16,17,87)*. In another approach, a molecular dynamics simulation is performed simultaneously for an ensemble of conformers, such that the restraints are not required to be fulfilled by each individual conformer but only by the ensemble as a whole *(88)*.

7.5. Relaxation Matrix Refinement

Both spin diffusion and the internal mobility of proteins influence the NOE intensities from which distance restraints are derived for the structure calculation. Complete relaxation matrix refinement *(22–24,26)* can, in principle, take these factors into account and thus allow to make a more quantitative use of the NOE data as with the initial rate approximation *(20)* and the semi-quantitative calibration of distance restraints *(1)*. However, there is the danger of overinterpretation of the data because many of the parameters entering the relaxation matrix in practice can not be derived from experiments. In particular, many assumptions about internal and overall motions of the protein enter the relaxation matrix *(13)*. Two different methods of complete relaxation matrix refinement are in use: Either the relaxation matrix treatment is used to derive a "more precise" set of distance restraints, which is then used in a conventional structure calculation *(22,89)*, or the 3D structure may be refined directly with respect to the observed NOE intensities *(23,24,90)*. The second approach conceptually is more attractive but also much more time-consuming. In analogy to the practice in X-ray crystallography, *R*-factors can be defined on the basis of the relaxation matrix that measure the agreement between the NOESY spectrum and the 3D structure *(91–93)*.

References

1. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
2. Braun, W., Bösch, C., Brown, L. R., Gō, N., and Wüthrich, K. (1981) Combined use of proton–proton overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim. Biophys. Acta* **667,** 377–396.
3. Braun, W., Wider, G., Lee, K. H., and Wüthrich, K. (1983) Conformation of glucagon in a lipid-water interphase by ^{1}H nuclear magnetic resonance. *J. Mol. Biol.* **169,** 921–948.
4. Arseniev, A. S., Kondakov, V. I., Maiorov, V. N., and Bystrov, V. F. (1984) NMR solution spatial structure of 'short' scorpion insectotoxin I_5A. *FEBS Lett.* **165,** 57–62.
5. Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wüthrich, K., and Kaptein, R. (1984) Spatial arrangement of the three α helices in the solution structure of *E. coli lac* repressor DNA-binding domain. *FEBS Lett.* **174,** 243–247.
6. Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) Solution conformation of proteinase inhibitor IIa from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182,** 295–315.
7. Gordon, S. L. and Wüthrich, K. (1978) Transient proton–proton Overhauser effects in horse ferrocytochrome c. *J. Am. Chem. Soc.* **100,** 7094–7096.
8. Dubs, A., Wagner, G., and Wüthrich, K. (1979) Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. *Biochim. Biophys. Acta* **577,** 177–194.
9. Wagner, G. and Wüthrich, K. (1982) Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. *J. Mol. Biol.* **160,** 343–361.
10. Solomon, I. (1955) Relaxation processes in a system of two spins. *Physiol. Rev.* **99,** 559–565.

11. Abragam, A. (1961) *Principles of Nuclear Magnetism,* Clarendon, Oxford.
12. Noggle, J. H. and Schirmer, R. E. (1971) *The Nuclear Overhauser Effect,* Academic, New York.
13. Macura, S. and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Mol. Phys.* **41,** 95–117.
14. Neuhaus, D. and Williamson, M. P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis,* VCH, New York.
15. Anil-Kumar, Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95,** 1–6.
16. Torda, A. E., Scheek, R. M., and van Gunsteren, W. (1989) Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Lett.* **157,** 289–294.
17. Torda, A. E., Scheek, R. M., and van Gunsteren, W. (1990) Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* **214,** 223–235.
18. Kalk, A. and Berendsen, H. J. C. (1976) Proton magnetic relaxation and spin diffusion in proteins. *J. Magn. Reson.* **24,** 343–366.
19. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford.
20. Anil-Kumar, Wagner, G., Ernst, R. R., and Wüthrich, K. (1981) Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: implications for studies of protein conformation. *J. Am. Chem. Soc.* **103,** 3654–3658.
21. Dobson, C. M., Olejniczak, E. T., Poulson, F. M., and Ratcliffe, R. G. (1982) Time development of proton nuclear Overhauser effects in proteins. *J. Magn. Reson.* **48,** 97–110.
22. Keepers, J. W. and James, T. L. (1984) A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.* **57,** 404–426.
23. Yip, P. and Case, D. A. (1989) A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra. *J. Magn. Reson.* **83,** 643–648.
24. Mertz, J. E., Güntert, P., Wüthrich, K., and Braun, W. (1991) Complete relaxation matrix refinement of NMR structures of proteins using analytically calculated dihedral angle derivatives of NOE intensities. *J. Biomol. NMR* **1,** 257–269.
25. Denk, W., Baumann, R., and Wagner, G. (1986) Quantitative evaluation of cross-peak intensities by projection of two-dimensional NOE spectra on a linear space spanned by a set of reference resonance lines. *J. Magn. Reson.* **67,** 386–390.
26. Bystrov, V. F. (1976) Spin-spin coupling and the conformational states of peptide systems. *Prog. NMR Spectrosc.* **10,** 41–81.
27. Karplus, M. (1963) Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85,** 2870,2871.
28. DeMarco, A., Llinás, M., and Wüthrich, K. (1978) Analysis of the ^{1}H-NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers* **17,** 617–636.

29. DeMarco, A., Llinás, M., and Wüthrich, K. (1978) ^{1}H-^{15}N spin–spin couplings in alumichrome. *Biopolymers* **17,** 2727–2742.
30. Pardi, A., Billeter, M., and Wüthrich, K. (1984) Calibration of the angular dependence of the amide proton–C$^{\alpha}$ proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. *J. Mol. Biol.* **180,** 741–751.
31. Kessler, H., Gehrke, M., and Griesinger, C. (1988) Two-dimensional NMR spectroscopy: background and overview of the experiments. *Angew. Chem. Int. Ed.* **27,** 490–536.
32. Neuhaus, D., Wagner, G., Vašák, M., Kägi, J. H. R., and Wüthrich, K. (1985) Systematic application of high-resolution, phase-sensitive two-dimensional ^{1}H-NMR techniques for the identification of the amino-acid-proton spin systems in proteins. *Eur. J. Biochem.* **151,** 257–273.
33. Griesinger, C., Sørensen, O. W., and Ernst, R. R. (1985) Two-dimensional correlation of connected NMR transitions. *J. Am. Chem. Soc.* **107,** 6394–6396.
34. Neri, D., Otting, G., and Wüthrich, K. (1990) New nuclear magnetic resonance experiment for measurements of the vicinal coupling constants $^{3}J_{HN\alpha}$ in proteins. *J. Am. Chem. Soc.* **112,** 3663–3665.
35. Szyperski, T., Güntert, P., Otting, G., and Wüthrich, K. (1992) Determination of scalar coupling constants by inverse Fourier transformation of in-phase multiplets. *J. Magn. Reson.* **99,** 552–560.
36. Spera, S. and Bax, A. (1991) Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113,** 5490–5492.
37. de Dios, A. C., Pearson, J. G., and Oldfield, E. (1993) Secondary and tertiary structural effects on protein NMR chemical shifts: An ab initio approach. *Science* **260,** 1491–1496.
38. Brüschweiler, R., Blackledge, M., and Ernst, R. R. (1991) Multi-conformational peptide dynamics derived from NMR data: A new search algorithm and its application to antamanide. *J. Biomol. NMR* **1,** 3–11.
39. Billeter, M., Braun, W., and Wüthrich, K. (1982) Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Computation of sterically allowed proton–proton distances and statistical analysis of proton–proton distances in single crystal protein conformations. *J. Mol. Biol.* **155,** 321–346.
40. Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J., and Wüthrich K. (1991) Structure determination of the *Antp(C39→S)* homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217,** 531–540.
41. Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., and Gronenborn, A. M. (1986) The three-dimensional structure of a1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5,** 2729–2735.
42. Wüthrich, K., Billeter, M., and Braun, W. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169,** 949–961.

43. Güntert, P., Braun, W., Billeter, M., and Wüthrich, K. (1989) Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Am. Chem. Soc.* **111,** 3997–4004.
44. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1990) ^{1}H-NMR stereospecific assignments by conformational data-base searches. *Biopolymers* **29,** 813–822.
45. Wider, G., Lee, K. H., and Wüthrich, K. (1982) Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Glucagon bound to perdeuterated dodecylphosphocholine micelles. *J. Mol. Biol.* **155,** 367–388.
46. Wüthrich, K., Wider, G., Wagner, G., and Braun, W. (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.* **155,** 311–319.
47. Senn, H., Werner, B., Messerle, B. A., Weber, C., Traber, R., and Wüthrich, K. (1989) stereospecific assignment of the methyl ^{1}H NMR lines of valine and leucine in polypeptides by non-random ^{13}C labelling. *FEBS Lett.* **249,** 113–118.
48. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wüthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labelling. *Biochemistry* **28,** 7510–7516.
49. Hyberts, S. G., Märki, W., and Wagner, G. (1987) Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. *Eur. J. Biochem.* **164,** 625–635.
50. Weber, P. L., Morrison, R., and Hare, D. (1988) Determining stereo-specific ^{1}H nuclear magnetic resonance assignments from distance geometry calculations. *J. Mol. Biol.* **204,** 483–487.
51. Güntert, P., Braun, W., and Wüthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217,** 517–530.
52. Brünger, A. T. (1992) *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.
53. Havel, T. F. and Wüthrich, K. (1984) A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H-^{1}H proximities in solution. *Bull. Math. Biol.* **46,** 673–698.
54. Havel, T. F. and Wüthrich, K. (1985) An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J. Mol. Biol.* **182,** 281–294.
55. Blumenthal, L. M. (1970) *Theory and Applications of Distance Geometry*, Chelsea, New York.
56. Crippen, G. M. (1977) A novel approach to the calculation of conformation: distance geometry. *J. Comp. Physiol.* **24,** 96–107.

57. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry—dynamical simulated annealing calcula1tions. *FEBS Lett.* **229,** 317–324.
58. Metzler, W. J., Hare, D. R., and Pardi, A. (1989) Limited sampling of conformational space by the distance geometry algorithm: Implications for structures generated from NMR data. *Biochemistry* **28,** 7045–7052.
59. Havel, T. F. (1991) An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog. Biophys. Mol. Biol.* **56,** 43–78.
60. Kuszewski, J., Nilges, M., and Brünger, A. T. (1992) Sampling and efficiency of metrix matrix distance geometry: A novel partial metrization algorithm. *J. Biomol. NMR* **2,** 33–56.
61. Braun, W. and Gō, N. (1985) Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm. *J. Mol. Biol.* **186,** 611–626.
62. Braun, W. (1987) Distance geometry and related methods for protein structure determination from NMR data. *Q. Rev. Biophys.* **19,** 115–157.
63. Momany, F. A., McGuire, R. F., Burgess, A. W., and Scheraga, H. A. (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occuring amino acids. *J. Phys. Chem.* **79,** 2361–2381.
64. Némethy, G., Pottle, M. S., and Scheraga, H. A. (1983) Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occuring amino acids. *J. Phys. Chem.* **87,** 1883–1887.
65. Powell, M. J. D. (1977) Restart procedures for the conjugate gradient method. *Math. Progr.* **12,** 241–254.
66. Noguti, T. and Gō, N. (1983) A method for rapid calculation of a second derivative matrix of conformational energy for large molecules. *J. Phys. Soc. Jpn.* **52,** 3685–3690.
67. Abe, H., Braun, W., Noguti, T., and Gō, N. (1984) Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles in proteins. General recurrent equations. *Comput. Chem.* **8,** 239–247.
68. Kline, A. D., Braun, W., and Wüthrich, K. (1988) Determination of the complete three-dimensional structure of the α-amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **204,** 675–724.
69. Güntert, P. and Wüthrich, K. (1991) Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR* **1,** 446–456.
70. Verlet, L. (1967) Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159,** 98–103.
71. Allen, M. P. and Tildesley, D. J. (1987) *Computer Simulation of Liquids,* Clarendon, Oxford.

72. Brünger, A. T. and Nilges, M. (1993) Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q. Rev. Biophys.* **26,** 49–125.
73. Nilges, M., Kuszewski, J., and Brünger, A. T. (1991) Sampling properties of simulated annealing and distance geometry, in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy,* Plenum, New York, pp. 451–455.
74. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett.* **239,** 129–136.
75. van Gunsteren, W. F. and Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual,* Biomos, Groningen.
76. Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Chandra Singh, U., Weiner, P., and Kollman, P. A. (1991) AMBER 4.0, University of California, San Francisco.
77. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4,** 187–217.
78. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23,** 327–341.
79. Berendsen, J. H. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690.
80. Laskowski, R. A., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–291.
81. McLachlan, A. D. (1979) Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128,** 49–79.
82. Billeter, M., Schaumann, T., Braun, W., and Wüthrich, K. (1990) Restrained energy refinement with two different algorithms and force fields of the structure of the a-amylase inhibitor tendamistat determined by NMR in solution. *Biopolymers* **29,** 695–706.
83. Güntert, P., Berndt, K. D., and Wüthrich, K. (1993) The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. *J. Biomol. NMR* **3,** 601–606.
84. Weiner, P. K., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7,** 230–252.
85. McCammon, J. A. and Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids,* Cambridge University Press, Cambridge.
86. Brooks, C. L., Karplus, M., and Pettitt, B. M. (1988) *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics,* Wiley, New York.
87. Torda, A. E., Brunne, R. M., Huber, T., Kessler, H., and van Gunsteren, W. (1993) Structure refinement using time-averaged *J*-coupling constant constraints. *J. Biomol. NMR* **3,** 55–66.

88. Scheek, R. M., Torda, A. E., Kemmink, J., and van Gunsteren, W. F. (1991) Structure determination by NMR: The modeling of NMR parameters as ensemble averages, in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*, Plenum, New York, pp. 209–217.
89. Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., and Kaptein, R. (1989) Iterative procedure for structure determination from proton–proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson.* **82,** 290–308.
90. Borgias, B. A. and James, T. L. (1988) COMATOSE, a method for constrained refinements of macromolecular structure based on two-dimensional nuclear Overhauser spectra. *J. Magn. Reson.* **79,** 493–512.
91. Gonzalez, C., Rullmann, J. A. C., Bonvin, A. M. J. J., Boelens, R., and Kaptein, R. (1991) Toward an NMR *R* factor. *J. Magn. Reson.* **91,** 659–664.
92. Thomas, P. D., Basus, V. J., and James, T. L. (1991) Protein solution structure determination using distances from 2D NOE experiments: Effect of approximations on the accuracy of derived structures. *Proc. Natl. Acad. Sci. USA* **88,** 1237–1241.
93. Withka, J. M., Srinivasan, J., and Bolton, P. H. (1992) Problems with, and alternatives to, the NMR *R* factor. *J. Magn. Reson.* **98,** 611–617.
94. Ottiger, M., Szyperski, T., Luginbühl, P., Ortenzi, C., Luporini, P., Bradshaw, R. A., and Wüthrich, K. (1994) The NMR solution structure of the pheromone Er-2 from the ciliated protozoan *Euplotes raikovi. Prot. Sci.* **3,** 1515–1526.
95. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P., and Clore, G. M. (1991) The immunoglobulin binding domain of streptococcal protein G has a novel and highly stable polypeptide fold. *Science* **253,** 657–661.
96. Clore, G. M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A. M. (1990) Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **29,** 1689–1696.
97. Berndt, K. D., Güntert, P., and Wüthrich, K. (1993) The NMR solution structure of dendrotoxin K from the venom of *Dendroaspis polylepis polylepis. J. Mol. Biol.* **234,** 735–750.
98. Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program Xeasy for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **5,** 1–10.
99. Szyperski, T., Güntert, P., Stone, S. R., and Wüthrich, K. (1992) The NMR solution structure of hirudin(1–51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J. Mol. Biol.* **228,** 1193–1205.
100. Rydel, T. J., Tulinsky, A., Bode, W., and Huber, R. (1991) Refined structure of the hirudin–thrombin complex. *J. Mol. Biol.* **221,** 583–601.
101. Eccles, C., Güntert, P., Billeter, M., and Wüthrich, K. (1991) Efficient analysis of protein 2D NMR spectra using the software package EASY. *J. Biomol. NMR* **1,** 111–130.

6

Studies of Protein–Ligand Interactions by NMR

David J. Craik and Jacqueline A. Wilce

1. Introduction

NMR spectroscopy is recognized widely as an invaluable tool for the structural characterization of biological macromolecules with molecular weights of less than approx 25 kDa. The quality of structures obtainable using NMR spectroscopic methods is comparable with those derived from X-ray crystallography but, in addition, NMR offers the possibility of obtaining quantitative information on molecular flexibility. A particularly important aspect of the dynamics of macromolecules is that of ligand binding. Such binding can be accompanied by conformational changes in either the ligand, the macromolecule, or both and, in many cases, such dynamic changes are crucial to the functioning of the macromolecular system. This chapter is concerned with the use of NMR to define the nature of specific protein–ligand interactions. Although the focus is on interactions of ligands with proteins, rather than with other biological macromolecules, such as DNA or membranes, many of the techniques applicable to studies of protein–ligand interactions generally are applicable also to other macromolecular interactions. The topic of protein–ligand interactions has also been addressed from a number of different viewpoints in several other recent reviews *(1–7)*.

1.1. Importance of Protein–Ligand Interactions

Protein–ligand interactions are integral to a wide range of biological processes, including hormone, neurotransmitter or drug binding, antigen recognition, and enzyme–substrate interactions. Fundamental to each of these interactions is the recognition by a ligand of a unique binding surface where it binds in a defined way to carry out its function. Through an understanding of these specific interactions it may be possible to design or discover analogous ligands with altered binding properties and therefore to intervene in the bio-

From: *Methods in Molecular Biology, Vol. 60: Protein NMR Techniques*
Edited by: D. G. Reid Humana Press Inc., Totowa, NJ