Journal of Biomolecular NMR, 1 (1991) 257–269
ESCOM

J-Bio NMR 020

Complete relaxation matrix refinement of NMR structures of proteins using analytically calculated dihedral angle derivatives of NOE intensities

John E. Mertz*, Peter Güntert, Kurt Wüthrich and Werner Braun

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

Received 25 March 1991
Accepted 7 May 1991

Keywords: Relaxation matrix refinement; Analytically calculated derivatives of NOE intensities; Distance geometry; R-factor; Protein structure calculation

SUMMARY

A new method for refining three-dimensional (3D) NMR structures of proteins is described, which takes account of the complete relaxation pathways. Derivatives of the NOE intensities with respect to the dihedral angles are analytically calculated, and efficiently evaluated with the use of a filter technique for identifying the dominant terms of these derivatives. This new method was implemented in the distance geometry program DIANA. As an initial test, we refined 30 rigid distorted helical structures, using a simulated data set of NOE distance constraints for a rigid standard α-helix. The final root-mean-square deviations of the refined structures relative to the standard helix were less than 0.1 Å, and the R-factors dropped from values between 7% and 32% to values of less than 0.5% in all cases, which compares favorably with the results from distance geometry calculations. In particular, because spin diffusion was not explicitly considered in the evaluation of 'exact' $^1H-^1H$ distances corresponding to the simulated NOE intensities, a group of nearly identical distance geometry structures was obtained which had about 0.5 Å root-mean-square deviation from the standard α-helix. Further test calculations using an experimental NOE data set recorded for the protein trypsin inhibitor K showed that the complete relaxation matrix refinement procedure in the DIANA program is functional also with systems of practical interest.

INTRODUCTION

For the first phase of the determination of the three-dimensional (3D) structure of biological macromolecules in solution from NMR data, one relies usually on distance geometry methods

Abbreviations: RMSD, root-mean-square deviation; NOE, nuclear Overhauser enhancement; NOESY, 2-dimensional nuclear Overhauser enhancement spectroscopy; CPU, central processing unit.

*On sabbatical leave from Cray Research, Inc., 655E Lone Oak Road, Eagan, MN 55121, U.S.A.

(Braun et al., 1981, 1983; Havel and Wüthrich, 1984, 1985; Braun and Gō, 1985; Wüthrich, 1986; Braun, 1987; Kaptein et al., 1988; Bax, 1989; Clore and Gronenborn, 1989a; Kuntz et al., 1989). The distance geometry structures are then refined either by restrained energy minimization (Billeter et al., 1990; Schaumann et al., 1990) or by restrained molecular dynamics calculations (Kaptein et al., 1985; Brünger et al., 1986). In the preparation of the input of distance constraints corresponding to the measured NOE intensities, both spin diffusion effects and the influence of internal mobility are usually either neglected or implicitly accounted for by a conservative calibration of the relation between NOEs and ^1H–^1H distances (Braun et al., 1981; Wüthrich, 1986; Kline et al., 1988). A quantitative assessment of the consequences of this 'initial slope approximation' on the NMR structures calls for a complete relaxation matrix treatment with dynamic molecular models. This problem has been discussed extensively on the level of ^1H–^1H distances or NMR relaxation rates (Clore and Gronenborn, 1989b; Madrid et al., 1989; Baleja et al., 1990; Borgias et al., 1990; Koehl and Lefèvre, 1990; Post et al., 1990), but proper answers will have to rely on the development of robust refinement programs based on complete relaxation matrix treatments. Currently available procedures (Dobson et al., 1982; Keepers and James, 1984; Olejniczak et al., 1986; Borgias and James, 1988; Boelens et al., 1989; Yip and Case, 1989; Baleja et al., 1990; Borgias et al., 1990) work either in distance space or in Cartesian space. For example, Keepers and James (1984) and Boelens et al. (1989) proposed distance space methods that yield an improved set of distances from the measured NOE data and seem to be quite robust with respect to initial trial structures. However, to improve the atomic coordinates of the molecular structure, these investigators still rely on distance geometry or restrained molecular dynamics. For the real-space methods (Borgias and James, 1988; Yip and Case, 1989; Baleja et al., 1990) practical experience with the refinement of protein structures still has to be gathered, and a gradient-based refinement technique has been described only very recently (Yip and Case, 1989). As a further contribution to the techniques available for investigations of the aforementioned fundamental problems, the present paper proposes the use of analytically calculated derivatives of the NOE intensities with respect to the intervening dihedral angles implemented for structure refinements with the program DIANA (Güntert et al., 1991a).

The method introduced here takes the complete relaxation pathways into account. Evaluation of the derivatives is based on the rapid calculation of derivatives in dihedral angle space by Abe et al. (1984) and on the calculation of derivatives in Cartesian coordinates as described by Yip and Case (1989). It has some similarity to previous relaxation matrix treatments (Borgias and James, 1988), but a decisive advantage of the present approach lies in the fact that it uses *analytically* calculated derivatives in dihedral angle space. We describe here the mathematical basis for the analytical calculation of the derivatives and its implementation in the program DIANA. We then show that the extended version of DIANA can be used in practice for refinement of proteins, first with a simulated data set of NOE distance constraints for a rigid helical polypeptide, and then with a trial refinement of a group of distance geometry structures of the trypsin inhibitor K using an experimental data set.

METHODS

Evaluation of derivatives with respect to dihedral angles

In a NOESY experiment, the cross relaxation during the mixing time τ_m can be described by the Bloch equations (Solomon, 1955; Macura and Ernst, 1980)

$$\frac{d\tilde{M}}{dt} = R\tilde{M}(t) \tag{1}$$

$\tilde{M}(t)$ is the difference between the magnetization at time t and the thermal equilibrium magnetization, and the tilde denotes that the sum of the magnetizations of a group of equivalent spins is considered. The relaxation rates, R_{st}, are the result of zero-quantum, single-quantum and double-quantum transitions induced by dipole-dipole interactions among all protons of the polypeptide chain, except that the leakage relaxation rate, R_s^l, accounts for all other relaxation mechanisms (Tropp, 1980; Ernst et al., 1987)

$$R_{st} = \sqrt{n_s n_t}\,\frac{\kappa_1}{r_{st}^6}\,(s \neq t) \quad R_{ss} = n_s \sum_{t \neq s} \frac{\kappa_2}{r_{st}^6} + R_s^l \tag{2}$$

where

$$\kappa_1 = \frac{\gamma^4 \hbar^2 \tau_c}{10}\left(\frac{6}{1 + 4\omega^2\tau_c^2} - 1\right)$$
$$\kappa_2 = \frac{\gamma^4 \hbar^2 \tau_c}{10}\left(1 + \frac{3}{1 + \omega^2\tau_c^2} + \frac{6}{1 + 4\omega^2\tau_c^2}\right) \tag{3}$$

The indices s and t run over all groups of chemical shift-equivalent spins, and n_s is the number of equivalent spins per group (e.g., $n_s = 1$ for single protons, $n_s = 3$ for methyl groups). ω is the Larmor frequency, and τ_c is the effective rotational correlation time. In all calculations described in this paper, R_s^l is set to zero. Since leakage relaxation can to a good approximation be treated as independent of the molecular conformation, it can readily be added for the treatment of actual experimental data. Equations (2) and (3) ensure that the relaxation matrix is symmetric, i.e., $R_{st} = R_{ts}$. The solution of Eq. (1) is obtained by diagonalizing the rate matrix, R (Dobson et al., 1982; Keepers and James, 1984; Olejniczak et al., 1986)

$$\tilde{M}(\tau_m) = O e^{-\lambda \tau_m} O^{-1} \tilde{M}(0) \tag{4}$$

where λ is the diagonal matrix of eigenvalues λ_i, and O is the orthogonal matrix of eigenvectors of the relaxation matrix R. The calculated NOE cross-peak intensities, Icij, are the elements of the matrix $O e^{-\lambda \tau_m} O^{-1}$

$$I_{ij}^c = \sum_k O_{ik} O_{kj}^{-1} e^{-\lambda_k \tau_m} \tag{5}$$

To refine the structure against the measured NOE intensities, the following term T will be minimized:

$$T = \sum_{i,j} k(I^o_{ij} - qI^c_{ij})^2 \tag{6}$$

where the sum runs over the observed NOE intensities, and I^o_{ij} and I^c_{ij} are the observed and the calculated NOE cross-peak intensities, respectively. The scaling factor q is a proportionality constant between the measured and calculated NOE intensities, and is chosen such that the target function is minimized for given values of I^o_{ij} and I^c_{ij}:

$$q = \frac{\sum_{i,j} I^o_{ij} I^c_{ij}}{\sum_{i,j} I^c_{ij} I^c_{ij}} \tag{7}$$

As is described in the following section, the full relaxation matrix minimization was implemented in the distance geometry program DIANA (Güntert et al., 1991a), which works in dihedral angle space. To this end, the term T in Eq. (6) has to be written as a function of the dihedral angles. Its derivative with respect to a dihedral angle ϕ_a thus becomes

$$\frac{\partial T}{\partial \phi_a} = \sum_{i,j} h_{ij} \frac{\partial I^c_{ij}}{\partial \phi_a} \tag{8}$$

where $h_{ij} = -2kq(I^o_{ij} - qI^c_{ij})$. For distance geometry calculations in dihedral angle space it is essential to have a powerful method for calculating derivatives with respect to dihedral angles (Braun and Gō, 1985). Abe et al. (1984) have shown that, for a symmetric pair potential $\varphi(r_{\alpha\beta})$, where $r_{\alpha\beta} = |\mathbf{r}_\alpha - \mathbf{r}_\beta|$ and $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ are the position vectors of the atoms α and β, the dihedral angle derivatives can efficiently be evaluated through general recurrent equations. Parts of this method can be adapted for use with the more complex situation of Eq. (6).

Using Eq. (5) in Abe et al. (1984), we express Eq. (8) as

$$\frac{\partial T}{\partial \phi_a} = -\mathbf{e}_a \cdot \sum_{\substack{\alpha \in M_a \\ \beta \in \overline{M}_a}} c_{\alpha\beta}(\mathbf{r}_\alpha \wedge \mathbf{r}_\beta) - (\mathbf{e}_a \wedge \mathbf{r}_{\in(a)}) \cdot \sum_{\substack{\alpha \in M_a \\ \beta \in \overline{M}_a}} c_{\alpha\beta}(\mathbf{r}_\alpha - \mathbf{r}_\beta) \tag{9}$$

where $\mathbf{e}_a$ is a unit vector along the rotatable bond a, and $\mathbf{r}_{\in(a)}$ is the end point of the rotatable bond a. M_a is the set of all those atoms for which the coordinates are affected by a change in ϕ_a, and $\overline{M}_a$ is the set of all other atoms in the molecule. The calculation of the derivative therefore reduces to the evaluation of $c_{\alpha\beta} = \frac{1}{r_{\alpha\beta}} \frac{d\varphi(r_{\alpha\beta})}{dr_{\alpha\beta}}$:

$$c_{\alpha\beta} = \sum_{i,j} \frac{h_{ij}}{r_{\alpha\beta}} \frac{\partial I^c_{ij}}{\partial r_{\alpha\beta}} \tag{10}$$

Following Yip and Case (1989), the expression $\frac{\partial I_{ij}^c}{\partial r_{\alpha\beta}}$ is

$$\frac{\partial I_{ij}^c}{\partial r_{\alpha\beta}} = \sum_{r,s,t,u} O_{ir} O_{rs}^{-1} \frac{\partial R_{st}}{\partial r_{\alpha\beta}} O_{tu} O_{uj}^{-1} f_{ru} \tag{11}$$

where

$$f_{ru} = \frac{e^{-\lambda_r \tau_m} - e^{-\lambda_u \tau_m}}{\lambda_r - \lambda_u} (r \neq u) \quad f_{rr} = -\tau_m e^{-\lambda_r \tau_m} \tag{12}$$

The elements of $\frac{\partial R_{st}}{\partial r_{\alpha\beta}}$ are evaluated with the use of Eq. (2), and from Eqs. (10) and (11) we obtain

$$c_{\alpha\beta} = -6 r_{\alpha\beta}^{-8} \sum_{i,j} \sum_{r,u} h_{ij} O_{ir} O_{uj}^{-1} g_{ru}^{\alpha\beta} f_{ru} \tag{13}$$

where

$$g_{ru}^{\alpha\beta} = \sqrt{n_\alpha n_\beta} \kappa_1 (O_{r\alpha}^{-1} O_{\beta u} + O_{r\beta}^{-1} O_{\alpha u}) + \kappa_2 (n_\alpha O_{r\beta}^{-1} O_{\beta u} + n_\beta O_{r\alpha}^{-1} O_{\alpha u}) \tag{14}$$

The summations over r and u are both over the dimension of the relaxation matrix. Because $g_{ru}^{\alpha\beta} = g_{ur}^{\alpha\beta}$, Eq. (13) can be rewritten as

$$c_{\alpha\beta} = -6 r_{\alpha\beta}^{-8} \sum_{i,j} \sum_{r} h_{ij} \left[O_{ir} O_{rj}^{-1} g_{rr}^{\alpha\beta} f_{rr} + \sum_{u<r} (O_{ir} O_{uj}^{-1} + O_{iu} O_{rj}^{-1}) g_{ru}^{\alpha\beta} f_{ru} \right] \tag{15}$$

Combined with Eq. (9) this expression yields the desired derivatives. As has been shown previously (Abe et al., 1984; Braun and Gō, 1985), Eq. (9) can be efficiently evaluated if $c_{\alpha\beta} = c_{\beta\alpha}$. Since $g_{ru}^{\alpha\beta} = g_{ru}^{\beta\alpha}$, it is apparent that this condition holds.

Implementation in the program DIANA

Using the expressions derived in the previous section, we implemented the refinement procedure in the program DIANA (Güntert et al., 1991a). The target function to be minimized in the expanded version of DIANA consists of the sum of the previously described DIANA target function (Eq. (6) in Güntert et al., 1991a) and the term of Eq. (6) in this paper.

The memory required by the modified DIANA program is about 6 megawords for a good experimental NOE distance constraint set of a protein with 60 amino-acid residues. A significant portion of the memory is used to store $\sqrt{n_\beta} O_{r\alpha}$ and $\sqrt{n_\alpha} O_{\beta u}$. If memory becomes a limiting factor, these values can be recomputed and are not stored, but in this CPU-intensive scheme the calculation is approximately three times slower than for the memory-intensive version.

In order to make the gradient calculation practical with respect to CPU time, a filter technique

is used that avoids the computation of negligibly small contributions to the target function term of Eq. (6) and its gradient. The details of this filter technique and of the treatment of diastereotopic substituents are given in the following sections.

Efficient gradient calculation using a filter technique

As Yip and Case (1989) pointed out, the analytical expression for the derivatives is computationally demanding, as the time needed to evaluate the derivatives is determined by the number of expressions computed in the summations of Eq. (15):

$$n_r n_p n_i + \frac{n_r(n_r - 1)}{2} n_p n_i \tag{16}$$

where

$$\begin{aligned} n_r &= \text{dimension of the relaxation matrix} \\ n_p &= \text{number of spin pairs in the relaxation matrix} \\ n_i &= \text{number of intensities in the observed data set} \end{aligned} \tag{17}$$

Fortunately, the computation time can be reduced based on the sparsity of nonvanishing terms in the summations over both the dimensions in the matrix, r and u, and the number of spin pairs, α and β. This is achieved by continuous monitoring of the magnitudes of $g_{ru}^{\alpha\beta}$ in Eq. (14) and the terms

$$O_{ir} O_{rj}^{-1} f_{rr} \tag{18}$$

$$(O_{ir} O_{uj}^{-1} + O_{iu} O_{rj}^{-1}) f_{ru} \tag{19}$$

from Eq. (15). Only the values are retained that contribute significantly to the final derivatives in the summation. In practice, this is achieved by defining a cutoff, δ_1, so that only absolute values of the expressions in Eqs. (18) and (19) greater than δ_1 are included in the summation of Eq. (15). An analogous cutoff, δ_2, is defined for $g_{ru}^{\alpha\beta}$ of Eq. (14). With this filter technique, computation of the analytical derivatives is more efficient than a numerical calculation using a difference approximation, since the numerical computation is dominated by the diagonalization of the relaxation matrix R, and two diagonalizations are required for each dihedral angle. To further document the efficiency of this filter technique, the Results section will present CPU times measured in actual structure refinements (Table 1).

Treatment of pairs of diastereotopic substituents without individual assignments

In the program DIANA, chemical-shift-equivalent spins can be treated as pseudoatoms (Wüthrich et al., 1983), and the different representations of the spins to be used in Eqs. (1–3) for different types of protons are specified in the input list of NOE intensities. For example, the three protons in a methyl group are treated as one pseudoatom positioned at the geometric center of the protons. Experimental NOE intensities that do not contribute to the dihedral angle derivatives

TABLE 1
INFLUENCE OF THE FILTER PARAMETERS δ_1 AND δ_2 ON THE CPU TIME AND ON THE ACCURACY OF THE CALCULATION OF THE DERIVATIVES WITH RESPECT TO DIHEDRAL ANGLES IN EQ. (15)

	CPU time (s)[a]			Deviation[b]	
$S_{10} \rightarrow S_0$[c]	$\delta_2 = 10^{-6}$	$\delta_2 = 10^{-4}$	$\delta_2 = 10^{-2}$	RRMSD	max (rad^{-1})
$\delta_1 = 10^{-5}$	15.1	14.6	12.4	0.000	0.005
$\delta_1 = 10^{-4}$	8.1	7.8	6.8	0.006	0.055
$\delta_1 = 10^{-3}$	2.9	2.8	2.6	0.049	0.558
Inhibitor K[d]	$\delta_2 = 10^{-5}$	$\delta_2 = 10^{-3}$	$\delta_2 = 10^{-1}$		
$\delta_1 = 10^{-5}$	1583	1472	699	0.009	4.300
$\delta_1 = 10^{-4}$	388	332	168	0.045	12.49
$\delta_1 = 10^{-3}$	37	36	28	0.297	104.0

[a] All CPU times are for the calculation of the gradient of the target function, which is the most time-consuming step in each iteration. They were measured on a Cray X/MP using one processor.

[b] The relative root-mean-square deviation (RRMSD) is defined as

$$\frac{\sqrt{\sum_a |g_a^f - g_a^e|^2}}{\sqrt{\sum_a |g_a^e|^2}} \cdot g_a^f$$

and is the derivative of the target function with respect to ϕ_a obtained using the filter values indicated. g_a^e is the exact value obtained without filtering for the example $S_{10} \rightarrow S_0$, and an estimate in the case of the inhibitor K, respectively. The maximal deviation, max, is defined as $\max_a |g_a^f - g_a^e|$, and is given in units of rad^{-1}. The numbers listed in the Table are for the combinations of the δ_1-values with $\delta_2 = 10^{-2}$ for the example $S_{10} \rightarrow S_0$, and with $\delta_2 = 10^{-3}$ for the inhibitor K, respectively (For the values presented in Table 1, the effect of δ_1 on RRMSD and max was much larger than the effect of δ_2).

[c] Refinement of one of the conformers in the group S_{10} vs. the simulated data set for S_0 (see text). The error assessments RRMSD and max are relative to the derivatives calculated without any filtering, which took 40 s CPU time.

[d] The error assessments are relative to the derivatives calculated with $\delta_1 = 10^{-5}$ and $\delta_2 = 10^{-5}$.

(e.g., cross peaks between geminal protons) are automatically eliminated. Unless specifically requested otherwise, the substituents of prochiral centers (β-methylene protons and the methyl groups of leucine and valine) and the ring protons of tyrosine and phenylalanine are treated individually. When separate NOE intensities can be measured for two diastereotopic substituents that have not been stereospecifically assigned, the target function is calculated twice for the two possible individual assignments, and the assignment that yields the smaller target function is used. If the polypeptide studied includes multiple pairs of diastereotopic ligands without stereospecific assignments, each permutation of two individual assignments is treated as being independent of the individual assignments for the substituents in all other prochiral centers. Experience showed that it is sufficient to perform this permutation process once in every ten minimization steps.

RESULTS

This section reports on initial tests of the modified version of the DIANA program that includes the term of Eq. (6). For these tests we used either a helical polypeptide with a simulated

data set of NOE intensities, or a globular protein with 57 amino-acid residues, the protease inhibitor K, with an experimental data set.

Simulated refinement of a 15-residue α-helix

The polypeptide segment used here was

-Glu-Leu-Ala-His-Met-Ala-Asn-Gln-Ala-Ala-Glu-Ala-Ile-Leu-Lys- (20)

which corresponds to an α-helix in the designed α-protein FELIX (Hecht et al., 1990). The observed data set of NOE intensities, I^{o}_{ij}, was generated by calculating these intensities for the rigid geometry of a conformation S_0 of the polypeptide (20), where the polypeptide backbone forms a standard α-helix, and the side chains adopt any one of the sterically allowed spatial arrangements. This molecular geometry was obtained by minimizing the target function in the standard DIANA program (Güntert et al., 1991a) with distance constraints enforcing the $O_i \cdots HN_{i+4}$ hydrogen bonds, van der Waals constraints, and dihedral angle constraints for a regular α-helical polypeptide backbone, i.e., $\varphi = -57 \pm 1°$ and $\psi = -47 \pm 1°$. For the amino-acid side chains, only the van der Waals constraints were imposed. In addition, three groups of ten molecular geometries each, S_{10}, S_{20} and S_{30}, were generated, starting from randomized conformers of the polypeptide (20) and minimizing the standard DIANA target function against data sets consisting of the van der Waals constraints and backbone dihedral angles constraints, where φ_i and ψ_i were allowed to vary individually over ranges of $\pm 10°$, $\pm 20°$ and $\pm 30°$, respectively, about the standard α-helix values used for S_0.

To explore parameter sets that would allow performing the relaxation matrix refinements with reasonable use of computing time, we first examined the effects of the aforementioned filtering in the calculation of Eq. (15) on the derivatives and the computation time. Values of δ_1 and δ_2 ranging from 10^{-6} to 10^{-2} were used (Table 1). The derivatives were calculated in the refinement of one of the S_{10} conformers against the S_0 data set. The NOESY mixing time was taken to be 40 ms, and only dipolar spin-spin interactions with distances shorter than 5.0 Å were considered. While the errors in the derivatives increased significantly for increasing values of δ_1, it was found that the various values of δ_2 had little effect on the errors (not shown in Table 1), although the computation time is reduced for larger values of δ_2. In the following studies with the helical polypeptide (20), we used values of $\delta_1 = 10^{-4}$ and $\delta_2 = 10^{-4}$. As an alternative, which would be reminiscent of the general strategy of the variable target function approach used by DIANA (Güntert et al., 1991a), one could start the refinement with relatively large values of δ_1 and δ_2, and then gradually decrease these values during the minimization. This should further improve the efficiency of the computation.

Using the aforementioned filters, each of the 30 structures in the groups S_{10}, S_{20} and S_{30} was refined against a data set consisting of all NOE intensities calculated for S_0 and all van der Waals constraints for the polypeptide (20), using the conjugate-gradient minimizer in DIANA. Twenty-nine of the structures were brought directly to the S_0 structure. One of the S_{30} conformers was initially trapped in a local minimum. By altering the balance of the contributions of the van der Waals terms and the NOE intensity terms to the target function, the conformation went to the S_0 structure, too. Table 2 shows a summary of this simulation. In the refined data, the small devia-

TABLE 2
RELAXATION MATRIX REFINEMENT OF THE POLYPEPTIDE (20) AGAINST THE DATA SET S_0 FOR A STANDARD α-HELIX, STARTING WITH TEN DISTORTED α-HELICES FROM EACH OF THE THREE GROUPS S_{10}, S_{20} AND S_{30}[a]

Refinement[b]	RMSD (Å)[c]		R-factor (%)[d]
	Backbone	All heavy atoms	
S_{10}	0.52 (0.33...0.69)	0.69 (0.45...0.85)	10.1 (7.19...12.3)
$S_{10} \rightarrow S_0$	0.00 (0.00...0.01)	0.02 (0.01...0.03)	0.09 (0.07...0.14)
S_{20}	0.94 (0.46...2.02)	1.34 (0.69...3.34)	19.1 (15.8...24.1)
$S_{20} \rightarrow S_0$	0.01 (0.00...0.01)	0.05 (0.04...0.07)	0.20 (0.10...0.38)
S_{30}	1.08 (0.67...2.38)	1.65 (0.96...3.90)	26.2 (18.6...32.1)
$S_{30} \rightarrow S_0$	0.01 (0.00...0.02)	0.08 (0.05...0.11)	0.28 (0.44...0.50)
DG refined	0.45 (0.45...0.45)	0.61 (0.60...0.61)	13.8 (13.8...13.8)

[a] See text for a precise definition of S_0, S_{10}, S_{20} and S_{30}.

[b] S_{10} are the starting structures in the group S_{10}, $S_{10} \rightarrow S_0$ are the same structures after refinement against the data set S_0. *DG refined* indicates that all 30 conformers of the groups S_{10}, S_{20} and S_{30} were refined against a set of distance constraints derived from the NOE intensity data set S_0 without explicit allowance for spin diffusion.

[c] All RMSD values are averages of the pairwise RMSDs between the individual conformers in each group relative to the structure S_0. The pairs of conformers were superimposed either for minimal RMSD calculated for the backbone atoms N, C^α and C′, or for all heavy atoms.

[d] $R = \frac{\sum_{i,j}|(I^o_{ij} - wI^c_{ij})|}{\sum_{i,j}I^o_{ij}} \times 100$

tions from the S_0 structure are due to the limited number of minimization steps performed and not to the lack of complete convergence.

In order to more closely approximate the situation of an experimental set of NOE intensities, we added random variations to the exact data set S_0. Thereby each of the NOE intensities in S_0, I^o_{ij}, was modified as

$$I^p_{ij} = I^o_{ij}(1 + v_{ij}) \tag{21}$$

where v_{ij} are random, uniformly distributed variables in the interval $[-a,a]$. All calculations started from a conformer in the group S_{30}, and a was incremented from 0.025 to 0.25 in steps of 0.025. As expected, the initial and final R-factors increased with increasing random errors in the NOE intensity data set, i.e. from 30.0 to 40.9%, and from 2.1 to 19.7%, respectively. Likewise, the RMSDs between the refined structures and S_0 increased with increasing random errors in the intensities (from 0.04 to 0.32 Å for the backbone atoms, and from 0.11 to 0.52 Å for all heavy atoms), but the deviations were smaller than what one would obtain by generating structures randomly and selecting structures with similar R-factors. For example, the final R-factor in the refinement with $a = 0.25$ was around 20%, which is in the range of the unrefined S_{20} conformers (Table 2), but the corresponding RMSD values of 0.32 Å for the backbone atoms and 0.52 Å for all heavy atoms are significantly smaller than the range of RMSD values for the conformers of the set S_{20} (Table 2).

To assess the benefits of using the complete relaxation matrix refinement relative to a much faster distance geometry refinement that does not explicitely allow for spin diffusion, we computed a set of NOE distance constraints from the simulated NOE intensities for the helical structure S_0 of the polypeptide (20). Exact distance constraints, i.e., equal upper and lower bounds, were derived assuming the relationship

$$I = \sum_{j=1}^{m} \frac{\alpha_j}{d^j}, \quad \alpha_j \geq 0 \tag{22}$$

between the NOE intensity, I, and the corresponding $^1H - {}^1H$ distance, d, of the helical structure S_0. Separate fits for the coefficients α_j were made for NOE intensities between single protons, and NOEs involving methyl groups, respectively (Fig. 1). For given NOE intensities, the corresponding distances were then calculated using these calibration curves by solving Eq. (22) for the distance d with a Newton method. Using this distance constraint set for S_0, all 30 structures from the three groups S_{10}, S_{20} and S_{30} were minimized with the program DIANA. Weighting factors of 0.1 were used for the NOE distance constraints, and during the calculation the weight for the van der

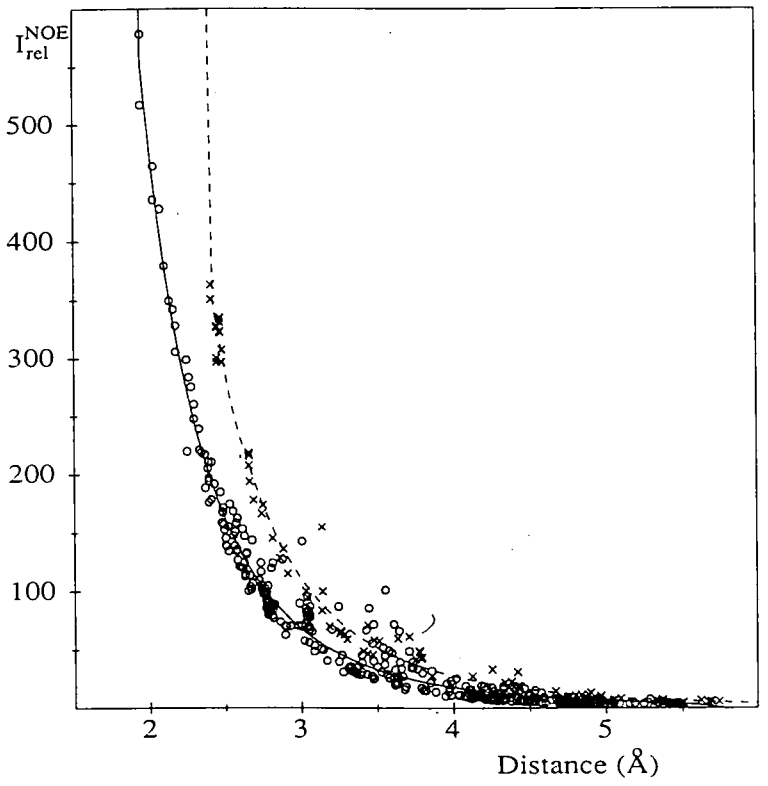

Fig. 1. Plot of the relative NOESY intensities versus the corresponding ^1H-^1H distances for the standard α-helix structure S_0 of the polypeptide (20) (see text). Circles refer to pairs of single protons, crosses to pairs that include at least one methyl group. The solid line and the dashed line are the best-fit curves for the NOE intensity vs. ^1H-^1H distance calibration for these two situations (see text).

Waals constraints was increased from 1.0 to 2.0 to 5.0. All 30 structures converged to the same final conformation, with pairwise RMSDs between different solutions of 0.01 Å for the backbone and 0.07 Å for all heavy atoms. However, this refined structure differed from the target structure S_0 by RMSD values of 0.45 Å for the backbone atoms, and 0.61 Å for all heavy atoms.

Trial relaxation matrix refinements of experimental NMR structures of the inhibitor K

The studies with this globular protein (M = 6000) were added to obtain an initial impression of the performance of the modified DIANA program in a 'real-life situation'. In the data set for the inhibitor K (K. Berndt, P. Güntert and K. Wüthrich, unpublished work), there are 922 measured NOE intensities that contributed to Eq. (6), and a total of 376 proton spins. The NOESY spectra were recorded with a mixing time, τ_m, of 40 ms. By neglecting dipolar interactions for distances longer than 5.0 Å, we found that values of $\delta_1 = 5 \times 10^{-4}$ and $\delta_2 = 10^{-2}$ provided a good balance between a reasonable amount of computation time and accuracy of the calculated derivatives. Ten conformers of the inhibitor K obtained from standard distance geometry calculations with the program DIANA (Güntert et al., 1991a) were selected as starting conformations for a complete relaxation matrix refinement with these parameters, and each conformer was minimized using the conjugate gradient algorithm in DIANA for 250 iterations. The program performed these calculations with a CPU time of about 50 s per iteration, and on average the R-factor was reduced by 22%.

DISCUSSION

Empirical criteria, such as the comparison of corresponding NMR structures in solution and X-ray crystal structures (e.g., Kline et al., 1988; Billeter et al., 1989) have clearly demonstrated that good-quality NMR structures of proteins can be obtained based on the initial slope approximation (Gordon and Wüthrich, 1978; Wagner and Wüthrich, 1979; Anil Kumar et al., 1981) for the interpretation of NOESY data. Nonetheless, as mentioned in the Introduction, in this approach both spin diffusion and internal mobility of the molecular structure can, in principle, contribute to systematic deviations of the NMR structure from the actual molecular conformation. This was also confirmed by the tests performed in this paper with the polypeptide (20) using distance geometry calculations (Table 2). The extension of the software package DIANA by term (6) should be useful in assessing the aforementioned systematic deviations, and has the potential of producing structures that are devoid of such shortcomings. With regard to practical applications, it is of particular interest that with the extended version of the program DIANA, both complete relaxation matrix refinement and distance geometry structure calculations can be performed within a single program.

The basis of the program DIANA is the efficient minimization of all terms of the target function in dihedral angle space. Compared to minimization in Cartesian coordinate space, the number of degrees of freedom is thus significantly reduced. For example, in the 15-residue polypeptide chain (20) there are 69 and 768 degrees of freedom in dihedral angle space and in Cartesian space, respectively, and in the inhibitor K the corresponding numbers are 252 and 3315. Although the complete evaluation of the derivatives is nonetheless computationally demanding, the presently introduced filter technique contributes greatly to making the complete relaxation matrix refine-

ment method a practical procedure. Its functionality has been demonstrated with the examples listed in Table 1. In particular, the simulated refinements of the helical polypeptide (20) illustrate that the technique has a large convergence radius. Nonetheless, a complete relaxation matrix refinement is about 100 times more CPU-intensive than a standard distance geometry calculation, and it remains to be seen whether the improvements of the structure determination warrant this extra expense. The following are some indications of ways in which a complete relaxation matrix refinement with the extended version of the program DIANA could profitably be applied in conjunction with other computational techniques.

Previously, it has been proposed (Keepers and James, 1984; Boelens et al., 1989) that relaxation matrix refinement might be used to improve the NOE distance constraints for subsequent distance geometry or restrained molecular dynamics calculations. On general grounds, we would suggest rather to use distance geometry calculations with conservative upper limit distance constraints for the initial structural interpretation of the NMR data (Kline et al., 1988; Güntert et al., 1991b), and then subject a set of distance geometry structures to a complete relaxation matrix refinement. In the aforementioned trial refinements of a group of distance geometry structures of the inhibitor K, we observed that the intraresidue NOE intensities were to a large extent satisfied at the expense of the interresidue NOEs. The reason for this behaviour is the heavy weight with which high NOE intensities are treated in the term T of Eq. (6). These observations suggest that a combined input of conservatively calibrated distance constraints *and* the experimental NOE intensities might be advantageous, since the distance constraints would effectively prevent falsification of weak NOE intensities by the relaxation matrix treatment. This combined input strategy could be further refined, for example, by eliminating in the minimization with the term (6) those NOE intensities which arise from flexible parts of the protein molecule (Dyson et al., 1988; Kessler et al., 1988; Torda et al., 1990), as would be indicated by additional measurements of T_1, T_2 or scalar coupling constants.

In the present implementation of the complete relaxation matrix refinement in the extended DIANA program, we greatly simplified the problem to be solved by assuming that all NOE intensities might be fitted to a single rigid structure, and that one isotropic rotational tumbling time can be used for all protons. Clearly, the computational demands will be further increased when more realistic models allowing for anisotropic overall rotation and additional intramolecular motions are used. Although it remains to be seen whether such calculations are warranted in the day-to-day determination of NMR structures of proteins, they should be applied to properly evaluate the effects of using less demanding, approximate treatments. The presently described extended version of the DIANA program should be a useful tool in such endeavours.

ACKNOWLEDGEMENTS

This work was supported by Cray Research, Inc., the Interdisziplinäres Projektzentrum für Supercomputing (IPS) of the ETH Zürich, and the Schweizerischer Nationalfonds (project 31.25174.88). J.E. Mertz thanks Dr. Hans-Peter Lüthi for the hospitality during his sabbatical time at the IPS. The use of the Cray X-MP/28 of the ETH Zürich is gratefully acknowledged. We thank Mr. R. Marani for the careful processing of the manuscript.

REFERENCES

Abe, H., Braun, W., Noguti, T. and Gō, N. (1984) *Comput. Chem.*, **8**, 239-247.

Anil Kumar, Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Am. Chem. Soc.*, **103**, 3654-3658.

Baleja, J.D., Moult, J. and Sykes, B.D. (1990) *J. Magn. Res.*, **87**, 375-384.

Bax, A. (1989) *Annu. Rev. Biochem.*, **58**, 223-256.

Billeter, M., Kline, A.D., Braun, W., Huber, R. and Wüthrich, K. (1989) *J. Mol. Biol.*, **207**, 677-687.

Billeter, M., Schaumann, T., Braun, W. and Wüthrich, K. (1990) *Biopolymers*, **29**, 695-706.

Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1989) *J. Magn. Res.*, **82**, 290-308.

Borgias, B.A. and James, T.L. (1988) *J. Magn. Res.*, **79**, 493-512.

Borgias, B.A., Gochin, M., Kerwood, D.J. and James, T.L. (1990) *Progr. Nucl. Magn. Res. Spectrosc.*, **22**, 83-100.

Braun, W. (1987) *Q. Rev. Biophys.*, **19**, 115-157.

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

Braun, W., Bösch, C., Brown, L.R., Gō, N. and Wüthrich, K. (1981) *Biochim. Biophys. Acta*, **667**, 377-396.

Braun, W., Wider, G., Lee, K.H. and Wüthrich, K. (1983) *J. Mol. Biol.*, **169**, 921-948.

Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3801-3805.

Clore, G.M. and Gronenborn, A.M. (1989a) *Crit. Rev. Biochem. Mol. Biol.*, **24**, 479-564.

Clore, G.M. and Gronenborn, A.M. (1989b) *J. Magn. Res.*, **84**, 398-409.

Dobson, C.M., Olejniczak, E.T., Poulson, F.M. and Ratcliffe, R.G. (1982) *J. Magn. Res.*, **48**, 97-110.

Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. and Wright, P.E. (1988) *J. Mol. Biol.*, **201**, 161-200.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford.

Gordon, S.L. and Wüthrich, K. (1978) *J. Am. Chem. Soc.*, **101**, 5152-5153.

Güntert, P., Braun, W. and Wüthrich, K. (1991a) *J. Mol. Biol.*, **217**, 517-530.

Güntert, P., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1991b) *J. Mol. Biol.*, **217**, 531-540.

Havel, T.F. and Wüthrich, K. (1984) *Bull. Math. Biol.*, **46**, 673-698.

Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.*, **182**, 281-294.

Hecht, M.H., Richardson, J.S., Richardson, D.C. and Ogden, R.C. (1990) *Science*, **249**, 884-891.

Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and van Gunsteren, W.F. (1985) *J. Mol. Biol.*, **182**, 179-182.

Kaptein, R., Boelens, R., Scheek, R.M. and van Gunsteren, W.F. (1988) *Biochemistry*, **27**, 5389-5395.

Keepers, J.W. and James, T.L. (1984) *J. Magn. Res.*, **57**, 404-426.

Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W.F. and Berendsen, H.J.C. (1988) *J. Am. Chem. Soc.*, **110**, 3393-3396.

Kline, A.D., Braun, W. and Wüthrich, K. (1988) *J. Mol. Biol.*, **204**, 675-724.

Koehl, P. and Lefèvre, J.F. (1990) *J. Magn. Res.*, **86**, 565-583.

Kuntz, I.D., Thomason, J.F. and Oshiro, C.M. (1989) *Methods Enzymol.*, **177**, 159-203.

Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95-117.

Madrid, M., Mace, J.E. and Jardetzky, O. (1989) *J. Magn. Res.*, **83**, 267-278.

Olejniczak, E.T., Gampe, R. and Fesik, S. (1986) *J. Magn. Res.*, **67**, 28-41.

Post, C.B., Meadows, R.P. and Gorenstein, D.G. (1990) *J. Am. Chem. Soc.*, **112**, 6796-6803.

Schaumann, Th., Braun, W. and Wüthrich, K. (1990) *Biopolymers*, **29**, 679-694.

Solomon, I. (1955) *Phys. Rev.*, **99**, 559-565.

Torda, A.W., Scheek, R.M. and van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223-235.

Tropp, J. (1980) *J. Chem. Phys.*, **72**, 6035-6043.

Wagner, G. and Wüthrich, K. (1979) *J. Magn. Res.*, **33**, 675-680.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

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

Yip, P. and Case, D.A. (1989) *J. Magn. Res.*, **83**, 643-648.