J. Mol. Biol. (1985) **186**, 435–455

Solution Conformation of a Heptadecapeptide Comprising the DNA Binding Helix F of the Cyclic AMP Receptor Protein of *Escherichia coli*

Combined use of ^{1}H Nuclear Magnetic Resonance and Restrained Molecular Dynamics

G. Marius Clore, Angela M. Gronenborn

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

Axel T. Brünger and Martin Karplus

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

(Received 8 July 1985)

A nuclear magnetic resonance study on a heptadecamer (17-mer) peptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli* is presented under solution conditions (*viz.* 40% (v/v) trifluorethanol) where it adopts an ordered helical structure as judged by circular dichroism. Using a combination of two-dimensional nuclear magnetic resonance techniques, complete resonance assignments are obtained in a sequential manner. From the two-dimensional nuclear Overhauser enhancement spectra, a set of 87 approximate distance restraints is derived and used as the basis for three-dimensional structure determination with a restrained molecular dynamics algorithm in which the interproton distances are incorporated into the total energy function of the system in the form of an additional effective potential term. The convergence properties of this approach are tested by starting from three different initial structures, namely an α-helix, a β-strand and a 3-10 helix. In all three cases, convergence to an α-helical structure is achieved with a root mean square difference of <3 Å for all atoms and <2 Å for the backbone atoms.

1. Introduction

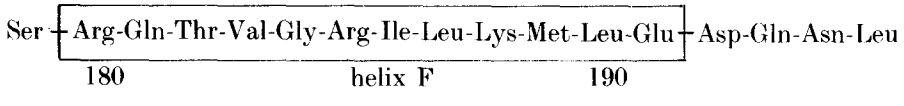
The cAMP receptor protein (CRP†) of *Escherichia coli* is involved in the regulation of transcription of at least 20 genes by binding in the presence of cAMP to specific DNA target sites located at the 5′ end of the respective gene: in some cases this results in the stimulation of transcription (e.g. in the *lac* operon), whereas in others it represses transcription (e.g. the CRP structural gene: for a review, see de Crombrugghe *et al.*, 1984). The mechanisms whereby CRP achieves its effects are unknown.

Within the C-terminal domain of CRP there is a helix–turn–helix motif comprising helices E (residues 168 to 175) and F (residues 180 to 191) (McKay & Steitz, 1981; McKay *et al.*, 1982), which shares a considerable degree of structural homology with a similar motif in three other sequence-specific DNA binding proteins, namely the *cro* (Anderson *et al.*, 1981), λ CI (Pabo & Lewis, 1982) and *lac* (Kaptein *et al.*, 1985) repressors. In the crystal structure of CRP, the two F helices from each subunit constitute a symmetrical pair of antiparallel α-helices that protrude from the surface of the protein and are separated from each other by 34 Å (McKay & Steitz, 1981). Model-building

† Abbreviations used: cAMP; cyclic AMP; CRP, cAMP receptor protein of *Escherichia coli*; n.m.r., nuclear magnetic resonance; c.d., circular dichroism; NOE, nuclear Overhauser effect; NOESY, 2-dimensional NOE spectroscopy; COSY, two-dimensional homonuclear *J* correlated spectroscopy; u.v., ultraviolet; TFE, trifluorethanol; r.m.s., root-mean-square; RM, restrained energy minimized; FM, free energy minimized; RD, restrained dynamics; FD, free dynamics.

studies have suggested that the two F helices can interact directly with two successive major grooves of right-handed *B*-DNA (Weber & Steitz, 1984). This is supported by genetic studies on mutations altering the DNA specificity of CRP that have shown that Glu181, the second residue of helix F, interacts with G·C base-pairs at positions 7 and 16 of the recognition site (Ebright *et al.*, 1984*a*,*b*).

Although X-ray crystallography could, in principle, provide extensive structural detail on the specific CRP–DNA interaction, crystallization of such a complex has not been possible to date. A complementary approach involves n.m.r. spectroscopy in solution, and in particular NOE measurements to demonstrate the proximity of protons in space (Noggle & Schirmer, 1971; Redfield & Gupta, 1971; Poulsen *et al.*, 1980; Wagner & Wüthrich, 1982; Gronenborn & Clore, 1982) followed by a refinement procedure to determine the three-dimensional solution structure from this data. CRP, however, is too large ($M_r \sim 46{,}000$) for such detailed structural studies by n.m.r. spectroscopy. One possible way to overcome problems due to the size of the protein may be to examine the interaction of the relevant portion of the protein, namely a small peptide comprising helix F, with a synthetic DNA oligonucleotide consisting of a specific target site. As a first step in this direction, detailed n.m.r. studies on a duplex DNA undecamer comprising the specific target site of CRP in the *gal* operon have been carried out (Clore & Gronenborn, 1984*a*,*b*,*c*), and its three-dimensional solution structure refined on the basis of interproton distance data by restrained least-squares minimization (Clore & Gronenborn, 1985*a*). In this paper, the second step is presented, namely an n.m.r. study on the solution structure of a heptadecamer peptide extending from residue 179 to residue 195:



which spans the sequence of helix F.

First, we show by c.d. measurements, that the 17-mer peptide, although a random coil in the absence of any helix-stabilizing solvent, adopts an ordered helical conformation at low TFE concentrations. Second, using two-dimensional n.m.r. spectroscopy, complete resonance assignments are obtained. A qualitative interpretation of the NOESY data confirms the c.d. measurements, and indicates that, at 40% (v/v) TFE, the 17-mer peptide is completely helical. Third, a set of approximate interproton distances is obtained from the NOE data and used as a basis of structure refinement by restrained molecular dynamics. This requires that the empirical energy potentials now available for molecular dynamics are sufficiently accurate to determine the conformation of a peptide in the presence of NOE interproton distance restraints. This is demonstrated by presenting three paths of refinements starting from three quite different initial structure, namely an α-helix, a 3–10 helix and an extended β-strand. In each case convergence to an α-helical structure is achieved, with an r.m.s. difference of less than 2 Å between the backbone atoms and less than 3 Å between all atoms of the three structures. On the other hand, restrained energy minimization with a standard conjugent gradient algorithm is shown not to yield this type of convergence. This ability of restrained molecular dynamics to search a wide range of conformational space and converge to a single type of structure demonstrates the power of this method of solution structure refinement based on interproton distance data. It suggests that additional tests of this approach, particularly to known structures, and its application to larger and more complex systems, would be of great interest.

2. Experimental Methods

The heptadecamer peptide was custom-synthesized by Cambridge Research Biochemicals (Cambridge, U.K.) using classical solid-phase methods and purified by reverse phase high-pressure liquid chromatography on a μ-Bondapak C_{18} column. Purity was assessed both by high-pressure liquid chromatography and fast atom bombardment mass spectroscopy, which gave a clean positive ion spectrum with $M+H^+$ at m/z 2001.

The standard buffer used in both n.m.r. and c.d. experiments was 10 mM-KCl, 3 mM-potassium phosphate (pH 6·6) and 0·02 mM-EDTA either in ^2H_2O or H_2O. In the n.m.r. experiments, the solutions also contained 40% (v/v) d_3-TFE. All n.m.r. experiments were carried out at 0°C and c.d. experiments at 5°C.

Circular dichroism spectra were recorded digitally from 320 to 200 nm using a JASCO JVI-C spectropolarimeter equipped with a J-DPY data processor, with a sensitivity of 0·5 millidegrees/cm and a time constant of 4 s. Cells with a pathlength of 2 mm were used.

All n.m.r. experiments were recorded on a Bruker AM 500 spectrometer equipped with digital phase shifters, an ASPECT 3000 computer and an array processor. For measurements in H_2O the solvent resonance was suppressed by selective irradiation during the relaxation delay and, in the case of the NOESY spectra, during the mixing time as well (Wider *et al.*, 1984). The 2-dimensional spectra in H_2O and ^2H_2O were recorded with sweep widths of 5208 Hz and 2732 Hz, respectively, with the carrier placed in the middle of the spectrum. The digital resolution employed was 5 Hz per point for the pure phase absorption NOESY spectra and 10 Hz per point for the absolute value COSY and relayed COSY spectra. This was achieved by appropriate zero-filling in the t_1 dimension only. In all cases the 2-dimensional spectra were symmetrized (Bauman *et al.*, 1981).

NOESY spectra (Jeener *et al.*, 1979; Anil Kumar *et al.*, 1980) at 2 mixing times (150 ms and 300 ms) were recorded in the pure phase absorption mode using the time proportional phase incrementation method (Redfield & Kunz, 1975; Bodenhausen *et al.*, 1980) as described by Marion & Wüthrich (1983). Appropriate phase cycling was used for the suppression of axial peaks and of cross-peaks due to coherence transfer *via* multiple quantum coherence transfer; in addition, a 10% random variation in the mixing time τ_m was used to eliminate zero-quantum coherence transfer (Macura *et al.*, 1981): 160 transients were collected for each of 600 increments with a relaxation delay of 1 s between successive transients. An initial phase correction was carried out during transformation with a final adjustment after completion of the 2-dimensional transform.

Absolute value COSY (Aue *et al.*, 1976; Nagayama *et al.*, 1980) and relayed COSY (Eich *et al.*, 1982; Wagner, 1983) spectra were also recorded, with appropriate phase cycling to provide quadrature information in the *F*1 dimension and to eliminate axial peaks.

Initial model building and displaying of all structures was carried out on the Evans & Sutherlands PS 330 Colour Graphics system interfaced to a VAX 11/780 computer using the interactive molecular graphics program FRODO (Jones, 1978, 1982) modified for the PS 330 by Dr Jim Pflugrath.

Energy minimization and molecular dynamics calculations were carried out on a CRAY 1A computer (Max-Planck Institut für Plasma Physik, Garching) using the program CHARMM (Brooks *et al.*, 1983) optimized for the CRAY (Brünger & Karplus, unpublished results). All analysis of structures, however, was carried out on the VAX 11/780, and the displaying of molecular dynamics trajectories was carried out using a modified version of the function network of FRODO interfaced with CHARMM.

Empirical energy potentials for the energy minimizations and molecular dynamics calculations were taken from Brooks *et al.* (1983), where the polypeptide atoms are represented by an extended atom model for non-polar groups and polar hydrogen atoms are treated explicitly. This potential was modified in order to treat the hydrogen atom attached to the C^{α} carbon explicitly. In addition, the hydrogen-bonding potential used by Brooks *et al.* (1983) was replaced by the appropriate parameterization of the Lennard-Jones and electrostatic non-bonding energy terms used in the CHARMM program version 19 (W. Reiher & M. Karplus, unpublished results). Solvent molecules were not explicitly included in the simulation, but the effect of solvent was approximated by multiplying the electrostatic energy term by a $(1/r)$ screening function (Brooks *et al.*, 1983). This is a particularly appropriate approximation for the polypeptide as all portions are about equally accessible to the solvent. The non-bonded interactions were switched off, using a cubic switching function, between 6·5 Å and 7·5 Å, with pairs up to 8 Å included in the non-bonded list. The potential energy term representing the NOE-restraints, described in detail in Results and Discussion section (d), was added to the total energy function of the system.

Integration of the equations of motion was performed by use of a Verlet integrator algorithm (Verlet, 1967) with initial velocities assigned to a Maxwellian distribution at the appropriate temperature. During heating, cooling and thermalization, the velocities were reassigned every 0·2 ps. The time step of the integrator was 0·001 ps and the non-bonded interaction lists were updated every 0·02 ps.

3. Results and Discussion

(a) *Circular dichroism*

The far u.v., c.d. spectrum of the 17-mer peptide in the standard buffer containing different concentrations of TFE (from 0 up to 66% (v/v)) is shown in Figure 1(a). In the absence of TFE, the CD spectrum shows very low negative intensity at 222 nm, indicating the virtual absence of any α-helical content. Except for the relatively low

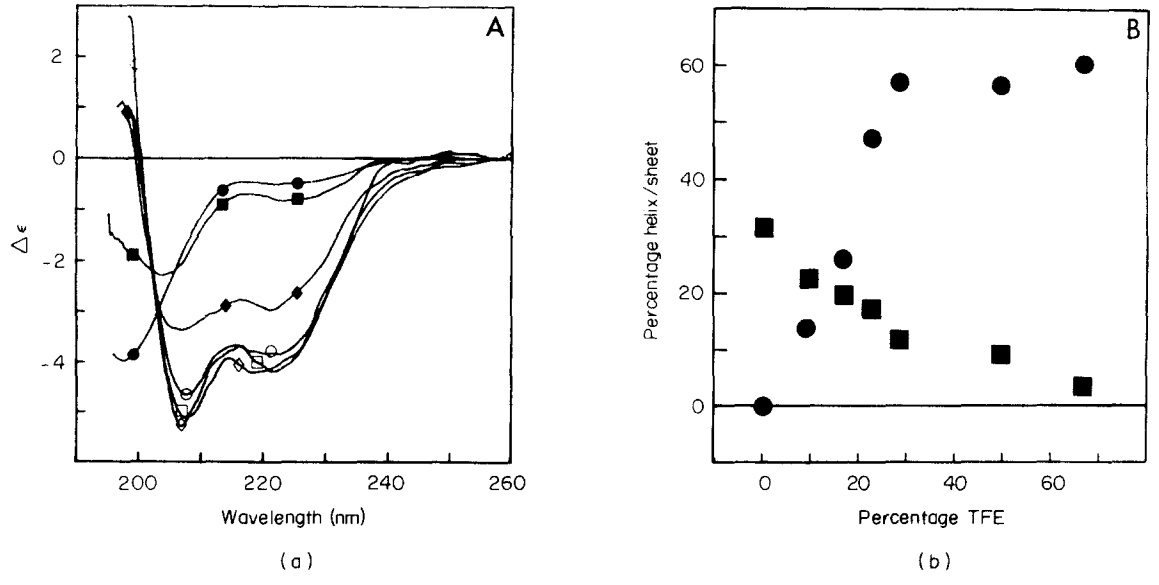


Figure 1. (a) Far u.v. c.d. spectra of the 17-mer peptide (concentration, 50 μM) at 5°C, as a function of TFE concentration: 0% (●), 9·1% (■), 16·6% (◆), 28·3% (○), 50% (□) and 66·6% (◇) (v/v). (b) Results of a multicomponent analysis of the data in (a) using the program CONTIN (Provencher & Glockner, 1981): α-helix (●) and β-sheet (■).

negative intensity at 198 nm and the presence of a broad negative band in the 215 to 225 nm region, it resembles most closely that of a random coil. Increasing concentrations of TFE, a well-known promoter of α-helical structure, leads to the expected spectrum for an α-helix, with prominent negative bands at approximately 207 nm and 222 nm. The results of a secondary structure analysis of the c.d. data using the program CONTIN (Provencher & Glockner, 1981) are shown in Figure 1(b). This analysis confirms the absence of α-helix and indicates a β-sheet content of 32% when no TFE is present. As the TFE concentration is increased, conversion to an α-helix accompanied by the virtual elimination of the β-sheet content is achieved. Two features of this transition should be noted. First, the concentration of TFE (30% (v/v)) required to complete the transition is unusually low, indicating that a relatively small perturbation of the solvent conditions is all that is required to induce an α-helical structure of the 17-mer peptide. Second, the CONTIN analysis indicates a maximal helical content of 60%, the remaining percentage being random coil. This figure, however, should be taken only as an approximate guide, as end-effects may make a significant contribution to the c.d. spectrum of the 17-mer peptide and the analysis itself is based on data derived from large proteins.

(b) *Nuclear magnetic resonance spectroscopy*

On the basis of the c.d. data, all n.m.r. experiments were carried out in the presence of 40% (v/v) d_3-TFE, a condition ensuring the complete conversion of the 17-mer peptide to an ordered helical structure. Sequence-specific assignments were carried out in a sequential manner by means of two-dimensional n.m.r. spectroscopy, as described (Wüthrich *et al.*, 1982; Wüthrich, 1983; Billeter *et al.*, 1982; Wagner & Wüthrich, 1982; Strop *et al.*, 1983; Zuiderweg *et al.*, 1983). In short, this involves first identifying spin systems through use of COSY and relayed coherence transfer experiments, followed by NOESY experiments to identify through space short (<5 Å) interproton contacts of which the inter-residue $C^{\alpha}H_i$–NH_{i+1} (d_1), NH_i–NH_{i+1} (d_2) and $C^{\beta}H_i$–NH_{i+1} (d_3) connectivities are the most important for the purpose of sequential resonance assignment. An example of a phase-sensitive NOESY spectrum in H_2O showing d_1 and d_2 connectivities is shown in Figure 2 and the assignments of the proton resonances are given in Table 1. It should be noted that, with the exception of the C^{α} and C^{β} protons of Ser1, there is only one resonance position per proton. We can therefore conclude that for the segment extending from residue 2 to residue 17 there is either only one conformation present or, if more than one, they are in fast exchange on the chemical shift scale. In the case of Ser1, however, we find two resonance positions of approximately equal intensity for both the C^{α} and C^{β} protons, indicating two distinct conformations for the N-terminal Ser in slow exchange on the chemical scale. From the separation of 35 Hz between the two positions of the C^{α} proton, it can be deduced that the exchange rate between the two conformations of Ser1 is less than 200 per second.

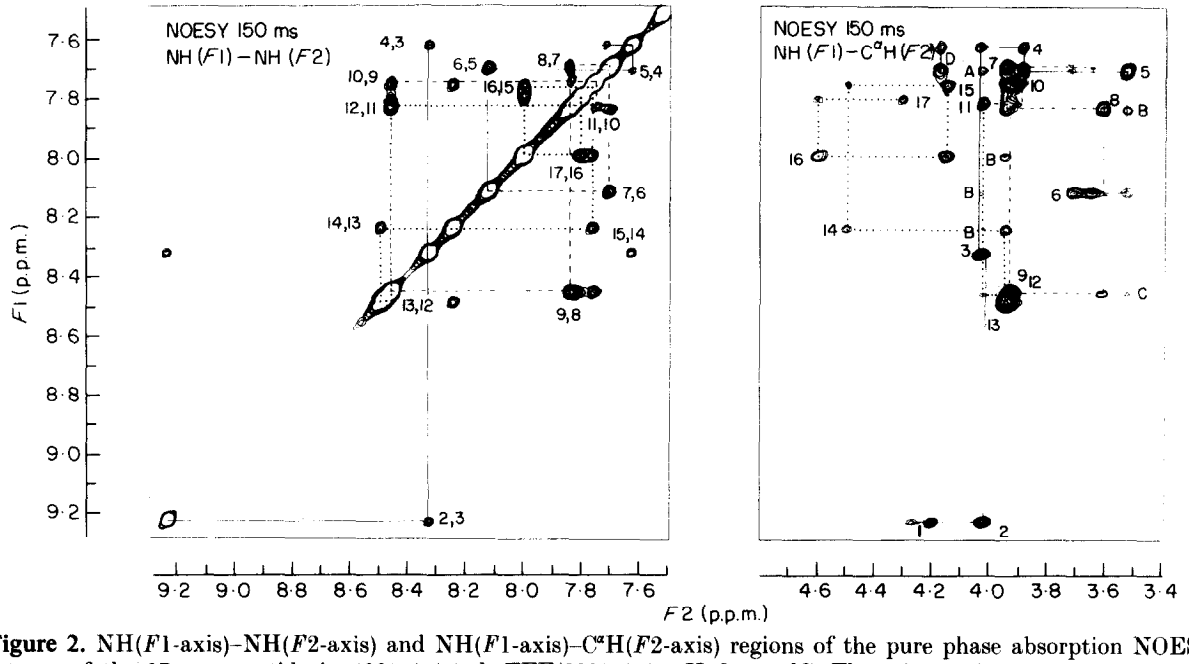


Figure 2. NH($F1$-axis)–NH($F2$-axis) and NH($F1$-axis)–C^{α}H($F2$-axis) regions of the pure phase absorption NOESY spectrum of the 17-mer peptide in 40% (v/v) d_3-TFE/60% (v/v) H_2O at 0°C. The mixing time is 150 ms and the concentration of 17-mer peptide is 2·5 mM. p.p.m., parts per million. The numbered cross-peaks in the NH($F1$)–NH($F2$) region represent NH_i–NH_{i+1} connectivities; the numbered cross-peaks in the NH($F1$)–C^{α}H($F2$) region represent $C^{\alpha}H_i$–NH_i connectivities; the notation for the lettered cross-peaks in the NH($F1$)–C^{α}H($F2$) region is as follows: A, $C^{\alpha}H_i$–NH_{i+2}; B, $C^{\alpha}H_i$–NH_{i+3}; C, $C^{\alpha}H_i$–NH_{i+4}; D, $C^{\beta}H_i$–NH_{i+1}. The lines (continuous, dotted and broken) in the NH($F1$)–C^{α}H($F2$) region delineate the $C^{\alpha}H_i$–NH_{i+1}–$C^{\alpha}H_i$ connectivities.

Table 1

Proton resonance assignments of the 17-mer peptide in 40% (v/v) d_3-trifluorethanol, 60% H_2O at 0°C

Residue	NH	Resonance $C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	Others
Ser1	—	4·28, 4·21	4·03, 3·81			
Arg2	9·22	4·02	1·75, 1·75	1·62, 1·55	3·09, 3·09	$N^{\varepsilon}H$ 7·11, guanidinium 6·77, 6·36
Glu3	8·32	4·03	1·95, 1·91	2·37, 2·12		
Thr4	7·61	3·88	4·17			$C^{\gamma}H_3$ 1·16
Val5	7·70	3·53	1·97			$C^{\gamma 2}H_3$ 0·92, 0·81
Gly6	8·11	3·72, 3·65				
Arg7	7·69	3·94	1·87, 1·71	1·54, 1·54	3·07, 3·07	$N^{\varepsilon}H$ 7·15, guanidinium 6·77, 6·36
Ile8	7·83	3·61	1·90	1·63, 0·99		$C^{\gamma}H_3$ 0·70, $C^{\delta}H_3$ 0·78
Leu9	8·45	3·93	1·82, 1·36	1·72		$C^{\gamma}H_3$ 0·70, 0·70
Lys10	7·74	3·92	1·84, 1·67	1·37, 1·37	1·64, 1·56	$N^{\varepsilon}H_3$ 7·49
Met11	7·82	4·02	2·18	2·59, 2·37		SCH_3 1·95
Leu12	8·45	3·13	1·82, 1·36	1·72		$C^{\gamma}H_3$ 0·70, 0·70
Glu13	8·48	3·95	2·13, 2·04	2·58, 2·36		
Asp14	8·23	4·49	2·95, 2·89			
Gln15	7·75	4·15	2·00, 1·97	2·38, 2·14		$N^{\delta}H_2$ 7·35, 6·54
Asn16	7·99	4·59	2·80, 2·63			$N^{\gamma}H_2$ 7·53, 6·74
Leu17	7·81	4·30	1·63, 1·57	1·49		$C^{\gamma}H_3$ 0·82, 0·74

† Two resonance positions of approximately equal intensity are found for both the C^{α} and C^{β} protons of Ser1, indicating the presence of 2 distinct conformations for this residue in slow exchange on the chemical shift scale. The $C^{\alpha}H$, $C^{\beta}H$ pairs are as follows: (1) 4·28 and 4·03 parts per million; (2) 4·21 and 3·81 parts per million. In both conformations, the 2 C^{β} protons have the same chemical shifts.

As discussed in detail by Wüthrich *et al.* (1984), the secondary structure can be deduced by a qualitative visual inspection of the NOESY spectrum. Looking at Figure 2, it may be seen that there are strong NH_i–NH_{i+1} (d_2) (Fig. 2(a)) and weak $C^{\alpha}H_i$–NH_{i+1} (d_1) (Fig. 2(b)) NOE connectivities for all 16 inter-residue steps. $C^{\beta}H_i$–NH_{i+1} (d_3) NOE connectivities are also observed for all inter-residue steps, with the exception of those between Gly6 and Arg7 (as glycine does not possess C^{β} protons) and between Lys10 and Met11. The general pattern of NOESY cross-peak intensities (a_{ij}) is as follows: $a_{NH_i\text{-}NH_{i+1}} \sim a_{C^{\alpha}H_i\text{-}NH_i} > a_{C^{\beta}H_i\text{-}NH_{i+1}} > a_{C^{\alpha}H_i\text{-}NH_{i+1}}$. These observations are indicative of an α-helical structure (Wüthrich *et al.*, 1984).

(c) *Interproton distances*

For short mixing times τ_m, the magnitude of the cross-peak intensity a_{ij} between two protons i and j is given by (Wagner & Wüthrich, 1979; Macura & Ernst, 1980; Dobson *et al.*, 1982; Keepers & James, 1984; Clore & Gronenborn, 1985*b*):

$$a_{ij}(t) \sim \sigma_{ij}t, \qquad (1)$$

where σ_{ij} is the cross-relaxation rate between the two protons. Because σ_{ij} is, in turn, proportional to r_{ij}, distance ratios, or distances, if one distance is already known, can be obtained from the equation:

$$r_{kl}/r_{ij} = (\sigma_{ij}/\sigma_{kl})^{1/6} \sim [a_{ij}(t)/a_{kl}(t)]^{1/6}, \qquad (2)$$

providing the effective correlation times of the i–j and k–l vectors are approximately the same. Note that the approximate relationship in equation (2) involving the cross-peak intensities a_{ij} and a_{kl} remains valid up to values of t three to four times longer than the initial rate approximation given in equation (1) (Clore & Gronenborn, 1985*b*). Equation (2) can, in principle, be used to determine interproton distances within an accuracy of ±0·2 Å (Clore & Gronenborn, 1985*b*). Indeed, this has been done successfully for the haem–methione moiety of ferrocytochrome *c* (Senn *et al.*, 1984), for small ligands bound to macromolecules (Gronenborn & Clore, 1982; Gronenborn *et al.*, 1984*a*,*b*; Clore *et al.*, 1984*a*) and for oligonucleotides (Gronenborn *et al.*, 1984*c*; Clore *et al.*, 1984*b*; Clore & Gronenborn, 1984*c*). However, in the case of proteins, such accuracy is, in general, not feasible. First, stereoselective resonance assignments of the methylene protons usually can not be made. Second, suitable internal reference distances are often not available owing to the poor spectral dispersion of the methylene proton pairs.

In the case of the 17-mer peptide, both limitations mentioned above apply, so that only approximate distances could be obtained. Our approach is based on that of Wüthrich and his collaborators, who use the known sequential distances [$r_{NH_i\text{-}NH_{i+1}}$, $r_{C^{\alpha}H_i\text{-}NH_{i+1}}$ and $r_{C^{\beta}H_i\text{-}NH_{i+1}}$] in the different secondary structures found in proteins (Braun *et al.*, 1983; Williamson *et al.*, 1985). Rather than employing their distance geometry approach (Braun *et al.*, 1983; Havel & Wüthrich, 1984, 1985; Williamson *et al.*, 1985), we have chosen to include the NOE data as restraints in a molecular dynamics simulation, as described in section (d), below.

The distances are calibrated on the basis of the relative cross-peak intensities, essentially as described by Williamson *et al.* (1985), with a few minor modifications, from the data at a mixing time

of 150 milliseconds as follows (see eqn (2)): for $a_{ij} \gtrsim 100$, $r_{ij} = 2{\cdot}5$ ($\pm 0{\cdot}5$) Å; for $a_{ij} = 40$ to 100, $r_{ij} = 3{\cdot}0$ ($\pm 0{\cdot}5$) Å; for $a_{ij} = 10$ to 40, $r_{ij} = 3{\cdot}5$ ($\pm 0{\cdot}5$) Å; and for $a_{ij} = 5$ to 10, $r_{ij} = 4$ (± 1) Å. (Note that a number of one-dimensional NOE experiments with irradiation of a few selected resonances for different times was used to check that the approximation in equation (2) was valid at 150 ms.) In the case of NOE values involving protons for which stereospecific assignments could not be made, the distances were referred to the directly bonded carbon atom and appropriate corrections for distances and errors were made on stereochemical grounds, taking into account the CH bond length, the van der Waals' radii of the C and H atoms and the errors in the estimation of the interproton distances. For example, consider an inter-residue NOE between two methyl groups corresponding to an $(\langle r^{-6} \rangle)^{-1/6}$ distance of 2·5 ($\pm 0{\cdot}5$) Å. From this distance one can deduce that the protons of the two methyl groups are pointing towards each other. Taking into account the tetrahedral geometry of the methyl group, it follows that the distance between the carbon atoms of the two methyl groups lies within the range 4(± 1) Å. Similarly, consider the C_i^{β}–NH_{i+1} distance. The $C^{\beta}H_i$–NH_{i+1} cross-peak intensities correspond to an $(\langle r^{-6} \rangle)^{-1/6}$ distance within the range 2·5 to 3·5 Å. Taken together with the knowledge that the NH_i–NH_{i+1} cross-peak intensities correspond to a distance range of 2 to 3 Å, model building then shows that the C_i^{β}–NH_{i+1} distance must lie within the range 3(± 1) Å. This approach, rather than generating a pseudo-atom (Wüthrich *et al.*, 1983), was used as it greatly simplifies the molecular dynamics simulations and enables one to reduce the number of atoms in the system, thereby shortening the computational time required. In fact, only polar and C^{α} protons were included in the calculations, other protons and their attached carbon atoms being represented by a single extended atom (Gelin & Karplus, 1975; Brooks *et al.*, 1983). The complete set of structurally useful NOE values (numbering a total of 87) together with the approximate distances derived from them are given in Table 2.

(d) *Restrained molecular dynamics*

At the present time three approaches have been tried to solve three-dimensional solution structures of macromolecules on the basis of experimental interproton distances determined by NOE measurements: (1) the use of distance geometry algorithms based on triangulation (Braun *et al.*, 1983; Arseniev *et al.*, 1983; Havel & Wüthrich, 1984, 1985; Williamson *et al.*, 1985); (2) restrained least-squares refinement (Clore & Gronenborn, 1985*a*; Clore *et al.*, 1985*b*,*c*); and (3) restrained molecular dynamics (Kaptein *et al.*, 1985). The distance geometry algorithm approach, in principle, enables one to generate a structure, provided the chirality of the structure is known (Crippen & Havel, 1978; Kuntz *et al.*, 1979; Wako & Scheraga, 1982; Havel & Wüthrich, 1984). This approach has been tested with model data derived from the crystal structure of bovine pancreatic trypsin inhibitor (Havel & Wüthrich, 1985) and has been applied with some success to lipid-bound glucagon (Braun *et al.*, 1983), insectotoxin I_5A (Arseniev *et al.*, 1983) and proteinase IIA from bull seminal plasma (Williamson *et al.*, 1985). The method appears to be able to define the global features (namely overall size, shape and folding) reasonably well if sufficient information is available, but leaves the local backbone conformation poorly determined (Havel & Wüthrich, 1985). This is due to a number of factors, including local convergence problems, and the absence of non-bonded interactions (i.e. van der Waals' and electrostatic) and torsion potentials in the algorithm. To overcome this problem, Williamson *et al.* (1985) have suggested applying a combination of energy minimization and interactive molecular graphics to the structures generated by the distance geometry algorithm. The second approach involves the restrained least-squares refinement of an initial starting model on the basis of the interproton distances subject to geometric (*viz.* bond length, bond angle and planarity) restraints with or without the inclusion of van der Waals' repulsion terms. This has been applied with some success to oligonucleotides and, within certain limits, the convergence properties of the method are good (Clore & Gronenborn, 1985*a*; Clore *et al.*, 1985*a*,*b*). Thus, this approach is capable of achieving convergence to a single *B*-DNA structure starting, for example, from two different *B*-DNA structures, but is incapable of achieving convergence to a *B*-DNA structure starting from an initial *A*-DNA structure. In the case of oligonucleotides, this problem is not very severe, as the structural type (*viz.* *A*, *B* or *Z*) can be determined unambiguously, not only from a qualitative interpretation of the NOE data but also from independent c.d. data. The final approach involves the application of restrained molecular dynamics. This involves the simultaneous solution of the classical equation of motion for all atoms of an assembly for a suitable time period (McCammon *et al.*, 1974, 1979; Karplus & McCammon, 1983), with the experimental interproton distances incorporated into the total energy function of the system in the form of effective potentials. A preliminary study using this approach has been carried out on the *lac* repressor head piece (Kaptein *et al.*, 1985). Only a single initial structure was used in this study so that the global convergence properties of the method could not be assessed.

The approach used to determine the solution structure of the 17-mer peptide is restrained molecular dynamics. Details of the methodology for the calculations, including equilibration and thermalization, have been described (Brooks *et al.*, 1983). The empirical energy function used in the molecular dynamics studies consists of bond, bond angle, torsional and non-bonding (i.e. van der

Table 2
Distance restraints (Å) derived from the NOESY spectra

A. *Sequential*

	NH_i–NH_{i+1} (d_2)	$C^\alpha H_i$–NH_{i+1} (d_1)	C^β_i–NH_{i+1} (d_3)	Others
Ser1–Arg2	2·5±0·5			
Arg2–Glu3	2·5±0·5	3·5±0·5	3±1	
Glu3–Thr4	2·5±0·5	3·0±0·5	3±1	
Thr4–Val5	3·0±0·5	3·5±0·5	3±1	$C^{\gamma 2}$–NH $3.5^{+0.5}_{-1}$, C^β–$C^{\gamma 1}$ 4±1
Val5–Gly6	2·5±0·5	3·5±0·5	3±1	$C^{\gamma 2}$–NH 3±1
Gly6–Arg7	2·5±0·5	3·5±0·5†		C^α–C^β 4±1
Arg7–Ile8	3·0±0·5	3·5±0·5	3±1	C^γ–NH 4±1
Ile8–Leu9	2·5±0·5	3·5±0·5	3±1	C^δ–NH 3±1
Leu9–Lys10	2·5±0·5	3·5±0·5	3±1	
Lys10–Met11	3·0±0·5	3·5±0·5		C^γ–NH 4±1
Met11–Leu12	2·5±0·5	3·5±0·5	3±1	
Leu12–Glu13	2·5±0·5	3·5±0·5	3±1	C^β–C^αH 4±1
Glu13–Asp14	2·5±0·5	3·5±0·5	3±1	
Asp14–Gln15	2·5±0·5	3·5±0·5	3±1	
Gln15–Asn16	2·5±0·5	3·0±0·5	3±1	
Asn16–Leu17	2·5±0·5	3·5±0·5	3±1	

B. *Intraresidue*

	NH–C^αH	Others
Arg2	2·5±0·5	NH–C^δ 4±1
Glu3	2·5±0·5	
Thr4	2·5±0·5	
Val5	2·5±0·5	
Gly6	2·5±0·5†	
Arg7	2·5±0·5	NH–C^δ 4±1, NH–C^γ 4±1
Ile8	2·5±0·5	NH–$C^{\gamma 1}$ $3.5^{+0.5}_{-1}$
Leu9	2·5±0·5	
Lys10	2·5±0·5	NH–C^γ 4±1
Met11	2·5±0·5	
Leu12	2·5±0·5	
Glu13	2·5±0·5	
Asp14	2·5±0·5	
Gln15	2·5±0·5	NH–$N^\delta H'$ 4±1, C^α–$N^\delta H'$ 3±0·5
Asn16	2·5±0·5	NH–$N^\gamma H'$ 3±1
Leu17	2·5±0·5	

C. *Long range*

Arg2 (C^αH)–Val5 ($C^{\gamma 1}$)	3±0·5	
Glu3 (C^αH)–Val5 (NH)	$4^{+1}_{-0.5}$	Glu3 (C^αH)–Gly6 (NH) $4^{+1}_{-0.5}$
Thr4 ($C^{\gamma 2}$)–Ile8 ($C^{\gamma 2}$)	4±1	
Val5 (C^αH)–Ile8 (NH)	$4^{+1}_{-0.5}$	Val5 (C^αH)–Leu9 (NH) $4^{+1}_{-0.5}$
Val5 (C^αH)–Ile8 ($C^{\gamma 2}$)	3±0·5	Val5 (C^αH)–Ile8 (C^β) 3·5±0·5
Met11 (C^αH)–Asp14 (NH)	$4^{+1}_{-0.5}$	
Glu13 (C^αH)–Asn16 (NH)	$4^{+1}_{-0.5}$	

The distances and errors were estimated from the relative intensities of the cross-peaks in the NOESY spectra as described in the text. In those cases where stereospecific assignments could not be made, the H is not specified and the distances are related to the attached carbon atom with appropriate corrections being made on the basis of stereochemical considerations.

† This distance applies to both C^α protons of Gly6.

Waals' and electrostatic) interaction terms. The NOE effective potential terms, introduced as restraints, are represented by a skewed biharmonic effective potential of the form:

$$E_{\mathrm{NOE}}(r_{ij}) = \begin{cases} c_1(r_{ij}-r^{\circ}_{ij})^2, & \text{if } r_{ij} > r^{\circ}_{ij}, \\ c_2(r_{ij}-r^{\circ}_{ij})^2, & \text{if } r_{ij} < r^{\circ}_{ij}, \end{cases} \quad (3)$$

where r°_{ij} and r_{ij} are the target and calculated interproton distances, respectively, and c_1 and c_2 are force constants given by:

$$c_1 = \frac{kTS}{2(\Delta^+_{ij})^2} \qquad c_2 = \frac{kTS}{2(\Delta^-_{ij})^2}, \quad (4)$$

where k is the Boltzman constant, T is the absolute temperature, S is a scale factor, and Δ^+_{ij} and Δ^-_{ij} are the positive and negative error estimates of r_{ij}, respectively. In the molecular dynamics calculations, the scale factor S in

Table 3

Energies of the initial, the free and restrained energy minimized, and the average free and restrained dynamics structures

	Energy, kJ mol^{-1} (number of terms)											r.m.s. difference between calculated and target restraints (Å)
	Total	Potential	Kinetic	Bond (194)	Angle (304)	Torsion (108)	Improper (50)	Electrostatic	Van der Waals'	Restraints‡ (87)	Temperature (K)	
Initial structures												
I	7886	7886	†	79	1542	79	85	−342	5998	445		0·65
II	8866	8866	†	79	127	199	6	−252	1287	7420		2·43
III	6418	6418	†	79	1542	71	68	−380	3958	1080		0·97
Free energy minimization§												
I	−1057	−1057	†	22	88	47	10	−1053	−217	[605]		0·79
II	−740	−710	†	13	123	74	15	−932	−70	[6974]		2·46
III	−940	−940	†	22	110	76	12	−1141	−90	[925]		0·99
Restrained energy minimization												
I	461	461	†	94	214	74	17	−548	104	504		0·68
II	3973	3973	†	145	373	204	40	−472	357	3330		1·65
III	749	749	†	103	264	96	24	−610	171	704		0·76
Free dynamics average§,‖												
I	−715 (0·7)	−1363 (29)	648 (29)	194 (22)	347 (27)	171 (16)	63 (12)	−1927 (31)	−213 (30)	[657]	296 (13)	0·78
II	−253 (1·3)	−928 (33)	677 (34)	198 (21)	400 (31)	222 (24)	72 (13)	−1709 (56)	−113 (33)	[4679]	309 (16)	1·87
III	−744 (0·5)	−1396 (29)	652 (29)	195 (22)	356 (28)	152 (17)	66 (13)	−2006 (31)	−157 (31)	[815]	296 (13)	0·86
Restrained dynamics average‖												
I	−639 (0·3)	−1237 (25)	598 (25)	178 (21)	351 (27)	164 (17)	63 (11)	−1931 (36)	−223 (29)	158 (11)	273 (12)	0·35
II	−466 (2·4)	−1128 (30)	662 (30)	185 (21)	409 (32)	183 (24)	69 (13)	−2006 (38)	−171 (35)	204 (13)	302 (14)	0·42
III	−564 (0·9)	−1198 (29)	634 (28)	181 (20)	370 (29)	174 (18)	73 (13)	−1973 (33)	−217 (33)	195 (14)	288 (13)	0·39

The symbols I, II and III refer to the structures obtained starting off from the α-helix, β-strand and 3-10 helix initial structures, respectively. The average dynamics structures are the averages over the last 8 ps of the second dynamics run (see the text).

† The initial and the energy minimized structures have zero kinetic energy.

‡ The scale factor S in eqn (4) used for the restraints was 5. Thus the value of the force constants in eqns (3) and (4) corresponding to error estimates of 0·5 Å and 1 Å in the distances are 24·9 kJ mol^{-1} Å^{-2} and 6·23 kJ mol^{-1} Å^{-2}, respectively.

§ The total and potential energies for the free energy minimized structures and the total, potential and kinetic energies for the average free dynamics structures do not include the restraints energy component. The latter are given in square brackets as they were not included in these calculations.

‖ The energies for the dynamics structures are obtained by averaging the energies over the structures from the last 8 ps of the second dynamics run. The r.m.s. fluctuations in the energy terms are given in parentheses.

equation (4) was set to 5 so that error estimates of 0·5 Å and 1 Å in the distances correspond to force constants of 24·9 and 6·23 kJ mol^{-1} $Å^{-2}$, respectively. These force constants are relatively weak, and the choice of distance ranges and errors is such that there is sufficient overlap between adjacent distance ranges to ensure that the uncertainties in the distance measurements will not have a strong effect on the end result.

The strategy we have employed to test the restrained molecular dynamics method is the following.

First, three regular secondary structure types, namely an α-helix (initial I), an extended β-strand (initial II) and a 3–10 helix (initial III) were generated using the interactive computer graphics program FRODO (Jones, 1978, 1982). The ϕ, ψ angles were −57°, −47° for the α-helix, −119°, 113° for the β-strand and −15°, −70° for the 3–10 helix. The difference between these structures is sufficiently large to provide a good test of the global convergence properties of the method. This is true particularly for the β-strand starting structure, since the energy barrier for the conversion of a β-strand to either an α-helix or a 3–10 helix is quite large (Ramachandran *et al.*, 1966; Ramachandran & Sasisekharan, 1968; Schultz & Schirmer, 1979).

Second, the three initial structures were subjected to 1000 cycles of restrained energy minimization (RM) (with $S = 5$ in eqn (4)) to generate structures RMI, RMII and RMIII and 1000 cycles of free energy minimization (FM) (i.e. without the interproton distance restraints) to generate structures FMI, FMII and FMIII.

Third, restrained dynamics calculations were carried out starting from structures RMI, RMII and RMIII, and free dynamics calculations from structures FMI, FMII and FMIII. The dynamics calculations were divided into six parts:

(1) Two picoseconds of equilibration, during which time the structure was heated up from 200 K to 300 K in steps of 10 K every 0·2 picosecond. In the restrained equilibration S was initially set to 0·25 and increased by 0·25 every 10 K up to a value of 1·0; for the unrestrained runs, S was set to zero.

(2) Two picoseconds of thermalization in which the initial velocities at 300 K were reassigned every 0·1 picosecond. The value of S was kept at 1·0 during this phase of the restrained dynamics.

(3) Ten picoseconds of dynamics (referred to as the first dynamics run) without adjusting the temperature of the system. For this phase, and all further phases of the restrained dynamics, S was set to 5·0. Considerable heating occurred during this step of the restrained dynamics trajectory. Since the interproton distance restraints are poorly satisfied in the initial structure, the reduction in the interproton restraint energy is accompanied by a corresponding increase in kinetic energy; this occurs within three picoseconds. In the case of the α-helix and 3–10 helix, the temperature increased to 600 K and in the case of the β-strand to 1100 K at three picoseconds and remained stable thereafter. Because of the harmonic bonding terms in the empirical energy function, these high temperatures do *not* result in any bond breaking nor in any significant distortion of bond angles. However, there are changes in the softer degrees of freedom (e.g. torsional angles). In the free dynamics, the heating effect, as expected, was much less

Table 4

The r.m.s. difference for all atoms and for the backbone atoms between the initial (IniI, IniII, InIII), free (FMI, FMII, FMIII) and restrained (RMI, RMII, RMIII) energy minimized, the average free (FDI, FDII, FDIII) and restrained (RDI, RDII, RDIII) dynamics structures

	Overall r.m.s. difference (Å)														
	InI	InII	InIII	FMI	FMII	FMIII	RMI	RMII	RMIII	FDI	FDII	FDIII	RDI	RDII	RDIII
InI		9·60	3·77	0·53	9·58	3·70	0·58	9·07	3·80	2·56	5·90	3·55	3·52	5·90	3·55
InII	8·67		7·80	9·77	0·45	8·02	9·67	1·49	9·23	9·54	8·75	9·23	9·24	9·74	9·83
InIII	2·60	6·64		3·94	7·78	0·67	3·89	7·33	0·88	4·43	5·65	3·79	4·57	5·43	5·22
FMI	0·32	8·82	2·76		7·77	3·85	0·51	9·24	3·95	2·51	5·99	3·58	3·59	4·29	3·59
FMII	8·67	0·37	6·63	8·81		8·00	9·64	1·50	7·88	9·50	8·72	9·20	9·22	9·72	9·81
FMIII	2·52	6·84	0·60	2·65	6·83		3·80	7·56	0·69	4·32	5·64	3·67	4·47	5·64	5·11
RMI	0·37	8·62	2·58	0·40	8·62	2·47		9·14	3·86	2·55	6·02	3·63	3·55	4·35	3·59
RMII	8·06	1·58	6·08	8·19	1·63	6·28	8·02		7·42	9·00	8·22	8·74	8·73	9·24	9·34
RMIII	2·59	6·67	0·68	2·73	6·66	0·45	2·53	6·11		4·43	5·76	3·82	4·50	5·42	5·20
FDI	0·95	8·78	2·83	0·88	8·78	2·74	0·95	8·13	2·80		5·86	3·21	2·85	3·59	3·26
FDII	4·74	7·62	4·26	4·76	7·62	4·25	4·76	6·93	4·24	4·82		5·82	6·09	6·02	5·92
FDIII	2·38	8·27	2·69	2·36	8·26	2·59	2·32	7·67	2·64	2·07	4·37		3·80	4·18	4·11
RDI	1·51	8·43	2·77	1·62	8·43	2·70	1·55	7·82	2·72	1·61	4·66	1·98		2·71	2·84
RDII	2·35	8·88	3·53	2·32	8·88	3·45	2·40	8·24	3·50	2·04	5·08	2·35	1·79		3·11
RDIII	1·77	8·90	3·28	1·80	8·89	3·18	1·84	8·27	3·22	1·72	4·99	2·78	1·45	1·89	

Backbone r.m.s. difference (Å)

The symbols I, II and III refer to the structures obtained starting off from the α-helix, β-strand and 3-10 helix initial structures, respectively. The average dynamics structures are the averages over the last 8 ps of the second dynamics run (see the text).

pronounced, the temperature rising only to 400 K, suggesting that the original equilibration time was too short.

(4) A cooling step in which the structures were cooled down to 300 K at a rate of 10 K every 0·2 picosecond (except in the case of the β-strand restrained dynamics, where cooling was carried out at a rate of 10 K every 0·1 picosecond).

(5) Two picoseconds of rethermalization, in which the initial velocities at 300 K were reassigned every 0·1 picosecond.

(6) Ten picoseconds of further dynamics (known

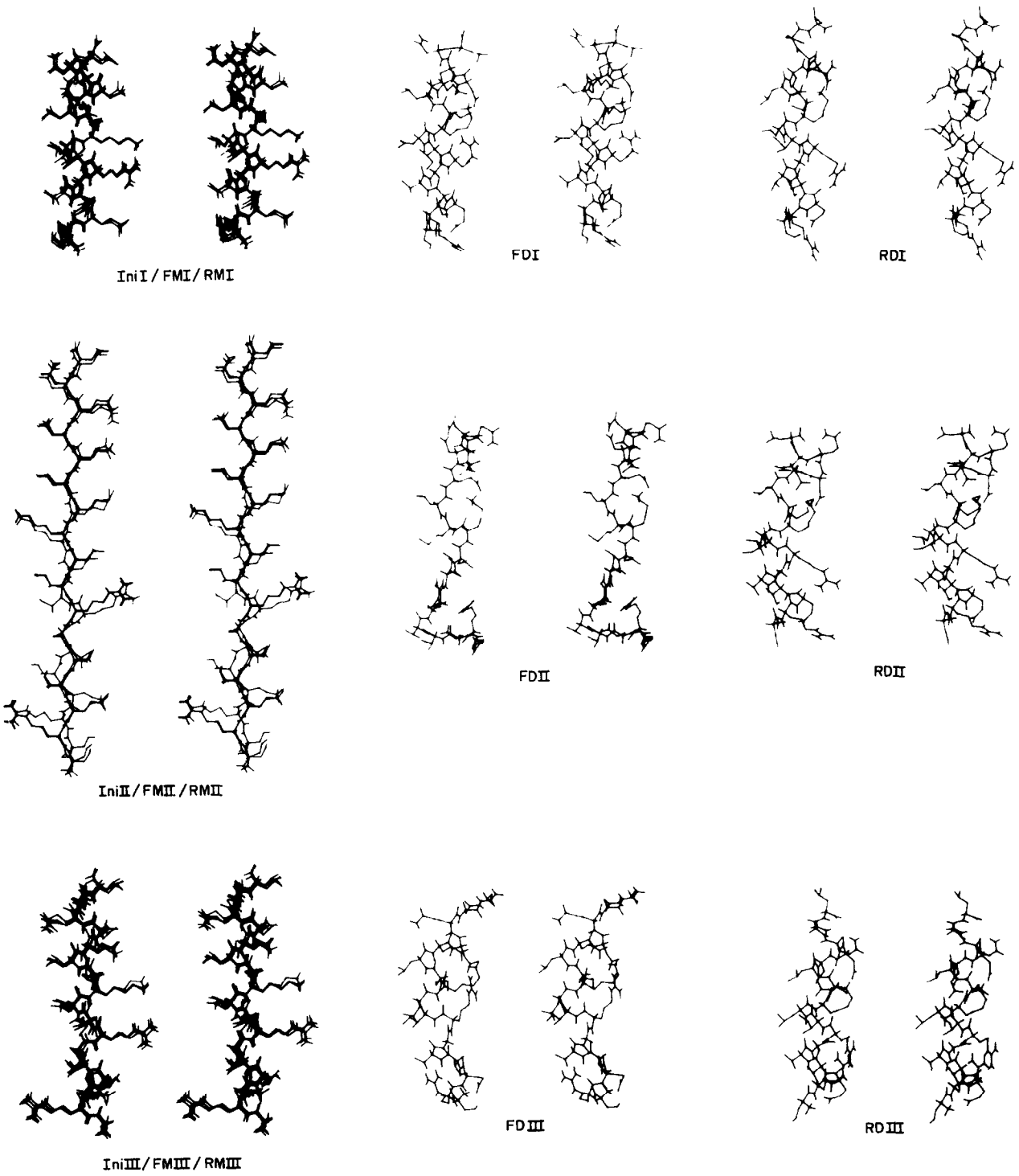


Figure 3. Stereoviews of the initial (IniI, IniII, IniIII), free (FMI, FMII, FMIII) and restrained (RMI, RMII, RMIII) energy minimized, and average free (FDI, FDII, FDIII) and restrained (RDI, RDII, RDIII) dynamics structures. The symbols I, II and III refer to structures obtained starting off from α-helix, β-strand and 3-10 helix initial structures, respectively. For the 3 sets, the initial and the free and restrained energy minimized structures are superimposed to show that both the free and restrained energy minimizations result in only very small changes from the initial structures.

as the second dynamics run). In this run, the temperature in all cases remained stable around 300 K. The final averaged structures for the restrained (RDI, RDII and RDIII) and free (FDI, FDII and FDIII) dynamics were obtained by averaging the co-ordinate trajectories over the last eight picoseconds of this run.

The energies of the initial, free and restrained energy minimized, and average free and restrained dynamics structures are given in Table 3; the r.m.s. difference over all atoms and over backbone atoms between these structures are given in Table 4. Stereoviews of the structures are shown in Figure 3 and r.m.s. differences between all atoms, backbone atoms and side-chain atoms as a function of residue number are shown in Figures 4 and 5 for some combinations involving the initial structures and the average free and restrained dynamic structures. Several interesting features emerge:

(1) Neither restrained nor free energy minimization results in any gross alteration of the initial structures (see particularly Fig. 3). However, it should be noted that the internal energy, which is very high in the model-built structures is reduced to a satisfactory value by free energy minimization, with only relatively small changes in the local structure. By contrast, in the presence of the NOE restraints, the energies of all the structures remain rather high after minimization; this is particularly evident in the β-strand structure. Apparently the NOE restraints have prevented the structural adjustments required to obtain satisfactory van der Waals' and electrostatic energies. Moreover, restrained minimization results in only a small improvement in the restraints potential in the cases of the 3–10 helix and β-strand, and a slight increase in the restraint potential energy in the case of the α-helix. These results suggest that the model-built structure, even for the α-helix, is in the neighbourhood of a local energy minimum that is inconsistent with the NOE restraints and from which the system cannot escape by energy minimization techniques.

(2) The free dynamics do produce structural changes. The α-helix remains an α-helix, but alterations in side-chain positions occur. The 3–10 helix is converted to an α-helix, in part because the van der Waals' interactions are more favourable in an α-helix relative to those of a 3–10 helix; also it is known that 3–10 helices are generally stabilized by interactions with bound water molecules (Rose *et al.*, 1983). However, the r.m.s. difference between FDI and FDIII is still large (3·2 Å for all atoms;

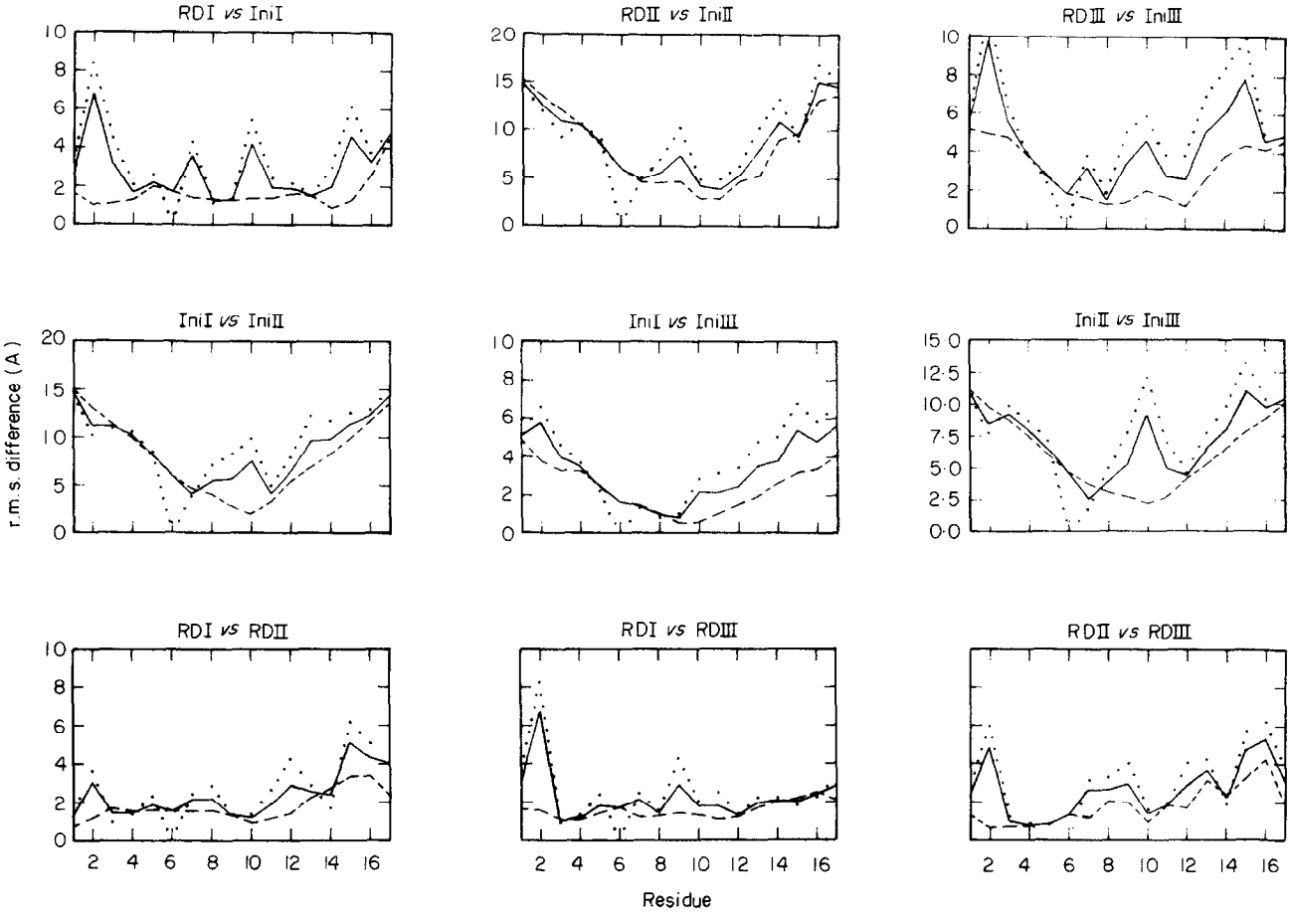


Figure 4. The r.m.s. difference (Å) for all (——), backbone (- - - -) and side-chain (· · · · ·) atoms as a function of residue number for various pairs of structures involving the initial (IniI, IniII, IniIII) and average restrained dynamics (RDI, RDII, RDIII) structures.

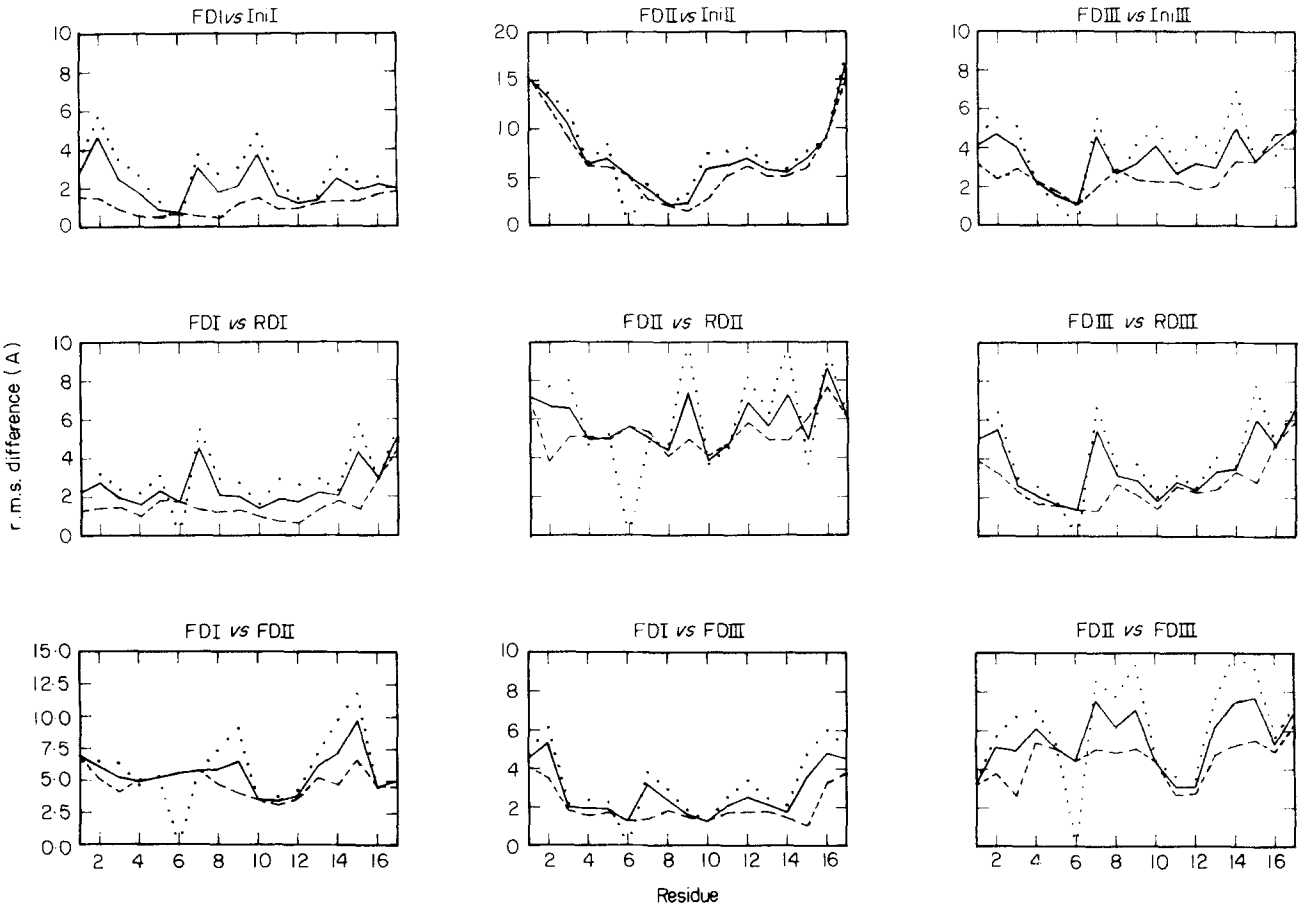


Figure 5. The r.m.s. difference (Å) for all (——), backbone (- - - -) and side-chain (· · · · ·) atoms as a function of residue number for various pairs of structures involving the initial (IniI, IniII, IniIII), average free dynamics (FDI, FDII, FDIII) and average restrained dynamics (RDI, RDII, RDIII) structures.

2·1 Å for the backbone atoms) and not significantly reduced over that between the initial structures I and III (3·8 Å for all atoms; 2·6 Å for the backbone atoms). In the case of the β-strand (FDII), something rather different happens: the structure is compressed and becomes locked in an essentially random coil configuration (also see Fig. 7 and the Ramachandram plot in Fig. 8). In all three cases, however, little improvement occurs in the agreement with the experimental interproton distances.

(3) In all three cases, the restrained dynamics result in convergence to an α-helical structure satisfying all experimental interproton distance restraints, within the error limits of their estimation. The similarity between the average restrained structures is highlighted by Figure 6, which shows stereoviews of the C^{α}, the backbone and all atoms of the three structures superimposed. The convergence in backbone conformations is also illustrated by a plot of the ϕ and ψ angles of the structures as a function of residue number (Fig. 7). For the initial α-helix, the dihedral angle plots indicate that changes in the backbone are small. Comparison of the ϕ and ψ angles of RDI and RDIII only shows a significant difference at residue 7 for ϕ and residue 6 for ψ. Comparison of the ϕ and ψ angles of RDII *versus* the other two structures only shows a significant difference at residues 14 and 15 for ϕ, and at residues 11, 13 and 14 for ψ. These are apparent on examination of Figure 6(a) and (b). The convergence to an α-helix is also illustrated by a Ramachandran plot in Figure 8.

(4) The time period over which major structural changes occur during the first dynamics run is much shorter for the restrained than for the free dynamics. This is illustrated by snapshots of the trajectories in Figures 9 and 10 for the two most dramatic cases, namely the first restrained and free dynamics runs, respectively, starting from the β-strand (initial structure II). In the restrained dynamics, the transition to a structure with the length of an α-helix is achieved within 1·5 picoseconds. In contrast, significant changes in structure in the free dynamics occur only after about four picoseconds. However, rather large structural oscillations continue to take place in the restrained structure over the remainder of the trajectory and the final structure at 1100 K is different from the RDII result.

(5) The dynamics provide information not only on structure but also on mobility. This is illustrated

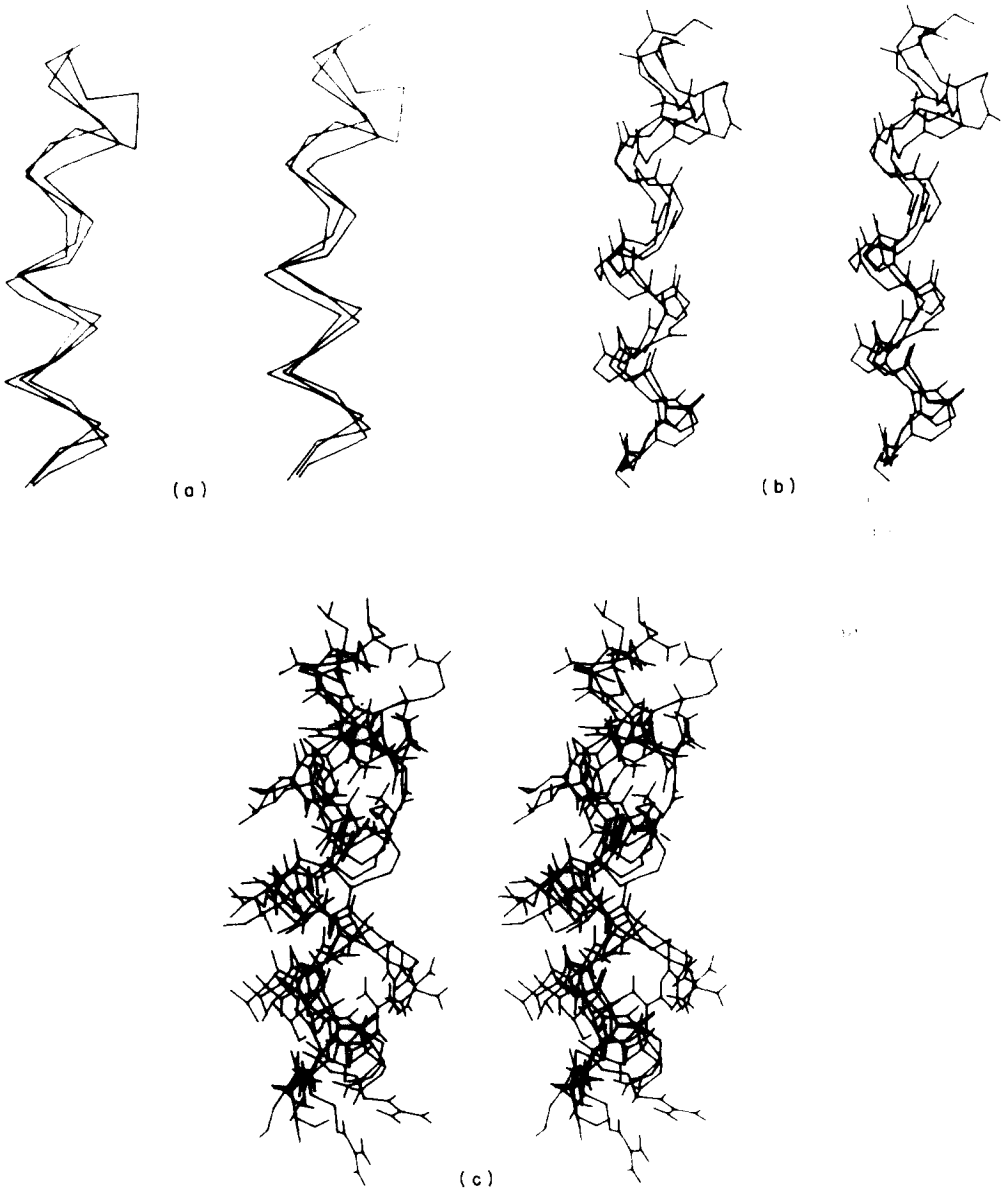


Figure 6. Best fit superposition of (a) the C^{α} atoms (connected by virtual bonds), (b) the backbone atoms (excluding protons) and (c) all atoms of the 3 average restrained dynamics structures, RDI, RDII and RDIII.

in Figure 11 by a plot of r.m.s. fluctuations in backbone, side-chain and all atoms as a function of residue number for the six average dynamics structures. As expected, the mobility of the side-chains is slightly higher than that of the backbone. The magnitude of the r.m.s. fluctuations (between 0·5 Å and 1 Å) are broadly similar for all the structures, with the exception of FDII, where they are much larger (between 1 Å and 3 Å). This is not surprising as FDII is close to a random coil and has some drifting motions superimposed on the local fluctuations of the atoms. A visual impression of the motions in structures RDI, RDII, RDIII and FDII is given in Figure 12, where stereoviews of snapshots taken at two, four, six, eight and ten picoseconds in the second dynamics run are superimposed for each structure. A comparison of Figure 12 with Figure 6(c) also reveals that the magnitude of the motions in the average restrained dynamics structures, although slightly smaller, is not very different from the magnitude of the difference between the three average restrained dynamics structures.

(e) *Structural features of the average restrained dynamics structures*

A striking feature of this α-helical 17-mer peptide is that it possesses two distinct faces, namely a hydrophobic face and a hydrophilic face. This is illustrated in Figure 13 by a space-filling representation of the average restrained dynamics structure RDI. This makes excellent sense in terms of the position of helix F in CRP: the hydrophobic face of the helix is directed towards the protein

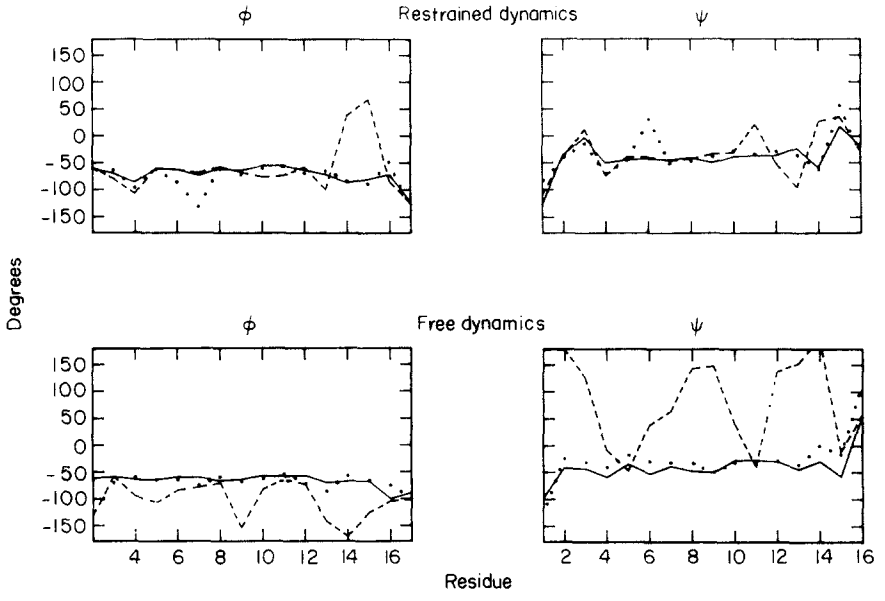


Figure 7. ϕ and ψ angles plotted as a function of residue number for the average restrained and free dynamics structures. (——), RDI and FDI; (-----), RDII and FDII; (· · · · ·), RDIII and FDIII.

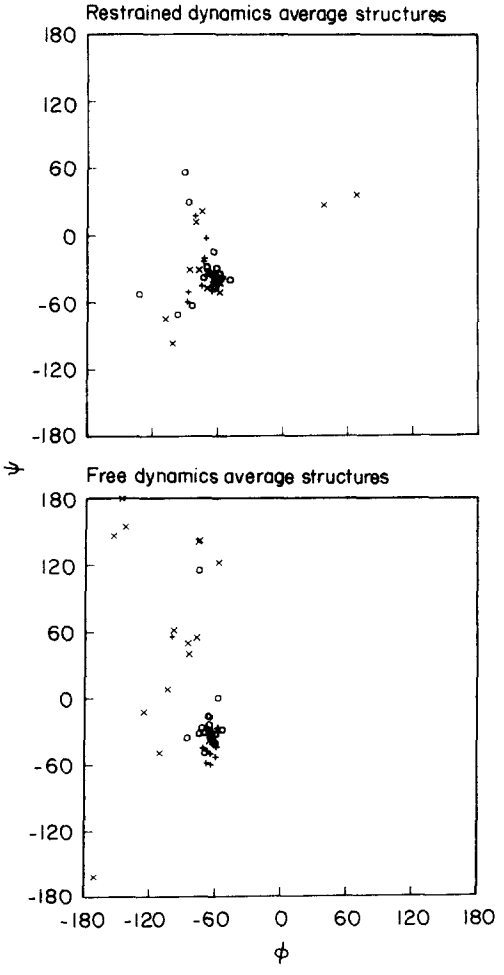


Figure 8. Ramachandran (ϕ, ψ) plot for the average restrained and free dynamics structures. Symbols: (+), RDI and FDI; (×), RDII and FDII; (○), RDIII and FDIII.

interior, whereas the hydrophilic face is exposed to the solvent and free to interact with a DNA target site.

The helix itself is stabilized by a network of polar and non-polar interactions. The hydrophobic interactions involving methyl groups occur between the following pairs of residues: Thr4 and Ile8, Val5 and Ile8, Ile8 and Met11, Leu9 and Leu12, and Met11 and Leu12. Interestingly, these residues occur in relative positions 1, 2, 5 (Thr4, Val5 and Ile8; Ile8, Leu9 and Leu12) and 1, 4, 5 (Ile8, Met11 and Leu12). This is in complete agreement with statistical analyses of non-polar group distributions in proteins which have shown that the occurrence of non-polar triplets in helices with these relative positions is significantly higher than would be expected on the basis of amino acid composition (Lim, 1974*a*,*b*; Palau & Puigdomenech, 1974). In addition to the backbone hydrogen bonding involving the backbone CO and NH groups, the polar interactions common to all three average restrained dynamics structures are as follows: between the terminal NH_3^+ of Ser1 and the carboxylate of Glu3, the backbone CO of Glu3 and the guanidinium group of Arg7, the $N^{\epsilon}H_3^+$ of Lys11 and the carboxylate groups of Glu13 and Asp14, and between the backbone CO of Glu13 and the $N^{\delta}H_2$ of Asn16. Additional interactions include those between the O^{δ} of Ser1 and the $N^{\epsilon}H$ and guanidinium group of Arg2 in RDI, the $O^{\gamma}H$ of Ser1 and the carboxylate of Glu3 in RDI and RDII, the $N^{\epsilon}H$ of Arg2 and the carboxylate of Glu3 in RDI, the backbone CO of Glu3 and the $N^{\epsilon}H$ of Arg7 in RDI and RDIII, the $O^{\gamma 1}H$ of Thr4 and the carboxylate of Glu3 in RDIII, and the $N^{\epsilon}H_2$ of Glu15 and the C-terminal carboxylate of Leu17 in RDII.

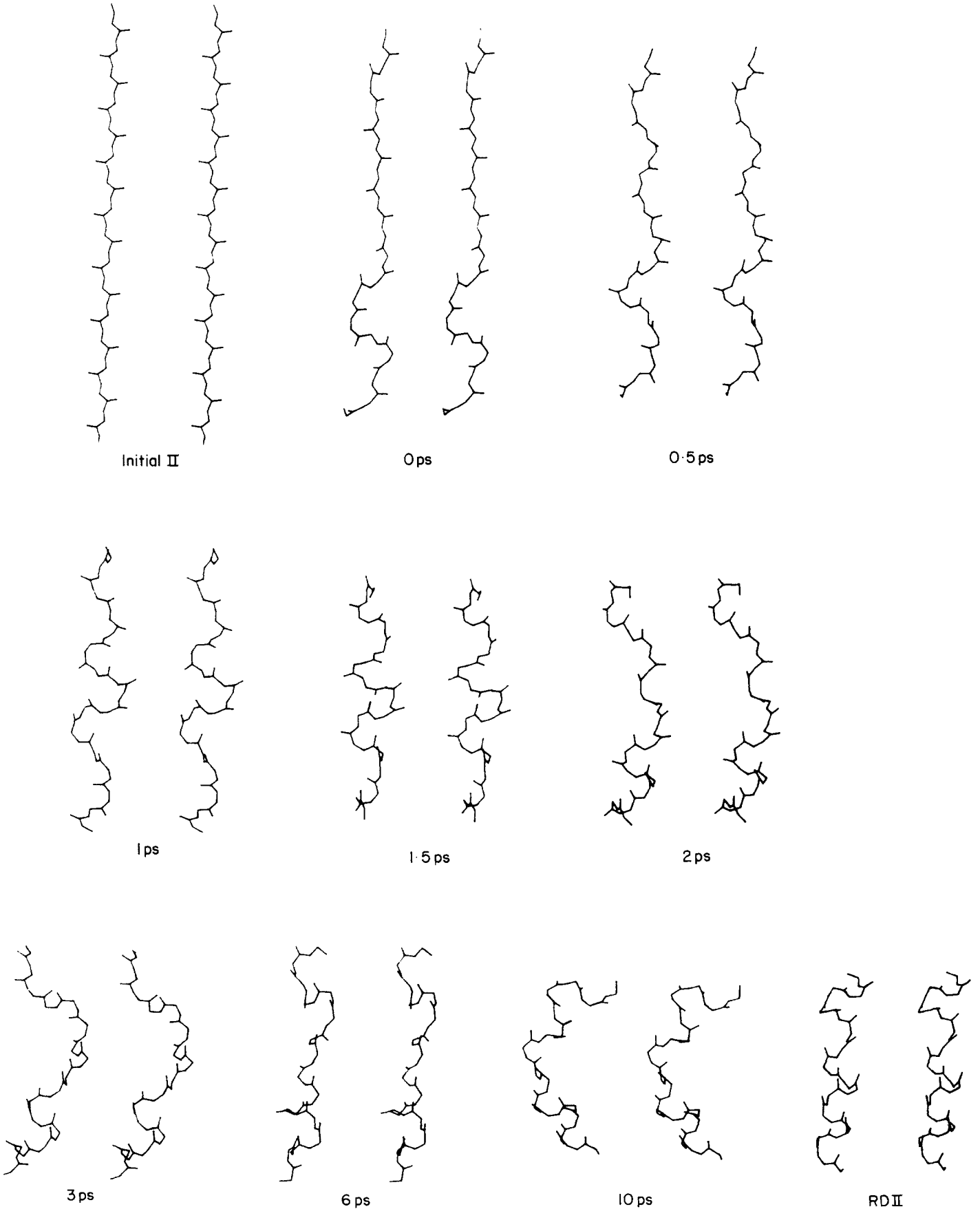


Figure 9. Snapshots of the trajectory of the first restrained dynamics run showing the conversion of the β-strand (initial structure II) to an α-helix. The structure at 0 ps is the structure obtained after equilibration and thermalization. RDII is the average restrained dynamics structure of the second dynamics run. For the sake of clarity only the backbone atoms (excluding protons) are shown.

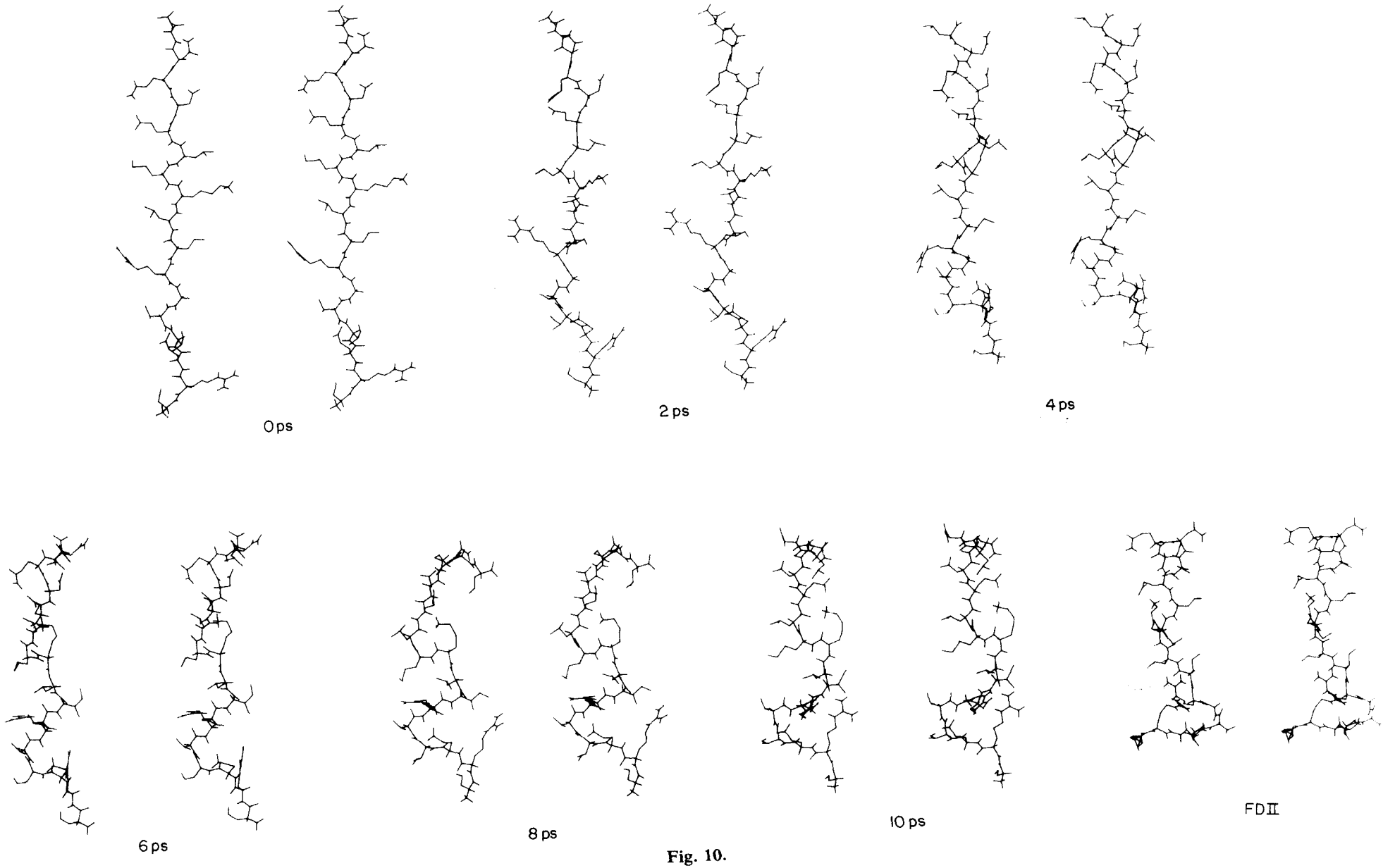


Fig. 10.

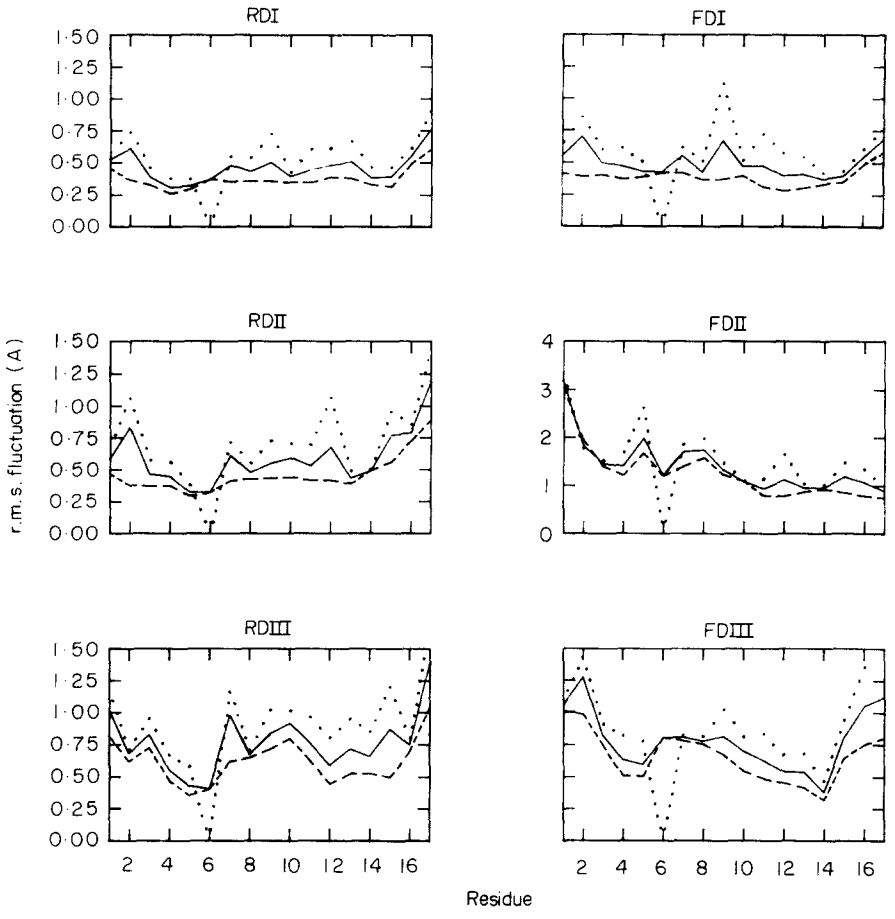


Figure 11. The r.m.s. fluctuations (Å) of all (——), backbone (-----) and side-chain (· · · · ·) atoms as a function of residue number for the average restrained (RDI, RDII, RDIII) and free (FDI, FDII, FDIII) dynamics structures.

There are two other interesting structural features of the average restrained dynamics structures. First, two of the structures, namely RDI and RDII, are bent with the hydrophobic face on the convex surface and the hydrophilic face on the concave surface, whereas the third structure, RDIII, is straight. This is best seen in Figures 12 and 13. The bending of helices in proteins has been discussed by Blundell *et al.* (1983). Second, the orientation of the side-chains of Ser1 and Arg2 are quite different in RDI and RDII on the one hand and RDIII on the other. In RDI and RDII the side-chain of Ser1 is directed towards the side-chains of Arg2 and Glu3, whereas in RDIII it is directed away from them. These two conformations of Ser1 may reflect the two slow-exchanging conformations of Ser1 observed in the n.m.r. studies. In addition, whereas the side-chain of Arg2 is pointed towards the N terminus in RDI and RDII, it is directed towards the C terminus in RDIII, with its guanidinium group making an electrostatic interaction with the backbone CO of Gly6 (note the oxygen of the latter no longer points in the direction of the helix axis but is at an angle of approximately 70° to it in RDIII: this is reflected by deviations in the ϕ and ψ angles involving Gly6; see Fig. 7). Since there were no interproton distance restraints involving the side-chains of these two residues, these differences arise from energy terms intrinsic to the molecular dynamics calculation itself, in particular the electrostatic and van der Waals' interactions.

4. Concluding Remarks

In this paper we have demonstrated the power of restrained molecular dynamics, based on interproton distance restraints derived from NOE measurements, as a method of three-dimensional structure determination in solution. Within the limits of the rather imprecise distant restraints used, the initial starting conformation (α-helix, β-strand or 3–10 helix) has little influence on the

Figure 10. Snapshots of the trajectory of the first free dynamics run showing the conversion of the β-strand (initial structure II) to a compressed random coil. The structure at 0 ps is obtained after equilibration and thermalization. FDII is the average free dynamics structure of the second dynamics run.

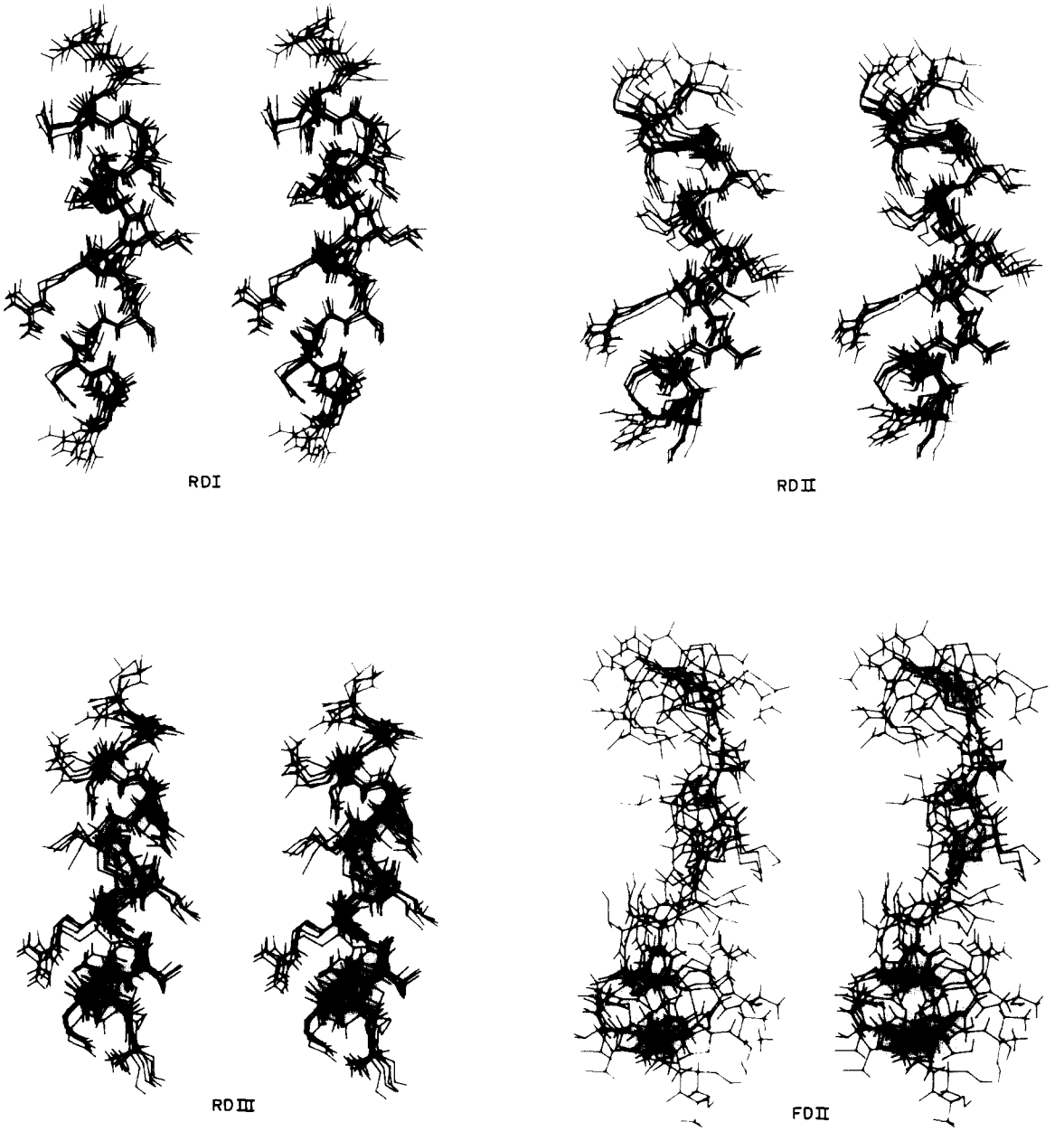


Figure 12. Best fit superposition of the structures at 2, 4, 6, 8 and 10 ps of the second dynamics run for the restrained dynamics structures (RDI, RDII, RDIII) and 1 free dynamics structure (FDII).

final global state (α-helix). This contrasts with the results of energy minimization. However, it should be noted that the present study concerns itself with a particularly good case for such global convergence, since the distance restraints from the NOE data are very effective in defining α-helices (Wüthrich *et al.*, 1984). Also, it must be mentioned that there is no independent evidence that the details of the predicted structure are correct. Not surprisingly, the positions of the backbone atoms are better defined than those of the side-chains, as relatively few distance constraints involving side-chain atoms were used. This situation might improve when applying the method to whole proteins, as the available positions of side-chains within the protein interior would be limited by the strict packing requirements and non-bonding interactions on the one hand and by NOE restraints connecting neighbouring side-chains in space on the other.

Inspection of the energy terms of the average restrained dynamics structures reveals that the energies increase in the order RDI < RDIII < RDII. (Note that this could also indicate a not as yet complete equilibration that may take longer than 20 ps.) That is to say, the closer the initial structure to the actual structure, the lower the energy terms (both overall and individual) of the average restrained dynamics structure. This may be of significance when the method is applied to whole proteins (e.g. as in the case of the *lac* repressor head piece; Kaptein *et al.*,

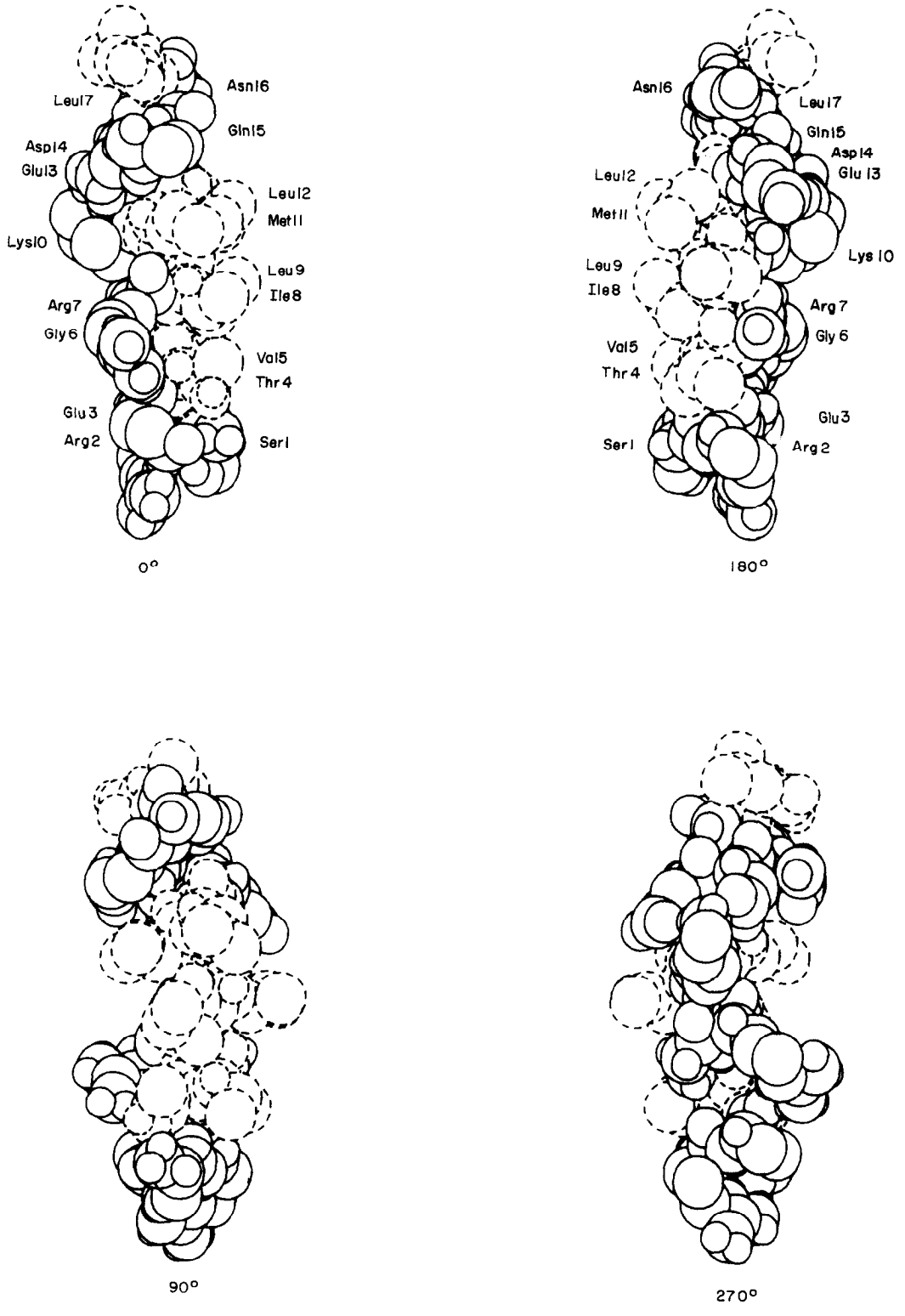


Figure 13. Views of a space-filling representation of the average restrained dynamics structure RDI, illustrating the 2 distinct faces of the helix, namely the hydrophobic and hydrophilic faces. The atoms of those residues whose side-chains have methyl groups (namely, Thr4, Val5, Ile8, Leu9, Met11, Leu12 and Leu17) are represented by broken circles, whereas the atoms of all other residues are represented by continuous circles. The 4 views represent successive rotations of 90° about the helix axis. These pictures were produced by a computer program written by Lesk & Hardman (1982).

1985). In such cases, convergence would be expected to be most efficient if a reasonable initial structure derived by model building, rather than some random structure, were used to start the calculation. This may be possible in proteins composed mainly of α-helices and β-sheets, since such secondary structure elements are easily identified by visual inspection of the NOESY

spectra (Wüthrich *et al.*, 1984). Moreover, under ideal conditions, it is possible that the approximate three-dimensional structure of the protein can then be deduced by interactive molecular graphics on the basis of the observed long-range contacts (i.e. short interproton distances between residues far apart in the sequence) (Zuiderweg *et al.*, 1984). However, the uniqueness and accuracy of such structure determinations has yet to be determined.

What are the relative merits of the restrained dynamics and distance geometry algorithm approaches? A disadvantage of the latter is that given the availability of only a set of either rather imprecise or incomplete distance restraints, the structures obtained may possess energetically unfavourable features. This is particularly so as long as the distance geometry algorithms do not take into account non-bonded (*viz.* van der Waal's and electrostatic) interactions and torsion potentials. Moreover, the subsequent application of energy minimization (either free or restrained) to the structures may not be fruitful owing to the limited global convergence properties of the energy minimization algorithms. However, since the distance geometry algorithms are relatively fast, trials with several initial structures can be done easily. Restrained molecular dynamics, on the other hand, takes into account all energy terms, can overcome energy barriers between false minima and is capable of converging to structures with lower energies. It suffers from the fact that it is expensive in terms of computer time. Thus, for the small system of only 17 amino acid residues and 195 atoms presented here, each complete dynamics run took approximately 30 min on a CRAY 1 computer, which is equivalent to about 35 to 40 hours on a VAX 11/780 computer. It is therefore clear that even for small proteins it would only be feasible to start from a few initial structures, given the present speed and capabilities of computers commonly available for these calculations. However, if one is interested in refining only a particular part of a larger structure such as, for example, the DNA binding domain of CRP bound to DNA, use of the recently developed stochastic molecular dynamics method (Brooks & Karplus, 1983; Brunger *et al.*, 1984; A. Brunger, C. L. Brooks & M. Karplus, unpublished results; Brooks *et al.*, 1985) in combination with the NOE restrained effective potential could greatly reduce the computing time for such larger systems.

We thank the Max-Planck Institut für Plasma Physik (Garching) for computing facilities on the CRAY 1 computer. We also thank Dr S. R. Martin for help with the c.d. spectra, Dr B. M. Pettit for assistance in modifying the CHARMM potential and Dr J. Pflugrath for discussion and criticism. This work was supported by the Max-Planck-Gesellschaft (GMC and AMG).

References

Anil Kumar, Ernst, R. R. & Wüthrich, K. (1980). *Biochem. Biophys. Res. Commun.* **95**, 1–6.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981). *Nature (London)*. **220**, 754–759.

Arseniev, S. A., Kondakov, V. I., Maiorov, V. N. & Bystrov, V. F. (1984). *FEBS Letters*, **165**, 57–62.

Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976). *J. Chem. Phys.* **64**, 2229–2246.

Bauman, R., Wider, G., Ernst, R. R. & Wüthrich, K. (1981). *J. Magn. Reson.* **44**, 402–406.

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

Blundell, T., Barlow, D., Borkakoti, N. & Thornton, J. (1983). *Nature (London)*. **306**, 281–283.

Bodenhausen, G., Vold, R. L. & Vold, R. R. (1980). *J. Magn. Reson.* **37**, 93–106.

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., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.

Brooks, C. L. & Karplus, M. (1983). *J. Chem. Phys.* **79**, 6312–6325.

Brooks, C. L., Brünger, A. & Karplus, M. (1985). *Biopolymers*, **24**, 843–865.

Brünger, A., Brooks, C. L. & Karplus, M. (1984). *Chem. Phys. Letters*, **105**, 495–500.

Clore, G. M. & Gronenborn, A. M. (1984*a*). *Eur. J. Biochem.* **141**, 119–129.

Clore, G. M. & Gronenborn, A. M. (1984*b*) *FEBS Letters*, **172**, 219–225.

Clore, G. M. & Gronenborn, A. M. (1984*c*) *FEBS Letters*, **175**, 117–123.

Clore, G. M. & Gronenborn, A. M. (1985*a*). *EMBO J.* **4**, 829–835.

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

Clore, G. M., Gronenborn, A. M., Piper, E. A., McLaughlin, L. W., Graeser, E. & van Boom, J. H. (1984*a*). *Biochem. J.* **221**, 737–750.

Clore, G. M., Gronenborn, A. M. & McLaughlin, L. W. (1984*b*). *J. Mol. Biol.* **174**, 163–173.

Clore, G. M., Gronenborn, A. M., Moss, D. S. & Tickle, I. J. (1985*a*). *J. Mol. Biol.* **185**, 219–226.

Clore, G. M., Gronenborn, A. M. & McLaughlin, L. W. (1985*b*). *Eur. J. Biochem.* **151**, 153–165.

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

de Crombrugghe, B., Busby, S. & Buc, H. (1984). In *Biological Regulation and Development* (Yamamoto, K., ed.), vol. 3B, pp. 129–167, Plenum Press, New York.

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

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984*a*). *Nature (London)*, **311**, 232–235.

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984*b*). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 7274–7278.

Eich, G., Bodenhausen, G. & Ernst, R. R. (1982). *J. Amer. Chem. Soc.* **104**, 3731–3732.

Gelin, B. R. & Karplus, M. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 2002–2006.

Gronenborn, A. M. & Clore, G. M. (1982). *J. Mol. Biol.* **157**, 155–160.

Gronenborn, A. M., Clore, G. M. & Jeffery, J. (1984*a*). *J. Mol. Biol.* **172**, 559–572.

Gronenborn, A. M., Clore, G. M., McLaughlin, L. W., Graeser, E., Loiber, B. & Giege, R. (1984*b*). *Eur. J. Biochem.* **145**, 359–364.

Gronenborn, A. M., Clore, G. M. & Kimber, B. J. (1984*c*). *Biochem. J.* **221**, 723–736.
Havel, T. F. & Wüthrich, K. (1984). *Bull. Math. Biol.* **46**, 673–698.
Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 281–294.
Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). *J. Chem. Phys.* **71**, 4546–4553.
Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.
Jones, T. A. (1982). In *Computational Crystallography* (Sayre, D., ed.), pp. 303–317, 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. (1983). *Annu. Rev. Biochem.* **52**, 263–300.
Keepers, J. W. & James, T. L. (1984). *J. Magn. Reson.* **53**, 104–124.
Kuntz, I. D., Crippen, G. M. & Kollman, P. A. (1979). *Biopolymers*, **18**, 939–957.
Lesk, A. M. & Hardman, K. D. (1982). *Science*, **216**, 539–540.
Lim, V. I. (1974*a*). *J. Mol. Biol.* **88**, 857–872.
Lim, V. I. (1974*b*). *J. Mol. Biol.* **88**, 873–894.
Macura, S. & Ernst, R. R. (1980). *Mol. Physiol.* **41**, 95–117.
Macura, S., Huang, Y., Suter, D. & Ernst, R. R. (1981). *J. Magn. Reson.* **43**, 259–281.
Marion, D. & Wüthrich, K. (1983). *Biochem. Biophys. Res. Commun.* **113**, 967–974.
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.
McKay, D. & Steitz, T. (1981). *Nature (London)*, **290**, 744–749.
McKay, D., Weber, I. & Steitz, T. (1982). *J. Biol. Chem.* **257**, 9518–9524.
Nagayama, K., Anil Kumar, Wüthrich, K. & Ernst, R. R. (1980). *J. Magn. Reson.* **40**, 321–334.
Noggle, J. H. & Schirmer, R. E. (1971). *The Nuclear Overhauser Effect – Chemical Applications*, Academic Press, New York.
Pabo, C. O. & Lewis, M. (1982). *Nature (London)*, **298**, 433–446.
Palau, J. & Puigdomenech, P. (1974). *J. Mol. Biol.* **88**, 457–469.
Poulsen, F. M., Hosch, J. C. & Dobson, C. J. (1980). *Biochemistry*, **19**, 2596–2607.
Provencher, S. W. & Glockner, J. (1981). *Biochemistry*, **20**, 33–37.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Advan. Protein Chem.* **23**, 283–437.
Ramachandran, G. N., Venkatachalam, C. M. & Krimm, S. (1966). *Biophys. J.* **6**, 849–872.
Redfield, A. G. & Gupta, K. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 405–419.
Redfield, A. G. & Kunz, S. D. (1975). *J. Magn. Reson.* **19**, 250–254.
Rose, G. D., Young, W. B. & Gierasch, L. M. (1983). *Nature (London)*, **304**, 654–657.
Schulz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, Springer-Verlag, New York.
Senn, H., Billeter, M. & Wüthrich, K. (1984). *Eur. Biophys. J.* **11**, 3–15.
Štrop, P., Wider, G. & Wüthrich, K. (1983). *J. Mol. Biol.* **166**, 641–667.
Verlet, L. (1967). *Physiol. Rev.* **159**, 98–105.
Wagner, G. (1983). *J. Magn. Reson.* **55**, 151–156.
Wagner, G. & Wüthrich, K. (1979). *J. Magn. Reson.* **33**, 675–680.
Wagner, R. & Wüthrich, K. (1982). *J. Mol. Biol.* **160**, 343–361.
Wako, H. & Scheraga, H. A. (1982). *J. Protein Chem.* **1**, 85–117.
Weber, I. & Steitz, T. A. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3973–3977.
Wider, G., Macura, S., Anil Kumar, Ernst, R. R. & Wüthrich, K. (1984). *J. Magn. Reson.* **56**, 207–234.
Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.
Wüthrich, K. (1983). *Biopolymers* **22**, 131–138.
Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). *J. Mol. Biol.* **155**, 311–319.
Wüthrich, K., Billeter, M. & Braun, W. (1983). *J. Mol. Biol.* **169**, 949–961.
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.
Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wüthrich, K. & Kaptein, R. (1984). *FEBS Letters*, **174**, 243–247.

Edited by G. A. Gilbert