J. Mol. Biol. (1986) **191**, 523–551

Application of Molecular Dynamics with Interproton Distance Restraints to Three-dimensional Protein Structure Determination

A Model Study of Crambin

G. Marius Clore[1], Axel T. Brünger[2], Martin Karplus[2] and Angela M. Gronenborn[1]

[1] *Max-Planck-Institut für Biochemie D-8033 Martinsried bei München, F.R.G.*

[2] *Department of Chemistry, Harvard University 12 Oxford Street, Cambridge, MA 02138, U.S.A.*

(Received 27 February 1986, and in revised form 9 June 1986)

The applicability of restrained molecular dynamics for the determination of three-dimensional protein structures on the basis of short interproton distances (<4 Å) that can be realistically determined from nuclear magnetic resonance measurements in solution is assessed. The model system used is the 1·2 Å resolution crystal structure of the 46 residue protein crambin, from which a set of 240 approximate distance restraints, divided into three ranges ($2{\cdot}5 \pm 0{\cdot}5$, $3{\cdot}0^{+0{\cdot}5}_{-1{\cdot}0}$ and 4 ± 1 Å), is derived. This interproton distance set comprises 159 short-range ($|i-j| \leq 5$) and 56 ($|i-j| > 5$) long-range inter-residue distances and 25 intra-residue distances. Restrained molecular dynamics are carried out using a number of different protocols starting from two initial structures: a completely extended β-strand; and an extended structure with two α-helices in the same positions as in the crystal structure (residues 7 to 19, and 23 to 30) and all other residues in the form of extended β-strands. The root-mean-square (r.m.s.) atomic differences between these two initial structures and the crystal structure are 43 Å and 23 Å, respectively. It is shown that, provided protocols are used that permit the secondary structure elements to form at least partially prior to folding into a tertiary structure, convergence to the correct final structure, both globally and locally, is achieved. The r.m.s. atomic differences between the converged restrained dynamics structures and the crystal structure range from 1·5 to 2·2 Å for the backbone atoms and from 2·0 to 2·8 Å for all atoms. The r.m.s. atomic difference between the X-ray structure and the structure obtained by first averaging the co-ordinates of the converged restrained dynamics structures is even smaller: 1·0 Å for the backbone atoms and 1·6 Å for all atoms. These results provide a measure with which to judge future experimental results on proteins whose crystal structures are unknown. In addition, from an examination of the dynamics trajectories, it is shown that the convergence pathways followed by the various simulations are different.

1. Introduction

The determination of the three-dimensional structure of a protein in solution by nuclear magnetic resonance (n.m.r.)† spectroscopy proceeds in three stages: (1) the sequential assignment of resonances by means of through-bond and through-space connectivities (Wagner & Wüthrich, 1982; Wüthrich *et al.*, 1982; Billeter *et al.*, 1982; Strop *et al.*, 1983; Zuiderweg *et al.*, 1983); (2) the derivation of an approximate set of interproton distances from the nuclear Overhauser enhancement (NOE) data (Wagner & Wüthrich, 1979; Dobson *et al.*, 1982; Clore & Gronenborn, 1985); and (3) the determination of the three-dimensional structure on the basis of these distances. Given the limitations in number, accuracy and range (<5 Å) of the interproton distances, this third step is not

† Abbreviations used: n.m.r., nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement or effect; NOESY, two-dimensional NOE spectroscopy; r.m.s., root mean square; RD, restrained dynamics; FD, free dynamics.

0022-2836/86/190523-29 $03.00/0

straightforward, and it is with this particular problem that this paper is concerned.

To date, two distance geometry approaches have been proposed for determining three-dimensional structures on the basis of interproton distance data. The first method is based on the use of the metric matrix (Crippen & Havel, 1978; Kuntz *et al.*, 1979) whose elements are composed of the interproton distance estimates and distances derived from the known covalent structures (namely, bond lengths and angles) and where the atoms are represented as hard spheres of appropriate van der Waals' radius (Havel & Wüthrich, 1984, 1985). The second and more recently introduced method involves the least-squares minimization of a target function, comprising the interproton distance estimates and soft van der Waals' repulsion terms, in torsion angle space; i.e. bond lengths and angles are kept totally fixed during the minimization and only the torsion angles are treated as variables (Braun & Go, 1985). The results of both methods with model data from the crystal structure of bovine pancreatic trypsin inhibitor are encouraging, since the overall shape, size and folding of the polypeptide chain are reasonably well reproduced (Havel & Wüthrich, 1985; Braun & Go, 1985). However, when interproton distance sets that could be realistically obtained experimentally are used, the structures tend to be slightly expanded relative to the X-ray structure and quite large deformations in the local backbone structures are apparent. One possibility of overcoming this problem is to use a method that not only possesses a large radius of convergence (as do the distance geometry methods) but also introduces energetic considerations during the entire course of the structure determination. Such a method is restrained molecular dynamics. This involves the simultaneous solution of the classical equations of motion for all atoms for a suitable time at an appropriate temperature (McCammon *et al.*, 1974, 1979; Karplus & McCammon, 1983) with the interproton distance incorporated into the total energy function of the system in the form of effective potentials (Clore *et al.*, 1985; Kaptein *et al.*, 1985; Nilsson *et al.*, 1986). In two recent papers we have shown that restrained molecular dynamics can determine with precision the secondary structures of a peptide (Clore *et al.*, 1985) and an oligonucleotide (Nilsson *et al.*, 1986). In this paper we explore the utility of restrained molecular dynamics to determine three-dimensional protein structures; a preliminary report of part of this work has been published (Brünger *et al.*, 1986).

Using the 1·2 Å resolution crystal structure of the small protein crambin (Hendrickson & Teeter, 1981) as a model system, we first derive a set of 240 approximate distance restraints that could be realistically obtained experimentally. A number of restrained molecular dynamics simulations using several different protocols are then carried out starting from two initial structures: a completely extended β-strand, and an extended structure with two α-helices in the same positions as in the crystal structure and all other residues in the form of extended β-strands. With one exception, convergence from both initial structures to structures close to the X-ray structure is achieved. The r.m.s. differences between the final, converged, average-restrained dynamics structures on the one hand and the crystal structure on the other range from 1·5 to 2·2 Å for the backbone atoms and from 2·0 to 2·8 Å for all atoms. When the co-ordinates of the converged dynamics structures are averaged, the r.m.s. difference between the resulting structure and the X-ray structure is even smaller: 1·0 Å for the backbone atoms and 1·6 Å for all atoms. In one case, however, an incorrectly folded structure is obtained. By comparing the properties of the incorrectly folded structure with the correctly folded ones, a set of criteria is derived that enables one to distinguish correctly from incorrectly folded structures. Finally, the restrained dynamics trajectories are analysed and the folding of the polypeptide chain from the initial to the final structures *via* a number of different convergence pathways is directly visualized.

2. Methodology

(a) *Energy calculations*

All energy minimization and molecular dynamics calculations were carried out on a CRAY 1A computer using the program CHARMM (Brooks *et al.*, 1983) optimized for the CRAY (A. T. Brünger, unpublished results). Analysis of structures was carried out on a VAX 11/780, and displaying of molecular dynamics trajectories was carried out using a modified version of the function network of FRODO (Jones, 1978, 1982) interfaced with CHARMM on an Evans & Sutherland PS 330 colour graphics system.

Empirical energy potentials for the energy minimization and molecular dynamics calculations were taken from Brooks *et al.* (1983) and modified in order to treat all hydrogen atoms explicitly (D. States & M. Karplus, unpublished results). A potential energy term representing the interproton distance restraints, $E_{\mathrm{NOE}}(r_{ij})$, was added to the total energy of the system in the form of a quadratic biharmonic potential given by Clore *et al.* (1985):

$$E_{\mathrm{NOE}}(r_{i,j}) = \begin{cases} c_1(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} > r_{ij}^0 \\ c_2(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} < r_{ij}^0, \end{cases} \quad (1)$$

where $r_{i,j}$ and $r^0_{i,j}$ are the calculated and target interproton distances respectively, and c_1 and c_2 are force constants given by:

$$c_1 = \frac{k_{\mathrm{B}}TS}{2(\Delta_{ij}^+)^2}, \qquad c_2 = \frac{k_{\mathrm{B}}TS}{2(\Delta_{ij}^-)^2}, \quad (2)$$

where k_{B} is the Boltzman constant, T the absolute temperature, S a scale factor, and Δ_{ij}^+ and Δ_{ij}^- the positive and negative error estimates, respectively, of r_{ij} Solvent molecules were not included explicitly in the simulations, but the effect of solvent was approximated by multiplying the electrostatic energy term by a $(1/r)$ screening function (Brooks *et al.*, 1983). The non-bonded 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 non-bonded list.

Integration of the equations of motion was performed using a Verlet integrator algorithm (Verlet, 1967) with initial velocities assigned to a Maxwellian distribution at the appropriate temperature. The time step of the integrator was 0·001 ps and the non-bonded interaction lists were updated every 0·02 ps.

(b) *Interproton distance estimates*

Protons were added to the crystal structure of crambin, according to standard stereochemistry using the HBUILD facility (Brünger, unpublished results) in CHARMM. In choosing a suitable set of interproton distances, particular care was taken to select only those distances that could be realistically obtained experimentally by means of NOE measurements. In this respect it is important to bear in mind that the magnitude of the cross-relaxation rate, and hence of the pre-steady state NOE at short times t, is proportional to $\langle r^{-6} \rangle$. As a result, the size of the measured effects falls off rapidly as the mean $(\langle r_{ij}^{-6} \rangle)^{-1/6}$ distance increases and becomes essentially undetectable for $(\langle r_{ij}^{-6} \rangle)^{-1/6} > 5$ Å. For this reason only distances less than 4 Å were considered. From a complete listing of all interproton distances <4 Å, only a subset was chosen. This choice was based on 3 criteria. We included in our data set only distances between protons: (1) whose resonances could be readily assigned using the presently available 2-dimensional n.m.r. technology; (2) whose resonances could be expected to be reasonably well separated in the n.m.r. spectrum; and (3) whose corresponding cross-peaks could be expected to occur in relatively non-crowded regions of the 2-dimensional NOE (NOESY) spectrum. Because accurate quantification of cross-peak intensities in NOESY spectra is difficult, the distances were divided into 3 classes corresponding to strong, medium and weak NOEs (Braun *et al.*, 1983; Williamson *et al.*, 1985; Clore *et al.*, 1985); (1) for all $r_{ij} \leq 2{\cdot}7$ Å, r_{ij} was set to $2{\cdot}5(\pm 0{\cdot}5)$ Å; (2) for all r_{ij} in the range $2{\cdot}7$ Å $< r_{ij} \leq 3{\cdot}4$ Å, r_{ij} was set to $3{\cdot}0(+0{\cdot}5/-1{\cdot}0)$ Å; and for all r_{ij} in the range $3{\cdot}4 < r_{ij} \leq 4{\cdot}0$ Å, r_{ij} was set to $4{\cdot}0(\pm 1{\cdot}0)$ Å. For distances involving either methylene protons (for which stereospecific assignments are rarely made) or methyl protons, single $(\langle r^{-6} \rangle)^{-1/6}$ mean distances were used as this is the quantity that can be related directly to the experimental NOE. Thus, for example, consider the 2 distances $r_{C^{\beta}H1-j}$ and $r_{C^{\beta}H2-j}$ from the C^{β} methylene protons of a given residue to a proton j. Then the single $(\langle r^{-6} \rangle)^{-1/6}$ mean distance included in the restraints list is simply given by:

$$(\langle r^{-6} \rangle)^{-1/6} = \{[(r_{C^{\beta}H2-j})^{-6} + (r_{C^{\beta}H2-j})^{-6}]/2\}^{-1/6}. \quad (3)$$

(Note that forces were obtained for all protons by taking the partial derivatives of eqn (3) with respect to each coordinate.) This quantity is particularly useful as it is heavily weighted towards the distance with the smaller value. This is readily illustrated by the following example. Consider the case where one of the two individual distances has a value of 2 Å; then the maximum value that the single $(\langle r_{i,j}^{-6} \rangle)^{-1/6}$ mean distance can have is 2·24 Å, no matter how large the value of the other individual distance. In the case of tyrosine and phenylalanine rings it is usually the case that only a single averaged resonance is seen for the Hδ1 and Hδ2 protons and for the Hε1 and Hε2 protons, due to rapid ring flipping. Consequently, distances involving the Hδ1/Hδ2 and Hε1/Hε2 protons can also be treated as single $(\langle r^{-6} \rangle)^{-1/6}$ mean distances. In the present case, however, we chose to treat these distances individually in order to ensure for purposes of comparison that the Tyr and Phe rings converged to the same orientation as in the crystal structure; i.e. we wished to avoid obtaining a set of structures in which the orientations of the Tyr and Phe rings differed from one structure to the next by 180° rotations about the C^{β}–C^{γ} bond. In addition, for proton pairs involving resonances in potentially crowded regions (e.g. NOEs involving C^{β}, C^{γ} and methyl protons) only distances corresponding to medium and strong NOEs were included in the restraints list.

A complete listing of all the interproton distance restraints used in the calculations is given in Table 1. This comprises a total of 240 restraints less than 4 Å composed of 25 intra-residue distances, 159 short range (i.e. between $|i-j| \leq 5$ residues) and 56 long range ($|i-j| > 5$) inter-residue distances. This set of distances superimposed on the backbone atoms of the crambin crystal structure is shown in Fig. 1. It should be noted that this set of distances represents a conservative estimate of the number that could actually be obtained experimentally.

(c) *Calculational strategy*

The strategy employed in the calculations was to start from 2 initial structures: (1) a completely extended β-strand known as Ini I and (2) an extended structure known as Ini II, having residues 7 to 19 and 23 to 30 in the form of α-helices as in the crystal structure and the other residues in the form of extended β-strands. It should be noted that an initial structure such as Ini II would be a perfectly reasonable starting point for a structure determination based on NOE distances, as the secondary structure elements can be readily delineated by a qualitative interpretation of the NOE data (Wüthrich *et al.*, 1984). Standard geometry was used to generate the extended β-strand and α-helical elements of the starting structures using ϕ, ψ values of $-113°$, $180°$ and $-57°$, $-47°$, respectively. The side-chains of the initial structures were placed in an extended geometry (see Fig. 2). The atomic r.m.s. differences between the initial structures and the crystal structure are given in Table 2, and the superposition of the initial structures on the crystal structure is shown in Fig. 2. Both initial structures were first subjected to 500 cycles of conjugate gradient (Powell, 1977) energy minimization with no NOE restraints but constrained to the initial structures by weak harmonic constraints at 20 kcal/mol per Å^2 and decreasing to 2 kcal/mol per Å^2 (Bruccoleri & Karplus, 1986).

The calculations then proceeded in two stages: a structure determination stage (stage 1) and a refinement stage (stage 2). The specific protocols, together with the notation of the resulting structures, are given in Table 3. For the structure determination stage, 3 different methods were used. In method A the structures were subjected to a number of phases, each comprising 5 ps of restrained dynamics (initial velocities assigned at 300 K) followed by 500 cycles of restrained energy minimization, the scale factor for the restraints effective potential being increased at each stage as the NOE restraints improved. The purpose of the energy minimization step is to reduce the large bond and angle energies at the end of each dynamics run. This arises as a consequence of the large increase in kinetic energy (and hence temperature), concomitant with the large decrease in the restraints energy during the course of each dynamics run; the total energy for this isolated system remains constant. In order to explore the sensitivity of the method to different

Table 1

Interproton distance restraints (Å) derived from the X-ray structure of crambin and used in the restrained molecular dynamics simulations

A. *Short-range (|i − j| ≤ 5) inter-residue interproton distances*

Sequential	NH(i)–NH($i+1$)	$C^{\alpha}H(i)$–NH($i+1$)	$C^{\beta}H(i)$–NH($i+1$) (only $r \leq 3.4$ Å)
T1, T2		2·5	
T2, C3		2·5	3
C3, C4		2·5	
P5, S6		3	
S6, I7		2·5	2·5
I7, V8	3	4	2·5
V8, A9	3	4	2·5
A9, R10	3	4	3
R10, S11	3	4	3
S11, N12	3	4	3
N12, F13	2·5	4	3
F13, N14	3	4	3
N14, V15	3	4	3
V15, C16	2·5	4	3
C16, R17	3	4	3
R17, L18	2·5	3	
P19, G20		3	
G20, T21	2·5	3	
P22, E23		2·5	
E23, A24	3	4	3
A24, I25	2·5	4	3
I25, C26	2·5	4	3
C26, A27	2·5	4	3
A27, T28	3	4	3
T28, Y29	2·5	3	
Y29, T30	2·5	4	3
T30, G31	2·5	3	
G31, C32	2·5	3	3
C32, I33		2·5	3
I33, I34		2·5	
I34, I35		2·5	
P36, G37		4	
G37, A38		2·5	
A38, T39	2·5	3	
T39, C40		2·5	
P41, G42		2·5	3
G42, D43	3	3	
D43, Y44	2·5	3	
Y44, A45	3	2·5	
A45, N46	2·5	3	

NH(i)–NH(j)

S6, A9	4
S6, R10	4
V15, R17	4
I25, A27	4
A27, Y29	4
T30, C32	4

$C^{\beta}H(i)$–NH(j) (only $r \leq 3.4$ Å)

C4, S6	3
A9, S6	3
P22, A24	3
P22, I25	3
P41, D43	3

$C^{\alpha}H(i)$–$C^{\beta}H(j)$ (only $r \leq 3.4$ Å)

I7, R10	3
V8, S11	3
A9, N12	3
R10, F13	2·5
S11, N14	3
N12, V15	3
N14, R17	3

$C^{\alpha}H(i)$–NH(j)

I7, R10	3
I7, S11	4
V8, S11	4
A9, N12	4
A9, F13	4
R10, N14	4
S11, N14	4
N12, V15	3
N12, C16	4
F13, C16	3
N14, R17	3
V15, R17	4
V15, L18	3
C16, L18	4
R17, G20	3
R17, T21	3

Ala $C^{\beta}H(i)$–X(j) (only $r \leq 3.4$ Å)

A9, C4C^{β}H	3
A9, P5C^{γ}H	3
A9, P5C^{δ}H	3
R17, T21	3
E23, C26	3
A24, A27	3
C26, Y29	2·5
A27, C32	3

Leu $C^{\gamma}H(i)$–X(j)

L18, V15	2·5

Thr $C^{\gamma}H(i)$–X(j) (only $r \leq 3.4$ Å)

T1, T2NH	3
T21, P22C^{δ}H	3
T21, I25C^{β}H	2·5
T21, I25C^{γ}H	3
T30, Y29C^{β}H	3
T30, Y29C^{δ}2H	3

Leu $C^{\delta}H1(i)$–X(j)

L18, V15C^{α}H	3

L18, G20 4
P22, A24 4
E23, C26 3
E23, A27 4
A24, A27 3
I25, A27 4
A27, G31 4
A27, C32 3
T28, T30 4
T28, G31 3
I35, G37 4
G37, T39 4
G42, Y44 4
D43, A45 4
Y44, N46 4

Phe $C^{\delta}1H(i)$–X(j)

F13, N14NH	4
F13, N14C^{α}H	3

Tyr $C^{\delta}1H(i)$–X(j)

Y44, P41C^{γ}H	2·5
Y44, D43NH	4

Tyr $C^{\delta}2H(i)$–X(j)

Y29, C26C^{α}H	4

Val $C^{\gamma}H1(i)$–X(j)

V8, A9NH	4

Ile $C^{\gamma}_{M}H(i)$–X(j)

I34, I35NH	3

B. *Intra-residue interproton distances*

Val	V8	V15
$C^{\alpha}H$–$C^{\gamma}H1$	3	3
$C^{\alpha}H$–$C^{\gamma}H2$	2·5	2·5
NH–$C^{\gamma}H2$	3	3

Arg	R10	R17
$C^{\beta}H$–$N^{\epsilon}H$	2·5	3
$C^{\gamma}H$–$N^{\epsilon}H$	2·5	2·5

Thr: T30	
NH–$C^{\gamma}H$	3

Glu: E23	
$C^{\alpha}H$–$C^{\gamma}H$	2·5
NH–$C^{\gamma}H$	2·5

Ile 17, I25, I33, I34, I35 (only $r \leq 3.4$ Å)

$C^{\alpha}H$–$C^{\gamma}_{M}H$	3

Leu: L18	
$C^{\alpha}H$–$C^{\gamma}H$	3
NH–$C^{\gamma}H$	2·5

Phe: F13	
$C^{\beta}H$–$C^{\delta}1H$	2·5
$C^{\beta}H$–$C^{\delta}2H$	2·5

Tyr: Y29	
$C^{\alpha}H$–$C^{\delta}1H$	2·5

Tyr: Y44	
NH–$C^{\delta}1H$	3

C. *Long-range inter-residue interproton distances*

NH(i)–NH(j)

C3, I33	3
C4, N46	3

$C^{\beta}H(i)$–NH(j)

C4, R10	3

$C^{\alpha}H(i)$–$C^{\beta}H(j)$ (only $r \leq 3.4$ Å)

C4, C32	3
I7, N46	3
A9, T30	3
F13, C26	3
I33, A27	3
A38, T1	3

Phe $C^{\delta}2H(i)$–X(j)

F13, C26C^{β}H	3
F13, A27C^{α}H	3
F13, T30C^{γ}H	4
F13, C32NH	4
F13, C32C^{β}H	3

Phe $C^{\epsilon}1H(i)$–X(j)

F13, T2C^{γ}H	4
F13, 23C^{γ}H	4

Arg $N^{\epsilon}H(i)$–X(j)

R10, C4C^{β}H	3

$C^{\alpha}H(i)$–NH(j)

T2, I35	4
C4, I33	3
P5, A45	4
P5, N46	4
I34, C3	4

$C^{\alpha}H(i)$–$C^{\alpha}H(j)$

T2, I34	2·5
C4, C32	2·5
P5, Y44	2·5

Phe $C^{\epsilon}2H(i)$–X(j)

F13, E23C^{α}H	4
F13, C26C^{β}H	3
F13, A27NH	3
F13, A27C^{α}H	3
F13, T30C^{γ}H	4
F13, C32C^{β}H	4

Tyr $C^{\delta}2H(i)$–X(j)

Y29, C16C^{β}H	3
Y44, C4C^{α}H	3
Y44, P5C^{α}H	4
Y44, P5C^{δ}H	2·5
Y44, C32C^{α}H	3
Y44, I33NH	4
Y44, I33C^{γ}H	3

Ala $C^{\beta}H(i)$–X(j) (only $r \leq 3.4$ Å)

A9, T30C^{β}H	3
A27, F13C^{ϵ}2H	4
A27, I34C^{γ}H	3

Thr $C^{\gamma}H(i)$–X(j) (only $r \leq 3.4$ Å)

T1, A38C^{α}H	3
T2, F13C^{ζ}H	2·5
T30, N12C^{β}H	3
T30, F13NH	3
T30, F13C^{α}H	2·5
T30, F13C^{β}H	3

Phe $C^{\zeta}H(i)$–X(j)

F13, T2C^{β}H	4
F13, E23C^{α}H	3
F13, E23C^{β}H	3
F13, E23C^{γ}H	2·5
F13, C26C^{β}H	4

Tyr $C^{\epsilon}2H(i)$–X(j)

Y44, P5C^{β}H	3
Y44, P5C^{δ}H	3
Y44, C32C^{α}H	4
Y44, I33C^{γ}_{M}H	3

Note that for all methylene and methyl protons a single $(\langle r^{-6} \rangle)^{-1/6}$ mean distance is used (see the text). The error limits on the distances are as follows: ±0·5 ångström unit for $r_{ij} = 2.5$ Å, +0·5/−1 Å for $r_{ij} = 3$ Å, and ±1 ångström unit for $r_{ij} = 4$ Å. The amino acids are identified by the one-letter code. $C^{\gamma}_{M}H$ refers to the C^{γ} methyl group of isoleucine.

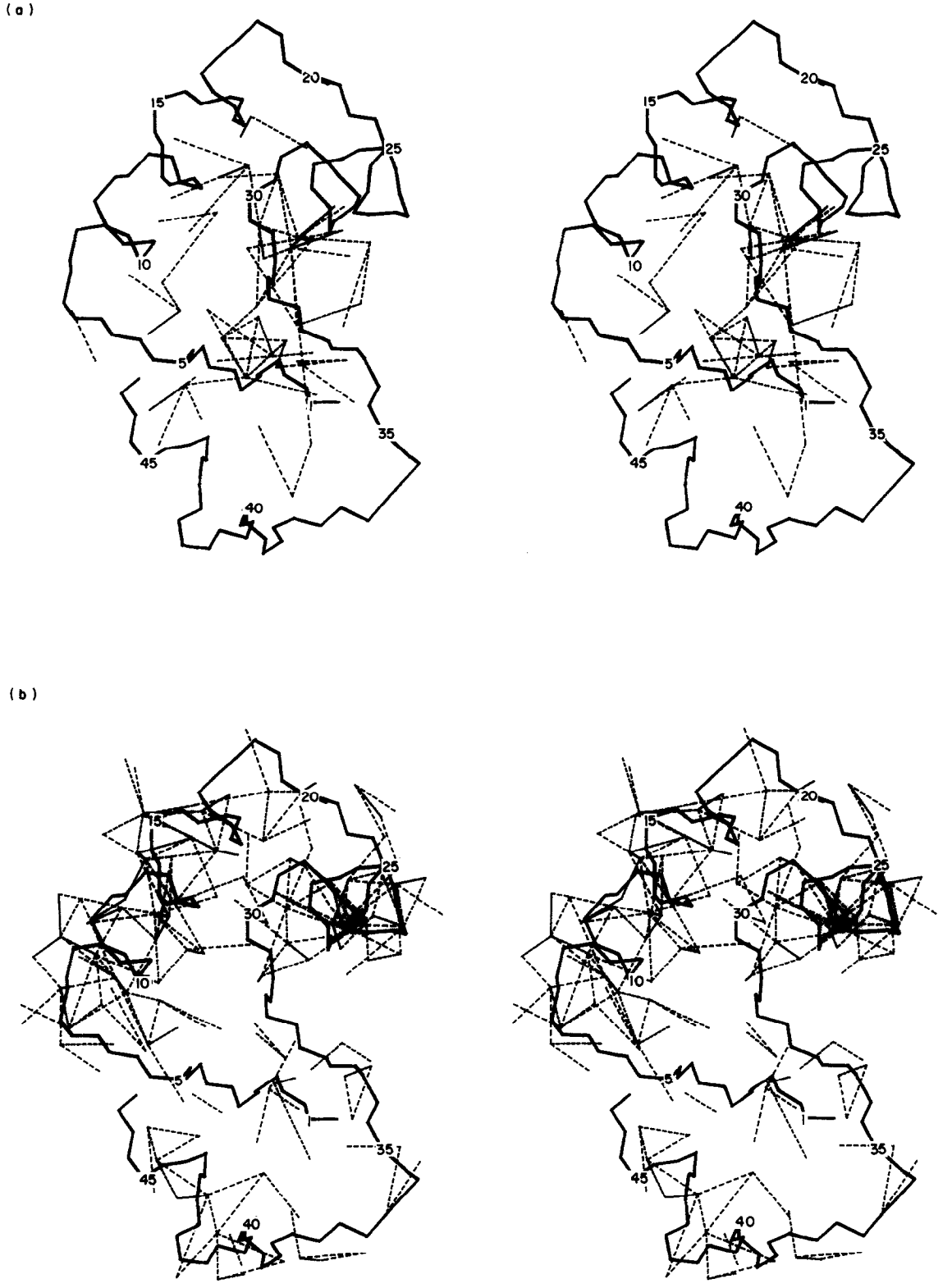


Figure 1. Stereo views of the long (a) and short (b) interproton distance restraints shown as broken lines on a framework comprising the backbone (N, C^{α}, C) atoms of the X-ray structure of crambin.

folding pathways as well as to avoid problems associated with excessive heating, 2 other methods were implemented. In method B, the structures were subjected to 21 ps of restrained dynamics in which, every 0·6 ps, the velocities were reassigned at 300 K and the scale factor for the NOE restraints was doubled up to a particular maximum value. In method C, the structures were subjected to 18 ps of restrained dynamics in which the scale factor for the NOE restraints was also doubled every 0·6 ps up to a particular maximum value and the velocities, rather than being reassigned every 0·6 ps, were simply scaled by a factor of 0·75 when the temperature exceeded 500 K.

It is important to stress that *no* additional restraints

Table 2
Atomic r.m.s. differences (Å) between the initial, the X-ray and the average free dynamics structures

	r.m.s. difference (Å) for all atoms				
	Ini I	Ini II	FDA	FDB	X-ray
Ini I		22·9	42·7	42·7	42·6
Ini II	22·9		23·5	23·5	23·3
FDA	42·7	22·9		1·5	1·6
FDB	42·7	22·8	1·1		1·4
X-ray	42·6	22·6	1·2	1·0	

r.m.s. difference (Å) for backbone atoms (N, C^{α}, C, O).

corresponding to the 3 disulphide bridges present in crambin were included in the structure determination stage and that the full van der Waals' radii were used for the SH groups of the 6 Cys residues. The reason for this is that we wanted to establish how well the 3-dimensional structure could be determined from the interproton distance restraints alone, particularly as not all proteins possess disulphide bridges and, in a real case, information on the exact location of the disulphide bridges may not be available.

The structure refinement stage involved 2 phases, each comprising a period of 5 ps thermalization (reassigning the velocities at 300 K every 0·2 ps) followed by 12 ps of restrained dynamics at 300 K. In the first phase, only the interproton distance restraints were included in the restraints energy. In the second phase, 3 S–S distance restraints corresponding to the 3 disulphide bonds were added to the restraints energy and the van der Waals' radii for the SH groups were replaced by the van der Waals' radii of the sulphur atoms alone. The average restrained dynamics structures were then obtained by averaging the co-ordinates of the trajectories for the last 10 ps of the 2 refinement phases. The scale factor S in eqn (2) was set to 6 for both phases such that error estimates of 0·5 Å and 1 Å in the interproton distance correspond to force constants of 7·15 and 1·79 kcal/mol (1 kcal = 4·18 kJ), respectively.

The details of the various dynamics run together with the notation of the resulting structures are given in Table 3.

In addition, 2 free dynamics simulations (i.e. with no interproton distance restraints and the scale factor S in eqn (2) set to 0) were carried out starting from the crystal structure. These give a measure of the conformational space included in a normal dynamics simulation for comparison with the deviations from the crystal structure for the various restrained dynamics structures. Each comprised 1 ps of equilibration, during which time the temperature was increased from 200 K to 300 K in steps of 10 K every 0·1 ps, 2 ps of thermalization in which the velocities were reassigned every 0·2 ps at 300 K, and 12 ps of dynamics at 300 K without adjusting the temperature of the system. In the first free dynamics simulation, the disulphide bonds were replaced by reduced SH groups and resulted in structure FDA; in the second, resulting in structure FDB, the disulphide bonds were retained. Both FDA and FDB are the average free dynamics structures obtained by averaging the trajectory over the last 10 ps of each dynamics run. It should be noted that the purpose of these free dynamics simulations was simply to see which regions deviate most from the X-ray structure. For this purpose, the usual long equilibration and thermalization periods (10 to 15 ps) required to stabilize the system fully were not needed.

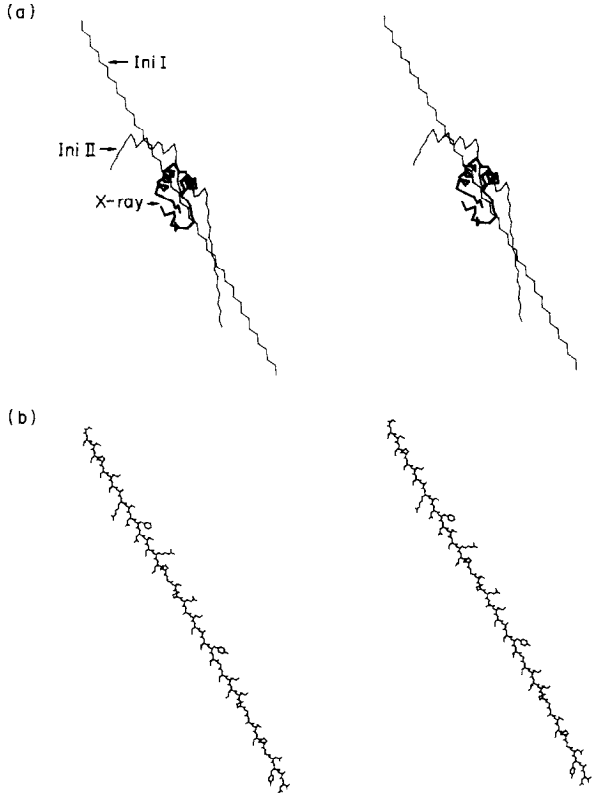


Figure 2. (a) Superposition of the 2 initial structures (Ini I, Ini II) and the X-ray structure. (b) View of the initial structure Ini I with side-chains.

3. Results and Discussion

(a) *Convergence and uniqueness*

In considering the problem of three-dimensional structure determination on the basis of the short (i.e. ≤ 4 Å) interproton distances that can be measured from NOE data, a very important element concerns the question of whether these distances, taken as a whole, possess sufficient information content to define "uniquely" the correct three-dimensional structure. That is to say, is it possible to obtain a set of structures close to the correct structure, as defined by the X-ray structure and, if not, can one discriminate unambiguously between correctly and incorrectly folded structures? To address this problem as well as to assess the sensitivity of our method to particular conditions, we carried out a number of calculations using six different protocols (see Table 3).

With the exception of the restrained dynamics run resulting in structure RDIA (see Fig. 3), all the restrained molecular dynamics simulations from both initial structures converge to structures that satisfy the interproton distance restraints within the errors specified (r.m.s. difference of the interproton distances in the range 0·3 to 0·4 Å) and are very close to the X-ray structure both locally and globally (r.m.s. atomic difference in the range 1·5 to 2·2 Å for the backbone atoms and 2·0 to 2·8 Å

Table 3

Protocols of the restrained dynamics simulations

Restrained dynamics simulation	IA	IB	IC	II	ID	IE
Initial structure	Ini I	Ini I	Ini I	Ini II	Ini I	Ini I
Stage 1: structure determination† (no S–S restraints)	Method A‡				Method B§	Method C‖, ¶
	5 ps, all, $S = 0{\cdot}01$ 5 ps, all, $S = 0{\cdot}2$ 5 ps, all, $S = 5$	5 ps, short, $S = 1$ 5 ps, all, $S = 0{\cdot}05$ 5 ps, all, $S = 5$	2 ps, short, $S = 1$ 5 ps, all, $S = 0{\cdot}01$ 5 ps, all, $S = 1$ 5 ps, all, $S = 5$	5 ps, all, $S = 0{\cdot}01$ 5 ps, all, $S = 1$ 5 ps, all, $S = 5$	Velocities reassigned every 0·6 ps at 300 K 0–9 ps: short, $S = 0{\cdot}02 \rightarrow 2{\cdot}5$ 9–18 ps: all, $S = 0{\cdot}002 \rightarrow 10$	Velocities scaled by a factor of 0·75 when $T > 500$ K 0–6 ps: short, $S = 0{\cdot}02 \rightarrow 2{\cdot}5$ 6–18 ps: all, $S = 0{\cdot}002 \rightarrow 10$
Stage 2: refinement ($S = 6$)						
Phase 1 structures (no S–S restraints; 17 ps)	RDIA	RDIB	RDIC	RDII	RDID	RDIE
Phase 2 structures (with S–S restraints; 17 ps)	—	RDIB′	RDIC′	RDII′	RDID′	RDIE′

† The duration of each phase, the restraints included and the scale factor S in eqn (2) are given for each restrained dynamics simulation. "All" signifies that all interproton distance restraints are included in the restraints energy, and "short" that only the short-range restraints ($|i-j| \leq 5$) are included.

‡ The initial velocities are assigned at 300 K. Each phase is followed by 500 cycles of energy minimization with weak harmonic constraints and with the same restraints and S value as in the current phase.

§ The scale factor S is doubled every 0·6 ps up to the specified value.

‖ The initial velocity is assigned at 300 K.

¶ Each phase of the refinement stage consists of a 5 ps thermalization period in which the velocities are reassigned every 0·2 ps at 300 K followed by 12 ps of restrained dynamics. The average restrained dynamics structures are the averages obtained from the trajectories over the last 10 ps of each restrained dynamics run. In phase 1, only the interproton distance restraints are included in the restraints energy; in phase 2, three additional S–S distance restraints corresponding to the disulphide bonds between residues 3 and 40, 4 and 32, and 16 and 26, are also included in the restraints energy. These three S–S distances are set to 2·0 Å with error limits of $\pm 0{\cdot}5$ Å.

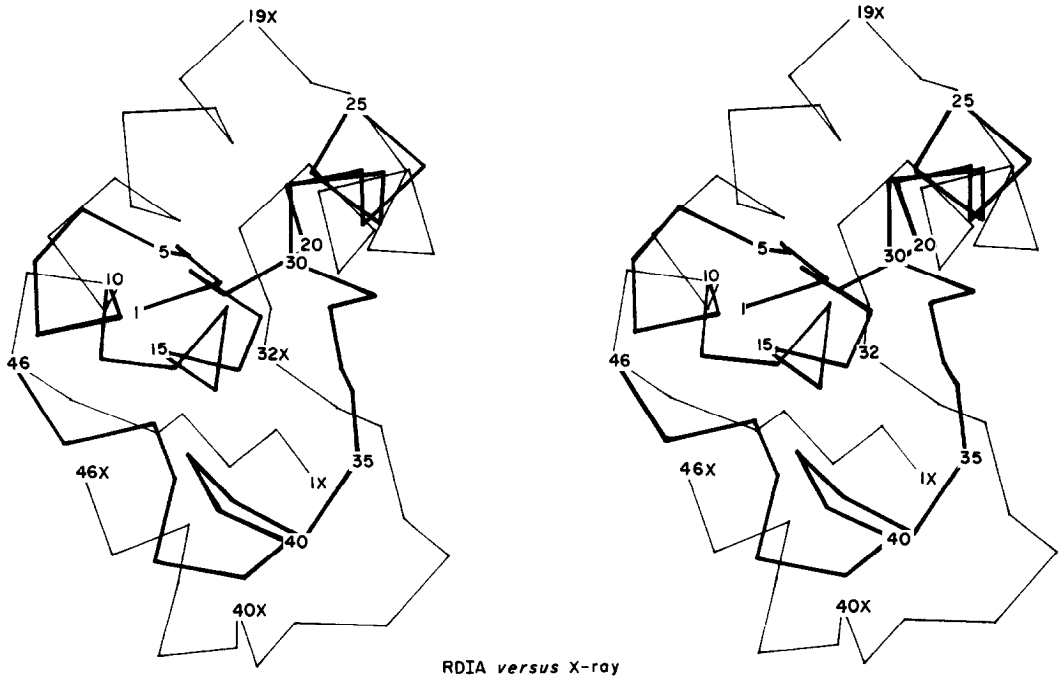
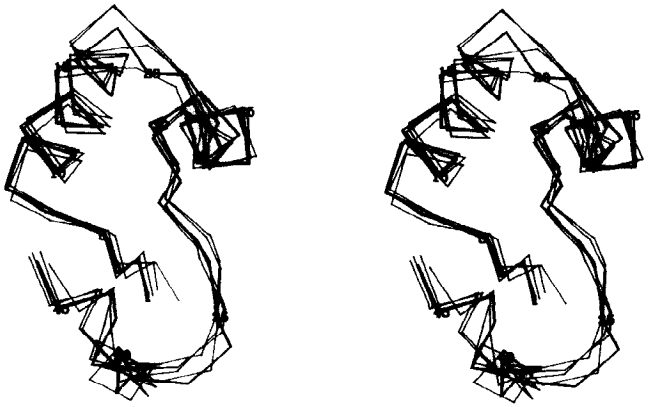


Figure 3. Superposition of the incorrectly folded restrained dynamics structure RDIA resulting from the restrained dynamics simulation in which all restraints were applied simultaneously from the start (thick line) with the X-ray structure (thin line). Only the C^{α} atoms are shown. The C^{α} atoms of RDIA are labelled by numbers only, while those of the X-ray structure are labelled by numbers followed by the letter X.

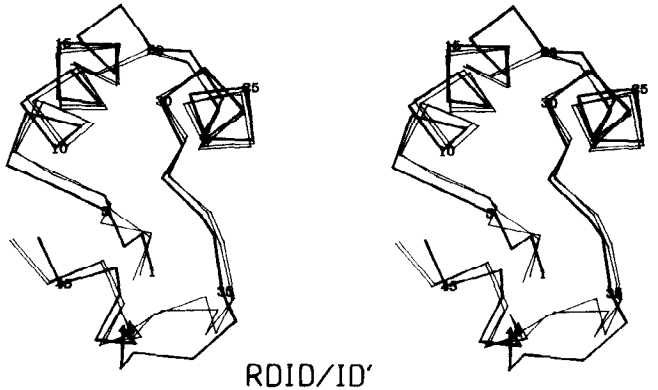
for all atoms). This can be assessed from the stereoviews of the superpositions of the restrained dynamics structures with the X-ray structures shown in Figures 4 and 5, from the atomic r.m.s. difference plots as a function of residue number shown in Figure 6 and from the data in Tables 4 to 7. Tables 4 and 5 list the atomic r.m.s. differences between the restrained dynamics structures on the one hand and the crystal and free dynamics structures on the other, and between the restrained dynamics structures themselves, respectively. Table 6 shows the r.m.s. differences of the interproton distances and the radii of gyration for all the structures, and Table 7 gives the number of restraints violations for the restrained dynamics structures. In addition, the energies of all the structures are given in Table 8.

From the data presented, it is clear the convergence can be achieved using a number of different protocols. In the case of the simulations starting out from the completely extended β-strand (Ini I), however, convergence only occurs if the α-helical elements are at least partially delineated prior to the folding process. In other words, it is desirable to apply the short-range restraints initially followed by the later inclusion of the long-range restraints. If all restraints are applied simultaneously from the start, partially incorrect folding can occur, as found in structure RDIA (Fig. 3). In this structure residues 15 to 22 are apparently forced to lie too close to the second α-helix (residues 23 to 30) and adjacent β-strand (residues 30 to 35) as a result of the application of the long-range restraints, thereby preventing the formation of the first α-helix (residues 7 to 19) and the correct placement of residues 7 to 22 relative to the other residues. Thus, residues 7 to 22 lie on the wrong side of the second α-helix. Interestingly, this incorrect folding could be corrected partially by simply manipulating some of the ϕ, ψ torsion angles of residues 21 to 24. In the light of this result it is therefore particularly important to establish criteria that enable one to distinguish correctly from incorrectly folded structures. Such criteria are the restraints statistics and the non-bonded energy. Thus, in the case of RDIA the restraints energy (Table 8) is an order of magnitude larger and the r.m.s. difference of the interproton distances (Table 6) significantly higher, particularly for the long-range restraints, than the corresponding values for the other restrained dynamics structures. In addition, the number of restraints violations (Table 7) for RDIA is large, comprising 20% of the total number of restraints compared to 1% or less for the other restrained dynamics structures. Equally significant is the finding that the value for the total non-bonded energy (i.e. sum of van der Waals', electrostatic and hydrogen bonding terms) is $\approx$300 kcal/mol larger than for the other restrained dynamics structures (see Table 8). This

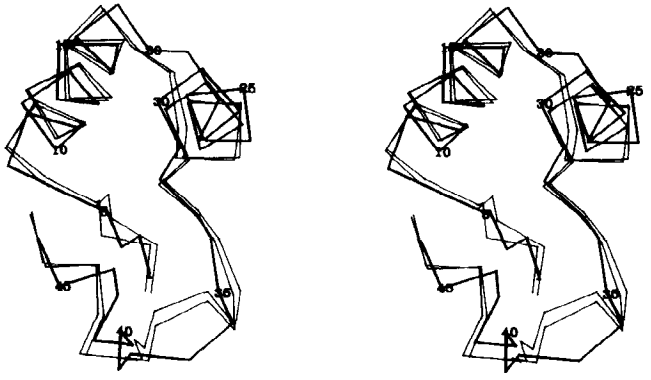
Figure 4. Best-fit superpositions of the C^{α} atoms of groups of average free and restrained dynamics structures (thin lines) on the X-ray structure (thick lines). In each superposition, only the dynamics structures are labelled.



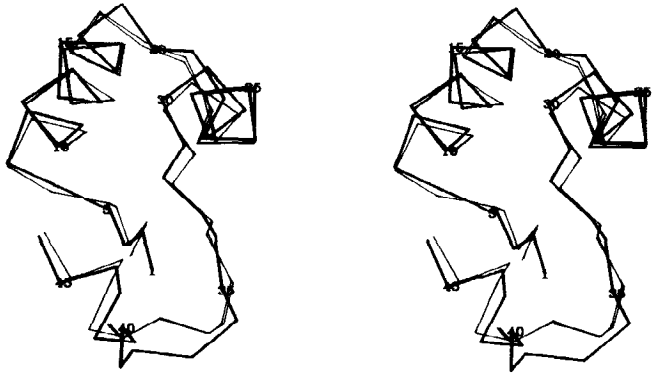
RDII/II', RDIB/IB'



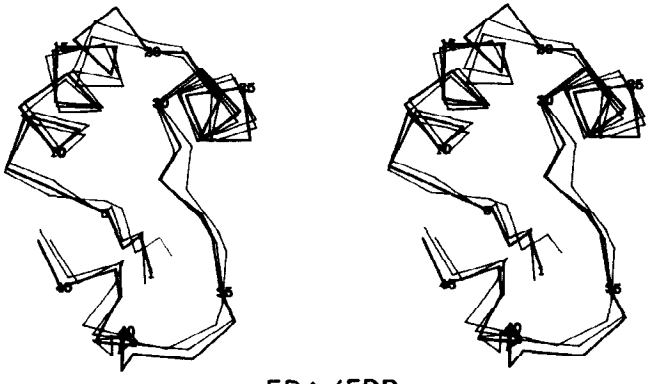
RDID/ID'



RDIC/IC'



RDIE/IE'



FDA/FDB

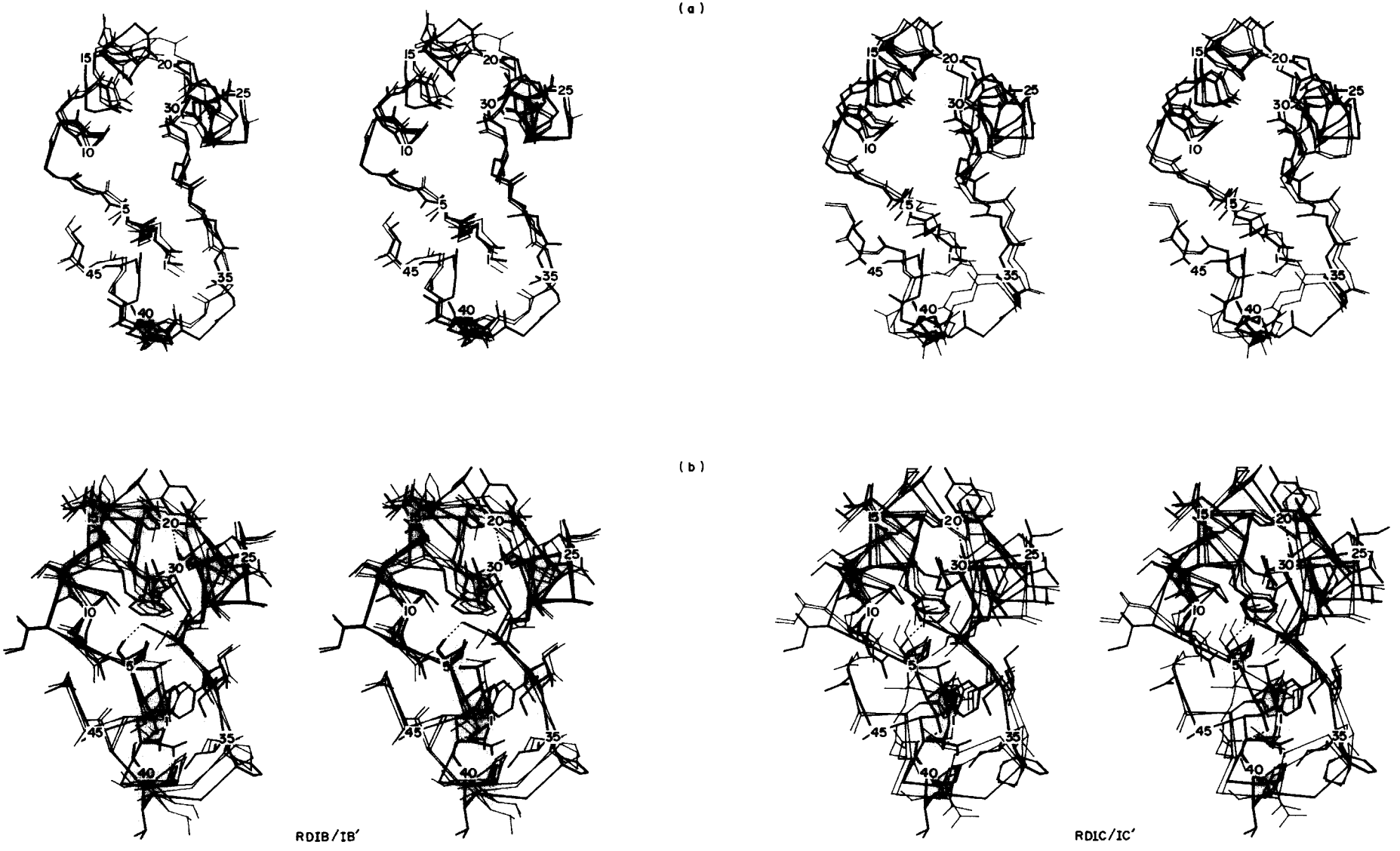


Fig. 5.

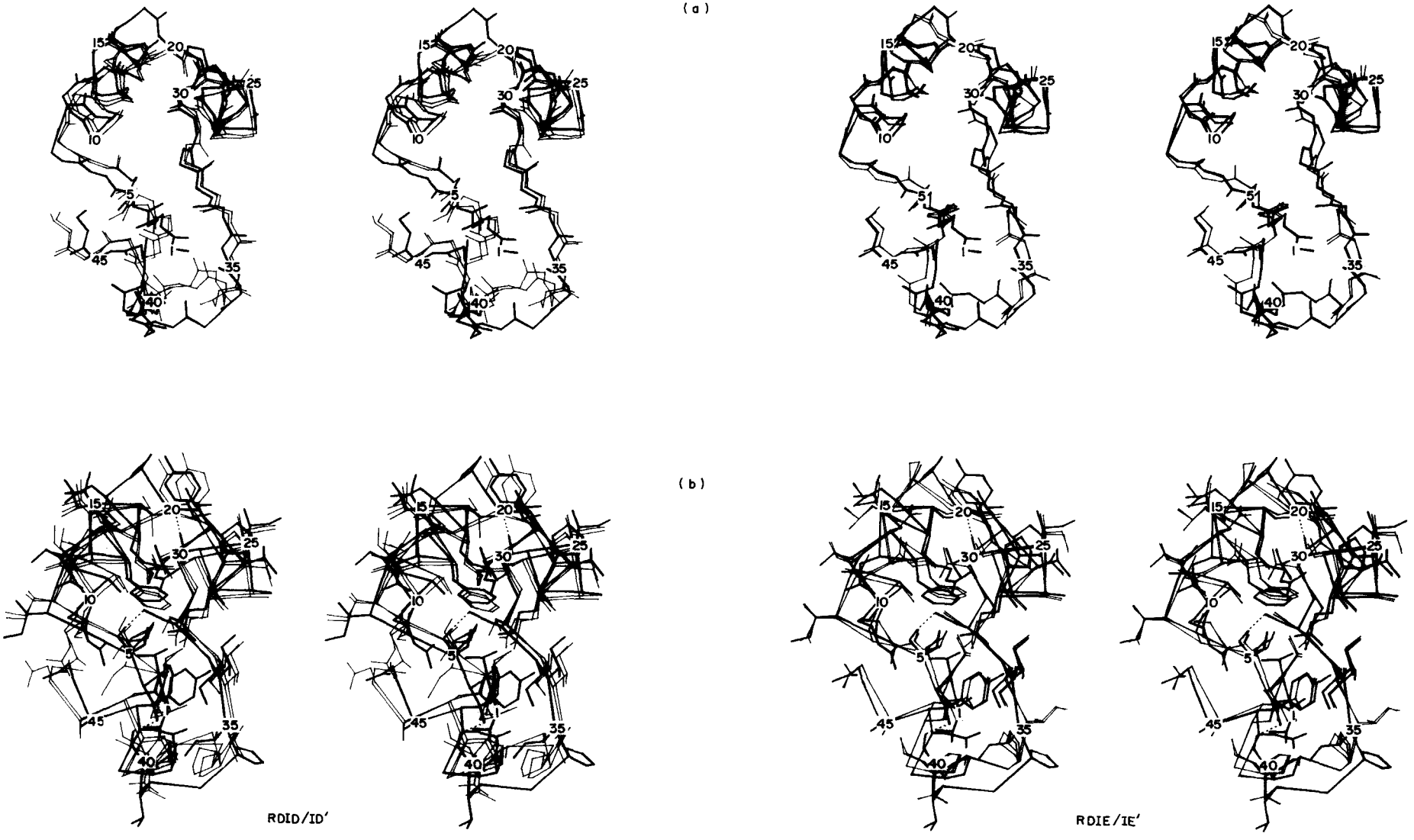


Fig. 5, *cont.*

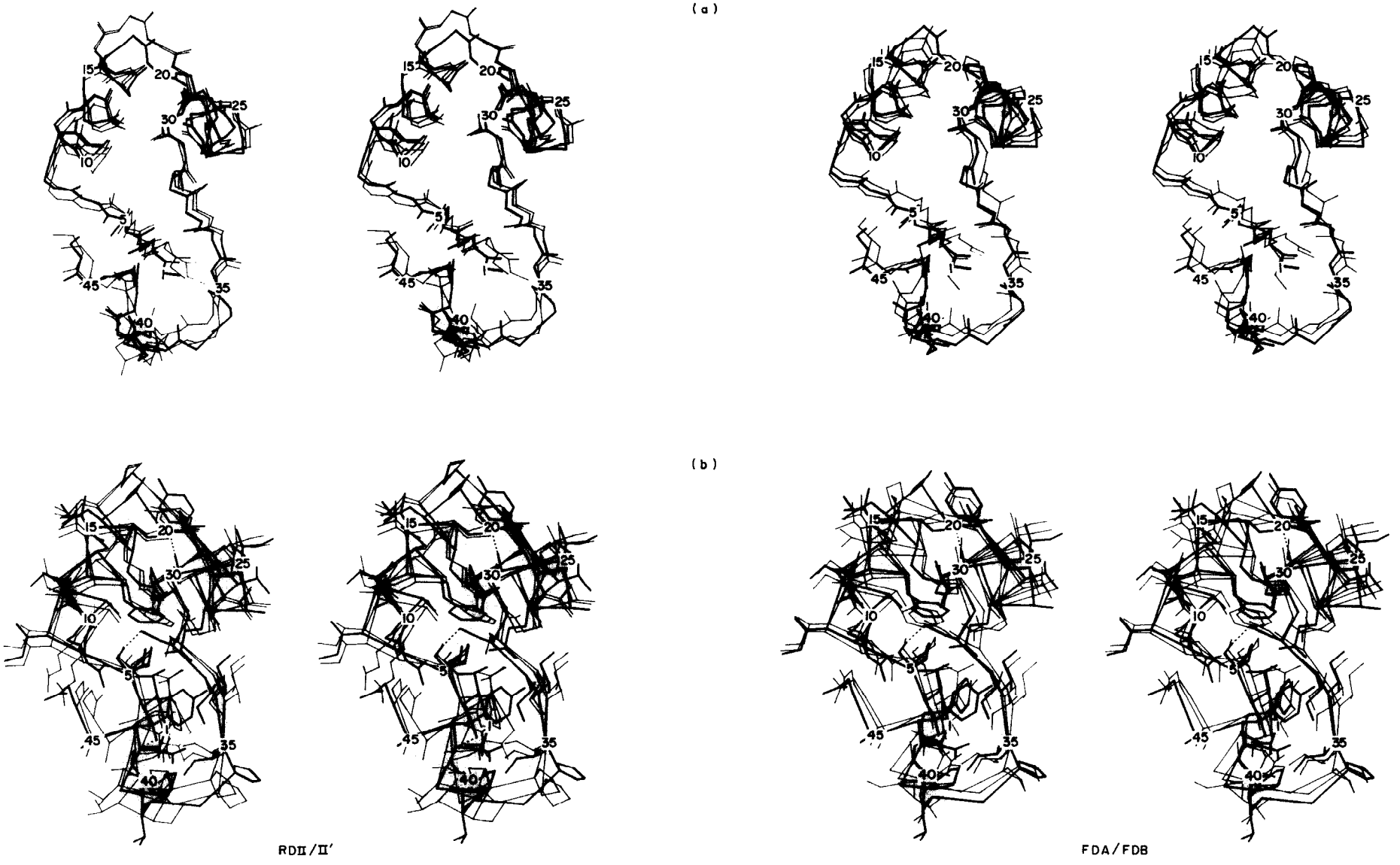


Fig. 5, *cont.*

Table 4

Atomic r.m.s. differences (Å) between the average restrained dynamics structures on the one hand and the average free dynamics and X-ray structures on the other

	r.m.s. difference (Å)					
	All atoms			Backbone atoms (N, C^{α}, C, O)		
	FDA	FDB	X-ray	FDA	FDB	X-ray
RDIA	6·1	6·0	6·4	5·5	5·4	5·7
RDIB	2·4	2·3	2·1	1·9	1·8	1·6
RDIB′	2·3	2·3	2·0	1·9	1·9	1·6
RDIC	2·1	2·8	2·8	2·6	2·1	2·2
RDIC′	2·8	2·4	2·5	2·5	2·0	2·1
RDID	2·7	2·4	2·5	2·2	1·6	1·9
RDID′	2·7	2·4	2·6	2·1	1·7	1·9
RDIE	2·4	2·3	2·2	1·9	1·6	1·6
RDIE′	2·4	2·3	2·2	1·8	1·6	1·6
RDII	2·4	2·4	2·1	1·7	1·8	1·5
RDII′	2·2	2·4	2·0	1·5	1·8	1·5
RDave†	2·4	2·2	1·9	1·6	1·4	1·2

† The structure RDave is the structure obtained by first averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID and RDIE and RDII, and then subjecting the resulting average structure to restrained energy minimization (see the text).

arises as a consequence of the inclusion of the restraints energy in the total energy function. Thus, any decrease in the restraints energy is accompanied by a large increase in the non-bonding energy, with the result that the RDIA dynamics simulation is trapped in an incorrectly folded structure from which it cannot escape. It is important to stress, however, that in the absence of the restraints energy term, energy function calculations cannot by themselves distinguish a correctly folded structure from an incorrectly folded one (Novotny *et al.*, 1984). Indeed, when a further 500 steps of conjugate gradient energy minimization carried out on RDIA without the restraint energy (i.e. with S in eqn (2) set to 0), the non-bonded energy improves by ~300 kcal/mol, although the accompanying atomic r.m.s. shift is small (<1 Å).

Recently, Eisenberg & McLachlan (1986) proposed a method for calculating the free energy of solvation of protein structures from their atomic co-ordinates and used this to estimate the relative stabilities of different protein conformations. Using the same natural and incorrectly folded structures of an immunoglobulin V_L domain and haemerythrin investigated by Novotny *et al.* (1984), they found that the correctly folded structures were stabilized by 17 to 34 kcal/mol relative to the incorrectly folded structures. When their method is applied to our structures, we find that the free energies of solvation for all the restrained dynamics structures, including the incorrectly folded RDIA structure, and the X-ray structure, are 19 to 24 kcal/mol smaller than that for the initial structure Ini I. There is no significant difference, however, between RDIA and the correctly folded restrained dynamics structures, the solvation free energy of RDIA lying in the middle of the range, ~22 kcal/mol smaller than that for Ini I. Thus, in this particular case, the solvation free energy does not provide a useful guide in distinguishing correctly from incorrectly folded structures. This may be due to the fact that crambin is not a globular protein and as such does not possess a large hydrophobic core.

Table 5

Atomic r.m.s. differences (Å) between the average restrained dynamics structures

	r.m.s. difference (Å) for all atoms											
	RDIA	RDIB	RDIB′	RDIC	RDIC′	RDID	RDID′	RDIE	RDIE′	RDII	RDII′	RDave†
RDIA		6·4	6·4	6·4	6·2	6·0	5·8	6·2	6·1	6·6	6·4	6·3
RDIB	5·8		0·8	2·6	2·5	2·8	2·9	2·5	2·5	2·6	2·6	2·5
RDIB′	5·9	0·6		2·8	2·6	2·9	3·0	2·5	2·5	2·5	2·5	2·5
RDIC	5·8	2·1	2·3		0·9	2·7	2·8	2·7	2·6	2·9	3·2	2·7
RDIC′	5·7	2·2	2·3	0·8		2·3	2·5	2·8	2·5	2·7	3·0	2·5
RDID	5·2	2·4	2·5	2·1	1·7		1·0	2·6	2·6	2·7	3·0	2·5
RDID′	5·6	2·5	2·6	2·3	1·9	0·8		2·6	2·5	2·7	2·9	2·5
RDIE	5·4	1·8	1·9	2·1	2·2	2·0	1·9		0·3	2·7	2·6	2·4
RDIE′	5·4	1·8	1·9	2·1	2·1	2·0	1·9	0·2		2·7	2·6	2·4
RDII	6·0	1·9	1·8	2·4	2·2	2·4	2·4	2·0	2·0		1·2	0·9
RDII′	5·8	2·0	1·8	2·7	2·5	2·6	2·5	2·0	2·1	1·1		1·6
RDave†	5·7	1·7	1·7	2·1	1·9	1·8	1·8	1·9	1·9	1·0	1·5	

r.m.s. difference (Å) for backbone atoms (N, C^{α}, C, O)

† The structure RDave is the structure obtained by first averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID and RDIE and RDII, and then subjecting the resulting average structure to restrained energy minimization (see the text).

Figure 5. Best fit superpositions of (a) the backbone (N, C^{α}, C, O) atoms and (b) the C^{α} and side-chain atoms of pairs of average free and restrained dynamics structures (thin lines) on the X-ray structure (thick lines). In each superposition, only the dynamics structures are labelled.

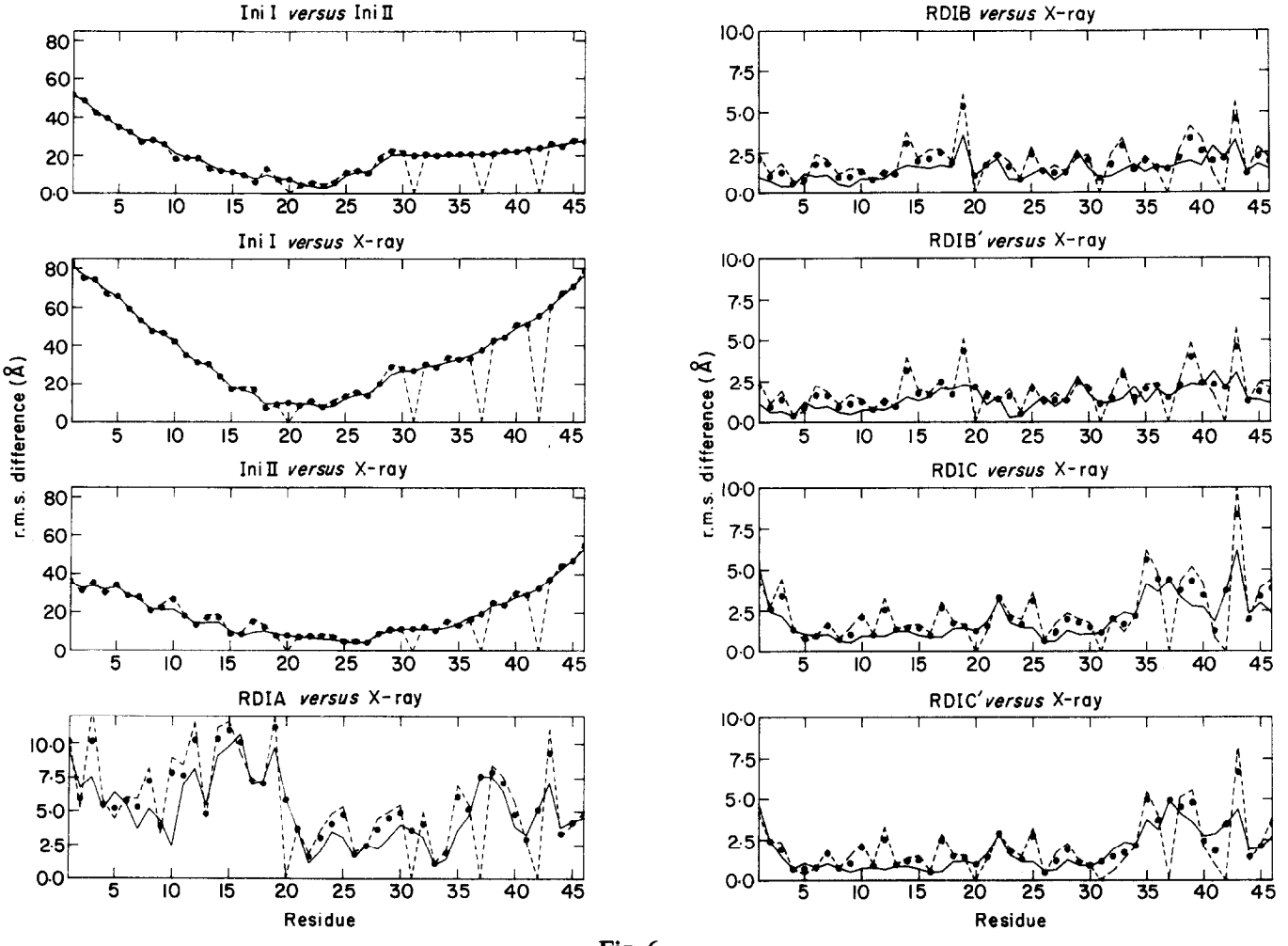


Fig. 6.

(b) *Characteristics of the converged structures*

The converged restrained dynamic structures (RDIB/IB', RDIC/IC', RDID/ID', RDIE/IE' and RDII/II') have a number of points in common.

The overall shape and size of the protein as well as the polypeptide fold are well reproduced. This is readily apparent from the superpositions of the restrained dynamics structures with the X-ray structure shown in Figures 4 and 5. Despite the approximate nature of the distance restraints, there are no signs of expansion of the restrained dynamics structures relative to the X-ray structure. Indeed, the radii of gyration of the restrained dynamics structures range from 9·2 to 9·6 Å, which is close to that of the X-ray structure (9·6 Å) and to those of the average free dynamics structures (9·3 to 9·5 Å) obtained by starting off from the X-ray structure. Also of interest is the fact that the radii of gyration tend to be smaller and are never larger than that of the X-ray structure, the relative reduction ranging from 0% to 5%. This is in contrast to the results of the distance geometry methods using similar interproton distance data sets, which yielded slightly expanded structures (Havel & Wüthrich, 1985; Braun & Go, 1985), and emphasizes the important role played by the non-bonding energies in the restrained dynamics simulations. In particular, it would appear that the distance restraints are essential in guiding the restrained dynamics into the correct region of conformational space and the empirical energy function is important in ensuring the correct stereochemistry and non-bonded interactions, which determine the size and shape of the protein.

The secondary structure elements, in particular the α-helices, the β-strands and β-sheets, and the turns, are all formed correctly, and their relative orientations with respect to each other are the same as in the X-ray structure. This holds, for example, for the angle (≈35°) and cross-over point (at residues 3 and 33) between the β-strand from residues 1 to 4 and that from 32 to 35, and for the orientation and angle (≈30°) of the long axes of the two helices relative to each other (see Fig. 4). The local conformations of the α-helices and β-strands are also well reproduced in general, as is readily seen by a comparison of the ϕ (Fig. 7) and ψ (Fig. 7) backbone torsion angles. Turning to the backbone hydrogen bonds (assigned using the criterion of a hydrogen bonding energy of < -1 kcal/mol; see Table 9), we note that all the hydrogen bonds present in helix 1 (residues 7 to 19) of the X-ray structure are also present in all the

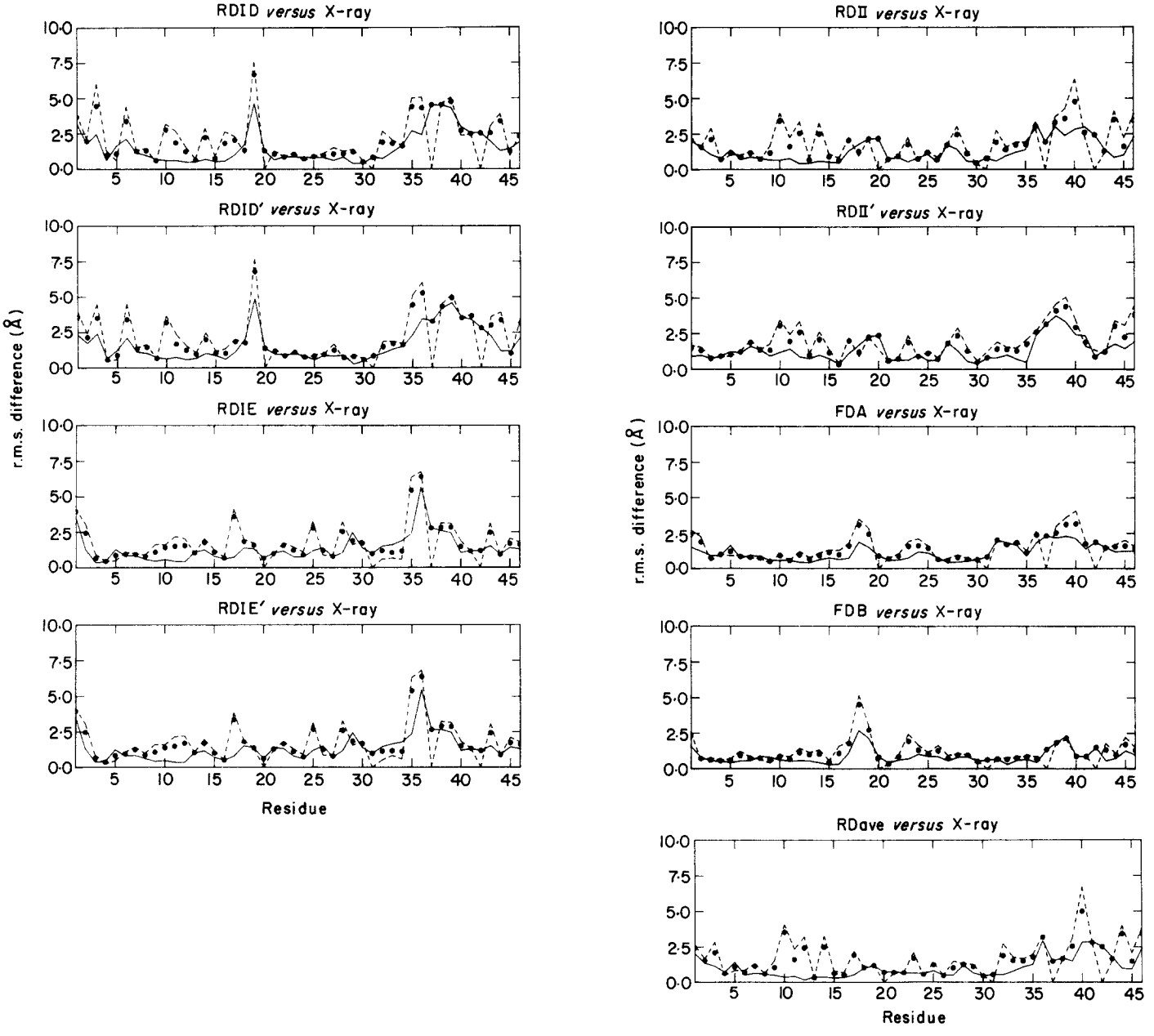


Figure 6. Atomic r.m.s. difference for all (●), the backbone (——), and the side-chain (— — —) atoms as a function of residue number for various pairs of structures involving the initial, the average free and restrained dynamics, and the X-ray structures.

restrained dynamics structures. In the case of helix 2 (residues 23 to 30) and the parallel/antiparallel β-sheet (composed of residues 44 to 46, 1 to 4 and 32 to 35), however, the correspondence of the backbone hydrogen bonds is not exact in the restrained dynamics structures. This is possibly due to the short length of these secondary structures in such a way that their geometry is easily distorted from ideal geometry and probably affected by the conformations of the adjacent turns. In this respect it should be stressed that the non-bonding energy terms play a significant role in defining the local backbone conformation of the secondary structure elements, as the classification of the short-range distance restraints into the three approximate ranges used here (namely, 2·5(±0·5), 3·0(+0·5/−1·0) and 4(±1·0) Å) does not in itself impose a rigid constraint. For example, the maximum variation in the $C^{\alpha}H(i)$–$NH(i+1)$ distance is only 2·2 to 3·6 Å; similarly, that for the $NH(i)$–$NH(i+1)$ distance is only 2·5 to 4 Å.

Only two small regions of the backbone are relatively poorly defined, as judged from the superposition of the restrained dynamics structures with the X-ray structure (Figs 4 and 5), the atomic r.m.s. difference plots (Fig. 6), and the ϕ (Fig. 7) and ψ (Fig. 8) backbone torsion angle plots. The first, which is only ill-defined in structures RDIB/RDIB' and RDID/RDID', involves residues 18 to 19 at the junction of the first helix and the

Table 6

r.m.s. differences of the interproton distances restraints and of the three S–S distances corresponding to the disulphide bridges present in crambin, together with the radii of gyration, for the initial, the average restrained dynamics, the average free dynamics and the X-ray structures

	r.m.s. difference (Å) between calculated and target distances					
	Interproton distances					
			Inter-residue			
Structure	All (240)	Intra-residue (25)	Short-range (159)	Long-range (56)	S–S distance for disulphide bonds (3)†	Radius of gyration (Å)
Ini I	39·2	1·07	4·48	78·6	92·7	45·9
Ini II	20·9	1·11	1·84	42·1	46·5	27·6
RDIA	0·91	0·42	0·79	1·28	8·49	8·55
RDIB	0·37	0·21	0·40	0·35	3·24	9·49
RDIB′	0·37	0·22	0·40	0·32	0·89	9·57
RDIC	0·36	0·30	0·36	0·41	3·78	9·28
RDIC′	0·35	0·20	0·36	0·39	0·86	9·27
RDID	0·34	0·25	0·32	0·43	3·41	9·32
RDID′	0·37	0·21	0·34	0·48	0·50	9·30
RDIE	0·33	0·16	0·35	0·34	1·64	9·18
RDIE′	0·34	0·17	0·36	0·36	1·04	9·18
RDII	0·39	0·30	0·38	0·43	3·73	9·61
RDII′	0·36	0·30	0·34	0·43	0·82	9·42
RDave‡	0·37	0·27	0·37	0·41	4·7	9·57
FDA	0·76§	0·49	0·67	1·04	2·87	9·45
FDB	0·76§	0·42	0·59	1·16	0·01	9·31
X-ray	0·26	0·17	0·27	0·26	0·04	9·64

† The 3 S–S distances corresponding to the disulphide bonds between residues 3 and 40, 4 and 32, and 16 and 26 are only included in the restraints energy for the following structures: RDIB′, RDIC′, RDID′, RDIE′, RDII′ and FDB.

‡ The structure RDave is obtained by averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID, RDIE and RDII, and subjecting the resulting average structure to restrained energy minimization (see the text).

§ The larger r.m.s. differences of the interproton distances for the free dynamics simulations principally arise from distances involving the aromatic rings and a few surface side-chains whose positions have changed slightly relative to their positions in the X-ray structure.

turn joining the two helices, and can be attributed to the ψ angle of residue 18, which is in a g^+ conformation instead of the correct g^- conformation as in the X-ray structure and the other restrained dynamics structures. The second comprises the turn from residues 35 to 40. This is probably not surprising as there are few distance restraints in this region (see Fig. 1). Interestingly, the region from residues 35 to 40 has higher than average backbone thermal factors in the crystal structure ($B = 9$ to 11 Å^2 compared to 3 to 6 Å^2 for the other residues; Hendrickson & Teeter, 1981).

The positions of the side-chains show a considerable improvement relative to their extended orientation in the initial structure, and, on the whole, are reasonably close to those in the X-ray structure. Indeed, the number of residues with a χ angle in a different conformational range from that found in the crystal structure ranges from four to seven, compared to 28 for the initial structures. This can be seen from the superpositions of the side-chains shown in Figure 5. In general, however, the atomic r.m.s. differences for the side-chains are slightly larger than for the corresponding backbone atoms (Fig. 6). This is not very surprising given that there are only 25 intra-residue restraints and that the inter-residue restraints involving non-aromatic side-chains are also few in number.

Table 7

Number of violations of the interproton distance restraints for the average restrained dynamics structures

	Violations >0·5 Å		Violations >2·0 Å	
Structure	Short-range ($\|i-j\| \leq 5$)	Long-range	Short-range ($\|i-j\| \leq 5$)	Long-range
RDIA	21	24	2	2
RDIB	0	0	0	0
RDIB′	0	0	0	0
RDIC	0	0	0	0
RDIC′	0	0	0	0
RDID	0	0	0	0
RDID′	0	1	0	0
RDIE	0	0	0	0
RDIE′	0	0	0	0
RDII	2	1	0	0
RDII′	0	0	0	0
RDave†	0	0	0	0

The violations are defined relative to the upper and lower limits of a particular distance range. For example, for the distance range 4(±1) Å, a violation of >0·5 Å occurs if the calculated distance is either less than 2·5 Å or greater than 5·5 Å.

† The structure RDave is the structure obtained by first averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID, RDIE and RDII, and then subjecting the resulting average structure to restrained energy minimization (see the text).

Table 8

Energies for the initial, the average restrained dynamics, the average free dynamics and the X-ray structure

	Energy (kcal/mol)†									
Structure	Total	Potential	Bond (652)	Angle (1183)	Torsion (320)	Improper (143)	van der Waals'	Electro-static	H-bond	Restraints (240)
Ini I‡	$2{\cdot}11\times10^{6}$	1920	595	413	194	0	776	−58	0	$2{\cdot}11\times10^{6}$
Ini II‡	$5{\cdot}98\times10^{5}$	1759	596	413	240	0	646	−136	0	$5{\cdot}98\times10^{5}$
RDIA	1184	239	60	390	322	25	25	−550	−34	945
RDIB	−408	−535	20	200	171	22	−156	−726	−65	127
RDIB′	−382	−511	20	203	184	19	−145	−731	−60	129
RDIC	−398	−523	17	202	207	18	−163	−734	−70	125
RDIC′	−343	−465	19	219	215	18	−151	−720	−65	122
RDID	−370	−474	18	202	185	28	−141	−689	−77	104
RDID′	−285	−422	19	208	212	28	−120	−680	−79	137
RDIE	−427	−537	18	197	181	18	−162	−721	−67	110
RDIE′	−398	−531	19	200	176	18	−147	−732	−64	133
RDII	−425	−543	16	179	159	21	−165	−685	−67	118
RDII′	−428	−542	18	189	172	21	−166	−718	−57	114
RDave§	−428	−532	16	172	167	21	−167	−675	−66	104
FDA‖	[126]	−679	17	150	148	13	−179	−739	−91	[805]
FDB‖	[1]	−691	16	152	165	18	−181	−773	−89	[692]
X-ray	−27	−91	81	151	271	0·3	−213	−327	−54	64

† For the average free and restrained dynamics structures the energies are those obtained after subjecting the average structures to 500 cycles of restrained energy minimization (with $S = 6$) and additionally constrained to their original structures by weak harmonic constraints. This procedure was used to correct for minor distortions in the covalent structure (namely, bond lengths and angles) produced by the averaging procedure and resulted in only very small atomic r.m.s. shifts (<0·2 Å for all atoms). The total energy is the sum of the potential and restraints energies, and the potential energy is made up of all the other bonded and non-bonded energy terms. The number of terms for the bond, angle, torsion, improper and restraints energy terms is given in parentheses.

‡ The initial geometries for the 2 initial structures were generated using the FRODO (Jones, 1978, 1982) dictionary and differ slightly from those of CHARMM.

§ The structure RDave is obtained by averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDIE and RDII and subjecting the resulting average structure to restrained energy minimization (see the text).

‖ For the free dynamics structures FDA and FDB the restraints were not included in the energy function. For this reason the total energy and the restraints energy are given in square brackets.

Considering the first side-chain torsion angle χ (Fig. 9), the largest differences involve either surface residues or side-chains whose positions are not restricted within a narrow region of conformational space either by the interproton distance restraints or by the packing requirements within the protein interior. The latter arise from the non-bonded interactions in the energy function (Gelin & Karplus, 1975).

Finally, convergence is achieved *without* the inclusion of S–S distance restraints corresponding to the disulphide bridges present in crambin (namely, between residues 3 and 40, 4 and 32, and 16 and 26). Furthermore, the subsequent inclusion of the three S–S distance restraints in the second phase of the refinement stage has only a very minor effect on the structure, resulting in atomic r.m.s. shifts of <1 Å (see Table 5). The minor nature of the changes can also be appreciated from the superpositions of the average restrained dynamics structures from the first phase of the refinement stage where *no* S–S distance restraints are present (e.g. RDIB, RDIC, etc. . . .) with those from the second phase of the refinement stage (e.g. RDIB′, RDIC′, etc. . . ., respectively) shown in Figures 4 and 5, as well as from the r.m.s. difference and torsion angle plots (Figs 6 to 9). This finding indicates that the interproton distance restraints together with the empirical energy function, including the non-bonding energy terms, are sufficient in this case to determine the three-dimensional structure of the protein alone. Thus, in appropriate cases, the restrained dynamics methodology should be applicable to proteins that do not possess disulphide bonds as well as to proteins that do, even if the location of their disulphide bridges has not been determined.

(c) *Further refinement possibilities*

The atomic r.m.s. differences between the converged restrained dynamics structure using different protocols tend to be slightly larger than the corresponding r.m.s. differences with the X-ray structure (Tables 4 and 5). This is readily appreciated by a comparison of the superposition of five of the restrained dynamics structures shown in Figure 10(a) with the superpositions of the restrained dynamics structures on the crystal structure shown in Figures 4 and 5. This suggests that on convergence the restrained dynamics simulations explore a region of conformational space close to the X-ray structure analogous to that explored in the free dynamics simulations. If this is indeed the case, then the mean of the converged restrained dynamics structures should be even closer to the X-ray structure than the restrained dynamics structures themselves. This is exactly what is

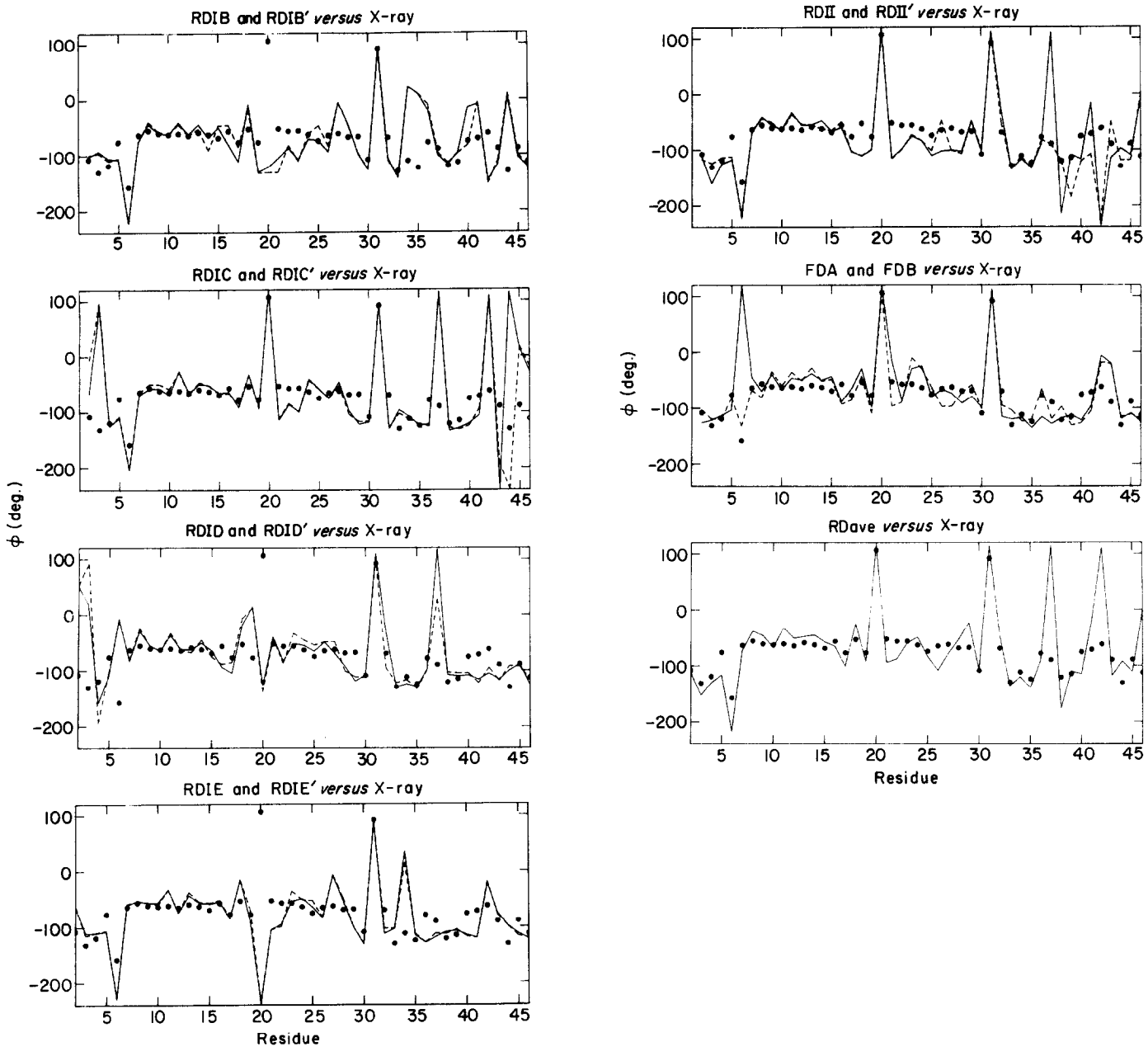


Figure 7. Comparison of the ϕ backbone torsion angles of the average free and restrained dynamics structures with the X-ray structure. (●) X-ray structure; (——) RDIB, RDIC, RDID, RDIE, RDII and FDA; (– – –) RDIB', RDIC', RDID', RDIE', RDII' and FDB.

observed. The r.m.s. difference between the X-ray structure and the structure generated by averaging the co-ordinates of the five restrained dynamics structures RDIB, RDIC, RDID, RDIE and RDII, is 1·0 Å for the backbone atoms and 1·6 Å for all atoms. This does not, however, imply that the five restrained dynamics structures represent a random sample of structures taken from a normal distribution about the X-ray structure. If this were the case, the average of the five structures would be expected to differ from the X-ray structure by r.m.s. differences of 0·8 Å and 1·0 Å for the backbone atoms and all atoms, respectively. (These values are given by $\sqrt{(\text{r.m.s. } d_i^2/N)}/\sqrt{N}$, where r.m.s. d_i is the r.m.s. difference for structure i, and N is the number of structures.) Rather it implies that the structure about which the restrained dynamics structures are normally distributed is close to but not identical with the X-ray structure. That it is not identical with the X-ray structure reflects imperfections in the present empirical energy function. These imperfections are likely to be even larger for the distance geometry approaches (Havel & Wüthrich, 1985; Braun & Go, 1985) as the only potential term considered is a soft van der Waals' repulsion term.

Not surprisingly, the average structure has a number of very bad contacts. We therefore subjected it to a total of 1240 cycles of restrained energy minimization (with the restraints scale

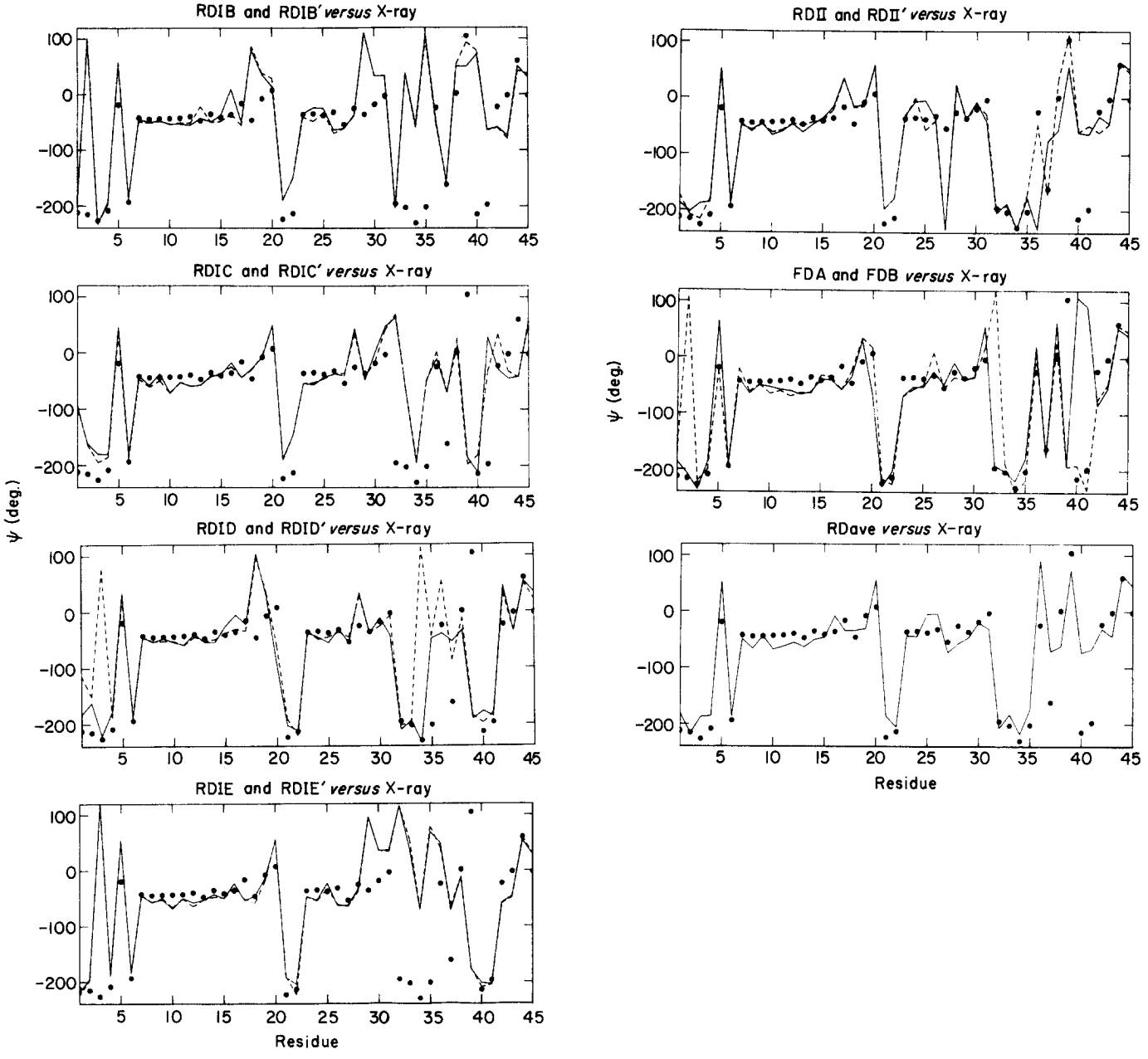


Figure 8. Comparison of the ψ backbone torsion angles of the average free and restrained dynamics structures with the X-ray structure. The symbols are the same as those for Fig. 7.

factor set to 6) in a two-step procedure in order to overcome the very high positive van der Waals' energy: this comprised 40 cycles with the van der Waals' radii reduced by a factor of 2 (as the initial van der Waals' energy with full van der Waals' radii was too high for the minimization program), followed by 1200 cycles with the full van der Waals' radii. The resulting structure, known as RDave, is shifted from the original averaged structure by 0·8 Å for the backbone atoms and by 1·0 Å for all atoms, has values for the bonding, non-bonding and restraints energy terms comparable to those for the other restrained dynamics structures (Table 8), and satisfies the restraints within their error limits with no violations (Tables 6 and 7). The best-fit superposition of the backbone atoms of RDave with the X-ray structure is shown in Figure 10(b). The r.m.s. difference between RDave and the X-ray structure is 1·2 Å for the backbone atoms and 1·9 Å for all atoms, which, though slightly worse than the values for the original average structure, is better than for any of the restrained dynamics structures (Table 4 and Fig. 6). The improvement in the r.m.s. difference for the backbone atoms is also reflected in an improvement in the ϕ, ψ backbone torsion angles (Figs 7 and 8) as well as an improvement in representation of the backbone hydrogen bonds for the second α-helix and the parallel/antiparallel β-sheet (Table 9). In comparison to the three best restrained dynamics structures, however, there is no large improvement in the positions of the side-chains as judged from the side-chain r.m.s.

Table 9

Backbone hydrogen bonds present in the X-ray and average free and restrained dynamics structures

$D_i(N)$–$A_j(O)$	X-ray	FDA	FDB	RDIB	RDIB′	RDIC	RDIC′	RDID	RDID′	RDIE	RDIE′	RDII	RDII′	RDave†
A. *Short-range* ($\|i-j\|<5$)														
9, 6						+								
10, 6	+	+	+	+	+	+	+	+	+			+	+	+
11, 7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
12, 8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
13, 9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
14, 10	+	+	+	+	+	+	+	+	+	+	+	+	+	+
15, 11	+	+	+	+	+	+	+	+	+	+	+	+	+	+
16, 12	+	+	+	+	+	+	+	+	+	+	+	+	+	+
17, 13	+	+	+	+	+	+	+	+	+	+	+	+	+	+
18, 14		+	+	+	+	+	+	+	+	+	+			
20, 17		+	+	+	+	+	+	+	+					+
25, 22		+	+						+		+			
26, 22	+							+		+	+	+	+	+
26, 23			+											
27, 23	+	+	+	+	+					+	+			+
28, 24	+	+	+			+	+			+	+			+
29, 25	+	+	+										+	
30, 26						+	+			+	+	+	+	
31, 27	+		+							+	+			
32, 27		+		+	+	+			+					
39, 36											+			
44, 41			+					+	+					
45, 42		+	+											
46, 44												+		
Number	13	16	18	12	11	14	12	12	13	13	15	11	11	12
B. *Long-range*														
1, 35													+	
1, 36													+	
1, 37					+									
1, 38								+	+					
2, 37						+								
3, 33	+	+	+										+	+
4, 44										+	+			
4, 46	+	+	+	+				+	+					
33, 3	+	+			+	+		+	+			+		+
35, 1	+	+	+	+		+	+					+	+	+
46, 4	+	+	+	+	+	+	+	+	+			+	+	+
Number	5	5	4	3	3	4	2	4	4	1	1	3	5	4

Only interactions contributing −1 kcal/mol or more are listed.

† The structure RDave is the structure obtained by first averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID and RDIE and RDII, and then subjecting the resulting average structure to restrained energy minimization (see the text).

difference plots (Fig. 6) and the χ side-chain torsion angle plots (Fig. 9).

(d) *The convergence pathway*

Figures 11 to 14 depict the pathways taken by the four restrained dynamics simulations IB, IC, ID and IE (structure determination stage) from the completely extended β-strand to the correctly folded end state.

In the initial phase of the simulation (namely, phase 1 for simulations IB and IC; the first six and nine picoseconds for simulations IC and ID, respectively; see Table 3), only the short-range restraints are applied. The two α-helices are rapidly formed, the exact speed depending on how the scale factor S for the restraints (see eqn (2)) is applied. When S is immediately set to a value of 1·0 from $t=0$ (cf. phase 1 of the IB and IC simulations) the α-helices are formed within two picoseconds. For simulations ID and IE, where S is gradually increased, starting from a value of 0·02 and doubled every 0·6 picosecond up to a maximum value of 2·5, the α-helices take four and six picoseconds, respectively, to form, at which time the structures are very similar to that at two picoseconds for simulations IB and IC.

To examine tertiary structure formation a comparison of the latter part of the first phase of simulations IB and ID is of interest. In the case of simulation ID, the linear polypeptide with two α-helices formed by six picoseconds is stable and essentially unchanged at nine picoseconds (Fig. 13). In contrast, the similar structure formed at two picoseconds in the IB simulation folds up over the subsequent three picoseconds, such that at five

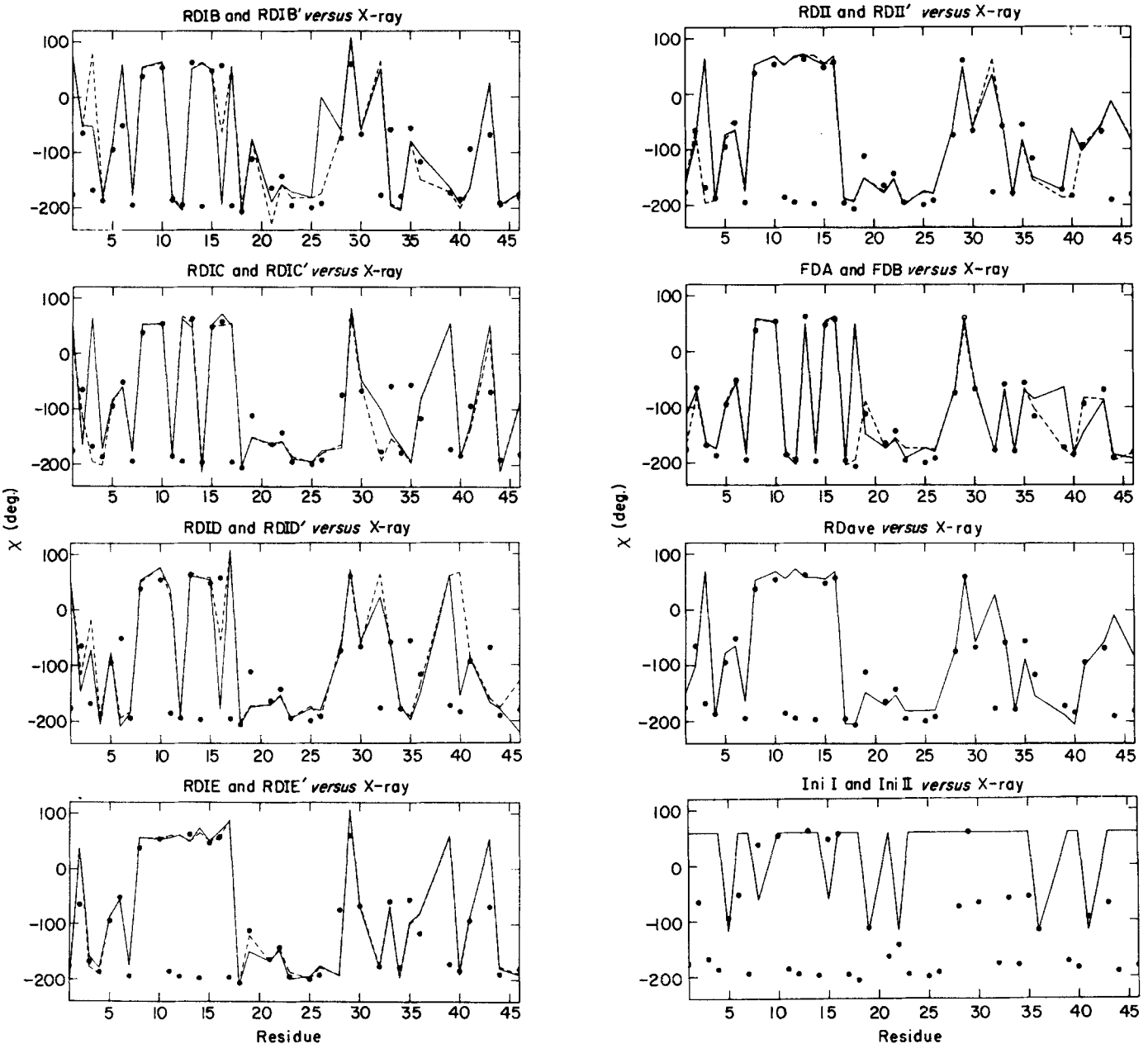


Figure 9. Comparison of the first side-chain torsion angles χ of the average free and restrained dynamics structures with the X-ray structure. The symbols are the same as those for Fig. 7. A comparison of the side-chain torsion angles χ of the initial structures (——) with the X-ray structure (●) is also shown.

picoseconds a structure only slightly larger than the final structure is obtained (Fig. 11). The folding of the polypeptide chain at this stage of the IB simulation is, of course, not correct. Nevertheless, the structure has some common features with the "native" structure. For example, the positioning of the two α-helices relative to each other is similar. The folding that occurs from two to five picoseconds during the first phase of the IB simulation can be attributed entirely to non-bonded interactions coupled with the very high atomic velocities. The latter arise from the large increase in kinetic energy and temperature (up to 3000 K) accompanying the rapid decrease in the restraints energy from an initial value of ~3300 kcal/mol at zero picoseconds to ~200 kcal/mol at two picoseconds. (Note that from 2 to 5 ps, the restraints energy decreases by a further 25 kcal/mol.) In other words, the difference between the IB and ID simulations is that the former is at high kinetic energy, whereas the latter is always at "normal" kinetic energy, except for short times.

When the long-range restraints are added to the total energy function (phases 2 and 3 of the IB simulation, phases 2 to 4 of the IC simulation, 9 to 21 ps for the ID simulation, and 6 to 18 ps for the IE simulation), the polypeptide chain begins to move towards the correct native structure. In all cases, once the scale factor S reaches a sufficiently high value to ensure that all the restraints are

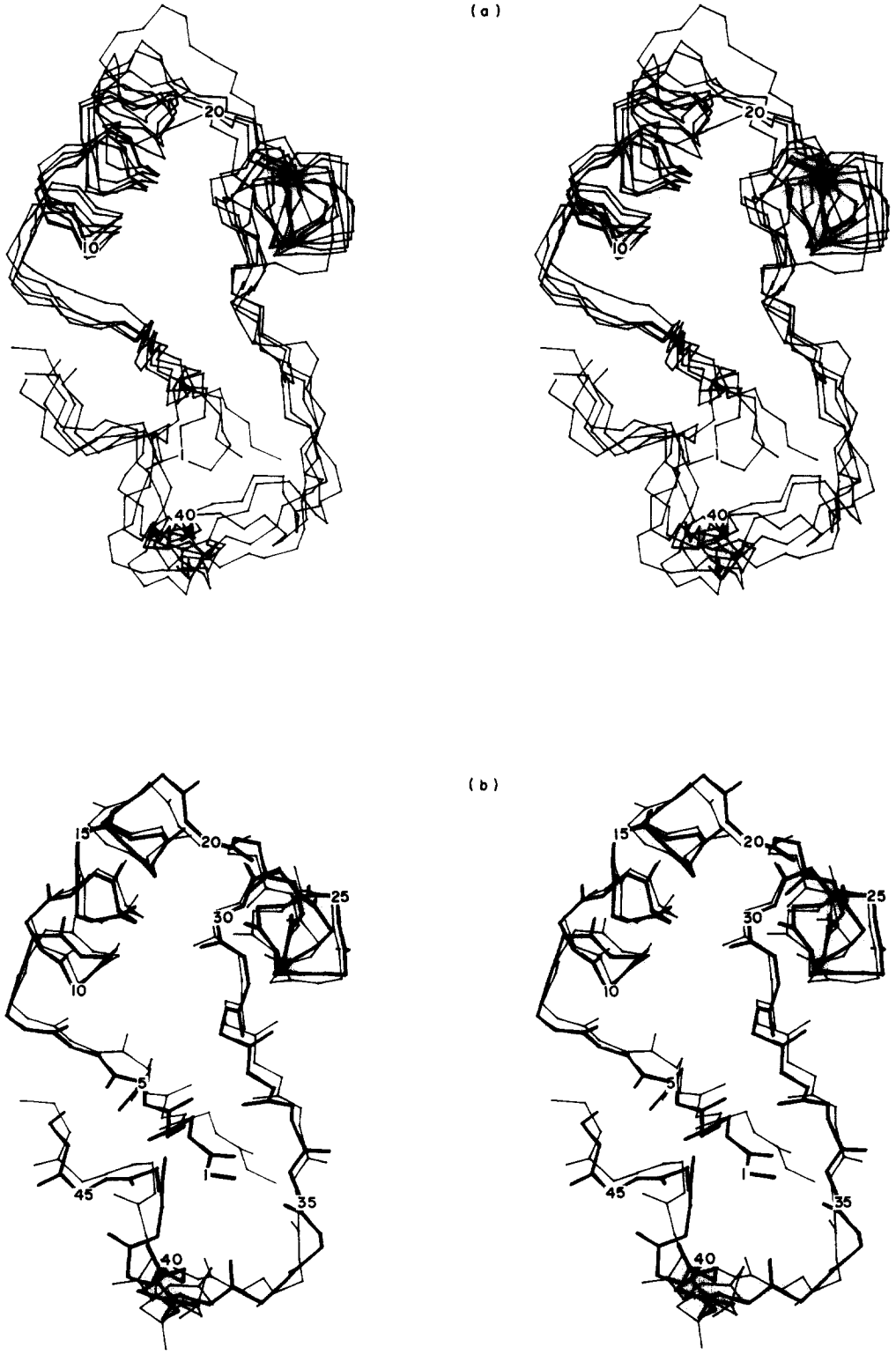


Figure 10. (a) Best-fit superposition of the backbone atoms (N, C^{α}, C) of the 5 average restrained dynamics structures RDIB, RDIC, RDID, RDIE and RDII. (b) Best-fit superposition of the backbone atoms (N, C^{α}, C, O) of RDave and the X-ray structure. RDave is the structure obtained by averaging the co-ordinates of the restrained dynamics structures RDIB, RDIC, RDID, RDIE and RDII, and then subjecting the resulting average structure to restrained energy minimization (see the text).

satisfied within their error limits, the folding process is rapidly completed. What is immediately apparent from a comparison of these simulations is that the folding pathways from the linear polypeptide with the two α-helices, although broadly similar, are by no means the same and there are no common structures formed that could be described as "folding intermediates". During this folding

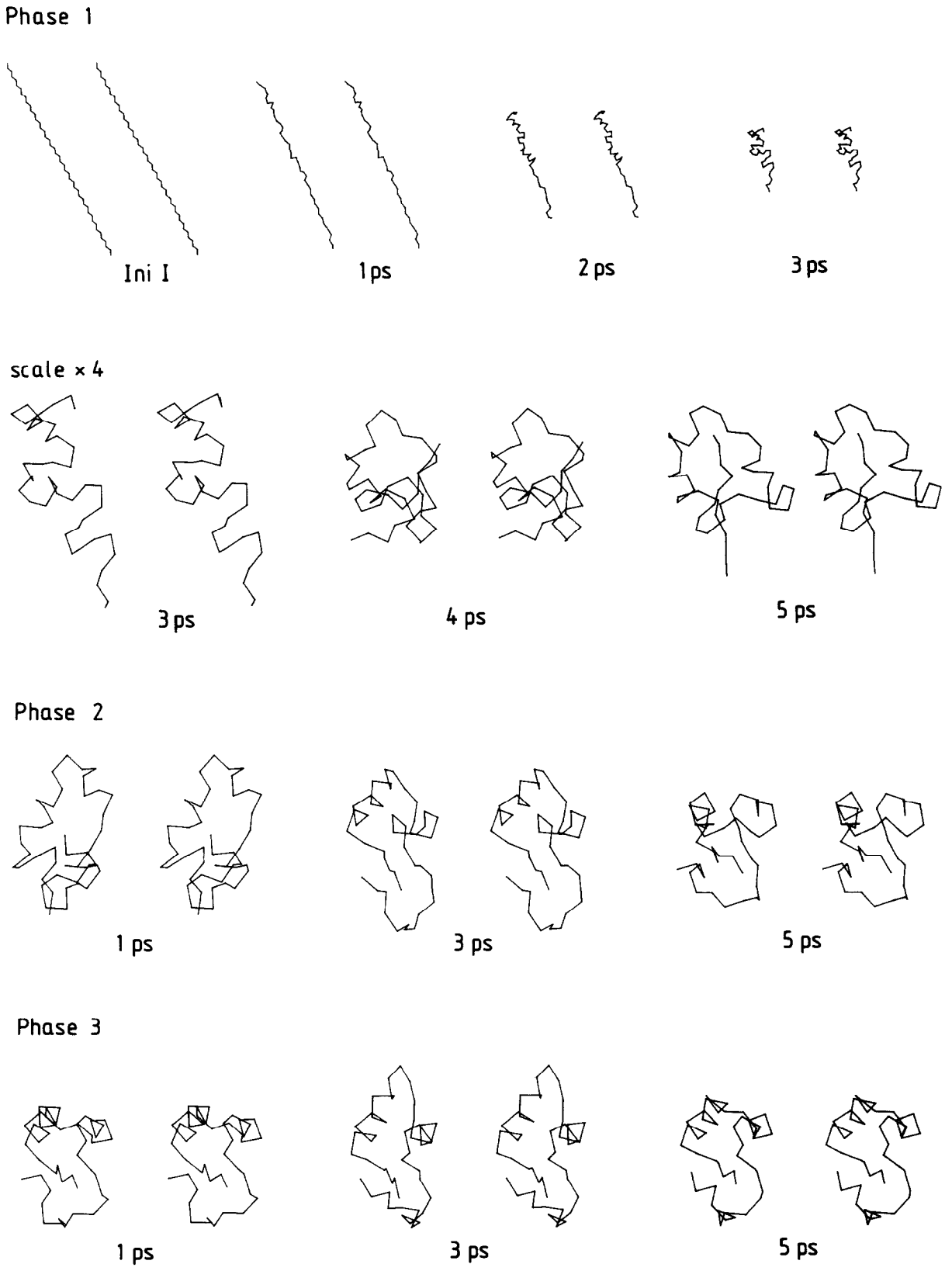


Figure 11. Snapshots of the trajectory for the 3 phases comprising the structure determination stage of the IB restrained dynamics simulation. During phase 1 only the short-range restraints are applied with $S = 1$ (where S is the scale factor for the restraints in eqn (2)): thereafter, all the restraints are included in the restraints energy with S set to 0·5 and 5 in phases 2 and 3, respectively. For clarity, only the C^{α} atoms are shown.

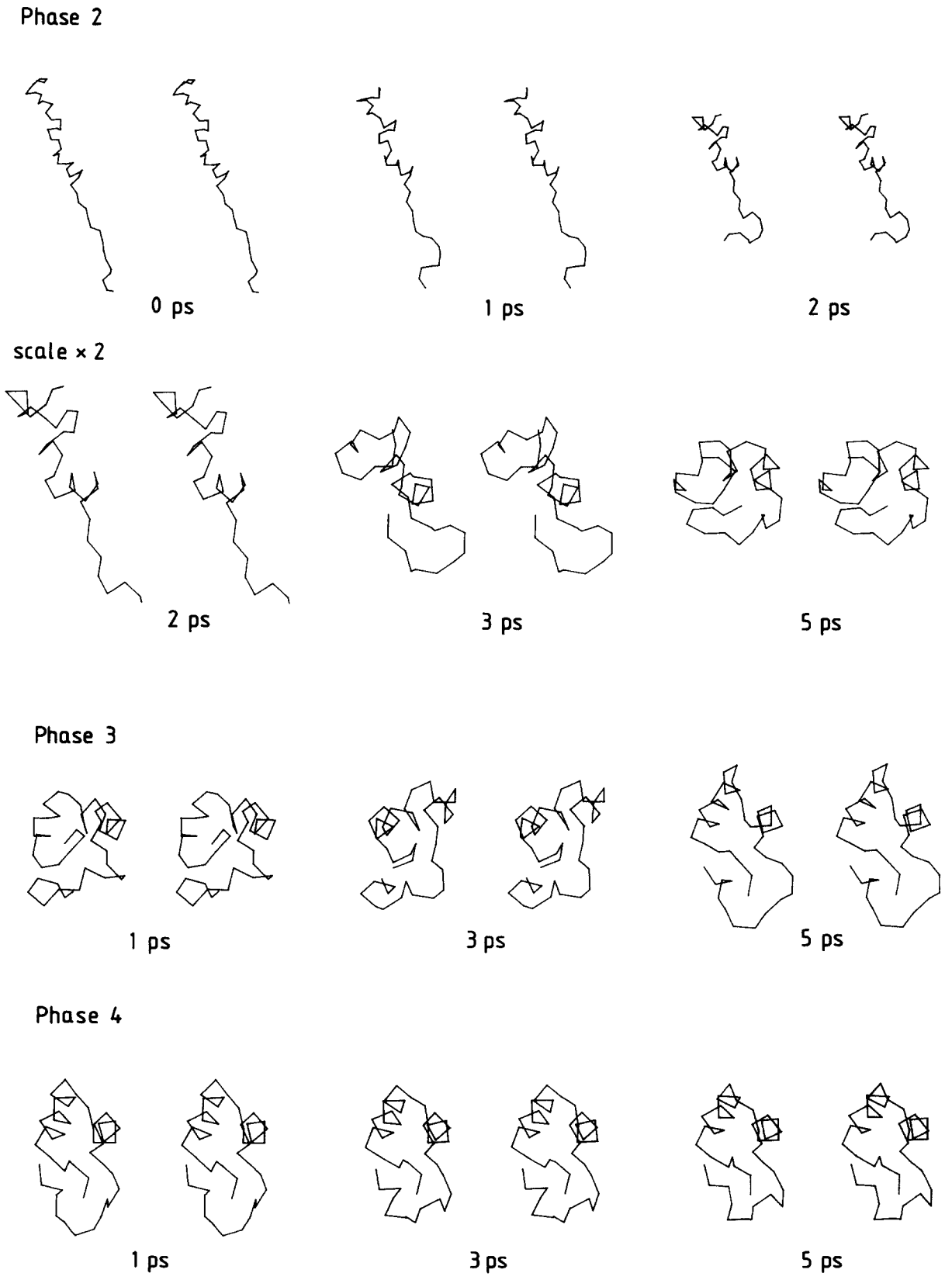


Figure 12. Snapshots of the trajectory for phases 2 to 4 comprising the structure determination stage of the IC restrained dynamics simulations. All the restraints are applied during these 3 phases with the restraints scale factor S (cf. eqn (2)) set to 0·01, 1 and 5 in phase 2, 3 and 4, respectively. Phase 1 of the IC simulation consists of the first 2 ps of phase 1 of the IB simulation. For clarity only the C^{α} atoms are shown.

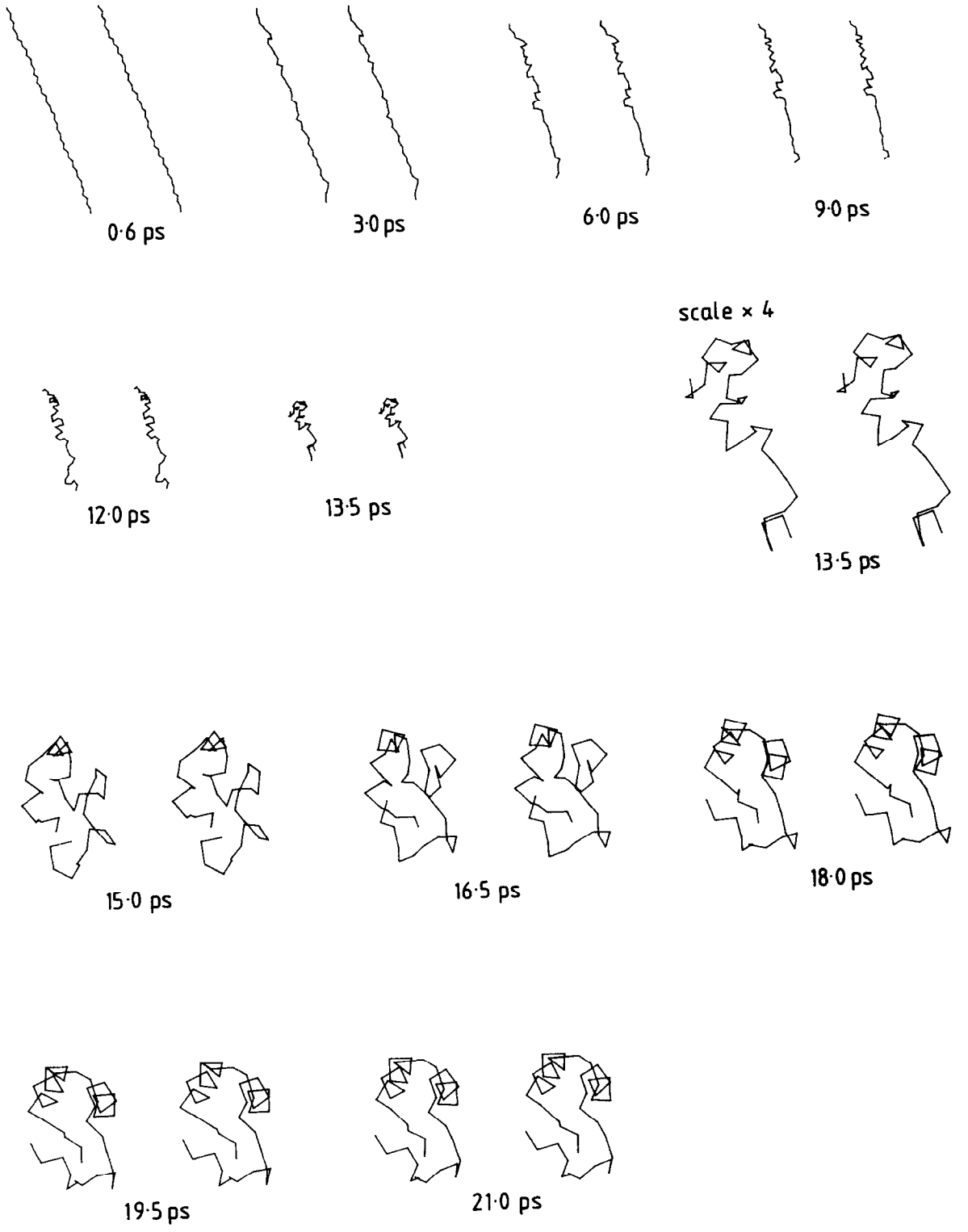


Figure 13. Snapshots of the trajectory comprising the structure determination stage of the ID restrained dynamics simulation. Only the short-range restraints are applied in the first 9 ps; thereafter, all the restraints are included in the restraints energy. For clarity only the C^{α} atoms are shown.

process, one of the two α-helices can unwind and then reform; i.e. for the IB, ID and IE simulations this involves the second α-helix (see the structures at 5 ps of phase 2 (Fig. 11), at 15 and 16·5 ps (Fig. 13), and at 12 ps (Fig. 14), respectively), while for the IC simulation this involves the first α-helix (see the structure at 5 ps of phase 2 (Fig. 12)).

An examination of the time courses of some of the energy terms is also instructive. These are depicted in Figures 15 and 16 for the ID and IE simulations, respectively. The principal driving forces at each stage is, of course, the reduction in the restraints energy for any given value of the scale factor S. This is manifested by a continuous

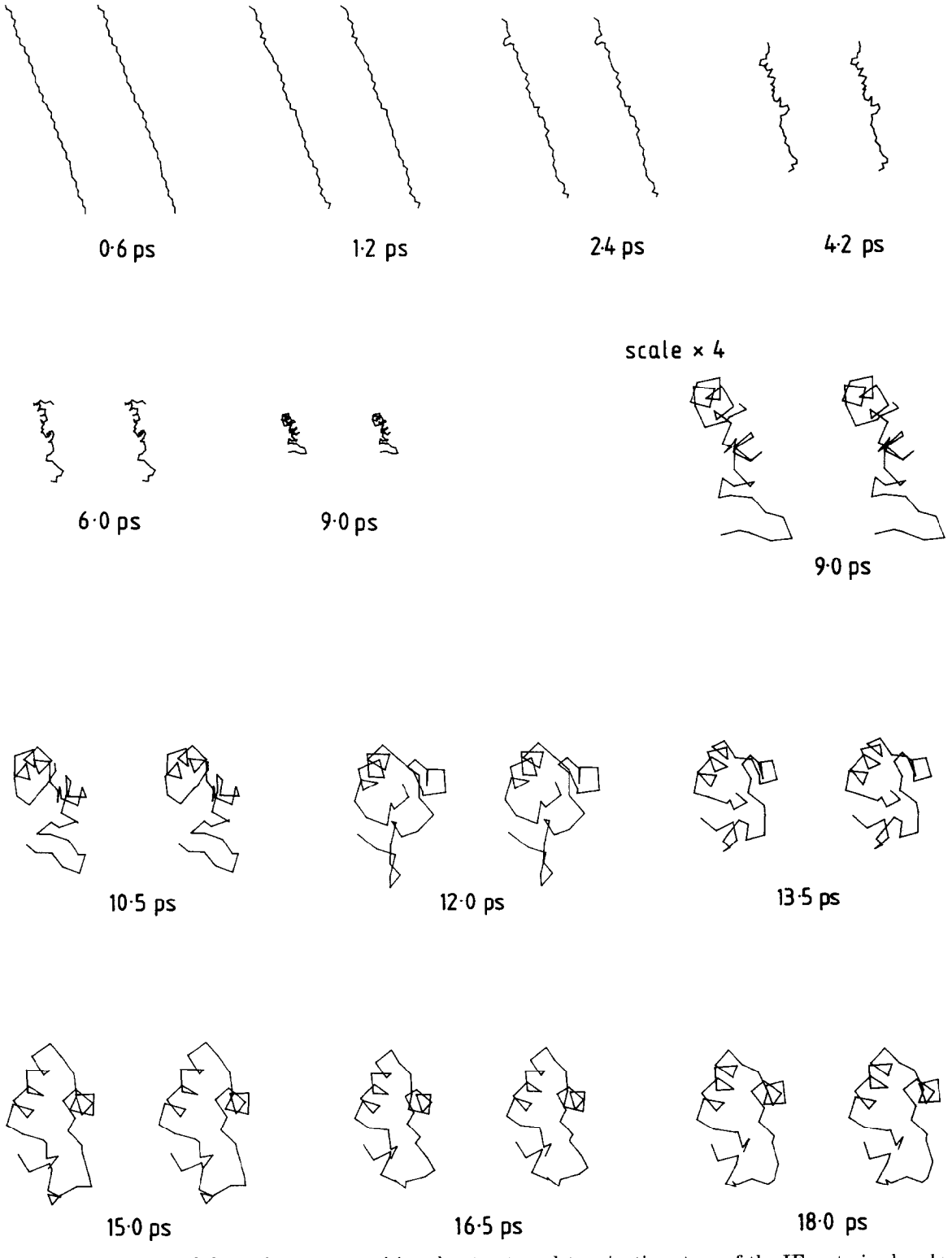


Figure 14. Snapshots of the trajectory comprising the structure determination stage of the IE restrained molecular dynamics simulation. Only the short-range restraints are applied in the first 6 ps; thereafter, all the restraints are included in the restraints energy. For clarity only the C^{α} atoms are shown.

decrease in the overall r.m.s. difference of the interproton distances. It is noteworthy, however, that, in both the ID and IE simulations, there is a transient small increase in the r.m.s. difference of the short-range interproton distances associated with the rapid decrease in the r.m.s. difference of the long-range interproton distance between 12 and 15 picoseconds. This arises from distortions in the secondary structure, such as helix unwinding (see above), which occur during the folding process. As to the molecular energy function terms, both the electrostatic and hydrogen-bonding components

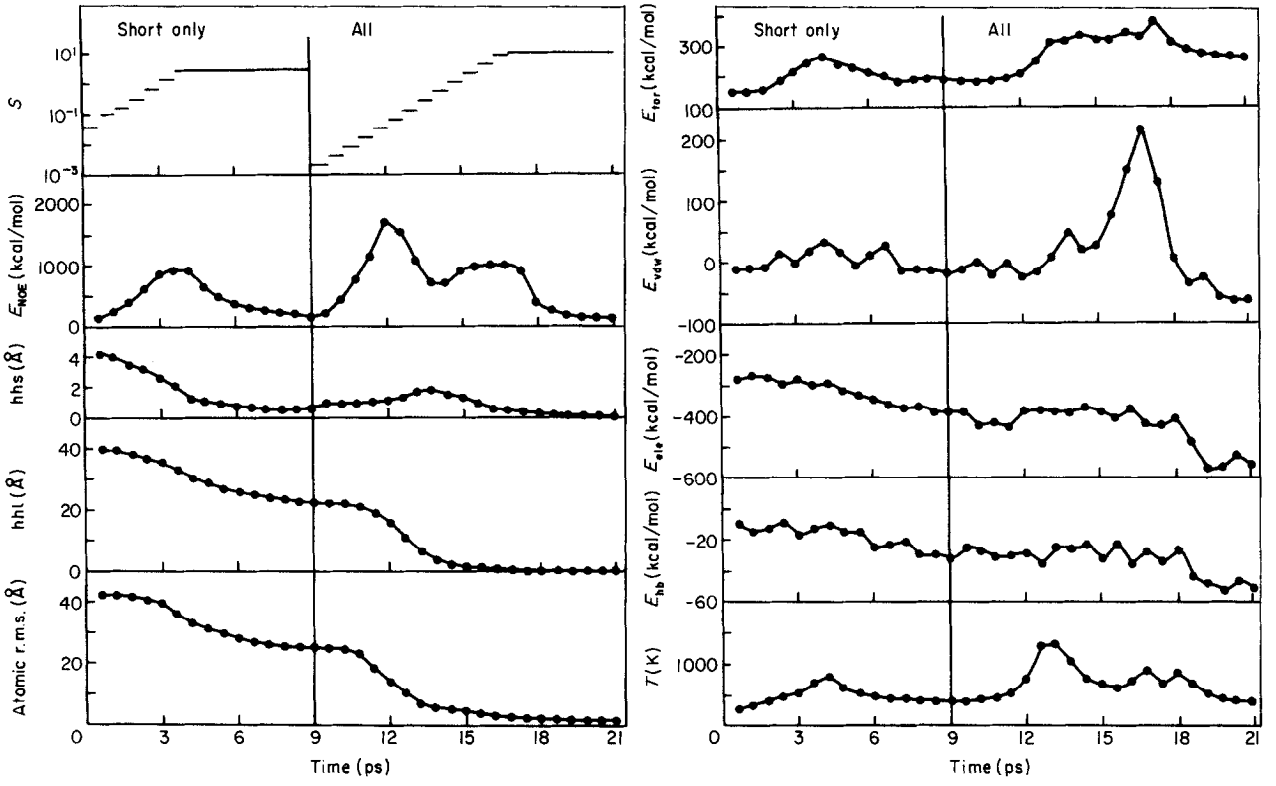


Figure 15. Time dependence of various parameters of interest during the structure determination stage of the ID restrained molecular dynamics simulation. The notation is as follows: hhs and hhl are the r.m.s. differences between the calculated and target values of the short- and long-range interproton distance restraints, respectively; atomic r.m.s. represents the atomic r.m.s. difference with respect to the X-ray structure; S is the scale factor for the restraints; E_{NOE}, E_{tor}, E_{vdw}, E_{ele} and E_{hb} are the restraints, torsion, van der Waals', electrostatic and hydrogen bonding energies, respectively; T is the absolute temperature.

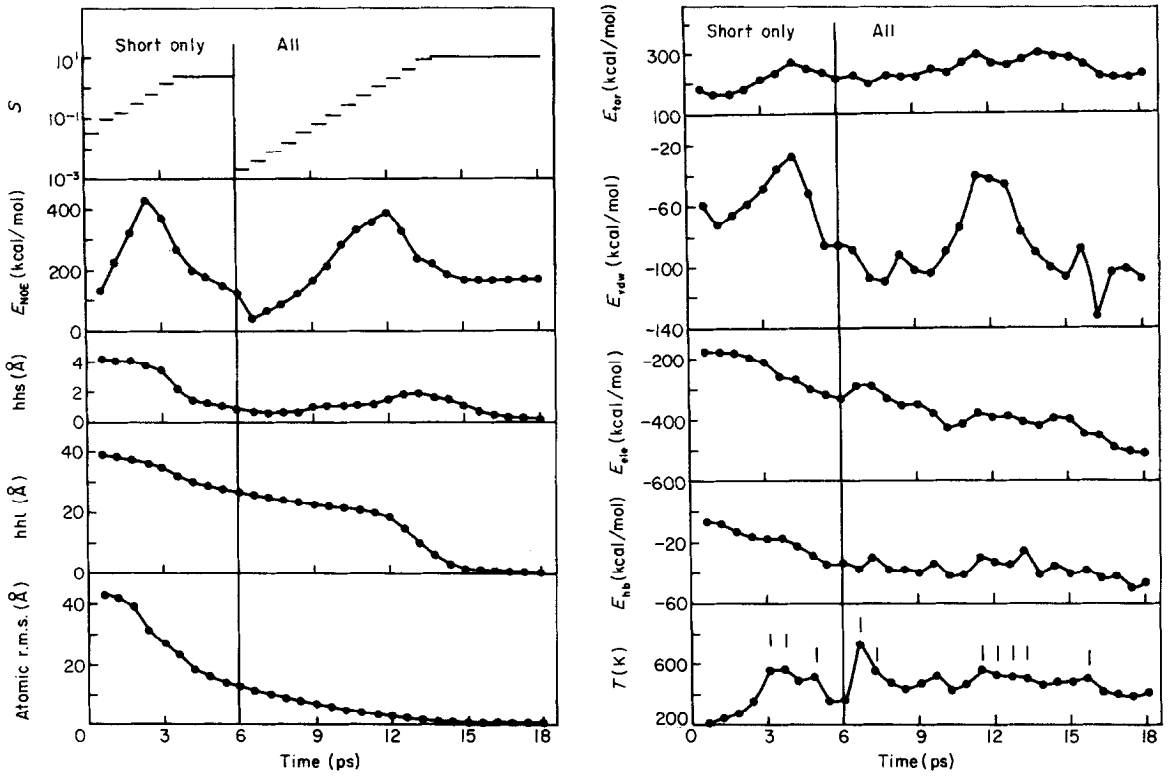


Figure 16. Time dependence of various parameters of interest during the structure determination stage of the IE restrained molecular dynamics simulation. The notation is the same as that in Fig. 15.

undergo a progressive decrease during the course of the simulations with only minor fluctuations. The van der Waals' and torsion energies, however, exhibit quite distinctive minima and maxima. In the first phase, during which time only the short-range restraints are applied, the torsion energy increases to a maximum at around four picoseconds and then decreases. A corresponding but smaller change in the van der Waals' energy is also seen. These maxima in the torsion and van der Waals' energy terms are associated with the transition of the ϕ, ψ backbone torsion angles from the extended β-strand to the α-helix domain. During the second phase of the simulations, when all the restraints are applied, the torsion potential slowly increases to a plateau which is maintained for five to seven picoseconds before decreasing again (at about 18 and 16 ps for the ID and IE simulations, respectively) as the correctly folded structure is finally formed. A distinctive maximum in the van der Waals' energy is also observed at 16·8 picoseconds for the ID simulations and 12 picoseconds for the IE simulation. This appears to be associated with the transition accompanying the refolding of the second α-helix and the correct formation of the turn connecting the two α-helices.

4. Concluding Remarks

In this paper we have shown that restrained molecular dynamics is a powerful method for determining three-dimensional protein structures on the basis of a set of approximate interproton distance restraints of the type that can be obtained from NOE measurements. Convergence to the correctly folded final structure, both globally and locally, can be achieved from completely unfolded initial structures. In addition, the method does not appear to be sensitive to the protocol used, providing this permits the secondary structure elements, and in particular the α-helices, to be formed at least partially prior to their folding into a tertiary structure. This may have its parallel in actual protein folding. From the methodological viewpoint, this requirement can be achieved in one of two ways: either by applying only the short-range ($|i-j| \leq 5$) restraints initially followed by the latter addition of long-range ($|i-j| > 5$) restraints, or by using an initial structure with preformed α-helical elements. The latter can be easily built in practice as the secondary structure elements can be readily determined and delineated by a qualitative interpretation of the NOE data, and in particular of the NOE connectivities involving the NH, $C^{\alpha}H$ and $C^{\beta}H$ protons (Wüthrich *et al.*, 1984). Thus, molecular dynamics with an empirical energy function used during the entire course of the structure determination ensures that the interproton distance restraints guide the folding to final structures, which not only exhibit the appropriate size, shape and folding of the polypeptide chain but also have approximately correct local stereochemistry and non-bonding interactions. The differences between the converged restrained dynamics structures provide a measure of the region of conformational space over which the interproton distance restraints can be satisfied. For the interproton distance set used here this region of space is approximately of the same magnitude as that sampled by long (100 to 300 ps) free dynamics simulations starting out from X-ray structures (Karplus & McCammon, 1979; Levy *et al.*, 1985). These results therefore provide a measure by which to judge future experimental results on proteins whose crystal structures are unknown. In particularly, if a set of restrained dynamics simulations, starting either from different initial structures or from the same structure but guiding the dynamics through different folding pathways results in a series of different structural types, all of which satisfy the interproton distances within their error limits, then the information content of the experimental interproton distance data can be deemed insufficient to determine the three-dimensional structure of the protein. Conversely, if convergence to a "unique" structural set that satisfies the interproton distances is achieved, then one can be confident that a realistic and accurate picture of the actual solution structure has been obtained and that the region of conformational space occupied by the global energy minimum has been located.

In applications for which no crystal structure is available, it would be appropriate to use, in addition to the interproton distance restraints, other information that can be derived from n.m.r. data. This includes the alignment of the individual β-strands within parallel and antiparallel β-sheets on the basis of interstrand NH–NH and $C^{\alpha}H$–$C^{\alpha}H$ NOE connectivities (Wüthrich *et al.*, 1984; Kline & Wüthrich, 1985). This enables one to define the backbone hydrogen bonds in these secondary structures and to add a set of corresponding H–O and N–O distance restraints (Williamson *et al.*, 1985). In addition, the ranges of some of the ϕ and χ torsion angles can be ascertained from the $^{3}J_{H,N^{\alpha}}$ and $^{3}J_{\alpha,\beta}$ coupling constants, respectively (Pardi *et al.*, 1985; Williamson *et al.*, 1985), permitting torsion angle restraints to be introduced into the total energy function in the form of torsion angle potentials.

Finally, the excellent agreement between the converged restrained dynamics structures and the X-ray structure suggests a novel approach to solving X-ray structures of small proteins for which no suitable isomorphous heavy-atom derivatives are available. First an approximate solution structure is obtained by the combined use of n.m.r. and restrained molecular dynamics. The converged restrained dynamics structures are then used as starting structures for Patterson search techniques for obtaining first the orientation and position of the molecule in the crystal unit cell (Lattman, 1985) and then the initial X-ray phases to calculate an initial electron density map. The feasibility of this approach is being examined in test calculations for

crambin, starting with the available restrained dynamics structures.

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.). We thank the Max-Planck Institut für Plasma Physik (Garching) and the NSF Supercomputer Initiative (University of Minnesota Computing Center) for CRAY 1 computing facilities, and Rita Sergeson for typing the manuscript.

References

Billeter, M., Braun, W. & Wüthrich, K. (1982). *J. Mol. Biol.* **155**, 321–346.

Braun, W. & Go, N. (1985). *J. Mol. Biol.* **186**, 611–626.

Braun, W., Wider, G., Lee, K. H. & Wüthrich, K. (1983). *J. Mol. Biol.* **169**, 921–948.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swamanithan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.

Bruccoleri, R. E. & Karplus, M. (1986). *J. Comp. Chem.* **7**, 165–175.

Brunger, A. T., Clore, G. M., Gronenborn, A. M. & Karplus, M. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3801–3805.

Clore, G. M. & Gronenborn, A. M. (1985). *J. Magn. Reson.* **61**, 158–164.

Clore, G. M., Gronenborn, A. M., Brunger, A. T. & Karplus, M. (1985). *J. Mol. Biol.* **185**, 435–455.

Crippen, G. M. & Havel, T. F. (1978). *Acta Crystallogr. sect. A*, **34**, 282–284.

Dobson, C. H., Olejniczak, E. T., Poulsen, F. M. & Ratcliffe, R. G. (1982). *J. Magn. Reson.* **48**, 87–110.

Eisenberg, D. & McLachlan, A. D. (1986). *Nature (London)*, **319**, 199–203.

Gelin, B. R. & Karplus, M. (1975). *Biochemistry*, **18**, 1255–1268.

Havel, T. F. & Wüthrich, K. (1984). *Bull. Math. Biol.* **46**, 673–698.

Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 281–294.

Hendrickson, W. A. & Teeter, M. A. (1981). *Nature (London)*, **290**, 107–112.

Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.

Jones, T. A. (1982). In *Computational Crystallography* (Sayer, D., ed.), pp. 303–371, Clarendon Press, Oxford.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). *J. Mol. Biol.* **182**, 179–182.

Karplus, M. & McCammon, J. (1979). *Nature (London)*, **277**, 578–581.

Karplus, M. & McCammon, J. (1983). *Annu. Rev. Biochem.* **52**, 263–300.

Kline, A. & Wüthrich, K. (1985). *J. Mol. Biol.* **185**, 503–507.

Kuntz, I. D., Crippen, G. M. & Kollman, P. A. (1979). *Biopolymers*, **18**, 939–957.

Lattman, E. (1985). In *Methods of Enzymology, Diffraction Methods for Biological Macromolecules* (Wyckhoff, H. W., Hirs, C. H. W. & Timasheff, S. N., eds), vol. 115, part B, Academic Press, New York.

Levy, R. M., Sheridan, R. P., Keepers, I. W., Dubey, G. S., Swaminathan, S. & Karplus, M. (1985). *Biophys. J.* **48**, 509–518.

McCammon, J. A., Gelin, B. R. & Karplus, M. (1974). *Nature (London)*, **267**, 585–590.

McCammon, J. A., Wolynes, P. G. & Karplus, M. (1979). *Biochemistry*, **18**, 927–942.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T. & Karplus, M. (1986). *J. Mol. Biol.* **188**, 455–475.

Novotny, J., Bruccoleri, R. & Karplus, M. (1984). *J. Mol. Biol.* **177**, 787–818.

Pardi, A., Billeter, M. & Wüthrich, K. (1985). *J. Mol. Biol.* **180**, 741–752.

Powell, M. J. D. (1977). *Mathemat. Progam.* **12**, 241–254. 241–254.

Strop, P., Wider, G. & Wüthrich, K. (1983). *J. Mol. Biol.* **166**, 641–667.

Verlet, L. (1967). *Phys. Rev.* **159**, 98–105.

Wagner, G. & Wüthrich, K. (1979). *J. Magn. Reson.* **33**, 675–680.

Wagner, G. & Wüthrich, K. (1982). *J. Mol. Biol.* **160**, 343–361.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.

Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). *J. Mol. Biol.* **155**, 311–319.

Wüthrich, K., Billeter, M. & Braun, W. (1984). *J. Mol. Biol.* **180**, 715–740.

Zuiderweg, E. R. P., Kaptein, R. & Wüthrich, K. (1983). *Eur. J. Biochem.* **137**, 279–292.

Edited by M. F. Moody