Volume 213, number 2, 269–277 FEB 04500 March 1987

A comparison of the restrained molecular dynamics and distance geometry methods for determining three-dimensional structures of proteins on the basis of interproton distances

G. Marius Clore, Michael Nilges, Axel T. Brünger*, Martin Karplus* and Angela M. Gronenborn

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG and **Department of Chemistry, Harvard University, 12 Oxford Street Cambridge, MA 02138, USA*

Received 28 October 1986; revised version received 19 December 1986

A direct comparison of the metric matrix distance geometry and restrained molecular dynamics methods for determining three-dimensional structures of proteins on the basis of interproton distances is presented using crambin as a model system. It is shown that both methods reproduce the overall features of the secondary and tertiary structure (shape and polypeptide fold). The region of conformational space sampled by the converged structures generated by the two methods is similar in size, and in both cases the converged structures are distributed about mean structures which are closer to the X-ray structure than any of the individual structures. The restrained molecular dynamics structures are superior to those obtained from distance geometry as regards local backbone conformation, side chain positions and non-bonding energies.

Protein structure; Tertiary structure; NMR; Interproton distance; Distance geometry; Restrained molecular dynamics

1. INTRODUCTION

Two alternative approaches, both with very large radii of convergence, have been described for the determination of three-dimensional structures of proteins on the basis of interproton distance data sets that can be obtained from NMR measurements: (i) distance geometry methods based solely on the use of distance and planarity restraints comprising interproton distances, bond lengths, bond angles, planes and soft van der Waals repulsion terms [1–3]; (ii) a restrained molecular dynamics method in which the classical equations of motion for all atoms are solved subject to a total energy function [4] that consists of bond, angle, torsion, planarity and non-bonding (van der Waals, electrostatic and hydrogen-bonding) potentials supplemented by effective potential terms representing the interproton distance restraints [5–12].

Using model interproton distance data the two distance geometry methods, one based on the metric matrix [1,2], the other on a minimization in torsion angle space with fixed bond lengths and bond angles [3], have been tested on bovine pancreatic trypsin inhibitor, whereas the restrained molecular dynamics method has been tested on crambin [5,7]. The general features that emerge from these model studies are as follows: (i) the overall global three-dimensional structures of the proteins examined are well reproduced by both the

Correspondence address: G.M. Clore, Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG

Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; rms, root mean square; DG, distance geometry structures; RD, restrained molecular dynamics structures; 1 kcal, 4.18 kJ

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

distance geometry and restrained molecular dynamics methods; (ii) quite large deformations in the local backbone structure of regular secondary structure elements are apparent in the structures generated using the distance geometry methods but not in those obtained by restrained molecular dynamics; and (iii) the distance geometry structures tend to be slightly expanded relative to the X-ray structure whereas the restrained dynamics structures are slightly contracted. As the distance geometry and restrained molecular dynamics methods have not been tested using identical data, a quantitative assessment of their relative performance is difficult. In this paper, we therefore present a direct comparison of the distance geometry and restrained molecular dynamics methods. The distance geometry algorithm used in this study is the program DISGEO [1,2,13] which is based on the metric matrix. The model system employed is crambin [14] with exactly the same interproton distance set that we previously used in our restrained molecular dynamics study, comprising a total of 240 interproton distances <4 Å made up of 56 long range ($|i-j|>5$) interresidue distances, 170 short range ($|i-j|\leqslant 5$) interresidue distances and 24 intraresidue distances [5,7].

2. CALCULATIONAL STRATEGY

The distance geometry calculations proceeded in 4 phases. In phase 1 a complete set of bounds on the distances between all atoms of the molecule was determined by triangulation from the NOE interproton distance restraints and the distance and planarity restraints obtained from the primary structure. The latter consisted of assumed exact distances between all covalently bonded and geminal pairs of atoms, as well as lower limits on the distances between all pairs of atoms more than three bonds apart (these were assumed to be no smaller than the sum of the atom hard sphere radii). In the restrained molecular dynamics calculations the interproton NOE distances were represented as effective biharmonic potentials and divided into three distance ranges: $2.5(+0.5/-0.5)$ Å, $3.0(+0.5/-1.0)$ Å and $4.0(+1.0/-1.0)$ Å [5,7]. These ranges correspond to lower and upper bounds of 2–3 Å, 2–3.5 Å and 3–5 Å, respectively, in the distance geometry calculations. In the case of the methylene and methyl protons, appropriate corrections to these ranges were made for the pseudo-atom representation used by DISGEO as described in [15]. In both approaches no restraints corresponding to the three disulphide bridges present in crambin were included in the input data. In phase 2 a set of random substructures was embedded, consistent with the bounds corresponding to distances between a subset of all atoms composed of all the main chain C, Cα, N and CαH atoms together with all non-terminal Cβ and Cγ atoms. This was followed by phase 3 in which a set of initial structures which approximately fit all the data, were computed. This involved choosing approximate distances at random within the triangle limits between all pairs of atoms not in the substructures, given the distances between all atoms in the substructures: this procedure, known as metrization in distance space, was followed by the generation of cartesian coordinates from these distances. Finally, the structures from phase 3 were subjected to 1500 cycles of restrained least squares refinement with respect to all the distances in phase 4. Nine structures were calculated. Of these one contained inconsistent distances in the substructure produced during phase 2 and failed to reach the final phase 4. In the case of the remaining 8 structures, the pseudo atoms were replaced by real atoms and all hydrogen atoms were built on, using the HBUILD algorithm (Brünger, unpublished results) in CHARMM [16], to generate a set of structures known as DG(i) which could be directly compared to the restrained molecular dynamics (RD) structures. In addition, the DG structures were subjected to 1000 cycles of restrained energy minimization carried out as described in [7] to yield the structures DGm(i). The interproton distance restraints and their force constants used in the restrained energy minimization were identical to those used in the refinement stage of the restrained molecular dynamics calculations [5,7].

3. RESULTS

Seven of the 8 DG structures converged to similar structures with an average atomic rms difference between them of 1.8 ± 0.2 Å for the backbone (C, Cα, N, O) atoms (see fig.1). In addition, they satisfy the distance restraints within the errors specified, exhibiting no distance violations

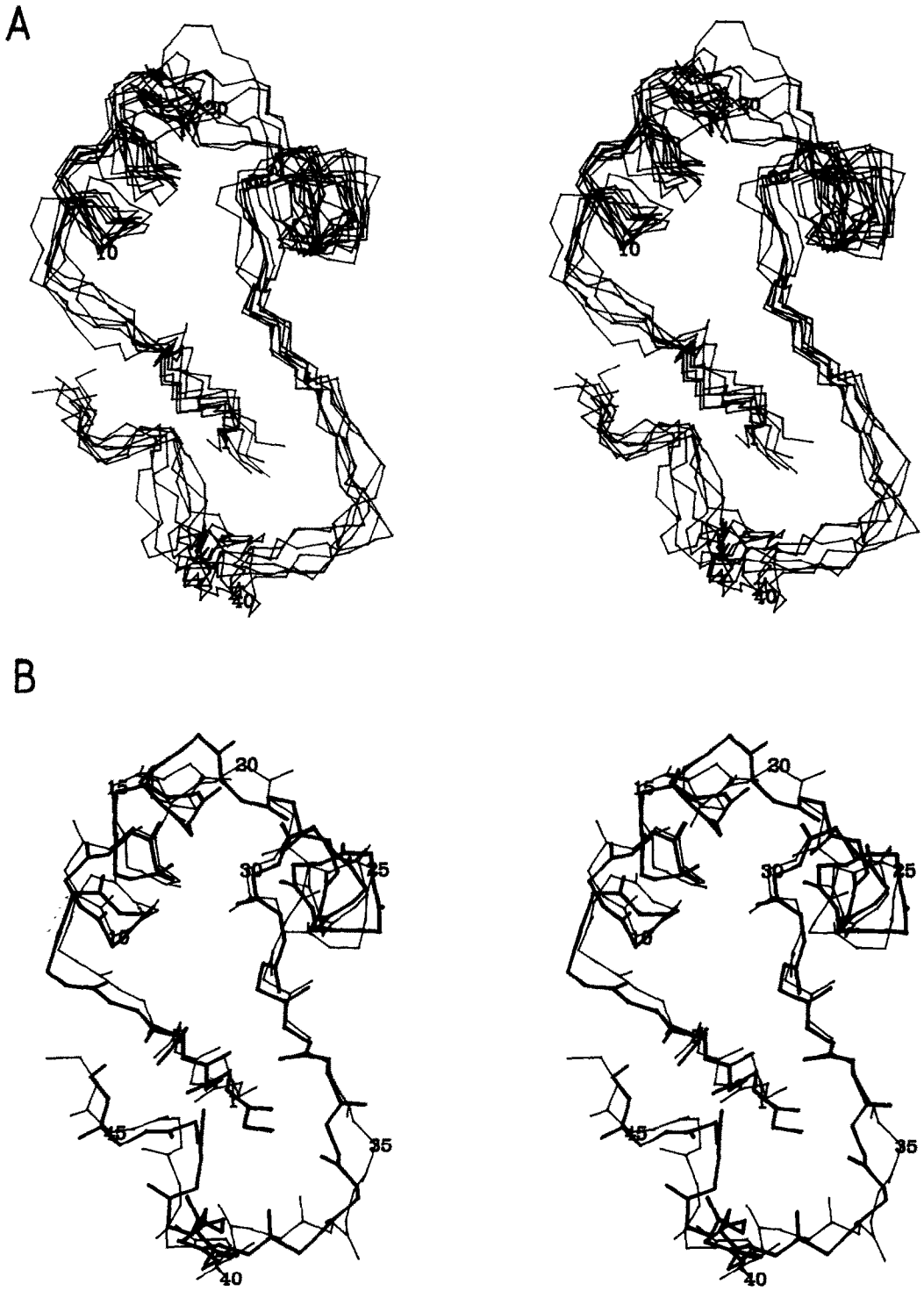

Fig.1. (A) Best fit superposition of the backbone (N, Cα, C) atoms of the 7 converged DG structures. (B) Best fit superposition of the backbone atoms (N, Cα, C, O) of the mean $\overline{DG}$ (thin line) structure and the X-ray structure (thick line). $\overline{DG}$ is the structure obtained by averaging the coordinates of the 7 converged DG structures.

>1.0 Å and an average number of violations between 0.5 Å and 1.0 Å of 1.9 ± 2. The average atomic rms difference between the remaining structure, DG(4), and the other structures is 6.6 ± 0.3 Å for the backbone atoms. In contrast to the other structures, however, DG(4) fails to satisfy the distance restraints with 29 violations between 0.5 Å and 1.0 Å and 12 violations >1.0 Å, indicating that it is incorrectly folded. Consequently, DG(4) was excluded from further consideration. Restrained energy minimization of the DG structures results in only small atomic rms shifts (≤1 Å), and no change in the average atomic rms difference between the structures.

The average atomic rms differences between the converged DG structures and the converged RD structures are comparable (table 1). Thus the region of conformational space sampled by the converged structures generated by the two methods is similar in size. A comparison of the average atomic rms difference for the backbone and side chain atoms of the DG and RD structures with respect to the X-ray structure is shown in fig.2 as a function of residue number. The difference in the average atomic rms difference for the backbone atoms is relatively small: the RD structures are slightly closer to the X-ray structure in the region extending from residues 4 to 31 comprising the two helices (residues 7–19, and 23–30), whereas the DG structures tend to be slightly closer to the X-ray structure for the N- and C-terminal segments. For the side chains, however, the RD structures are significantly closer to the X-ray structure. Indeed, in only 8 of the 42 side chains is the average atomic rms difference for the DG structures smaller than that for the RD structures. Restrained energy minimization of the DG structures has little or no effect on the atomic rms differences with respect to the X-ray structure.

As noted previously [2], structures calculated using distance geometry algorithms have a tendency to be slightly expanded relative to the X-ray structure. This is also the case in the present calculations. The average radius of gyration for the DG

Table 1

Atomic rms differences between various structures

Structure	Atomic rms difference (Å)							
	Backbone atoms only				All atoms			
	X-ray	$\overline{DG}$	$\overline{DGm}$	$\overline{RD}$	X-ray	$\overline{DG}$	$\overline{DGm}$	$\overline{RD}$
<DG>	1.7 (± 0.1)	1.2 (± 0.1)	1.3 (± 0.1)	1.6 (± 0.1)	2.7 (± 0.2)	1.8 (± 0.1)	1.9 (± 0.1)	2.4 (± 0.2)
<DGm>	1.6 (± 0.2)	1.3 (± 0.1)	1.2 (± 0.1)	1.5 (± 0.2)	2.6 (± 0.3)	1.8 (± 0.1)	1.7 (± 0.1)	2.2 (± 0.1)
<RD>	1.8 (± 0.3)	1.7 (± 0.1)	1.7 (± 0.1)	1.4 (± 0.1)	2.3 (± 0.3)	2.3 (± 0.1)	2.2 (± 0.2)	1.7 (± 0.1)
$\overline{DG}$	1.3		0.5	1.1	2.1		0.6	1.5
$\overline{DGm}$	1.1	0.5		0.9	1.9	0.6		1.3
$\overline{RD}$	1.0	1.1	0.9		1.6	1.5	1.3	
$\overline{(DG)}$m	1.3	0.8	0.8	1.2	2.1	1.2	1.2	1.7
$\overline{(DGm)}$m	1.3	1.1	0.8	1.2	2.1	1.3	1.1	1.7
$\overline{(RD)}$m	1.2	1.3	1.2	0.8	1.9	2.0	1.8	1.1

The notation for the structures is as follows: <DG>, <DGm> and <RD> comprise the 7 converged distance geometry structures, the restrained energy minimized structures derived from the DG structures, and the 5 converged restrained molecular dynamics structures, respectively. The latter consist of structures RDIB, RDIC, RDID, RDIE and RDII of [7]. $\overline{DG}$, $\overline{DGm}$ and $\overline{RD}$ are the structures obtained by averaging the coordinates of the DG, DGm and RD structures, respectively. $\overline{(DG)}$m, $\overline{(DGm)}$m and $\overline{(RD)}$m are the structures obtained by restrained energy minimization of $\overline{DG}$, $\overline{DGm}$ and $\overline{RD}$, respectively

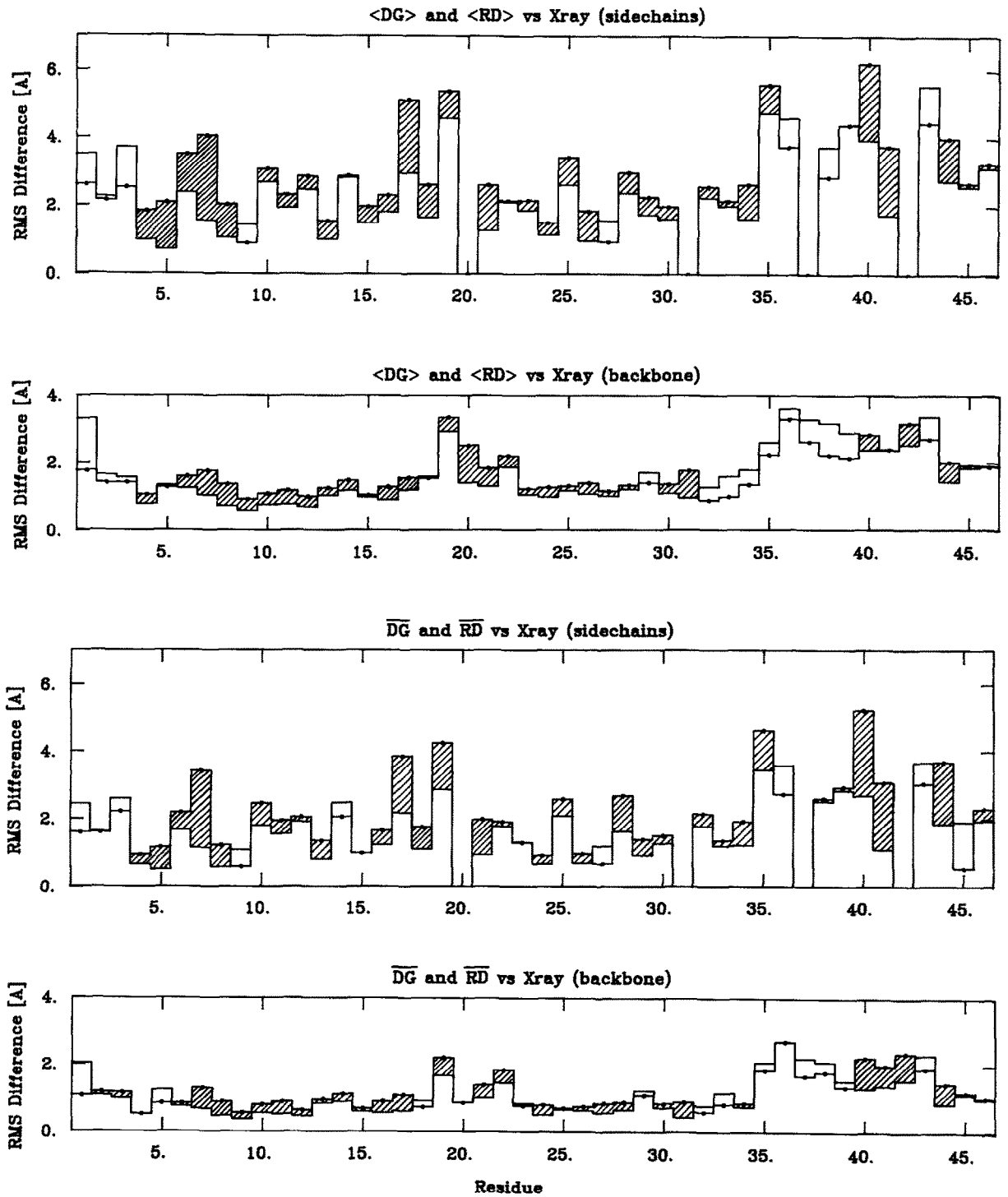

Fig.2. Comparison of the average atomic rms differences of the converged DG (—●—) and RD (——) structures and of the atomic rms differences of the mean $\overline{DG}$ (—●—) and $\overline{RD}$ (——) structures versus the X-ray structure. Plots for the backbone (N, Cα, C, O) and side chain atoms are shown separately, and the shaded areas indicate the regions where the <DG> and $\overline{DG}$ structures are worse than the <RD> and $\overline{RD}$ structures, respectively.

structures is 9.92 ± 0.08 Å compared to a value of 9.64 Å for the X-ray structure. The radius of gyration, however, is reduced on restrained energy minimization of the DG structures to 9.58 ± 0.15 Å. The RD structures, on the other hand, tend to be slightly compressed relative to the X-ray structure with an average radius of gyration of 9.38 ± 0.17 Å. This can be attributed to the attractive component of the van der Waals energy.

In order to assess the quality of the local backbone conformation we have calculated the average angular rms difference for the ϕ,ψ torsion angles of the DG, DGm and RD structures versus the X-ray structure for the two α-helical regions (residues 7–19, and 23–30) separately as well as for the whole molecule. Angles that are in an entirely different conformation (i.e. $|\phi - \phi_{\text{x-ray}}| > 90°$ or $|\psi - \psi_{\text{x-ray}}| > 90°$) are excluded from the mean and counted separately as angular violations. The results are summarized in table 2. It is clear from these data that the local backbone conformation is better reproduced in the RD structures than in the DG and DGm structures. This is manifested by smaller average angular rms differences for the RD structures indicating that those angles which are in the correct conformational region are closer to the X-ray values. Indeed the angular rms differences for the RD structures are very similar to those for a free dynamics structure (FDA) obtained starting from the X-ray structure. In addi-

Table 2

ϕ,ψ angular rms differences and violations for the converged DG and RD structures and for the structures derived from them

Structure	Angular rms difference (°) (angular violations)					
	Helix 1 (residues 7–19)		Helix 2 (residues 23–30)		All residues	
	ϕ	ψ	ϕ	ψ	ϕ	ψ
<DG>	33 ± 11	31 ± 5	41 ± 12	35 ± 4	38 ± 5	39 ± 3
	(1.9 ± 1.5)	(1.7 ± 1.4)	(1.0 ± 1.5)	(0.9 ± 1.2)	(8.4 ± 3.4)	(6.0 ± 2.8)
<DGm>	34 ± 6	31 ± 5	36 ± 8	31 ± 9	38 ± 3	37 ± 4
	(1.3 ± 1.4)	(1.1 ± 1.3)	(0.4 ± 0.8)	(0.6 ± 0.8)	(6.9 ± 4.0)	(4.9 ± 2.0)
<RD>	20 ± 5	18 ± 3	26 ± 3	25 ± 2	29 ± 4	33 ± 3
	(0.6 ± 0.9)	(0.4 ± 0.5)	(0.0 ± 0.0)	(0.6 ± 0.5)	(4.2 ± 2.0)	(4.6 ± 1.5)
$\overline{\text{DG}}$	35	25	33	29	36	33
	(0)	(1)	(0)	(0)	(3)	(2)
$\overline{\text{(DG)}}$m	35	26	30	28	35	30
	(1)	(1)	(0)	(0)	(2)	(3)
$\overline{\text{(DGm)}}$	26	23	24	24	37	32
	(1)	(1)	(0)	(0)	(2)	(3)
$\overline{\text{(DGm)}}$m	26	24	30	29	32	32
	(2)	(1)	(0)	(0)	(5)	(3)
$\overline{\text{RD}}$	27	20	14	17	28	29
	(0)	(0)	(0)	(0)	(2)	(3)
$\overline{\text{(RD)}}$m	17	15	30	28	29	35
	(0)	(0)	(0)	(0)	(2)	(3)
FDA	20	21	19	16	28	34
	(0)	(0)	(0)	(0)	(0)	(0)

The angular violations are defined as the number of angles for which the difference between the values for the calculated structures and the X-ray structure is greater than 90°; these angles are not included in the calculation of the angular rms difference. The notation for the structures is the same as that in table 1. FDA is a free dynamics average structure obtained starting from the X-ray structure [7]

tion, the number of angular violations (i.e. the number of angles in the wrong conformation) in the RD structures is smaller than that in the DG and DGm structures. It is also interesting to note that the ϕ,ψ angles of the first helix are better reproduced than those of the second helix. This is possibly due to the fact that the second helix is somewhat irregular [14].

As there are no terms in the present distance geometry calculations which correspond to the dihedral and non-bonding (i.e. van der Waals, electrostatic, hydrogen bonding) energy terms present in the restrained molecular dynamics calculations, one would expect these energy terms to be significantly larger for the DG structure than for the RD structures. This is indeed the case as can be seen from the energies listed in table 3. Restrained energy minimization of the DG structures results in a significant lowering (by ~800 kcal/mol) of the non-bonded energy terms although the total non-bonded energy of the DGm structures is still ~200 kcal/mol higher than that of the RD structures.

If n structures are normally distributed in a random manner about a mean, then the rms error in the coordinates of the calculated average structure is given approximately by rmsd/$\sqrt{n}$, where rmsd is the average atomic rms difference between the n structures and the average structure. Thus, if the true mean of the n structures is identical to the X-ray structure then the atomic rms difference between the calculated average structure and the X-ray structure should be a factor of ~$\sqrt{n}$ smaller than the average atomic rms difference between the n structures and the X-ray structure. The mean structure $\overline{DG}$ (see fig.1B) obtained by averaging the coordinates of the converged DG structures has an atomic rms standard error of ~0.5 Å for the backbone atoms. The atomic rms difference between $\overline{DG}$ and the X-ray structure is 1.3 Å for the backbone atoms. This value is clearly smaller than the average atomic rms difference between the DG structures and the X-ray structure (see table 1) but about twice as large as expected if the mean about which they were distributed was identical to the X-ray structure. The average structures $\overline{DGm}$ and $\overline{RD}$ are both slightly closer to the X-ray structure than $\overline{DG}$ (table 1) but the differences between them and the X-ray structure are still larger than if they were randomly distributed about a mean corresponding to the X-ray structure.

Table 3

Energies of the converged DG and RD structures and for the structures derived from them

Structure	Energy (kcal/mol)								
	Total	Bond (652)	Angle (1183)	Torsion (320)	Improper (143)	van der Waals	Electro-static	H-bond	NOE restraints (240)
<DG>	1516±453	57±6	266±44	304±12	0.005±0.003	230±336	−106±27	−12±4	777±145
<DGm>	64±101	23±2	225±22	194±14	20±3	−123± 16	−550±45	−47±9	193± 45
<RD>	−406± 23	18±2	196±10	181±18	21±4	−157± 10	−711±22	−69±5	117± 10
$\overline{DG}$	$>10^6$	3.6×10^4	6126	360	0.34	$>10^6$	−2630	1.0×10^5	417
$\overline{DGm}$	$>10^6$	3.3×10^4	5887	345	21	$>10^6$	−2293	−15	138
$\overline{RD}$	$>10^6$	2.2×10^4	4201	349	46	$>10^6$	−1361	9912	100
$\overline{(DG)}$m	−180	20	197	198	23	−148	−540	−41	111
$\overline{(DGm)}$m	−234	21	200	217	21	−140	−614	−48	109
$\overline{(RD)}$m	−428	16	172	167	21	−167	−675	−66	104

The notation for the structures is the same as that in table 1. The number of terms for the bond, angle, torsion, improper and NOE restraints energy terms is given in parentheses. All energy calculations were carried out using the program CHARMM [16] with the same all hydrogen empirical energy potential and biharmonic NOE potential as those used in [7]. The values and errors of the NOE distance restraints are given in table 1 of [7] and their energy contribution is calculated using the same values of force constants as in the refinement stage of the restrained molecular dynamics calculations (see table 3 of [7])

The atomic rms differences between $\overline{DG}$ and $\overline{RD}$ and between $\overline{DGm}$ and $\overline{RD}$ are approximately the same as that between the three average structures and the X-ray structure (see table 1). In this respect it is worth pointing out that the mean structures about which the DG, DGm and RD structures are distributed would not be expected to be identical either to the X-ray structure or to each other for two reasons. First, the distance restraints used do not represent exact distances. Second, the empirical energy function used in the calculations contain imperfections, including the neglect of solvent. These are clearly more severe for the distance geometry calculations than for the restrained molecular dynamics ones as the former ignore the non-bonding energy terms, with the exception of a crude representation of the repulsive component of the van der Waals energy. Restrained energy minimization of the DG structures partially overcomes the deficiencies in the DG structures by locating the closest local subminimum with the global minimum region, but, in contrast to restrained molecular dynamics, does not necessarily locate the lowest energy local subminima.

The effect of averaging the DG, DGm and RD structures is not only reflected in an improvement in the atomic rms difference with respect to the X-ray structure but also in an improvement in other parameters as well, in particular in the ϕ,ψ backbone torsion angles. Thus the average structures have fewer ϕ,ψ angular violations and a lower angular rms difference than the individual structures (see table 2).

The average structures are poor in stereochemical terms, not only with respect to bond lengths and angles, but also as regards non-bonded contacts (see table 3). For this reason $\overline{DG}$ and $\overline{DGm}$ were subjected to 1240 cycles of restrained energy minimization to generate the structures $\overline{(DG)}$m and $\overline{(DGm)}$m, respectively, using the same two step procedure (viz. van der Waals radii set to half their normal values for the first 40 cycles and then restored to their normal values for the remaining cycles) as that used previously for the restrained energy minimization of $\overline{RD}$ [7]. This procedure results in atomic rms shifts of ~1 Å and only minimal changes in either the atomic rms differences with respect to the X-ray structure (table 1) or in the ϕ,ψ torsion angles (table 2). The energies, however, are considerably improved and are approximately the same as those for the respective individual structures (table 3).

4. CONCLUDING REMARKS

In conclusion the results presented in this paper indicate that the general features of the tertiary structure, namely overall shape and fold, are well reproduced by both the distance geometry and restrained molecular dynamics methods. The conformational space sampled by the converged structures generated by the two methods is similar in size and the converged structures are distributed about mean structures which are closer than the individual structures to the X-ray structure. From a structural viewpoint, the restrained molecular dynamics method, however, appears to be superior to the distance geometry method in three respects: the local backbone conformations and side chain positions are better reproduced and the non-bonding energies are smaller. The DG structures can be slightly improved by restrained energy minimization, although the resulting structures are still not as good as the RD ones. This suggests that even larger improvements in the DG structures could be obtained by subjecting them to a few picoseconds of restrained molecular dynamics refinement, as was recently found to be the case in the determination of the solution structures of α1-purothionin [10], phoratoxin [11] and hirudin [12]. Although these differences are small, they may be important, for example in making use of the NOE derived structures to solve crystal structures directly by Patterson search methods [17]. Despite the advantages of the restrained molecular dynamics methods, one caveat in its use should be mentioned: namely the computational time required per structure is 4–5-times longer than that for the distance geometry method. As a result it is advantageous to be able to carry out the structure determination stage of the restrained molecular dynamics calculations on a supercomputer. Bearing these considerations in mind, we suggest that the most efficient strategy for determining three-dimensional structures of proteins on the basis of NOE distance data at the present time involves a combination of the two approaches: namely distance geometry for the structure determination stage followed by restrained molecular dynamics for the refinement stage.

ACKNOWLEDGEMENTS

This work was supported by the Max-Planck Gesellschaft and grant no.Cl 86/1-1 from the Deutsche Forschungsgemeinschaft (G.M.C. and A.M.G.). M.N. acknowledges a pre-doctoral fellowship of the Max-Planck Gesellschaft.

REFERENCES

[1] Havel, T.F. and Wüthrich, K. (1984) Bull. Math. Biol. 46, 673–698.
[2] Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 281–294.
[3] Braun, W. and Go, N. (1985) J. Mol. Biol. 186, 611–626.
[4] Karplus, M. and McCammon, J. (1983) Annu. Rev. Biochem. 52, 263–300.
[5] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) Proc. Natl. Acad. Sci. USA 83, 3801–3805.
[6] Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1985) J. Mol. Biol. 186, 435–455.
[7] Clore, G.M., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) J. Mol. Biol. 191, 523–551.
[8] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) J. Mol. Biol. 182, 179–182.
[9] Nilsson, L., Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1986) J. Mol. Biol. 188, 455–475.
[10] Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) EMBO J. 5, 2729–2735.
[11] Clore, G.M., Sukumaran, D.K., Nilges, M. and Gronenborn, A.M. (1987) Biochemistry, in press.
[12] Clore, G.M., Sukumaran, D.K., Nilges, M., Zarbock, J. and Gronenborn, A.M. (1987) EMBO J., in press.
[13] Havel, T.F. (1986) DISGEO, Quantum Chemistry Program Exchange program no.507, Indiana University.
[14] Hendrickson, W.A. and Teeter, M.M. (1981) Nature 290, 107–113.
[15] Wüthrich, K., Billeter, M. and Braun, W. (1983) J. Mol. Biol. 160, 949–961.
[16] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem. 4, 187–217.
[17] Brünger, A.T., Campbell, R.L., Clore, G.M., Gronenborn, A.M., Karplus, M., Petzko, G.A. and Teeter, M.M. (1987) Science, in press.