

SAMPLING PROPERTIES OF SIMULATED ANNEALING AND DISTANCE GEOMETRY

Michael Nilges, John Kuszewski, and Axel T. Brünger
The Howard Hughes Medical Institute and
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, CT 06511

ABSTRACT

Properties of two different methods for calculating three-dimensional structures of macromolecules from nuclear magnetic resonance data, distance geometry, and simulated annealing, are studied. It is shown that a simulated annealing refinement of structures generated with distance geometry is sufficient to remove bias introduced in the distance geometry stage. A new efficient simulated annealing protocol which does not require initial structures from distance geometry is presented. The influence of distance selection on quality and distribution of structures generated with distance geometry is studied.

INTRODUCTION

Several recent studies indicate that sampling properties of metric matrix distance geometry (DG) algorithms are insufficient in a more fundamental way than previously thought¹⁻³. In particular, it has been suggested that the distance geometry approach may not be able to find folds of the polypeptide chain which are compatible with the NOE derived distance data³, and that a simulated annealing (SA) stage following the distance geometry stage may not suffice to insure appropriate sampling of the conformational space².

We investigate in the present study if the SA stage as implemented in the hybrid DG-SA method⁴ is sufficient to remove bias introduced in the distance geometry stage. Further, a new SA protocol is presented which does not require starting structures obtained with a DG program and thus avoids any DG-related sampling problems; and is simpler and more efficient than earlier protocols^{5,6}. We also study the effects of metrization on the distribution and quality of structures produced by DG.

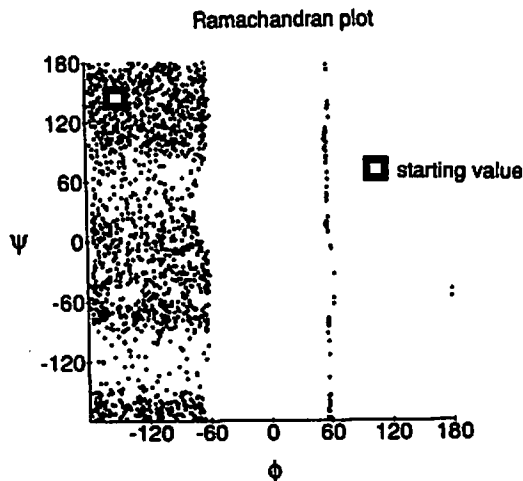


Figure 1

SAMPLING OF SA REFINEMENT

A particularly biased starting structure is used to study the amount of randomization by the SA stage of the DG-SA hybrid method, namely, a polyalanine 20mer in an extended conformation ($\phi = -139^\circ$, $\psi = 135^\circ$). 100 structures were calculated in the absence of any tertiary distance information. The only difference between calculations was the random number seed used to assign initial velocities. While it is intuitively clear that molecular dynamics at high temperature would randomize the conformation it has not been shown in detail for the particular conditions used in the SA refinement in (4) (very low van der Waals interaction initially, short dynamics run at 1000° K).

As a reference set we generated 100 structures by randomly assigning values for ϕ and ψ , where sterically forbidden regions in the $\phi - \psi$ map were excluded. Only structures which had little van der Waals (vdW) overlap were accepted. The mean end-to-end distance of the structures generated in this way (≈ 28 Å) agrees with values obtained by others on the same system (J. Thomason, personal communication).

Fig. 1 shows a Ramachandran plot for all 100 calculated SA structures. The square in the top left corner indicates the starting conformation. As can easily be seen, all of the allowed region is visited. In Fig. 2, we compare how randomly the structures are distributed in space globally. For each C^α atom in the sequence, the number of C^α atoms within a distance of 7.6 and 11.4 Å (2 and 3 $C_i^\alpha - C_{i+1}^\alpha$ distances) is counted. This number is averaged over 100 structures. The two curves are almost identical, the random structures are only slightly more densely packed than the SA structures. This also indicates

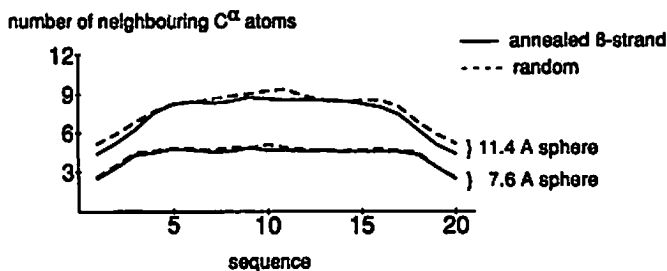


Figure 2

Table 1. average properties of $[\text{Ala}]_{20}$ conformations

calculation	r.m.s. bonds* [Å]	r.m.s. angles* [°]	D_{1-20} † [Å]	r.m.s.d.* [Å]
DG, no metrization	12.8	68.1	48.2	11.7
DG, ordered metr.	0.7	48.2	7.0	3.4
DG, random metr.	0.6	47.8	5.5	3.0
SA	0.0	0.0	28.5	6.7
random ϕ, ψ	0.0	0.0	27.5	6.4

*root mean square deviation from ideal value

†average distance from first to last C^α

#average pairwise root mean square difference

that the missing attractive component in the vdW interaction does *not* lead to overly extended structures.

Further properties of the calculated structures are compiled in Table 1, together with those of the DG structures. All investigated properties of structures generated with SA, starting from a single starting conformation (an extended β -strand), are nearly indistinguishable from those of random structures. The short dynamics run at high temperature together with the employed force field and the variation of the vdW force constant thus ensures very good sampling and seems easily sufficient to remove bias in DG structures. Our findings are in contrast to (2), where a different SA algorithm⁷ was used. One reason for the sampling properties of our SA protocol⁴ is probably the way the vdW weight is varied during the refinement. Due to its very low initial value, the structures can rotate almost freely around rotatable bonds.

A NEW SA PROTOCOL FOR STRUCTURE DETERMINATION

One requirement for employing a calculational strategy like the DG-SA hybrid method is that the general fold of the molecule is correctly determined

by the DG phase. Re-folding on a very large scale cannot be expected during the short dynamics run, and alternate folds as described in (3) will indeed be missed if they are not picked up in the DG phase. For this reason, a new SA protocol was developed with the aim to ensure good global sampling of the conformational space. The protocol is outlined briefly below.

1. generate starting conformation: ϕ and ψ random, $\chi_i = 180^\circ$
2. 15 ps dynamics at 1000° K
NOE pseudo potential with "soft" asymptotic behaviour⁵
switch between harmonic and asymptotic region at constant value of 0.5 Å
slope of the asymptote set to 0.1
weight on the quartic vdW term⁴ set to a very low value (0.001 — 0.01 kcal mol⁻¹ Å⁻⁴) to allow atoms to pass through each other
3. 10 ps of dynamics at 1000° K
tilt asymptote of the NOE potential from 0.1 to 1
increase weight on the vdW interaction to 0.1 kcal mol⁻¹ Å⁻⁴.
4. cooling and minimization as described before⁴.

The protocol relies mainly on the fact that the overall fold of a protein is rather well determined by the NOE distance restraints even if the vdW interaction is very low, and does not attempt to fold the polypeptide chain from an extended strand. The protocol is distributed with the refinement program X-PLOR⁸. One calculation for NP-5 (478 atoms) with data from (9) takes 22 minutes on a Convex C2 computer.

INFLUENCE OF METRIZATION ON DG STRUCTURE

The effects of different ways to generate the distance matrix on the distribution and quality of structures produced by DG were studied on a polyalanine 20mer without any tertiary distance restraints. Three sets of calculations were performed. 10 structures were calculated without metrization¹⁰, which takes the effect of a randomly chosen distance on the remaining distance bounds into account. The results of metrization depends on the order of choosing distances. In the second set of calculations, the distances were simply chosen in sequence, in the third set, in random order.

The results, which are essentially in agreement with those of others^{2,11}, are compiled in Table 1. Structures calculated without metrization show large deviations from input bounds, as evidenced by the large r.m.s. deviation from bond lengths. Metrization improves the quality of the structures

in this respect drastically. The pairwise r.m.s. differences for the DG structures with metrization is too small. This is a consequence of the too tight packing of the DG structures, and indicates that scaling of the coordinates after embedding may be necessary.

Inspection of the structures on a graphics screen revealed that for structures calculated with ordered metrization the N-terminus is packed considerably more tightly than the C-terminus, a consequence of the ordered way in which distances are chosen. This artefact is removed by randomizing the order of choosing distances.

REFERENCES

1. A. T. Brünger, G. M. Clore, A. Gronenborn, and M. Karplus, (1987), *Protein Eng.* **1**, 399-406 (1987).
2. W. J. Metzler, D. R. Hare, and A. Pardi, *Biochemistry* **28**, 7045-052 (1989).
3. R. M. Levi, D. A. Bassolino, D. B. Kitchen, and A. Pardi, *Biochemistry* **28**, 9361-9372 (1989).
4. M. Nilges, G. M. Clore, and A. M. Gronenborn, *FEBS Lett.* **229**, 317-324 (1988).
5. M. Nilges., A. M. Gronenborn, A. T. Brünger, and G. M. Clore, *Protein Eng.* **2**, 27-38 (1988).
6. M. Nilges, G. M. Clore, and A. M. Gronenborn, *FEBS Lett.* **239**, 129-136 (1988).
7. W. Nerdal, D. R. Hare, and B. R. Reed, *J. Mol. Biol.* **201**, 717-739 (1988).
8. A. T. Brünger X-Plor version 2.1, User Manual, Yale University (1990).
9. A. Pardi, D. R. Hare, M. E. Selsted, R. D. Morrison, D. A. Bassolino, and A. C. Bach, *J. Mol. Biol.* **201**, 625-636 (1988).
10. T. F. Havel and K. Wüthrich, *Bull. Math. Biol.* **46**, 673-698 (1984).
11. T. F. Havel, *Biopolymers* **29**, 1565-1585 (1990).