

# Molecular Dynamics Simulations of Peptide Fragments from Hen Lysozyme: Insight into Non-native Protein Conformations

Lorna J. Smith<sup>1\*</sup>, Alan E. Mark<sup>2</sup>, Christopher M. Dobson<sup>1</sup>  
and Wilfred F. van Gunsteren<sup>1,2</sup>

<sup>1</sup>Oxford Centre for Molecular Sciences and New Chemistry Laboratory, University of Oxford, South Parks Road Oxford OX1 3QR, England

<sup>2</sup>Department of Physical Chemistry, Swiss Federal Institute of Technology Zürich ETH-Zentrum, 8092 Zürich Switzerland

Molecular dynamics simulations of four peptides taken from the hen lysozyme sequence have been used to generate models for non-native protein conformations. Comparisons between the different peptides and with experimental data for denatured lysozyme and peptide fragments provides insight into the characteristics of the conformational ensembles populated in these non-native states and the dependence of their structural features on the amino acid sequence. For the denatured conformers populated local contacts dominate in determining the properties observed in the trajectories, all four peptides showing similar characteristics. These include a significant increase in the number of main-chain O(*i*)–NH(*i* + 2) hydrogen bonds and hydrogen bonds involving side-chain groups, this increase compensating to a large extent for the loss of hydrogen bonds involved in helical or  $\beta$ -sheet secondary structure in the native fold, and the generation of a population of collapsed states with local clusterings of hydrophobic groups. The hydrophobic clusters enable at least partial burial of many side-chains exposed by the loss of tertiary contacts on denaturation and provide models that may explain the experimentally observed protection of amides from hydrogen exchange and the existence of residual secondary structure in non-native species of lysozyme. The results suggest that this approach has an important role to play in aiding the interpretation of experimental data for conformationally disordered non-native states of proteins.

© 1998 Academic Press

**Keywords:** molecular dynamics; computer simulation; protein folding; NMR spectroscopy; denatured protein.

\*Corresponding author

## Introduction

There is increasing interest in characterising unfolded and partially folded protein conformations as it has been recognised that such states can give considerable insight into issues such as protein stability and folding, and the relationship between the sequence and three-dimensional structure of a protein (Dill & Shortle, 1991; Dobson, 1992; Shortle, 1996a; Smith *et al.*, 1996). In addition, non-native states are involved in a range of important biological processes, including the transport of

proteins across membranes and protein degradation within the cell (Dill & Shortle, 1991; Ptitsyn, 1995). They are also involved in protein aggregation which is of significance with regard to our understanding of amyloid associated diseases such as Alzheimer's and the spongiform encephalopathies (Thomas *et al.*, 1995). Recently a number of proteins that are unfolded under physiological conditions but are biologically active have also been identified (Gast *et al.*, 1995; Kriwacki *et al.*, 1996; Weinreb *et al.*, 1996; Penkett *et al.*, 1997). Defining the conformational properties of states that are not fully folded is, however, challenging as they are in general ensembles of interconverting conformers. This can complicate both the experimental measurements, particularly in the case of NMR techniques if there is relatively slow interconversion between the conformers, and also the

Abbreviations used: MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; RMSD, root-mean-square difference.

E-mail address of the corresponding author:  
lorna.smith@chemistry.oxford.ac.uk

interpretation of any parameters that are extracted, since averaging of the data across the conformational ensemble must be taken into account (Shurtle, 1996b; Smith *et al.*, 1996). It is, therefore, of interest to complement experimental investigations with theoretical studies that can propose models for non-native states for comparison with experimental data. This is the approach we employ here by exploring how molecular dynamics simulations could be used effectively in this way, analysing the properties of partly folded and unfolded conformations present in a number of simulation trajectories.

In addition to characterising the conformational properties of non-native states it is important to rationalise the dependence of their observed structural features on their sequence. As we move from folded towards more unfolded structures local interactions between residues that are close in the primary sequence can be expected to become increasingly dominant in determining the conformations adopted. However, long range contacts between residues that are more distant in the sequence, often of a non-specific nature, may still play an important role particularly in stabilising the more compact states of a given conformational ensemble. Here we investigate possible factors that define the conformational distributions in non-native states by using a protein dissection approach in conjunction with MD simulations. This protein dissection strategy, in which a series of peptide fragments is characterised whose sequences together encompass that of the protein of interest, has been employed in a number of experimental studies attempting to understand protein structure and folding (e.g. see Oas & Kim, 1988; Dyson *et al.*, 1992a,b; Kemmink & Creighton, 1993; Itzhaki *et al.*, 1995; Muñoz *et al.*, 1995). A similar approach has also been used to probe the intrinsic stabilities of helices from the myoglobin sequence by MD simulations (Hirst & Brooks, 1995). Studies using MD techniques have also been performed for peptide fragments related to the ribonuclease A sequence (Tirado-Rives & Jorgensen, 1991; Soman *et al.*, 1993). Here we perform MD simulations of peptide fragments of lysozyme starting from the native protein co-ordinates. Elevated temperature is used to promote conformational transitions. We consider the sequence dependence of the characteristics of the states formed, examining in particular the similarities and differences in the behaviour of the same sequence in both smaller and larger fragments of the protein.

The work uses the sequence and structure of the protein hen lysozyme. This is an appropriate system for study as both native and denatured states of the protein and its folding behaviour have been extensively characterised by experiment (Blake *et al.*, 1965; Smith *et al.*, 1993; Dobson *et al.*, 1994; Radford & Dobson, 1995; Chen *et al.*, 1996; Schwalbe *et al.*, 1997). Peptide fragments from the lysozyme sequence have also been studied exper-

imentally (Ueda *et al.*, 1994; Yang *et al.*, 1995, 1996; Bolin *et al.*, 1996) and MD simulations of the native protein and its unfolding have been performed (Mark & van Gunsteren, 1992; Hünenberger *et al.*, 1995; Smith *et al.*, 1995). There are therefore opportunities for comparisons of the peptide simulations with an extensive range of data. In this work MD simulations of four peptide fragments of hen lysozyme are analysed (Figures 1 to 3). The first two peptides correspond to the  $\alpha$  and  $\beta$  domains of the protein. The  $\alpha$  domain consists of two chains (residues 1 to 35 and 85 to 129) linked by two disulphide bridges (6 to 127, 30 to 115). In the native structure it comprises four  $\alpha$ -helices (A, residues 5 to 14; B, residues 25 to 34; C, residues 89 to 100; D, residues 109 to 114) and a C-terminal  $3_{10}$  helix (residues 120 to 123). The  $\beta$  domain (residues 36 to 84) in contrast has a  $\beta$ -sheet (residues 41 to 60), long loop and a short  $3_{10}$  helix (residues 80 to 83) in the native fold and contains one disulphide bridge (64 to 80). As the  $\alpha$  domain consists of two chains, this domain has been further divided and each of the constituent chains has been studied separately (AB and CD peptides).

## Results

MD simulations have been run for all four peptides for 800 ps at 300 K and at higher temperatures to promote unfolding (500 K and 700 K for  $\alpha$  and  $\beta$  domains; 500 K for AB and CD peptides) starting from the structures the peptides adopt in the intact native lysozyme fold (referred to throughout as the native conformation). The change in radius of gyration and root-mean-square difference from the native conformations of the four peptides during each of these simulations is shown in Figures 4 and 5. For all four peptides the secondary structure present in the native conformation is preserved throughout the simulations at 300 K, although some changes are seen in the rela-

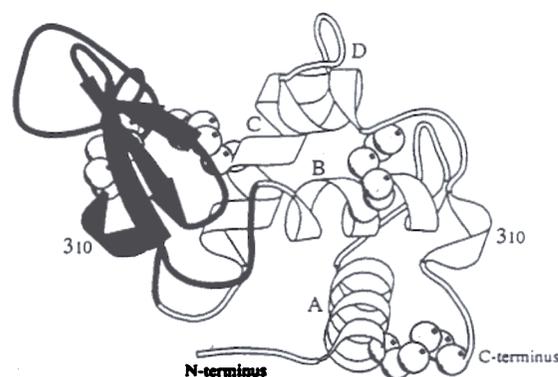
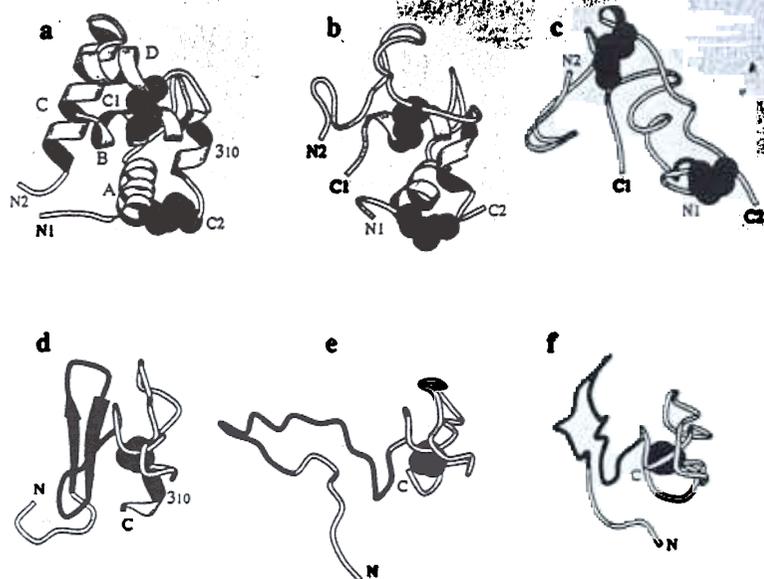


Figure 1. Schematic diagram showing the crystal structure of hen lysozyme. The  $\alpha$  domain (residues 1 to 35; 85 to 129) is shown in white and the  $\beta$  domain (residues 36 to 84) in black. The helices and N and C termini of the protein are labelled. The positions of disulphide bridges is also indicated. The diagram was generated using the program MOLSCRIPT (Kraulis, 1991).

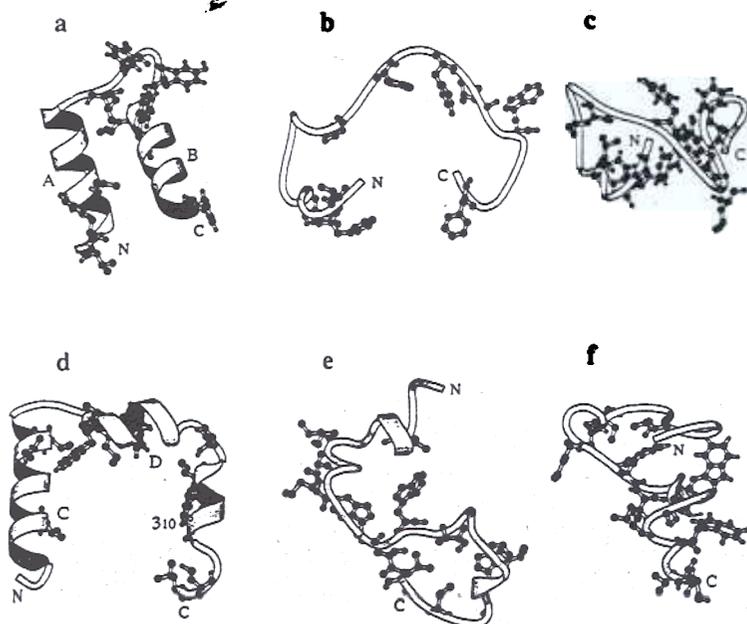


**Figure 2.** Structures from the  $\alpha$  and  $\beta$  domain peptide simulations at 700 K. a,  $\alpha$  Domain native conformation; b,  $\alpha$  domain after 400 ps; c,  $\alpha$  domain after 800 ps; d,  $\beta$  domain native conformation; e,  $\beta$  domain after 670 ps; f,  $\beta$  domain after 800 ps. The secondary structure shown is that identified in each of the structures using the method of Kabsch & Sander (1983). The helices present in the native conformations and the N and C termini are labelled. For the  $\beta$  domain residues 41 to 60, corresponding to the  $\beta$ -sheet region in native lysozyme, are shown in black. The atoms involved in disulphide bridges are indicated by black spheres. The diagram was generated using the program MOLSCRIPT (Kraulis, 1991).

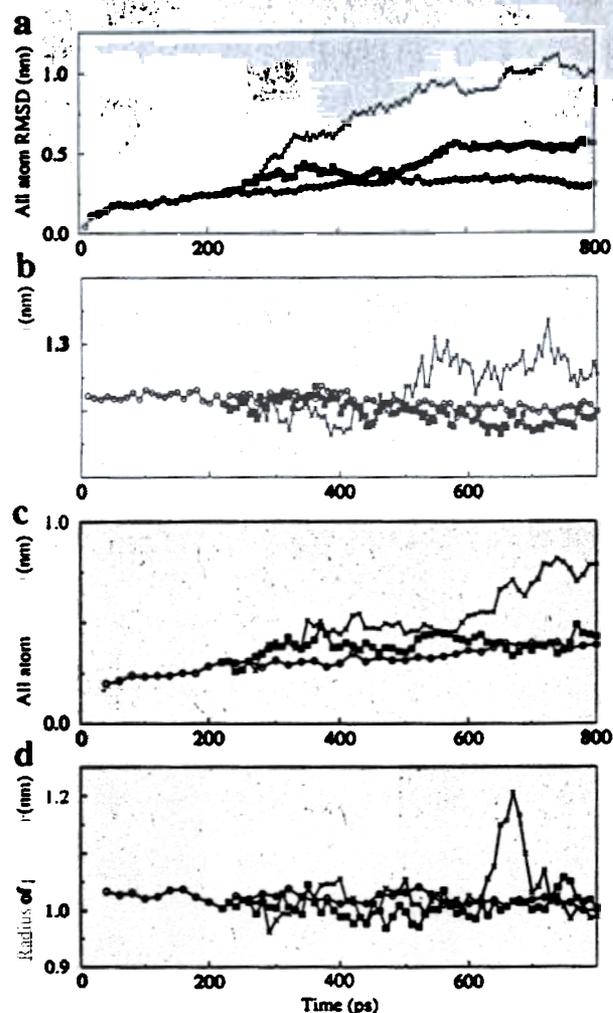
tive orientations of the secondary structure units particularly for the AB and CD peptides. Here, however, the analysis centres on the simulations performed at 700 K for the  $\alpha$  and  $\beta$  domains and those at 500 K for the AB and CD peptides where significant conformational changes have been observed. The analysis concentrates first on the partly folded species present during unfolding and then on the characteristics of the denatured conformers formed. Example structures taken from the simulations are shown in Figures 2 and 3, the characteristics of these conformations being given in Table 1.

In analysing the peptide simulations the conformational disorder characteristic of non-native states has to be taken into account. In order to gain a complete picture of the conformational ensembles that comprise non-native states the

simulations would need to be repeated numerous times. As in the work reported here only one simulation has been performed for each peptide at a given temperature we concentrate on comparisons between the various peptide fragments. The differences observed between the peptides give insight into the breadth of the conformational ensemble accessible to non-native states and the similarities identify the features that are likely to be representative of the full ensemble. In the case of the denatured state this approach also enables the dependence of the results on the sequences and the starting native conformations used in the simulations to be assessed. In the comparisons we also only consider persistent features. This is because the length of the trajectories calculated and the transient nature of some of the non-native species



**Figure 3.** Structures from the AB and CD peptide simulations at 500 K. a, AB peptide native conformation; b, AB peptide after 600 ps; c, AB peptide after 1000 ps; d, CD peptide native conformation; e, CD peptide after 600 ps; f, CD peptide after 1000 ps. The secondary structure shown is that identified in each of the structures using the method of Kabsch & Sander (1983). The helices present in the native conformations and the N and C termini are labelled. The side-chains of Val2, Phe3, Leu8, Leu17, Tyr20, Tyr23, Leu25, Trp28 and Phe34 in the AB peptide and Val92, Ile98, Val99, Met105, Trp108, Val109, Trp111, Thr118, Val120, Trp123, Cys127 and Leu129 in the CD peptide are shown. The diagram was generated using the program MOLSCRIPT (Kraulis, 1991).

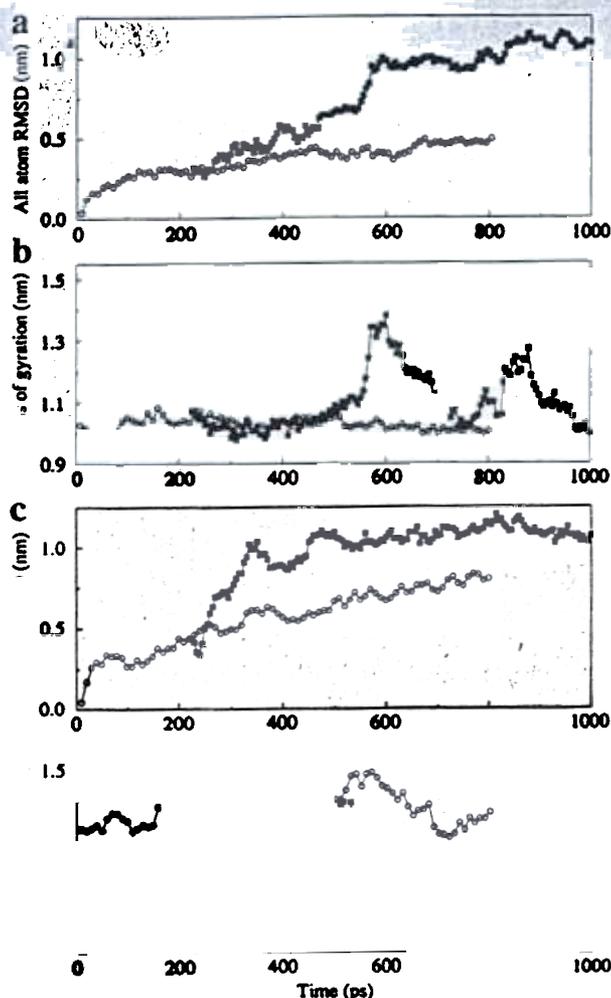


**Figure 4.** Variations in all atom root-mean-square difference from the native conformation and radius of gyration through the  $\alpha$  domain (a and b) and  $\beta$  domain (c and d) peptide simulations at 300 K (open circles), 500 K (filled squares) and 700 K (crosses). The values are for instantaneous structures taken at 10 ps intervals through the 300 K simulations (20 ps for the  $\beta$  domain) and at 5 ps intervals through the 500 K and 700 K simulations (10 ps intervals for the  $\beta$  domain).

formed result in limited statistics for the conformational analysis of more short-lived species.

#### Unfolding behaviour of the peptides during the high temperature simulations

In the simulation of the  $\alpha$  domain, the C helix is lost rapidly at 700 K (Figure 6). A compact core involving the A, B, D and  $3_{10}$  helices is, however, retained until approximately 600 ps. After this time the only significant residual helicity is in the C-terminal region (corresponding to the native D and  $3_{10}$  helices). In the simulation of the  $\beta$  domain the hydrogen bonds between the second and third strands of the  $\beta$ -sheet are lost within 50 ps of reaching 500 K. However, some of those between



**Figure 5.** Variations in all atom root-mean-square difference from the native conformation and radius of gyration through the AB (a and b) and CD (c and d) peptide simulations at 300 K (open circles) and 500 K (filled squares). The values are for instantaneous structures taken at 10 ps intervals through the 300 K simulations and at 5 ps intervals through the 500 K simulations.

the first and second strands (notably 54NH-42O, 44NH-52O, 52NH-44O) persist until approximately 600 ps. After this point the native  $\beta$ -sheet is completely absent (Figure 7). A sudden increase is observed in the radius of gyration of the  $\beta$  domain at about 650 ps. Residues 42 to 53, which form the first two strands of the  $\beta$ -sheet in the native conformation, swing away from the main body of the peptide; within 50 ps these residues return to pack against the peptide core in a non-native conformation with a concurrent reduction in the radius of gyration (Figure 2e and f).

The AB and CD peptides both unfold within 1 ns at 500 K. The AB peptide retains its native-like structure until approximately 550 ps. There is then a rapid expansion with almost all the native-like secondary structure and tertiary contacts being lost (Figure 3b). The CD peptide, in contrast, rapidly

**Table 1.** Characteristics of peptide conformations populated during the simulations

	$\alpha$ Domain 700 K (80 residues)		$\beta$ Domain 700 K (49 residues)		AB peptidic 500 K (35 residue)		CD peptidic 500 K (45 residues)	
	X-ray	400-450 ps	X-ray	650-700 ps	X-ray	600-650 ps	X-ray	600-650 ps
Solvent accessible surface								
Total	51.3	50.8	34.4	40.3	32.4	38.4	42.3	37.1
Non-polar side	25.5	26.9	19.6	21.5	17.9	22.9	24.1	21.4
Polar side	15.5	15.5	9.1	11.5	8.1	8.0	10.4	8.6
Buried residues <sup>a</sup>	24	26	9	7	5	1	1	8
Radius of gyration	1.22	1.18	1.03	1.15	1.03	1.26	1.31	1.09
Long range contacts <sup>b</sup>								
Total <sup>c</sup>	61 (76)	42 (53)	17 (35)	12 (24)	14 (40)	4 (11)	7 (15)	11 (24)
Native-like	-	18	-	5	-	0	-	0
Hydrogen bonds <sup>d</sup>								
All (intra-peptide)	71 (89)	58 (73)	44 (90)	26 (53)	34 (97)	18 (51)	34 (76)	29 (64)
O(i)-NH(i+2)	5 (6)	8 (10)	3 (6)	7 (14)	2 (6)	6 (17)	3 (7)	7 (16)
O(i)-NH(i+3,4,5)	48 (60)	30 (38)	11 (22)	2 (4)	22 (63)	5 (14)	25 (56)	7 (16)
$\beta$ -Sheet or $\beta$ -bridge	2 (3)	2 (3)	9 (18)	4 (8)	2 (6)	2 (6)	0	1 (2)
Main-chain to solvent <sup>d</sup>	68 (85)	68 (85)	62(126)	52 (106)	24 (69)	32 (91)	33 (73)	28 <sup>e</sup> (62)

Data for the simulations are the mean values from ten instantaneous structures taken at 5 ps intervals through the trajectories.

The solvent accessible surface area is given in nm<sup>2</sup> and the radius of gyration in nm.

<sup>a</sup> Number of residues for which the percentage solvent accessibility was less than 20%.

<sup>b</sup> Number of residue pairs (i to i + 5 separation or greater) between which one or more hydrophobic contacts have been identified.

<sup>c</sup> Number per 100 residues given in parentheses.

<sup>d</sup> Hydrogen bonds between the main-chain NH or CO groups and solvent water molecules. For the X-ray structure the conformations present after the first 10 ps of the MD were used; up to this point all the protein atoms had been restrained to their positions in the X-ray co-ordinates.

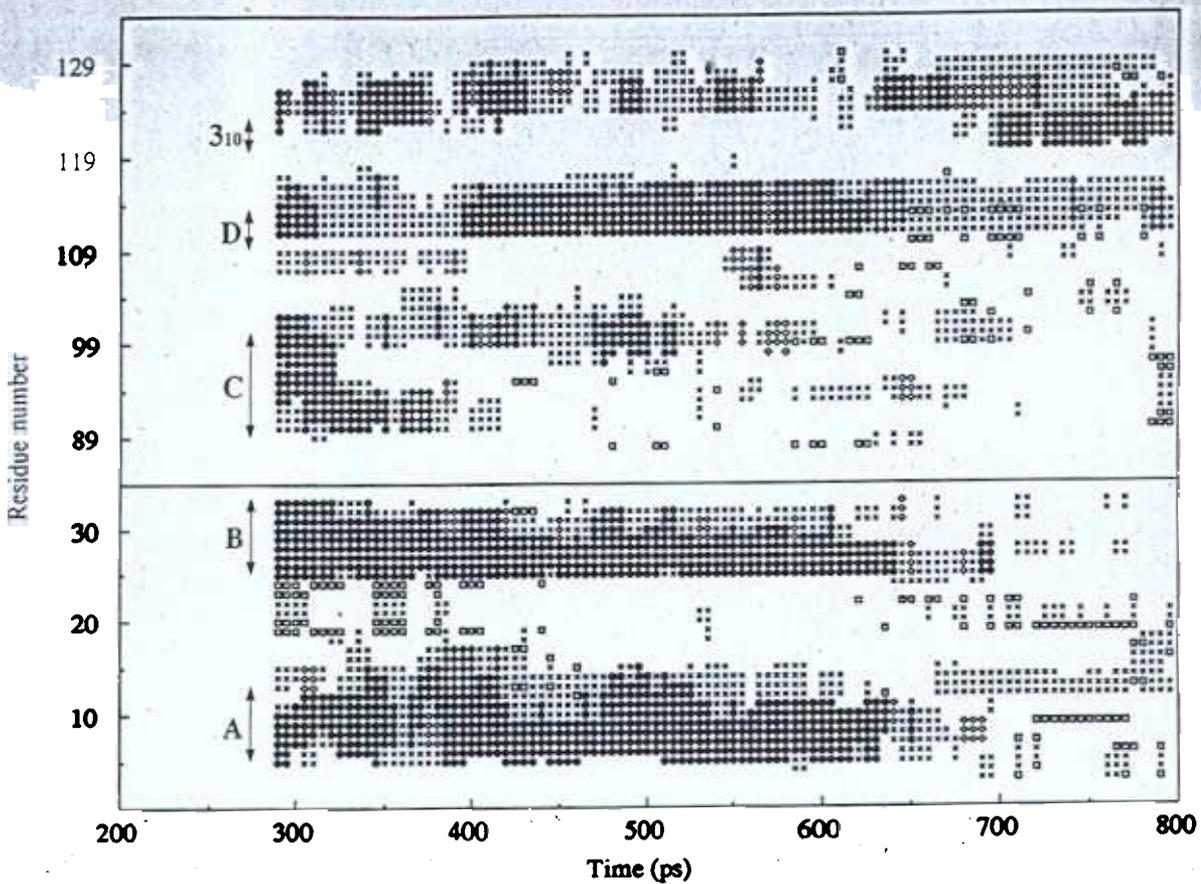


Figure 6. Summary of the secondary structure identified (according to the Kabsch & Sander (1983) procedure) in the  $\alpha$  domain simulation at 700 K. The positions of the helices present in native lysozyme are indicated. Symbols used: filled circle,  $\alpha$ -helix; open diamond,  $3_{10}$ -helix; filled diamond,  $\pi$ -helix; cross, turn; dot, bend; open square,  $\beta$ -strand or bridge.

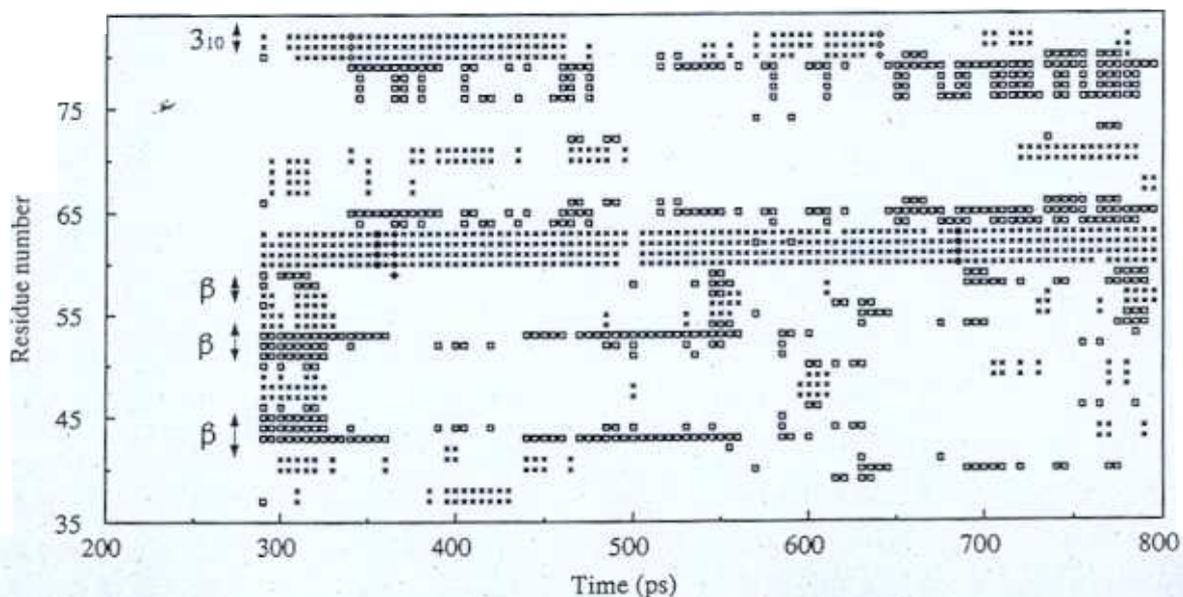
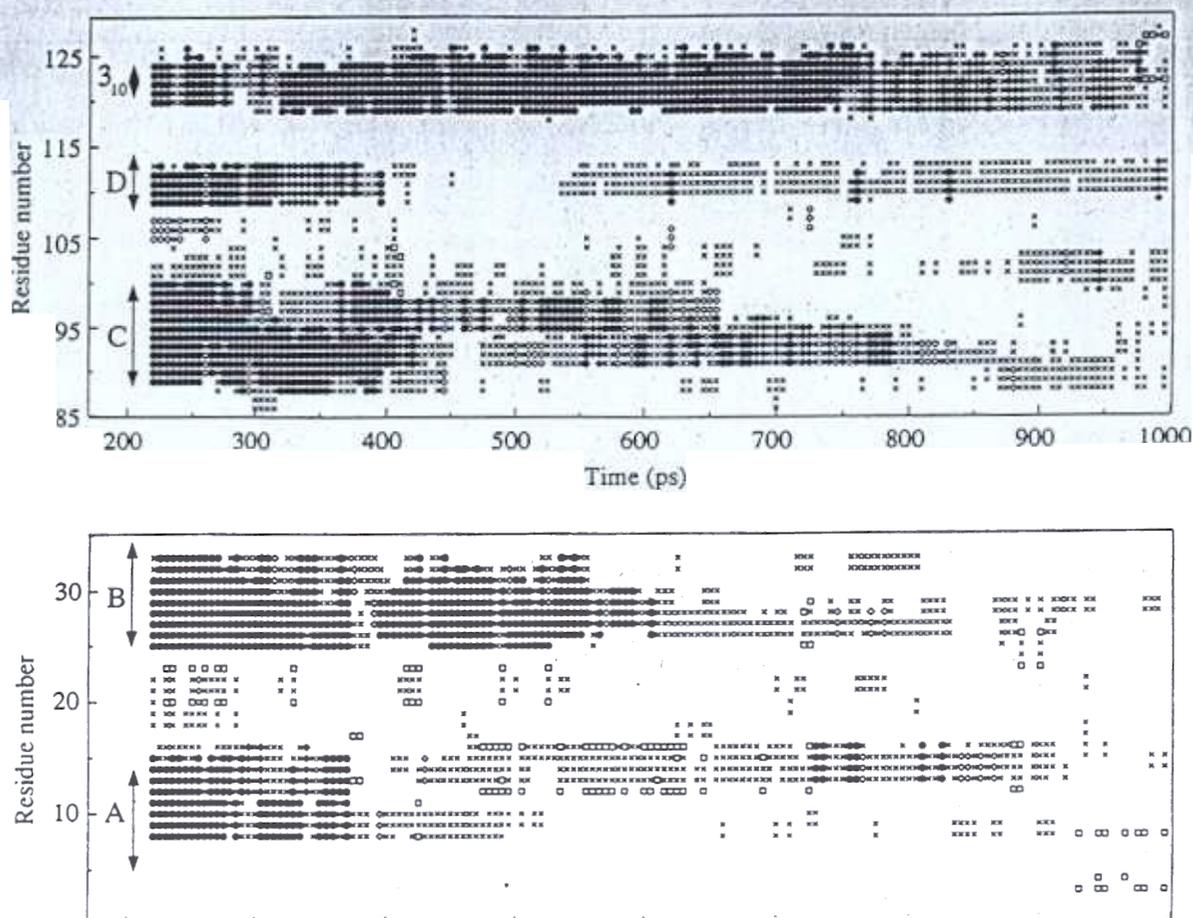


Figure 7. Summary of the secondary structure identified (according to the Kabsch & Sander (1983) procedure) in the  $\beta$  domain simulation at 700 K. The positions of the  $\beta$ -strands and  $3_{10}$  helix present in native lysozyme are indicated. Symbols used: filled circle,  $\alpha$ -helix; open diamond,  $3_{10}$ -helix; filled diamond,  $\pi$ -helix; cross, turn; dot, bend; open square,  $\beta$ -strand or bridge.



**Figure 8.** Summaries of the secondary structure identified (according to the Kabsch & Sander (1983) procedure) in the AB (lower plot) and CD (upper plot) peptide simulations at 500 K. The positions of the helices present in native lysozyme are indicated. Symbols used: filled circle,  $\alpha$ -helix; open diamond,  $3_{10}$ -helix; filled diamond,  $\pi$ -helix; cross, turn; dot, bend; open square,  $\beta$ -strand or bridge.

collapses into a more compact state at 500 K with the loss of much of the native C helix. There is, however, some persistent helicity or turns in the regions corresponding to the native D and  $3_{10}$  helices throughout the simulation (Figure 8). For example, as illustrated in Table 2 in the simulation

of the CD peptide between 600 and 800 ps, two and three  $O(i)-NH(i+4)$  hydrogen bonds involving residues corresponding to the native D and  $3_{10}$  helices, respectively, are populated significantly ( $\geq 40\%$ ). Interestingly, although a persistent helix is identified by the Kabsch & Sander (1983) criteria

**Table 2.** Population of  $O(i)-NH(i+3)$  and  $O(i)-NH(i+4)$  hydrogen bonds and mean  $\alpha H(i)-NH(i+3)$  interproton distances in the CD peptide simulation from 600 to 800 ps at 500 K

Residue ( <i>i</i> )	Percentage hydrogen bond occupancy		Mean $\alpha H(i)-NH(i+3)$ distance ( $r^{-3}$ ) <sup>-1/3</sup> (nm)
	( <i>i, i+3</i> )	( <i>i, i+4</i> )	
	6		0.49
	0		0.49
	27		0.46
	19		0.44
<b><math>3_{10}</math> helix region</b>			
119	11	71	0.48
120	3	48	0.45
121	31	45	0.48
122	17	25	0.44

only for the  $3_{10}$  and not for the D helix sequence during this period of the simulation, mean  $\alpha H(i)$ -NH( $i+3$ ) interproton distances less than 5 Å are observed for residues in both of these sequences. Short  $\alpha H(i)$ -NH( $i+3$ ) distances are often used to recognise helices by NMR techniques through the observation of the corresponding NOEs (Wüthrich, 1986). These results suggest that the observation of  $\alpha H(i)$ -NH( $i+3$ ) NOEs is by itself insufficient to prove conclusively the presence of a helical element.

It is interesting to consider some of the similarities between the unfolding behaviour and the distributions of conformations sampled in the simulations of the intact  $\alpha$  domain and the separate AB and CD peptides. The two dissections performed on the native lysozyme structure in creating these peptide fragments differ significantly in character. The dissection to give the separate  $\alpha$  and  $\beta$  domains results in only a limited loss of tertiary contacts. Most of the effects are localised at the termini of the chains, in the vicinity of the 76-94 disulphide bridge that in the intact protein links the two domains and in the residues that form the second and third strands of the  $\beta$ -sheet and the C helix in the native conformation. In contrast the dissection of the  $\alpha$  domain to give the separate AB and CD peptides gives a significant loss of contacts along the length of both chains. There is also a considerable increase in the exposure of a number of hydrophobic side-chains, including Leu17, Trp28, Ile98, Met105 and Trp108, that cluster within the major hydrophobic core in the intact protein.

In comparing the trajectories it can be seen that in both the  $\alpha$  domain and the CD peptide simulations the C helix is lost rapidly (although some helicity remains for longer in the C-terminal region of this helix in the CD peptide simulation). This loss of the C helix is associated with the burial of a number of hydrophobic side-chains that were exposed in both peptide fragments in their native conformations, but which in the intact lysozyme fold lie on the surface of this helix that packs on to the  $\beta$  domain. The changes on the loss of the C helix in the two peptides are not, however, identical. For example, the percentage accessibility of Ile98 in the  $\alpha$  domain reduces from 59% in the native conformation to 1% after 450 ps at 700 K. This residue stays exposed in the CD peptide simulation at 450 ps (accessibility 72%) but the accessibility of Ala90 falls from 72% in the native conformation to 15% (28% in the  $\alpha$  domain) after 450 ps. These results, therefore, suggest that contacts between the  $\alpha$  and  $\beta$  domains may play an important role in stabilising the native C helix, particularly through enabling the burial of hydrophobic groups.

A second similarity between the trajectories is in the unfolding behaviour of the A and B helices. In the AB peptide the loss of these helices correlates with the loss of the native tertiary contacts between them (particularly contacts between Leu8, Met12,

Leu25, Trp28 and Val29); the number of residue pairs involved in long range hydrophobic contacts falls by 60% between 500 and 550 ps in the 500 K trajectory. Similar behaviour is seen in the  $\alpha$  domain, the A and B helices unfolding in a concerted manner after approximately 600 ps at 700 K with the loss of key contacts particularly those between Met12, Leu25 and Trp28. Therefore both in the presence and absence of a significant number of other tertiary interactions, the contacts between the A and B helices seem to be of key importance in maintaining these structural units in the simulations.

A final similarity in the behaviour of the intact  $\alpha$  domain and the shorter peptides in the simulations is the persistence of helicity or turns in the regions corresponding to the native D and  $3_{10}$  helices. Interestingly in the  $\alpha$  domain this appears to be associated with interactions between the AB and CD chains (for the D helix contacts between Trp108, Trp111 and Trp28, Val29, and for the  $3_{10}$  helix contacts between Val120, Trp123, Ile124 and Phe3, Leu8, Leu25). However, in the CD peptide, which in the absence of the AB chain collapses into a compact structure, residues involved in residual turns in the D helix region form interactions with residues in the C helix region (between Trp108, Trp111 and Val92, Val99) while those in the  $3_{10}$  helix region form local interactions involving Val120, Trp123 and Ile124. Here, therefore, we observe native-like secondary structure stable in two different non-native conformations.

#### Comparisons with experimental data

Experimental studies of the folding behaviour and partially folded states formed by lysozymes from a variety of species suggest that conformations are populated which have a partially structured core in the  $\alpha$  domain in the absence of a persistent C helix (Radford & Dobson, 1995). For example, equine lysozyme forms a molten globule state at pH 2 in which there is significant hydrogen exchange protection for amides in the A, B and D helices but not the C helix (Morozova *et al.*, 1995; Morozova-Roche *et al.*, 1997). Similarly, in the kinetic refolding of human lysozyme hydrogen exchange protection develops very rapidly for amide protons in the A, B and C-terminal  $3_{10}$  helices (Hooke *et al.*, 1994). Hen lysozyme does not, however, adopt a highly populated molten globule state under any of the conditions in which it has been currently studied experimentally, and on refolding hydrogen exchange protection develops co-operatively for the amide in all  $\alpha$ -helices (Radford *et al.*, 1992a). However, the denatured states of the protein comprise an ensemble of conformers ranging from compact to extended conformations and analysis of the pattern of NOEs for hen lysozyme denatured in 8 M urea suggests that the more compact conformers in this ensemble may contain the native A, B, D and  $3_{10}$  helices but not the C helix (Schwalbe *et al.*, 1997).

In the light of these experimental results it is interesting to consider the characteristics of the states formed in the 700 K  $\alpha$  domain unfolding simulation prior to 600 ps when a core consisting of the native A, B, D and  $3_{10}$  helices persists in the absence of the C helix (Figure 2b). As an example, the properties of the states present from 400 to 450 ps at 700 K are summarised in Table 1. Helix C is essentially absent from the structures at this time. The other helices are frayed at their termini with 60% of the number of  $O(i)$ - $NH(i+4)$  hydrogen bonds identified in the native structure being present. There are also significant differences in the tertiary contacts compared with those in the native conformation. Only 30% of the residue pairs involved in long range hydrophobic contacts in the native conformation at the start of the simulation remain involved in such contacts. Overall the total number of residue pairs involved in long range hydrophobic contacts present is also reduced from 61 to 42. Considering these differences it is, therefore, interesting that the peptide conformation is compact (radius of gyration 1.18 nm compared with 1.22 nm in the native conformation) with a similar number of buried residues to the native conformation (26 residues have a percentage accessibility less than <20% compared to 24 residues in the native conformation). States such as this, therefore, show many of the characteristics expected for a structure within the ensemble of conformers defining the molten globule state where there is some persistent native-like secondary structure and the structures are collapsed but lack persistent tertiary interactions (Ptitsyn, 1995).

A range of peptide fragments from the lysozyme sequence have been probed experimentally, including those corresponding to the AB and CD peptides used in the simulations (Yang *et al.*, 1995, 1996). The most striking result from these experimental studies is that whereas the AB peptide is essentially unstructured, in the CD peptide there is evidence for extensive helicity in the regions corresponding in the native state to the D and  $3_{10}$  helices. Shorter peptide fragments have also been characterised experimentally (Yang *et al.*, 1996). A peptide whose sequence corresponds to the native  $3_{10}$  helix alone (residues 116 to 129) displays significant helicity in aqueous solution, but a peptide with the D helix sequence (residues 105 to 115) is unstructured. This indicates that non-native interactions stabilise the observed helicity in the D and  $3_{10}$  region of the CD peptide fragment. It is, therefore, interesting to compare these results with the simulation. As discussed above in the  $3_{10}$  region of the CD peptide the formation of very local interactions between the side-chains of Val120, Trp123 and Ile124 is observed in the simulation together with persistent helicity. This suggests that these local interactions could play a significant role in stabilising the helicity and would provide an explanation for the experimental observation of helicity persisting in this peptide, both in the CD peptide and in the shorter peptide corresponding

to residues 116 to 129. Similarly the turns present in the D helix region in the simulation are associated with non-native interactions, particularly with Val92 and Val99 in the C helix. If these interactions between residues in the C and D helix regions are important in stabilising the turns this again would provide an explanation of the experimental results where helicity is only observed for this D helix region in the full CD peptide.

### Characteristics of the denatured conformations populated in the high temperature simulations

A wide range of conformations is accessible to a denatured polypeptide chain. Although it is not possible to sample this ensemble adequately in the length of the simulations, the structures populated by the molecules in the high temperature simulations when the native conformations have been significantly disrupted (referred to here as denatured conformers) have been analysed to gain insight into the characteristics of the full denatured ensemble. For the  $\alpha$  and  $\beta$  domain the comparison concentrates on the 700 K simulations after 600 ps. Despite the different sequences and the different structures adopted by the two peptides at the start of the simulations many similarities are observed in the characteristics of the denatured conformers present at the end of the trajectories. Both the peptides adopt conformations that have a similar total solvent accessible surface area (mean accessibility per residue of 0.71 to 0.72 nm<sup>2</sup>) at the end of the simulations, with 26% of the residues still being significantly buried (accessibility less than 20%; Table 1). Interestingly for the  $\alpha$  domain there is a significant increase in the mean accessibility compared to that for the native conformation, most of the increase coming from the exposure of non-polar side-chains (30% increase in accessibility). In the  $\beta$  domain there is only a slight increase in the total accessibility, an increased exposure of polar side-chains being compensated for by a decreased exposure of non-polar side-chains. This difference, in part, reflects the greater proportion of residues with polar side-chains in the  $\beta$  domain sequence compared with the  $\alpha$  domain (41% and 51% of the residues in the  $\alpha$  and  $\beta$  domain sequences, respectively, are polar) and the key role that hydrophobic contacts play in stabilising the native  $\alpha$  domain conformation (particularly involving the aromatic residues Trp28, Trp108 and Trp111, which cluster in the hydrophobic core) while both hydrophobic and polar contacts together with hydrogen bonds stabilise the native  $\beta$  domain conformation.

Similar results are observed if the numbers of residue pairs involved in long range hydrophobic contacts in the two domains are compared. In the native conformations the number of residue pairs involved in these contacts (per 100 residues) in the  $\alpha$  domain (76) is approximately double that in the

$\beta$  domain (35). In both cases the number of residues involved in long range hydrophobic contacts reduces during the trajectory, but a more significant reduction is seen for the  $\alpha$  domain (40% reduction) than for the  $\beta$  domain (18% reduction; Table 1). The majority of long range contacts that remain in the denatured conformers involve aromatic residues in hydrophobic clusters, particularly around the disulphide bridges. Thus, for example, in the  $\alpha$  domain there are contacts between Phe3 and Leu8 at the N-terminus and Val120, Trp123 and Ile124 at the C terminus, these two regions being kept in close proximity by the 6–127 disulphide bridge. Similarly in the  $\beta$  domain the side-chains of Ile58, Trp63, Ile78 and Leu83 cluster around the 64–80 disulphide bridge.

Analysis of the hydrogen bonds present in the two domains shows that overall there is only a slight reduction in the number of intrapeptide hydrogen bonds present through the 700 K trajectories. This is particularly the case for the  $\beta$  domain where the reduction in the number of main-chain hydrogen bonds is compensated by an increase in the number of hydrogen bonds involving side-chains (44 hydrogen bonds in the native conformation and 43, on average, in the denatured conformers present at the end of the simulation). Despite significant differences between the main-chain hydrogen bonding pattern in the native conformations of the two peptides, the distributions of intrapeptide main-chain hydrogen bonds in the denatured conformers are similar. In particular in both the  $\alpha$  and  $\beta$  domains short range  $O(i)$ – $NH(i+2)$  hydrogen bonds dominate (20 hydrogen bonds per 100 residues). Of the other hydrogen bonds present, the  $\alpha$  domain has a greater population of  $O(i)$ – $NH(i+3)$ ,  $O(i)$ – $NH(i+4)$  and  $O(i)$ – $NH(i+5)$  hydrogen bonds (15 per 100 residues), reflecting the persistence of helicity in the C-terminal D and  $3_{10}$  helix regions. Interestingly the  $\beta$  domain has a slightly higher proportion of hydrogen bonds defining  $\beta$ -bridges or strands (8 per 100 residues) than the  $\alpha$  domain but these hydrogen bonds do not occur in the region corresponding to the native  $\beta$ -sheet but mostly between the residues adjacent to the 64–80 disulphide bridge. For the  $\alpha$  domain there is an increase in the number of hydrogen bonds from the peptide main-chain to solvent molecules on denaturation. The reverse is seen, however, for the  $\beta$  domain; this reflects the significant number of peptide–water hydrogen bonds present in the native conformation of this peptide particularly involving residues in the long loop region.

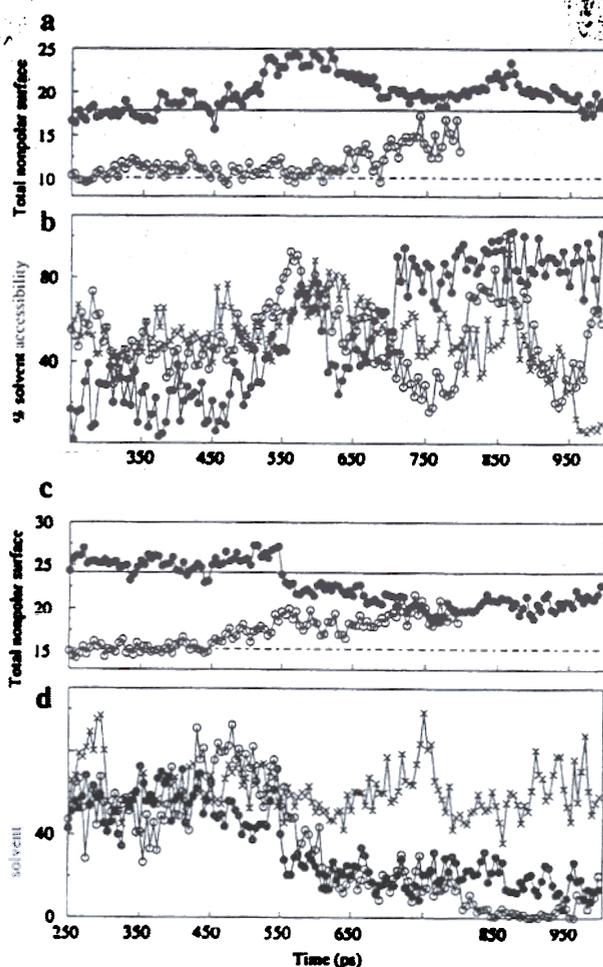
Even after the native conformation has been significantly disrupted the  $\alpha$  and  $\beta$  domains are restricted to relatively compact conformations by the disulphide bridges. It is, therefore, interesting to compare the characteristics of the  $\alpha$  and  $\beta$  domain denatured conformers with those present at the end of the AB and CD peptide trajectories where no disulphide bridges link distant parts of the sequence. Due to the considerable number of

hydrophobic contacts between these two chains in the intact lysozyme fold the native conformations present at the start of the simulations of these two peptides have a significantly higher solvent accessible surface area and a lower number of hydrophobic contacts per residue than either the  $\alpha$  or  $\beta$  domains. During the simulations neither of the peptides manages to obtain an accessible surface area or a number of hydrophobic contacts comparable to the chains in the  $\alpha$  domain adopting its native conformation but the peptides can be seen to be populating conformations that maximise the burial of hydrophobic side chains.

The AB peptide explores a range of compact states which contain different clusters of hydrophobic side chains but have a similar solvent accessible surface area ( $\sim 0.90$  nm<sup>2</sup> per residue). In the native conformation there are contacts between the hydrophobic residues in the helices (particularly Leu8, Met12, Leu25, Trp28, Val29). After 760 ps, however, a compact conformer with Tyr20 clustering together with residues at the N and C termini of the chain (Val2, Phe3, Phe34) is adopted while structures between 950 and 1000 ps contain contacts between the side chains of Tyr20, Tyr23 and Trp28 (Table 1 and Figure 3c). This variety of hydrophobic clusters can be seen to result in burial of different hydrophobic side-chains from solvent at different times during the trajectory (Figure 9). In between these compact states, extended conformations are populated (Figure 3b and Table 1) with a significantly higher accessible surface area ( $\sim 1.1$  nm<sup>2</sup> per residue).

The CD peptide, in contrast, collapses into states with radii of gyration ( $\sim 1.1$  nm) significantly less than that for the peptide in its native conformation (1.3 nm). Subsequent to this collapse more extended conformations are not populated. Looking at the characteristics of the denatured conformers present from 950 to 1000 ps at 500 K, there are considerably more long range hydrophobic contacts (contacts between 11 residue pairs) than in the native conformation at the start of the trajectory (contacts between seven residues pairs). The solvent accessible surface area is also reduced ( $0.82$  nm<sup>2</sup> per residue compared to  $0.94$  nm<sup>2</sup> in the native conformation). Throughout the simulation compact conformers are populated with a range of characteristics. However, the majority of them contain hydrophobic contacts between Trp108 or Trp111 and residues with hydrophobic side-chains at the N terminus of the sequence (Val92, Ile98, Val99). A second group of hydrophobic contacts is also seen in the vicinity of Val120, Trp123 and Ile124. Thus, both in the presence or absence of disulphide bridges linking the polypeptide chain, hydrophobic clusters are a common feature of the denatured conformers populated in the simulations.

Comparisons of the hydrogen bond pattern in the denatured conformers present in the AB and CD peptide simulations (Table 1) shows that the



**Figure 9.** a, Variation in the total accessibility (in nm<sup>2</sup>) of the non-polar side-chains of residues 1 to 35. Filled circles show the accessibility of these residues in the AB peptide simulation at 500 K and open circles the accessibility of these residues in the  $\alpha$  domain simulation at 700 K. The accessibility of non-polar side-chains for residues 1 to 35 in the  $\alpha$  domain and AB peptides adopting the native lysozyme-fold is indicated by a broken and continuous line, respectively. b, Variations in the percentage accessibility of Tyr20 (open circles), Tyr23 (crosses) and Val29 (filled circles) during the AB peptide simulation at 500 K. c, Variation in the total accessibility (in nm<sup>2</sup>) of the non-polar side-chains of residues 85 to 129. Filled circles show the accessibility for these residues in the CD peptide simulation at 500 K and open circles the accessibility for these residues in the  $\alpha$  domain simulation at 700 K. The accessibility of non-polar side-chains for residues 85 to 129 in the  $\alpha$  domain and CD peptides adopting the native lysozyme fold is indicated by a broken and continuous line respectively. d, Variations in the percentage accessibility of Val92 (open circles), Trp108 (filled circles) and Val120 (crosses) during the CD peptide simulation at 500 K. In all cases the values shown are for instantaneous structures taken at 5 ps intervals from the simulations.

AB peptide has a higher population of O(i)-NH(i+2) hydrogen bonds (average of 20 per 100 residues from 950 to 1000 ps), fewer hydrogen

bonds between main-chain groups more distant in the sequence and more hydrogen bonds between the main-chain and solvent molecules than the CD peptide. The CD conformers have a significant number of O(i)-NH(i+3,4,5) hydrogen bonds (average of 16 per 100 residues), reflecting the helicity in the C-terminal part of this chain as also observed in this region in the denatured  $\alpha$  domain conformers.

#### Comparison with experimental data

Studies of lysozyme denaturation by urea show differences in the extent of unfolding at a given urea concentration as monitored by the radius of gyration from X-ray solution scattering data and the far UV circular dichroism signal at 222 nm (Chen *et al.*, 1996). It has been suggested that this difference reflects unfolding of the  $\beta$  domain prior to full unfolding of the  $\alpha$  domain, the unfolding of the  $\beta$  domain giving a much greater increase in the overall mean radius of gyration (~0.5 nm) than unfolding of the  $\alpha$  domain (~0.2 nm). Such an interpretation is supported by the results of the MD simulations. Much of the  $\beta$ -sheet in the  $\beta$  domain is lost very rapidly at 700 K and this peptide then explores conformations with radii of gyration up to 0.17 nm larger than that of the  $\beta$  domain in its native conformation (Figure 4d). The  $\alpha$  domain, in contrast, exhibits far more persistent secondary structure even in the absence of long range contacts, and even by the end of the simulation the most extended denatured conformers sampled only have a radius of gyration 0.11 nm larger than that of the  $\alpha$  domain adopting its native conformation.

The limited increase in the radius of gyration of the  $\alpha$  domain after disruption of the native conformation reflects, at least in part, the constraints provided by the disulphide bridges. For the AB peptide, however, in the absence of any disulphide bridges holding the chain into a relatively compact conformation, similar behaviour is observed to that of a freely jointed or random flight chain used to model the global characteristics of a random coil (Kuhn, 1934; Smith *et al.*, 1996). A range of conformations is explored after the native conformation is disrupted, including both extended conformers with radii of gyration up to 40% larger than that for the native conformation and compact conformers with different non-native clusterings of hydrophobic groups and radii of gyration equivalent to that in the native conformation (Figure 5). Some of these conformers contain fluctuating turns similar to those identified experimentally in peptides containing nascent helices (Dyson & Wright, 1991) and predicted theoretically by the statistical model for a random coil (Smith *et al.*, 1996). This behaviour contrasts significantly with that of the CD peptide, which rapidly collapses into compact states in the trajectory at 500 K, in agreement with experimental NMR data which suggests that the AB peptide, unlike the CD peptide discussed pre-

viously, is essentially unstructured (Yang *et al.*, 1996).

Despite its predominantly unstructured nature NOEs are observed experimentally for the AB peptide between some of the aromatic and hydrophobic side-chains (Yang *et al.*, 1996). In particular there are NOEs between Tyr20, Tyr23, Trp28 and the side-chain methyl groups of Val2, Leu8, Leu17 or Leu25; in these cases, however, residue-specific assignments could not be made because of resonance overlap. Analysis of the distances between these residues in the denatured peptide from the trajectory (600 to 1000 ps at 500 K) shows that there are short mean distances,  $(\langle r^{-3} \rangle)^{-1/3}$ , between the methyl groups of Leu25 and the ring protons of Trp28 (e.g. mean average distance between C<sup>δ1</sup> of Leu25 and H<sup>ε1</sup> of Trp28 is 0.44 nm) which would result in readily observable NOEs between these residues. Distances in the range 0.7 to 0.8 nm are observed between the methyl groups of Leu25 and the aromatic ring protons of Tyr20 and Tyr23 and between the methyl groups of Val2 and the ring protons of Tyr20. These relatively short distances reflect the adoption of hydrophobic clusters in the more compact states of the conformational ensemble during the simulation and show that the hydrophobic clusters would be consistent with the observed experimental data.

The observation of hydrophobic clusters in the denatured conformers populated in all the four simulation trajectories is of interest as such clusters are the major cause of discrepancies observed experimentally for denatured polypeptide chains from predictions for a random coil (Dobson, 1992; Wüthrich, 1994; Shortle, 1996b). For example in the case of reduced lysozyme denatured in urea, clusters are proposed in the vicinity of Trp62 and 63, Trp108 and 111 and Trp123 on the basis of deviations of chemical shifts, coupling constants, NOEs and relaxation rates from random coil predictions (Schwalbe *et al.*, 1997). Of particular interest for comparison with experimental data is the cluster observed in the  $\beta$  domain peptide simulation in the region of the 64–80 disulphide bridge involving Ile58, Trp63, Cys64, Ile78, Cys80 and Leu83. Experimental analysis of the hydrogen exchange behaviour of the backbone amide protons in thermally denatured lysozyme shows that five residues have protection factors greater than 6 (Radford *et al.*, 1992b). Four of these residues are those involved in this hydrophobic cluster (Ile58, Trp63, Cys64 and Ile78). The amide protons of Trp63, Cys64 and Ile78 are also observed to be protected rapidly from exchange during the refolding of lysozyme, an experimental observation that initially was unexpected as these residues are not within a region of secondary structure in the native fold (Radford *et al.*, 1992a). However, the involvement of the residues in a persistent hydrophobic cluster in the denatured protein could explain these experimental observations.

## Discussion

The interpretation of experimental data for non-native protein conformations in structural terms must take into account the complex structural heterogeneity of the majority of these states which results in averaging of the experimental parameters across the conformational ensemble (Dobson *et al.*, 1991; Smith *et al.*, 1996). To aid the analysis of experimental results one possible strategy is to compare the experimental data with predictions from theoretical models for non-native states (Fiebig *et al.*, 1996). Here we have examined whether MD simulations can be successfully employed to generate such models by considering the properties of non-native and denatured conformers formed in unfolding simulations of four peptide fragments from the hen lysozyme sequence.

The conformers populated during these trajectories display many of the features that would be expected from experimental studies of such states and from knowledge of the factors governing protein structures. For example, simulations performed with different sequences starting from significantly different conformations form denatured states with closely similar overall properties. Thus, at the end of the trajectories at 700 K the  $\alpha$  and  $\beta$  domain peptides populate conformers which have, for example, closely similar values of the solvent accessible surface area per residue and hydrogen bonding patterns. The hydrogen bonding patterns reflect the short range nature of the contacts that are dominant in determining the conformational distributions in these non-native states. A significant increase is observed on denaturation in the numbers of O(*i*)–NH(*i* + 2) hydrogen bonds and hydrogen bonds involving side-chain groups as longer range main-chain hydrogen bonds defining helices or  $\beta$ -sheets are lost.

Due to the broad nature of the conformational ensemble adopted by non-native states of proteins, it is difficult for structural models to represent adequately the full conformational ensemble. This is the case here where the sampling is limited by the length of the simulations. However, for the denatured states formed at least, the close similarity of many of the overall properties observed in the simulations of the different peptides suggests that these features may be characteristic of the full conformational ensemble. Further MD studies are, however, needed to examine these findings in more detail. Of particular interest would be the comparison of these simulations performed in aqueous solution with those performed in other solvent systems; the polarities and hydrogen bonding properties of the solvent will clearly have a significant effect on the features of the non-native states as demonstrated by experimental studies of proteins denatured in solvents such as methanol and trifluoroethanol (Harding *et al.*, 1991; Buck *et al.*, 1993; Fan *et al.*, 1993).

Although local contacts dominate in defining the characteristics of denatured states this does not

preclude a significant population of the conformers in the ensemble displaying persistent structural features, particularly those that appear to be associated with hydrophobic clusters. This is demonstrated clearly in the simulation trajectories reported here where hydrophobic clusters are observed in the denatured conformers adopted by all four peptides involving groups that are close in the sequence or brought into close proximity by the disulphide bridges present in the  $\alpha$  and  $\beta$  domain peptides. These clusters enable burial of hydrophobic surfaces thus compensating, to some extent, for the loss of long range tertiary contacts on the disruption of the native conformations. Some of the clusters are populated only transiently while others are more persistent, those that may be involved in stabilising the residual native-like secondary structure in the C-terminal part of the  $\alpha$  domain sequence by non-native interactions being of particular interest.

The observation of persistent structural features together with local clusters of hydrophobic side-chains is of interest with regard to our understanding of protein folding. Such persistent contacts reflect deviations from simple random coil behaviour and are thought to be of potential importance in defining the characteristics of the ensemble of conformers present in the early stages of folding (Dyson & Wright, 1993; Smith *et al.*, 1996). Through a reduction of the conformational space that is energetically accessible they determine the initial features of the folding landscape and thus play a significant role in the overall folding process (Dobson *et al.*, 1998). It is particularly intriguing here that both native and non-native interactions appear to be involved in generating such persistent structure. Non-native local contacts may be important in helping restrict conformational space explored by individual molecules, at later stages in the folding process these contacts being replaced by longer range native contacts. By providing a framework for interpreting experimental data and models for both the local and global properties of non-native states, an approach such as that explored here can therefore provide considerable insight into the ensembles of conformers defining non-native states.

## Methods

Simulations of four peptide fragments from the hen lysozyme sequence ( $\alpha$  domain,  $\beta$  domain, AB peptide and CD peptide), performed in the presence of explicit solvent (water) with periodic boundary conditions, are presented. The simulations and analysis were carried out using the GROMOS package of programs (van Gunsteren & Berendsen, 1987; van Gunsteren *et al.*, 1996). The GROMOS force field for solvent simulations, parameter set 37C4, was used with a modification to the interaction between water oxygen and the carbon atoms of the protein and the explicit inclusion

of aromatic hydrogen atoms (partial charge 0.14e) as described by Smith *et al.* (1995). The starting solute (peptide) co-ordinates in each case were taken from the crystal structure of triclinic hen lysozyme (Ramanadham *et al.*, 1987), entry 2LZT of the protein data base, with modifications to the chain termini and cysteine residues as summarised in Table 3; the native lysozyme sequence numbering was used for all four peptides. No crystallographic water molecules or counter ions were included in the simulations. Truncated octahedron periodic boundary conditions were used in each case. The dimensions of the periodic box were chosen so that no non-hydrogen solute atom in the starting structure lies within 1.1 nm of the square box wall in the case of the  $\alpha$  and  $\beta$  domain peptides and within 1.4 nm and 1.2 nm of the square box wall for the AB and CD peptides, respectively. The number of solute atoms and simple point charge (SPC) water molecules (Berendsen *et al.*, 1981) used in each of the simulations is summarised in Table 3.

For each peptide a simulation at 300 K was performed at constant pressure (1atm) for 800 ps. The temperature and pressure were maintained by weak coupling to an external bath (Berendsen *et al.*, 1984) (temperature coupling relaxation time 0.1 ps; pressure coupling relaxation time 0.5 ps). To allow initial relaxation of the water around the peptides, the co-ordinates of all the peptide atoms were positionally restrained for the first 10 ps of the simulations at 300 K and the positions of the C $\alpha$  atoms alone were restrained for a further 10 ps. Branching from the 300 K simulation trajectories after 220 ps (with temperature increases of 50 K per 10 ps until the required temperature was reached) constant volume simulations were performed at 500 K for all peptides and at 700 K for the  $\alpha$  and  $\beta$  domain peptides. The simulations were run until 800 ps (1000 ps for the AB and CD peptides). The temperature was again maintained by weak coupling to an external bath. Bond lengths were constrained to equilibrium values during the simulations using the SHAKE procedure (Ryckaert *et al.*, 1977). Non-bonded interactions were treated using a twin range method (van Gunsteren & Berendsen, 1990). Within the short-range cutoff of 0.8 nm all interactions were determined at every step. Longer range electrostatic interactions within a cutoff range of 1.4 nm were updated at the same time as the pair list was generated (every 10fs). Time steps of 2 fs at 300 K and 1 fs at 500 K and 700 K were used for integrating the equations of motion. Analysis was performed using trajectory co-ordinates and energies written to disk every 0.05 ps at 300 K and every 0.025 ps at 500 K and 700 K.

The solvent accessible surface areas of individual structures taken from the simulation trajectories were calculated using the program NACCESS (Hubbard & Thornton, 1993). The total accessible area of polar (oxygen and nitrogen) and non-polar (carbon and sulphur) side-chain atoms in the struc-

**Table 3. Characteristics of the peptide fragments used in the MD simulations**

Peptide	$\alpha$ Domain	$\beta$ Domain	AB peptide	CD peptide
Corresponding residues in native lysozyme *	1-35; 85-129	36-84	1-35	
N termini	Chain 1 NH <sub>3</sub> <sup>+</sup> Chain 2 CH <sub>3</sub> -C(O)-N(H)-	CH <sub>3</sub> -C(O)-N(H)-	NH <sub>3</sub> <sup>+</sup> -	
C termini	Chain 1 -CO <sub>2</sub> -NH <sub>2</sub> Chain 2 -CO <sub>2</sub> <sup>-</sup>	-CO <sub>2</sub> -NH <sub>2</sub>	-CO <sub>2</sub> -NH <sub>2</sub>	
Cysteine residues	Cys94 changed to Ala	Cys76 changed to Ala	Cys 6 -CH <sub>2</sub> -N(H)-C(O)-CH <sub>3</sub> Cys30 not modified	Cys94 changed to Ala Cys115 -CH <sub>2</sub> -N(H)-C(O)-CH <sub>3</sub> Cys127 not modified
Disulphide bonds	6-127; 30-115	64-80	-	-
Number of protein atoms	829	504	380	461
Number of water molecules	3532	2737	3428	4081
Initial box dimensions (nm)	$x = y = z = 6.26$	$x = y = z = 5.62$	$x = y = z = 5.99$	$x = y = z = 6.37$

\* Hen lysozyme amino acid sequence: 1 to 35, KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFE; 36 to 84, SNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALL; 85 to 129, SSDITASVNC AKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGRL.

tures and the percentage accessibility of individual residues (accessibility relative to that in Ala-X-Ala peptides) were analysed. Regions of secondary structure and hydrogen bonds were identified using the program DSSP (Kabsch & Sander, 1983). The hydrophobic contacts present in the structures taken from the trajectories were analysed using the program NAOMI (Brocklehurst & Perham, 1993) which identifies the residue pairs between the side-chains of which a hydrophobic contact is made (contacts involving the carbon atoms and their covalently bonded protons in the side-chains of the amino acid residues A, F, I, L, V, W, Y, M, H, K, P and T). The variation in the total number of residue pairs whose side-chains are involved in long range hydrophobic contacts (residues with an  $i$  to  $i + 5$  separation or greater) and the number of such residue pairs that are native-like (contacts that occur between the same pair of residues as are observed in the native lysozyme structure independent of whether the same atoms are involved) have been considered.

## Acknowledgements

This is a contribution from the Oxford Centre for Molecular Sciences which is supported by the UK Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council and the Medical Research Council. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Research Institute. L.J.S. is a Royal Society Research Fellow. W.F.v.G. gratefully acknowledges financial support given by the Underwood Fund.

## References

- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular Forces* (Pullman, B., ed.), pp. 331–342, Reidel, Dordrecht.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
- Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). Structure of hen-egg white lysozyme: a three dimensional Fourier synthesis at 2.0 Å resolution. *Nature*, **206**, 757–761.
- Bolin, K. A., Pitkeathly, M., Miranker, A., Smith, L. J. & Dobson, C. M. (1996). Insight into a random coil conformation and an isolated helix: structural and dynamical characterisation of the C-helix peptide from hen lysozyme. *J. Mol. Biol.* **262**, 443–453.
- Brocklehurst, S. M. & Perham, R. N. (1993). Prediction of the three dimensional structure of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine-cleavage system: a new automated method for the prediction of protein tertiary structure. *Protein Sci.* **2**, 626–639.
- Buck, M., Radford, S. E. & Dobson, C. M. (1993). A partially folded state of hen egg-white lysozyme in trifluoroethanol. Structural characterization and implications for protein folding. *Biochemistry*, **32**, 669–678.
- Chen, L. L., Hodgson, K. O. & Doniach, S. (1996). A lysozyme folding intermediate revealed by solution X-ray scattering. *J. Mol. Biol.* **261**, 658–671.
- Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.
- Dobson, C. M. (1992). Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* **2**, 6–12.
- Dobson, C. M., Hanley, C., Radford, S. E., Baum, J. & Evans, P. A. (1991). Characterization of unfolded and partially folded states of proteins by NMR spectroscopy. In *Conformations and Forces in Protein Folding*. (Nall, B. T. & Dill, K. A., eds), pp. 175–181, AAAS Publications, Washington, DC.
- Dobson, C. M., Evans, P. A. & Radford, S. E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37.
- Dobson, C. M., Sali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37**, 868–893.
- Dyson, H. J. & Wright, P. E. (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Chem.* **20**, 519–538.
- Dyson, H. J. & Wright, P. E. (1993). Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* **3**, 60–65.
- Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A. & Wright, P. E. (1992a). Folding of peptide fragments comprising the complete sequence of proteins: models for initiation of protein folding I. Myohe-merythrin. *J. Mol. Biol.* **226**, 795–817.
- Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A. & Wright, P. E. (1992b). Folding of peptide fragments comprising the complete sequence of proteins: models for initiation of protein folding II. Plastocyanin. *J. Mol. Biol.* **226**, 819–835.
- Fan, P., Bracken, C. & Baum, J. (1993). Structural characterization of monellin in the alcohol-denaturation state by NMR: evidence for the  $\beta$ -sheet to  $\alpha$ -helix conversion. *Biochemistry*, **32**, 1573–1584.
- Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J. & Dobson, C. M. (1996). Towards a description of the conformation of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem.* **100**, 2661–2666.
- Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Müller-Frohne, M., Zirwer, D., Czarneci, J. & Damaschun, G. (1995). Prothymosin  $\alpha$ : a biologically active protein with a random coil conformation. *Biochemistry*, **34**, 13211–13218.
- Harding, M. M., Williams, D. H. & Woolfson, D. N. (1991). Characterization of a partially denatured state of a protein by 2D NMR: reduction of the hydrophobic interactions in ubiquitin. *Biochemistry*, **30**, 3120–3128.
- Hirst, J. D. & Brooks, C. L. (1995). Molecular dynamics simulations of isolated helices of myoglobin. *Biochemistry*, **34**, 7614–7621.
- Hooke, S. D., Radford, S. E. & Dobson, C. M. (1994). The refolding of human lysozyme: a comparison with the structurally homologous hen lysozyme. *Biochemistry*, **33**, 5867–5876.

- Hubbard, S. J. & Thornton, J. M. (1993). NACCESS computer program, Departments of Biochemistry and Molecular Biology, University College London.
- Hünenberger, P. H., Mark, A. E. & van Gunsteren, W. F. (1995). Computational approaches to study protein unfolding: hen egg-white lysozyme as a case study. *Proteins: Struct. Funct. Genet.* **21**, 196–213.
- Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., Prat Gay, G. & Fersht, A. R. (1995). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* **254**, 289–304.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Kemmink, J. & Creighton, T. E. (1993). Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. *J. Mol. Biol.* **234**, 861–878.
- Kraulis, P. J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
- Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. (1996). Structural studies of p21<sup>(waf1/cip1/ink1)</sup> in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl Acad. Sci. USA*, **93**, 11504–11509.
- Kuhn, W. (1934). Über die Gestalt Fadenförmiger Moleküle in Lösungen. *Kolloid-Z.* **68**, 2–15.
- Mark, A. E. & van Gunsteren, W. F. (1992). Simulation of the thermal denaturation of hen egg-white lysozyme - trapping the molten globule state. *Biochemistry*, **31**, 7745–7748.
- Morozova, L. A., Haynie, D. T., Arico-Muendel, C., van Dael, H. & Dobson, C. M. (1995). Structural basis of the stability of a lysozyme molten globule. *Nature Struct. Biol.* **2**, 871–875.
- Morozova-Roche, L. A., Arico-Muendel, C. C., Haynie, D. T., Emelyanenko, V. I., Van Dael, H. & Dobson, C. M. (1997). Structural characterisation and comparison of the native and A-states of equine lysozyme. *J. Mol. Biol.* **268**, 903–921.
- Muñoz, V., Serrano, L., Jimenez, M. A. & Rico, M. (1995). Structural analysis of peptides encompassing all  $\alpha$ -helices of 3  $\alpha/\beta$ -parallel proteins—che-Y, flavodoxin and p21-ras. Implications for  $\alpha$ -helix stability and the folding of  $\alpha/\beta$ -parallel proteins. *J. Mol. Biol.* **247**, 648–669.
- Oas, T. G. & Kim, P. S. (1988). A peptide model of a protein folding intermediate. *Nature*, **336**, 42–48.
- Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G. & Smith, L. J. (1997). NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol.* **274**, 152–159.
- Pittsytyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83–229.
- Radford, S. E. & Dobson, C. M. (1995). Insights into protein folding using physical techniques: studies of lysozyme and  $\alpha$ -lactalbumin. *Phil. Trans. Roy. Soc. ser. B*, **348**, 17–25.
- Radford, S. E., Dobson, C. M. & Evans, P. A. (1992a). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, **358**, 302–307.
- Radford, S. E., Buck, M., Topping, K. D., Evans, P. A. & Dobson, C. M. (1992b). Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins: Struct. Funct. Genet.* **14**, 237–248.
- Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1987). SRLSQ refinement of triclinic lysozyme. *Acta Crystallog. sect. A*, **43**, 13 (Suppl.).
- Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341.
- Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S., Smith, L. J. & Dobson, C. M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, **36**, 8977–8891.
- Shortle, D. (1996a). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27–34.
- Shortle, D. (1996b). Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* **6**, 24–30.
- Smith, L. J., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1993). Structure of hen lysozyme in solution. *J. Mol. Biol.* **229**, 930–944.
- Smith, L. J., Mark, A. E., Dobson, C. M. & van Gunsteren, W. F. (1995). Comparison of MD simulations and NMR experiments for hen lysozyme. Analysis and local fluctuations, cooperative motions and global changes. *Biochemistry*, **34**, 10918–10931.
- Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. (1996). The concept of a random coil. Residual structure in peptides and denatured proteins. *Folding Design*, **1**, R95–R106.
- Soman, K. V., Karimi, A. & Case, D. A. (1993). Molecular dynamics analysis of a ribonuclease C-peptide analogue. *Biopolymers*, **33**, 1567–1580.
- Thomas, P. J., Qu, B. H. & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20**, 456–459.
- Tirado-Rives, J. & Jorgensen, W. L. (1991). Molecular dynamics simulations of the unfolding of an  $\alpha$ -helical analogue of ribonuclease A S-peptide in water. *Biochemistry*, **30**, 3864–3871.
- Ueda, T., Nakashima, A., Hashimoto, Y., Miki, T., Yamada, H. & Imoto, T. (1994). Formation of  $\alpha$ -helix 88–98 is essential in the establishment of higher-order structure from reduced lysozyme. *J. Mol. Biol.* **235**, 1312–1317.
- van Gunsteren, W. F. & Berendsen, H. J. C. (1987). *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh 14, 9747 AG Groningen, The Netherlands.
- van Gunsteren, W. F. & Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics - methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Eng.* **29**, 992–1023.
- van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P. & Tironi, I. G. (1996). *Biomolecular Simulations: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland.
- Weinreb, P. H., Zhen, W. G., Poon, A. W., Conway, K. A. & Lansbury, P. T. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*, **35**, 13709–13715.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Wüthrich, K. (1994). NMR assignment as a basis for structural characterisation of denatured states of

globular proteins. *Curr. Opin. Struct. Biol.* **4**, 93-99.

Yang, J. J., Buck, M., Pitkeathly, M., Kotik, M., Haynie, D. T., Dobson, C. M. & Radford, S. E. (1995). Conformational properties of four peptides spanning the sequence of hen lysozyme. *J. Mol. Biol.* **252**, 483-491.

Yang, J. J., van den Berg, B., Pitkeathly, M., Smith, L. J., Bolin, K. A., Keiderling, T. A., Redfield, C., Dobson, C. M. & Radford, S. E. (1996). Native-like secondary structure in a peptide from the alpha-domain of hen lysozyme. *Folding Design*, **1**, 473-484.

Edited by B. Honig

(Received 2 December 1997; received in revised form 5 March 1998; accepted 5 March 1998)