Three-dimensional structure of some synthetic calmodulin fragments using high resolution NMR, CD, IR and molecular modelling

L. Gomathi*†, Thomas Fairwell, Gopal A. Krishna**, James A. Ferretti**, and S. Subramanian***

*Regional Sophisticated Instrumentation Centre and Department of Chemistry, Indian Institute of Technology, Madras 600 036, India
†Present address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India
**National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA

High resolution two-dimensional NMR complemented by Fourier transform IR and circular dichroism has been employed for the elucidation of three-dimensional solution structure of several synthetic peptides corresponding to the calcium binding domains of the 148 residue protein calmodulin. The synthetic fragments correspond to the amino acid sequences in the calcium binding domain of the native protein as well as permutation and combination of loops and helices among the four binding sites. The results indicate that the characteristic helix–loop–helix (EF hand motif) is largely lost in fragmenting the peptide, although parts of the peptides show secondary structural characteristics that imply rapid equilibrium between loose helical and beta turn structure. The results also bring about a better ordering of the structure of calcium binding site I when the middle loop is exchanged for the loop in the site IV. Attempts have also been made to simulate the solution structure of some of the fragments using molecular modelling. The modelling studies also bring about a distinct increase in the secondary structural content upon exchanging the loop between calcium binding sites I and IV.

CALCIUM is involved in regulating a variety of cellular enzyme systems and in most types of cell motility[1–3]. Intracellular calcium levels control a variety of biochemical processes including all cytoplasmic movements that are mediated by myosin interaction. It is now established that calcium-binding proteins act as a mediators of calcium regulation in all calcium-dependent systems. Examples of some of the calcium-binding proteins are troponinC[4–6], parvalbumin[7,8], calmodulin (CaM)[9–11], intestinal calcium-binding protein[12,13], etc. The unique common feature to all the calcium-binding proteins is the presence of one or more of EF hand[14] structural motifs that is shown to regulate a variety of calcium-dependent biological functions. These motifs have now been identified in a score of different groups of proteins.

The protein CaM occupies a special position because of its high affinity and specificity for calcium binding and the fact that it is strictly conserved and functionally preserved throughout the animal and plant kingdom[15]. It serves as an intracellular calcium receptor, mediates calcium level in nucleotide and glycogen metabolism, secretion, motility and calcium transport. CaM is also a dynamic component of the mitotic apparatus. CaM exists as a monomer of molecular weight 16,700 D. It resists denaturation in the calcium-bound state. It is well established both by NMR[16,17] and X-ray crystallography[18,19] that it contains four calcium-binding sites. There is considerable controversy in the literature regarding the relative affinities of individual sites[20,21]. It is also known that binding of calcium to any one of the sites results in a definite conformational change. This conformational change is necessary in CaM to regulate many of the enzyme systems[15] such as cyclic nucleotide metabolism, protein phosphorylation, myosin light chain kinase, calcium flux, microtubule assembly/disassembly, etc. With a single bound calcium, CaM activates phosphodiesterase whereas microtubule depolymerization requires all the four calcium atoms. It is the ability to adopt different conformations depending upon different fractional occupancy of calcium-binding sites that is responsible for making CaM involved in such diverse biological processes. It has been shown from earlier studies of Kretsinger[7], that the calcium-binding site in all such proteins including CaM comprises of a characteristic helix–loop–helix motif (EF hand). Each motif contains two α-helices oriented at approximately 90°, linked by a loop with precisely spaced negatively charged residues, and a central glycine. The anionic residues and a main chain oxygen form a distorted pentagonal bypyramid which coordinates a calcium or a magnesium ion. It is also known that synthetic calcium-binding proteins that mimic EF hands assume similar structure in solution and also bind calcium with a low affinity. Sequence comparisons among calcium-binding

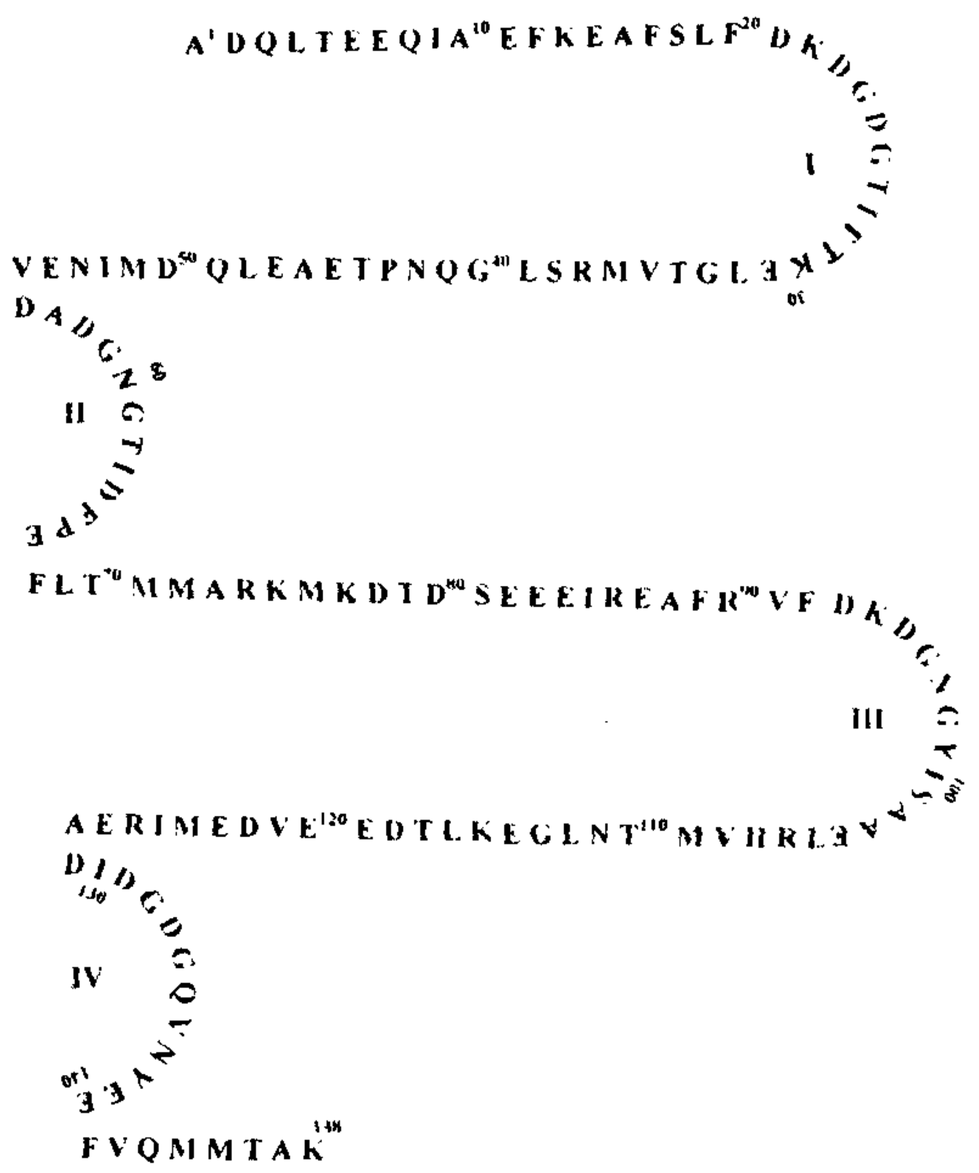


Figure 1. Schematic representation of the native calmodulin molecule showing the four calcium-binding sites I to IV.

proteins show a high homology in the calcium-binding region as well as close structural similarity in the EF hands. A schematic diagram of CaM is shown in Figure 1. All the four calcium-binding sites have different affinities for calcium as well as for other ions. The precise relative affinities for the four sites have not yet been established unequivocally. Different conclusions were reached using Tb^{3+}-mediated fluorescence[22–24], ^{1}H (ref. 25), ^{43}Ca (ref. 26), ^{113}Cd NMR (ref. 27), gel filtration and equilibrium dialysis methods[28–31].

Several attempts have been made using the conventional one-dimensional NMR to study the relative affinities of binding sites and to look at the conformational changes induced by calcium on both synthetic and native CaM fragments[32–37]. Early one-dimensional NMR studies were based on calcium-induced changes in chemical shifts of protons from the calcium-binding loop region as well as in one-dimensional difference NOE. It has been reported that resonances of aromatic and methyl region of CaM are very sensitive to calcium-induced conformational changes[38–40]. Calcium binding has also been studied by Dalgarno *et al.*[41,42] on two tryptic fragments namely TR1C and TR2C (residues 1–77 and 78–148 respectively) spanning the first and the second domains. One-dimensional NMR studies show that the two tryptic fragments are very similar to those in the native CaM. Also the calcium-free tryptic domains have similar structure to the calcium-free CaM. Analysis of chemical shift patterns and NOE showed that each half of the CaM can be modelled respectively to parvalbumin and intestinal calcium-binding protein.

Ikura[43] reported one-dimensional NMR studies on the kinetics of conformational change upon calcium binding. It was found that the N terminal and C terminal halves of the CaM have different calcium affinities and have different rates of conformational exchange, implying that the two halves play a different role in the mode of biological activity. It was also found that the rates of activation for conformational change, obtained by NMR were in agreement with previous studies based on ^{43}Ca NMR as well as stopped flow fluorescence studies[44]. Combined two-dimensional NMR and computer modelling techniques applied to the apo TR2C tryptic fragment of CaM enabled complete assignment of all the aromatic residues and high field shifted methyl groups[45]. The assignments were based on 2D NOE and computer modelling.

2D NMR studies[46] were also used to assign the extremely down field shifted amide proton resonances in the ^{1}H NMR spectrum of calcium-bound as well as calcium-free tryptic fragments of CaM. It was noted that the amide protons of Gly-25, Gly-61 and Gly-98 involved in sites I, II and III showed a low field shift whereas that of Gly-134 showed a different behaviour. This has been interpreted as due to positive calcium-binding cooperativity between sites III and IV.

Attempts have been made by Motta *et al.*[47] to look at calcium-dependent structural studies of synthetic analogs of calcium-binding loop I of bovine brain CaM. The linear and cyclic analogs of the dodecapeptide corresponding to residues 20–31 were studied using several one- and two-dimensional NMR techniques. Chemical shifts and $^{3}J_{\alpha CH\text{-}NH}$ indicated that, both the cyclic and linear peptides are flexible and assume multiple conformations in rapid equilibrium on the NMR time scale. Addition of calcium showed only minor changes in aqueous solution for both the peptides. However, the influence of calcium was enhanced in more hydrophilic mixtures such as TFE/water.

A complete three-dimensional structural elucidation of the *Drosophila* CaM has been reported by Ikura *et al.*[17] using 2D, heteronuclear 3D and isotope labelling techniques. Using heteronuclear triple resonance 3D NMR, they were able to obtain complete sequential assignments of the ^{1}H, ^{13}C and ^{15}N in the backbone. Their results showed that the secondary structure in the two globular domain of *Drosophila* CaM in solution is essentially identical to that of X-ray crystal structure of *mammalian* CaM[19] which consists of two pairs of helix–loop–helix motifs per domain. NMR results also confirmed the existence of a short antiparallel β-sheet between the two loops in each domain. The only difference between X-ray data and NMR results is the absence of a central helix from Phe-65 to Phe-92 found in the crystal structure.

Table 1. The synthetic calmodulin fragments studied in this work

Peptide	Amino acid sequence
Peptide 1–1–1 (MW = 3485) (32 residues)	^{10}AEFKEAFSLF^{19} ^{20}DKDGDGTITTKE^{31} ^{32}LGTVMRSLGQ$^{41(*)}$
Peptide 4–1–4 (MW = 3160) (28 residues)	^{121}VDEMIREA^{128} ^{20}DKDGDGTITTKE^{31} ^{141}FVQMMTAK^{148}
Peptide 1–4–1 (MW = 3560) (32 residues)	^{10}AEFKEAFSLF^{191} ^{129}DIDGDGQVNYEE^{140} ^{32}LGTVMRSLGQ^{41}
Peptide 4–4–4 (MW = 3235) (28 residues)	^{121}VDEMIREA^{128} ^{129}DIDGDGQVNYEE^{140} ^{141}FVQMMTAK^{148}

*The number indicates the amino acid positions in the native CaM sequence

Both CD and ORD have been used to measure the change in secondary structural content of CaM upon calcium binding[48]. Among the indirect methods of probing the calcium-binding sites in proteins, one of the most sensitive and useful probes is the carbocyanin dye, stains-all[49]. The binding of stains-all generates characteristic absorption peaks and induces CD signals. Several synthetic helix–loop calcium-binding sites of CaM and some of their modified analogs have been studied using stains-all as a probe[50]. It was shown that in the calcium-binding site I, replacement of Asp by Asn abolishes the binding of stains-all whereas a similar change in site IV only attenuates the stains-all binding. Similarly, replacement of Tyr in site IV with Trp increases the dye-binding affinity.

We describe here our attempts to study the three-dimensional structure of some synthetic CaM fragments corresponding to the various calcium-binding sites in isolation to see whether these have reasonable extent of secondary structure. The peptides chosen are labelled as 1–1–1, 1–4–1, 4–4–4 and 4–1–4, meaning that 1–1–1 corresponds to the first calcium-binding site characterized by helix1–loop1–helix1 and 1–4–1 would correspond to the same fragment as 1–1–1 except that the loop residue of calcium-binding site 1 is replaced by the corresponding loop from calcium-binding site IV. The four peptides we have undertaken for study are listed in Table 1. Both the native CaM and the tryptic fragments of CaM, each containing two of the binding sites have already been studied by a variety of techniques. We thought it would be an idea to further fragment the CaM molecule into its individual calcium-binding sites and to study these fragments by CD and some of them by high resolution 2D NMR. Such a fragmentation of the large peptide may result not only in the loss of three-dimensional structure but may also have calcium-binding affinities very different from the corresponding sites in the native protein. The detailed study of the secondary structure of these fragments would also reveal the importance of allosteric effects of one binding site on the other. Since only limited time was available on a 600 MHz NMR machine, not all fragments could be studied by NMR. In order to supplement the NMR data, CD studies were carried out for these fragments in TFE/H_2O mixture both in the presence and absence of calcium. We have also looked at the secondary structural contents using FT IR, in two of the fragments, 1–1–1 and 1–4–1, in TFE/H_2O and DMSO, with and without calcium. Finally some of the synthetic fragments have been studied by molecular modelling.

Experimental

The CaM fragments were synthesized using modified Merrifield solid phase method[51]. The crude peptide was purified by reverse-phase HPLC and the purity and identity were confirmed by analytical HPLC and amino acid sequencing.

CD studies were carried out on a JASCO J600 spectropolarimeter. The instrument was calibrated with an aqueous solution of d_{10}-camphor sulphonic acid[52]. Cells with 0.02 cm optical path length were used. All the four peptides 1–1–1, 1–4–1, 4–4–4 and 4–1–4 were measured in TFE/H_2O (60:40, v/v) mixture, maintaining a peptide concentration of 100 μM. Measurements were also made in presence of calcium with peptide-to-calcium ratios of 1:20 and 1:40. The signal-to-noise ratio was improved by coadding eight scans.

NMR experiments were carried out using a JEOL GSX 400 spectrometer and a Bruker AMX 600 spectrometer. Measurements were made at 40°C with a peptide concentration of 4–5 mM in DMSO-d_6. A spectral width of ~7500 Hz was used. The residual water peak in DMSO was suppressed using a presaturation pulse of 2 s duration. Two-dimensional NMR spectra namely DQF COSY, TOCSY and NOESY were obtained generally with 1 K complex data points, 32 transients, 256 t_1 increments and were zero filled to 2K × 2K points. In some experiments, the parameters were changed depend-

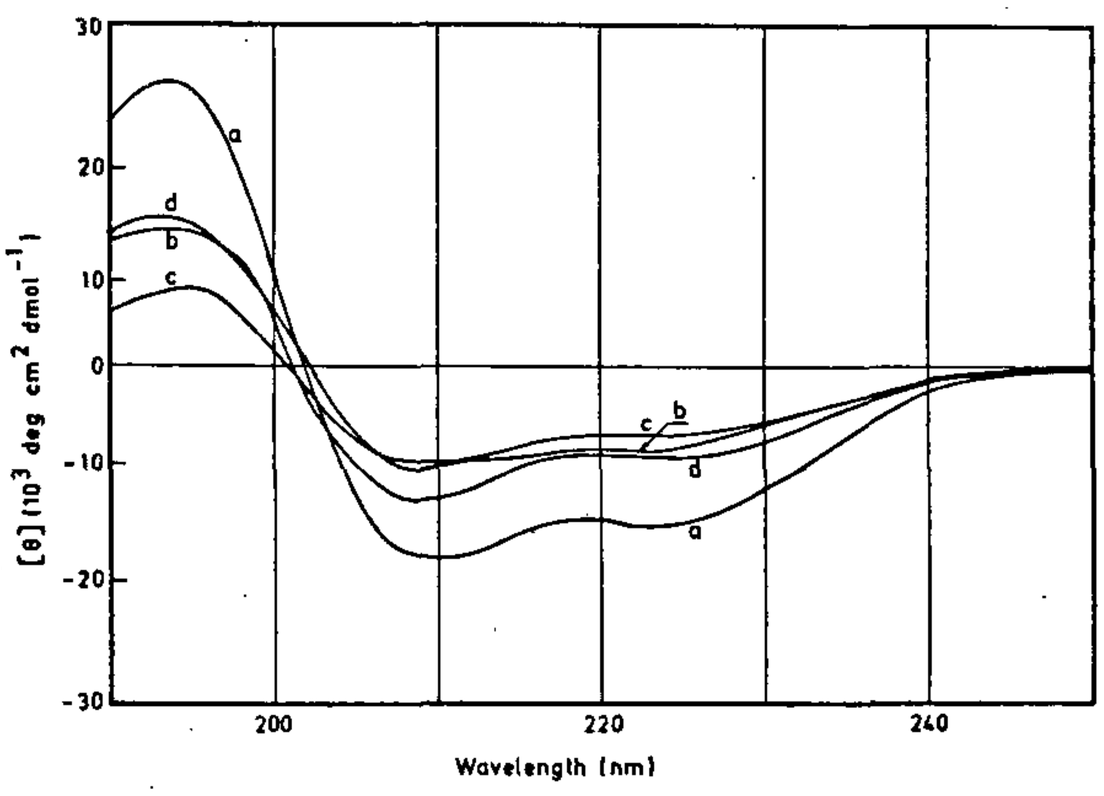
ing upon the requirements. Before Fourier transformation, the time domain signals were weighted by sine square or sine bell window functions in both dimensions. NOESY mixing times were optimized after doing several 2D experiments and a mixing time of 400 ms was found to be adequate. Similarly for the 2D TOCSY experiment, a mixing time of 60 ms was used. All 2D NMR experiments were carried out in the phase-sensitive mode and processed according to the procedure of States *et al.*[53]

Infrared spectra were recorded on Bruker IFS 66 V FT IR spectrometer. Experiments were carried out for the peptide 18R in DMSO and TFE. In order to study the nature of calcium binding on these peptides, IR spectra were taken after adding $CaCl_2$ to solutions of the peptide 1–1–1 and 1–4–1 in DMSO and TFE. The concentration of peptides in the solutions was approximately 1 mg/ml of the solvent. The ratio of concentration of calcium to peptide was 10:1. IR spectra were recorded using a variable pathlength cell with Irtan windows. All the measurements were made with the pathlength of 0.4 mm. Spectra were taken at a resolution of 2 cm^{-1}. The spectra obtained in transmittance mode were converted to absorbance mode. The solvent spectrum was subtracted from the spectrum of the peptide in the solvent. The IR band between 1500 cm^{-1} and 1800 cm^{-1}, amide I region, was considered. Second derivative analysis was carried out to find the number of component bands and their approximate peak positions. Based on the second derivative analysis, band fitting was carried out with Voigt function using Levenberg–Marquardt algorithm[54,55]. The RMS error was found to be below 0.01 for all the band-fitted spectra.

Results and discussion

CD studies

The CD spectra of all the four fragments in the absence of calcium are shown in Figure 2. In TFE/H_2O mixture (60:40, v/v), the spectra show two negative maxima around 205 and 222 nm and a positive one around 195 nm. These features are characteristic of the peptides in the right-handed α-helical conformation[56]. A numerical least squares analysis shows nearly 30–40% of helical content for all the fragments, with the 4–4–4 and 4–1–4 having maximum of approximately 39% and the 1–4–1 and 1–1–1 have 29% and 33% respectively. The β-sheet content also follows the same trend with the 4–4–4 and 4–1–4 having more than that of 1–4–1 and 1–1–1. A plot of molar ellipticity *vs* peptide-to-calcium ratio is shown in Figure 3, depicting the trends in the secondary structural content as a function of calcium concentration. It is quite difficult to quantitatively interpret the CD data although some qualitative trends can be derived. In the calcium-free situation, the 4–4–4 fragment has the maximum molar ellipticity of ~14,000 with the 4–1–4 and 1–1–1 showing a reduced ellipticity of ~8,000 and that of 1–4–1 being ~6,000 deg.cm²/dmol. Upon adding calcium, the native CaM sequences corresponding to 4–4–4 and 1–1–1 show a marked decrease in the ellipticity whereas that of the loop exchanged ones 4–1–4 and 1–4–1 show an increase in ellipticity. The trend is reversed for 1–1–1 and 4–1–4 when the calcium concentration ratio is increased beyond 1:20, although those of 4–4–4 and 1–4–1 show only a mar-



(a) Peptide 4-4-4, (b) Peptide 4-1-4, (c) Peptide 1-4-1, (d) Peptide 1-1-1

Figure 2. CD spectra of the calmodulin fragments in a mixture of TFE/water (60:40, v/v) in the absence of calcium.

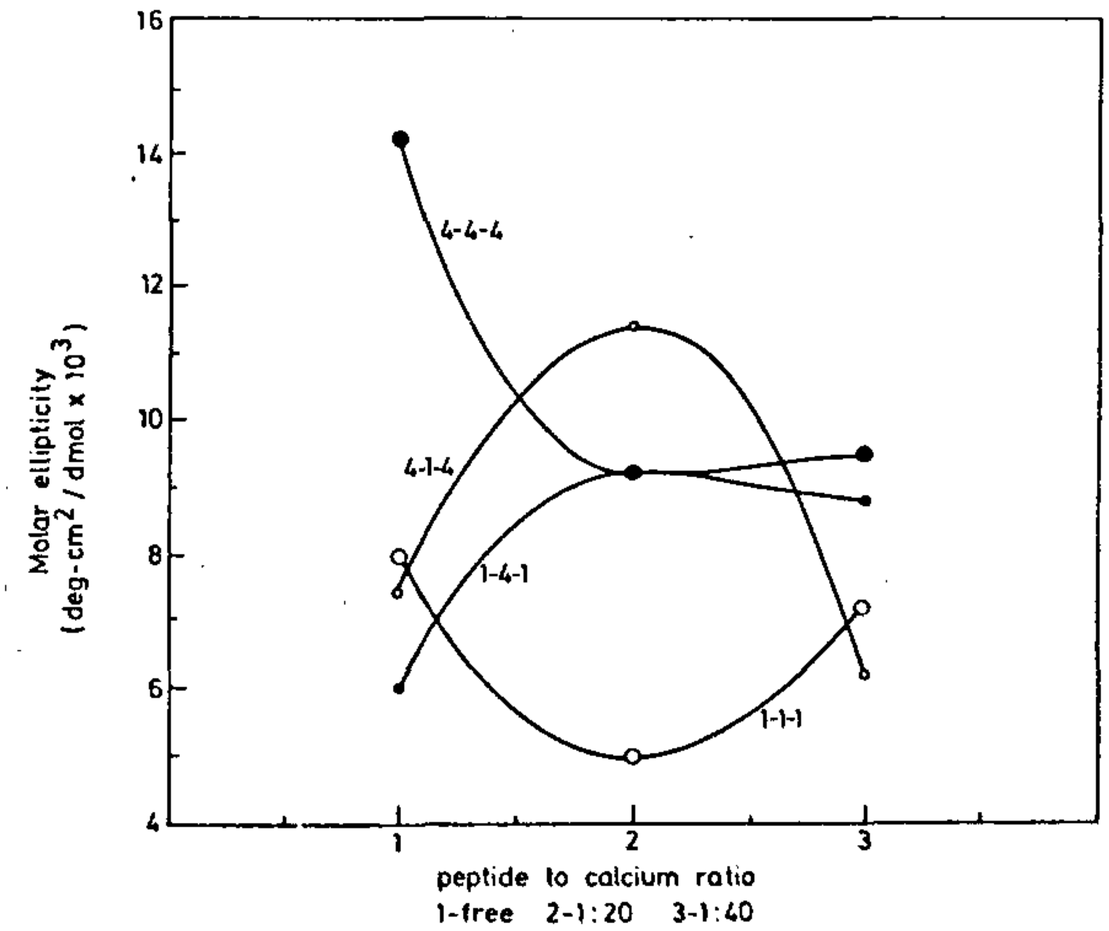


Figure 3. A plot of molar ellipticity versus peptide to Ca ratio for the four fragments mentioned in Table 1.

ginal change. When we compare the ellipticities at a peptide-to-calcium ratio of 1:20, the loop exchanged 4–1–4 and 1–4–1, having ellipticities higher than the free peptide, with the native fragments showing reduced structure. The increase in ellipticity of loop exchanged fragments could be taken as an indication of higher calcium-binding capability compared to that of the native fragments 4–4–4 and 1–1–1. Although it is difficult to quantify the ellipticity changes in terms of calcium ion affinity, the relative changes in ellipticity can be roughly equated with affinity for calcium and these are larger for 1–4–1, 4–4–4 and 4–1–4 than for 1–1–1. It should be mentioned, however, that these results pertain to the TFE/H_2O medium. Therefore the main conclusion from CD results is that a definite change in structural content is brought about by these loop exchanges.

NMR studies

The analysis of NMR spectra is based on the sequential resonance assignment procedure developed by Wüthrich *et al.*[57]. This is a two-stage process. The first step in this process involves the assignment of each cross peak in the *fingerprint* region to a specific amino acid type. The experiments, COSY[58], DQF COSY[59,60], and TOCSY[61] give information about the through bond scalar coupled spin systems. Amino acid residues with unique coupling patterns could be identified clearly whereas the other cross peaks can only be assigned to more general classes of amino acid types. Complete spin system assignment depends on the resolved coupling patterns and on the topology of the spin system under consideration. In the second stage, the cross peaks are assigned to the specific amino acid residues in the peptide sequence using NOE data.

Peptide 1–1–1

Analysis of the spin system. The peptide 1–1–1 consists of 32 amino acid residues. The DQF COSY spectrum for this peptide is shown in Figure 4 *a*. The assignment of the cross peaks NH-αCH; αCH-βCH; and the βCH-$(\gamma, -\gamma')CH_3$ to the residue Val is shown in this figure. The majority of the cross peaks have the same antiphase structure in the fingerprint region of the COSY spectrum. It is difficult to assign the amino residue type by its appearance in the COSY spectrum alone. The only residue that can be easily identified from the fingerprint region is Gly. It has a unique pair of cross peaks which result from the coupling of the NH proton to the two αCH protons. Of the 32 residues, the fingerprint region clearly shows at least 22 of the NH-αCH connectivities. The remaining residues could not be located either due to severe overlap or due to the corresponding peaks being absent in the fingerprint region (Figure 4 *b*). This

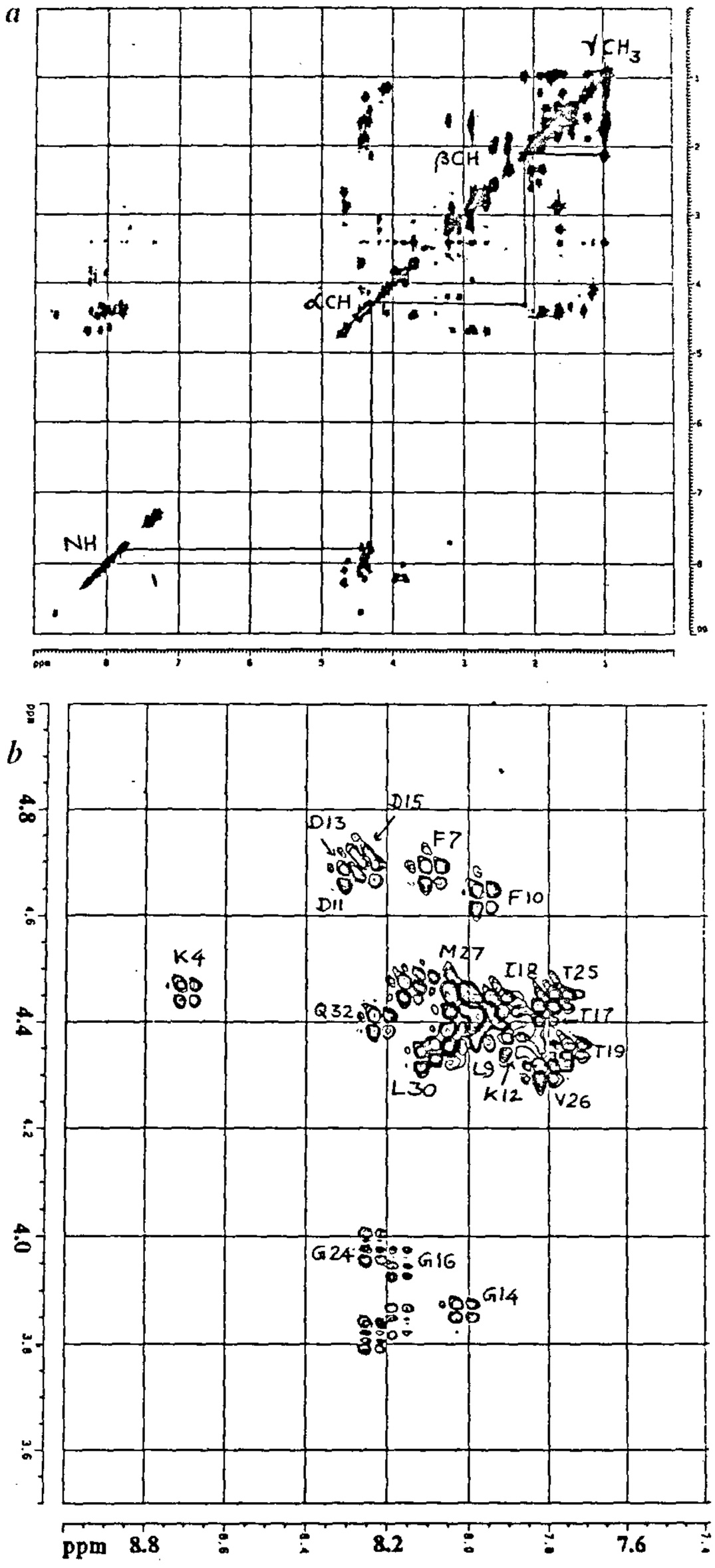


Figure 4. *a*, 600 MHz phase sensitive DQF COSY spectrum of the peptide 1–1–1 in DMSO at 40°C. The COSY connectivities of valine is shown; *b*, 600 MHz phase sensitive DQF COSY spectrum of the peptide 1–1–1 in DMSO at 40°C, showing the expanded fingerprint region. The NH-C_aH cross peaks have been labelled with the one letter amino acid codes.

might be due to the very low intensity arising out of a small NH-αCH coupling constant or due to the large line width. Also the overlap of antiphase components in

DQF COSY spectrum leads to cancellation of most of the cross peak intensity[62].

The residues Ala, Thr, Val, Ile, Gly have unique spin systems and can be identified easily. Other spin systems are classified either as AMX spin systems (phenylalanines, aspartates, and serines) or as long side chain spin systems (leucines, lysines, arginines, methionines, glutamines and glutamates). The peptide 1–1–1 has 12 amino acid residues with unique spin systems, 8 of AMX type and 12 long side chain residues.

Unique side chain spin systems (4Gly, 2Ala, 1Val, 4Thr, 3Leu, and 1Ile). There are four glycines at sequence position 14, 16, 24 and 31. Two of the glycines are identified based on their characteristic cross peaks in the COSY spectrum. One more glycine residue, with degenerate αCH protons, is identified from the sequential NOE data. Two alanines at sequence positions 1 and 6 are identified from the COSY and TOCSY spectra. The assignment of the αCH proton is difficult due to severe overlap. One valine and all the four threonines are assigned unambiguously from the COSY and TOCSY spectra. The identifications of Ile and Leu residues are based on the appearance of the NH-δCH3 cross peaks. Unambiguous assignment is achieved only from the analysis of NOESY spectrum. Thus out of 15 residues of this type, eleven residues are assigned unambiguously (Figure 5).

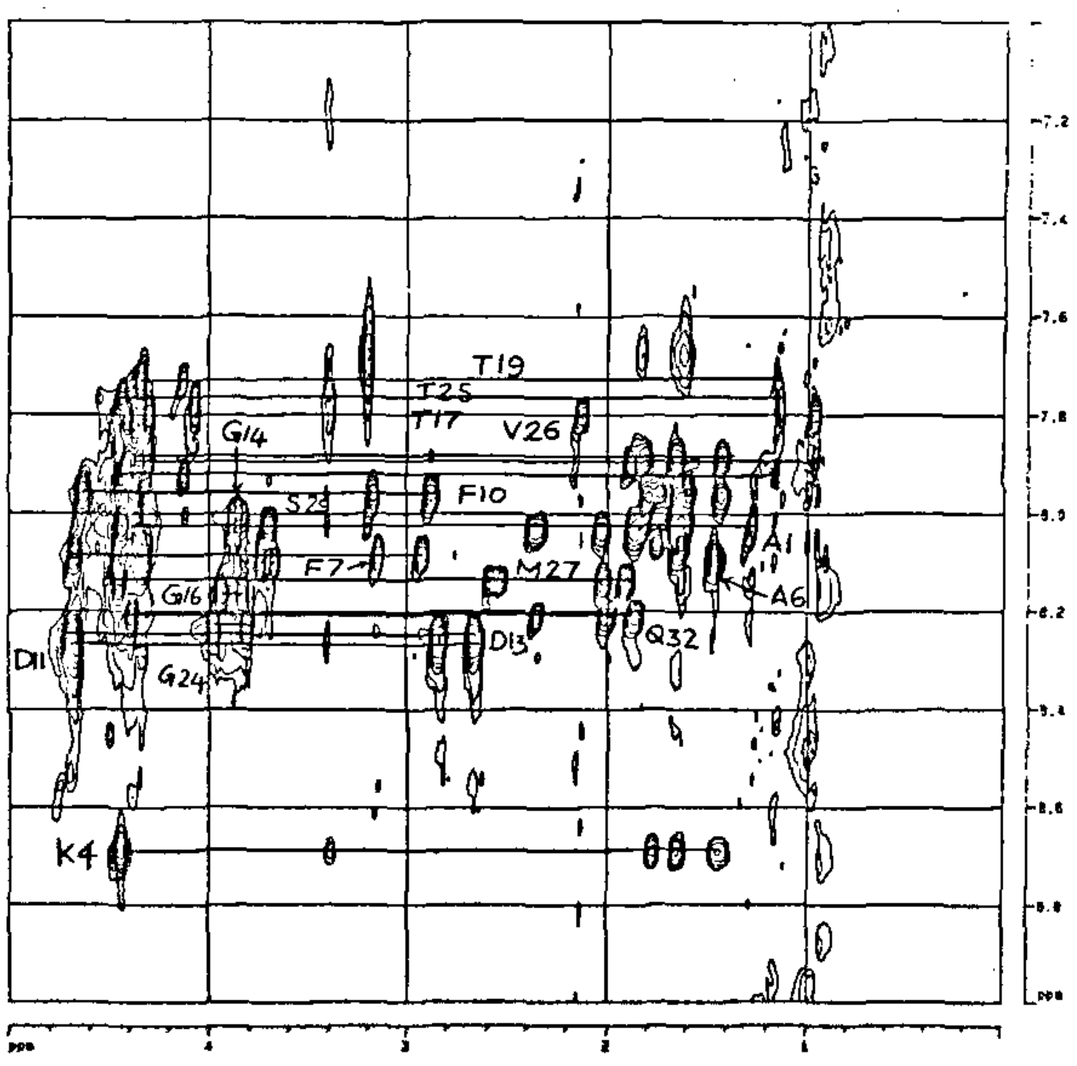


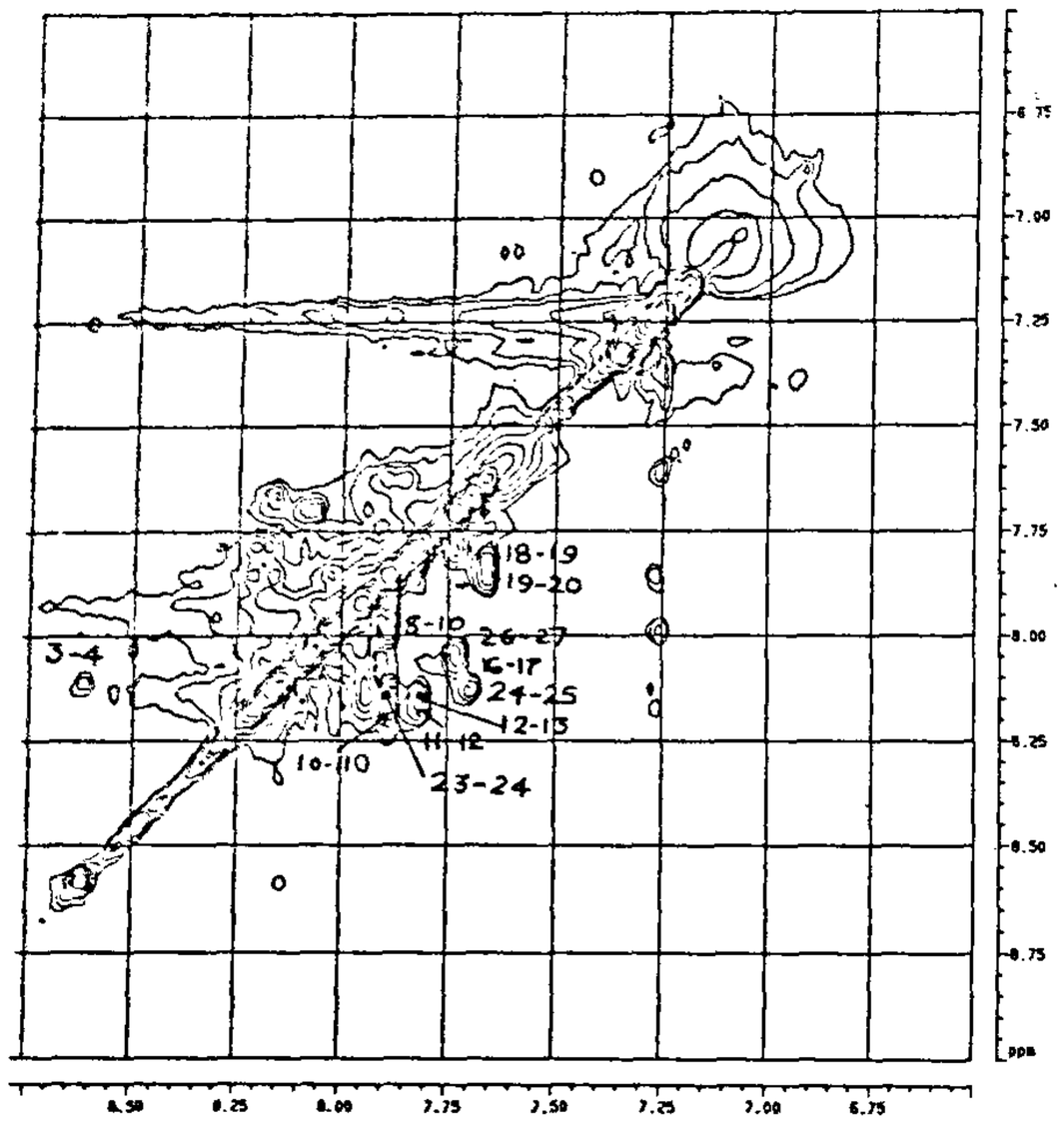
Figure 5. Part of the 600 MHz phase sensitive TOCSY spectrum of the peptide 1–1–1 in DMSO at 40°C with spin lock duration of 60 ms. *Some of the coupling networks have been labelled.*

AMX spin systems (3Asp, 3Phe, 2Ser). These residues are distinguished from other spin systems primarily by the chemical shift positions of β protons and by the recognition of the characteristic cross-peak patterns. The only aromatic residues present in this system are phenyl alanines. Two of the three Phe are identified from the possible NOE cross peaks between the aromatic ring protons and the aliphatic protons of the same residue. NH resonance of one of the Phe is missing, but its position is identified from the NOE data. Two serines are distinguished from the aspartate residues due to the low field resonances of β protons. Seven out of eight residues of this type are thus identified.

Long side chain spin systems (3Glu, 1Met, 1Gln, 3Lys and 1Arg). These spin systems were grouped into one class partly on the basis of chemical shift values of the βCH protons and partly on the patterns of the αCH-βCH cross peaks which are distinct from those of the AMX spin systems. The identification of the additional resonances belonging to the spin system helps in distinguishing the type of amino acid, i.e., lysines from glutamates, methionines, etc.

Sequential assignment. The sequential assignment of the peptide 1–1–1 dissolved in DMSO solution is based on the analysis of NOESY spectra with a mixing time of 400 ms. The general approach is to analyse the peaks in the NH–NH region and that in the fingerprint region (Figure 6 *a*, *b*). The unambiguous assignment of the residues valine, glycines and threonines helps in the identification of the sequence position of the other residues. Analysis of the fingerprint region of the NOESY spectrum helps in the assignment of the peptide segments, Phe-3–Lys-4; Asp-11–Lys-12–Asp-13–Gly-14–Asp-15; Gly-16–Thr-17; Gly-24–Thr-25; Val-26–Met-27. The peak position of the Phe-3 is located on the basis of its NOE with Lys-4. The other two phenylalanines are at positions 7 and 10. The residue at position 10 is identified on the basis of the NOE cross peak between the ring proton and the γCH proton of Leu-9. Thus all the three phenylalanines are located clearly. Analysis of the NH–NH region of the NOESY spectrum (Figure 6 *a*) confirms the sequential assignment. The residue Leu-23 is assigned based on the NH–NH cross peak between Leu-23 and Gly-24. The unambiguous assignment of the two lysines, at sequence positions 4 and 12 enables the assignment of the remaining one to Lys-21. The presence of weak NH–NH cross peak between Lys-21 and another long side chain containing residue helps in the assignment of Glu-22. Identification of the Leu-9 and Leu-23 made the assignment of the remaining leucine residue to Leu-30. The summary of through space connectivities inferred from 2D NOESY for the peptide 1–1–1 is given in Figure 7.

a



b

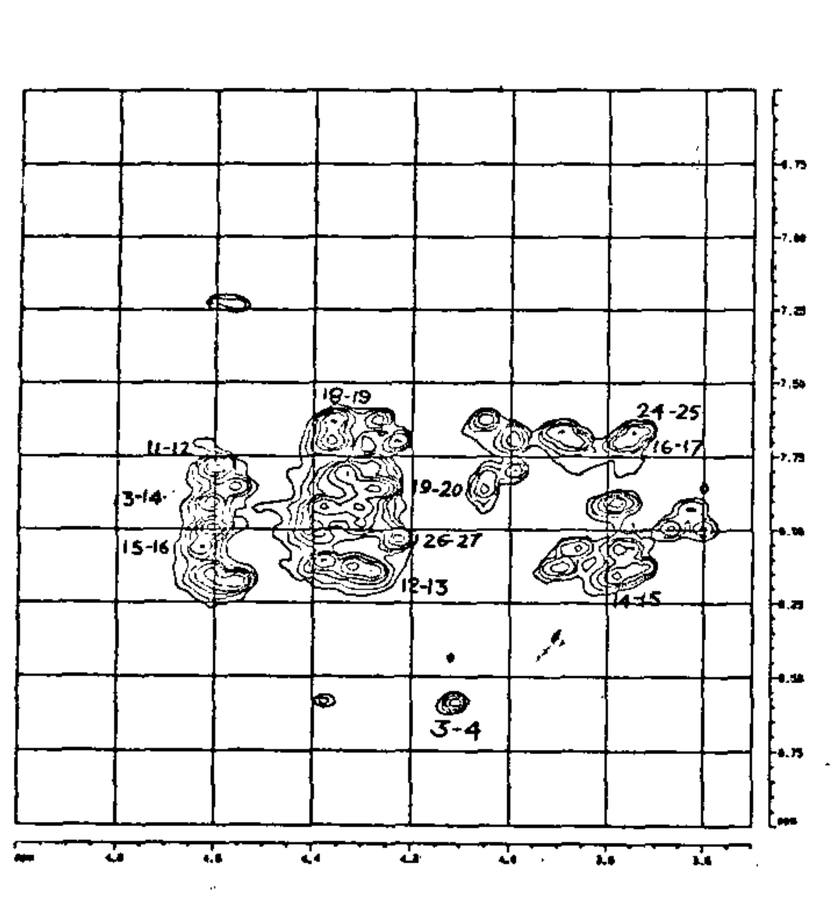


Figure 6. *a*, 600 MHz phase sensitive NOESY spectrum of the peptide 1–1–1 in DMSO at 40°C with a mixing time of 400 ms showing the amide region. *b*, 600 MHz phase sensitive NOESY spectrum of the peptide 1–1–1 in DMSO at 40°C with a mixing time of 400 ms showing the fingerprint region.

Secondary structural elements. It has been demonstrated that the regular secondary structural elements of proteins can be identified by visual inspection of the patterns of sequential connectivities observed in a given region of the primary structure[63]. The presence of NH_i–NH_{i+1} and αCH_i–NH_{i+3} connectivities is indicative of α-helices, while the appearance of αCH_i–NH_{i+1} is indicative of extended structures such as β-strands. Type I and type II β-turn can also be identified by their pattern of sequential connectivities.

The presence of strong NH–NH connectivities from the Phe-10 to Asp-15 and Gly-16 to Thr-20 indicates that the peptide exhibits helical structure in these regions. The NH resonance of all the three aspartates appear almost close to each other, making the assignment of the αCH_i–NH_{i+3} cross peaks between the residues, Asp-11–Gly-14; Lys-12–Asp-15; and Asp-13–Gly-16, difficult. The absence of αCH_i–NH_{i+3} for the peptide segment Gly-16 to Thr-20 indicates that this part is characterized by a loose helical segment. The strong αCH_i–NH_{i+1} and a weak αCH_i–αNH_{i+1} between Asp-11 to Asp-15 indicate the possibility of an extended structure. All these point to a possible conformational equilibrium between an extended and helical structure. This fact is again supported by the presence of weak sequential αCH_i–αCH_{i+1} connectivities for the residues Lys-12 to Asp-15. The presence of strong αCH_i–NH_{i+1} and NH_i–NH_{i+1} cross peaks between the residues Phe-3–Lys-4; Gly-24–Thr-25; and Val-26–Met-27 indicates the possibility of a turn structure. The presence of NH_i–NH_{i+2} and βCH_i–NH_{i+2} connectivities between the residues Phe-7 to Phe-10 indicates the occurrence of a β-turn structure in this region. The end residues are not well characterized due to lack of NOE information.

Peptide 1–4–1

The primary structure of the peptide 1–4–1 resembles that of the peptide 1–1–1 except that the residues at positions 12, 17, 18, 19, 20, 21 are different. This peptide has thirteen amino acid residues with unique spin systems (2Val, 1Thr, 2Ala, 4Gly, 1Ile, 3Leu), ten of AMX spin systems (3Phe, 1Tyr, 3Asp, 1Asn, 2Ser) and nine with long side chain spin systems (1Met, 4Glu, 2Gln, 1Lys, 1Arg). Figure 8 shows the fingerprint region of the DQF COSY spectrum. At the contour level chosen, the characteristic NH–αCH cross peaks can be located for 25 amino acid residues. As discussed previously, some of the cross peaks are missing or might occur in the crowded region. The various amino acid residues identified through TOCSY patterns are shown in Figure 9. Peaks corresponding to one glycine and one alanine are missing in the spectrum. The assignment of Leu, Glu, Gln is difficult at this stage. The assignment of the NOESY cross peaks in the fingerprint region and the NH–NH region helps in the identification of the sequence positions of the various amino acid residues. The summary of the NOE connectivities observed is presented in Figure 10.

The peptide 1–4–1 exhibits almost similar conformation as that of the peptide 1–1–1. In the peptide 1–4–1, the three aspartate residues are well resolved unlike in the peptide 1–1–1. The NH–NH region of the peptide

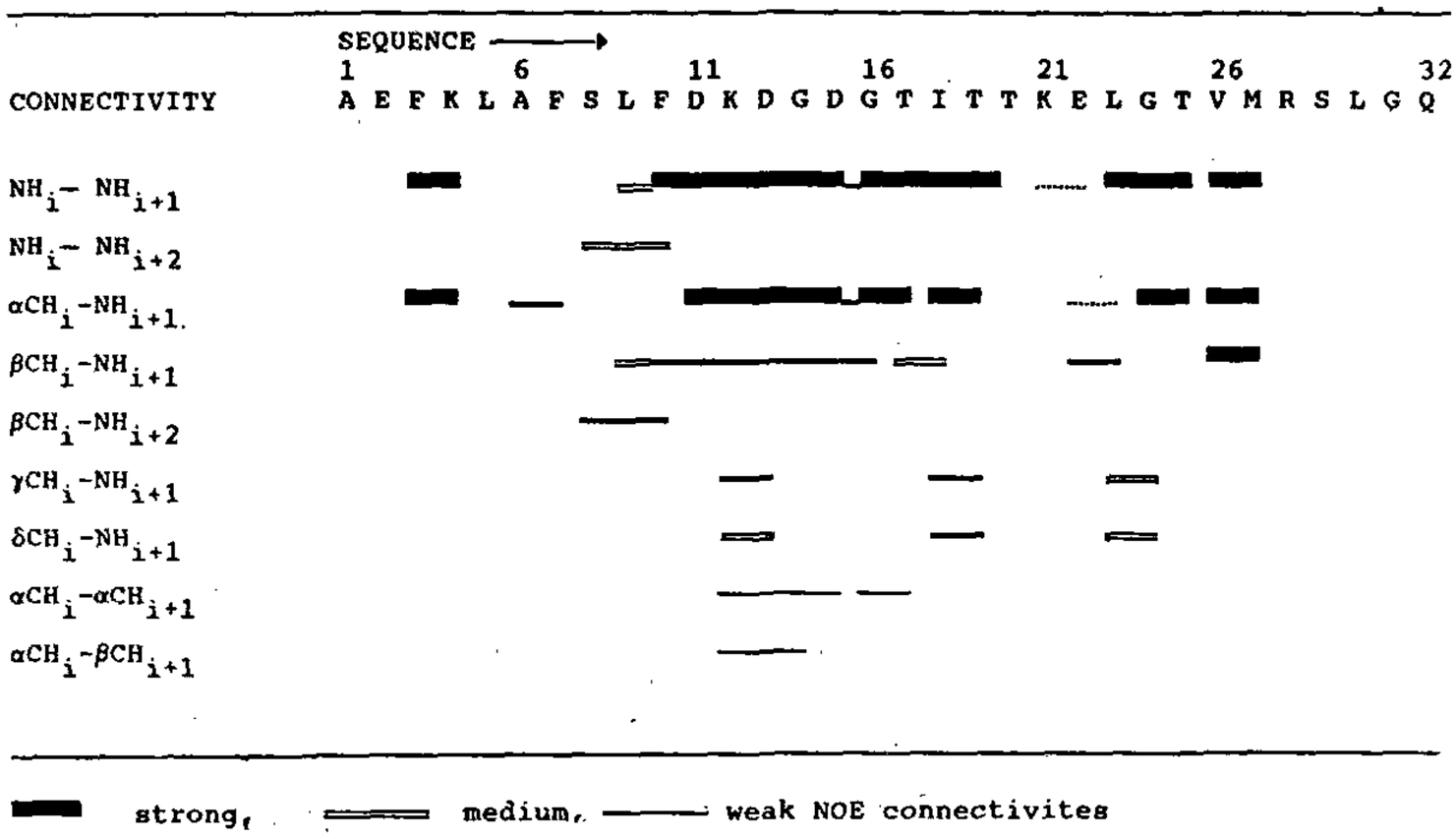
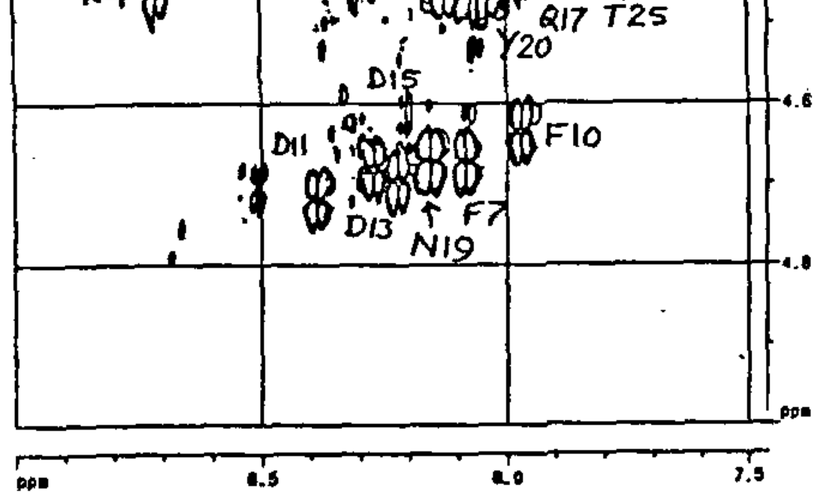


Figure 7. Schematic diagram showing the various through space connectivities derived from NOE data for the peptide 1–1–1.



ire 8. 600 MHz phase sensitive DQF-COSY spectrum of peptide 1–4–1 in DMSO at 40°C showing the fingerprint region.

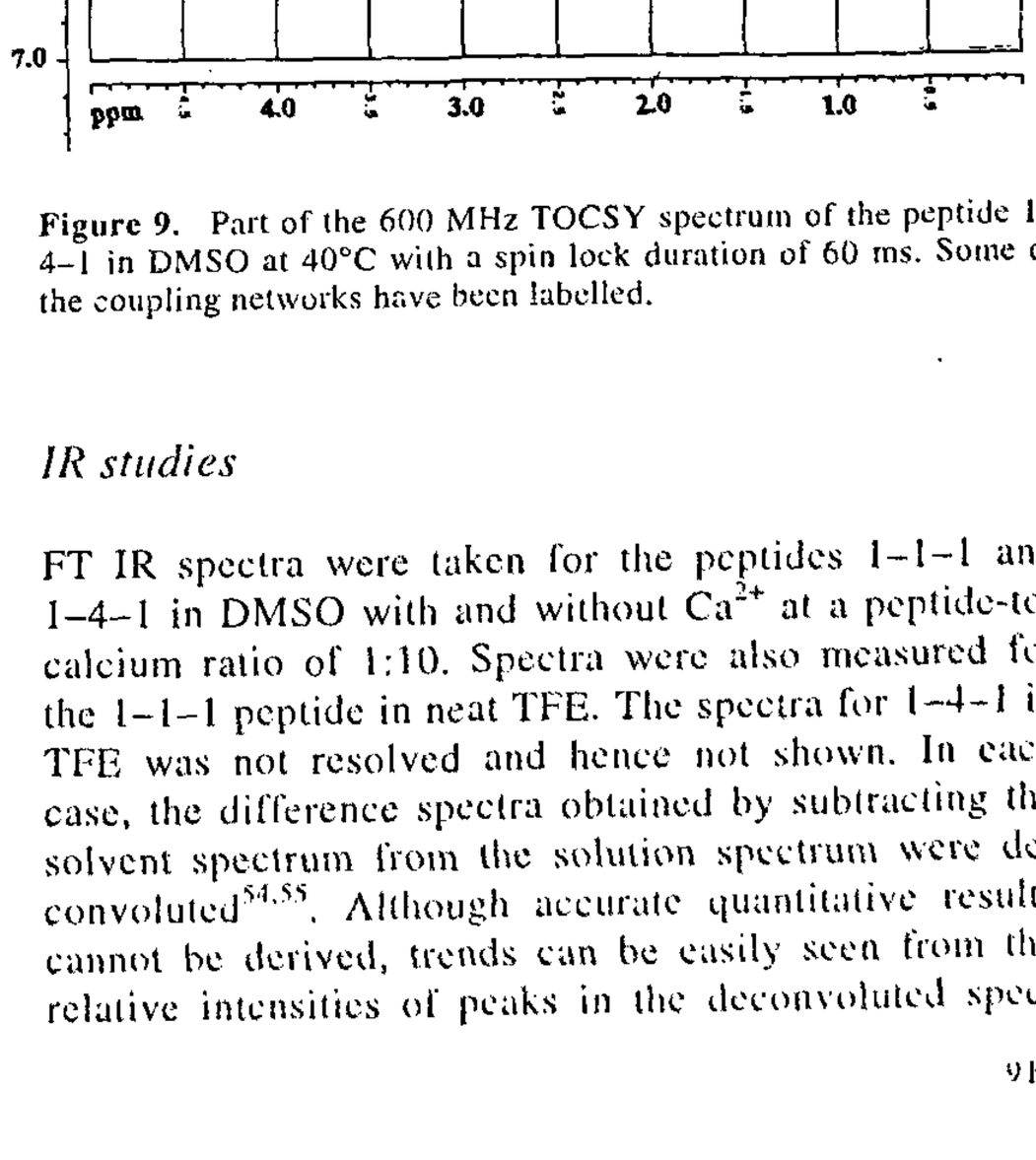


Figure 9. Part of the 600 MHz TOCSY spectrum of the peptide 1–4–1 in DMSO at 40°C with a spin lock duration of 60 ms. Some of the coupling networks have been labelled.

:–1 was characterized by a well resolved continuous :tch of NH–NH cross peaks and this is probably be-se this peptide exists in a folded conformation. The stence of equally intense αCH_i–NH_{i+1} and the ab-ce of αCH_i–NH_{i+3} and αCH_i–βCH_{i+3} connectivities eal that this peptide might exist in dynamic equilib-m between various conformers.

IR studies

FT IR spectra were taken for the peptides 1–1–1 and 1–4–1 in DMSO with and without Ca^{2+} at a peptide-to-calcium ratio of 1:10. Spectra were also measured for the 1–1–1 peptide in neat TFE. The spectra for 1–4–1 in TFE was not resolved and hence not shown. In each case, the difference spectra obtained by subtracting the solvent spectrum from the solution spectrum were deconvoluted[54,55]. Although accurate quantitative results cannot be derived, trends can be easily seen from the relative intensities of peaks in the deconvoluted spec-

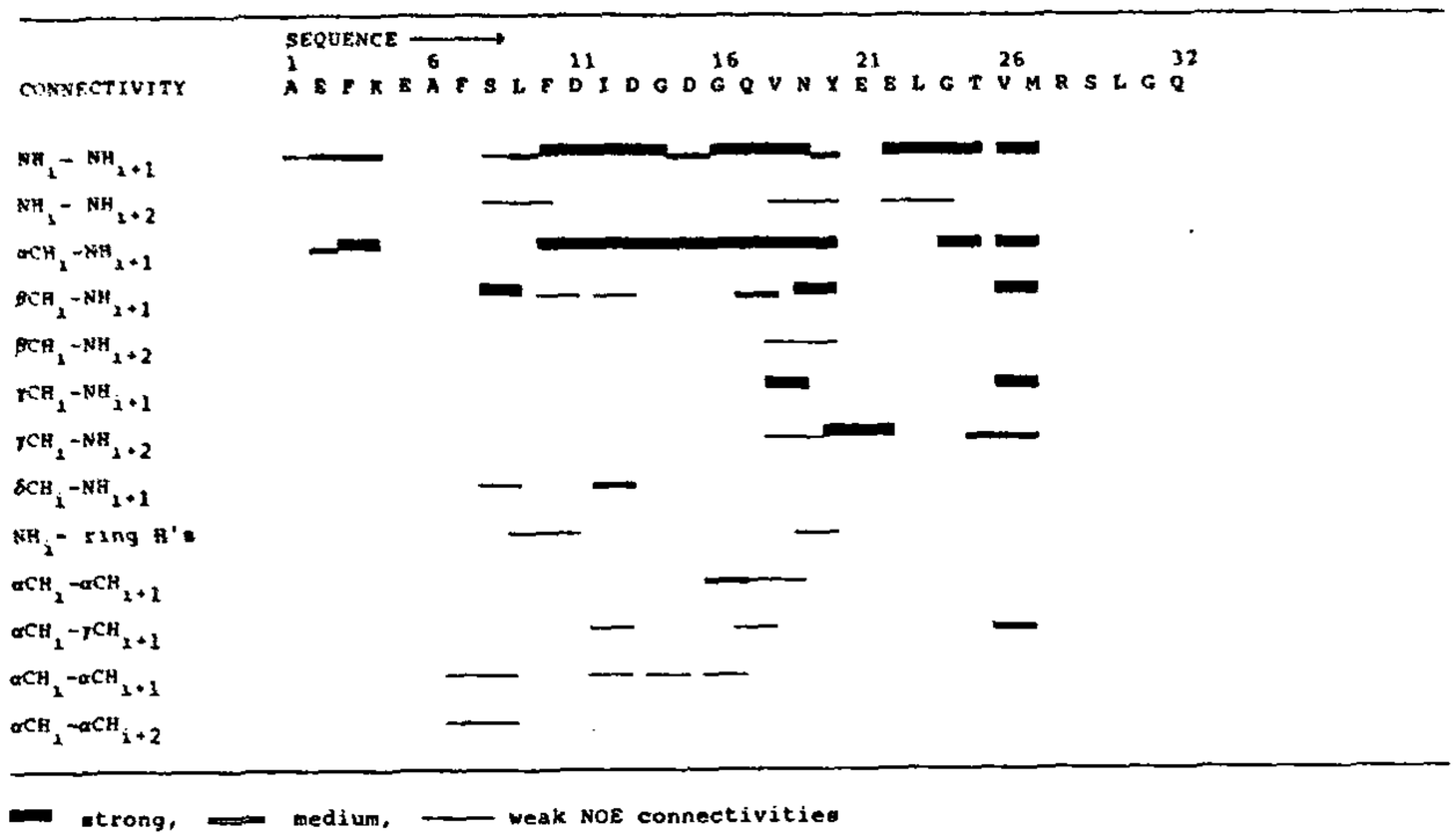


Figure 10. Schematic diagram showing the various through space connectivities derived from NOE data for the peptide 1–4–1.

trum corresponding to the various structural elements. These are summarized below.

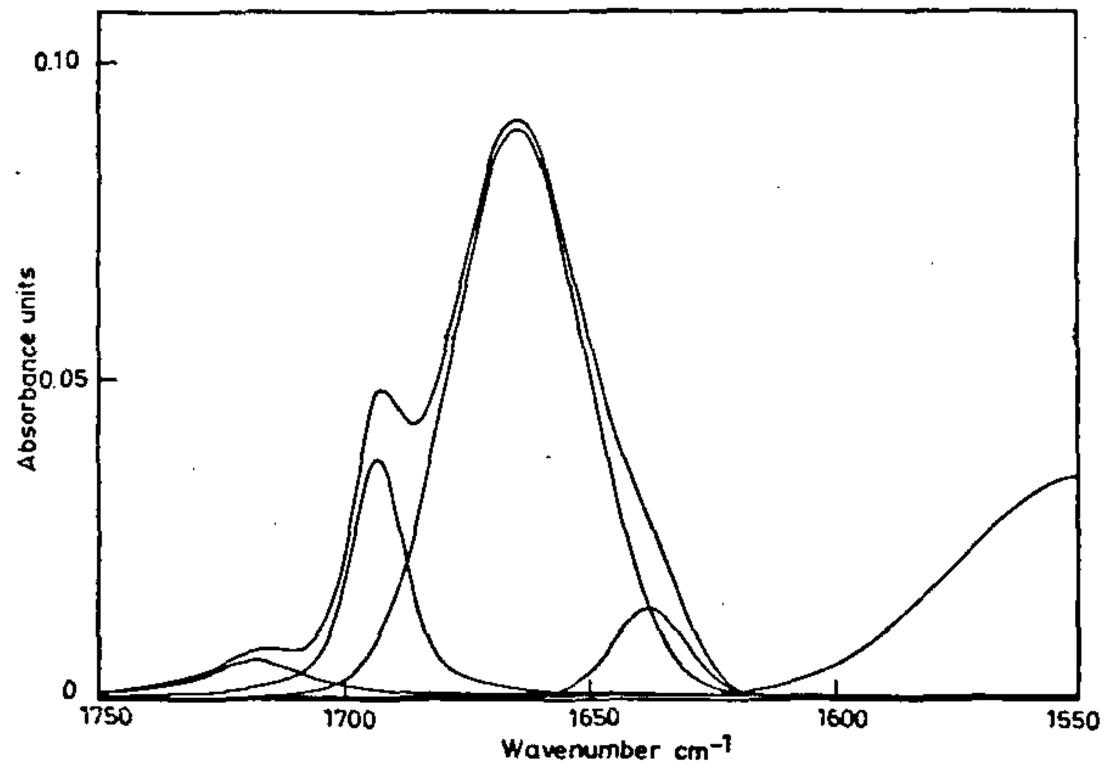
For the peptide 1–1–1 in DMSO (Figure 11 *a*) there are three prominent peaks at 1638 cm^{-1}, 1693 cm^{-1} and 1664 cm^{-1} respectively in the order of increasing intensity. This means that the peptide adopts an extended structure with some of the backbone adopting a β-turn structure. The band at 1693 cm^{-1} indicates the presence of antiparallel strands in some part of the molecule. In the presence of Ca^{2+} (Figure 11 *b*), there is a marked reduction in the peak at 1693 cm^{-1} with a relative increase of the peak at 1638 cm^{-1}. There is also a relative reduction in the intensity of band at 1664 cm^{-1}. The presence of calcium is seen to induce an increase in the turn structure.

For the loop exchanged peptide, 1–4–1 in DMSO (Figure 12 *a*) in the absence of calcium, the prominent peaks appear at 1638 cm^{-1}, 1659 cm^{-1}, 1666 cm^{-1} and weak peaks at 1679 and 1692 cm^{-1}. The strongest peak in terms of area is the one at 1638 cm^{-1}. The spectra can be interpreted in terms of the system adopting mostly β-turn structure. In the presence of calcium, there is a marked change in the overall spectral profile (Figure 12 *b*). The deconvoluted spectrum clearly shows a marked increase of the peak at 1641 cm^{-1} and a considerable decrease in the peak at 1693 cm^{-1}. The peaks at 1669 cm^{-1} and 1664 cm^{-1} do not change in intensity. This can be interpreted in terms of considerable increase in the β-turn structure with a concomitant decrease in the antiparallel β-sheet structure. The FT IR spectrum of the peptide 1–1–1 was also carried out in the structure-forming solvent, TFE. The results indicate that in addition to α-helical content, parts of this molecule should adopt an antiparallel β-sheet structure as well as β-turn structure. The introduction of calcium causes an increase in turn structure with a concomitant decrease in the helical content. It is to be noted that molar ellipticity at 222 nm for the peptide 1–1–1 shows a similar trend, i.e., decrease of helicity upon adding calcium.

In summary, FT IR results have shown that there is no preponderance of helicity in DMSO, for any of these fragments, nor is there an induction of α-helical structure upon adding calcium. The structure-forming tendency of TFE has been brought out clearly, by both CD and FT IR studies.

Molecular modelling studies

Our attempt to study the structure of some of the synthetic calmodulin fragments by molecular modelling is summarized in the following. The detailed two-dimensional NMR[64] study of the first calcium binding domain of 1–1–1 and its loop exchanged analog 1–4–1 where the spectra are more or less completely assigned, show that in DMSO these two peptides do not seem to adopt the helix–loop–helix conformation. This would mean that the breaking up of the CaM molecule into its four individual calcium-binding sites has serious consequences on the secondary structures. The overall three-dimensional structure as well as its calcium-binding affinity could be controlled by allosteric effects. Perhaps this is an important result which points to the fact that it is not possible to precisely compare the relative binding affinities of different calcium domains in CaM. This is because one site may influence the other with respect to the exact conformation and therefore binding



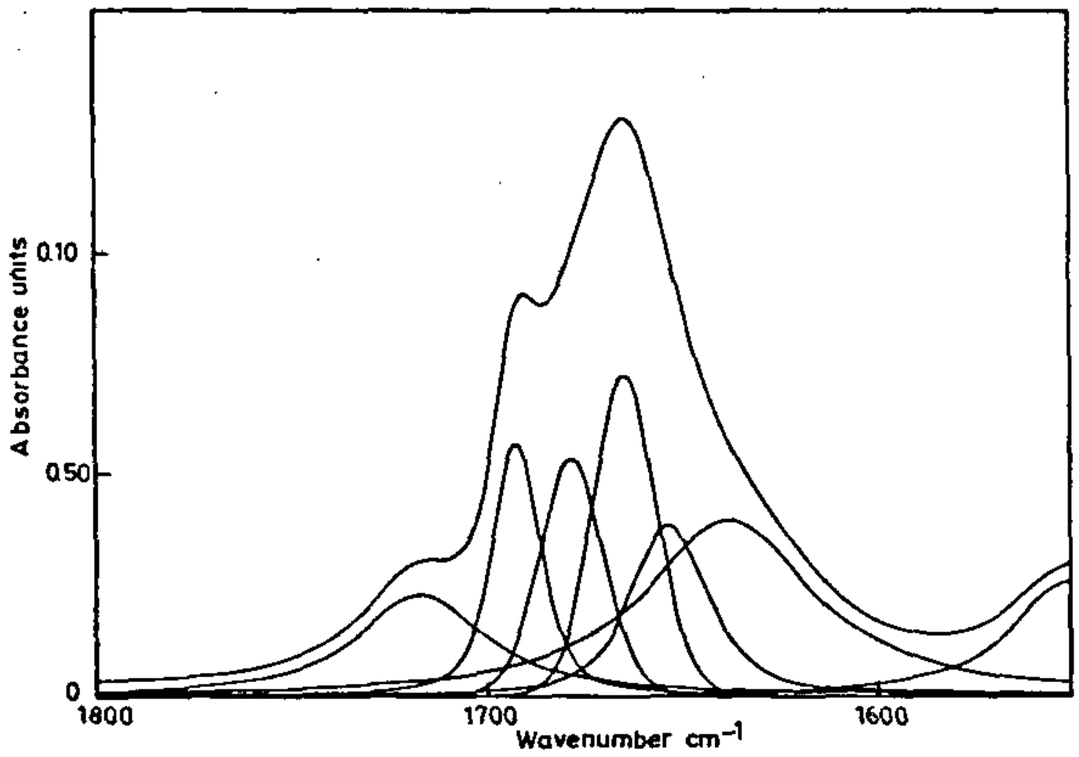
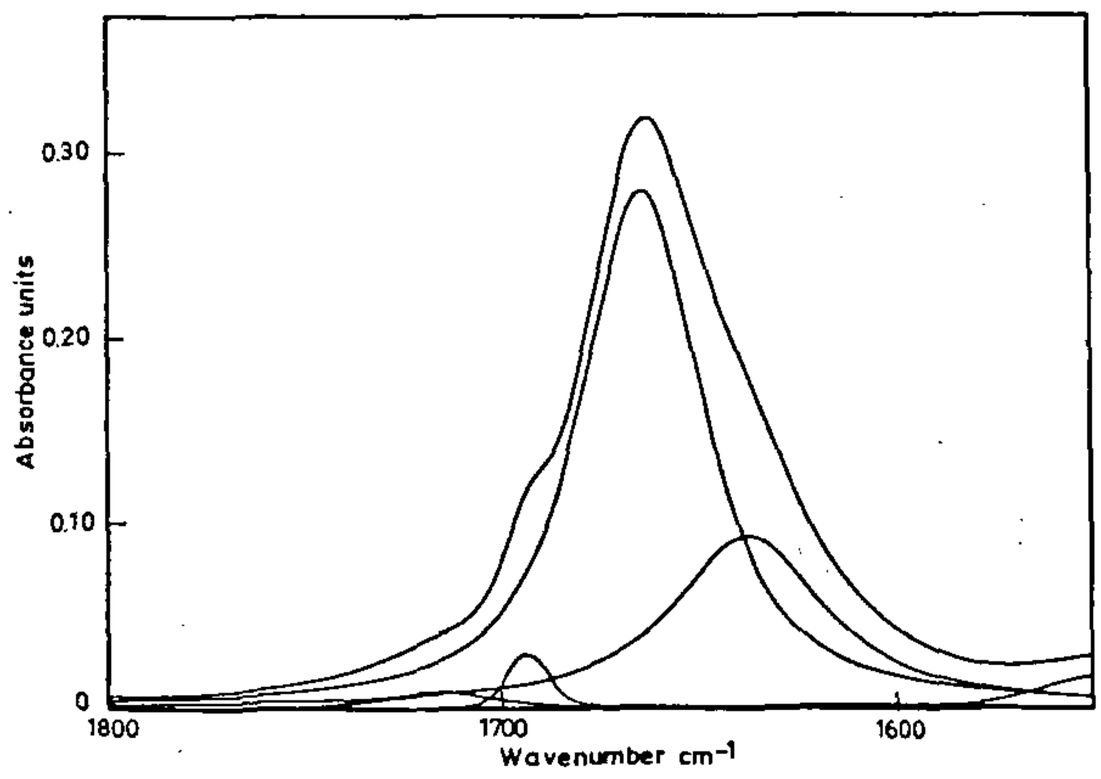


Figure 11 *a, b.* Difference FT IR spectra of the peptide 1–1–1 in DMSO solution corresponding to (*a*) free peptide in the absence of calcium and (*b*) that with a peptide-to-calcium ratio of 1:10. The spectra were obtained by subtracting the spectrum of the solvent from that of the solution. The figure includes the experimental (outer) spectrum and the deconvoluted resolved peaks.



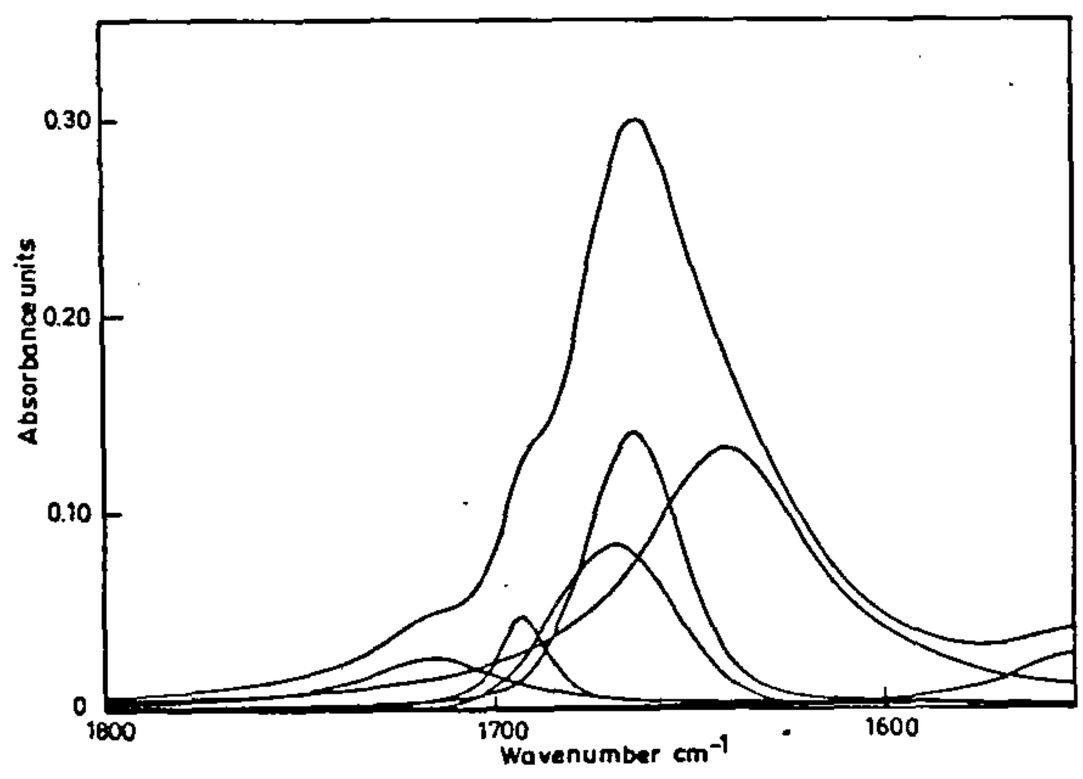


Figure 12 *a, b.* Difference FT IR spectra of the peptide 1–4–1 in DMSO solution corresponding to (*a*) free peptide in the absence of calcium and (*b*) that with a peptide-to-calcium ratio of 1:10. The spectra were obtained by subtracting the spectrum of the solvent from that of the solution. The figure includes the experimental (outer) spectrum and the deconvoluted resolved peaks.

efficiency. Having undertaken the studies of fragments by 2D NMR, we thought it would be logical to also look into the possible conformation of these fragments using molecular modelling. To this end, we have attempted a study of the synthetic CaM fragments 4–4–4, 4–1–4 (EF hands) as well as a shorter peptide (4–4) representing only the loop and helix portion of site 4 of CaM (Table 2). There are several standard modelling packages available such as QUANTA/CHARMM[65] BIOSYM[66], AMBER[67,68], etc. We had access only to QUANTA/CHARMM program and that too with limited computer time. Using the available facilities, the above three systems have been explored.

Table 2. The amino acid sequences for the peptides 4–4, 4–4–4 and 4–1–4

Peptide	Amino acid sequences
Peptide **4–4**	^{112}LGEKLTDEE^{120} ^{121}VDEMIREA^{128}*
(32 residues)	^{129}**DIDGDGQVNYEE**^{140} ^{141}FVQ^{143}
Peptide **4–4–4**	^{121}VDEMIREA^{128} ^{129}**DIDGDGQVNYEE**^{140}
(28 residues)	^{141}FVQMMTAK^{148}
Peptide **4–1–4**	^{121}VDEMIREA^{128} ^{20}**DIDGDGTITTKE**^{31}
(28 residues)	^{141}FVQMMTAK^{148}

*The number indicates the respective position of the amino acid residues in the native CaM. The residues 121–128 and 141–148 corresponding to the E and F helices of the calcium-binding site 4; 20–31 and 129–140 correspond to the loop region of the calcium binding site 1 and 4 respectively.

Secondary structural elements and the polypeptide chain geometries. The twenty normal amino acids that form the building blocks of proteins and polypeptides can be classified broadly into aliphatic (Ala, Val, Leu and Ile), nonpolar (Gly, Pro, Cys and Met), aromatic (His, Phe, Tyr and Trp), polar (Asn, Glu, Ser and Thr) and charged (Lys, Arg, Asp and Glu) types. At the

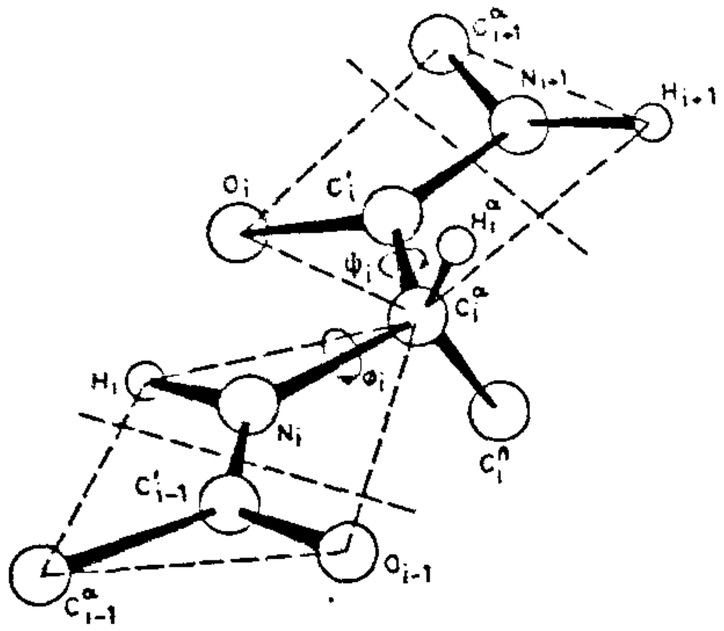


Figure 13. Perspective drawing of the α-L-polypeptide chain illustrating the torsion angles ϕ and ψ. The chain is drawn in the planar all-*trans* form ($\phi = \psi = 180°$). The dashed lines delineate the extent of the *i*th amino acid residue.

Table 3. Polypeptide chain geometries

Bond	Bond length	Bond	Bond angle
C^{α}–C′	1.53 Å	C^{α}–C′–N	113°
C′–N	1.32 Å	C′–N–C^{α}	123°
N–C^{α}	1.47 Å	N–C^{α}–C′	110°
C′=O	1.24 Å		
N–H	1.00 Å		
C^{α}–C^{β}	1.54 Å		
C^{α}–H^{α}	1.07 Å		

Table 4. Dihedral angles ϕ and ψ for typical secondary structures in polypeptides

Structure	ϕ	ψ
Right-handed α-helix	–57°	–47°
Left-handed helix	+57°	+47°
Parallel-chain pleated sheet	–119°	+113°
Anti-parallel chain pleated sheet	–139°	+135°

normal pH of 7, usually the α amino group is protonated and the carbonyl is deprotonated so that they exist as zwitter ions[69]. Fortunately, proteins and nucleic acids are unbranched polymers and unlike synthetic high polymers they can be arranged into a head-to-tail fashion. In the case of proteins, the absolute sequence of residues constitutes the *primary structure*. The primary sequence does not imply a knowledge of the conformation of any individual residue or the polymer backbone.

The secondary structural elements consist of the common motifs such as α-helix, parallel and antiparallel β-sheets, β turns of type I and II and occasionally a more compact 3_{10} helix in some globular proteins. All the above secondary structural motifs can be understood in terms of two of the important basic structural facts. The peptide bond –CO–NH– invariably occurs in the *trans planar* configuration and this can be rationalized in terms of resonance stabilization that imparts a partial double bond character to the C–N bond. The all-*trans* planar structure fixes the distance between successive α carbon atoms at 3.8 Å and the conformation of the whole backbone chain is completely governed by the rotations about the N–C_{α} and C_{α}–C′ bonds, where C′ is the carbonyl carbon. Torsional motion about these bonds are the dihedral angles, ϕ and ψ respectively. This means that if one can determine all the ϕ_i and ψ_i values, the complete conformation of the backbone of the protein can be uniquely described. Figure 13 gives the schematics of a polypeptide chain in the all-*trans* form and the standard bond lengths and bond angles involving atoms of the backbone skeleton are given in Table 3.

In the planar *trans* configuration, for an α-L-polypeptide, $\phi_i = \psi_i = 180°$. The three-dimensional conformation of the polypeptide backbone as a whole is specified by enumerating all ϕ_i and ψ_i values. A helical form is generated by making all $\phi_i = -57°$ and all $\psi_i = -47°$. The various torsion angles characteristic of the regular secondary structures encountered in polypeptides and proteins are summarized in Table 4.

Ramachandran diagrams. Although a polypeptide can, in principle, adopt an infinite array of backbone conformation, each corresponding to a unique set of ϕ, ψ values, a majority of the hypothetical conformations can be excluded from considerations based on unfavourable steric overlaps. Ramachandran and coworkers[70,71] were the first to investigate this problem.

Thus the distances C_i^{α} to O_i, C_i^{α} to N_{i+1} or C_i^{α} to C_{i+1}^{α} depend solely on the angle ψ_i (Figure 13). By carefully examining the steric overlap, one can easily eliminate certain ranges of values of ψ_i. On critical examination of the scale model, additional values of ψ_i can be eliminated by considering overlaps between atoms separated by ϕ_i and ψ_i, i.e. O_i and O_{i-1}. The resulting steric restrictions which depend simultaneously on the values of ϕ_i and ψ_i make the torsion angles about N_i–C_i^{α} and C_i^{α}–C_1' interdependent. With the amide groups fixed in a *trans* form, rotation within a given residue is interdependent, whereas they are sterically independent of rotation in the neighbouring residues. After consideration of short-range interactions among the atoms in the immediate vicinity of the peptide bonds, Ramachandran *et al.*[70,71] developed a set of diagrams which show the energetically-allowed ranges of ϕ and ψ that are sterically favoured.

Figure 14 *a* shows the steric diagram for the glycine residue (no side chain groups) showing the allowed values of these angles. The diagram is centrosymmetric with respect to any line passing through the points $\phi = \psi = 0$, as a consequence of the symmetry of the

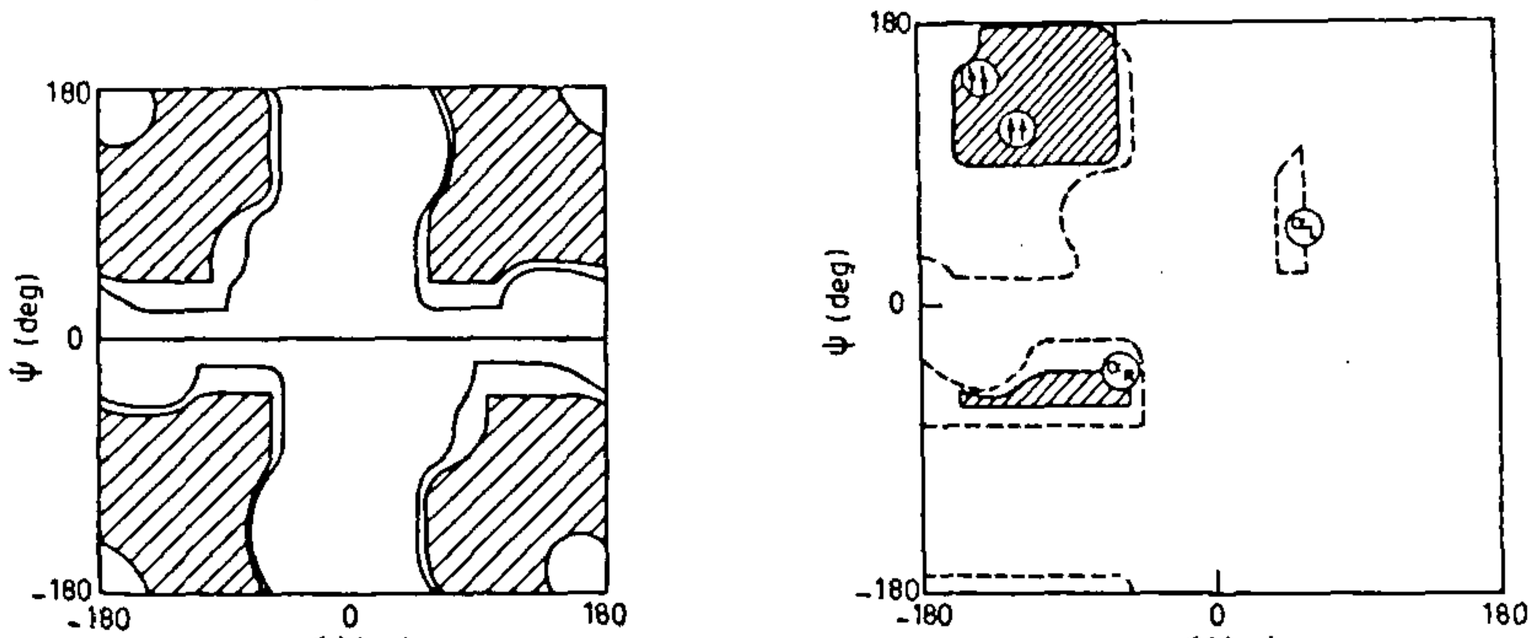


Figure 14. *a,* Steric contour diagram for a glycyl residue in a polypeptide chain. Shaded zones show normal and unshaded zones show 'outer limit' contours; *b,* The same for L-alanyl in a polypeptide chain. Right handed (α_R), left handed (α_L) helices, parallel and antiparallel sheets are marked.

glycine residue. As expected the *cis* configuration about either bond ($\phi = 0$, $\psi = 0$) is sterically forbidden. Figure 14 *b* is a similar diagram for alanine ($R = CH_3$) which is, as expected, clearly asymmetric; and the sterically allowed domain is considerably smaller because of additional unfavourable contact with the methyl side chain. Similar inference can be derived for peptides with larger and branched side chains, where the sterically allowed areas will be somewhat restricted than for alanine. The Ramachandran diagrams are the best starting point that give approximate indication of the allowed conformation of the amino acid residues. They do not indicate the gradations in energy within the sterically permissible regions and to estimate this, it is necessary to compute the potential energy as a function of ϕ and ψ. While rigorous quantum mechanical evaluation of such potential functions is beyond the reach of the currently-available computing power, elegant methods based on classical Newtonian mechanics have been developed, which provide reasonable insight into three-dimensional structures of macromolecules.

Molecular modelling and minimization dynamics

Energy minimization. An *ab initio* calculation based on the solution of the relativistic Schrödinger equation[72] of a large biomolecule is a formidable problem. It is, however, possible to describe the molecular mechanics and dynamics of large biomolecules based on empirical data that implicitly take into account the relativistic and quantum effects. The starting point is the non-relativistic Schrödinger[73] equation within the Born–Oppenheimer approximation[74]. The electronic motions are at least three orders of magnitude faster than nuclear motions, and hence it is possible to solve the electronic Hamiltonian for a fixed position of nuclear coordinates. That is,

$$\mathcal{H}\Phi(R, r) = \varepsilon\Phi(R, r), \tag{1}$$

where $\mathcal{H}$ is the total Hamiltonian of the system, Φ is the total wave function and ε is the total energy, which is a function of both electron coordinates r and nuclear coordinates R. The energy ε which is a function of R is the so-called *potential energy surface*. We can now, based on Born–Oppenheimer approximation, describe the motion of the nuclei on this potential energy surface as

$$\mathcal{H}\Psi(R) = E\Psi(R). \tag{2}$$

Most *ab initio* and semi-empirical codes written for small molecules solve eq. (1) for the electronic wave function and energy as a function of coordinates. Solving eq. (2) is important to understand the time-evolution of a molecule. Since it is very difficult to solve the potential energy surface using eq. (1), an empirical approach to the potential energy surface is taken, especially in the case of large biomolecules. The solution of eq. (2) is called quantum dynamics, but because of the large mass of biomolecules, quantum effects are insignificant and hence the quantum mechanical equation (2) can be replaced by Newton's equation,

$$\frac{dE}{dR} = m\frac{d^2R}{dt^2}. \tag{3}$$

The solution of eq. (3) using empirical derivation of the potential energy surface constitutes the molecular dynamics. Molecular mechanics on the other hand, focuses on the energies for a particular static geometry of the molecule. Using the minimum energy conformation, calculation of transition states, relative energies, force constants (vibrational energies) etc., is carried out.

CHARMM uses a flexible and comprehensive energy function which is a summation of many additive individual energy terms. The potential energy for any molecular structure can be computed given a starting

structure of a polypeptide or a protein along with a set of parameters and the Cartesian coordinates. In addition to the calculation of the total energy of the molecule, the forces acting on each atom of the molecule can also be computed and the latter is important during the molecular dynamics simulations.

In CHARMM, the energy functional is based on additive internal coordinate terms and pairwise non-bonded terms. The internal energy terms include the bond length (E_b), bond angle (E_θ), dihedral angle (E_ϕ) and the improper tortional potentials (deviation from planarity, E_i).

The harmonic approximation is used for E_b and is given by

$$E_b = \sum k_b (r - r_0)^2, \tag{4}$$

where k_b is the force constant and r_0 is the equilibrium bond length. The three-atom bond angle term is likewise approximated using a harmonic potential as

$$E_\theta = \sum k_\theta (\theta - \theta_0)^2, \tag{5}$$

where k_θ is the bending force constant and θ_0 is the equilibrium bond angle.

The dihedral angle corresponding to the torsion energy term is a four-atom potential based on the dihedral angle between the plane of first three atoms and the plane of the next three atoms, and is given by

$$E_\phi = \sum k_\phi [1 + \cos(n\phi - \delta)], \tag{6}$$

where k_ϕ is the torsional force constant, ϕ is the equilibrium dihedral angle and n is the periodicity of the potential.

The improper torsional potential maintains chirality about a tetrahedral extended heavy atom and also maintains planarity about sp^2 hybridized atoms.

$$E_i = \sum k_i (i - i_0)^2, \tag{7}$$

where k_i is the force constant and i_0 is the equilibrium improper angle.

The external potential energy terms take into account of electrostatic (E_{el}) and van der Waals interaction (E_{vdW}). These terms are additively calculated for individual atom-pairs which are not directly bonded.

Electrostatic energy arises via the classical coulombic forces among polar and charged atoms. The electrostatic forces act at a larger range compared to van der Waals forces and are calculated using the monopole approximation as

$$E_{el} = \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{8}$$

where q_i and q_j are the net charges on two non-bonded atoms, i and j, separated by a distance r_{ij} and ε_0 is the dielectric constant of the medium. In other modelling programs, a combination of polarization effects and higher moments (dipole, quadrupole etc.) have been employed.

The van der Waals forces between a pair of non-bonded atoms include electronic-shell repulsion forces, nuclear–nuclear repulsion forces and London dispersion forces[75] due to dipole-induced dipole interaction that are attractive and act at a larger range than the repulsive forces. These are best described using the Lennard-Jones 6–12 potential[76] and are accounted for by setting

$$E_{vdW} = \sum_{i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right), \tag{9}$$

where A_{ij} and B_{ij}s are parametrized based on atom polarizabilities and the effective number of valence electrons.

CHARMM has also some special energy terms which are certain constraints[77,78] that restrict certain types of molecular deformations during the manipulation of the structure through minimization dynamics. These include fixed constraints (allowing certain atoms or interatomic distances fixed throughout the simulation), atom harmonic constraints (allowing only limited displacement of certain atoms), dihedral constraints (fixing certain torsional motions, while observing other motions in the rest of the molecule) and NOE constraints (the NMR-derived experimental internuclear distances being used to fix certain interatomic distances in the molecule).

Once the potential energy functions corresponding to bonded and non-bonded atoms have been described, the next step is to find a set of coordinates for the molecule under consideration that corresponds to the minimum potential energy. The basis of the minimization is equation (3) which states that the derivative of the energy with respect to coordinates (r, θ, ϕ, i, cf equations [4]–[7]) provides the force acting on each atom and the global minimum in the potential energy will be achieved when there is no net force acting on any of the atoms.

There are several methods to perform the minimization of the potential energy function. The *Steepest descent method*[79] calculates the force vector acting on each atom and moves the atom in the direction of the force by a step size (~0.1 to 0.2 Å) and repeats for all atoms until a minimum in the total potential energy is arrived at. The step size can be adjusted to accelerate the convergence, and when the minimum is overshot, the stepsize is reduced. Convergence is not guaranteed but the steepest descent strategy improves a very poor starting conformation.

The second method is the *conjugate gradient technique*[80], and has better convergence characteristics than the steepest descent technique. It essentially consists of an iterative procedure, which uses the previous history of minimization step and the current gradient to determine the next step. It can be shown that the method converges to the minimum energy in N steps for a quadratic energy surface with N number of degrees of freedom. The main drawback is that with a very poor starting conformation, the conjugate gradient technique is likely to

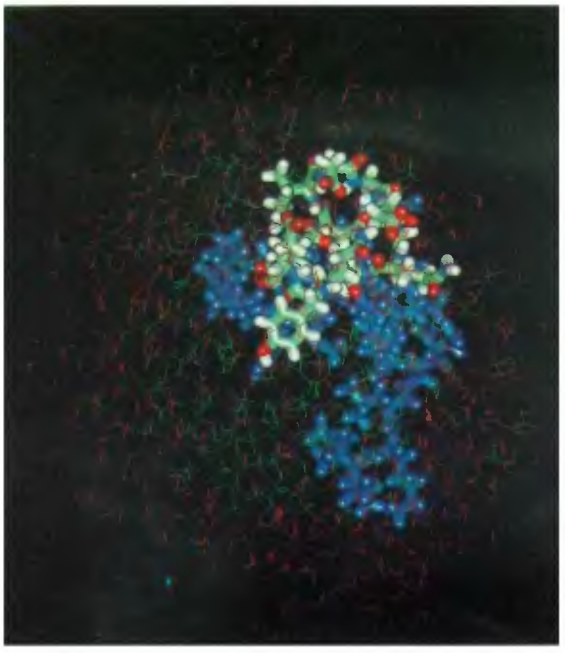


Figure 15. ***a***, Licorice model of the energy optimized and thermally equilibrated structure of the helix-loop (4–4) fragment of CaM arrived at using QUANTA/CHARMM. The solvent molecules are also shown (red: water; green: methanol); ***b***, The ribbon model of the energy optimized and thermally equilibrated structure of the helix-loop (4–4) fragment of CaM. The helix region corresponding to the crystal structure of the native CaM is shown in blue and the loop region in pink.

Figure 16. ***a***, Licorice model of the energy optimized and thermally equilibrated structure of the helix-loop-helix (4-4-4) fragment of CaM arrived at using QUANTA/CHARMM. The solvent molecules are also shown (red: water; green: methanol), ***b***. The ribbon model of the energy optimized and thermally equilibrated structure of the helix–loop–helix (4–4–4) fragment of CaM. The helix region corresponding to the crystal structure of the native CaM is shown in blue and the loop region in pink.

a

b

generate numerical overflows. CHARMM uses an algorithm which has a better interpolation scheme and an improved automatic stepsize selection.

A third alternative is the *conjugate gradient Powell* method, which is an improved conjugate gradient algorithm. A fourth method that is available in CHARMM is the *Newton Raphson minimization* procedure[81]. It involves the computation of the derivative of the gradients of the potential energy functional which is a matrix of size N^2 for N degrees of freedom. The procedure is to find a point where the gradient will be zero, assuming a quadratic potential. CHARMM accomplishes this by diagonalizing the second derivative matrix and finding the optimum step size along each eigenvector. The method as implemented in CHARMM includes a facility to avoid saddle points and the procedure converges rapidly. The major disadvantage is the requirement of large memory and long computational time.

A fifth method, the *adapted basis Newton Raphson* method, applies the Newton Raphson algorithm to a subspace of a coordinate vector, spanned by the displacement coordinates of the last position. At each step, the residual gradient vector is calculated and used to add steepest descent step into the Newton Raphson step. This is the most commonly used method for large molecules.

The minimization of the total energy will critically depend on the list of nonbonded atoms and their coulombic and van der Waals interaction energy must be calculated prior to energy evaluation. Specification of the list of the nonbonded atoms and the cut off distance are to be given or taken from the default values. Creating a list of hydrogen bonds is an option, but not recommended for standard minimization.

Molecular dynamics. A completely-minimized structure arrived at using one of the above procedures represents the molecular system at 0 K. To achieve the actual temperature at which the system exists (298 K, normally) it is important to add kinetic energy slowly and uniformly at periodic time intervals by randomly assigning velocities to each atom. Usually this process known as heating can be carried over, in several pico seconds depending on the size of the system. The essential requirement is that Newton's equation of motion is to be solved to determine the changes in atomic coordinates and velocities as a function of time such that the mean kinetic energy of the molecular system containing N atoms satisfies the

Figure 17. *a*, Licorice model of the energy optimized and thermally equilibrated structure of the loop exchanged analog (4–4–4) fragment of CaM arrived at using QUANTA/CHARMM. The solvent molecules are also shown (red: water; green: methanol); *b*, The ribbon model of the energy optimized and thermally equilibrated structure of the loop exchanged analog (4–4–4) fragment of CaM. The helix region corresponding to the crystal structure of the native CaM is shown in blue and the loop region in pink.

following equation,

$$\frac{1}{2}\sum m_i\langle v_i^2\rangle = \frac{3}{2}Nk_BT. \quad (10)$$

All dynamics simulations begin with an initial structure sometimes derived from X-ray data; the data from X-ray structure can often exhibit considerable steric overlap among atoms due to the solid state packing effects. Therefore, these starting structures from X-ray data, need to be relaxed, using the above minimization procedure.

The first step in dynamics is *heating* which is accomplished by assigning random initial velocities according to the Gaussian distribution appropriate for the low temperature. The temperature is gradually increased by assigning larger random velocities to each atom at predetermined time steps.

Equilibration is achieved by allowing the system to evolve for a period of time by integrating the equation of motion until the average temperature and the structure remain stable. This is facilitated by periodically reassigning the velocities appropriate to the desired temperatures. The procedure is continued until the various statistical properties of the system become time independent.

The final molecular dynamics simulation takes the equilibrated structure obtained above as the starting point. In a typical simulation the trajectory traces the motion of the macromolecule, through a period of about 10 ps. Provision is also made, just as in the minimization procedure, to update the nonbonded interaction periodically.

Constraints may also be used during molecular dynamics simulations when information is available for certain rigid parts of the molecule, where the internuclear distances are not likely to change. Additional constraints may also be imposed based on distance information derived from 2D NOESY.

It is also possible to incorporate solvent molecules in the whole minimization and dynamics process. Most modelling packages use information available from the previous X-ray crystallographic studies. The major source of X-ray crystallography data is the Cambridge Protein Data Bank (PDB)[82–84]. Usually the program has the provision to match a test sequence to other closely related homologs from already available X-ray crystallographic data. Although there may not be an exact sequence identity available in the PDB, even closely related homologs will be a good starting point for minimizing the energy.

As a starting exercise the energy minimization was attempted for a synthetic CaM fragment which corresponds to the fourth calcium binding site of native CaM except that we have chosen an incomplete EF hand comprising of 32 residues, corresponding to a helix-loop fragment.

The starting geometry for this fragment was taken from the PDB file containing the NMR NOE derived structure of *Drosophila* CaM[17]. To this structure we added 15 Å radius of methanol excluding those areas where methanol is not likely to be associated. Further an 8 Å radius extent of water was incorporated. The structure of the molecule in methanol and water was minimized to get a starting low temperature geometry. This geometry showed the presence of a small helix in the expected region. In order to get the geometry at room temperature the system was subjected to dynamic simulation. For this the minimum energy geometry was heated to 300 K, equilibrated at that temperature, and the CHARMM dynamics was carried out for 60 ps.

The results (Figure 15 *a*) indicate that at the N terminal and the C-terminal ends, the peptide 4–4 adopts a random coil structure with the middle existing in α-helical conformation, as expected from the structure of the native CaM. This is better illustrated in the ribbon model shown in Figure 15 *b*.

After this, modelling on the EF hand corresponding to 4–4–4 as well as the loop exchanged 4–1–4 (Table 2) was carried out. The coordinates for the starting geometry for these two fragments were also taken from the 2D NMR-derived structure of CaM. For the loop exchanged fragment, the data were manipulated so as to splice the loop-4 from 4–4–4 and the loop-1 of the fragment 1–1–1 was attached. In both these fragments solvents, methanol and water were incorporated as before. As mentioned previously the modelling was carried out using the normal procedure of minimization, heating, equilibration, production, etc. The resulting structures are given in Figures 16 and 17. It is at once evident that the helical content of the loop exchanged 4–1–4 is higher than that of 4–4–4. Secondly, in 4–1–4 the flanking helical regions are well characterized right up to both the extremes. In the case of the 4–4–4, on viewing from the N terminal end the α-helical conformation extends to almost one third of the *otherwise* loop region. The ribbon models in the Figures 16 *b* and 17 *b* faithfully reproduce the EF hand motifs in both the fragments.

While the molecular modelling studies have predicted well the secondary structures of these synthetic CaM fragments reasonably in consonance with their structure in the native CaM crystal as well as the solution structure of *Drosophila* CaM, NMR results of CaM fragments indicate only an average structure probably involving a rapid interconversion between helix and extended structure. The peptides under consideration, being small, might exist in a number of conformations characterized by very small barriers and a study of all these minimum energy conformations has not been undertaken. The solution structure of calcium-free CaM has also been reported[85].

Conclusions

Two-dimensional NMR coupled with FT IR and CD can be used effectively to understand the three-dimensional

structure of fair-sized biomolecules. The data derived from experiment can be used in a judicious way to get reasonable approximations to three-dimensional solution structure of large biomolecules such as proteins, nucleic acids and peptides using molecular modelling for which a number of softwares are available. The above approach has been employed to study the three-dimensional structure of several synthetic fragments of calmodulin corresponding to the calcium-binding sites. The changes in secondary structural content upon exchanging the loops of different binding domains have also been addressed. Such an approach holds good promise to our understanding of the three-dimensional structure of large biomolecules and will help us in unravelling structure/property correlation.

1. Rasmussen, H., Goodman, D. B. P., Friedmann, N., Allen, J. A. and Kurokawa, K., in *Handbook of Physiology, Endocrinology* (ed. Aurbach, G. D.), Williams-Wilkins, Baltimore, 1976, pp. 225.
2. Rubin, R. P., *Calcium and the Secretory Process*, Plenum, New York, 1974.
3. Dedman, J. R., Brinkley, B. R. and Means, A. R., *Adv. Cyclic Nucleotide Res.*, 1979, **11**, 131.
4. Herzberg, O. and James, M. N. G., *Nature*, 1985, **313**, 653–659.
5. Herzberg, O. and James, M. N. G., *Biochemistry*, 1985, **24**, 5298–5302.
6. Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M. and Sundaralingam, M., *J. Biol. Chem.*, 1988, **263**, 1628–1647.
7. Kretsinger, R. H. and Nockolds, C. E., *J. Biol. Chem.*, 1973, **248**, 3313–3326.
8. Kumar, V. D., Lee, L. D. and Edwards, B. F. P., *Biochemistry*, 1990, **29**, 1404–1412.
9. Cox, J. A., *Biochem. J.*, 1988, **249**, 621–629.
10. Klee, C. B., Crouch, T. H. and Richman, P. G., *Annu. Rev. Biochem.*, 1980, **49**, 489–515.
11. Klee, C. B. and Vanaman, T. C., *Adv. Protein Chem.*, 1982, **35**, 213–321.
12. Szebenyi, D. M. E., Obendorf, S. K. and Moffat, K., *Nature*, 1981, **294**, 327–332.
13. Szebenyi, D. M. E. and Moffat, K., *J. Biol. Chem.*, 1986, **261**, 8761–8777 and references therein.
14. Claudia, A., da Silva, R. and Reinach, F. C., *TIBS*, 1991, **16**, 53–57.
15. Means, A. R. and Dedman, J. R., *Nature*, 1980, **285**, 73–77.
16. Ikura, M., Kay, L. E. and Bax, A., *Biochemistry*, 1990, **29**, 4659–4667.
17. Ikura, M., Spera, S., Barbato, G., Kay, L. E., Krinks, M. and Bax, A., *Biochemistry*, 1991, **30**, 9216–9228.
18. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. and Cook, W. J., *Nature*, 1985, **315**, 37-40.
19. Babu, Y. S., Bugg, C. E. and Cook, W. J., *J. Mol. Biol.*, 1988, **204**, 191–204.
20. Wang, C. L. A., *Biochem. Biophys. Res. Commun.*, 1985, **130**, 426–430.
21. Wang, C. L. A., Leavis, P. C. and Gergely, J., *Biochemistry*, 1984, **23**, 6410–6415.
22. Killhoffer, M. C., Demaille, J. G. and Gerard, D., *FEBS Lett.*, 1980, **116**, 269–272.
23. Wang, C. L. A., Aquaron, R. R., Leavis, P. C. and Gerard, D., *Eur. J. Biochem.*, 1982, **124**, 7–21.
24. Wallace, R. W., Tallant, E. A., Dockter, M. E. and Cheung, W. Y., *J. Biol. Chem.*, 1983, **257**, 1845–1854.
25. Ikura, M., Hiraoki, T., Hikichi, K., Mikuni, T., Yazawa, M. and Yage, K., *Biochemistry*, 1983, **22**, 2573–2579.
26. Forsén, S., Andersson, A., Drakenberg, T., Teleman, O., Thulin, E. and Vogel, H. J., in *Calcium Binding Proteins* (eds De Bernard, B., Sottocasa, G. L., Sandri, G., Carafoli, E., Taylor, A. N., Vanaman, T. C. and Williams, R. J.), Elsevier, Amsterdam, 1983, p. 121.
27. Andersson, A., Forsén, S., Thulin, E. and Vogel, H. J., *Biochemistry*, 1983, **22**, 2309–2313.
28. Burger, D., Cox, J. A., Comte, M. and Stein, E. A., *Biochemistry*, 1983, **23**, 1966–1971.
29. Tanokura, M. and Yamada, K., *J. Biochem. (Tokyo)*, 1984, **95**, 643–649.
30. Iida, S. and Potter, J. D., *J. Biochem.*, 1986, **99**, 1765.
31. Milos, M., Schaer, J. J., Comte, M. and Cox, J. A., *Biochemistry*, 1986, **25**, 6279–6287.
32. Marsden, B. J., Shaw, G. S. and Sykes, B. D., *Biochem. Cell Biol.*, 1990, **68**, 587–601.
33. Reid, R. E., *J. Biol. Chem.*, 1990, **265**, 5971–5976.
34. Shaw, G. S., Hodges, R. S. and Sykes, B. D., *Biochemistry*, 1991, **30**, 8339–8347.
35. Beckingham, K., *J. Biol. Chem.*, 1991, **266**, 6027–6030.
36. Babu, A., Su, H., Ryu, Y. and Gulati, J., *J. Biol. Chem.*, 1992, **267**, 15469–15474.
37. Borin, G., Ruzza, P., Rossi, M., Calderan, A., Marchiori, F., Peggion, E. and Temussi, P., *Biopolymers*, 1989, **28**, 353–369.
38. Seamon, K. B., *Biochemistry*, 1980, **19**, 207–215.
39. Krebs, J. and Carafoli, E., *Eur. J. Biochem.*, 1982, **124**, 619–627.
40. Ikura, M., Hiraoki, T., Hikichi, K., Mikuni, T., Yazawa, M. and Yagi, K., *Biochemistry*, 1983, **22**, 2568–2572 and 2573–2579.
41. Dalgarno, D. C., Klevit, R. E., Levine, B. A., Williams, R. J. P., Dobrowolski, Z. and Drabikowski, W., *Eur. J. Biochem.*, 1984, **138**, 281–289.
42. Klevit, R. E., Dalgarno, D. C., Levine, B. A. and Williams, R. J. P., *Eur. J. Biochem.*, 1984, **139**, 109–114.
43. Ikura, M., *Biochim. Biophys. Acta*, 1986, **872**, 195–200.
44. Teleman, A., Drakenberg, T. and Forsén, S., *Biochim. Biophys. Acta*, 1986, **873**, 204–213.
45. Aulabaugh, A., Niemczura, W. P., Blundell, T. L. and Gibbons, W. A., *Eur. J. Biochem.*, 1984, **143**, 409–418.
46. Ikura, M., Minowa, O., Yazawa, M., Yagi, K. and Hikichi, K., *FEBS Lett.*, 1987, **219**, 17–21.
47. Motta, A., Tancredi, T., Borin, G., Marchiori, F., Peggion, E. and Temussi, P., *Biopolymers*, 1988, **27**, 789–803.
48. Foffani, M. T., Battistutta, R., Calderan, A., Ruzza, P., Borin, G. and Peggion, E., *Biopolymers*, 1991, **31**, 671–681.
49. Caday, C. G. and Steiner, R. F., *J. Biol. Chem.*, 1985, **260**, 5985–5990.
50. Sharma, Y., Gopalakrishna, A., Balasubramanian, D., Fairwell, T. and Krishna, G., *FEBS Lett.*, 1993, **326**, 59–64.
51. Merrifield, R. B., *Adv. Enzymol.*, 1969, **32**, 221–296.
52. Schippers, P. H. and Dekkers, P. J. M., *Anal. Chem.*, 1981, **53**, 778–782.
53. States, D. J., Haberkorn, R. A. and Ruben, D. J., *J. Magn. Reson.*, 1982, **48**, 286–292.
54. Marquardt, D. W., *J. Soc. Ind. Appl. Math.*, 1963, **11**, 431.
55. Jacobs, D. A. H., in *The State of the Art in Numerical Analysis* (ed. Dennis, von J. E.), Academic Press, London, 1977, Kap. III.2.
56. Yang, J.-T., Wu, C.-S. C. and Martinez, H. M., *Methods Enzymol.*, 1986, **130**, 208.
57. Wüthrich, K., *NMR of Proteins and Nucleic Acids*, J. Wiley, New York, 1986.
58. Aue, W. P., Bartholdi, E. and Ernst, R. R., *J. Chem. Phys.*, 1976, **64**, 2229–2246.
59. Piantini, U., Sørenson, O. W. and Ernst, R. R., *J. Am. Chem. Soc.*, 1982, **104**, 6800–6801.
60. Rance, M., Sørenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. and Wüthrich, K., *Biochem. Biophys. Res. Commun.*, 1983, **117**, 479–485.

51. Braunschweiler, L. and Ernst, R. R., *J. Magn. Reson.*, 1983, **53**, 521–528.
52. Neuhaus, D., Wagner, G., Vašák, M., Kägi, J. H. R. and Wüthrich, K., *Eur. J. Biochem.*, 1985, **151**, 257–273.
53. Billeter, M., Braun, W. and Wüthrich, K., *J. Mol. Biol.*, 1982, **155**, 321–346.
54. Gomathi, L. and Subramanian, S., *Proc. Indian Acad. Sci. (Chem. Sci.)*, 1994, **106**, 1525–1536.
55. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M., *J. Comp. Chem.*, 1983, **4**, 187–217.
56. Hagler, A. T., in *Conformation in Biology and Drug Design, The Peptides* (ed. Meienhofer, J.), Academic Press, New York, 1985, vol. 7, pp. 213.
57. Weiner, P. K. and Kollman, P. A., *J. Comput. Chem.*, 1981, **2**, 287.
58. Weiner, S. J., Kollman, P. A., Naguyen, D. T. and Case, D. A., *J. Comp. Chem.*, 1986, **7**, 230.
59. Schulz, G. E. and Schirmer, R. H., *Principles of Protein Structure*, Springer, New York, 1979.
70. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, **7**, 95–99.
71. Ramachandran, G. N. and Sasisekaran, V., *Adv. Protein Chem.*, 1968, **23**, 283–438.
72. Schrödinger, E., *Ann. Physik*, 1926, **97**, 361.
73. Daudel, R., Lefebvre, R. and Moser, C., *Quantum Chemistry Methods and Applications*, Interscience, New York, 1959.
74. Born, M. and Oppenheimer, J. R., *Ann. Physik*, 1927, **84**, 457.
75. Gray, C. G. and Gubbins, K. E., *Theory of Molecular Fluids*, Clarendon Press, Oxford, 1984, vols 1–3.
76. Lennard-Jones, J. and Pople, J. A., *Proc. R. Soc.*, 1951, **A205**, 155.
77. Havel, T. and Wüthrich, K., *Bull. Math. Biol.*, 1984, **46**, 673–698.
78. Braun, W. and Go, N., *J. Mol. Biol.*, 1985, **186**, 611-626.
79. Fletcher, R., *Practical Methods of Optimization*, Wiley, Chichester, 1981.
80. Fletcher, R. and Reeves, C. M., *Comput. J.*, 1963, **7**, 149.
81. Ermer, O., *Structure and Bonding*, 1976, **27**, 161.
82. Baldwin, R. L., *TIBS*, 1989, **14**, 291.
83. Kanzman, W., *Adv. Protein Chem.*, 1959, **14**, 1.
84. Kyte, J. and Doolittle, R. F., *J. Mol. Biol.*, 1982, **157**, 105–132.
85. Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klec, C. B. and Bax, A., *Nature Struct. Biol.*, 1995, **2**, 768–783.

ACKNOWLEDGEMENT. We thank Dr Tsao Desiree and Dr James Gruschus of NHLBI, National Institute of Health, Bethesda, Maryland, USA for helpful discussions.