Molecular Dynamics Simulations of the Unfolding of Apomyoglobin in Water†

Julian Tirado-Rives* and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Received November 23, 1992; Revised Manuscript Received February 16, 1993

ABSTRACT: Molecular dynamics simulations of apomyoglobin have been conducted in aqueous solution for 350 ps at 25 °C and for 500 ps in two different runs at 85 °C. The structures obtained at the higher temperature display properties similar to those of molten globules. Close agreement is obtained between the computed structural models and experimental data on the helical content of both native apomyoglobin and the low-pH unfolding intermediate. The results also suggest explanations for the surprising observations on the effects of mutations at the interface of the A, G, and H helices. Detailed analyses of the final structures and the unfolding pathways at high temperature clearly show that the most stable α-helical regions are those in contact with other helices.

It is now accepted that globular, single-domain proteins typically encode in their amino acid sequence enough information to fold in vitro into a specific, native, three-dimensional structure (Baldwin, 1989; Jaenicke, 1987). The details and mechanism of this process are, however, not fully understood, and the many attempts to elucidate it have given rise to the so-called "protein folding" problem (King, 1989). Experimental studies in this area have been challenging since most folding/unfolding transitions are highly cooperative and often fit a simple two-state model in which only the native and unfolded forms are populated, and the actual existence of intermediates had until recently only been implied (Tanford, 1968; Privalov, 1979).

However, equilibrium intermediates have now been detected for several proteins under mild unfolding conditions, and although their complete three-dimensional conformations are not known, some of their physical properties have been measured and earned them the name of "molten globules" (Ptitsyn, 1987; Kuwajima, 1989). These intermediates are typically characterized by a relatively compact structure when compared to fully unfolded proteins but with higher side-chain conformational mobilities and lower contents of secondary and tertiary structure than the native, folded forms (Baldwin, 1991; Christensen & Pain, 1991). They have also been found in kinetic folding pathways through fast CD measurements and NMR experiments (Kim & Baldwin, 1990). In particular, experiments utilizing hydrogen isotopic exchange in combination with 2-D NMR have located sites protected from exchange at different times during refolding, presumably due to formation of secondary structure or to inaccessibility to the solvent. Some sites showing protection match the locations of α-helices in the native forms of several proteins, including apomyoglobin (Hughson et al., 1990), ribonuclease A (Udgaonkar & Baldwin, 1988), cytochrome *c* (Roder et al., 1988), α-lactalbumin (Baum et al., 1989), barnase (Bycroft et al., 1990; Matouschek et al., 1992), and lysozyme (Miranker et al., 1991; Radford et al., 1992; Lu & Dahlquist, 1992).

The theoretical study of protein folding and unfolding could also be valuable in providing insights on several key issues including the structures of folding intermediates, factors affecting the stability of the folded and unfolded forms, and the pathways followed during their interconversion. However, the limited work to date has mostly utilized simplified representations of the solute and the solvent to reduce computational demands. For example, lattice simulations feature a small number of interaction sites per residue and, additionally, restrict these sites to occupy discrete positions at the vertices of a regular lattice. With such models and avoidance of explicit treatment of the solvent, simulations of processes during long time spans are possible; and, in some simulations of protein folding, transient states displaying properties similar to those of molten globules have been observed [see, for instance, Chan and Dill (1989) and Skolnick and Kolinski (1990)].

Applications of molecular dynamics to the folding/unfolding problem using full atomic representations of both solute and solvent have mostly concentrated on smaller peptides. The simulations of an analog of S-peptide in water reproduced the experimentally observed stability of the helix at 278 K and unraveling at 358 K (Tirado-Rives & Jorgensen, 1991). Other similar studies included the formation and destruction of reverse turns at 300 K for the Tyr-Pro-Gly-Asp-Val pentapeptide (Tobias et al., 1991), the unfolding at 300 K of an 18-residue peptide corresponding to the N-terminus of the H helix of myoglobin (Soman et al., 1991), and a set of simulations at different temperatures of a capped tridecaalanine helix (Daggett & Levitt, 1992a). Molecular dynamics combined with statistical perturbation theory has been used to determine the free energy changes for folding and unfolding capped alanine and valine tripeptides (Tobias & Brooks, 1991) and for the folding of helical proline-substituted alanine polypeptides (Yun et al., 1991).

It is only in the last year that molecular dynamics studies of unfolding of complete proteins in water have appeared in the literature. Daggett and Levitt reported simulations of the native and reduced forms of BPTI at 298 and 423 K for 550 ps and of reduced BPTI at 498 K for 284 ps, at which time the trajectory became unstable (Daggett & Levitt, 1992b), while Mark and van Gunsteren chose to simulate egg white lysozyme at 300 K for 550 ps and at 500 K for 190 ps (Mark & van Gunsteren, 1992). Both of these studies utilized constant volume simulations at highly elevated temperatures as a means to decrease the computation time needed to observe the transitions. The final structure of the lysozyme simulation did not contain any appreciable nativelike secondary structure, while most of it was conserved for BPTI. Direct comparisons with experimental unfolding data were limited in these studies,

† Supported by NIH Grant GM 32136.

0006-2960/93/0432-4175$04.00/0 © 1993 American Chemical Society

Table I: Comparison of Different Properties of Apomyoglobin for the Different Simulations and the Crystal Structure

property	simulation 6C	simulation 6H	simulation 4H	crystal structure
temperature (K)	298	358	358	260
total time (ps)	350	500	500	
pH	6.0	6.0	4.2	6.0
		Properties of the Final Instantaneous Structure		
backbone rms deviation (Å)	5.41	6.19	9.12	
side-chain contacts	475	463	439	538
		Average Structure over the Last 50 ps		
backbone rms deviation (Å)	5.21	5.59	8.47	
radius of gyration (Å)	16.468	16.506	16.774	15.231
solvent accessible area (Å^2)	9003	9536	9858	7426
helical residues	78	57	54	119
% helicity	51	37	35	78
		Hydrogen Bonds during the Last 50 ps		
hydrogen bonds to solvent	485 ± 13	479 ± 11	501 ± 11	
total intramolecular	137 ± 6	126 ± 6	118 ± 8	178
backbone–backbone	100 ± 4	83 ± 4	83 ± 4	133
backbone–side chain	23 ± 3	29 ± 3	21 ± 4	25
side chain–side chain	14 ± 3	15 ± 3	14 ± 4	20
α-helical	54 ± 4	31 ± 4	40 ± 3	93
3_{10}-helical	32 ± 4	30 ± 3	25 ± 3	35
π-helical	4 ± 1	5 ± 1	4 ± 1	2
		Average Properties during the Last 50 ps		
density (g·cm^{-3})	1.038 ± 0.004	0.982 ± 0.004	0.983 ± 0.005	
helical residues	74 ± 5	50 ± 5	49 ± 6	119
% helicity	49 ± 3	33 ± 3	32 ± 4	78
backbone rms deviation (Å)	5.23 ± 0.16	5.66 ± 0.17	8.59 ± 0.36	
C$^\alpha$ fluctuations (Å)	0.9 ± 0.3	1.1 ± 0.3	1.2 ± 0.5	0.6 ± 0.2

and the employed temperatures near 225 °C may lead to unfolding pathways that are different than would be observed under experimentally more realistic conditions.

The present simulations were undertaken to study the thermally induced unfolding of apomyoglobin using full atomic representations for both the protein and the solvent water via molecular dynamics. The direction of unfolding was again chosen for study in order to take advantage of the fact that the structures of folded proteins are much better defined than those of their unfolded counterparts. In addition, following the unfolding process provides a convenient way to circumvent the potential difficulties caused by the likely existence of multiple minima in the folding pathway. Importantly, for apomyoglobin there are substantial experimental data available for comparison, and less elevated temperatures could be used in the simulations.

Apomyoglobin is produced from myoglobin by removal of the heme group. Though its crystal structure has not been reported, the crystal structure of myoglobin, shown in Figure 1, has been determined. It features eight distinct α-helices named A through H, no other secondary structure, and no disulfide bonds (Kuriyan et al., 1986). It is experimentally known that removal of the heme has a profound effect on the stability of the protein such that its melting temperature decreases from 353 to 334 K (Privalov et al., 1986; Griko et al., 1988), which facilitates the present simulations of the apoprotein under realistic conditions. NMR data on native apomyoglobin indicate that the hydrophobic core (Cocco & Lecomte, 1990; Cocco et al., 1992) and the A, B, E, G, and H helices (Hughson et al., 1990) retain characteristics similar to those in the holoprotein. These helices are also reported as stable in a previous molecular dynamics study of apomyoglobin in water at 312 K (Brooks, 1992). Apomyoglobin has lower helicity than the holoprotein (Harrison & Blout, 1965; Breslow et al., 1965), and it yields a stable intermediate on acid unfolding (Griko et al., 1988). Further CD studies of the latter process have shown that this intermediate (I form) has lower helical content, 35%, than the ca. 55% estimated for native apomyoglobin at neutral pH and the 78% for

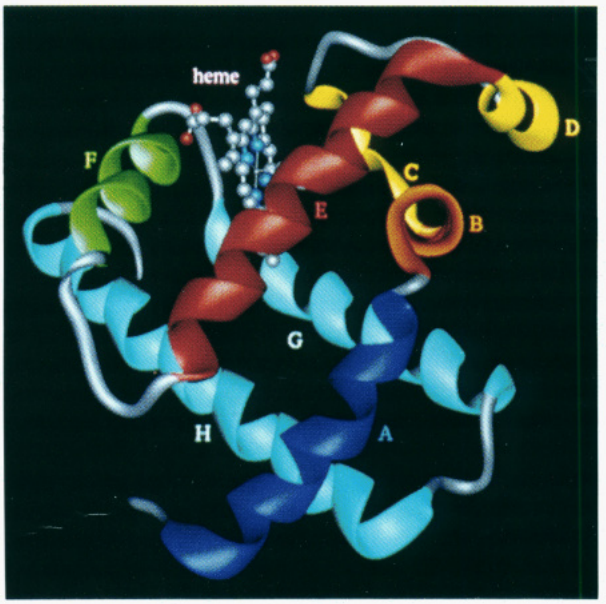

FIGURE 1: Ribbon representation of the crystal structure of myoglobin showing the eight α-helices and the position of the heme.

myoglobin (Hughson et al., 1990). There has also been some further characterization of the I state. Several sections of nativelike helices have been detected through the application of the isotopic exchange/2-D NMR technique (Hughson et al., 1990), and mutant forms of the protein have been produced to probe specific interactions (Hughson et al., 1991). These two studies have left uncertainty about the extent of nativelike interhelical interactions in the I state. The present results address this discrepancy by providing detailed models for the structures of native apomyoglobin and the I form that are also consistent with the CD data.

COMPUTATIONAL METHODS

All the simulations were conducted using the AMBER 3.0a program (Singh et al., 1986) on SGI-4D/35 computers in

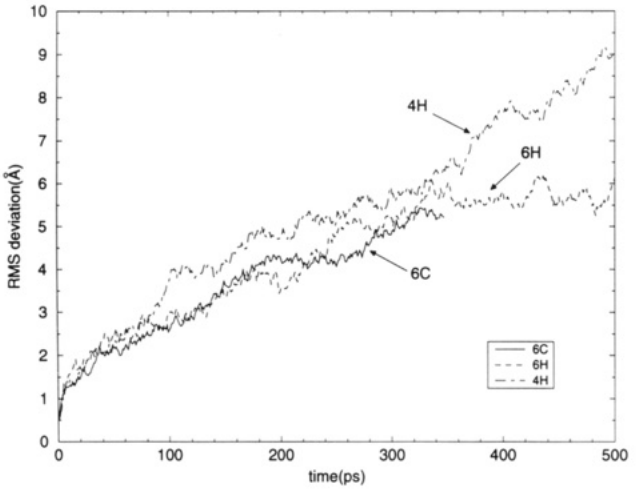

FIGURE 2: Plots of the rms deviations for the main-chain heavy atoms from the three simulations of apomyoglobin as a function of time relative to the crystal structure of myoglobin.

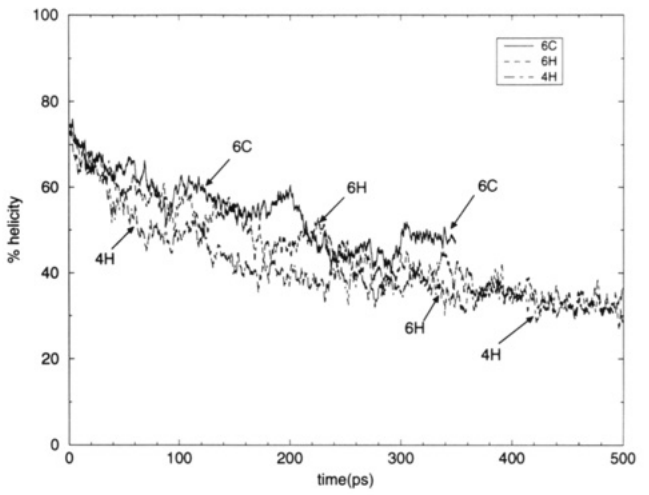

FIGURE 3: Helical content of apomyoglobin from the three simulations as a function of time.

these laboratories. The OPLS nonbonded parameters (Jorgensen & Tirado-Rives, 1988), including the all-atom parameters recently developed for aromatic rings (Jorgensen & Severance, 1990), were used for the protein atoms in conjunction with the TIP3P model for water (Jorgensen et al., 1983). As specified in the OPLS model, the dielectric constant was fixed at 1.0, and the scaling factors for the 1,4 nonbonded interactions were 8.0 for the Lennard-Jones and 2.0 for the electrostatic interactions. The energetics for angle bending and torsional motion were described with the AMBER force field (Weiner et al., 1984). During the simulations all bond lengths and the H–H distances in water were fixed at their equilibrium values using the SHAKE algorithm (Ryckaert et al., 1977) with a tolerance of 0.0004 Å, which allowed a time step of 2 fs. A nonbonded pair list was used to accelerate the calculations and was updated every 10 steps. This list was generated using a 9-Å residue-based cutoff to avoid splitting dipoles. The solution calculations utilized periodic boundary conditions to avoid edge effects. The pressure was fixed at 1 bar (0.987 atm) in all NPT dynamics simulations.

The initial coordinates were obtained from the X-ray structure of sperm whale myoglobin (Kuriyan et al., 1986) as deposited in the Brookhaven Protein Data Bank, entry 1MBC (Bernstein et al., 1977). The coordinates of the heme, the iron atom, crystallographic waters, and a phosphate were deleted, and two different protonation states of apomyoglobin were generated, depending on the pH. Although the precise

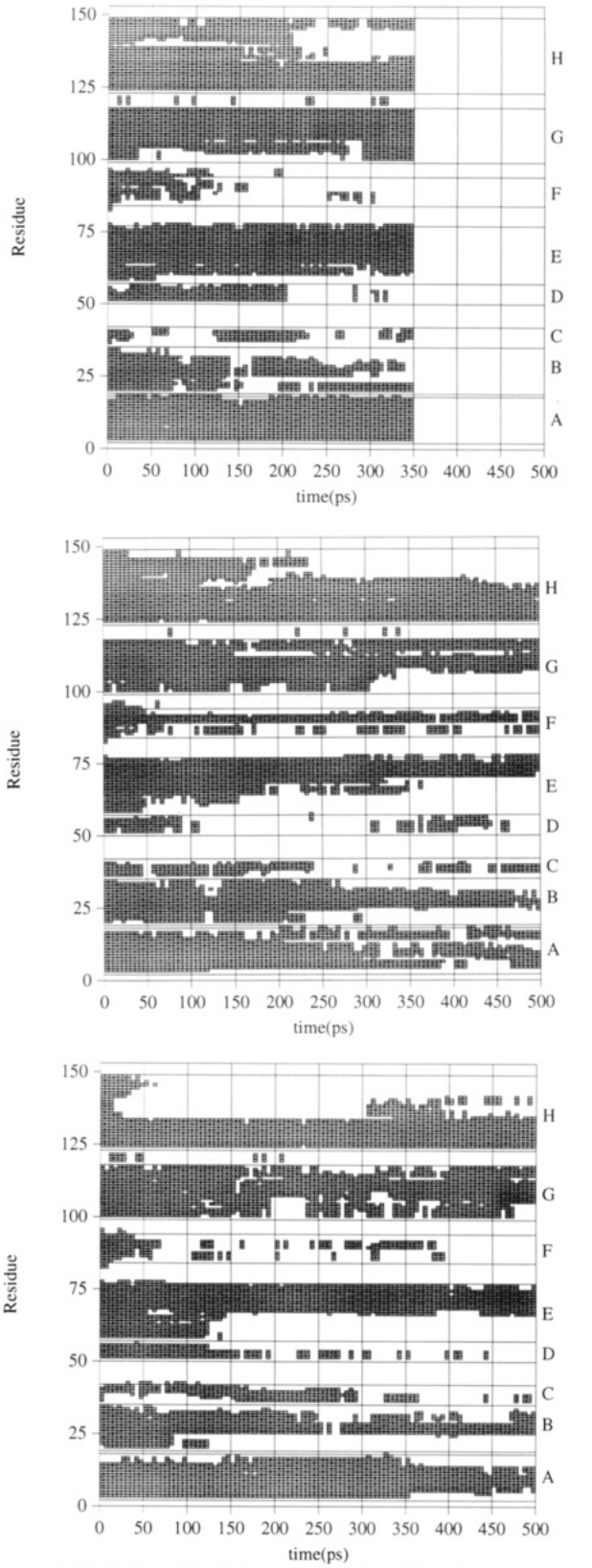

FIGURE 4: Evolution of the helices as a function of time for simulations 6C (top), 6H (middle), and 4H (bottom).

protonation states of all histidine side chains in apomyoglobin at pH 6.0 are not known, there is an uptake of two protons on titration down to pH 4 (R. Baldwin, personal communication). Thus, at pH 4.2 all histidines were protonated, while at pH 6.0 two histidines, His-24 and His-119, with low pK_a's were kept neutral (Cocco et al., 1992). The tautomeric states of these were chosen so as to preserve the hydrogen bond between them. All polar and aromatic hydrogen atoms were then added, and both structures were minimized in vacuo

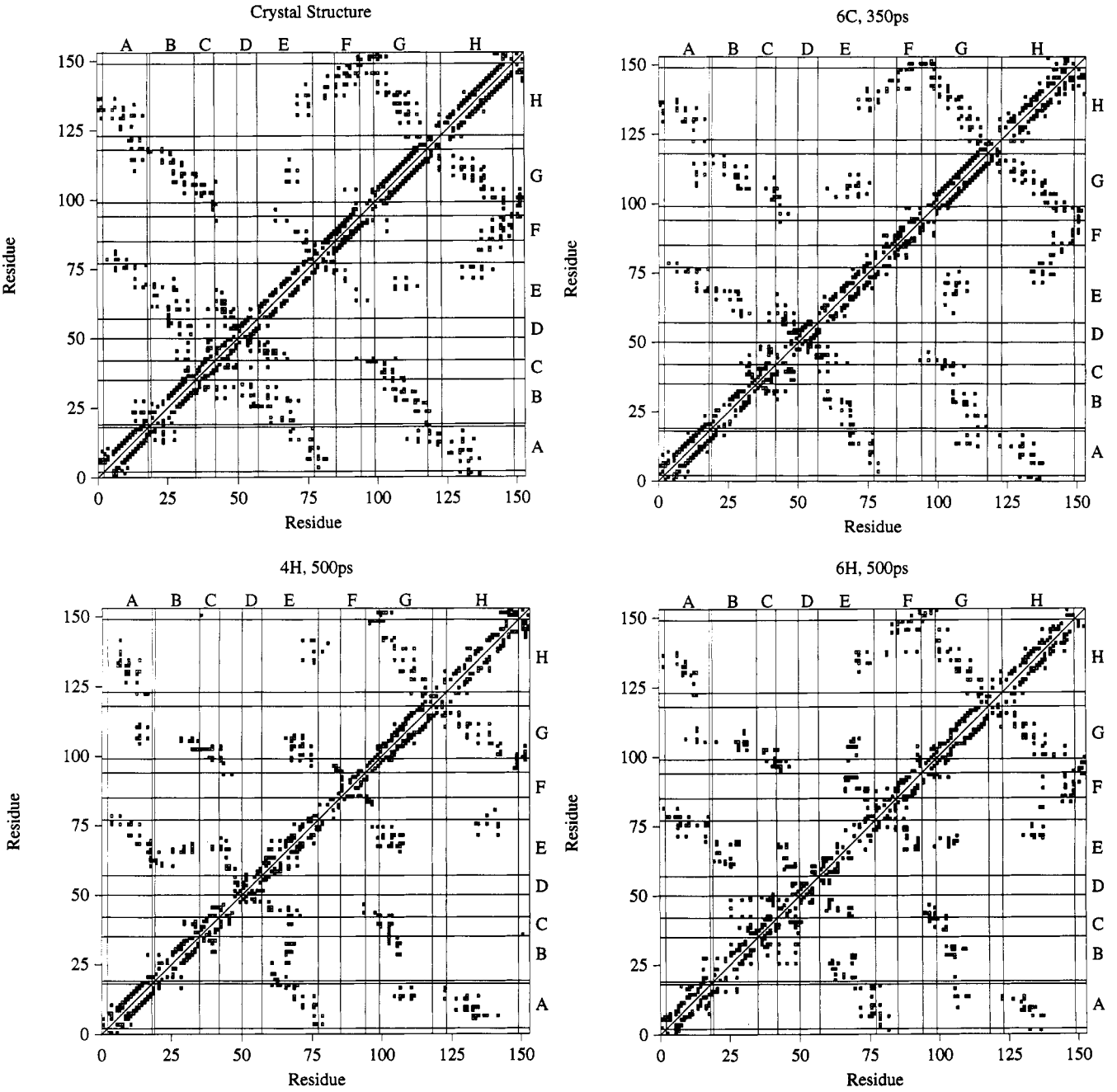

FIGURE 5: Side-chain contact maps for the crystal structure (top left) and the final structures from simulations 6C (top right), 4H (bottom left), and 6H (bottom right).

using a distance-dependent dielectric constant of 8.0 R. The final rms deviations from the myoglobin structure for the backbone atoms were 0.363 and 0.368 Å for the pH 6.0 and pH 4.2 structures, respectively.

Since apomyoglobin is highly charged under these conditions (+12 at pH 6.0 and +14 at pH 4.2), the corresponding numbers of chloride counterions were needed. In order to locate these in a fairly objective fashion, both structures were centered in rectangular boxes of water obtained by periodic translations in the x, y, and z directions of an equilibrated cube of TIP3P water. Water molecules closer than 1.5 Å, or farther than 8.0 Å in the x and y directions or 10.5 Å in the z direction from the closest protein atom, were then deleted to give initial systems containing the solute plus 5332 water molecules in a rectangular box of dimensions 62.3 × 57.8 × 50.6 Å.

The systems were then carefully prepared by running initially 100 cycles of steepest descent energy minimization at constant volume followed by 3 ps of constant volume molecular dynamics, during which time the temperature was gradually increased from an initial 100 K to 278 K. The resulting structures were equilibrated at 278 K for an additional 2 ps of NVT-MD. During these calculations only the water was allowed to move. At this stage the needed counterions were added by first selecting the 30–40 water molecules which had the lowest electrostatic energies and from that subset graphically choosing the 12 or 14 molecules that were closest to the surface of the protein and were not clustered together. The coordinates of the oxygen atoms were then used for the chloride counterions. The neutral systems were resolvated and equilibrated up to a temperature of 298 K in a manner similar to that described above and followed by a 1-ps NVT-MD run in which the protein atoms were allowed to move as well. The final boxes obtained in this fashion measured 62.3 × 57.9 × 52.9 Å and contained 5345 and 5348 water molecules at pH 6.0 and 4.2, respectively.

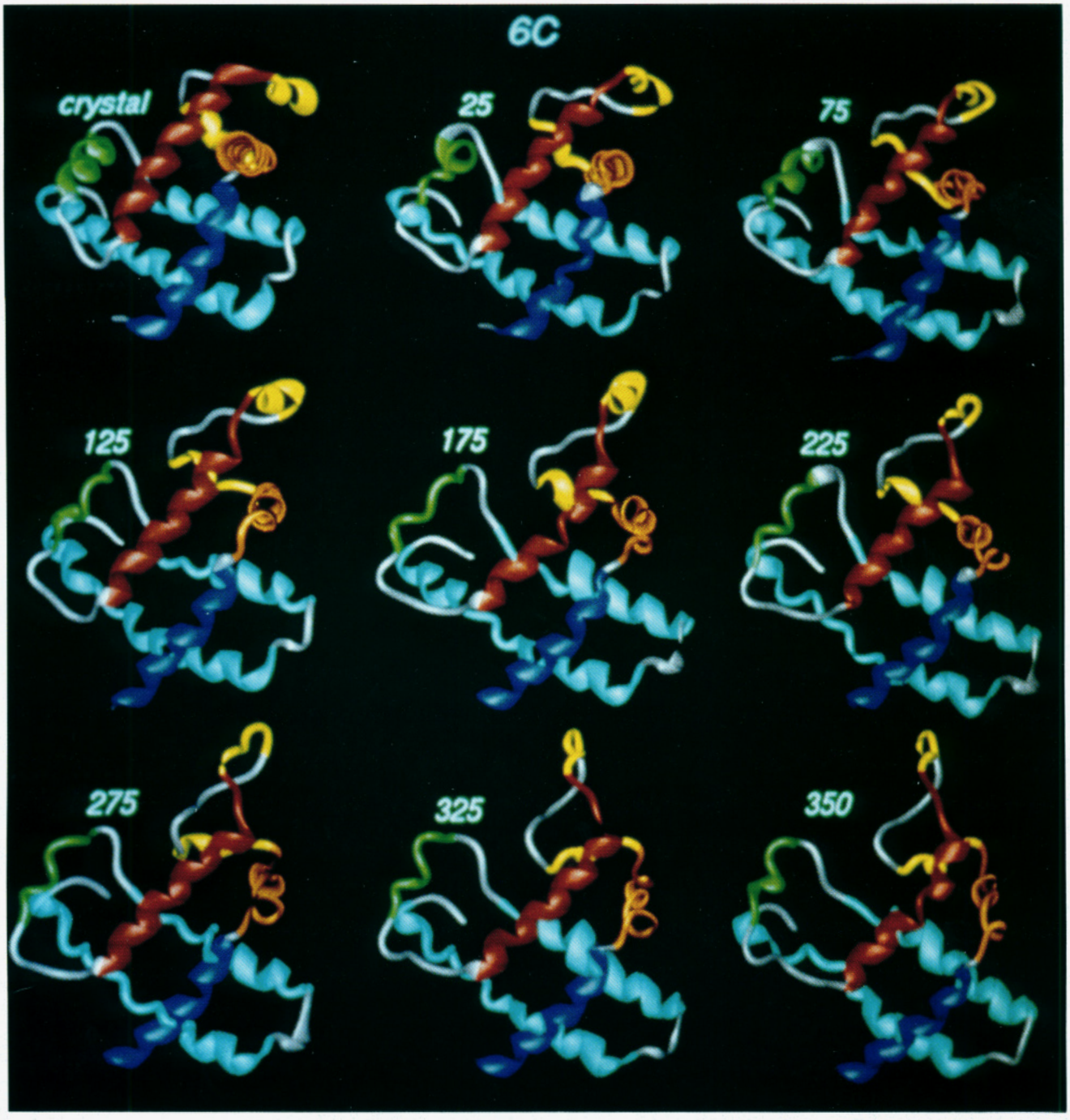

FIGURE 6: Ribbon representations of the crystal structure of myoglobin less the heme and of snapshots along simulation 6C at 298 K.

Three different NPT-MD simulations were started from the equilibrated systems, a single run (abbreviated 4H) at pH 4.2 was started at 100 K and heated to 358 K (85 °C) in 4 ps and then continued at that temperature until 500 ps. Two calculations were carried out at pH 6.0, a run at 358 K (abbreviated 6H) for 500 ps and a control run at room temperature (abbreviated 6C) at 298 K for 350 ps.

The values of the potential energy, rms deviation from the initial structure, and volume were monitored continuously in order to assess the evolution of the system. The coordinates, velocities, and energies were saved every 50 time steps (0.1 ps) for further analysis. In addition, due to the change in shape of the protein during its unfolding, it was necessary to check the systems for periodic distances becoming too small, i.e., the shortest separation between any protein atom and the image generated by the periodic boundary conditions of any other protein atom falling below a predetermined limit, 10.0 Å. When this condition was detected, the solvent molecules farthest away from the protein were reorganized by replicating enough water on both extremes of the Cartesian direction in which distances were becoming too short and by deleting the same number of water molecules on one or both of the other axes. This treatment was followed by 5 ps of NVT-MD with the protein fixed to allow reequilibration of the box edges before continuing the NPT runs. In this manner the total volume and number of molecules were conserved without allowing artifactual interactions of the protein with images of itself.

RESULTS

A brief summary of the results from the three simulations including comparisons to the crystal structure of the holoprotein is given in Table I. The properties of the average structures over the last 50 ps in the high-temperature simulations are strikingly similar to those experimentally determined for molten globules (Christensen & Pain, 1991).

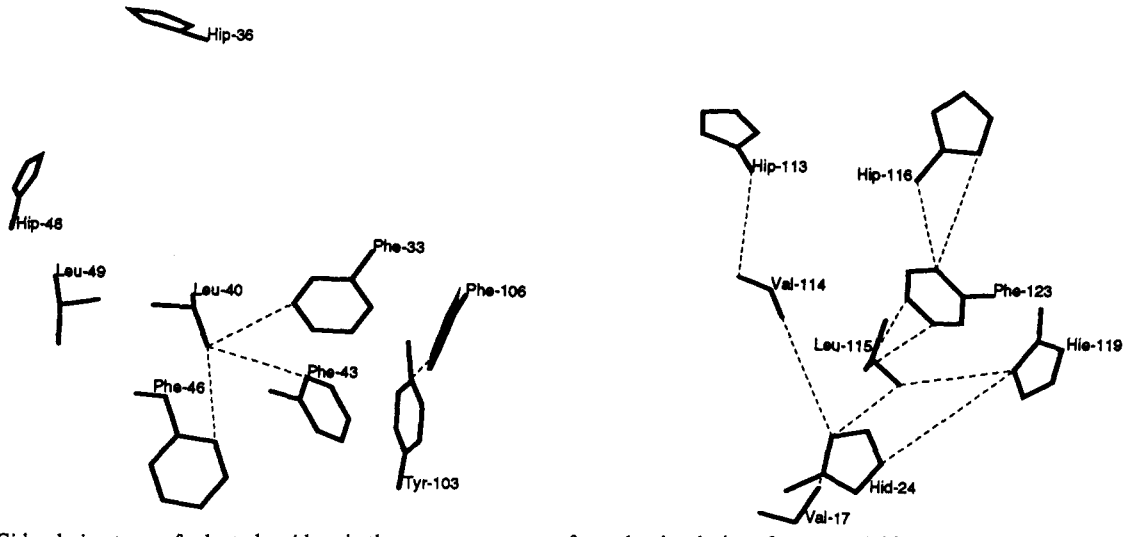

FIGURE 7: Side-chain atoms of selected residues in the average structure from the simulation of apomyoglobin at 298 K. Some distances smaller than 6 Å corresponding to experimentally observed NOE's are shown.

They are less compact than the folded proteins; their radii of gyration are 8–10% larger than for the crystal structure. They also contain less secondary structure, as evidenced by the lower helicity and number of backbone–backbone hydrogen bonds, although it must be noted that the latter are balanced by a higher number of hydrogen bonds to side-chain or solvent atoms. The structures from the high-temperature simulations also appear to display higher mobility as seen from the C^{α} fluctuations, although the additional kinetic energy available at the higher temperature makes this less clear. In view of the time course of the simulations, it is proposed that the final structures provide models for an early unfolding intermediate (or late folding intermediate) with the properties of a molten globule.

The rms deviations from the modified crystal structure for the main-chain atoms (N, C^{α}, C, and O) of the protein in the three simulations are plotted in Figure 2 as a function of simulation time. The deviations of the instantaneous structures at the end of each simulation are 5.41 Å after 350 ps for run 6C at 298 K and 6.19 and 9.12 Å after 500 ps for runs 6H and 4H at 358 K, respectively. Since the initial motions of the molecule are mostly dominated by effects of the loss of the heme, it is not surprising that the initial parts of all three simulations display similar deviations from the original structure. Nevertheless, the instantaneous structures from all three simulations at 350 ps differ significantly, their main-chain rms deviations ranging from 5.33 to 6.15 Å. Figure 3 shows the total helical content from the three simulations as a function of time. The criterion used here to define a helical residue is based on the distribution of dihedral angles in the simulations of helical polyalanine by Daggett and Levitt (1992a). It requires not only that the backbone dihedral angles of the given residue be within appropriate ranges ($-100° \leq \phi \leq -30°$ and $-80° \leq \psi \leq -5°$) but also that this requirement be met by at least three consecutive residues. For all three simulations, the helical content changes little after about 300 ps. It is interesting to notice that although both high-temperature simulations achieve a final state of similar helical content, simulation 4H, with the ionization state corresponding to the stable I form found on acid denaturation, unfolds initially faster than 6H and reaches its final value around 270 ps.

A more detailed analysis of the structural changes for the different simulations can be obtained from the representations in Figures 4 and 5. The local secondary structure is represented in Figure 4 by following the helical nature of each residue in 5-ps slices during the simulations. In these graphs each residue is represented by a filled block if it is helical in at least 50% of the structures saved during each 5-ps time slice and left blank otherwise. Figure 5 contains the side-chain contact maps for the crystal and the final structures from the three simulations. In building these contact maps, two nonconsecutive residues with any side-chain atoms within 7 Å of each other are connected by a filled block.

DISCUSSION

The final structure of simulation 6C, which corresponds to the native form of apomyoglobin at room temperature, still contains much of the secondary structure observed in the crystal structure of the holoprotein. Most notably, large segments of the A, E, G, and H helices and one or two turns of the B helix are quite stable (Figure 4). These helices were also found to be the most stable in a previous molecular dynamics simulation at 312 K (Brooks, 1992). In addition, a single turn of the C helix fluctuates between the folded and unfolded forms. There is complete decay of the short D and F helices and a break toward the C-terminal end of helix H. The average helical content over the last 50 ps of the simulation is $49 \pm 3\%$, which agrees closely with the 55% value determined by CD (Hughson et al., 1990). Several representative instantaneous structures from the simulation are shown in ribbon form in Figure 6, as generated by the Midas Plus program (Ferrin et al., 1988). These results help clarify the findings from the hydrogen-exchange experiments (Hughson et al., 1990). The slowly exchangeable protons identified for the holoprotein were also found largely to be protected in apomyoglobin at pH 6. This led to a dilemma in explaining the lower helicity from the CD data. However, probes could not be studied in the D and F helices, which are indeed unfolded in the simulation. This coupled with some fraying at other helical ends, which is also apparent in diminished protection factors, accounts for the reduced helicity of the native apoprotein.

Comparison of the contact maps (Figure 5) of the final structure of simulation 6C and of the crystal structure reveals that the major changes in the tertiary structure of native apomyoglobin are the disappearance of most contacts between both the unfolded D and F helices and the rest of the molecule and an increase in contacts between the CD loop with both the G helix and the FG loop. The latter change is a consequence of the collapse of these loops into the cavity left

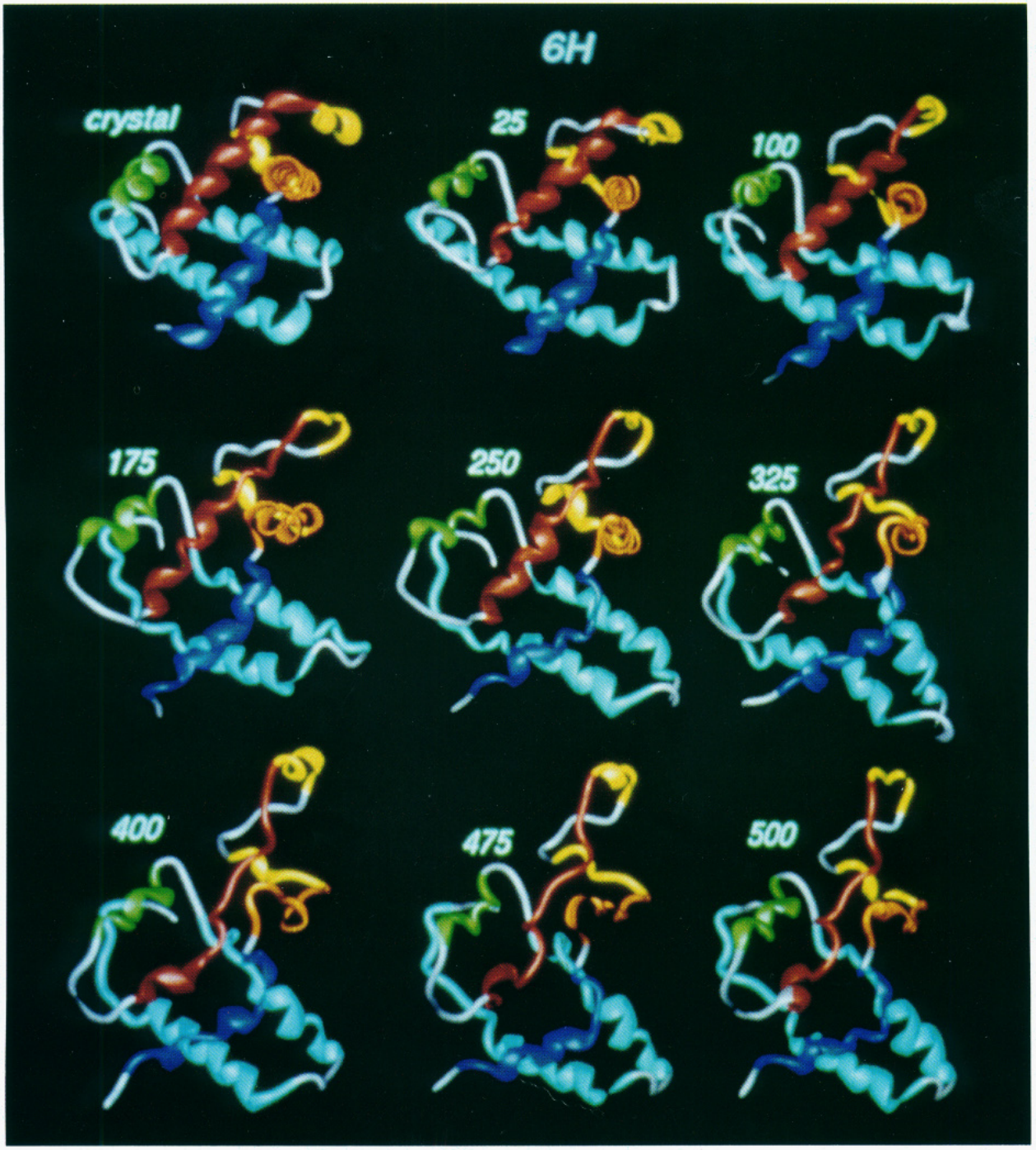

FIGURE 8: Ribbon representations of the crystal structure of myoglobin less the heme and of snapshots along simulation 6H for apomyoglobin at 358 K.

by the heme. The stable portion of helix B is in contact with helices E and G, while the stable portions of helix H are in contact with helices A and E and the EF loop.

The average structure over the last 50 ps of the room temperature simulation 6C also reproduces most of the spatial relationships determined by 2-D NMR on native apomyoglobin (Cocco & Lecomte, 1990; Cocco et al., 1992). Figure 7 shows the side chains of the residues found in the NMR experiments to form clusters in the hydrophobic core. All the distances indicated by dashed lines in this figure are computed to be in the 4.07–5.95-Å range. It can be seen that most of these residues, most particularly, the Phe-106 ↔ Tyr-103 pair from helix G, the Leu-40 ↔ Phe-46 ↔ Phe-43 ↔ Phe-33 cluster from the interface between helices B and C, and the CD loop, and Val-17 ↔ His-24 ↔ Val-114 ↔ His-113 ↔ His-119 ↔ Leu-115 ↔ Phe-123 ↔ His-116 from the interface between helices A, B, and G, are close enough to explain the observed NOE's. However, some additional NOE's are observed that appear to be out of range in the computed structures. In particular, His-36 is too far from Phe-106, and Phe-46, which although very close to Leu-40, is too far from Leu-49 and His-48 to be considered within NOE range. These residues are all located in either the C helix or the CD loop. As mentioned above, this region undergoes some of the most dramatic changes during the course of the simulation to close the void left by removal of the heme, including the unfolding of helix D.

Ribbon representations of instantaneous structures from the unfolding simulations at 358 K, 6H and 4H, are shown in Figures 8 and 9. The helical contents averaged over the last 50 ps are 32 ± 4% and 33 ± 3%, respectively, again matching well the helicity estimated from CD, 35% (Hughson

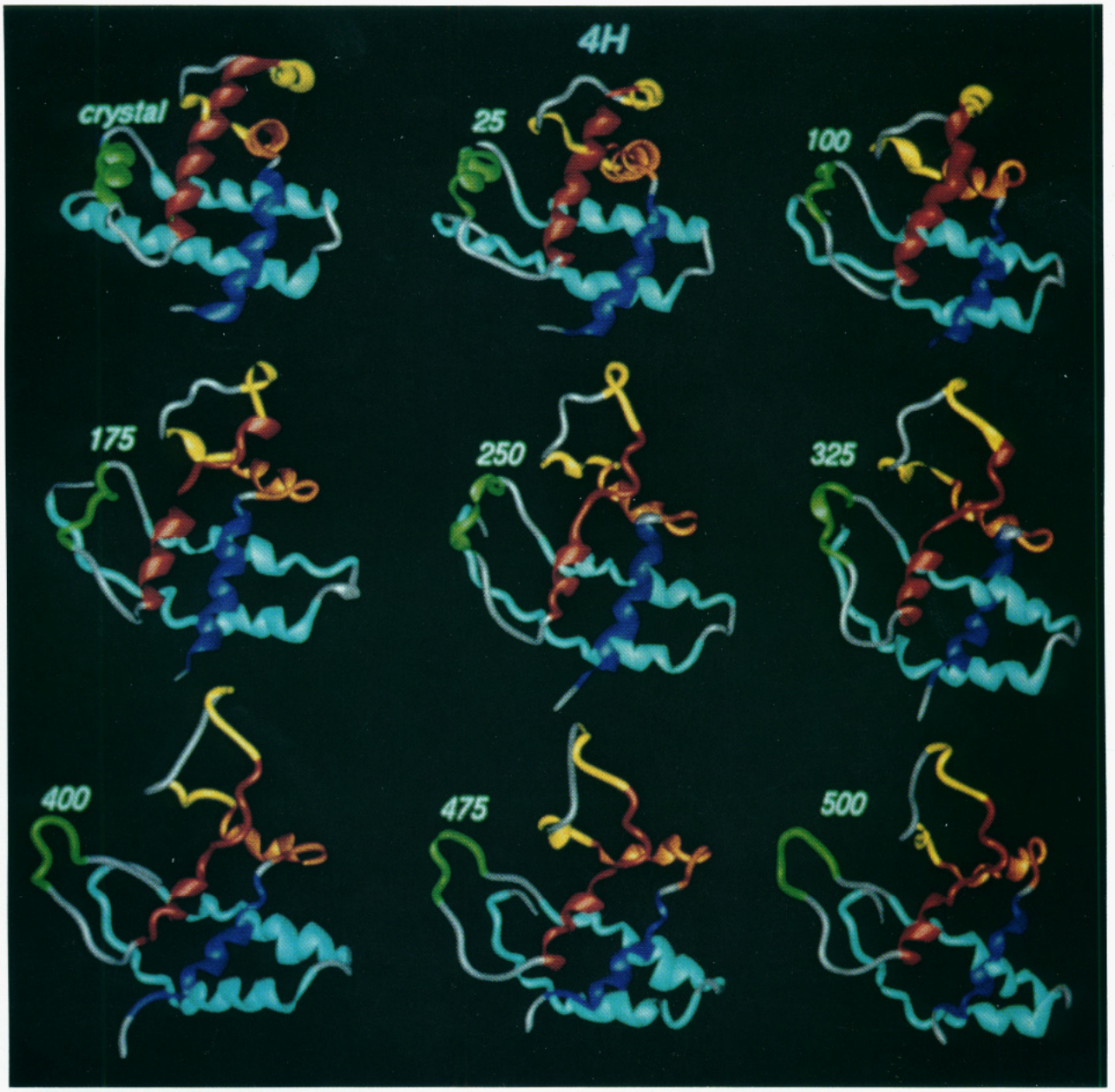

FIGURE 9: Ribbon representations as in Figure 8 including snapshots along simulation 4H for apomyoglobin at 358 K.

et al., 1990). This provides additional quantitative support for taking the final computed structures from the 358 K runs as models for the I state. Both final structures have common motifs; large fragments of helices A, B, E, G, and H are still present, and helix D is completely unfolded. The main differences between them are that helix F and a single turn of helix C are still present in 6H, but are absent in 4H. There is also some difference in the behavior of helix A. While in simulation 6H it has fragmented into two smaller helices at the N- and C-termini, the middle section (ca. 2.5 turns) is the only remaining part at the end of simulation 4H. This pattern of stability of helix A is paralleled by its side-chain contacts to helices G and H. The reduction in helical content observed by CD in the folding intermediate can be attributed not only to the disappearance of the small C and D helices but also to the partial unfolding of some of the larger helices.

The trajectories of the helices in all three simulations share some common patterns. At the secondary structure level, the most noticeable similarity is for helix H, which fragments rapidly into two smaller helices. In simulation 6C it remains that way, while in both high-temperature runs the helical segment closest to the C-terminus subsequently unfolds. Helices G and E, on the other hand, unravel partially starting from the N-terminus. At the more microscopic level, as each helical hydrogen bond is being broken or formed, the same transition pattern $\alpha \rightleftharpoons 3_{10} \rightleftharpoons$ no hydrogen bond observed in our previous studies of the S-peptide analog (Tirado-Rives & Jorgensen, 1991) is followed in 69% of the appearance/disappearance events.

The results of the simulations are also consistent with observations from the hydrogen isotopic exchange experiments (Hughson et al., 1990). For the acid unfolding intermediate I, only portions of helices A, G, and H and a single residue of helix B show protection factors of 10 or above. These helices are still present at the end of both simulations at 85 °C, 6H and 4H. The hydrogen-exchange experiments led to the proposal that the I form has the B, C, D, and E helices unfolded and that most of the remaining helical structure is in the vicinity of the junction of the A, G, and H helices. The low-pH simulation, 4H, largely supports this picture with the addenda that the experimentally unmonitored helix F is unfolded and that the C-terminal region of helix E, which contacts helices G and H, retains helicity. Some helicity near the center of helix B is also retained. Figures 8 and 9 show

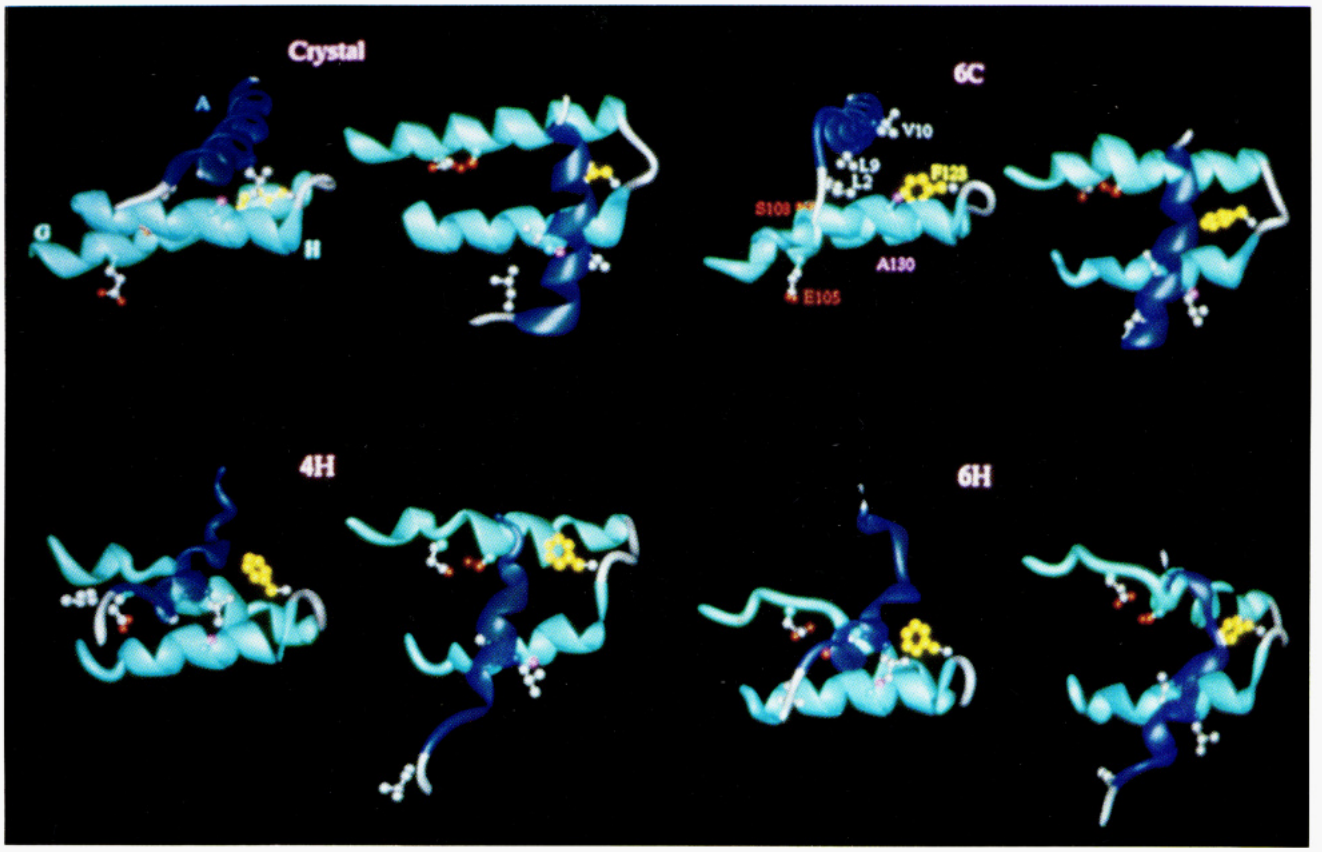

FIGURE 10: Orthogonal views of the region near the junction of the A, G, and H helices of the crystal structure (top left) and in the average structures over the last 50 ps from simulations 6C (top right), 4H (bottom left), and 6H (bottom right). The side chains of selected residues discussed in the text are shown.

that helix B at the end of the simulations is distorted, reoriented, and more solvent exposed, compared to the starting structure. With such structural change, it is possible that formally some helical character could remain, while the protons become more or less fully exchangeable.

Furthermore, the structures from the present simulations can be used to explain the puzzling experimental results for site-directed mutagenesis on apomyoglobin (Hughson et al., 1991). Since the results of the hydrogen-exchange/2-D NMR experiments indicated that the acid-stable I form contained helices A, G, and H, presumably in an arrangement similar to the holoprotein (Hughson et al., 1990), Baldwin and co-workers generated a series of modified proteins to probe the interactions among these helices. They replaced independently three residues, Ser-108 and Phe-123, whose side chains are located between helices G and H in the crystal structure of the holoprotein, and Ala-130, located in the interface between helices A and H (Figure 10). If the AGH region of the I form retains nativelike interactions, these mutations should cause a change in the relative stabilities of the I and the unfolded (U) form, while preserving the relationship in the native (N) to the I form. Surprisingly, when the helicities of the mutants were studied by following the changes in the CD as a function of pH, the mutations shifted the mid-transition pH for the conversion between the N and I forms by as much as 0.8 unit, mostly to higher values. On the other hand, the transition of I to the unfolded species (U) showed much smaller deviations to both higher and lower values. The implications are that the mutations mostly destabilize the native structure with respect to the intermediate form and that the I state does not retain nativelike interhelical interactions (Hughson et al., 1991).

These results can be more easily understood in light of the structures from the present simulations. The side chains of some of the pertinent residues are shown in the ribbon representations of the AGH region of the crystal structure of myoglobin and the average structures at the end of all three runs in Figure 10. The side chain of Ser-108 is between helices G and H in both the crystal and 6C, the native structure (N) of apomyoglobin. In the molten globules formed in simulations 6H and 4H, the original N-terminus of helix G has unraveled up to residues 106–107. Ser-108, although still mostly in the hydrophobic pocket formed by Asn-132, Leu-135, and Phe-138 from helix H, is also in contact with Arg-139 and is hydrogen bonded to Glu-105. Its replacement is then expected to affect both the relative stabilities of N vs I and I vs U, which is the experimentally observed result. The side chain of Phe-123 is completely buried in the interface region of helices G and H in the crystal structure. In the model for the N form provided by the 6C simulation, it has moved slightly out of this pocket, but in both the 6H and 4H average structures, it is completely out of the interface. Any perturbations on this group should, as observed experimentally, affect the N to I transition more than the I to U one. Ala-130 is a very similar case, albeit not as dramatic. In both the crystal and 6C structures, its side chain is very close to Leu-2, Leu-9, and Val-10, while in 4H and 6H, it has lost its contact with Leu-2, mainly due to partial unfolding of helix A, and it is in a less constrained environment. This would be reflected in a larger effect for the N to I transition, while I to U should be quite insensitive to changes in this side chain. The experimental results show changes in the N ↔ I mid-transition pH of 0.2–0.8 unit, while the largest shift in the I ↔ U transition is 0.1 pH unit. Thus, although the A, G, and H regions retain substantial helicity, the interhelical structure is found to drift significantly from the nativelike form upon partial unfolding.

Overall, the present study demonstrates that molecular dynamics simulations can be effectively utilized to study the

unfolding of a protein, apomyoglobin, in water in computationally manageable times using full atomic representations for both the solute and the solvent under experimentally accessible conditions. The simulations reveal a stable molecule with reduced secondary structure content at 298 K and extensive unfolding to produce expanded and less structured forms at 358 K. The resulting structures show excellent agreement with available experimental data in terms of secondary structure content and tertiary structure information. Consequently, it appears that the molecular dynamics simulations have provided a plausible view of the low-pH folding intermediate. Of course, it is desirable to run additional trajectories to obtain confirmation of the initial view, though the general consistency of the results from the 4H and 6H simulations is notable. The present unfolding pathway reveals rapid decay of less stable helices in a time frame that permits little overall swelling of the protein. Reversing the time sense, this could be taken as consistent with folding occurring by an initial hydrophobic collapse, followed by fine tuning the secondary and tertiary structure (Dill, 1985). However, this is very speculative, and the apparent stability of the region of confluence of the A, G, and H helices could also be taken as a potential nucleation site following, perhaps, transient formation of portions of these helices with simultaneous collapse (Kim & Baldwin, 1982). Caution is particularly advised in assigning relevance to unfolding pathways from very high temperature simulations (Daggett & Levitt, 1992b; Mark & van Gunsteren, 1992); helices are undoubtedly very unstable under such conditions and will decay rapidly, well before more global diffusion of the protein. Consequently, such calculations are prone to suggest folding models with late formation of secondary structure (Mark & van Gunsteren, 1992). Detailed analyses of the trajectories for the helical fragments in all three present simulations did show a correlation between side-chain contacts and stability. The most enduring helical fragments at elevated temperature are those in contact with other helices.

REFERENCES

Baldwin, R. L. (1989) *Trends Biochem. Sci. 14*, 291.

Baldwin, R. L. (1991) *Chemtracts: Biochem. Mol. Biol. 2*, 379.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F, Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535.

Breslow, E., Beychok, S., Hardman, K. D., & Gurd, F. R. N. (1965) *J. Biol. Chem. 240*, 304.

Brooks, C. L., III (1992) *J. Mol. Biol. 227*, 375.

Bycroft, M., Matouscheck, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488.

Chan, H. S., & Dill, R. A. (1989) *Macromolecules 22*, 4559.

Christensen, H., & Pain, R. H. (1991) *Eur. Biophys. J. 19*, 221.

Cocco, M. J., & Lecomte, J. T. J. (1990) *Biochemistry 29*, 11067.

Cocco, M. J., Kao, Y.-H., Phillips, A. T., & Lecomte, J. T. J. (1992) *Biochemistry 31*, 6481.

Daggett, V., & Levitt, M. (1992a) *J. Mol. Biol. 223*, 1121.

Daggett, V., & Levitt, M. (1992b) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5142.

Dill, K. A. (1985) *Biochemistry 24*, 1501.

Ferrin, T. E., Huang, L. E., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13.

Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., & Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127.

Harrison, S. C., & Blout, E. R. (1965) *J. Biol. Chem. 240*, 299.

Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544.

Hughson, F. M., Barrick, D., & Baldwin, R. L. (1991) *Biochemistry 30*, 4113.

Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol. 49*, 117.

Jorgensen, W. L., & Ravimohan, C. (1985) *J. Chem. Phys. 83*, 3050.

Jorgensen, W. L., & Tirado-Rives, J. (1988) *J. Am. Chem. Soc. 110*, 1657.

Jorgensen, W. L., & Severance, D. L. (1990) *J. Am. Chem. Soc. 112*, 4768.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983) *J. Chem. Phys. 79*, 926.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459.

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631.

King, J. (1989) *Chem. Eng. News 67* (15), 32.

Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133.

Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 56*, 87.

Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749.

Mark, A. E., & van Gunsteren, W. F. (1992) *Biochemistry 31*, 7745.

Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 837.

Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633.

Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167.

Privalov, P. L., Griko, Yu. V., Venyaminov, S. Yu., & Kutyshenko, V. P. (1986) *J. Mol. Biol. 190*, 487.

Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 273.

Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302.

Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327.

Singh, U. C., Weiner, P. K., Caldwell, J., & Kollman, P. A. (1986) *AMBER 3.0*, University of California, San Francisco.

Skolnick, J., & Kolinski, A. (1990) *Science 250*, 1121.

Soman, K. V., Karimi, A., & Case, D. A. (1991) *Biopolymers 31*, 1351.

Tanford, C. (1968) *Adv. Protein Chem. 23*, 122.

Tirado-Rives, J., & Jorgensen, W. L. (1991) *Biochemistry 30*, 3864.

Tobias, D. J., & Brooks, C. L., III (1991) *Biochemistry 30*, 6059.

Tobias, D. J., Mertz, J. E., & Brooks, C. L., III (1991) *Biochemistry 30*, 6054.

Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. J. (1984) *J. Am. Chem. Soc. 106*, 765.

Yun, R. H., Anderson, A., & Hermans, J. (1991) *Proteins: Struct., Funct., Genet. 10*, 219.