COMPUTER-SUPPORTED PROTEIN STRUCTURE DETERMINATION BY NMR

Peter Güntert

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

INTRODUCTION

The relationship between amino acid sequence, three-dimensional structure and biological function of proteins is one of the most intensely pursued areas of molecular biology and biochemistry. In this context, the three-dimensional (3D) structure has a pivotal role, as its knowledge is a prerequisite to understanding the physical, chemical, and biological properties of a protein (Schulz and Schirmer, 1979). Until 1984, structural information at atomic resolution could only be determined by X-ray diffraction techniques with protein single crystals (Blundell and Johnson, 1976). The introduction of nuclear magnetic resonance (NMR) spectroscopy as a technique for protein structure determination has made it possible to obtain structures at comparable resolution in solution (Wüthrich, 1986).

The utility of the NMR method for the study of molecular structures depends to a large part on the sensitive variation of the resonance frequency of a nucleus with the chemical structure, the conformation of the molecule, and the solvent environment (Ernst *et al.*, 1987). These so-called chemical shifts ensure the necessary spectral resolution, although they do not provide direct structural information. They arise because a given nucleus is shielded from the externally applied magnetic field to a different extent depending on its local environment. Three of the four most abundant elements in biological materials, *i.e.*, hydrogen, carbon and nitrogen, have naturally occurring isotopes with nuclear spin $I = 1/2$, and are therefore suitable for high-resolution NMR experiments in solution (Wüthrich, 1986). The proton (^{1}H) has the highest natural abundance (99.98%) and the highest sensitivity (due to its large gyromagnetic ratio) among these isotopes, and hence plays a central role in most NMR experiments with biopolymers. The low natural abundance of ^{13}C and ^{15}N (1.11% and 0.37% respectively) and low relative sensitivity of these nuclei normally require isotope enrichment. Today, this is routinely achieved by overexpression of proteins in isotope-labelled media, which is indispensable for structure determination with larger systems (Billeter *et al.*, 1993; Clore and Gronenborn, 1991). Structures of small proteins with molecular weight ≤ 10,000 can be efficiently determined by ^{1}H NMR alone.

The procedure used for the determination of three-dimensional protein structures by NMR is outlined in Fig. 1. It includes preparation of the protein, NMR experiments, obtain-

Statistical Mechanics, Protein Structure, and Protein Substrate Interactions
Edited by S. Doniach, Plenum Press, New York, 1994

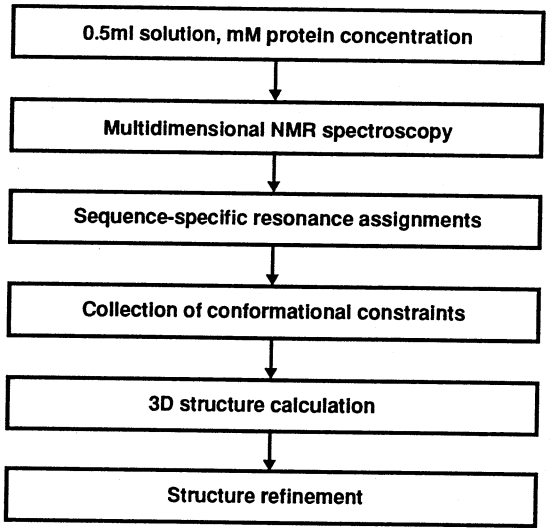

Figure 1. Outline of the procedure for NMR structure determination of proteins in solution.

ing assignments of the NMR lines to individual atoms in the polypeptide chain, calculation and refinement of the three-dimensional structure.

In practice, the data for the structure determination of a small protein are collected by performing homonuclear two-dimensional (2D) ^{1}H NMR experiments (Wüthrich, 1986). Typically, these spectra contain diagonal peaks (Fig. 2), with $\omega_1 = \omega_2$, which correspond to a conventional 1D spectrum. In addition there are a large number of off-diagonal cross peaks with $\omega_1 \neq \omega_2$. Each cross peak establishes a correlation between two protons. For example, in 2D nuclear Overhauser enhancement spectroscopy (NOESY; Anil-Kumar *et al.*, 1981) the cross peaks represent "trough-space" nuclear Overhauser effects (NOEs) and indicate that the two correlated protons are separated by a short distance (< 5 Å) in the three-dimensional structure. In 2D correlated spectroscopy (COSY; Rance *et al.*, 1983), the cross peaks represent "through-bond" scalar spin-spin couplings. For a protein, a COSY cross peak thus indicates that two protons are separated by at most three chemical bonds in the covalent structure. Using suitable NMR experiments with H_2O and/or D_2O solutions of the protein, one can obtain a complete or nearly complete set of sequence-specific resonance assignments (Wüthrich, 1986).

Once resonance assignments are available for all or nearly all polypeptide protons, one can start to collect conformational constraints for the determination of the 3D structure: NOEs which are converted into distance constraints, and spin-spin coupling constants which yield information on dihedral angles. The most important input for a structure calculation are upper bounds on ^{1}H-^{1}H distances derived from NOESY experiments (Fig. 3). The intensity of a NOESY cross peak is related to the distance r between the two correlated protons by

$$\mathrm{NOE} \propto \langle r^{-6} \rangle f(\tau_c) \qquad (1)$$

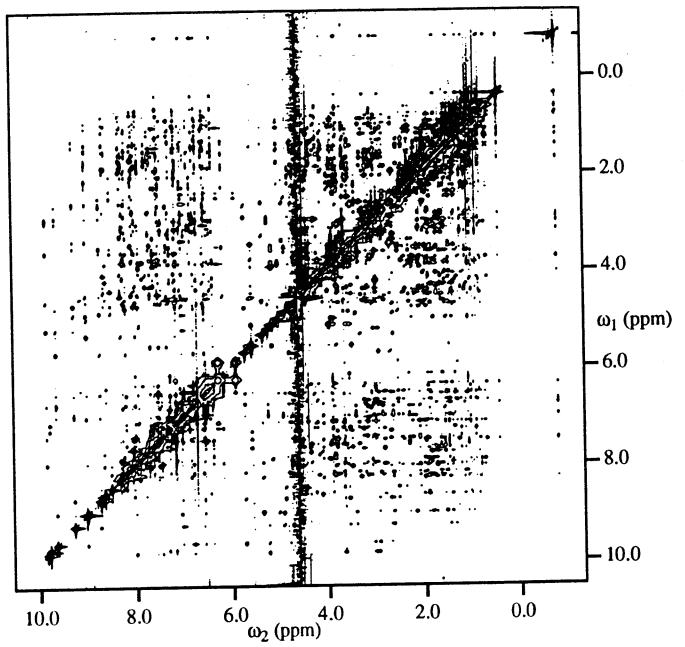

Figure 2. Two-dimensional ^{1}H NOESY spectrum of Toxin K in H_2O at pH 4.6 and 36 °C (protein concentration 10 mM, mixing time τ_m = 40 ms; Berndt *et al.*, 1993).

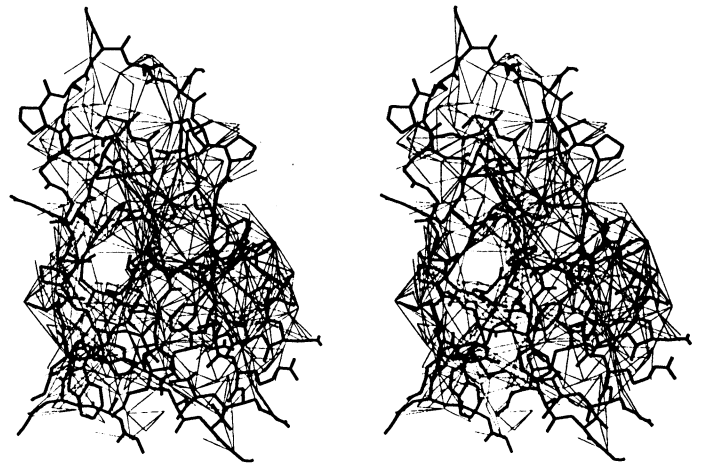

Figure 3. Stereo view of an all-heavy-atom representation of the protein Toxin K (bold lines). Each of the 809 upper distance constraints used in the structure calculation (Berndt *et al.*, 1993) is shown by a thin line connecting the two protons (or pseudo-atoms) involved in the constraint.

where the second term is a function of the correlation time, τ_c, which accounts for the influence of motional averaging processes on the observed NOE. $f(\tau_c)$ is governed by the combined effects of overall rotational tumbling and intramolecular mobility, and may vary for different locations in a protein molecule. Therefore only upper bounds on ^{1}H-^{1}H distances are usually derived from NOE measurements (Wüthrich, 1986). In a systematic analysis of the local conformation, measurements of vicinal spin-spin coupling constants $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ may be combined with intraresidual and sequential NOEs to restrict the dihedral angles ϕ, ψ, and χ^1 and to obtain stereospecific assignments for β-methylene protons (Güntert *et al.*, 1989).

STRUCTURE CALCULATION

The network of distance constraints between spatially proximate hydrogen atoms (Fig. 3) and angle constraints derived from the NMR experiments constitutes the input for the structure calculation. In the early stages of the development of the NMR method for protein structure determination it immediately followed from the nature of the conformational constraints that the techniques for structure determination from other experimental data, in particular X-ray diffractions, could not be adapted for the structural analysis of the NMR data, and hence new ways had to be developed. Initially, metric matrix distance geometry algorithms were used (Braun *et al.*, 1981; Havel and Wüthrich, 1984). Subsequent work included a variable target function algorithm (Braun and Gō, 1985), interactive molecular modelling using computer graphics (Billeter *et al.*, 1985), and restrained molecular dynamics (Brünger *et al.*, 1986) in conjunction with model building (Kaptein *et al.*, 1985) or distance geometry calculations (Clore *et al.*, 1985). The following procedures are currently mostly employed: Embedding using metric matrix distance geometry, followed by simulated annealing using molecular dynamics (e.g., Driscoll *et al.*, 1989; Lee *et al.*, 1989) and structure calculation using the variable target function algorithm, supplemented by energy minimization (e.g., Berndt *et al.*, 1992; Qian *et al.*, 1989).

NMR structures of proteins in solution are commonly presented as a group of conformers, each of which has been calculated individually from the same experimental input data. In a high-quality structure determination each individual conformer has small residual violations of the conformational constraints, and the root mean square deviations (RMSD) among all conformers in the group are small (Braun, 1987; Brünger and Nilges, 1993). In this chapter we describe a new strategy for the use of the variable target function program DIANA (Güntert *et al.*, 1991a) that enables efficient calculation of such groups of conformers.

Variable Target Function Algorithm

The basic idea of the variable target function algorithm (Braun and Gō, 1985) is to *gradually* fit an initially randomized starting structure to the conformational constraints collected with the use of NMR experiments, starting with intraresidual constraints only, and increasing the "target size" step-wise up to the length of the complete polypeptide chain. Advantages of the method are its conceptual simplicity and the fact that it works in dihedral angle space, so that the covalent geometry is preserved during the entire calculation.

The algorithm used by the program DIANA (Güntert *et al.*, 1991a) is based on the minimization of a variable target function $T(\phi_1, ..., \phi_n)$, where the n degrees of freedom are the dihedral angles $\phi_1, ..., \phi_n$ about rotatable bonds of the polypeptide chain. During the calculation the bond lengths, bond angles and chiralities of the covalent structure are kept fixed at the ECEPP standard values (Momany *et al.*, 1975; Némethy *et al.*, 1983). The target function T, with $T \geq 0$, is defined such that $T = 0$ if and only if all experimental distance and dihedral angle constraints are fulfilled and all non-

bonded atom pairs satisfy a check for the absence of steric overlap. $T(\phi_1, ..., \phi_n) \leq T(\phi'_1, ..., \phi'_n)$ if the conformation $(\phi_1, ..., \phi_n)$ satisfies the constraints better than the conformation $(\phi'_1, ..., \phi'_n)$. The problem to be solved is to find values $(\phi_1, ..., \phi_n)$ that yield low values of the target function.

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 constraints with respect to the polypeptide sequence are considered, and in subsequent rounds of calculations, constraints between atoms further apart with respect to the primary structure are included in a step-wise fashion (Fig. 4). 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 towards the end of the calculation (Fig. 5).

Two different kinds of constraints are considered by the target function (Wüthrich, 1986; Braun, 1987): 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 $\phi_a^{min} < \phi_a^{max} < \phi_a^{min} + 2\pi$. 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 constraint violation. I_o, I_u and I_v denote the sets of atom pairs αβ, 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_a the set of restrained dihedral angles, w_u, w_l, w_v and w_a weighting factors for the different types of constraints, $d_{\alpha\beta} = d_{\alpha\beta}(\phi_1, ..., \phi_n)$ 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 (Güntert *et al.*, 1991a):

$$T = \sum_{c=u,l,v} w_c \sum_{\alpha\beta \in I'_c} \left(\frac{\Theta_c(d_{\alpha\beta}^2 - c_{\alpha\beta}^2)}{2c_{\alpha\beta}} \right)^2 + w_w \sum_{a \in I_a} \left(1 - \frac{1}{2} \left(\frac{\Delta_a}{\Gamma_a} \right)^2 \right) \Delta_a^{2*} \qquad (2)$$

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 constraints between residues whose sequence numbers differ by not more than a given *minimization level L* (Fig. 4). The target function is continuously differentiable 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 constraints. Because only squared interatomic distances and no square roots have to be computed, the target function can be calculated rapidly.

A drawback 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 constraint violations, which is a typical local minimum problem. Because of the low yield of acceptable conformers, calculations have typically 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, the availability of approximately 10–20 "good" conformers to represent the solution conformation, and the available computing time (Kline *et al.*, 1988). 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 (Güntert *et al.*, 1991b), but for β-proteins with more complex topology the situation remained unsatisfactory. With the use of redundant dihedral angle constraints (REDAC) described in this chapter, a greatly improved yield of converged conformers is obtained also for β-proteins.

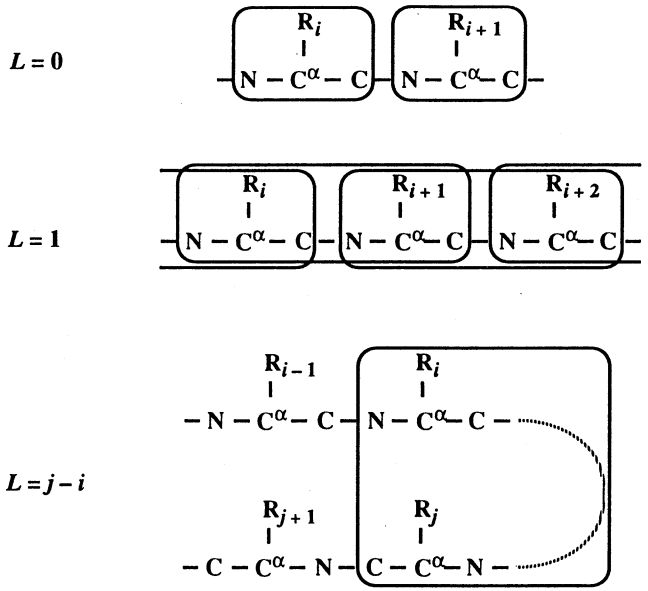

Figure 4. Illustration of the variable target function algorithm as implemented in the program DIANA (Güntert *et al.*, 1991). Boxes indicate examples of allowed ranges of active constraints at various minimization levels *L*.

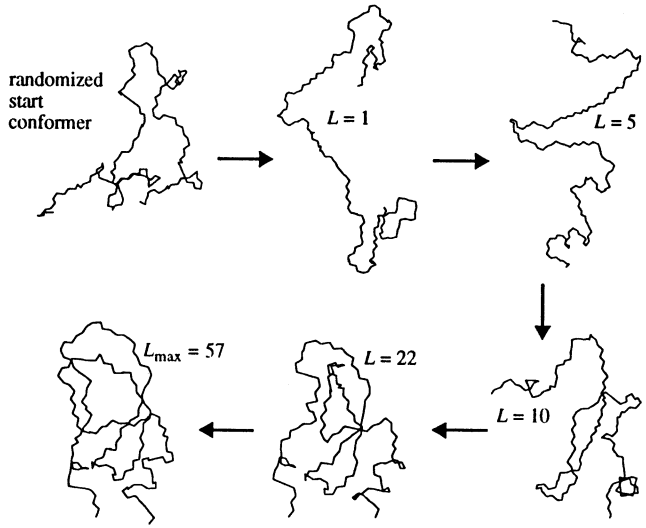

Figure 5. Randomized starting conformer, intermediate structures and final, converged structure of the protein Toxin K during variable target function minimization with the program DIANA. The backbone atoms N, C^{α}, and C' for all 57 residues of the random start conformation, and of the conformations at the end of the minimization levels $L = 1, 5, 10, 22$, and $L_{max} = 57$ are shown (Berndt *et al.*, 1993).

Use of Redundant Dihedral Angle Constraints

In Fig. 6 the use of DIANA with redundant dihedral angle constraints (REDAC; Güntert and Wüthrich, 1991) is outlined and placed in perspective with the "direct" variable target function method as proposed originally by Braun and Gō (1985). 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 further reduced 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 $m = 10$–20 acceptable conformers.

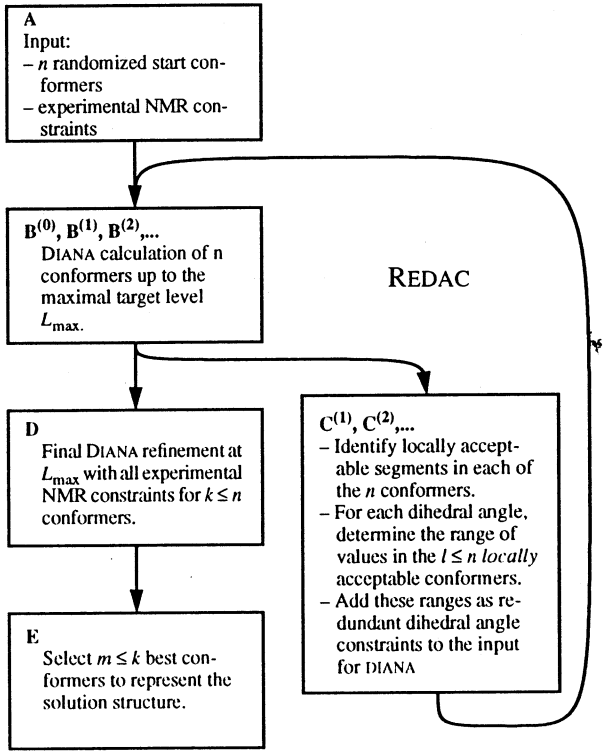

Figure 6. Flowchart outlining the course of a protein structure calculation with the program DIANA using either the "direct" way (A-$B^{(0)}$-D-E) or redundant dihedral angle constraints (REDAC; Güntert and Wüthrich, 1991) (A-$B^{(0)}$-[$C^{(1)}$-$B^{(1)}$-...]-D-E). Typically, the number of REDAC-cycles is 1 or 2.

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 due 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 neighbouring 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.

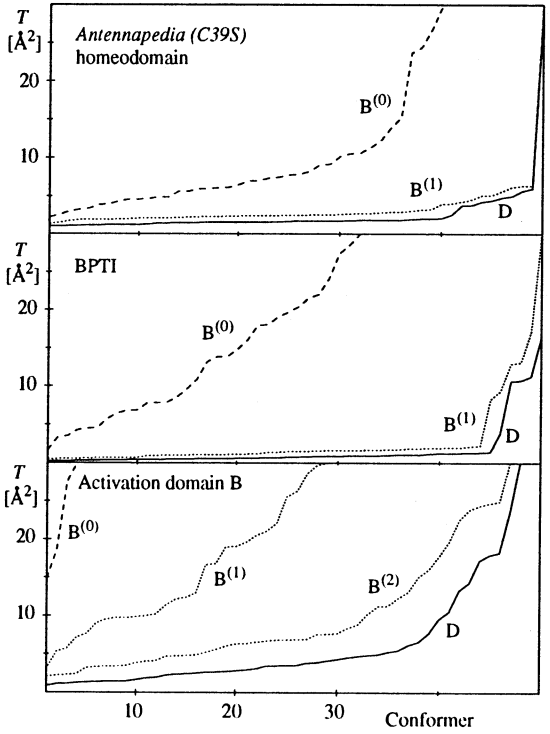

Figure 7. Distribution of the DIANA target function values, T, among the 50 conformers calculated using REDAC for each of the three proteins *Antennapedia(C39S)* homeodomain (HD), BPTI, and activation domain B (ADB). Along the horizontal axis the 50 conformers are ordered according to increasing target function value. For the HD and for BPTI the calculation consisted of the sequence of steps $A–B^{(0)}–C^{(1)}–B^{(1)}–D$; for ADB the sequence of steps was $A–B^{(0)}–C^{(1)}–B^{(1)}–C^{(2)}–B^{(2)}–D$. The letters identify the results at the end of the respective step in Fig. 6.

To compare the efficiency of DIANA calculations with and without use of REDAC, we calculated structures from high-quality experimental NMR data sets for the *Antennapedia (C39S)* homeodomain (HD; Güntert *et al.*, 1991b), the basic pancreatic trypsin inhibitor (BPTI; Berndt *et al.*, 1992), and the activation domain from porcine procarboxypeptidase B (ADB; Vendrell *et al.*, 1991). The HD is a typical α-protein, BPTI contains α and β secondary structure, and ADB has a more complex topology including a four-stranded β-sheet, two α-helices, and three loops that are only poorly determined by the NMR data (Vendrell *et al.*, 1991). For each protein we performed a structure calculation starting with $n = 50$ randomized conformations and using REDAC, and selected the $m = 20$ conformers with the smallest final

target function values for further analysis (Güntert and Wüthrich, 1991). For the HD and BPTI more than 40 conformers with final target function values at L_{max} below 2.1 Å^2 and 1.3 Å^2, respectively, were obtained after one REDAC cycle. For ADB two REDAC cycles were needed to yield a group of 20 conformers with target function values below 2.9 Å^2 (Fig. 7)

For the HD and BPTI the DIANA calculations were repeated with the direct approach (Fig. 6), with the aim of producing a group of 20 conformers of equal quality, i.e., with final target function values in the same range as the 20 best conformers obtained from 50 starting conformers with the use of REDAC. For the HD this was achieved with $n = 400$ starting conformers, for BPTI with $n = 2000$ (Table 1). For the ADB it was found that calculations without use of REDAC produced only one acceptable converged conformer from 400 starting conformers, so that we would have had to compute of the order of 8000 conformers in order to obtain a comparable result. Table 1 shows that for the three proteins the effective overall CPU time was reduced through the use of REDAC by factors of 5.7, 29, and about 100, respectively. The improved efficiency when using REDAC is particularly pronounced for the β-proteins, i.e., those proteins where the direct approach gives the lowest yields of acceptable structures.

Table 1. Efficiency of DIANA calculations with and without use redundant dihedral angle constraints (REDAC).

	Antennapedia(C39S) homeodomain		BPTI		activation domain B	
	direct	REDAC	direct	REDAC	direct[1]	REDAC
Number of start conformers,[2] n	400	50	2000	50	≈8000	50
CPU time (h)[3]	3.8	0.66	17.7	0.61	≈140	1.48

[1]Estimated (see text).
[2]The numbers n listed under "direct" were chosen so as to obtain the same number of acceptable conformers as with $n = 50$ and use of REDAC.
[3]Measured on a Cray Y/MP using one processor.

For both the HD and BPTI the DIANA structure calculations with and without REDAC gave nearly identical values for the final target function, the different types of residual constraint violations, and the global pairwise RMSDs (McLachlan, 1979) among the 20 best conformers. In addition, for both proteins the global RMSDs between the two groups of conformers calculated with and without use of REDAC are for all practical purposes identical to the RMSDs calculated between different conformers within each group (Güntert and Wüthrich, 1991).

CONCLUSIONS

The empirically found higher yield of good conformers with the use of REDAC 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 (Wüthrich, 1986), 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 already available 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. It further 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 is particularly well-determined by sequential and medium-range constraints (Wüthrich, 1986).

In conclusion, the success of DIANA structure calculations using REDAC is primarily due 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 during the entire duration of the structure calculation is used in obtaining the final result, whereas most of this information is discarded in the direct approach. Groups of conformers of equal quality are obtained with and without use of REDAC, and the only significant effect of the use of REDAC is a large reduction of the overall computation time (Table 1). The use of REDAC should therefore become the standard strategy for protein structure calculations with the program DIANA and, more generally, with all implementations of the variable target function algorithm.

ACKNOWLEDGMENTS

I thank Dr. Martin Billeter and Prof. Kurt Wüthrich for helpful discussions. Financial support by the Schweizerischer Nationalfonds (project 31.32033.91) and the use of the Cray Y/MP computer of the ETH Zürich are gratefully acknowledged.

REFERENCES

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. Amer. Chem. Soc.* 103:3654.

Berndt, K. D., Güntert, P., and Wüthrich, K., 1993, Nuclear magnetic resonance solution structure of dendrotoxin K from the venom of *Dendroaspis polylepis polylepis*, *J. Mol. Biol.* 234, in press.

Berndt, K. D., Güntert, P., Orbons, L. P. M. and Wüthrich, K., 1992, Determination of a high-quality NMR solution structure of the bovine pancreatic trypsin inhibitor (BPTI) and comparison with three crystal structures, *J. Mol. Biol.* 227:757.

Billeter, M., Engeli, M. and Wüthrich, K., 1985, Interactive program for investigation of protein structures based on ^{1}H NMR experiments, *J. Mol. Graph.* 3:79.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K., 1993, Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain-DNA complex, *J. Mol. Biol.* 234, in press.

Blundell, T. L. and Johnson, L. N., 1976, "Protein Crystallography," Academic, New York.

Braun W., 1987, Distance geometry and related methods for protein structure determination from NMR data, *Q. Rev. Biophys.* 19:115.

Braun, W. and Gō, N., 1985, Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *J. Mol. Biol.* 186:611.

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.

Brünger, A. T. and Nilges, M., 1993, Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy, *Q. Rev. Biophys.* 26:1.

Brünger, A. T., Clore, G. M., Gronenborn, A. M. and Karplus, M., 1986, Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin, *Proc. Nat. Acad. Sci., U.S.A.* 83:3801.

Clore, G. M. and Gronenborn, A. M., 1991, Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy, *Science* 252:1390.

Clore, G. M., Gronenborn, A. M., Brünger, A. T. and Karplus, M., 1985, Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of Escherichia coli. Combined use of ^{1}H nuclear magnetic resonance and restrained molecular dynamics, *J. Mol. Biol.* 186:435.

Driscoll, P. C., Gronenborn, A. M., Beress, L. and Clore, G. M., 1989, Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from sea anemone *Anemonia sulcata:* a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, *Biochemistry* 28:2188.

Ernst, R. R., Bodenhausen, G. and Wokaun, A., 1987, "The principles of nuclear magnetic resonance in one and two dimensions," Clarendon, Oxford.

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.

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. Amer. Chem. Soc.* 111: 3997.

Güntert, P., Braun, W. and Wüthrich, K., 1991a, 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.

Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. and Wüthrich K., 1991b, 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.

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.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. and van Gunsteren, W. F., 1985, A protein structure from nuclear magnetic resonance data. Lac repressor headpiece, *J. Mol. Biol.* 182:179.

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.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. and Wright, P. E., 1989, Three-dimensional solution structure of a single zinc finger DNA-binding domain, *Science* 245:635.

McLachlan, A. D., 1979, Gene duplication in the structural evolution of chymotrypsin, *J. Mol. Biol.* 128:49.

Momany, F. A., McGuire, R. F., Burgess, A. W. & 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.

Némethy, G., Pottle, M. S. & 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.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. and Wüthrich, K., 1989, The structure of the *Antennapedia* homeodomain determined by NMR in solution: comparison with prokaryotic repressors, *Cell* 59:573.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. and Wüthrich, K., 1983, Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Comm.* 117:479.

Schulz, G. E. and Schirmer, R. H., 1979, "Principles of Protein Structure," Springer, New York.

Vendrell, J., Billeter, M., Wider, G., Avilés, F. X. and Wüthrich, K., 1991, The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *EMBO J.* 10:11.

Wüthrich, K., 1986, "NMR of Proteins and Nucleic Acids," Wiley, New York.