Volume 229, number 2, 317–324 FEB 05624 March 1988

Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations

Michael Nilges, G. Marius Clore[+] and Angela M. Gronenborn[+]

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

Received 21 December 1987

A new hybrid distance space-real space method for determining three-dimensional structures of proteins on the basis of interproton distance restraints is presented. It involves the following steps: (i) the approximate polypeptide fold is obtained by generating a set of substructures comprising only a small subset of atoms by projection from multi-dimensional distance space into three-dimensional cartesian coordinate space using a procedure known as 'embedding'; (ii) all remaining atoms are then added by best fitting extended amino acids one residue at a time to the substructures; (iii) the resulting structures are used as the starting point for real space dynamical simulated annealing calculations. The latter involve heating the system to a high temperature followed by slow cooling in order to overcome potential barriers along the pathway towards the global minimum region. This is carried out by solving Newton's equations of motion. Unlike conventional restrained molecular dynamics, however, the non-bonded interactions are represented by a simple van der Waals repulsion term. The method is illustrated by calculations on crambin (46 residues) and the globular domain of histone H5 (79 residues). It is shown that the hybrid method is more efficient computationally and samples a larger region of conformational space consistent with the experimental data than full metric matrix distance geometry calculations alone, particularly for large systems.

Three-dimensional structure; Interproton distance; Nuclear Overhauser effect; Distance space; Real space; Distance geometry; Dynamical simulated annealing

1. INTRODUCTION

With the advent of a whole array of two-dimensional NMR techniques [1], it has become possible to obtain virtually complete resonance assignments as well as a large number of interproton distance restraints for small (<100 residues) proteins (reviews [2,3]). The latter form the basis for determining the three-dimensional structures of proteins in solution. To date two general classes of methods have been proposed for tackling of this problem. The first can be termed real space methods. These include restrained least-squares minimization in torsion angle space with either a variable target function [4] or a sequence of ellipsoids of constantly decreasing volume, each of which contains the minimum of the target function [5], and restrained molecular dynamics [6–10]. Because these methods operate in real space, great care has to be taken to ensure that incorrect folding of the polypeptide chain does not occur. In contrast, the folding problem does not exist in the second class of methods which operates in distance space and is generally referred to as metric matrix distance geometry [11–17]. Here,

Correspondence address: G.M. Clore, Laboratory of Chemical Physics, Building 2, Room 123, National Institute for Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[+] *Present address:* Laboratory of Chemical Physics, Building 2, Room 123, National Institute for Diabetes and Digestive and Kidney Diseases. National Institutes of Health, Bethesda, MD 20892, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies

the coordinates of the calculated structure are generated by a projection from $N(N-1)/2$ dimensional distance space (N, number of atoms) into three-dimensional coordinate space by a procedure known as embedding. As the distances, however, do not define the chirality of the structure, mirror images (local or global) of the correct structure can occur. In general, these can be easily rejected as the chirality of single amino acids (L) and helices (right-handed) is known. Nevertheless, distance space methods have a number of drawbacks: in particular, (i) inefficient sampling of the available conformational space [18–24]; (ii) long computational times required for the embedding of all atoms into three-dimensional space, which rise very rapidly as the size of the molecule increases [13]; and (iii) relatively poor stereochemistry of the final structures, particularly in terms of non-bonded contacts [18–23].

In this paper, we suggest a new and efficient approach based on a combination of distance and real space methods. Instead of folding the polypeptide chain from an extended strand, a substructure comprising only a subset of atoms is generated by a distance space method, and used as the starting point for subsequent real space calculations. The real space calculations involve the application of a powerful method, termed dynamical simulated annealing, which efficiently overcomes potential energy barriers on the path to the global minimum region by raising the temperature of the system initially followed by slow cooling. Thus, it is similar in spirit to high-temperature restrained molecular dynamics [6,8,25,26]. In contrast to restrained molecular dynamics calculations, we use purely geometric restraints and the non-bonded interactions are represented by a simple repulsion term. The latter replaces the dihedral, van der Waals, electrostatic and hydrogen-bonding potentials of the empirical energy function in conventional molecular dynamics. Two examples are used to illustrate the method: crambin and the globular domain of histone H5 (GH5) which have 46 and 79 residues, respectively. It is shown that the hybrid approach is significantly faster (by a factor of two in the case of GH5), samples a larger region of conformational space and produces higher quality structures than metric matrix distance geometry calculations alone.

2. METHODOLOGY AND CALCULATIONAL STRATEGY

The projection from high-dimensional distance space to three-dimensional coordinate space with the metric matrix distance geometry program DISGEO [14,16] is carried out in four phases. In phase 1, a complete set of bounds on the distances between all atoms of the molecule is determined by triangulation from the experimental restraints and from the distance and planarity restraints obtained from the primary structure. The latter consist 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 which are assumed to be no smaller than the sum of the hard-sphere van der Waals radii. The complete set of distance bounds is then subjected to a procedure known as bound smoothing which involves the selection of the smallest possible intervals between lower and upper bounds consistent with the triangle inequalities. Phase 1 takes a few minutes of CPU time and has to be performed only once for a given set of experimental data. In phase 2 a subset of atoms is embedded (e.g. N, C, $C^{\alpha}H$, C^{β}, non-terminal C^{γ} and C^{δ}, and a pseudoatom for the aromatic rings). The resulting substructure is subjected to conjugate gradient minimization, similar to that in phase 4 below. The distances between the atoms in the substructure are then relaxed and included as additional distance constraints for the embedding of all atoms in phase 3. As bound smoothing is performed for every single distance chosen within its allowed limits, the third phase is very time-consuming. Phase 2, on the other hand, only takes a few minutes of CPU time per substructure as it does not require the time-consuming checking of triangle inequalities. Phase 3 is then usually followed by a conventional conjugate gradient minimization of a target function which contains terms for bond lengths, angles, planes, chirality and NOE-distance restrains (phase 4).

The basis of simulated annealing involves raising the temperature of the system followed by slow cooling in order to overcome local minima and locate the global minimum region of the target function. In the original description of simulated annealing [27], the Metropolis algorithm [28] was used to simulate a system at a temperature T. In our application we make use of an integration algorithm to solve Newton's equations of motion in an analogous fashion to that used in molecular dynamics [29].

The total target function F_{tot} for which the global minimum region is searched comprises the following terms:

$$F_{\text{tot}} = F_{\text{covalent}} + F_{\text{repel}} + F_{\text{NOE}} \quad (1)$$

F_{tot} represents the effective potential energy in the dynamics calculation. This involves integration of Newton's equations of motion:

$$\frac{\partial^2 x_i}{\partial t^2} = -\frac{1}{m_i}\frac{\partial}{\partial x_i} F_{\text{tot}}(x_1,x_2,\ldots,x_n) \quad (2)$$

for all n atoms of the system. Thus, the temperature T at a time t is given by:

$$T_t = \frac{2}{k_B(3n-6)} \left(\sum_{i=1}^{n} m_i v_i^2/2\right)_t \quad (3)$$

F_{covalent} maintains correct bond lengths, angles, planes and chirality, and is given by

$$F_{\text{covalent}} = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{impropers}} k_\phi(\phi - \phi_0)^2 + \sum_{\omega} k_\omega(\omega - \omega_0)^2 \qquad (4)$$

The force constants of the energy terms for bonds (k_b), angles (k_θ), improper torsions (k_ϕ) and ω peptide bond dihedral angles are set to uniform high values to ensure near perfect stereochemistry of the single amino acid residues throughout the calculation, the respective values being 600 kcal·mol^{-1}·Å^{-2}, 500, 500 and 200 kcal·mol^{-1}·rad^{-2}. (Note that the peptide bond is assumed to be planar and *trans*, and that the improper torsion terms serve to maintain appropriate planarity and chirality.) There are no terms for the other dihedral angles at rotatable bonds as these effectively represent a non-bonded interaction.

The non-bonded interactions are described by a single repulsion term, F_{repel}, with a variable force constant k_{vdW}, to prevent unduly close non-bonded contacts:

$$F_{\text{repel}} = \begin{bmatrix} 0 & \text{if } r \geq r_{\min} \\ k_{\text{vdW}}(s^2 \cdot r_{\min}{}^2 - r^2)^2, & \text{if } r < r_{\min} \end{bmatrix} \qquad (5)$$

The values of $r_{\min}$ are the standard values of the van der Waals radii as represented by the Lennard-Jones potential used in the CHARMM empirical energy function [29]. s is set to 0.8 in the present calculations. The resulting hard-sphere radii are similar to those used in the various distance geometry programs [4,5,16].

The NOE distance restraints are represented by a square-well potential with the variable force constant k_{NOE} [18]:

$$E_{\text{NOE}} = \begin{bmatrix} k_{\text{NOE}}(r_{ij} - r_{ij}{}^{\text{u}})^2, \text{ if } r_{ij} > r_{ij}{}^{\text{u}} \\ 0 \qquad\qquad , \text{ if } r_{ij}{}^{1} \leq r_{ij} \leq r_{ij}{}^{\text{u}} \\ k_{\text{NOE}}(r_{ij} - r_{ij}{}^{1})^2 , \text{ if } r_{ij} < r_{ij}{}^{1} \end{bmatrix} \qquad (6)$$

where $r_{ij}{}^{\text{u}}$ and $r_{ij}{}^{1}$ are the values of upper and lower limits of the target distances, respectively.

The calculational strategy employed is illustrated by the flow chart in fig.1. First, a set of DISGEO [16] substructures (known collectively as ⟨Sub⟩) is generated with the subset of atoms listed above. All subsequent calculations are carried out with the program XPLOR (Brünger, A.T., unpublished; [6,9,25,26]). Amino acids in an extended conformation (ϕ = $-120°$, $\psi = 120°$, $\chi_i = 180°$) are first best fitted to the substructures residue by residue. 200 cycles of unrestrained (i.e. no NOE potential) Powell minimization with a very low force constant (k_{vdW} = 0.01 kcal·mol^{-1}·Å^{-2}) on the van der Waals repulsion term are then carried out to improve the covalent geometry prior to dynamical simulated annealing. The annealing schedule is carried out in two steps. Step 1 comprises 50 cycles of 75 fs dynamics each. The initial velocities are chosen from a Maxwell distribution at 1000 K. After each cycle the velocities are rescaled to 1000 K. During the first few cycles the value of force constant k_{NOE} is doubled at the beginning of each new cycle from an initial value of 0.1 to a maximum value of 50 kcal·mol^{-1}·Å^{-2}. To make rearrangements possible the initial value of k_{vdW} is very low (0.01 kcal·mol^{-1}·Å^{-2}). This is increased slowly by multiplying its value by a factor of $400^{1/50}$ prior to each cycle up to a final value of 4 kcal·mol^{-1}·Å^{-2}. These relative values of the force constants are somewhat arbitrary. They do, however, maintain nearly perfect covalent geometry of the structure, and ensure that no close contacts occur and that the NOE restraints are effectively introduced and satisfied. Step 2 of the simulated annealing schedule consists of 1.5 ps dynamics at 300 K with values of the force constants k_{NOE} and k_{vdW} equal to their final values at the end of step 1. This is followed by 200 cycles of restrained Powell minimization.

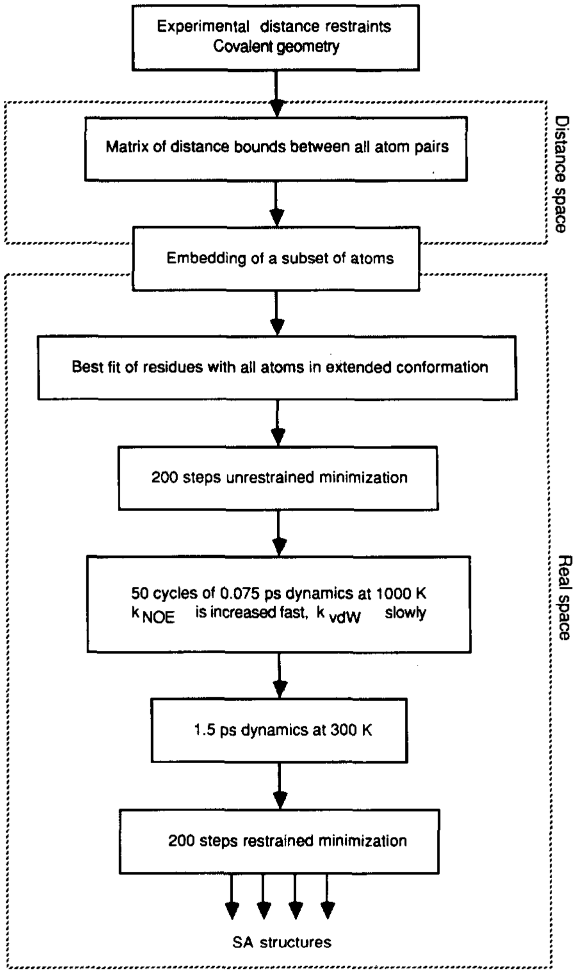
Experimental distance restraints
Covalent geometry
Matrix of distance bounds between all atom pairs
Embedding of a subset of atoms
Distance space
Best fit of residues with all atoms in extended conformation
200 steps unrestrained minimization
50 cycles of 0.075 ps dynamics at 1000 K
k NOE is increased fast, k vdW slowly
1.5 ps dynamics at 300 K
200 steps restrained minimization
SA structures
Real space

Fig.1. Flow chart of the strategy used in the hybrid distance geometry-dynamical simulated annealing calculations.

It is intuitively evident that raising the force constants has a similar effect to lowering the temperature. If all potentials were to be harmonic and all force constants were to be changed at the same rate, they would be shown to be exactly equivalent. Changing the force constants rather than the temperature, however, is more convenient as a variable step integrator is not required and the values of the various force constants can be altered differentially. Thus, increasing the force constant of F_{NOE} faster than that of F_{repel} proved to be more efficient than changing both at the same rate.

3. RESULTS OF CALCULATIONS ON CRAMBIN AND GH5

The results of the calculations on crambin and GH5 are summarized in tables 1,2 and fig.2. The same sets of NOE distances were employed as in our previous calculations [8,21]. In the case of crambin the NOE data set consisted of 240 interproton distances derived from the crystal structure [30]; for GH5 it comprised the 307 interproton distances derived from NOE measurements [21]. The distances were classified into three distance ranges, 1.8–2.7, 1.8–3.3 and 1.8–5.0 Å, corresponding to strong, medium and weak NOEs [31]. Distances to methyl, methylene and aromatic ring protons that were not assigned stereospecifically were calculated with respect to the average position of these protons and the upper limits of the corresponding restraints were corrected appropriately as described in [32]. In the

Table 1

NOE deviations, violations and energies, deviations of covalent geometry from ideality and van der Waals energies[a]

	NOE[b]			Deviations from ideality				Van der Waals energies[c] (kcal·mol^{-1})	
	RMS (Å)	Violation	F_{NOE} (kcal·mol^{-1})	Bonds (Å)	Angles (°)	Impropers (°)	ω (°)	F_{repel}	$E_{\mathrm{L\text{-}J}}$
Crambin									
⟨Sub⟩	1.15	74	10012	0.013	2.45	1.10	21.2	18428	$>10^6$
	± 0.18	± 9	± 3166	± 0.003	± 0.26	± 0.18	± 5.8	± 3603	
⟨SA⟩	0.06	0.2	24	0.006	2.08	0.35	3.4	33	−113
	± 0.01	± 0.4	± 9	± 0.001	± 0.24	± 0.10	± 0.9	± 8	± 16
⟨DG⟩	0.14	1.9	3126	0.017	3.79	0.15	0.48	929	230
	± 0.04	± 0.2	± 2326	± 0.001	± 0.29	± 0.05	± 0.28	± 167	± 336
$\overline{\text{SA}}$	0.08	1	46	0.428	22.48	0.30	2.1	8382	$>10^6$
$(\overline{\text{SA}})$r	0.04	0	9	0.005	2.00	0.29	3.3	29	−131
X-ray[d]	0	0	0	0.020	2.87	1.48	3.9	438	−213
GH5									
⟨Sub⟩	1.12	94	18462	0.014	2.80	1.08	26.7	47527	$>10^6$
	± 0.13	± 8	± 4304	± 0.001	± 0.37	± 0.21	± 3.9	± 7744	
⟨SA⟩	0.14	5	299	0.012	2.22	0.84	8.0	169	−100
	± 0.02	± 2	± 62	± 0.001	± 0.26	± 0.28	± 1.6	± 38	± 19
⟨DG⟩	0.57	50	4850	0.389	7.61	7.23	33.3	5846	17752
	± 0.06	± 9	± 831	± 0.585	± 0.52	± 1.98	± 8.5	± 763	± 26323
$\overline{\text{SA}}$	0.12	6	226	0.599	27.56	3.95	39.0	114450	$>10^6$
$(\overline{\text{SA}})$r	0.10	3	146	0.017	2.61	1.21	9.6	266	14

[a] Notation of structures: ⟨Sub⟩, converged substructures obtained from phase 2 of DISGEO, after best fitting of residues and subsequent unrestrained minimisation (see text); ⟨SA⟩, converged structures produced by dynamical simulated annealing starting from DISGEO substructures; ⟨DG⟩, final converged structures obtained with DISGEO alone (from [33], for crambin; from [21] for GH5); $\overline{\text{SA}}$, $\overline{\text{DG}}$, $\overline{\text{Sub}}$, mean structures obtained by averaging over the coordinates of individual SA, DG, Sub structures, respectively; $(\overline{\text{SA}})$r, structure obtained by restrained minimisation of the mean $\overline{\text{SA}}$ structure. There are 9 ⟨Sub⟩ and ⟨SA⟩ structures for crambin, and 10 for GH5; and 7 ⟨DG⟩ structures for crambin and 6 for GH5

[b] Interproton distance deviations calculated with respect to upper and lower limits of restraints. NOE violations: number of distances for which the difference between target and calculated values is >0.5 Å. NOE energy, F_{NOE}, calculated with a value of 50 kcal·mol^{-1}·Å^{-2} for the force constant k_{NOE} (cf. eqn 6)

[c] The van der Waals repulsion energy F_{repel} (cf. eqn 5) calculated with values of 4 kcal·mol^{-1}·Å^{-2} and 0.8 for force constant k_{vdW} and the van der Waals radius scale factor s, respectively. $E_{\mathrm{L\text{-}J}}$, full Lennard-Jones van der Waals energy calculated using the CHARMM empirical energy function [29]; it is not present in the target function for the dynamical simulated annealing calculations

Table 2

Atomic RMS distributions and shifts

	RMS difference (Å)	
	Backbone atoms (N,C$^\alpha$,C,O)	All atoms
Crambin		
⟨Sub⟩ vs $\overline{\text{Sub}}$	1.4 ± 0.2	–
⟨SA⟩ vs ⟨Sub⟩	2.6 ± 0.4	3.6 ± 0.3
⟨SA⟩ vs $\overline{\text{SA}}$	1.2 ± 0.3	1.7 ± 0.3
⟨DG⟩ vs $\overline{\text{DG}}$	1.2 ± 0.1	1.8 ± 0.1
$(\overline{\text{SA}})$r vs $\overline{\text{SA}}$	0.5	1.1
$\overline{\text{DG}}$ vs $\overline{\text{SA}}$	1.1	1.7
⟨SA⟩ vs X-ray	1.6 ± 0.2	2.4 ± 0.2
$\overline{\text{SA}}$ vs X-ray	1.0	1.7
$(\overline{\text{SA}})$r vs X-ray	1.1	1.9
$\overline{\text{DG}}$ vs X-ray	1.3	2.1
GH5		
⟨Sub⟩ vs $\overline{\text{Sub}}$	1.3 ± 0.1	–
⟨SA⟩ vs ⟨Sub⟩	4.0 ± 0.4	5.1 ± 0.3
⟨SA⟩ vs $\overline{\text{SA}}$	3.1 ± 0.3	3.8 ± 0.3
⟨DG⟩ vs $\overline{\text{DG}}$	2.1 ± 0.4	2.8 ± 0.7
$\overline{\text{SA}}$ vs $(\overline{\text{SA}})$r	1.4	2.1
$\overline{\text{SA}}$ vs $\overline{\text{DG}}$	2.6	3.4

Notation of structures as given in table 1

case of crambin 9 additional restraints for the three disulphide bridges were also included in the distance restraints list. 10 DISGEO substructures were generated for crambin and 50 for GH5. The coordinate generation was followed by 100 steps of conjugate gradient minimization. 9 crambin substructures and 36 GH5 substructures converged to the correct polypeptide fold. The remaining substructures were global mirror images. All 9 converged crambin substructures were subjected to simulated annealing; for GH5, the 10 converged substructures with the lowest value of the target function after minimization of the substructures were selected.

The quality of the calculated structures can be assessed in a quantitative manner by examining the deviations between the calculated and experimental interproton distances, the deviations of covalent geometry from ideality and the value of the van der Waals repulsion term F_{repel} (table 1). From these data it is clear that, in the case of the substructures, deviations between the experimental and calculated interproton distances are large and the stereochemistry is poor (table 1), although the overall polypeptide fold may be approximately correct (table 2). Considerable improvements are obtained by both dynamical simulated annealing and full DISGEO (i.e. phases 3,4) calculations to generate the ⟨SA⟩ and ⟨DG⟩ structures, respectively. In general, however, the final ⟨SA⟩ structures satisfy the NOE data better and display better stereochemistry than the final ⟨DG⟩ ones (table 1). This is particularly evident as the size of the protein increases (cf. GH5).

The mark of a good method is one that samples efficiently the conformational space consistent with the experimental data. This has two consequences. The atomic RMS distribution of regions that are poorly defined (either globally or locally) by the experimental data should be increased, while that of regions that are well defined by the data should be reduced as the agreement between the observed and calculated interproton distances is improved. The atomic RMS distributions and shifts of the structures before and after dynamical simulated annealing are summarised in table 2, and the backbone atomic RMS distribution of the individual SA structures about their mean as a function of residue number if shown in fig.2. GH5 contains an approximately equal mixture of well- and poorly defined regions [21]. The atomic RMS distributions of the GH5 substructures is approximately the same as those in the case of crambin (table 2), despite the fact that the latter is only half the size of the former and much better defined by the data [8,21]. Thus, as noted in [24], the relatively inefficient sampling of the available conformational space by metric matrix distance geometry calculations is already present in the substructure-generating phase. Although the atomic RMS distribution for GH5 is increased by ~50% after full DISGEO calculations, it is doubled by dynamical simulated annealing (table 2). A similar increase in atomic RMS distribution is also obtained after subjecting the ⟨DG⟩ structures to restrained molecular dynamics refinement [21]. In the case of crambin, however, the overall structure is well-defined both locally and globally by the data. As a result, the atomic RMS distributions of the final crambin DG and SA structures are in fact slightly smaller than that of the substructures (table 2).

A comparison with the X-ray structure in the case of crambin enables one to assess the effects of limitations in the number, range (<5 Å) and ac-

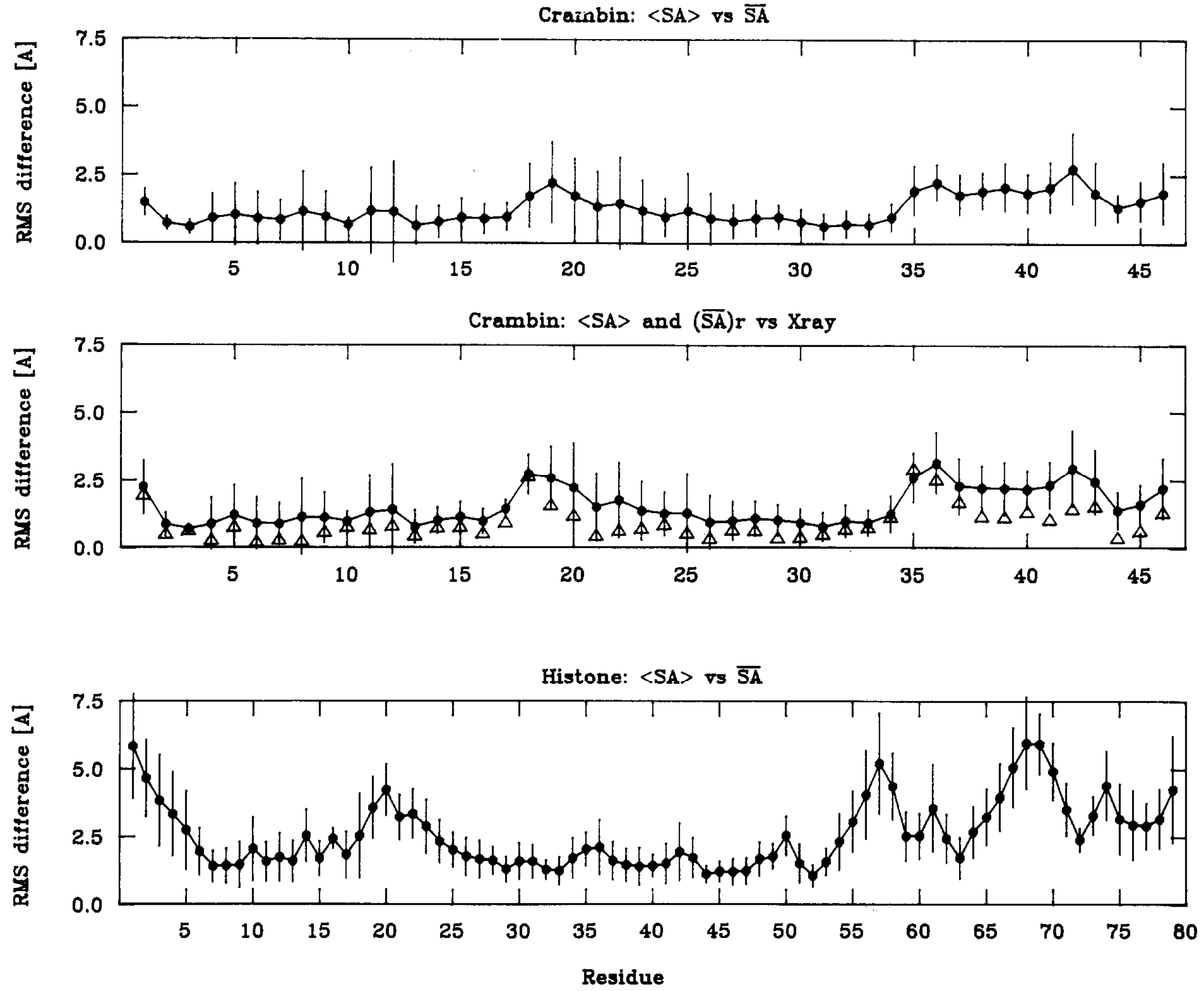

Fig.2. Backbone (N,C^{α},C,O) atomic RMS distributions of the individual SA structures about their mean $\overline{SA}$ structure for crambin and GH5, and backbone atomic RMS difference between the individual crambin ⟨SA⟩ and restrained minimized average ($\overline{SA}$)r structures, on the one hand, and the crambin X-ray structure, on the other. (●) Average RMS difference at each residue between the individual SA structures and either the mean structure or X-ray structure; the bars represent the corresponding standard deviations. (△) RMS difference between the crambin restrained minimized average structure ($\overline{SA}$)r and the crambin X-ray structure.

curacy of the interproton distance restraints on the quality of the structure determination. The average atomic RMS difference between the individual SA structures and the X-ray structure is similar to that for structures calculated by either full metric matrix distance geometry (table 2; [33]) or restrained molecular dynamics [8]. In addition, as in the case of other methods [8,33], the mean structure $\overline{SA}$, obtained by averaging the coordinates of the individual SA structures best fitted to each other, is closer to the X-ray structure than any of the individual SA structures (table 2 and fig.2). The mean structure is of course poor in terms of stereochemistry. This is easily corrected by subjecting it to restrained minimization to generate the structure ($\overline{SA}$)r (table 1). This procedure results in atomic RMS shifts of ≤1 Å and only minimal changes in the atomic RMS difference with respect to the X-ray structure (table 2).

A further feature of the hybrid method is its apparent computational efficiency relative to that of the full DISGEO calculations. This becomes increasingly obvious as the size of the system is increased. Thus, for proteins the size of crambin (46 residues), there is no significant difference in CPU times between both methods (~45 min per structure on a Vax 8550). For proteins the size of GH5 (79 residues), however, the hybrid approach is faster by a factor of 2 (viz. 6 h for phases 3,4 of DISGEO vs 3 h for the hybrid method on a Vax 8550). Additionally, the quality of the final SA structures does not depend on the quality of the substructures, once the correct fold is obtained. Thus, two of the substructures calculated for crambin contained distances inconsistent with the full input distance data and aborted during phase 3 of the DISGEO calculations. The final ⟨SA⟩ structures obtained from these same substructures, however, were of the same quality as the others. In our experience, the chances of obtaining inconsistent distances in the substructures increase with the size of the protein.

4. CONCLUDING REMARKS

The hybrid distance geometry-dynamical simulated annealing approach presented here combines computational speed with a reduced dependence on the quality of the substructures and an improved quality of the final structures. This enables one to calculate a large number of structures in a reasonable time frame (e.g. 32 h on a Vax 8550 for 10 structures in the case of the 79-residue GH5), thereby permitting one to sample efficiently the conformational space consistent with the experimental data and to obtain higher quality average structures (i.e. lower coordinate standard errors). The latter is important as calculations with model data on crambin (this paper and [8,33]), as well as with experimental data on potato carboxypeptidase inhibitor [22] and barley serine proteinase inhibitor 2 [34], indicate that the atomic RMS difference between the average structure and the X-ray structure becomes smaller as the number of calculated structures as well as their quality is increased.

The structures presented here are not refined by energy minimization or restrained molecular dynamics. This could of course be carried out. In addition, we note that further information can easily be included in the annealing phase. This is the subject of ongoing research. Examples are the inclusion of additional backbone torsion angle restraints deduced either from the pattern of NH(i)–NH($i+1$), C^{α}H(i)–NH($i+1$) and C^{β}H(i)–NH($i+1$) NOEs as proposed by Sherman et al. [35] or from a data bank search with short-range NOEs [36].

As the initial bound smoothing is carried out with all atoms, the hybrid approach still has the large memory requirements of a metric matrix distance geometry algorithm. This, however, should not present a real problem with modern computers. Further, as the accuracy required in the substructure generating phase is not very stringent, the memory requirements can be reduced by including fewer atoms in the residue library used to generate the substructures.

Acknowledgements: This work was supported by the Max-Planck Gesellschaft and grant 321/4003/0318909A from the Bundesministerium für Forschung und Technologie.

REFERENCES

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

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

[3] Clore, G.M. and Gronenborn, A.M. (1987) Protein Eng. 1, 275–288.

[4] Braun, W. and Go, N. (1985) J. Mol. Biol. 186, 611–626.

[5] Billeter, M., Havel, T.F. and Wüthrich, K. (1987) J. Comput. Chem. 8, 132–141.

[6] Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1985) J. Mol. Biol. 185, 435–455.

[7] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) J. Mol. Biol. 182, 179–182.

[8] Clore, G.M., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) J. Mol. Biol. 191, 523–551.

[9] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) Proc. Natl. Acad. Sci. USA 83, 3801–3805.

[10] Nilsson, L., Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1986) J. Mol. Biol. 188, 455–475.

[11] Crippen, G.M. and Havel, T.F. (1978) Acta Crystallogr. A34, 282–284.

[12] Kuntz, I.D., Crippen, G.M. and Kollman, P.A. (1979) Biopolymers 18, 939–957.

[13] Havel, T.F., Kuntz, I.D. and Crippen, G.M. (1983) Bull. Math. Biol. 45, 665–720.

[14] Havel, T.F. and Wüthrich, K. (1984) Bull. Math. Biol. 46, 673–698.

[15] Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 281–294.

[16] Havel, T.F. (1986) DISGEO, Quantum Chemistry Exchange, Program no.507, Indiana University.

[17] Sippl, M.J. and Scheraga, H.J. (1986) Proc. Natl. Acad. Sci. USA 83, 2283–2287.

[18] Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) EMBO J. 5, 2729–2735.

[19] Clore, G.M., Sukumaran, D.K., Nilges, M. and Gronenborn, A.M. (1987) Biochemistry 26, 1732–1745.

[20] Clore, G.M., Sukumaran, D.K., Nilges, M., Zarbock, J. and Gronenborn, A.M. (1987) EMBO J. 6, 529–537.

[21] Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K. and Zarbock, J. (1987) EMBO J. 1833–1842.

[22] Clore, G.M., Gronenborn, A.M., Nilges, M. and Ryan, C.A. (1987) Biochemistry 26, 8012–8023.

[23] Clore, G.M., Gronenborn, A.M., Kjaer, M. and Poulsen, F.M. (1987) Protein Eng. 1, 305–311.

[24] Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. and Wüthrich, K. (1987) J. Mol. Biol. 196, 611–640.

[25] Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.

[26] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1987) Protein Eng. 1, 399–406.

[27] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) Science 220, 671–680.

[28] Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A. and Teller, E. (1953) J. Chem. Phys. 21, 1087–1092.

[29] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., Sates, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem. 4, 187–217.

[30] Hendrickson, W.A. and Teeter, M.M. (1981) Nature 290, 107–112.

[31] Williamson, M.P., Havel, T.F. and Wüthrich, K. (1985) J. Mol. Biol. 182, 295–315.

[32] Wüthrich, K., Billeter, M. and Braun, W. (1983) J. Mol. Biol. 169, 949–961.

[33] Clore, G.M., Nilges, M., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1987) FEBS Lett. 213, 269–277.

[34] Clore, G.M., Gronenborn, A.M., James, M.N.G., Nilges, M., Kjaer, M., McPhalen, C.A. and Poulsen, F.M. (1987) Protein Eng. 1, 313–318.

[35] Sherman, S.A., Andrianov, A.M.. and Akhrem, A.A. (1987) J. Biomol. Struct. Dyn. 4, 869–884.

[36] Kraulis, P.J. and Jones, T.A. (1987) in: Structure, Dynamics and Function of Biomolecules (Ehrenberg, A. et al. eds) pp.118–121, Springer, Berlin.