*Proc. Natl. Acad. Sci. USA*
Vol. 83, pp. 3801–3805, June 1986
Biophysics

# Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to crambin

(nuclear magnetic resonance/nuclear Overhauser enhancement spectroscopy/restrained molecular dynamics/protein folding)

AXEL T. BRÜNGER*, G. MARIUS CLORE†, ANGELA M. GRONENBORN†, AND MARTIN KARPLUS*

*Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138; and †Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany

*Contributed by Martin Karplus, February 3, 1986*

ABSTRACT **Model calculations are performed to evaluate the utility of molecular dynamics with NMR interproton distance restraints for determining the three-dimensional structure of proteins. The system used for testing the method is the 1.5-Å resolution crystal structure of crambin (a protein of 46 residues) from which a set of 240 approximate interproton distances of less than 4 Å are derived. The convergence properties of the method are examined by using different dynamics protocols and starting from two initial structures; one is a completely extended β-strand, and the other has residues 7–19 and 23–30 in the form of α-helices (as in the crystal structure) with the remaining residues in the form of extended β-strands. In both cases global and local convergence to the correct final structure is achieved with rms atomic differences between the restrained dynamics structures and the crystal structure of 1.5–2.1 Å for the backbone atoms and 2.1–2.8 Å for all atoms; the averaged structure has backbone and all atom rms deviations of 1.3 and 1.9 Å, respectively. Further, it is shown that a restrained dynamics structure with significantly larger deviations (i.e., 5.7 Å for the backbone atoms) can be characterized as incorrect, independent of a knowledge of the crystal structure.**

With advances in NMR techniques (1, 2), it is now possible to obtain a large number of approximate interproton distances for certain proteins composed of less than 80 residues (3–5). These interproton distance data supply information that can be used in determining the three-dimensional structure of a protein. Because of the limitations on the number and the range (only <5 Å) of the experimentally available distances, these data by themselves are insufficient to obtain an accurate structure. One approach to the problem employs distance geometry algorithms (6) to generate a structure on the basis of the experimental distance estimates and the known covalent bond lengths and angles (7); the method has been tested by applying it to the bovine pancreatic trypsin inhibitor with model data generated from the crystal structure (8). An alternative approach is restrained molecular dynamics in which the interproton distance data are incorporated into the energy function of the system in the form of effective potentials (9–11). To date the restrained molecular dynamics method has been applied to a model of the *lac* repressor (9), to a 17mer peptide (10), and a DNA hexamer (11). To test the utility of restrained molecular dynamics for determining the tertiary structure of a protein, we here use model data derived from the 1.5-Å resolution crystal structure of crambin (12), a protein composed of 46 residues, for which as yet only very limited experimental NMR data are available (13). First, we derive a set of 240 approximate interproton distances (<4 Å) that can be realistically obtained from NMR experiments. We then carry out a set of restrained molecular dynamics calculations starting out with two initial structures: a totally extended β-strand and an extended structure with α-helices at the same positions as in the crystal structure. The average dynamics structures are compared with the crystal structure. Analysis of the results demonstrates the utility of restrained molecular dynamics for protein structure determination.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## METHODS

**Energy Calculations.** All energy minimization and molecular dynamics calculations were carried out using the program CHARMM (14) with an empirical energy function in which hydrogen atoms are treated explicitly. A potential energy term representing the interproton distance restraints was added to the total energy of the system in the form of a skewed biharmonic effective potential given by (10, 11)

$$E_{\mathrm{NOE}} = \begin{cases} c_1(r_{ij} - r_{ij}^{\circ})^2 \text{ , if } r_{ij} > r_{ij}^{\circ} \\ c_2(r_{ij} - r_{ij}^{\circ})^2 \text{ , if } r_{ij} < r_{ij}^{\circ} \end{cases} \quad [1]$$

where $r_{ij}$ and $r_{ij}^{\circ}$ are the calculated and target distances, respectively, and $c_1$ and $c_2$ are constants given by $c_1 = 0.5\ k_B TS\ (\Delta_{ij}^{+})^{-2}$ and $c_2 = 0.5\ k_B TS(\Delta_{ij}^{-})^{-2}$, where $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, $S$ is a scale factor, and $\Delta_{ij}^{+}$ and $\Delta_{ij}^{-}$ are the positive and negative error estimate of $r_{ij}$, respectively. Solvent molecules were not explicitly included in the calculations but the effect of solvent was approximated by multiplying the electrostatic energy term by $(1/r)$ (14). The nonbonded interactions were switched off, using a cubic switching function, between 6.5 and 7.5 Å, with pairs up to 8 Å included in the nonbonded list. Integration of the equations of motion was performed by use of a Verlet integration algorithm (15) with initial velocities assigned from a Maxwellian distribution at the appropriate temperature. The time step of the integrator was 1 fs, and the nonbonded interaction lists were updated every 20 fs.

**Interproton Distances.** In choosing a suitable interproton distance data set for the test calculations care was taken to include only distances that could be obtained experimentally from proton nuclear Overhauser enhancement (NOE) measurements. Since the magnitude of the presteady state NOE between two protons is proportional to $\langle r^{-6} \rangle$, the size of the measured effects falls off rapidly with distance and becomes essentially undetectable for $r > 5$ Å. Consequently, only distances less than 4 Å were considered in constructing the data set. Of these distances a subset was chosen that consisted of distances between protons with reasonably well-separated resonances that could be readily assigned with the available NMR technology. Because accurate quantifi-

Abbreviation: NOE, nuclear Overhauser enhancement.

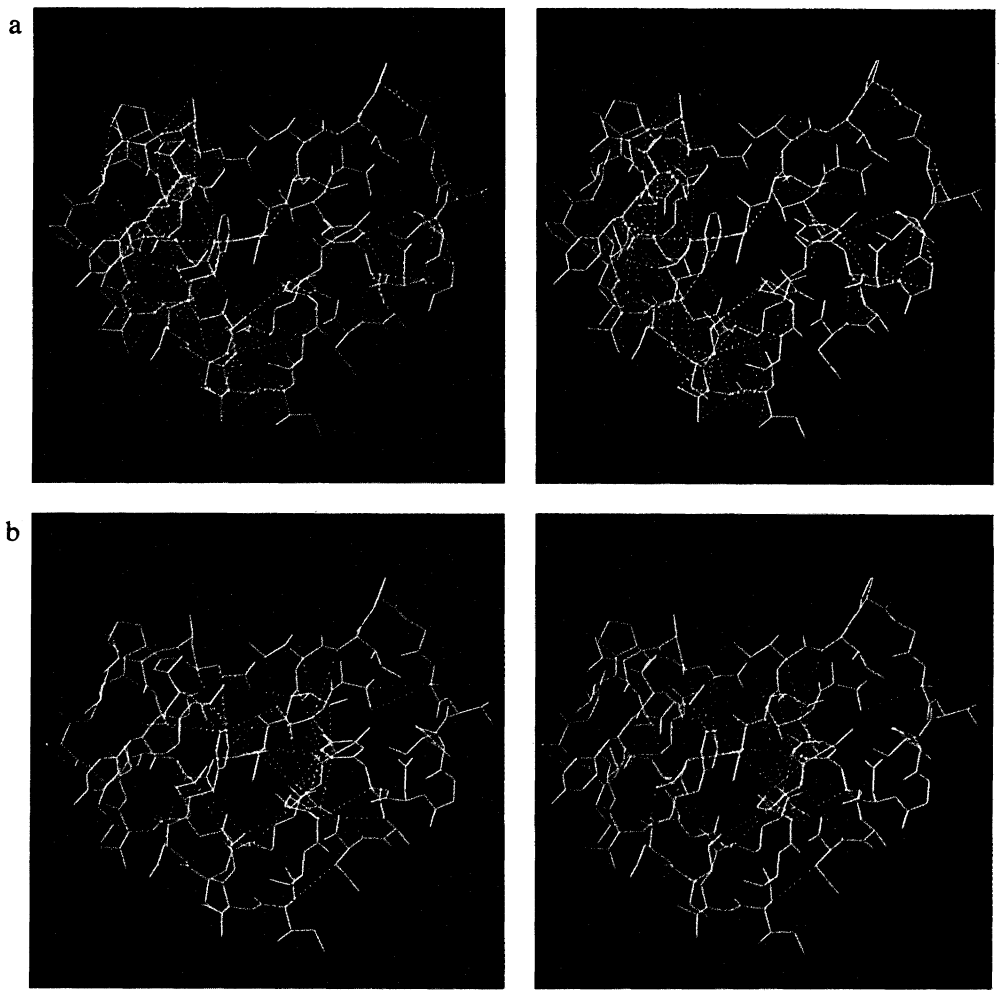


FIG. 1. Crambin distance restraints (stereoview). (*a*) Short-range and (*b*) long-range interproton distance restraints. The restraints are shown as dashed lines in green superimposed on the crystal structure; the (C, N, $C^{\alpha}$) backbone is shown as solid lines in orange and the sidechains are shown in pink.

cation of two-dimensional NOE spectra is difficult, the distances were divided into three classes corresponding to strong, medium, and weak NOEs (10, 16, 17): (*i*) for all $r <$ 2.7 Å, $r_{ij}$ was set to 2.5(0.5/0.5) Å; (*ii*) for all $r$ in the range 2.7 Å $< r <$ 3.4 Å, $r_{ij}$ was set to 3.0(0.5/1.0) Å; and (*iii*) for all $r$ in the range 3.4 Å $< r <$ 4.0 Å, $r_{ij}$ was set to 4.0(1.0/1.0) Å [the numbers in parenthesis correspond to the parameters $(\Delta_{ij}^{+}/\Delta_{ij}^{-})$]. For distances involving methylene protons (for which stereospecific assignments are rarely made) as well as those involving methyl protons, $(\langle r^{-6}\rangle)^{-1/6}$ mean distances were used as this is the quantity that is related to the experimental NOE. Further, for proton pairs involving resonances in potentially crowded regions (e.g., NOEs involving $\beta$, $\gamma$, and methyl protons) only distances corresponding to medium and strong NOEs were included. With these criteria, the following distances were selected (where $i$ and $j$ denote residue numbers): NH($i$)↔NH($j$), $C^{\alpha}$H($i$)↔NH($j$), $C^{\beta}$H($i$)↔NH($j$) (only $r < 3.4$ Å), $C^{\alpha}$H($i$)↔$C^{\alpha}$H($j$) (only $r < 3.4$

Table 1. Protocols for restrained molecular dynamics calculations

| Initial structure* | Ini I | Ini I | Ini I | Ini II |
|---|---|---|---|---|
| Stage 1 variations† | 5 ps, all, $S = 0.01$ | 5 ps, short, $S = 1.0$ | 2 ps, short, $S = 1.0$ | 5 ps, all, $S = 0.01$ |
| | 5 ps, all, $S = 1.0$ | 5 ps, all, $S = 1.0$ | 5 ps, all, $S = 0.01$ | 5 ps, all, $S = 1.0$ |
| | 5 ps, all, $S = 5.0$ | 5 ps, all, $S = 5.0$ | 5 ps, all, $S = 1.0$ | 5 ps, all, $S = 5.0$ |
| | | | 5 ps, all, $S = 5.0$ | |
| Stage 2 | 17 ps, all, $S = 6.0$ | 17 ps, all, $S = 6.0$ | 17 ps, all, $S = 6.0$ | 17 ps, all, $S = 6.0$ |
| Final structure‡ | RDIA | RDIB | RDIC | RDII |

For each calculation are specified the duration of the various phases of restrained molecular dynamics,whether all interproton distances were included (all) or only the short-range distances (short), and the scale factor $S$.

*Each initial structure was energy minimized [500 cycles of conjugent gradient minimization (19) with weak harmonic constraints (20)].

†Initial velocities at 300 K. Each phase was followed by 500 cycles of conjugent gradient energy minimization (19) with weak harmonic constraints (20) with the same $S$ and NOEs as in the current phase.

‡These structures represent the average restrained molecular dynamics structures obtained from trajectories over the last 10 ps of stage 2.

Table 2. rms differences between structures

| | All atoms, Å | | | | | | |
|---|---|---|---|---|---|---|---|
| | Ini I | Ini II | RDIA | RDIB | RDIC | RDII | X-ray |
| Ini I | | 22.9 | 43.3 | 42.6 | 42.5 | 42.5 | 42.6 |
| Ini II | 22.9 | | 24.8 | 23.3 | 23.4 | 23.2 | 23.3 |
| RDIA | 43.3 | 24.0 | | 6.4 | 6.4 | 6.6 | 6.4 |
| RDIB | 42.4 | 22.4 | 5.8 | | 2.6 | 2.6 | 2.1 |
| RDIC | 42.5 | 22.5 | 5.8 | 2.1 | | 2.9 | 2.8 |
| RDII | 42.4 | 22.3 | 6.0 | 1.9 | 2.4 | | 2.1 |
| X-ray | 42.6 | 22.6 | 5.7 | 1.6 | 2.1 | 1.5 | |
| | Backbone atoms, Å | | | | | | |

Table 3. rms difference between target and calculated interproton distances and nonbonded energies

| | rms difference between calculated and target interproton distances, Å | | | | |
|---|---|---|---|---|---|
| | | | Interresidue | | |
| Structure | All | Intra-residue | Short range | Long range | Nonbonded energy, kcal/mol* |
| Ini I | 39.2 | 1.07 | 4.48 | 78.6 | 784 |
| Ini II | 20.9 | 1.11 | 1.84 | 42.1 | 2530 |
| RDIA | 0.91 | 0.42 | 0.79 | 1.28 | −499 |
| RDIB | 0.37 | 0.21 | 0.40 | 0.35 | −848 |
| RDIC | 0.36 | 0.30 | 0.36 | 0.41 | −858 |
| RDII | 0.39 | 0.30 | 0.38 | 0.43 | −818 |
| X-ray | 0.26† | 0.17 | 0.27 | 0.26 | −593 |

Numbers of interproton distances were as follows: all, 240; intraresidue, 25; short-range interresidue, 159; long-range interresidue, 56. Values reported are the rms means of the interproton distance differences.

*The nonbonding energy is the sum of the van der Waals, electrostatic, and hydrogen-bonding energy.

†This value is not zero due to the difference of the actual distances and the set of distances (2.5, 3, and 4 Å) used in the calculations.

Å), $C^{\alpha}H(i) \leftrightarrow C^{\beta}H(j)$ (only $r < 3.4$ Å), distances involving the side chains of aromatic residues with other protons within 4.0 Å, distances involving the valine methyl protons with other protons within 3.4 Å, Glu $C^{\gamma}H(i) \leftrightarrow NH(j)/C^{\alpha}H(j)$ (only $r <$ 2.7 Å), and Val $C^{\gamma 2}H_3$/Ile $C^{\gamma 2}H_3$/Leu $C^{\gamma}H$/Leu $C^{\delta 1}H_3(i) \leftrightarrow$ $NH(j)/C^{\alpha}H(j)$. A total of 240 restraints were obtained; there were 184 short-range distances (i.e., $|i-j| \leq 5$), including 25 intraresidue distances and 56 long-range distances. This set of interproton distance restraints superimposed on the crystal structure of crambin is shown in Fig. 1.

**Calculational Strategy.** Two different initial structures were used (see Fig. 2*a*). The first was a completely extended $\beta$-strand (Ini I). The second was a structure in which residues 7–19 and 23–30 were $\alpha$-helical (as in the crystal structure but with stereotypic geometry), and all other residues were in the form of an extended $\beta$-strand (Ini II). The latter structure was chosen because $\alpha$-helices are easily delineated experimentally by a qualitative interpretation of the NOE data alone (18). For both initial structures the side chains were placed in an extended geometry.

The calculations then proceeded in two stages; the first stage involved the overall secondary and tertiary folding of the molecule and in the second stage the folded structure was refined. Details of the protocols used for the stages as well as the notation for the resulting structures are given in Table 1. The first stage comprised several phases, each consisting of restrained molecular dynamics with the scale factor $S$ being increased as the energy associated with the NOE restraints decreased. Several different first-stage protocols were used with Ini I to explore the dependence of the final structure on the way the restraints were introduced. After each phase, energy minimization was performed to reduce the structural distortions that accompany the increase in kinetic temperature due to the large decrease in the restraints energy. The second stage, in which all structures were treated identically,

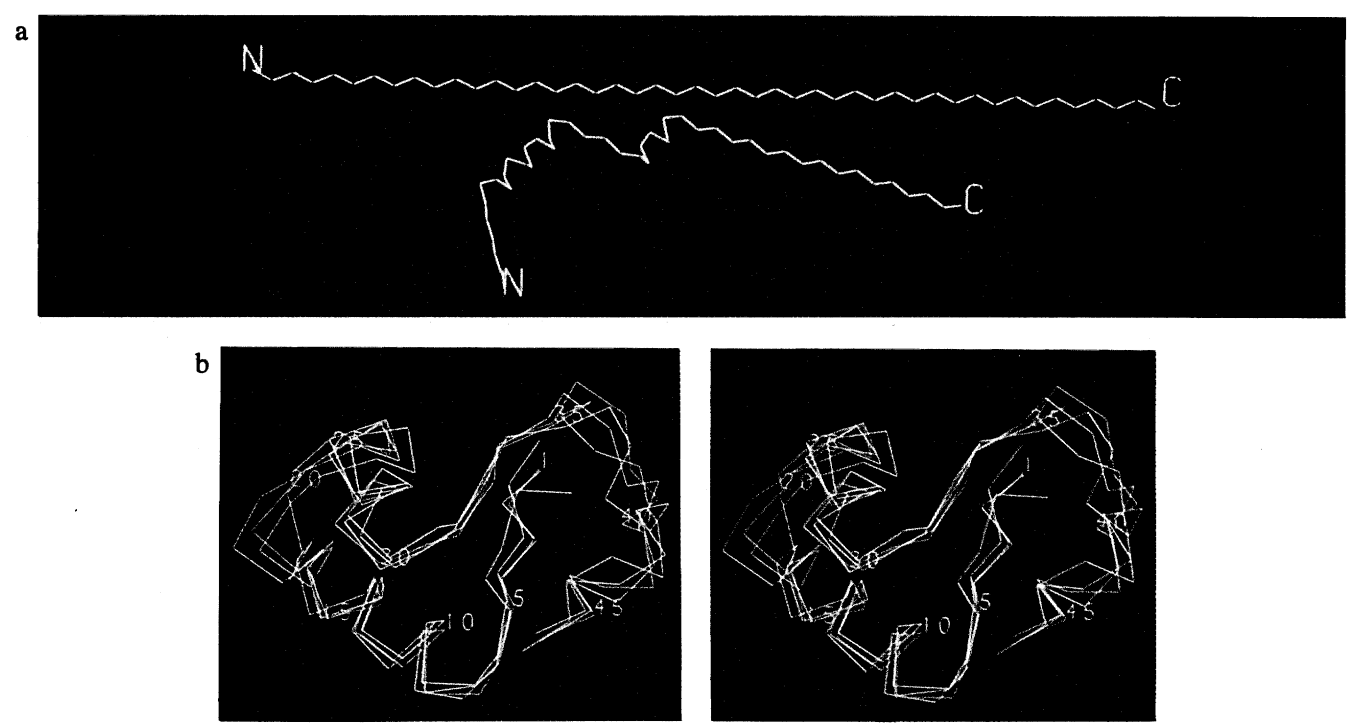


FIG. 2. Initial and final structures. (*a*) Initial structures, Ini I (blue), and Ini II (green). (*b*) Stereoview of the best fit superposition of RDIB (blue), RDIC (yellow), RDII (green) with the x-ray structure (red-orange)—only the $C^{\alpha}$ backbone is shown. For a description of the structures see Table 1.

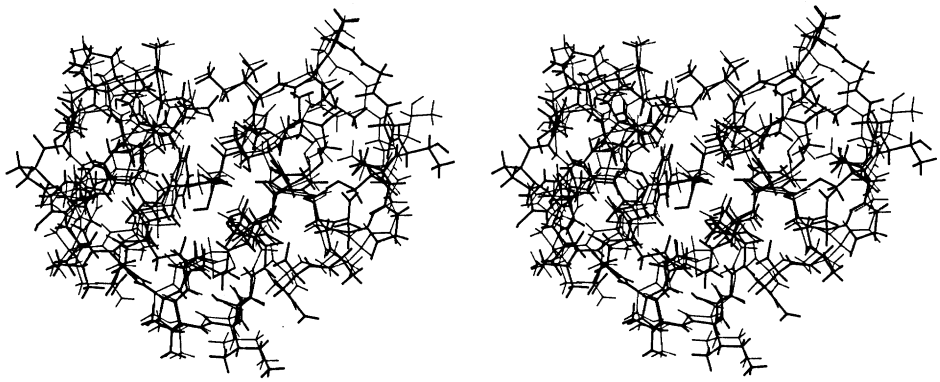

FIG. 3. Stereoview of the best fit superposition of all atoms from the average restrained molecular dynamics structures RDIB (thin lines) with the crystal structure of crambin (thick lines).

comprised a 5-ps equilibration and thermalization (14) followed by a 12-ps period of restrained molecular dynamics at 300 K. The temperature during this second stage remained constant at 300 K, and the structure did not undergo large conformational changes; the final 10 ps were used for computing average structures and their properties.

## RESULTS AND DISCUSSION

The results obtained from the molecular dynamics simulations with interproton distance restraints are given in Tables 2 and 3; the former lists the rms differences of the atomic positions among the various structures and the latter lists the rms deviations of the interproton distances from the values used as restraints.

Three out of four restrained molecular dynamics simulations from the two initial structures (RDIB, RDIC, and RDII) converged to final structures close to the crystal structure both locally and globally (Table 2 and Figs. 2b and 3); RDIA did not converge to the correct structure and is discussed separately. Further, the deviations between the target and calculated interproton distance restraints for the average dynamics structures all lie within the specified error limits (Table 3). This suggests that molecular dynamics can be a useful method for the determination of three-dimensional protein structure when restrained with approximate interproton distance data sets that can be obtained experimentally from NMR measurements. It appears that the NOE data are essential for guiding the dynamics into the correct region of conformational space and that the empirical energy function is important in determining the correct stereochemistry and the nonbonded interactions that govern the size and shape of the molecule. In particular, the radii of gyration of the three average restrained molecular dynamics structures, RDIB (9.49 Å), RDIC (9.28 Å), and RDII (9.61 Å) are close to that of the x-ray structure (9.64 Å). Comparing the three structures RDIB, RDIC, and RDII (Table 2), we see that each is slightly closer to the x-ray structure than it is to any of the other structures. This suggests that each of the converged restrained molecular dynamics simulations explores a somewhat different region in the neighborhood of the x-ray structure, analogous to the behavior of simulations starting from the x-ray structure itself. The minimized structure obtained from the average of RDIB, RDIC, and RDII is significantly closer to the x-ray structure than any of the individual structures; it has an rms difference from the x-ray structure of 1.3 Å and 1.9 Å for the backbone and all atoms, respectively.

Both the secondary and tertiary structure are well reproduced in the average restrained molecular dynamics structures. This is true for RDIB and RDIC derived from the extended chain (Ini I), as well as for RDII, derived from Ini II that already had the correct $\alpha$-helices. The $\beta$-sheet in crambin (residues 1–7 and 31–34) is formed correctly in all cases. Moreover, the secondary structural elements are oriented properly with respect to each other; e.g., the angle and crossover point between the $\beta$-strand from residues 1–7 and that from residues 31–34 are in accord with the x-ray structure, as are the relative positions of the two $\alpha$-helices. Residues 35–45 in the carboxyl-terminal arm show the largest variation among the refined structures (Fig. 2*b*). This is due to the fact that there are few NOE restraints in this region (Fig. 1).

The rms atomic differences involving the side chains are somewhat larger than those of the backbone atoms (Table 2). However, relative to the extended side chain orientation in both initial structures, the positions of the side chains show considerable improvement in the average restrained molecular dynamics structures; the final side chain positions of the RDIB structure are shown in Fig. 3. Given that few of the interproton distance restraints involve side chains, it appears that the positioning of side chains in the protein interior can be obtained, at least in part, from the nonbonded interactions in the energy function (21).

No restraints corresponding to the three disulfide bridges

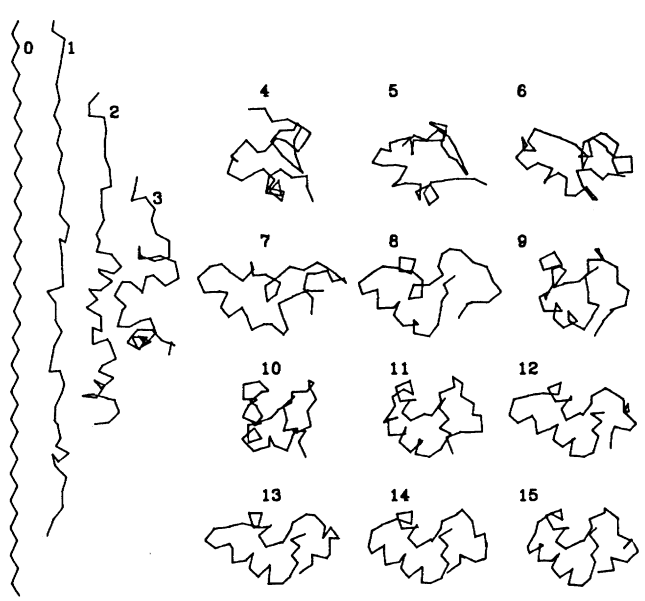


FIG. 4. Folding of crambin backbone by restrained molecular dynamics. Starting with structure Ini I (**0**) snapshots are shown at 1 ps intervals; the protocol is that leading to structure RDIB (**15**); only the $C^{\alpha}$ backbone is shown.

present in crambin were included in the calculation. This demonstrates that the interproton distance restraints together with the empirical energy function are sufficient to define the tertiary structure. Thus, a corresponding methodology should be applicable to proteins without disulfide bridges or proteins in which data for the disulfide bridges are not available. Of the three disulfide bonds present in crambin, only two are found to affect the local structure. The disulfide bonds between residues 3 and 40 and between residues 4 and 34 are significant for the form of the loop extending from residues 34 to 40; i.e., small structural differences appear in this region when S-S distance restraints are introduced into the molecular dynamics simulation (unpublished data). The presence or absence of the disulfide bond between residues 16 and 26 joining the two $\alpha$-helices has no effect on the structure.

The folding process as simulated by the restrained dynamics calculation is very rapid. This is illustrated in Fig. 4 by a series of snapshots taken from the simulations starting from Ini I and resulting in structure RDIB. At the end of the first 2 ps of restrained molecular dynamics, the secondary folding is essentially established while the molecule is still in an extended conformation; at 9 ps, the tertiary structure is approximately correct. Convergence to the final structure is achieved rapidly when the scale factor $S$ is raised to a sufficiently high level to ensure that all the restrained distances lie within their specified error ranges.

A very important element in evaluating the present results is concerned with their uniqueness; i.e., given the NOE interproton distance data, can a protocol be formulated that will yield a structure close to the correct structure or, if it does not, can a poor structure be distinguished from a good one? At present, energy function calculations, by themselves, are not able to do so (22). As a first approach to this question, we have examined two types of protocols starting with the structure Ini I; in one (RDIB and RDIC) only the short-range interresidue interproton distance data are included initially, while in the second (RDIA) all the NOE data are included throughout the simulation, even though a very small scale factor $S$ is used initially. Both RDIB and RDIC fold correctly and satisfy the interproton distance restraints, while RDIA does not (Tables 2 and 3). In structure RDIA residues 1–22 are on the wrong side of the second $\alpha$-helix (residues 23–30) and the $\beta$-strand (residues 31–35). Moreover, the first $\alpha$-helix (residues 7–19) is not present. Apparently residues 1–22 are forced to be too close to residues 23–35 as a result of the long-range restraints, so that there is insufficient room for the helix. If instead the $\alpha$-helices have formed before tertiary folding occurs (RDIB, RDIC, or RDII), satisfactory structures are obtained. That structure RDIA is incorrect is clear from the rms deviation of the interproton distances. In contrast to all the other restrained molecular dynamics structures, the final rms value for the interproton distances of RDIA fall significantly outside the error limit; this is particularly clear for the long- range results. In addition, the nonbonded energy for RDIA is more than 300 kcal/mol above that for the other restrained dynamics structures; i.e., in RDIA the reduction of the restraint energy is accompanied by a large increase in the nonbonded energy term. As a result, RDIA is trapped in an incorrectly folded structure from which it cannot escape. These comparisons suggest that large deviations of interproton distances and high nonbonded energy relative to other structures are indicative of an incorrect structure. Further studies are needed to verify this conclusion, to test other protocols, and to explore more fully the sensitivity of the method to the details of the folding dynamics.

It is of interest to consider whether the present results have any relation to actual protein folding dynamics. That a correctly folded structure is achieved only when the secondary structural elements are at least partly formed before the tertiary structure is introduced may have its parallel in protein folding, as suggested in the diffusion–collision model (23). The time scale of the simulated folding process is much faster than experimental estimates. At least part of the increase in the folding rate results from the formation of stable secondary structural elements and the avoidance of a complex search process by the introduction of NOE distance restraints.

This work was supported by the Max-Planck Gesellschaft (G.M.C. and A.M.G.) and the National Science Foundation (A.T.B. and M.K.). Computations were performed on a CRAY 1 at the Max-Planck Institut für Plasma Physik (Garching, F.R.G.) and at the University of Minnesota Computing Center.

1. Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 2229–2249.
2. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1976) *J. Chem. Phys.* **71,** 4546–4553.
3. Wagner, G. & Wüthrich, K. (1982) *J. Mol. Biol.* **160,** 343–361.
4. Wider, G., Macura, S., Kumar, A., Ernst, R. R. & Wüthrich, K. (1984) *J. Magn. Reson.* **56,** 207–234.
5. Strop, P., Wider, G. & Wüthrich, K. (1983) *J. Mol. Biol.* **166,** 641–667.
6. Crippen, G. M. & Havel, T. F. (1978) *Acta Crystallogr. Sect. A* **34,** 282–284.
7. Havel, T. F. & Wüthrich, K. (1984) *Bull. Math. Biol.* **46,** 673–698.
8. Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol.* **182,** 281–294.
9. Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985) *J. Mol. Biol.* **182,** 179–182.
10. Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1985) *J. Mol. Biol.* **186,** 435–455.
11. Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1986) *J. Mol. Biol.* **188,** in press.
12. Hendrickson, W. A. & Teeter, M. A. (1981) *Nature (London)* **290,** 107–112.
13. Lecomte, J. T. L. & Llinas, M. (1984) *J. Am. Chem. Soc.* **106,** 2741–2748.
14. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4,** 187–217.
15. Verlet, L. (1967) *Phys. Rev.* **159,** 98–105.
16. Braun, W., Wider, G., Lee, K. H. & Wüthrich, K. (1983) *J. Mol. Biol.* **169,** 921–948.
17. Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol.* **182,** 295–315.
18. Wüthrich, K., Billeter, M. & Braun, W. (1984) *J. Mol. Biol.* **180,** 715–740.
19. Powell, M. J. D. (1977) *Math. Program.* **12,** 241–254.
20. Bruccoleri, R. E. & Karplus, M. (1986) *J. Comp. Phys.*, in press.
21. Gelin, B. R. & Karplus, M. (1975) *Biochemistry* **18,** 1256–1268.
22. Novotny, J., Bruccoleri, R. E. & Karplus, M. (1984) *J. Mol. Biol.* **177,** 787–818.
23. Bashford, D., Weaver, D. L. & Karplus, M. (1984) *J. Biomol. Struct. Dyn.* **1,** 1243–1255.