J. Mol. Biol. (1993) 232, 600–619

Protein Unfolding Pathways Explored Through Molecular Dynamics Simulations

Valerie Daggett† and Michael Levitt

Beckman Laboratories for Structural Biology
Department of Cell Biology
Stanford University School of Medicine
Stanford, CA 94305–5400, U.S.A.

(Received 13 August 1992; accepted 8 March 1993)

Herein we describe the results of molecular dynamics simulations of the bovine pancreatic trypsin inhibitor (BPTI) in solution at a variety of temperatures both with and without disulfide bonds. The reduced form of the protein unfolded at high temperature to an ensemble of conformations with all the properties of the molten globule state. In this account we outline the structural details of the actual unfolding process between the native and molten globule states. The first steps of unfolding involved expansion of the protein, which disrupted packing interactions. The solvent-accessible surface area also quickly increased. The unfolding was localized mostly to the turn and loop regions of the molecule, while leaving the secondary structure intact. Then, there was more gradual unfolding of the secondary structure and non-native turns became prevalent. This same trajectory was continued and more drastic unfolding occurred that resulted in a relatively compact state devoid of stable secondary structure.

Keywords: protein unfolding; molecular dynamics; molten globule state

1. Introduction

Protein folding remains one of the most important unsolved problems in molecular biology. Although much is known of the structural details of the native, folded conformation of proteins, very little is known about the actual folding process, or how the linear polypeptide chain acquires its biologically active three-dimensional conformation. Characterization of the unfolding process is also equally important, both from the perspective of fully understanding a fundamental biochemical phenomenon, as argued persuasively by Fersht et al. (1990), and because of the probable link to protein degradation *in vivo* (Daggett, 1987). In addition, the transition state for folding appears to occur late in folding and is thought to be a distorted form of the native state (Goldenberg & Creighton, 1985; Matouschek *et al.*, 1989; Sancho *et al.*, 1991). Elucidation of the early steps in unfolding can provide much needed insight into this crucial step in folding.

The rapidity of folding in *vivo* and the possibility that a disordered polypeptide chain can assume a large number of conformations with almost equal energies make it likely that folding occurs *via* a limited number of kinetic pathways (Kim & Baldwin, 1982, 1990; Goldberg, 1985; Baldwin & Eisenberg, 1987). Intermediate states can direct folding and accelerate the transition to the native state by restricting the conformational space available for the unfolded state. Determination of the structural characteristics of intermediate states during folding are crucial for understanding the mechanism of these processes. Unfortunately, the cooperative nature of folding/unfolding results in only minute amounts of partially folded intermediates at equilibrium. Even using kinetic means, intermediates are populated only transiently and must be trapped in some way. Fortunately, such approaches have provided some structural information (Udgaonkar & Baldwin, 1988; Roder et *al.*, 1988; States *et al.*, **1987;** Bycroft *et* al., 1990; Matthews *et al.*, 1983).

Recently, the so-called molten globule state has received much attention; molten globules are compact, mobile structures with large amounts of secondary structure but diminished tertiary contacts relative to the native state. This state is interesting because it may be a common early intermediate during folding (Ptitsyn *et* al., 1990), and it

† Current address: Department of Medicinal Chemistry, BG-20, University of Washington, Seattle, WA 98195, U.S.A.

0022 2836/93/140600 20 $08.00/0

is an equilibrium intermediate under a variety of unfolding conditions (e.g. high temperature. extremes of pH.high concentrations of organic solvents) or upon the removal of a stabilizing ion (all reviewed by Ptitsyn. 1987 and Kuwajima, 1989). Recently, a molten globule was detected under folding conditions. In this case, it appears that the molten globule state is under kinetic control and is not an equilibrium intermediate (Baker et *al.*, 1992). However, the molecular details of even these relatively stable folding/unfolding intermediates remain difficult to characterize experimentally (Amir & Haas, 1988; Baum *et al.*, 1989; Dobson, 1992; Harding *et al.*, 1991; Hughson *et* al., 1990). The difficulties are principally due to the increased motion and lack of fixed structure throughout the molecules. Characterization of both the kinetic and equilibrium intermediates is also limited because the currently employed techniques can detect the presence of native-like structure but are unable to provide information about the formation of non-native conformations. Therefore, this is an area where experiment actually needs theory, as it is unlikely that experimental studies will ever yield structural information about protein folding in detail comparable with that available for native proteins.

For these reasons, we have performed molecular dynamics simulations of temperature-induced protein unfolding. The approach of studying unfolding in lieu of folding has an important 'computational advantage; namely, the -simulations proceed from a well-defined starting structure. In principle, this approach can delineate the details of intermediate structures and dynamic transitions that occur during the unfolding process. In addition, both native and non-native interactions can be observed during the simulation.

The studies presented here focus on bovine pancreatic trypsin inhibitor (BPTI†; Fig. 1), because it is a small, well-characterized protein. The technique of molecular dynamics has already been shown to be useful for realistic simulations of native states of proteins (Daggett & Levitt, 1991, 1993 and references therein), including BPTT (Levitt & Sharon, 1988; Levitt, 1989; Daggett & Levitt, **1992a).** Molecular dynamics (MD) simulations have also proved to be instructive in delineating the structures adopted by conformationally mobile peptides (Daggett *et* al., 1991*a*; DiCapua *et* al., **1990;**

† Abbreviations used: BPTI. bovine pancreatic trypsin inhibitor: MD, molecular dynamics; N, native state; MG. molten globule state: U. unfolded state; r.m.s. root-mean-square; r.m.s.d., r.m.s. displacement; N298. MD of oxidized, native BPTI at 298 K; N423, oxidized. native BPTI at 423 K: R423. reduced BPTI at 423 K; R498. reduced BPTI at 498 K; 1423 refers to two separate simulations at 423 K, that of the isolated β-sheet and the isolated C-terminal a-helix at 423 K; 1498, these same isolated fragments at 498 K. The secondary structure was broken down as follows al, residues 3 to 6: L1, 7 to 15; β1, 16 to 25; tl, 26 and 27; β2, 28 to 36; L2. 37 to 46; and α2, 47 to 56.

Figure 1. The main-chain trace of the BPTI crystal structure. The Cys residues are colored yellow with the native disulfide bonds shown (Cys30-Cys5 1, Cys5-Cys5s and Cysl4-Cys38). The different secondary structure boundaries are distinguished by color: al, residues 3 to 6 magenta; L1, 7 to 15 blue; β1, **16** to *25* cyan; t1, *26* to 2'7 green; β2, 28 to 36 cyan; L2, 37 to 46 blue; a2, 47 to 56 magenta.

Tirado-Rives & Jorgensen, 1991), the transitions that thev undergo (Daggett *et al.*, 1991*b*), and the denaturation process of peptides (Soman *et al.*, **1991;** Tirado-Rives & Jorgensen, **1991;** Daggett & Levitt, **19923).** Furthermore, various other theoretical approaches have now been used to study protein folding (Skolnick & Kolinski, 1989; Cove11 & Jernigan, 1990; Moult & Unger, 199 1; Hinds & Levitt, 1992). Although these strategies necessarily employ many simplifying approximations, they do provide insight into the problem. In any case, these past studies suggest that it should be possible to simulate the detailed conformational transitions experienced by a protein during the process of unfolding using molecular dynamics.

The major hindrance to using molecular dynamics for studying protein unfolding is the time-scale accessible during realistic solution simulations (presently picosecond to nanosecond) compared to the time-scale of folding/unfolding. While it is not known how fast proteins unfold, it is usually thought to be in the millisecond range: or faster, under denaturing conditions. This time-scale is not accessible by MD simulations; the simulations presented here require approximately one to two central processor unit hours per picosecond. Confronted with this problem, we chose to use high temperature as a means of accelerating the process.

Raising the temperature has a small effect on the velocities (atoms at 498 K move approx. 30% faster than at 298 K) but activated processes involving traversing energy barriers will be greatly accelerated. This acceleration depends on the enthalpy of activation and can give increases in speed of 10^3 ($\Delta H^{\ddagger}$ = 10 kcal/mol) to 10^9 ($\Delta H^{\ddagger}$ = 32 kcal/mol) at 498 K.

Various MD simulations lasting up to 550 ps were performed: native BPTI (with the 3 disulfide bonds intact, Fig. 1) at 298 K, N298, and at 423 K, N423; fully reduced BPTI (all disulfide bonds broken) at 423 and 498 K, R423 and R498, respectively; and isolated fragments of the protein (the central P-sheet and the C-terminal a-helix were simulated separately, Fig. 1) at 423 and 498 K, 1423 and 1498, respectively (Table 1). All simulations were carried out in a bath of water molecules with mobile counterions.

The simulations begin from the native state, and at the highest temperature, reduced BPTI (R498) unfolds in two phases. The first transition yields an ensemble of conformations with all of the properties of the molten globule state (e.g. N→MG), which is stable for approximately 250 ps (Daggett & Levitt, 1992a). To use Matthew's (1991) terminology this degree of unfolding corresponds to middle and late folding events. The second phase of unfolding results in more complete disruption of the structure. e.g. MG→U. Our focus in this paper is to characterize the early stages of the unfolding pathway, N→MG. A description of the resulting partially unfolded state (MG) has been presented (Daggett & Levitt, 1992*a*). Since both end points of these simulations appear to be experimentally reasonable, it is appropriate to follow the pathway between them by further analysis of these same trajectories. We have tried to avoid duplication of our earlier account but some is necessary to clarify points dealing with the unfolding process. The remainder of the folding pathway, from MG → U and structural characterization of the unfolded state will be presented elsewhere.

2. Methods

The starting point for the calculations was the X-ray crystal structure of BPTI (Deisenhofer & Steigemann, 1975). Details covering the simulations of BPTI and the BPTI fragments at different temperatures have been described (Daggett & Levitt, 1992*a*). All atoms were explicitly present) during the simulations. The protein and fragments were immersed in a box of water molecules and counterions were present to yield an electrically neutral system. The parameters, water model and details of the potential function are given elsewhere (Daggett *et* al., 1993).

3. Results and Discussion

The results of the unfolding simulations (N→MG) are presented according to the structural level considered, moving from the global to the semi-local and local level. We chose to follow the time-dependencies of those properties that show characteristic differences between the starting (X-ray) and final (molten globule) states. Description of the unfolding of the molten globule state and characterization of the unfolded state will be presented elsewhere (e.g. the 300 to 550 ps interval of the R498 simulation).

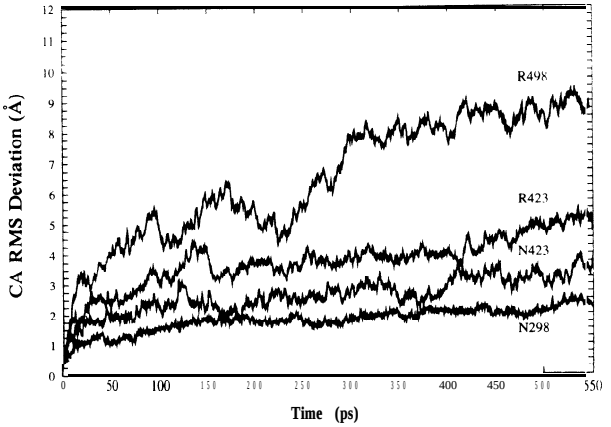


Figure 2. The average r.m.s. a-carbon displacements from the X-ray crystal structure for all residues in the BPTI models a's a function of time. Values were computed after optimum superposition of the coordinates to remove translational and rotational motion using the approach outlined by Kabsch (1976).

(a) *Global structure*

We define global structure at the level of the entire protein. This structure is probed. however. by both global and local means.

(i) *Deviations from the crystal structure*

Most large-scale changes in the overall root-mean-square (r.m.s.) a-carbon displacement from the crystal structure occurred within 50 ps (Fig. 2). From this point, there was a slower drift away from the crystal structure. The value for the control simulation (N298) was low, and the coordinate averaged r.m.s.d. (the average structure between 50 and 550 ps compared to the crystal structure) was 1·4 Å.

The average cc-carbon r.m.s. displacements after 50 ps were high in reduced BPTI and increased with temperature (Fig. 2 and Table 2). The increase in

Table 1

Summary of the different simulations performed

	Models		
Temperat ure (K)	Oxidized BPTI	Reduced BPTI	Isolated BPTI
298	**N298**	R298	
423	**N423**	R423	**1423**
498		R498	**1498**

N, oxidized; R. reduced; I, isolated, e.g. separate simulations of the excised /?-sheet and C-terminal a-helix.

Table 2

Overall properties of BPTI as a function of temperature and oxidation state of the disulfide bonds

Property	Crystal structure	N298	N423	R423	R498
Simulation time (ps)	0	550	550	550	284
<C^α r.m.s.d. > (Å)†		1·7 ± 0·3	2·7 ± 0·5	3·9 ± 0·8	5·1 ± 1·0
<C^α fluc> (Å)†	0·6 ± 0·2	0·6 ± 0·2	1·1 ± 0·4	2·1 ± 0·8	2·4 ± 0·7
d(C30–C51) (Å)‡	2·0	2·0 ± 0·1	2·0 ± 0·1	13·0 ± 3·4	14·4 ± 3.6
d(C5–C55) (Å)‡	2·0	2·0 ± 0·1	2·0 ± 0·1	11·1 ± 3·2	10·4 ± 3.2
d(C14–C38) (Å)‡	2·0	2·0 ± 0·1	2·0 ± 0·1	5·8 ± 1·7	7·9 ± 2·8
Overall %α§	33	28 ± 3	28 ± 4	27 ± 4	22 ± 4
Overall %β§	45	43 ± 2	42 ± 3	34 ± 6	45 ± 3
No. ring flips‖		0	4	16	15

† The average C^α r.m.s. displacement from the crystal structure after equilibration (50 to 550 ps) and the average C^α r.m.s. fluctuations about the mean structure during the last 150 ps of the simulation, except in the case of R498, where the averaging period was 134 to 284 ps, representing the molten globule state. Values are computed after optimum superposition of the coordinates (Kabsch, 1976) to remove translational and rotational motion. The C^α fluctuation for the crystal structure was estimated from the experimental B factors using the following relationship: C^α fluc = $(3B/8\pi^2)^{\frac{1}{2}}$.

‡ These are the distances between the sulfur atoms of the residues designated and represent the native disulfide bonds in BPTI.

§ These are the overall percentages of and β structure averaged during the simulation. The structure is defined by (ϕ,ψ) angles as described in the text.

‖ The total number of ring flips during the simulation where a flip is defined as a 180° rotation about χ_2 of the Phe or Tyr residues.

the C^α displacement in reduced BPTI was due to both unfolding and the increase in temperature in the absence of large-scale unfolding (compare N298, N423 and R423, Table 2 and Fig. 2). The simulations at high temperature also showed large fluctuations in the a-carbon positions in short periods of time, unlike the native protein, indicating increased mobility (Fig. 2). Two phases are clearly visible in the plot for the R498 simulation in Figure 2. Following the initial increase, the r.m.s. deviation stabilizes at approximately 5 Å, although there are large departures from this value; the region between 50 and 300 ps displays the properties of the molten globule state (Daggett & Levitt, 1992*a*). At approximately 300 ps there is another transition to deviations of approximately 9 Å; this region represents the unfolded state. Since our goal here is to characterize the N→MG process, the average values reported throughout this manuscript are taken only during the MG region of the R498 simulation and do not represent the second unfolding phase.

The deviation from the crystal structure observed during MD simulations was not distributed uniformly along the sequence (Fig. 3). Instead, the largest changes occurred in the loop and turn regions of the structure. Interestingly, the regions that deviated most from the crystal structure at room temperature serve as indicators of where the structure is likely to unfold at higher temperature (N 298s *versus* R498s, Fig. 3). These regions are also the most variable among the different BPTI crystal structures (Kossiakoff *et al.*, 1992). Not only did the loops experience the greatest movement in the structure but they also did so quickly. In comparing snapshots of reduced BPTI at 498 K early in the simulation to the crystal structure, it can be seen that the loops moved out away from the core of the structure and adopted slightly different conformations (Fig. 4). Later in the simulation the structure of the loops was even more distorted (Fig. 5). In contrast, the averaged structure over the last 50 ps of the N298 simulation looked very similar to the crystal structure (Fig. 5(a)). Reduced BPTI at high temperature shows that the loops have moved away from the core of the molecule with the segment: of secondary structure separated

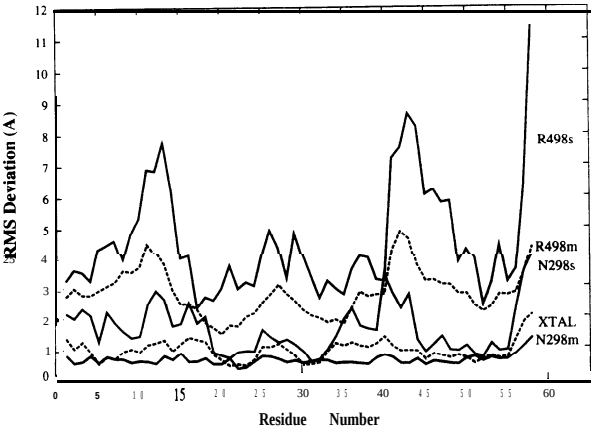


Figure 3. The average r.m.s. a-carbon displacements in the BPTI models from the X-ray crystal structure as a function of residue number, N298s and R498s, and the fluctuation about the mean structure during the last 150 ps of the simulation, N298m, and the last 150 ps of the MG region of the R498 simulation (150 to 300 ps). R498m. The experimental values derived from the crystallographic B factors are also given, XTAL.

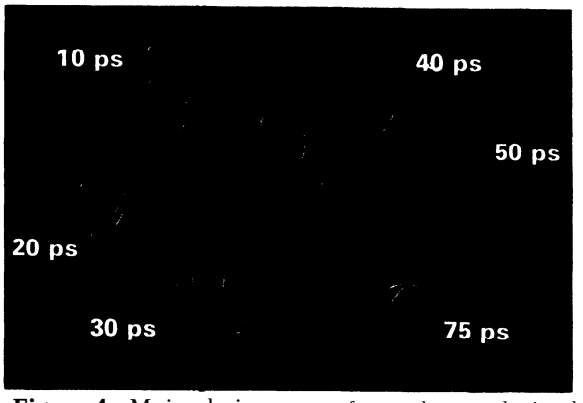


Figure 4. Main-chain traces of snapshots early in the simulation of reduced BPTI at 498 K (in green) compared to the crystal structure (in red).

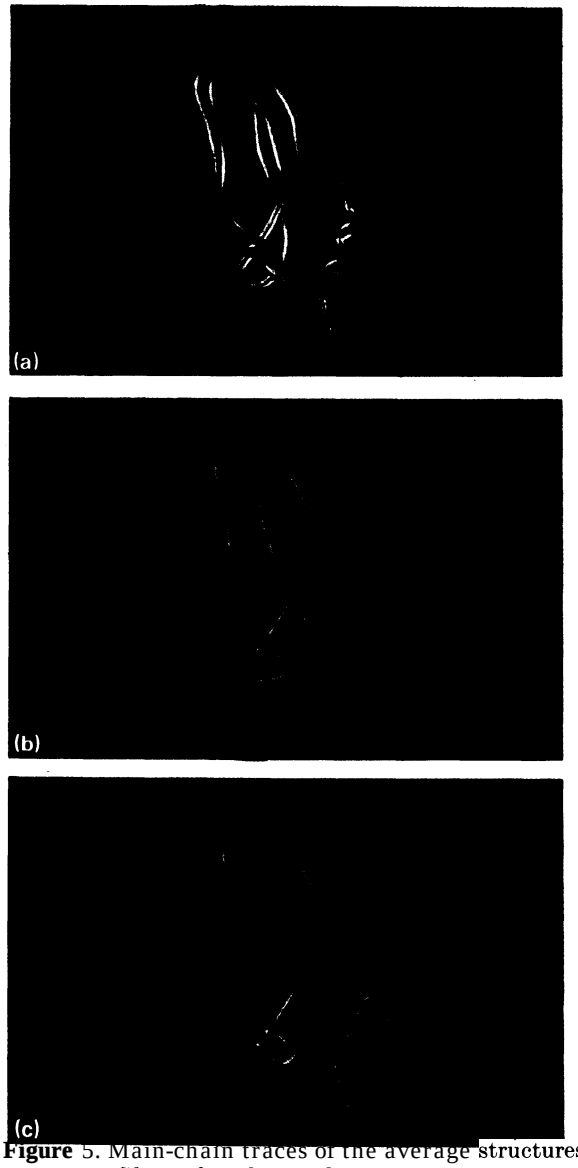


Figure 5. Main-chain traces of the average structures (over the last 50 ps of each simulation, in green, except for R498, which was from 250 to 300 ps) compared to the X-ray crystal structure (in red): (a) native BPTI at 298 K; (b) reduced BPTI at 423 K; (c) reduced BPTI at 498 K.

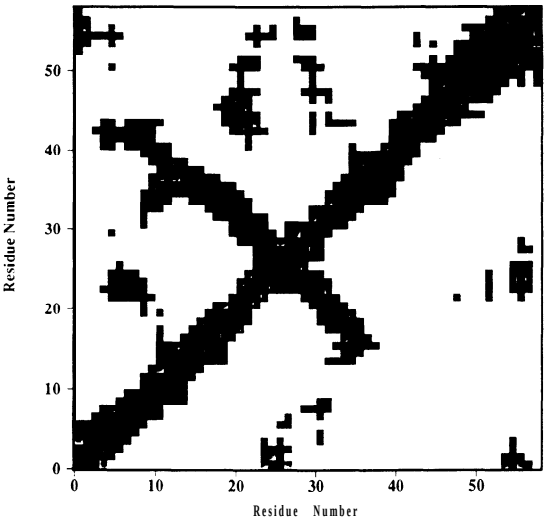


Figure 6. Combination contact map for native BPTI in the upper-left and reduced BPTI at 498 K after 175 ps of molecular dynamics in the lower-right diagonal. The darkest blocks represent C^{α}-C^{α} distances ≤ 5 Å, the next lightest shade is for $5\ \text{Å} < d \leq 7$ A. and the lightest shade is for $7\ \text{Å} < d \leq 10$ A.

in space and hence less well packed (Fig. 5(c)). Interestingly, at 498 K the twist of the /?-sheet was lost and. in fact, the sheet became extended compared to the control.

The a-carbon contact maps for structures at different temperatures differ considerably (Fig. 6, where the shading represents the distance of approach). In the crystal structure (top of Fig. 6), various tertiary contacts (off-diagonal contacts) between different portions of secondary structure are evident. In contrast, very few of these contacts were maintained in the partially unfolded structure of reduced BPTI at 498 K (bottom of Fig. 6). Most striking@. the contacts between residues in the loops were essentially absent. Only two types of non-native contacts were made in this particular structure at 498 K: the N terminus came into contact with β_1 and the C terminus (last 2 residues) bent back to interact with α_2. The central β-sheet was clearly maintained at 498 K, as shown by the opposing diagonal near the center of the Figure emanating from the main diagonal (Fig. 6). The extension of the diagonal representing the β-sheet in the crystal structure shows interactions between the main chain of Ll with $\beta 2$ and with L2. Traces of the main chain in Figure 5 also elucidate the changes that took place upon unfolding.

(ii) *Conformational sampling*

Just viewing a single average structure of the partially unfolded state can be misleading. Even after the large-scale changes occurred and the a-carbon r.m.s. deviation from the crystal structure had stabilized (Fig. 2), the reduced models at high

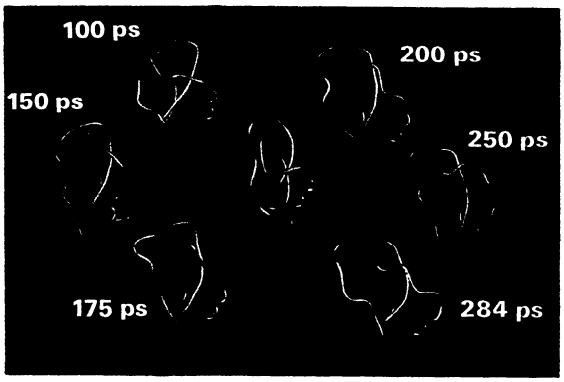


Figure 7. Main-chain traces of snapshots late in the simulation of reduced BPTT at **498** K (in green) compared to the crystal structure (in red).

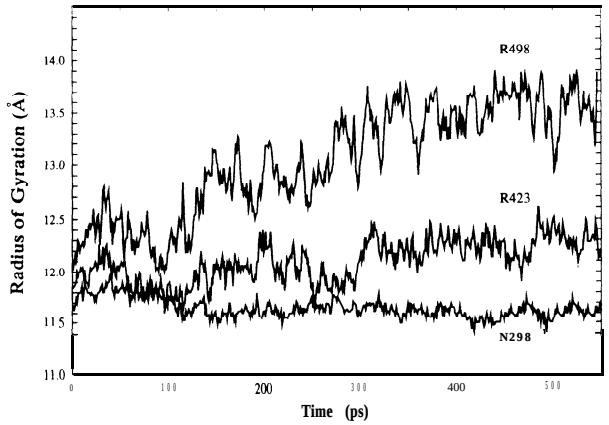


Figure 8. The radius of gvration as a function of time for the different BPTI mod&. The value for the starting structure was 11·7 Å before MD simulation.

temperature were very mobile and exhibited large fluctuations about their mean conformations. The fluctuations of the a-carbon atoms about their mean positions increased dramatically upon unfolding, from 0·6 Å for N298 to 2·4 Å for R498. although some of this increase is due to increasing the temperature (compare N298. N423 and R423, Table 2). The degree of motion observed in the N298m run is in good agreement with those values derived from the crystallographic R-factors (C" Auc = $(3B/8\pi^2)^{\frac{1}{2}}$, Table 2). In fact, the profiles of how these fluctuations are distributed along the sequence are also similar (compare the XTAL curve with N298m, Fig. 3). The cc-carbon fluctuations at high temperature indicate that the partially unfolded state was very heterogeneous and it can be seen that many distinct) conformations were sampled (Fig. 7), even while the overall a-carbon r.m.s. deviations for the displayed structures were very similar (Fig. 2).

(iii) Size

The radius of gyration serves as another measure of the global structure. At low temperature (N298), the radius of gyration of native BPTI was native-like throughout the simulation (Fig. 8). The value increased at high temperature, and. in fact, large changes occurred early in the simulation with further gradual increases with time. At 423 K the radius of gvration stabilized at approximately 12·3 Å. At 498 K, however, the value both fluctuated substantially over short increments of time and continued to increase throughout the simulation (Fig. 8). This radius of gyration was 9 % greater in the R498 simulation (0 to 300 ps) than in N298; this difference translates to a volume change of 25% if we assume that the protein is spherical. This effect is not due to the increase in temperature, N423 did not experience a volume increase relative to N298 (the average values were 11·6 Å and 11·7 Å for N298 and N423, respectively), while R423 expanded by 11 %. The increase in the radius of gyration may actually underestimate the volume increase because of the shape change involved in unfolding from the pear-shaped native molecule (Fig. 5(a)) to the more spherical reduced form at high temperature (Fig. 5(b) and (c)).

(iv) *Packing interactions*

Since the protein volume increased, the packing density decreased, or in other words, the packing of side-chains became disrupted and looser compared to the control. Many residue-residue contacts were lost within 10 to 25 ps (Fig. 9(a)). In addition, the tertiary structure was very dynamic at high temperature with the packing fluctuating dramatically with time even after reaching a stable value (Fig. 9(a)). The percentage of native contacts that remained intact during the simulations dropped to between 35 and 50% in the unfolded models and to 80% in the control simulations, where an inter-action was defined as occurring when two atoms; were within 4·5 Å of one another (Fig. 9(b)). The total number of tertiary contacts did not drop as drastically, however, because of the formation o-f non-native contacts: 13% of the total contacts were non-native in N298 and N423, while 30% were non-native in the reduced models.

Figure 10 shows the actual residue-residue contacts present in the BPTI crystal structure, and during the last 150 ps of the N298 run, and between 134 and 284 ps of the R498 simulation to further investigate the specific nature of the native and non-native interactions present during the MD simulations. These maps differ from the contact maps presented above, because, instead of taking an average structure and determining changes in the distances between the main-chain atoms, here we chose a fixed residue-residue distance and determined the time that each pair of residues was in contact ($\leq$4·5 Å separation). The degree of shading indicates the extent of time that the residues were in contact (darkest, longest time). Both native (top diagonal) and non-native (bottom diagonal) contacts are given for the simulations.

The interactions between the residues in the central P-sheet persisted in N298 and R498 (shown

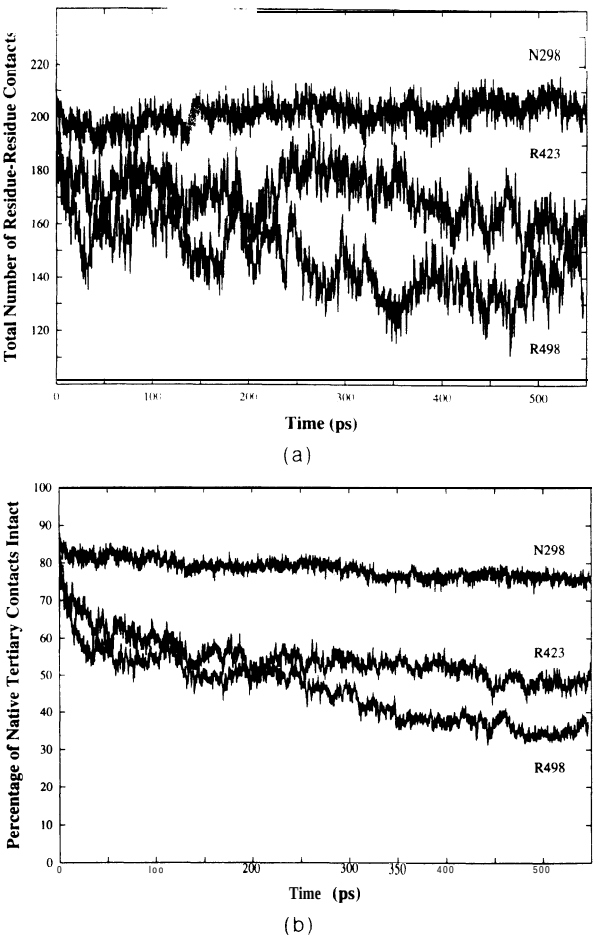


Figure 9. (a) The total number of tertiary contacts intact as a function of time. (b) The percentage of native tertiary contacts as a function of time. Residues are considered to be in contact if any atoms within the residues of interest were within 4·5 Å and when the residues are not neighbors.

as the cross-diagonal emanating from the main diagonal, Fig. 10). The interactions between residues in Ll and β and L2, shown as the extension of the /?-sheet diagonal, were present in N298 but lost in R498. The C-terminal a-helix was maintained during MD simulations, as shown bv broadening of the main diagonal. The other off-diagonal contacts in Figure 10 represent tertiary interactions between the secondary structure segments and loops. Essentiallv all of the tertiarv interactions present in" the crystal structure were maintained in the N298 simulation, although some residue-residue contacts were shifted by one to three residues (90% of the non-native interactions were within 1 to 3 residues of being native interactions: compare the upper and lower diagonals of Fig. 10(a) and (b)).

In contrast. in reduced BPTI at 498 K most of the native tertiary interactions *were* lost during the MD simulations. The interactions that did remain were not highly populated (Fig. 10(c)). The number of non-native contacts increased appreciably, though, compared to N298: 30% of the total

Table 3

The percentage of time that interactions between the aromatic rings were intact during the last 150 ps of the simulations and the MG portion of the R498 simulation

Residues involved	Crystal structure	N298	N423	R423	R498
F4/F22		0	0	26	39
F4/F33		0	0	0	39
F4/F45		5	26	40	29
Y10/F33		10	0	7	4
T10/T35	x	100	7	16	0
Y21/Y23		84	69	99	88
Y21/F33	x	100	100	78	96
Y21/F45	x	100	100	44	0
F22/F33	x	100	100	100	99
F22/F45	x	84	94	98	29
Y23/F45	x	52	70	42	0
F33/Y35	x	100	100	100	100
F33/F45		0	0	68	15
Y35/F45		0	86	68	18

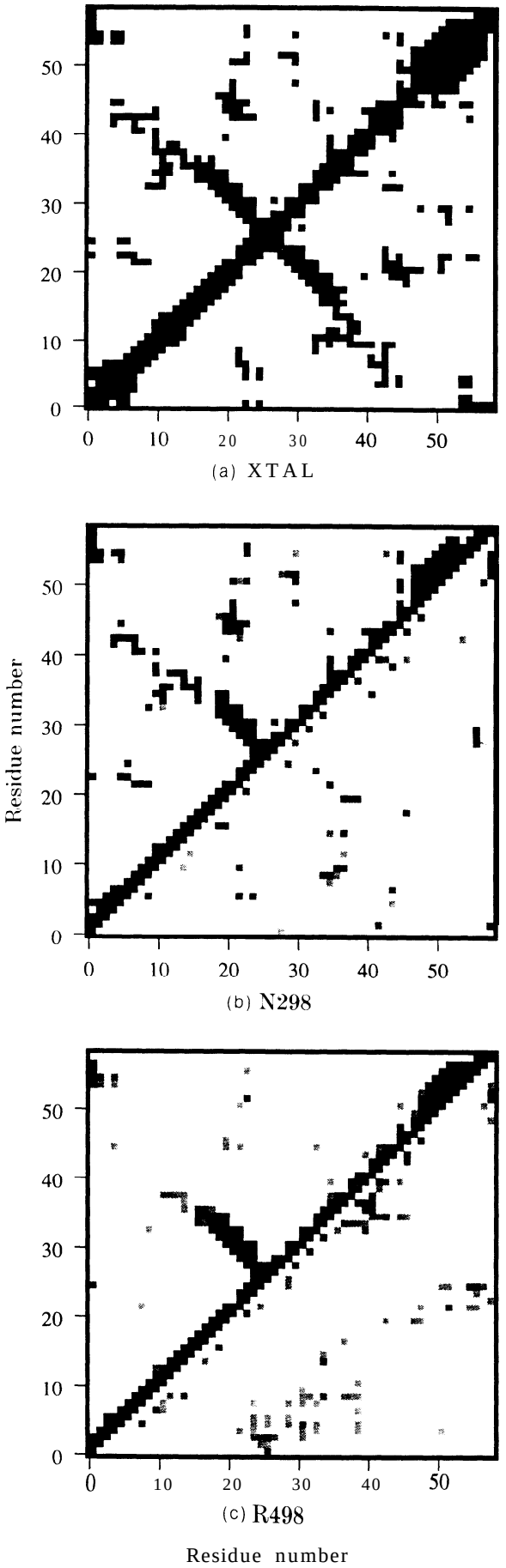
Residues were considered to be in contact if any atoms within the residues were within 4·5 Å.

number of contacts were non-native at 498 K *versus* 13 % in the simulation of native BPTI. Unlike N298, only 50% of the non-native contacts were conservative interactions similar to those found in the crystal structure. Most of the non-native contacts were short-lived and were quickly replaced by other non-native interactions.

As can be seen in Figure 10(c), many non-native contacts in R498 involved new interactions with the N-terminal region of the molecule (residues 1 to 10, comprising α1 and Ll). as a result of the change in position of the loop. Interestingly, the non-native contacts between Phe4 and Phe22 and Phe33 were populated 40 % of the time (Table 3), while many Tyr and Phe interactions were lost upon unfolding (2 of the 5 original Phe/Tyr interactions were maintained and one is between $i \rightarrow i+2$ neighbors in the /?-strand, Table 3). A non-native contact involving Tyr21 and Tyr23 was prevalent, however. Also, because of the change in position of L2, Ala25 now interacts with many' residues in α2 (Cys51, Met52, Cys55 and Ala56) up to 72% of the time. A new turn formed around Gly37 of R498 putting this residue, which previously was not interacting with any other residue, in contact with residues nearby, in L2 (Arg39, Ala40 and Lys41) 95% of the time.

(v) Hydrophobic core

The changes in the Tyr interactions are best illustrated graphically (Fig. 11). The distribution of the Phe and Tyr residues changed dramaticallv upon unfolding with the Tyr residues partitioned at or near the surface of the molecule to interact with water. Partially unfolded BPTI did maintain some features of the native hydrophobic core, however. Phe22 and Phe33 were most important in stabilizing the hydrophobic core and they were always in contact (Table 3 and Fig. **11).** Phe4 and Phe45 also interacted with these central residues some percent-



age of time and the non-native Phe4/Phe22 and Phe4/Phe33 interactions contributed to the stability of the hydrophobic core. These differences in the Tyr and Phe interactions highlight how the internal packing changed in the unfolded structures. As a result of these changes and expansion of the models. the occurrence of ring flips (180° rotation about χ_2) increased dramatically from 0 at 298 K to 15 to 16 in the reduced protein at high temperature (Table 2). This phenomenon is not just a result of the increase in temperature; the number of ring flips is much larger in the absence of the disulfide bridges (N423, 4 flips; R423, 16 flips).

Hydrophobic clusters remained during unfolding and within the partially unfolded conformations generated. These hydrophobic clusters contained a subset of the interactions of the native protein but were very dynamic. In addition, some non-native interactions were quite persistent and contributed to the stability of the hydrophobic core. This behavior is consistent with Lesk & Roses' (1981) hierarchic condensation model for folding, in which neighboring hydrophobic residues interact to form folding clusters. Such interactions can then aid collapse of secondary structure elements, thereby accelerating folding. Garvey *et al.* (1989) lent support to this model experimentally by showing that a hydrophobic cluster forms early in the folding of dihydrofolate reductase. **Heringa** & Argos **(1991)** also came to the conclusion that side-chain clusters are important in folding by identifying such clusters in crystal structures. **They** found that Tyr and Phe were highly preferred as clustering residues, as were many polar and charged residues; however, their work finds clusters in the native state and may not be indicative of their importance during the folding process.

In our simulations, the sources of stabilization of the hydrophobic core were predominantly the aromatic groups and Met52. In particular, the Phe/Phe interactions helped to hold together the C-terminal a-helix and the central b-sheet, even though the interactions were dynamic and not fixed in time. The Phe/Phe interactions were more important in stabilizing the partially unfolded form than were the Tyr/Tyr interactions. Serrano *et al.* **(1991)** have shown that the two interactions make identical contributions to protein stability but that Tyr is preferred to Phe at the surface. The Tyr residues did

Figure 10. Residue-residue contact maps. (a) Tertiary contacts within the crystal structure. (b) Tertiary contact map for the last 150 ps of the simulation of N298. (c) Tertiary contact map for the last 150 ps of the molten globule region of the R498 simulation. The upper diagonals represent native contacts and the lower regions contain non-native interactions. The positions are shaded according to the percentage of time that the residues were in contact: there are 4 levels of shading for interactions **present** $>10\%$ of the **time:** $80\% \leq t \leq 100\%$; $60\% \leq t < 80\%$; $40\% \leq t < 60\%$; **and** $10\% \leq t < 40\%$.

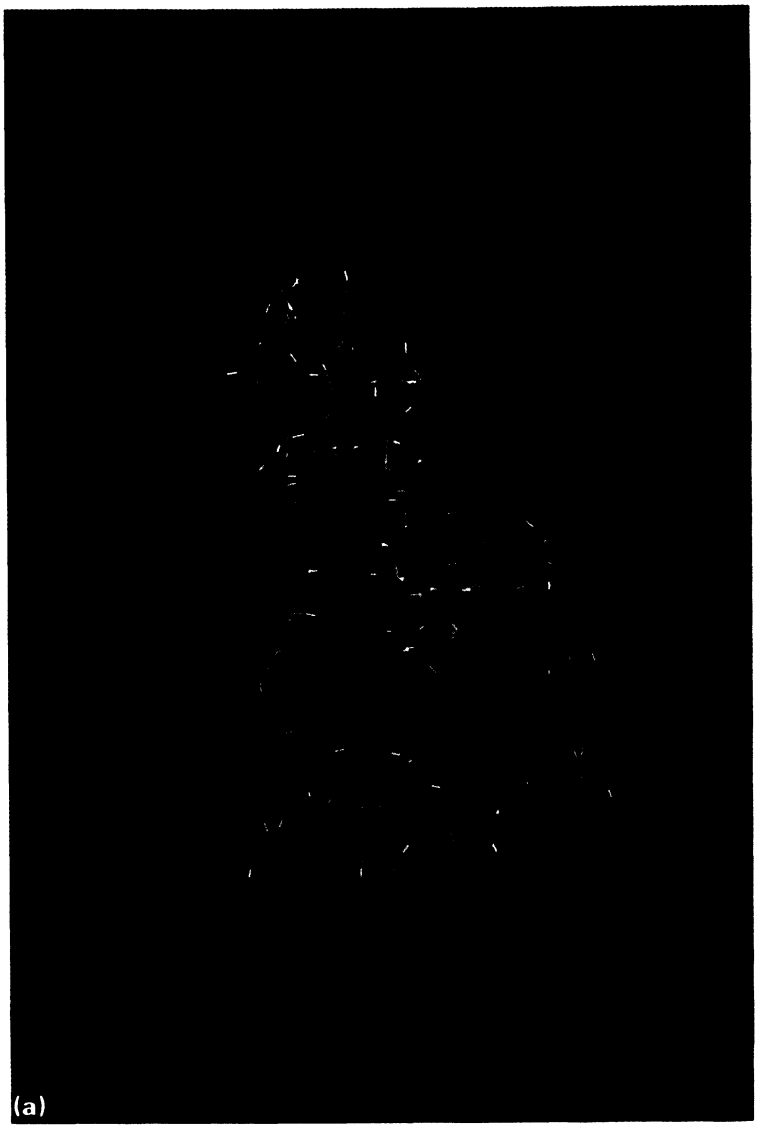


Fig. 11.

indeed partition themselves at the surface of the molecules in the high-temperature simulations.

Furthermore, it appears that hydrophobic clustering is not only important during folding and stabilization of folding intermediates but also within the denatured state. The temperature-denatured form of lysozyme provides an example of this phenomenon (Evans *et al.*, 1991). Interestingly, for this protein there is a residual hydrophobic effect giving rise to non-specific clusters involving particular larger hydrophobic alkyl and aromatic groups (Evans *et al.*, **1991).** We also observed such non-specific hydrophobic clustering in the unfolded state. which will be addressed elsewhere (unpublished results) .

(vi) *Disulfide bonds*

The distances between the sulfur atoms involved in disulfide bonds in native BPTI yield another measure of the overall structure. These distances increased dramatically in reduced BPTI at high temperature, especially for the disulfide bonds in the interior of the molecule, Cys30-Cys51 and Cys5-Cys55 (Fig. 1, Table 2); in addition, all of the non-native pairings of the thiol groups were long (data not presented). The average values for the different runs are quite similar prior to complete unfolding of R498; however, they do show different profiles with time (Fig. 12). These disulfide bonds are important in stabilizing BPTI while the 14–38 bond on the surface (Fig. **1)** can be selectively reduced without seriously affecting the protein s integrity (Marks *et al.*, **1987;** Goldenberg, **1988).** The C14–C38 distance also became long in the unfolded forms but, surprisingly, remained closer than those in the core, showing how much the bottom of the molecules (as depicted in Fig. 1) has opened upon unfolding.

(vii) *Secondary structure*

Even with the changes in the tertiary structure upon unfolding, the overall amount of secondary

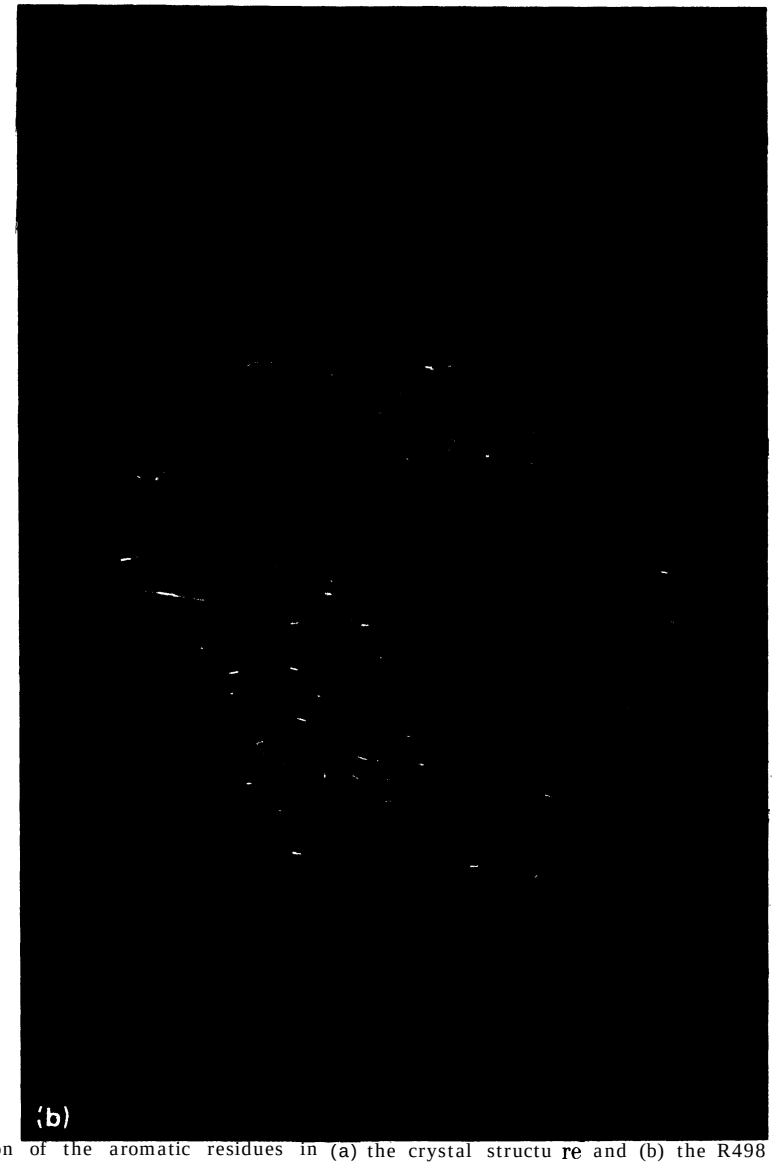


Figure 11. Distribution of the aromatic residues in (a) the crystal structu re and (b) the R498 average structure, The Tyr residues are shown in cyan and the Phe residues are given in magenta.

structure remained native-like, on average, with between 61 and 71% of the residues in either the a-helical or b-region of conformational space over time (Table 2). Secondary structure assignments are based on the backbone dihedral angles, ϕ and ψ. The residues are considered to reside in the helical region of conformational space when the values are within approximately 40° of the most commonly observed values in helical regions of various crystal structures ($-100° \leq \phi \leq -30°$; $-80° \leq \psi \leq -5°$, where the canonical values are $-63{\cdot}8(\pm 6{\cdot}6°)$ and $-41{\cdot}0(\pm 7{\cdot}2)°$ for ϕ and ψ, respectively: Presta & Rose, 1988). The actual boundaries were determined from the cluster of (ϕ,ψ) values observed in our previous simulations of polyalanine helices (Daggett & Levitt, 1992*a*). The β-region was bounded as follows: $-170° \leq \phi \leq -50°$, $80° \leq \psi \leq 190°$. This range is sufficiently broad to account for both parallel and antiparallel /?-conformations and was determined by the clustering of (ϕ,ψ) values in simulations of native BPTI at 298 K and in simulations of isolated β-sheets composed entirely of alanine residues (data not presented). Justification for the use of dihedral angles for quantifying secondary structure contents has been given by Daggett *et al.* (1991*b*).

Although the overall amounts of secondary structure remained large in reduced BPTI, the partitioning between α and p-structure shifted compared to native BPTI. The β content of reduced BPTI at 423 K dropped (Table 2). Upon raising the temperature further to 498 K, the amount of p-structure increased and slightly exceeded that of native BPTI, although the a-helix content decreased. Hydrogen bonds were not absolutely necessary to maintain the secondary structure but they were

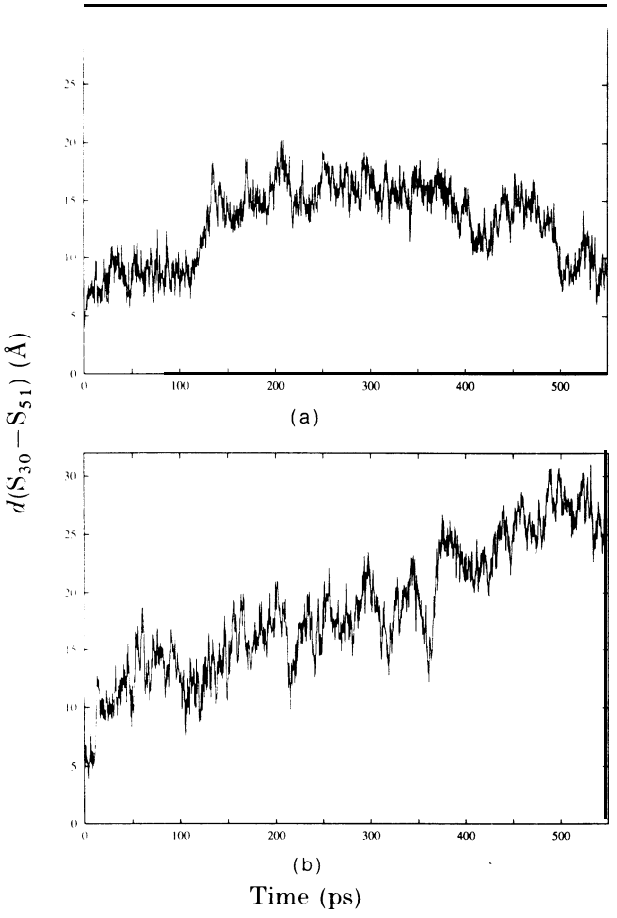


Figure 12. The sulfur distances between Cys30 and Cys51, which form an interior disulfide bond in native BPTI, as a function of time: (a) R423 and (b) R498.

prevalent (Daggett & Levitt. 1992*a*). Further discussion of details of the secondary structure is presented below.

(viii) *Summary*

Some of the global properties monitored during the simulations changed dramatically upon unfolding while others were essentially native-like. Clearly. the main-chain positions changed, and many of the a-carbon atoms had average r.m.s. displacements of over 8 Å with maximum displacements as great as ~ 19 Å for the MG state. The magnitude and position of the observed displacements indicates that the structures became only partially unfolded. It is difficult to estimate numerically the deviation that one would expect for a molten globule or other partially unfolded intermediate;. although we do obtain values for this state that are intermediate between N298 and the unfolded region of R498 (e.g. N298, 1·7 Å; R498 (MG, 50 to 300 ps), 5·1 Å; R498 (U, 300 to 550 ps), 8·5 Å). The unfolded state reached in the simulation remained collapsed (see Fig. 8) instead of displaying true random coil behavior and there is now evidence that even denatured proteins under a variety of conditions are collapsed with some degree of structure (Evans *et al.*, 1991; Dobson. 1991. 1992; Dill & Shortle, 1991). Nevertheless, our values seem reasonable for a molten globule. which is native-like in many respects, and is not expected to deviate considerably from the native state. Our results especially seem reasonable given that the magnitude of the change in the radius of gyration, volume and accessible surface area in the reduced forms of BPTI at high temperature are comparable with those observed experimentally for a wide variety of molten globules (Daggett & Levitt: 1992*a*; Daggett, **1993).**

In any case, there was a large change in the a-carbon atom positions early in the simulations (< 10 ps) followed by a more gradual drift from the crystal structure occurring throughout the remainder of the simulations. The radius of gyration, which also reflects the volume of the models, also changed very quickly with time. within a few picoseconds. The volume and contact maps are sensitive indicators of the changes in tertiary contacts, which were clearly disrupted upon unfolding. The tertiary structure began to fall apart quickly in the simulations, although some tertiary contacts persisted. The tertiary packing of N298 was essentially identical with that of the crystal structure. Native contacts that were lost were replaced by conservative interactions that involved shifting of one to three residues, thus maintaining the core and overall fold of the molecule. In contrast, many of the native tertiary contacts were lost in reduced BPTI at high temperature and those that were present were not highly populated. Also, many new non-native contacts were formed in the unfolded models, some of which preserved interactions similar to those found in the crystal structure while others represented new and different pairings. In contrast to the changes in the tertiary packing, the overall amount of the secondary structure changed very little, with any changes occurring gradually.

(b) *Ser.&local and local structure*

This section describes interactions between particular segments of secondary structure and within secondary structure elements. Thus, most of these descriptions still represent tertiary interactions but we are not concerned here with the state of the entire protein.

(i) *Secondary structure*

The extent, of structure within the α_2, β_1 and β_2 segments in BPTI (Fig. 1) was determined in the protein models and in isolated fragments of the protein to determine the inherent stability of the secondary structure in different environments. The secondary structure contents in particular regions of the structure were calculated as described above, based on (ψ,ϕ) values, with the added requirement that at least three successive residues fulfil the angular criterion at any one time. It should be

noted that the absolute stability of the secondary structure may be overestimated because of the short simulation times (only 200 ps for the fragments). Nevertheless. the various simulations can be compared to determine the effects of the protein scaffolding in stabilizing the secondary structure elements.

The C-terminal a-helix (α_2) was stable in all of the protein models during the first 200 ps of MD simulation (Table 4). The isolated helix was stable at 423 K; however, it melted out at 498 K. The helix was substantially more stable in the presence of the protein at both 423 and 498 K. BPTI is usually considered to have another helix at the N terminus (α_1, residues 3 to 6). This helix is distorted, however,, in the crystal structure and, by our definition., contains 0 % helix, which never improved during MD simulations.

In contrast, to α_2. both of the /?-strands adopted more β-structure in the isolated fragments than within native BPTI (Table 4). The β content of R423 was lower than that of N298 but increased to the native values with further heating to 498 K. These results suggest that the /?-sheet was constrained by the surrounding protein environ-- ment; b-structure in these segments was favored as the flexibility increased by removing much of the protein. All of the secondary structure elements were very dynamic at the local level, as shown by the large fluctuations in the secondary structure contents (Table 4); fluctuations of this magnitude (10 to 20 %) correspond to one or two residues changing their secondary structures.

We were concerned with the behavior of the major portions of secondary structure identified in the native protein and how they unfold/refold during the simulation (Fig. 13). Each of the three segments of secondary structure is discussed in turn, but plots are presented for onlv a few of the simulations for simplicity. The values are averaged over 1 ps intervals to smooth the curves, but thev still reflect loss of both semi-local (a few residues) and local (individual residues) structure from the segment. As can be seen in Figure 13, both levels of folding and refolding occurred.

The β-structure of β_1 in N298 and R498 was lower than in the crystal structure but was prevalent throughout the simulations (Fig. 13(a) and (b)). This segment was also quite stable in the isolated

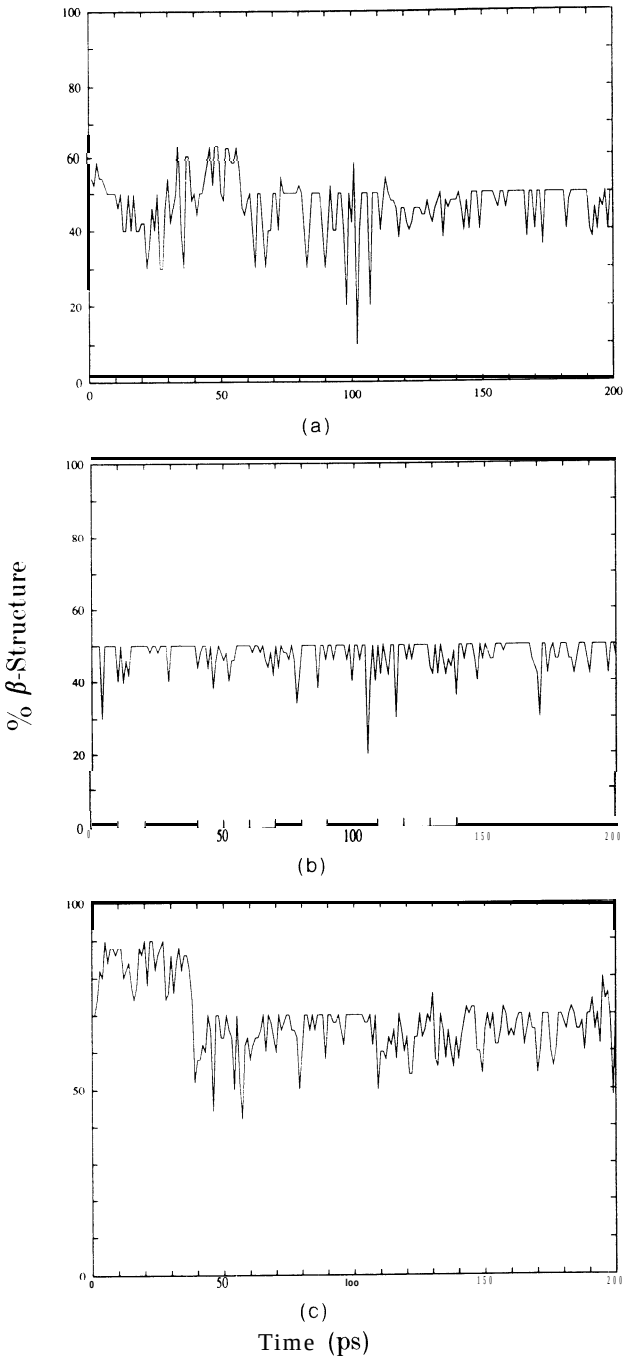


Fig. 13.

Table 4

The percentage of secondary structure in the three principal structural elements of BPTI both within the protein and within isolated fragments (50 to ZOO ps)

Secondary structure†	N298	N423	R423	R498	1423	1498
α_2	70 ± 9	72 ± 16	77 ± 10	65 ± 18	65 ± 24	6 ± 15
β_1	46 ± 13	46 ± 12	0 ± 0	47 ± 8	73 ± 10	65 ± 11
β_2	73 ± 9	70 ± 13	38 ± 20	70 ± 13	74 ± 13	70 ± 14

† The secondary structure segments are shown in Fig. 1 and the boundaries of the segments are given in the Figure legend and list of abbreviations.

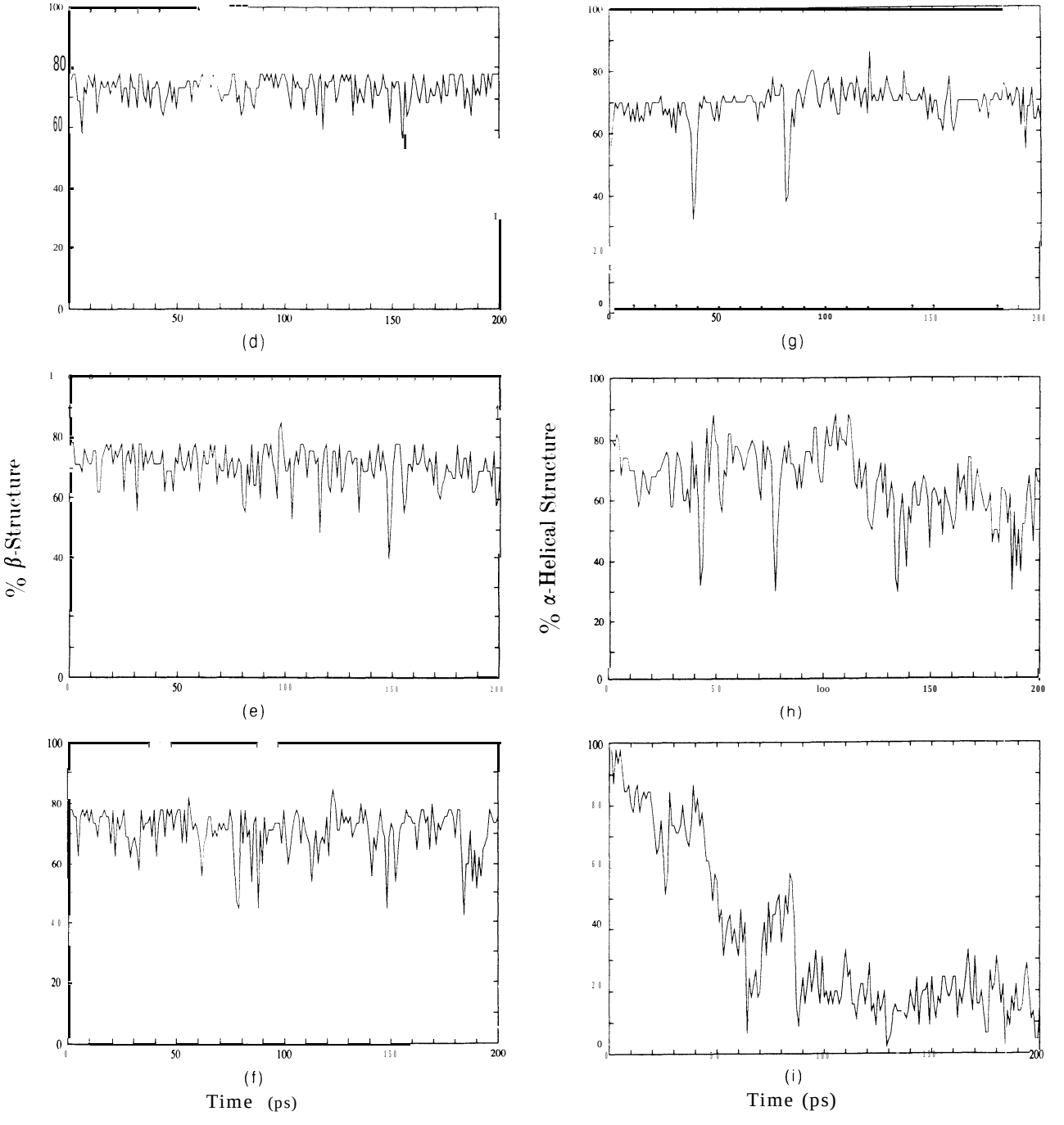


Figure 13. The percentage of secondary structure present in the 3 principal structural elements of BPTI as a function of environment', oxidation state, and extent, of the protein included. Secondary structure assignments were based on (ϕ,ψ) angles with the added constraint that it be repeating structure with the participation of at least 3 successive residues. The labeling is as follows: $\beta 1$ (a) N298, (b) R498, (c) 1498; $\beta 2$: (d) N298, (e) R498, (f) 1498: and $\alpha 2$ (g) N298, (h) R498 and (i) 1498.

P-sheet (Fig. 13(c)) at the same temperature, and exceeded the amounts of β-structure in native BPTI. The β_2 segment showed comparable amounts of structure in the same simulations (compare Fig. 13(d), (e) and (f)).

The C-terminal cc-helix behaved in a manner opposite to that of the P-strands. This helix was stable throughout the MD simulation of native BPTI, N298 (Fig. 13(g)). The helix began to melt out with increasing temperature in reduced BP'TI, although the helix content remained high (Fig. 13(h)). Most of the loss of structure was a result of fraying at the ends of the helix. The helical structure was completely lost in the isolated form of the helix at 498 K (Fig. 13(i)). When isolated. the residues are better able to sample the non-helical regions of conformational space (Fig. 14).

Experimentally, there is reason to believe that

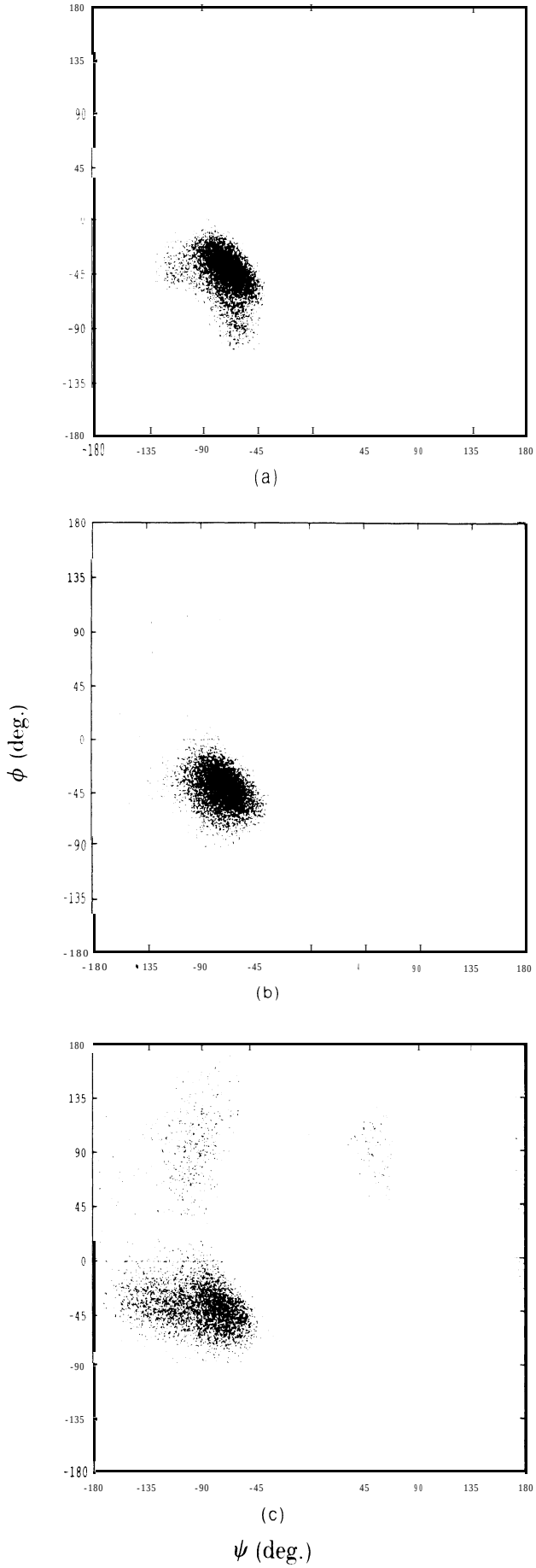


Figure 14. Ramachandran plots showing all (ϕ,ψ) values sampled by the C-terminal a-helix during the simulations: (a) N298, (b) R498 and (c) I498.

the results for the C-terminal a-helix are reasonable. The isolated C-terminal helix is moderately stable in solution, but less so than in the protein (Goodman & Kim, 1989). The results with the P-sheet are a bit more difficult to interpret. Experimentally, this sheet does not contain appreciable p-structure in isolation (Goodman & Kim, 1990) and P-sheet structures have historically been more difficult to construct than a-helical structures (reviewed by Richardson & Richardson, 1989). Our simulations of the isolated fragments were short and will overestimate the stability of these structures. The extreme stability observed for the P-sheet may be due to a very high activation barrier between N and U; this same activation barrier will also slow the folding of such structures. In addition, and most importantly. P-sheets tend to aggregate in solution and foil attempts to characterize them; we, of course. do not have this problem in the simulations.

Even given these shortcomings, the simulations of the reduced protein and isolated P-sheet exhibited higher stabilities of the p-structure than in native BPTI during the same period of time. This result is interesting, given that molten globules often show higher secondary structure contents than the corresponding native states (Ptitsyn, 1987). Also, the increased stability of the isolated sheets may highlight the importance of secondary structure formation early in folding. For example, construction of a fairlv uniform and stable P-sheet can act as a framework onto which further docking of structure can occur. When all, or more, segments have been added to the structure it may not be necessary for the P-sheet to be so stable. It can then bend and twist (as observed in the crystal structure) to better optimize tertiary interactions. If this mechanism does operate during folding: intermolecular aggregation of the P-sheets would probably be less of a problem than when they are isolated because of the presence of the rest of the protein.

(ii) *Turn formation*

We then investigated the occurrence of turns throughout the protein (Table 5). Turns were identified using Kuntz's (1972) definition. The entries in Table 5 represent the percentage of hairpin or near-hairpin turns whichever was more prevalent. The crystal structure shows a pronounced turn between the two P-sheets (residues 25 to 28, Fig. 1) and this is the only turn present in the crystal structure. Interestingly. this turn was highly populated in the simulation of native BPTI at 298 K but was less prevalent in reduced BPTI at high temperature. All of the high-temperature models picked up turns in the region of L2 (residues 34 to 41) that were not present in native BPTI. There was a weak turn in native BPTI between residues 41 and 44, which was present but shifted in the reduced models. Thus, the

Table 5

The percentage of time that turns were present between particular residues in the different BPTI models

Residues involved	N298	N423	R423	R498
8–11	**0**	**0**	6	0
1 l-14	0	8	0	0
16-19	0	0	0	9
24-27	0	0	0	5
25–28	94	87	78	66
28-3 1	0	0	6	0
30-33	0	0	9	0
34-37	0	0	0	6
3740	0	16	65	33
3841	0	0	0	8
40–43	0	29	54	11
4144	25	0	2	2

Turns were defined using the method of Kuntz (1972), which is based on the distance separating i and $i+3$ C^{α} atoms, d, and the angle θ formed by the vector connecting the i and i + **1** C^{α} atoms. $\mathbf{R}_{i,i+1}$, and that for the i+ 2 and $i+3$ atoms, $\mathbf{R}_{i+2,i+3}$. Using this approach. hairpin turns have $\leq 5{\cdot}5$ Å and $\theta \geq 135°$. Near-hairpin turns have either $d \leq 5{\cdot}5$ Å and $135° > \theta \geq 120°$, o r $5{\cdot}5$ Å $< d \leq 6{\cdot}5$ Å and $\theta \geq 135°$. The percentages given represent hairpin or near-hairpin turns, whichever were more prevalent. Any turn present $> 5\%$ of the time in any of the simulations is given. The only turn present in the crystal structure is between residues 25 and 28. Averages are taken after 50 ps and continued to 550 ps, with the exception that the R498 averages were taken to only 300 ps.

native turns were preserved but less populated in the partially unfolded models.

Many non-native turns were formed during unfolding, however, in the exposed surface loop region of the protein and in the /?-strands (this also appeared to be occurring at the ends of the helix, but. unfortunately, the methods used to identify helical and turn structures are not unique and there is overlap, making definitive analysis in this region difficult). These non-native turns were made up of hydrophilic residues, for the most part? with from one or two residues from the top five list of residues found in β-turns (Chou & Fasman, 1978; Jurka & Smith, 1987). The native turns contained two or three of the five most likely residues. Therefore, the non-native turns would be considered to have weak β-turn potential: however, they are similar in composition to β-turns found in native proteins, suggesting that the common methods to identify turns in native proteins may be applicable for finding potential turn sites important during the folding process.

(iii) *Correlated motion*

In an attempt to clarify the types of motion occurring during the unfolding of BPTI, a matrix of the cross-correlation coefficients, C_i, for a-carbon motion was computed using the following relationship:

$$C_i = <\Delta r_i\, \Delta r_j> / (<\Delta r_i^2><\Delta r_j^2>)^{1/2}, \qquad (1)$$

where Δr_i is the displacement of the ith atom from its mean position. Such an approach highlights how segments of the structure move relative to one another (Levitt *et al.*, 1985; Tilton *et al.*, 1988; Harte *et al.*, **1990).** Only correlation coefficients with an absolute value greater than 0·25 are depicted. Figure 15 contains plots for N298 and R498 du-ring the **150** to 200 ps interval of the simulation. This interval was chosen because it is in the molten globule region of R498 and equilibration, such as it is, should be complete. A positive correlation coieffi-

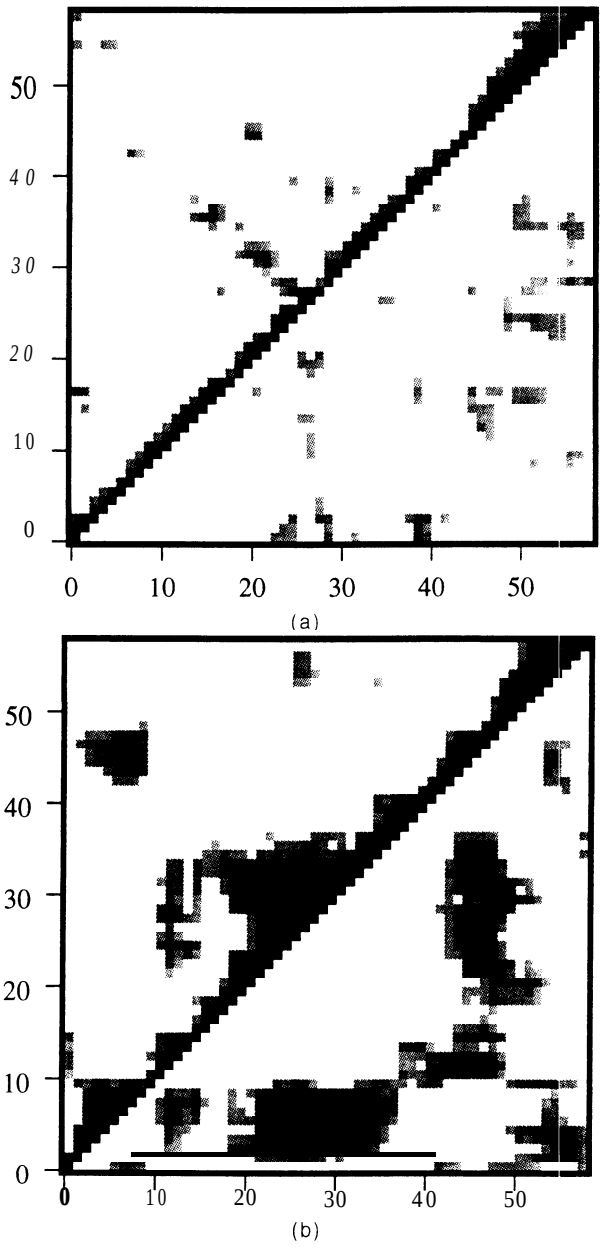


Figure 15. The correlation coefficients relating the motion of pairs of z-carbon atoms between 150 and ZOO ps: (a) N298 and (b) R498. The upper diagonal represents positive correlation coefficients while the lower diagonal contains negative correlations. The correlation coefficients between particular residues are shaded according to the magnitude of the motion: the darkest shade. $0{\cdot}75 < |C_i| \leq 1{\cdot}00$; the next darkest shade, $0{\cdot}50 < |C_i| \leq 0{\cdot}75$; the next darkest shade, $0{\cdot}35 < |C_i| 0{\cdot}50$; and the lightest shade $0{\cdot}25 < |C_i| \leq 0{\cdot}35$. Any correlation coefficient between $-0{\cdot}25$ and $0{\cdot}25$ for any 2 residues is left Hank.

cient indicates t the residues are moving together in the same direction (upper diagonal in Fig. 15). while a negative value represents residues moving in different directions (e.g. toward one another or in opposite directions. lower diagonal in Fig. 15). The magnitude of the coefficient indicates the extent of' correlation between the motion of two residues (the darkest shade in Fig. 15 represents highly correlated motion and the lightest shade is for weakly correlated residues).

In native BPTI the motion of' the residues was only weakly correlated, suggesting that the protein was, mostly, just vibrating about its mean conformation (Fig. 15(a)). Comparison of Figures 12 and 15 shows that residues in contact tend to have positive correlation coefficients (e.g. the residues within the ('-terminal helix. denoted by the broadening of the diagonal in the upper right-hand corner, and between the b-strands, represented by the off-diagonal projection emanating from the center of Fig. 12 and 15(a). the other regions represent) interactions between loops and the sheet or helix). The regions of negative correlation include residues in contact and those that are not. But, in all cases the motion was only weakly correlated. The observed weak correlation between residue movement and lack of long-range correlated displacements is consistent) with diffuse scattering experiments recently described by Badger (1992) and reviewed by Caspar & Badger (1991).

In contrast, reduced BPTI at'498 K showed very strong correlations between different residues and extended correlated regions (Fig. 15(b)). As in native BPTI, the motion of the /?-strands was correlated and the motion of the strands was linked to that of the turn. In addition, the motion of the residues within the strands was highly correlated (shown as a large broadening of the diagonal representing the β-sheet), unlike in native BPTI. The helical residues were highly correlated. and the motion of the C-terminal end of the protein became correlated with that of the helix. Most all of the tertiary interactions between different secondary structure segments were characterized by large negative correlation coefficients. The motion of the two loops was positively correlated. The motion of all regions of structure within BPTI was negatively correlated with the central β-sheet. Thus. this plot leaves the picture of the /?-strands moving in concert while the bulk of the structure has pulled away from the sheet and is oscillating, moving toward and away from the center. The large differences between R498 and X298 can be attributed to low-frequency correlated motions associated with the liquid-like molten globule, as opposed to the smaller amplitude vibrations associated with the solid-like native state.

(iv) *Summary*

These results suggest that the protein scaffolding is very important in stabilizing helical structure; in the absence of the protein, the C-terminal helix was only marginally stable. The scaffolding appears to be detrimental to the b-structure; partial release of structural constraints by either increasing the temperature or removing the rest of the protein, lead to increased b-content. The β-strands were mutually stabilizing. In our previous simulations of isolated /?-strands (not sheets) we found that the β conformation is inherently less stable than a-helical structure (Daggett & Levitt, 1992*a*), as expected. From the findings presented here. β-sheets may be inherently more stable than was thought originally. The advantage to forming fairly uniform and stable β-sheets during folding is that in so doing a large amount' of non-polar surface area is buried, creating a platform for further docking of secondary structure. As a result of these findings, we have investigated other β-sheets to test the generality of our findings and the importance of sequence in stabilization of the β-sheet (unpublished results).

In the case of the C-terminal a-helix. the protein aided in stabilization of the helical structure even when in the partially unfolded form (R498); the stability of the helix decreased and fraying occurred at the ends of the structure. but it was partially stable. In contrast, the isolated helix unfolded completely. In N298, the amount of structure within the secondary structure segments was very dynamic, with individual residues converting rapidly between different conformations. In the partiallv unfolded forms of BPTI, conformational conversions at the single-residue level occurred in addition to more larger scale unfolding and refolding involving three or four residues between successive structures in time. In some cases, complete unfolding of the sequence occurred.

In anv case? we found that there were substantial amounts of native structure present both during unfolding and in the partially unfolded state. The β-strands and a-helix were mutually stabilizing and remained relatively intact with unfolding occurring mostly in the turn and loop regions of the protein. In investigating the stability of this folding unit, we also simulated (Daggett & Levitt, 1992*a*) a BPTI fragment, containing the C-terminal a-helix and most of the β-sheet that had been synthesized and characterized by Oas & Kim (1988). In accord with their findings, the fragment contained native-like structure, demonstrating the importance of particular interactions, or the folding units themselves, not only to the native state but also to the folding process, as suggested by Levitt & Chothia (1976). This idea. was confirmed bv further studies from Kim's lab: another fragment was constructed and on the basis of its behavior it was concluded that much of the protein folding problem can be reduced to the identification and characterization of subdomains of native proteins (Staley & Kim, 1990).

Various non-native turns were adopted during the unfolding simulations. Dyson *et al.* (1988) have argued that the local amino acid sequence alone determines b-turn conformations, which are stabilized by short-range interactions. Therefore, turns are perfect candidates for protein folding initiation sites. Our results are in accord with this

hypothesis. Our results sugget that turns may be important targets of attack during unfolding' and the creation of non-native turns may aid in maintaining compact unfolded states by fragmenting the structure. Conversely, folding may be facilitated by the presence of these fragments. It has been proposed that β-turns are important in the early stages of folding (Lewis *et al.*, 1971: Dyson *et al.*, 1988). Although our simulations cannot address these early our results suggest that turns may be important also in the latter stages of folding.

4. Conclusions

An earlier description of features of the simulations presented here suggests that BPTI unfolds to a relatively stable compact set of conformations? which exhibits all of the known properties of the molten globule state (Daggett & Levitt, 1992*a*). The partially unfolded structures contain both ordered and disordered regions but have retained many features of the native state. The control simulation of oxidized BPTI at room temperature was well behaved with only minor deviations from the crystal structure. Some changes are to be expected since the crystal structure represents a single average conformation for a protein with minimal amounts of water present, while we have attempted to simulate the motion of a protein in solution. In any case. this is the longest reported solution simulation of a protein and it provides support for the use of empirical force-fields in modeling the native states of proteins. Furthermore, even though the potential functions were originally parameterized to reproduce crystal structures they work surprisingly well for modeling partially unfolded conformations. The ability to begin and end a simulation at experimentally reasonable points lends more credence to our description of the unfolding pathway than if our unfolded states had no experimental basis. This is one of the biggest problems with attempts to simulate the details of folding. Judicious choice of starting "unfolded" structures can dramatically affect the success of such schemes (Meirovitch & Scheraga, 1981). Unfortunately, the nature of the denatured state is unknown and vigorously debated, but we believe that it is safe to explore the pathway between two known states.

As explained at the outset, we use high temperature to both destabilize the folded form and accelerate the unfolding process by reducing the time required to cross the activation barrier between the folded and unfolded forms. We also use the reduced form of the protein, which is much less stable than the native conformation. In this way, we are able to simulate large-scale unfolding events on the subnanosecond time-scale. What are the consequences of running simulations at temperatures as high as 498 K? One change is to decrease the density to the value observed for liquid water at this temperature (0·829 g/ml at 225°C). This change is necessary to maintain the internal pressure of the surrounding water close to atmospheric pressure; without the density change, the high temperature run would subject the protein to huge compressive pressures. With the change, the volume available to the protein is the same as at lower temperatures and the rapid initial expansion cannot be attributed to a density change. We are currently exploring other methods of accelerating unfolding; including solvent changes that include high concentrations of methanol, urea and guanidinium hydrochloride.

Matthews (1991) has reviewed the status of experimental studies addressing the mechanism of protein folding. He separates the process into three stages: early events in folding: middle events. which are between the very earliest steps to the rate-determining step; and late events dealing with conversion of the transition state to the native conformation. During the earliest stage of folding. substantial secondary structure is formed (perhaps in less than a millisecond). The middle events are characterized by further formation of secondary structure and so-me tertiary structure, at least some of which has native-like character. During the final stage of folding, any necessary rearrangement and repacking of secondary structure and the core is accomplished. Dehydration of the hydrophobic surfaces also appears to play a critical role in formation of the transition state, and chain diffusion is correlated with this dehydration process.

The unfolding of BPTI in our simulations is consistent with the reverse of the sequence of events outlined by Matthews (1991). From the native state. the protein expanded rapidly, decreasing the packing density and disrupting various tertiary interactions, which also led to increased exposure of non-polar residues. During this period of time the secondary structure remained intact. This degree of unfolding is complementary to the events occurring during the final stage of folding. From this point, more gradual unfolding of both the secondary and tertiary structure occurred, which is analogous to the middle events in folding described above. Then, a variety of related conformations, representing the molten globule state, were sampled (Daggett & Levitt, 1992*a*). From the molten globule state, BPTI unfolded further to a very dynamic collapsed conformation with little to no long-range structure (unpublished results). Therefore, our simulations do not address the earliest events in folding leading to this collapsed state. We are currently exploring further unfolding from this collapsed unfolded state to address these events. In addition, we are slowly cooling the molten globule conformation in an attempt to refold the molecule. Refolding from this state should be both more tractable and more reasonable than from some arbitrary "random coil" structure because the conformational space that needs to be sampled is drastically diminished and, as suggested by Levitt & Warshel (1975), fine structure can then direct folding once a compact structure is reached.

Nevertheless, in viewing our results from the point of view of folding, we are left with a picture of' folding beginning from a collapsed state with some amount of secondary structure. Hence. our simulations cannot, address whether the framework model (Ptitsyn & Rashin, 1975; Kim & Baldwin, 1982) or the hydrophobic collapse model (Lesk & Rose. 1981: reviewed by Dill, 1990) is ultimately responsible for arrival at this state. Probably both are. But, from, and within. this state we observed propagation of the secondary structure. These segments of secondary structure can then diffuse together to form the final folded state, which occurs concomitantly with repacking and fine-tuning of the hydrophobic core. This last sequence of events is consistent with the framework model and the related diffusion-collision model (Karplus & Weaver, 1976: Cohen *et al.*, 1980; Bashford *et al.*, **1988).**

The transition state for folding/unfolding appears to approach the native state and contains significant' amounts of tertiary structure (Goldenberg & Creighton, 1985; Matouschek *et al.*, 1989: Sancho *et* al., 1991). Although we have not identified the actual transition state for unfolding in the simulation. at the point when the potential energy was maximal near the native state, the protein expanded, packing interactions decreased, the solvent,-accessible surface area increased, and the secondary structural units came apart slightly. This observation is consistent with experimental studies suggesting that dehydration of the hydrophobic residues is important in forming the transition state during folding (or hydration for unfolding; Hurle *et al.*, 1987), and that rearrangement /repacking of secondary and tertiary structure in the hydrophobic core occurs (Perry *et al.*, 1987), which can include diffusion of segments of secondary structure (Chrunyk & Matthews, 1990).

Weissman & Kim **(1991)** have re-examined the BPTI refolding pathway and found that the dominant intermediates are native-like, which is at odds with studies by Creighton **(1978, 1988). While** we have not simulated refolding, if folding were to occur from the molten globule state as the reverse of the unfolding steps simulated, our work suggests that native disulfide-bonded intermediates would be preferred, or at least that the non-native disulfide bonds involving residues **14** and 38 would be unlikely. We are currently investigating the potential importance and likelihood of non-native disulfide bonds in the MG→U process that, would act much earlier during folding. Weissman & Kim **(1991)** state that it is likely that the interactions stabilizing the native state are also in effect during protein folding. Our simulations show this directly. We found that many native interactions persisted during unfolding and in the partially unfolded state. Non-native interactions were also prevalent during unfolding and in stabilizing the resulting state; however, these non-native interactions were the same type of interactions that are observed in the native conformations of proteins.

We appreciate the financial support for this work provided by the National Institutes of Health (GM-41455 to M.L.) and the Jane Coffin Childs Memorial Fund for Medical Research (to V.D.). We thank David Hinds for the use of his contact map display program, in addition to solving many technical difficulties along the way. Figs 1. 4, 5 and 7 were made using UCSF MidasPlus (U.C.S.F.. San Francisco, CA; Ferrin *et al.*, **1988)** and Fig. **11** was made with MacImdad (M.A.G.. Stanford, CA: Levitt. **1990).**

References

Amir, D. & Haas, E. **(1988).** Reduced bovine pancreatic trypsin inhibitor has a compact structure. *Biochemistry,* **27**, *8889-8893.*

Badger, J. **(1992).** Flexibility in crystalline insulins. *Biophys. J.* **61**, **816-819.**

Baker, D.. Sohl, J. L. & Agard, D. (1992). A protein-folding reaction under kinetic control. *Nature (London,). 356. 263-265.*

Baldwin, R. L. & Eisenberg, D. **(1987).** Protein stability. In *Protein Engineering* (Oxender. D. L. & Fox, C. F., eds), pp. 127-148. Alan R. Liss. New York.

Bashford, D.. Cohen, F. E.. Karplus, M., Kuntz, E. D. & Weaver. D. L. **(1988).** Diffusion-collision model for the folding kinetics of' myoglobin. *Proteins, 4,* **21 l-227.**

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. **(1989).** Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig a-la&albumin. *Biochemistry, 28,* **7-13.**

Bycroft, M.. Matouschek, A., Kellis, J. T., Jr, Serrano, L. & Fersht, A. R. **(1990).** Detection and characterization of a folding intermediate in barnase by NMR. *Nature (London).* **346**. *488490.*

Caspar, D. L. D. & Badger. J. **(1991).** Plasticity of crystalline proteins. *Curr. Opin. Struct. Biol.* **1**, **877-882.**

Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advan. Enzymol. 47,* **45-61.**

Chrunyk, B. A. & Matthews, C. R. **(1990).** Role of diffusion in the folding of the a subunit of tryptophan synthase from *Escherichia coli. Biochemistry, 2 9 ,* **2149-2154.**

Cohen, F. E.. Sternberg, M. J. E., Phillips, D. C., Kuntz, I. D. & Kollman. I'. A. **(1980).** A diffusion-collision-adhesion model for the kinetics of myoglobin refolding. *Nature (London), 286,* **632-634.**

Covell, D. G. & Jernigan, R. L. **(1990).** Conformations of folded proteins in restricted places. *Biochemistry, 29, 3287-3294.*

Creighton, T. E. **(1978).** Experimental studies of protein folding and unfolding. *Progr. Biophys. Mol. Biol. 33,* **231-297.**

Creighton, T. E. **(1988).** Toward a better understanding of protein folding pathways. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5082–5086.

Daggett, V. **(1987).** Protein degradation: the role of mixed-function oxidases. *Pharmacol. Res.* **4**, *278-284.*

Daggett, V. **(1993).** A model of the molten globule state of CTF generated using molecular dynamics. I n *Techniques in Protein Chemistry* (Angeletti, R. H., ed.), vol. 4. in the press.

Daggett, V. & Levitt, M. **(1991).** A molecular dynamics simulation of the C-terminal fragment of the L7/L12

ribosomal protein in solution. *Chem. Phys.* **158,** 501–512.

Daggett. V. & Levitt. M. (1992*a*). A model for the molten globule state from molecular dynamics simulations. *Proc. Nat. Acad. Sci., U.S.A. 89,* **5142-5146.**

Dagget t . V. & Levitt. M. **(1992b).** Molecular dynamics simulations of helix denaturation. *J. Mol. Biol. 223.* 1121–1138.

Dagget t . V. & Levitt. M. (1993). Realistic simulations of native protein dynamics in solution and beyond. *Annu. Rev. Biophys. Biomol. Struct. 22.* 353–380.

Dagget t . V., Kollman. I? A. & Kuntz. I. **D.** (1991*a*). Molecular dynamics simulations of small peptides: dependence on dielectric model and pH. *Biopolymers,* **3 1,** 285–304.

Daggett. V., Kollman, P. A. & Kuntz, I. D. (1991*b*). **A** molecular dynamics simulation of polyalanine: an analysis of' equilibrium motions and helix-coil transitions. *Biopolymers,* **31, 1115-l 134.**

Deisenhofer. J. & Steigemann. W . (1975). Crystallographic refinetnent of the structure of bovine pancreatic trypsin inhibitor at 1·5 **A** resolution. *Acta. Crystallogr. sect. B. 3* **1. 238-250.**

DiCapua. F. M., Swaininathan. S. & Beveridge. D. L (1990). Theoretical evidence for destabilization of an a-helix by water insertion: molecular dynamics of hydrated decaalanine. *J. Amer. Chem. Soc.* **112,** 6768–6771.

Dill. K. (1990). Dominant forces in protein folding. *Biochemistry. 29,* **7133-7155.**

Dill. K. A. & Shortle, D. **(1991).** Denatured states of proteins. *Annu. Rev. Biochem. 60.* 795–825.

Dobson. C. M. (1991). Characterization of protein folding intermediates. *Curr. Opin. Struct. Biol.* **1, 22-27.**

Dobson. C. M. (1992). Unfolded proteins. compact states and molten globules. *Curr. Opin. Struct. Biol. 2,* **6-12.**

Dyson. J. H., Rance. M., Houghten, R. A., Lerner. R. A. & Wright. P. E. (1988). Folding of immunogenic pept ide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* 201. 161–200.

Evans. P. A., Topping. K. D., Woolfson. D. N. & Dobson. C. M. (199 1). Hydrophobic clustering nonnative states of a protein: interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins, 9. 248-266.*

Ferrin. T. E., Huang. C. C., Jarvis. L. E. & Langridge. R. **(1988). The** Midas display system. *J. Mol. Graph. 6.* 13–27.

Fersht, A. R., Kellis. J. T., Jr. Matouschek. A. T. E. L. & Serrano. L. (1990). Protein folding pathway. Reply to a letter. *Nature (London). 343,* 602.

Garvey. E. P., Swank, J. & Matthews. C. R. **(1989).** A hydrophobic cluster forms early in the folding of dvhydrofolate reductase. *Proteins: 6. 259-266.*

Goldberg. M. E. **(1985).** The second translation of the genetic message: protein folding and assem bly. *Trends. Biochem. Sci.* **10. 388-391.**

Goldenberg. D. P. (1988). Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. *Biochemistry. 27.* **248 l-2489.**

Golden berg. D. P. & Creighton. T. E. **(1985).** Energetics of protein structure and folding. *Biopolymers, 24.* 167–182.

Goodman, E. M. & Kim, P. S. (1989). Folding of a peptide corresponding to the a-helix in bovine pancreatic trypsin inhibitor. *Biochemistry, 28, 43434347.*

Goodman. E. M. & Kim, P. S. **(1990).** Tests of isolated b-sheet formation in a cyclic peptide from BPTI. *Curr. Res. Protein Chem.* Academic Press, Inc. 301-308.

Harding, M. M., Williams, D. H. & Woolfson, D. N. (199 1) . Characterization of a partially denatured state of a protein by two-dimensional NMR: reduction of the hydrophobic interactions in ubiquitin. *Biochemistry,* **31, 3120-3128.**

Harte. W. E., Swaminathan, S., Mansuri. M. M., Martin. J. C., Rosenberg, I. E. & Beveridge. D. L. (19190). Domain communication in the dynamical struc ture of human immunodeficiency virus **1** protease. *Proc. Nat. Acad. Sci., U.S.A. 87, 8864-8868.*

Heringa, J. & Argos, P. **(1991).** Side-chain clusters in protein structures and their role in protein folding. *J. Mol. Biol. 220,* **151-171.**

Hinds. D. & Levitt, M. **(1992).** A lattice model for protein structure prediction at low resolution. *Proc. Nat. Acad. Sci., U.S.A. 89. 2536-2540.*

Hughson. F., Wright, P. & Baldwin. R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science, 249,* 1544–1548.

Hurle. M. R., Michelotti, G. A., Crisanti. M. M. & Matthews, C. R. **(1987).** Characterization of a slow folding reaction for the a-subunit of tryptophan synthase. *Proteins. 2, 54-63.*

Jurka, J. &. Smith, T. F. **(1987).** β Turns in early evolution: chirality, genetic code. and biosynthetic pathwavs. *Cold Spring Harbor Symp. Quant. Biol. 52.* **407410.**

Kabsch. W. **(1976).** A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. sect. A.* **32, 922-923.**

Karplus, M. & Weaver, D. L. **(1976).** Protein-fold.ing dynamics. *Nature (London). 260, 404-406.*

Kim, P. S. & Baldwin, R. L. **(1982).** Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51. 459489.**

Kim. P. S. & Baldwin. R. L. **(1990).** Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem. 59,* 631–660.

Kossiakoff. A. A., Randal. M., Guenot. J. & Eigenbrot, C. (**1992).** Variability of conformations at crystal contacts in BPTI represent true low energy structure-correspondence among lattice packing and molecular dynamics structures. *Proteins Struct . E'unct. Genet.* **14. 65-74.**

Kuntz, I. D. **(1972).** Protein folding. *J. Amer. Chum. Soc. 94,* **4009-4012.**

Kuwajima. K. **(1989).** The molten globules state as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins, 6,* **87-103** and references therein.

Lesk, A. M. & Rose, G. D. **(1981).** Folding units in globular proteins. *Proc. Nat. Acad. Sci., U.S.A. '78.* **43044308.**

Levitt, M. (**1989).** Molecular dynamics of macromolecules in water. *Chem. Scripta,* **29A. 197-203.**

Levitt, M. **(1990).** MacImdad, Molecular Applications Group. Stanford, CA, Version 3.

Levitt, M. & Chothia, C. **(1976).** Structural patterns in globular proteins. *Nature (London).* **261. 552-558.**

Levitt, M. & Sharon, R. **(1988).** Accurate simulation of proteins in solution. *Proc. Nat. Acad. Sci., U.S.A. 135.* 7557-7561.

Levitt, M. & Warshel, A. **(1975).** Computer simulation of protein folding. *Nature (London). 253, 694-698.*

Levitt, M., Sander, C. & Stern, P. S. (1985). Protein

normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181**, 423–447.

Lewis, P. N., Momany, R. A. & Scheraga, H. A. (1971). Folding of polypeptide chains in proteins: a proposed mechanism of folding. *Proc. Nat. Acad. Sci., U.S.A.* **68**, 2293–2297.

Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D. & Anderson, S. (1987). Mutants of bovine pancreatic trypsin inhibitor cysteines 14 and 38 can fold properly. *Science*, **235**, 1370–1373.

Matouschek, A., Kellis, J. T., Jr, Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature (London)*, **340**, 122–126.

Matthews, C. R. (1991). The mechanism of protein folding. *Curr. Opin. Struct. Biol.* **1**, 28–35.

Matthews, C. R., Crisanti, M. M., Manz, J. T. & Gepner, G. L. (1983). Effect of a single amino acid substitution on the folding of the α subunit of tryptophan synthase. *Biochemistry*, **22**, 1445–1452.

Meirovitch, H. & Scheraga, H. A. (1981). Introduction of short-range restrictions in a protein-folding algorithm involving a long-range geometrical restriction and short-, medium-, and long-range interactions. *Proc. Nat. Acad. Sci., U.S.A.* **78**, 6584–6587.

Moult, J. & Unger, R. (1991). An analysis of protein folding pathways. *Biochemistry*, **30**, 3816–3824.

Oas, T. G. & Kim, P. S. (1988). A peptide model of a protein folding intermediate. *Nature (London)*, **336**, 42–48.

Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J. T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E. & Kraut, J. (1987). Effect of single amino acid replacements on the folding and stability of dihydrofolate reductase from *Escherichia coli*. *Biochemistry*, **26**, 2674–2682.

Presta, L. G. & Rose, G. D. (1988). Helix signals in proteins. *Science*, **240**, 1632–1641.

Ptitsyn, O. B. (1987). Protein folding: hypotheses and experiments. *J. Protein Chem.* **6**, 273–293 and references therein.

Ptitsyn, O. B. & Rashin, A. A. (1975). A model of myoglobin self-organization. *Biophys. Chem.* **3**, 1–20.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters*, **262**, 20–24.

Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1–14.

Richardson, J. S. & Richardson, D. C. (1989). The de novo design of protein structures. *Trends Biochem. Sci.* **14**, 304–309.

Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelliny and proton NMR. *Nature (London)*, **335**, 700–704.

Sancho, J., Meiering, E. M. & Fersht, A. R. (1991). Mapping transition states of protein unfolding by protein engineering of lipand-binding sites. *J. Mol. Biol.* **221**, 1007–1014.

Serrano, L., Bycroft, M. & Fersht, A. R. (1991). Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles. *J. Mol. Biol.* **218**, 465–475.

Skolnick, J. & Kolinski, A. (1991). Simulations of the folding of a globular protein. *Science*, **250**, 1121–1125.

Soman, K. V., Karimi, A. & Case, D. A. (1991). Unfolding of an α-helix in water. *Biopolymers*, **31**, 1351–1361.

Staley, J. P. & Kim, P. S. (1990). Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. *Nature (London)*, **344**, 685–688.

States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. (1987). Conformations of intermediates in the folding of the pancreatic trypsin inhibitor. *J. Mol. Biol.* **195**, 731–739.

Tilton, R. F., Jr, Singh, U. C., Kuntz, I. D., Jr & Kollman, P. A. (1988). Protein-ligand dynamics: a 96 picosecond simulation of a myoglobin-xenon complex. *J. Mol. Biol.* **199**, 195–211.

Tirado-Rives, J. & Jorgensen, W. L. (1991). Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. *Biochemistry*, **30**, 3864–3871.

Udgaondar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature (London)*, **335**, 694–699.

Weissman, J. S. & Kim, P. S. (1991). Reexamination of the folding of RPTT: predominance of native intermediates. *Science*, **253**, 1386–1393.

Edited by B. Honig