Volume 268, number 1, 141-145 FEBS 08656 July 1990

^{1}H NMR assignments of sidechain conformations in proteins using a high-dimensional potential in the simulated annealing calculations

J. Habazettl, C. Cieslar, H. Oschkinat and T.A. Holak

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG

Received 23 April 1990

A high-dimensional potential representing distance constraints for stereospecifically assignable diastereotopic proton or methyl pairs was incorporated into the dynamical simulated annealing protocol to calculate structures with stereospecifically determined sidechain conformations. The protocol is tested on nuclear magnetic resonance cross-relaxation data of a trypsin inhibitor from squash seeds, CMTI-I, and compared with two other methods of stereospecific assignment, the floating chirality and coupling constant methods. There is good agreement between the three methods in predicting the same stereospecific assignments. Because the high-dimensional potential uses more relaxed absolute distance constraints and also takes into account the relative distance constraint patterns, it avoids possible overinterpretation of the NOE data.

Nuclear magnetic resonance; Simulated annealing; Stereospecific assignment of prochiral centers; Protein; Structure determination

1. INTRODUCTION

Recent nuclear magnetic resonance (NMR) structure determinations of proteins have shown that significant improvements in the definition of amino acid side chains, and to a lesser extent of the protein backbone, can be obtained by stereospecific assignments of prochiral groups of protons [1-5]. Without stereospecific assignments, interproton distance constraints to prochiral groups of protons are usually incorporated into the calculations by means of a pseudoatom [6]. As a result, the information content of the data is significantly reduced because the distance constraints have to be weakened by a comparatively large correction term (about 1 Å for methylene groups. Under suitable circumstances, stereospecific assignments of β-methylene protons can be obtained from the analysis of intraresidue nuclear Overhauser effects (NOEs) and coupling constant data [7,8]. This method was improved by matching the experimental intraresidue NOEs, sequential interresidue NOEs, and $^3J_{NH\alpha}$ and $^3J_{\alpha\beta}$ coupling constants with those calculated for conformations present in a database [3]. The method is usually only applicable to unambiguous situations and often fails in large proteins when the coupling constants cannot be determined because of overlaps in the spectrum. An alternative strategy is to obtain the stereospecific assignments automatically during the structure calculation itself [1,2,9-11]. This is accomplished by making arbitrary assignments of prochiral protons and allowing the protons to exchange places during the calculation such that the conformation with the lower energy is chosen. This conformation fits all the NOE data. The method, called the floating stereospecific assignment, works well in cases where sufficient NOEs are available in terms of quantity and precision. The floating stereospecific method requires relatively tight boundaries for the distance constraints in order for the floating protons to flip yet remain in a particular conformation. Application of such tight constraints could produce structures for which not all possible conformations of a sidechain are sampled. The procedure described in this paper overcomes this limitation. All stereospecifically assignable NOEs belonging to the diastereotopic proton or methyl pairs are represented by a new n-dimensional potential, $F_{highdim}$. Two types of bounds are used in this potential. Restrictive bounds are used for the relative distance constraints within each NOE proton pair in the prochiral group. The absolute boundaries of the distance constraints, however, are more relaxed than those in the floating stereospecific assignments. Thus the high-dimensional potential should represent the experimental NOEs more realistically. In this study we apply the high-dimensional potential in the simulated annealing (SA) protocol to calculate structures of a small trypsin inhibitor from squash seeds, CMTI-I. We recently determined the three-dimensional structures of this inhibitor by the floating chirality method [10].

Correspondence address: T.A. Holak, Max-Planck-Institut für Biochemie, D-8033, Martinsried bei München, FRG

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

2. EXPERIMENTAL

The basic protocol used for the calculations has been presented previously [2,10,12] and consists of 5 stages. In stage 1, the coordinates of the substructures are calculated with the distance geometry program DISGEO [13]. In the second stage all atoms missing in the substructures are added [12]. Steps 3 and 4 consist of simulated annealing: raising the temperature of the system followed by slow cooling to overcome local minima and locate the region of the global minimum of the target function.

The NOE distance constraints are represented by the target function F_{NOE} which consists of two different potential functions:

$$F_{NOE} = \sum_{NOEs} F_{square} + \sum_{diastereotopic} F_{highdim}$$

F_{square} is a square well potential with variable force constant kNOE and was chosen for all uniquely assigned NOE constraints. Initially k_{NOE} was set to 0.01 kcal·mol^{-1}·Å^{-2} and increased to 50 kcal·mol^{-1}·Å^{-2} during the simulated annealing stage. All the n NOEs belonging to one diastereotopic proton or methyl pair are represented by a new n-dimensional potential $F_{highdim}$. This potential is the product of two n-dimensional potentials, each corresponding to one or two possible stereospecific assignments:

$$F_{highdim} = K_h \cdot \left(\sum_{i=1}^{dim/2} F_i^{ass1} \right) \cdot \left(\sum_{i=1}^{dim/2} F_j^{ass2} \right)$$

with

$$F_i^{ass1}(r_{xy1}^i, r_{xy2}^i) = C_i(\text{noe}_{i1}^x(r_{xy1}^i, r_{xy2}^i) + \text{noe}_{i1}^y(r_{xy1}^i, r_{xy2}^i) + \Delta\text{noe}_{i1}(r_{xy1}^i, r_{xy2}^i))$$

$$F_i^{ass2}(r_{xy1}^i, r_{xy2}^i) = F_i^{ass1}(r_{xy2}^i, r_{xy1}^i)$$

and r_{xy1}^i, r_{xy2}^i are the distances from atom x^i to the diastereotopic atoms $y1$ and $y2$, respectively. With noe_{i1}^x and noe_{i1}^y, the constraints between proton x^i and the diastereotopic pair $y1$ and $y2$ are introduced into the potential function using d_{L1}, d_{L2} and d_{ul}, d_{u2} as the lower and upper bounds, respectively. The term Δnoe_{il} defines the relative distance constraint of this NOE pair, constraining the difference of the two distances within a specific interval (d_{du}–d_{dL}):

$$\text{noe}_{i1}^x(r_{xy1}^i, r_{xy2}^i) = \begin{cases} 0 & \text{for } d_{L1} \le r_{xy1}^i \le d_{u1} \\ (d_{L1} - r_{xy1}^i)^2 & \text{for } r_{xy1}^i < d_{L1} \text{ and } d_{L1} - r_{xy1}^i \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim}(d_{L1} - r_{xy1}^i) & \text{for } r_{xy1}^i < d_{L1} \text{ and } d_{L1} - r_{xy1}^i > \text{lim} \\ (r_{xy1}^i - d_{u1})^2 & \text{for } r_{xy1}^i > d_{u1} \text{ and } r_{xy1}^i - d_{u1} \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim}(r_{xy1}^i - d_{u1}) & \text{for } r_{xy1}^i > d_{u1} \text{ and } r_{xy1}^i - d_{u1} > \text{lim} \end{cases}$$

$$\text{noe}_{i1}^y(r_{xy1}^i, r_{xy2}^i) = \begin{cases} 0 & \text{for } d_{L2} \le r_{xy2}^i \le d_{u2} \\ (d_{L2} - r_{xy2}^i)^2 & \text{for } r_{xy2}^i < d_{L2} \text{ and } d_{L2} - r_{xy2}^i \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim}(d_{L2} - r_{xy2}^i) & \text{for } r_{xy2}^i < d_{L2} \text{ and } d_{L2} - r_{xy2}^i > \text{lim} \\ (r_{xy2}^i - d_{u2})^2 & \text{for } r_{xy2}^i > d_{u2} \text{ and } r_{xy2}^i - d_{u2} \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim}(r_{xy2}^i - d_{u2}) & \text{for } r_{xy2}^i > d_{u2} \text{ and } r_{xy2}^i - d_{u2} > \text{lim} \end{cases}$$

The intensities of the NOE cross peaks were determined from volume integrals and were converted into the distance constraints as described previously [10]. For interproton distances less than 3.0 Å the absolute distance range ($d_{ul,2}$–$d_{L1,2}$) was set to 0.9 Å (the bounds to the distance constraints obtained from calibrations

$$\Delta\text{noe}_{i1}(r_{xy1}^i, r_{xy2}^i) = \begin{cases} 0 & \text{for } r_{xy1}^i + d_{dL} \le r_{xy2}^i \le r_{xy1}^i + d_{du} \\ D^2 & \text{for } r_{xy2}^i < r_{xy1}^i + d_{dL} \text{ and } D \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim} \cdot D & \text{for } r_{xy2}^i < r_{xy1}^i + d_{dL} \text{ and } D > \text{lim} \\ G^2 & \text{for } r_{xy2}^i > r_{xy1}^i + d_{du} \text{ and } G \le \text{lim} \\ \text{lim}^2 + 2\,\text{lim} \cdot G & \text{for } r_{xy2}^i > r_{xy1}^i + d_{du} \text{ and } G > \text{lim} \end{cases}$$

with

$$D = \frac{r_{xy1}^i - r_{xy2}^i + d_{dL}}{\sqrt{2}}$$

$$G = \frac{r_{xy2}^i - r_{xy1}^i - d_{du}}{\sqrt{2}}$$

were set to b–0.4 Å/+0.5 Å), the relative distance range (d_{du}–d_{dL}) to 0.4 Å; for proton distances from 3.1 Å to 3.6 Å the absolute and relative distance ranges were set to 1.5 Å (b–0.5 Å/1.0Å) and 0.6 Å, respectively; and for distances from 3.7 Å to 4.5 Å they were equal to 2.5 Å (b–1.0 Å/+1.5 Å) and 0.8 Å. For diastereotopic NOE pairs with two different distance constraints the bigger relative distance range was chosen. The absolute upper bound of methyl protons was further increased by 1 Å and the relative distance range by 0.5 Å. If there were only one NOE to a diastereotopic proton pair the upper bound for the missing NOE was set to the maximal possible value [2] and no relative distance restriction was made.

The full potential for one NOE pair is shown in Fig. 1. It is symmetric and has two minima (A and B) representing the two equally probable assignments of diastereotopic protons. For example, for two NOEs from the NH amide to protons β_1 and β_2 of a methylene group the coordinates for β_1, β_2 in A are $r_{NH\text{-}\beta1} < r_{NH\text{-}\beta2}$ (x = NH, $y_1 = \beta_1$, $y_2 = \beta_2$, Fig. 1); and for B $r_{NH\text{-}\beta1} < r_{NH\text{-}\beta2}$. The potential barrier between the minima can be set with the factor c_1 and gives the probability of the system for changing the assignment. For every high-dimensional potential c_i was chosen to set the barrier height to k_h·0.01 kcal·mol^{-1}·Å^{-2}. The force constant k_h was increased from 0.001 to 5 kcal·mol^{-1}·Å^{-2}. Another factor acts on the area outside the minima and the potential barrier to set the potential and the gradient

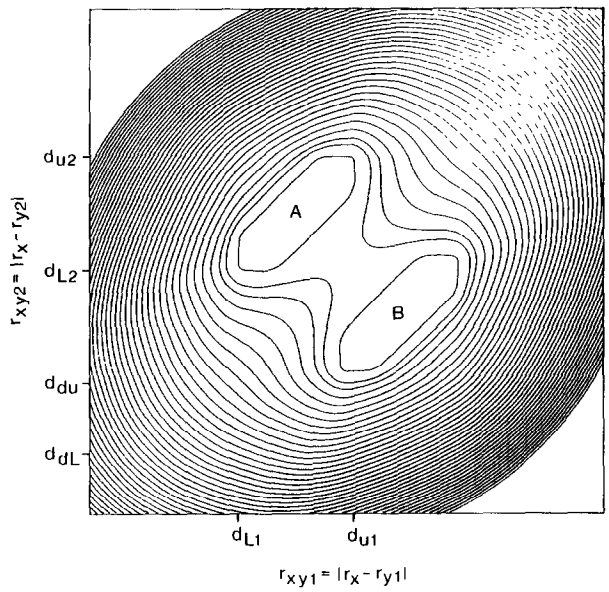

Fig. 1. Logarithmic plot of the high-dimensional potential for two NOEs with A and B the two minima. r_{xyl} is the distance from atom x to diastereotopic atom $y1$, r_{xy2} to $y2$, respectively.

comparable to the square well potential F_{NOE}. With the logarithmic representation of the potential in Fig. 1 it can be seen that this two-dimensional potential is also a quadratic function of the NOE violations. The gradient of $F_{highdim}$ for one variable is given by

$$\frac{\partial F_{highdim}}{\partial r^i_{xyl}} = k_h C_l \left\{ \left(\sum_{j=1}^{dim/2} F_j^{ass2} \right) \cdot \frac{\partial (noe^x_{i1} + noe^y_{i1} + \Delta noe_{i1})}{\partial r^i_{xyl}} + \left(\sum_{j=1}^{dim/2} F_j^{ass1} \right) \cdot \frac{\partial (noe^x_{i2} + noe^y_{i2} + \Delta noe_{i2})}{\partial r^i_{xyl}} \right.$$

with

$$noe^x_{i2}(r^i_{xy1}, r^i_{xy2}) = noe^x_{i1}(r^i_{xy2}, r^i_{xy1})$$
$$noe^y_{i2}(r^i_{xy1}, r^i_{xy2}) = noe^y_{i1}(r^i_{xy2}, r^i_{xy1})$$
$$\Delta noe_{i2}(r^i_{xy1}, r^i_{xy2}) = \Delta noe_{i1}(r^i_{xy2}, r^i_{xy1})$$

Step one of simulated annealing was carried out with 50 cycles of 75 fs dynamics at 1000 K. The temperature was coupled to the bath with the Berendsen method [14]. Step 2 consisted of 1.5 ps dynamics at 300 K. In stage 5 the structures are improved by 400 cycles of energy minimization.

3. RESULTS AND DISCUSSION

The high-dimensional potential function works by forcing a sidechain to rotate into the conformation that fits the pattern of all NOEs to the given proton. If there are n NOEs to the β_1 and β_2 protons, the potential is *n*-dimensional and has two energy minima. The two possible minima are a priori equivalent. Since the assignment of the β_1 and β_2 protons is fixed in the starting structures, the energies of the two minima are not equal, the assignment with the lowest penalty function is preferred. In the *n*-dimensional target function, all *n* NOEs are correlated and only one pathway of relative patterns of NOEs will be selected provided that there are enough NOEs to determine a sidechain conformation. Once a single or restricted conformation of the sidechain is secured, the stereospecific assignment of the β_1 and β_2 protons in terms of their chemical shifts can be accomplished by comparing the distribution of NOEs for three possible rotamers of the β-methylene group using the method described by Hyberts et al. [8]; or for γ,δ protons, by comparison with the floating stereospecific assignment [10].

The pseudoatom corrections are only used for protons with completely degenerate chemical shifts in our calculations. In all other cases, even when no preferred rotamer conformation is obtained from the calculations, the pseudoatom corrections are not necessary.

Table I

Comparison of the stereospecific assignments[a]

Residue	NOEs	High-dimensional potential	Floating assignment	Coupling constant
Val-2	12	(−217 ± 58)	(10 ± 53)	180 ± 60
Cys-3	12	−152 ± 35	−183 ± 4	180 ± 60
Leu-7	6	−162 ± 20	(−89 ± 31)	–[b]
Met-8	8	−162 ± 7	(74 ± 18)	–[b]
Glu-9	6	−65 ± 13	−57 ± 11	−60 ± 60
Cys-10	10	66 ± 32	45 ± 6	60 ± 60
Lys-12	8	−44 ± 6	(51 ± 7)	–[b]
	6	168 ± 11	179 ± 6	–[b]
Asp-13	14	(145 ± 88)	−71 ± 4	−60 ± 60
Ser-14	6	(64 ± 8)[d]	(60 ± 6)[d]	60 ± 60
Cys-16	6	−49 ± 8	−72 ± 19	−60 ± 60
Cys-20	18	−69 ± 9	−101 ± 28	−60 ± 60
Val-21	10	−73 ± 11	−65 ± 3	−60 ± 60 and 60 ± 60
Cys-22	14	171 ± 6	−160 ± 3	180 ± 60
Leu-23	10	−41 ± 33	−53 ± 18	−60 ± 60
	16	−175 ± 10	173 ± 5	–[b]
Glu-24	6	(38 ± 99)	−159 ± 13	–[c]
His-25	10	−75 ± 8	−41 ± 6	−60 ± 60
Tyr-27	10	(3 ± 85)	−64 ± 5	−60 ± 60
Cys-28	10	−47 ± 12	−61 ± 10	−60 ± 60

[a] The second column gives the number of NOEs to the prochiral protons obtained from the NOESY spectra. The average χ^1 angles (deg) for the SA structures are entered in columns 3–4 together with standard deviations; for Lys-12 and Leu-23 second entry refers to χ^2. 20 SA structures were calculated for each entry in columns of 3 and 4 with the slightly up-dated distance constraints table compared to that in [10]. The ranges for χ^1 torsion angle (5th column) are estimated as described in [7]. No stereospecific conformation could be determined for the χ angles given in parentheses.

[b] No unambiguous assignment was possible.

[c] See test.

[d] The chiral assignment was obtained by fixing the χ^1 angle to the value given in column 5 c.f. [10].)

Our input data is thus more accurate than that obtained from only intraresidue NOEs and J-coupling constants information, provided the same number of prochiral centers is determined by both methods.

The comparison of three different methods of the stereospecific assignments: high-dimensional potential, floating stereospecific, and coupling constant method, is presented in Table I. Our current coupling constant data is derived from the E. COSY spectrum [15] and is thus more comprehensive than published previously [10]. The second entry in Table I gives the number of NOEs to the prochiral protons obtained from the NOESY spectra. This number was ca. 4 for residues for which no stereospecific assignments were obtained. Table I shows that there is good agreement between all three methods in predicting the same stereospecific assignments. The high-dimensional potential seems to produce the most conservative results. For example, Asp-13, Glu-24 and Tyr-27 were not assigned stereospecifically by the high-dimensional method whereas they were by the floating chirality method. Asp-13 and Tyr-27 exhibit a single conformation based on the coupling constant data. For Glu-24, the E. COSY data indicates both coupling constants, $J_{\alpha\beta 1}$ and $J_{\alpha\beta 2}$, to be equal and larger than 5 Hz suggesting multiple conformations of the sidechain. The discrepancy between the results of the floating chirality assignment and coupling constant data for Glu-24 can be explained assuming that this residue occupies a certain range of side-chain conformations. The floating chirality method produces only a subset of all possible conformations. Such an explanation is supported by the fact that within each interaction between Glu-24 NH, Glu-24 $C_\alpha H$, H-25 NH, and the Glu-24 β-protons, the NOEs are quite different for the β_1 and β_2 protons (note that the stereospecific assignment can be achieved even in cases when the amino acid side chain is not in a single conformation but covers a range of conformations).

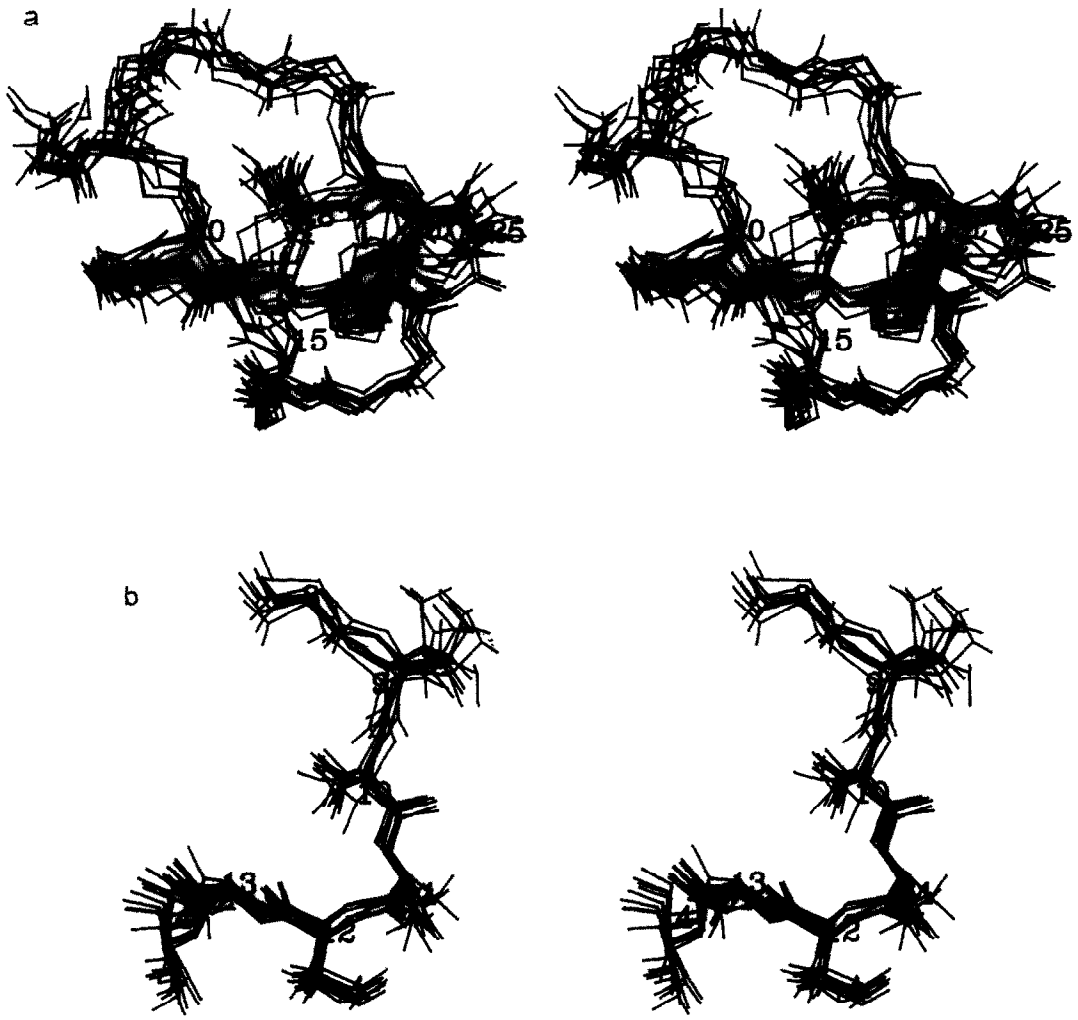

Fig. 2. (a) Stereoview of the backbone (N, C^α, C, O) atoms of the SA structures calculated with the high-dimensional best fitted to residues 2–29. (b) A stereoview of the fragment of residues 8–14. The β-protons (all residues) and γ-protons (residues 9 and 12) and the backbone atoms are shown.

Fig. 2a shows ten final calculated structures for CMTI-I and Fig. 2b exhibits a fragment of CMTI-I from residues 8–14. It is evident from Fig. 2 (cf. Fig. 4 and 7 of [10]) and from the data in Table I that the dispersion of the structures and the ranges of the sidechain torsion angles are greater in the structures shown here than in the structures calculated by the floating chirality method [10]. This can be traced to the generous absolute bounds used for the distance constraints in the high-dimensional potential. Thus the structures calculated here may represent a more realistic model of a molecule, allowing for additional flexibility in the structures, yet still producing structures with stereospecifically assigned sidechains. One of the advantages of the high-dimensional potential is that the input parameters provide an adjustable degree of structural variability. Since there is still no universally accepted procedure for conversion of NOE intensities to distances, the high-dimensional potential should accommodate a wide range of future interpretations of the input NOE data.

In conclusion, we have shown that the high-dimensional potential presents an effective way to calculate structures with stereospecifically determined sidechain conformations. The method works very well even for a small peptide, like CMTI-I, which does not have optimal numbers of NOEs connecting to the sidechain protons. For larger proteins there are usually many more such NOEs available. At the same time the effectiveness of the coupling constant method in determining the side-chain conformations for larger proteins diminishes, as coupling constants are less accurate due to the increased line-widths and spectral overlap (this problem could be alleviated with the introduction of a new technique based on the heteronuclear correlation spectra [16]). It is evident therefore that the *n*-dimensional potential is of considerable value in improving the precision of the structure determinations, especially of large proteins. As discussed previously [10] these different methods of stereospecific assignments can be used simultaneously.

The high-dimensional potential can also be used to sort out ambiguous assignments of the NOEs. If there is a connectivity between one residue and another and the latter has two possible residue assignments, the high-dimensional potential will choose the assignment with lower energy.

Acknowledgment: This work was supported by a research grant from the Bundesministerium für Forschung und Technologie (Grant no. 0318909A).

REFERENCES

[1] Weber, P.L., Morrison, R. and Hare, D. (1988) J. Mol. Biol. 204, 483–487.
[2] Holak, T.A., Nilges, M. and Oschkinat, H. (1989) FEBS Lett 242, 218–224.
[3] Guntert, P. Braun, W., Billiter, M. and Wüthrich, K. (1989) J. Am. Chem. Soc. 111, 3997–4004.
[4] Driscoll, P.C., Gronenborn, A.M., Beress, L. and Clore, G.M. (1989) Biochemistry 28, 2188–2198.
[5] Kline, A., Braun, W. and Wüthrich, K. (1988) J. Mol. Biol. 204, 657–724.
[6] Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) J. Mol. Biol. 155, 311–319.
[7] Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. and Wüthrich, K. (1987) J. Mol. Biol. 196, 611–639.
[8] Hyberts, S.G., Mäki, W. and Wagner, G. (1987) Eur. J. Biochem. 164, 625–635.
[9] Pardi, A., Hare, D.R., Selsted, M.E., Morrison R.D., Bassolino, D.A. and Bach II, A.C. (1988) J. Mol. Biol. 201, 625–636.
[10] Holak, T.A., Gondol, D., Otlewski, J. and Wilusz, T. (1989) J. Mol. Biol., 210, 635–648.
[11] Summers, M.F., South, T.L., Kim, B. and Hare D.R. (1990) Biochemistry 29, 329–340.
[12] Holak, T.A., Kearsley, S.K., Kim, Y. and Prestegard, J.H. (1988) Biochemistry, 27, 6135–6142.
[13] Havel, T.F. (1986) DISGEO, Quantum Chemistry Exchange, Program no. 507, Indiana University.
[14] Berendsen, H.J.C. and van Gunsteren, W.F. (1984) Molecular Liquids - Dynamics and Interactions, 475–500.
[15] Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1987) J. Magn. Reson. 75, 474–492.
[16] Montelione, G.T. and Wagner, G. (1989) J. Am. Chem. Soc. 111, 5474.