Biochemistry

 Volume 31, Number 34 September 1, 1992

Accelerated Publications

Simulation of the Thermal Denaturation of Hen Egg White Lysozyme: Trapping the Molten Globule State

Alan E. Mark and Wilfred F. van Gunsteren*

Department of Physical Chemistry, ETH-Zentrum, CH-8092 Zürich, Switzerland

Received March 16, 1992; Revised Manuscript Received June 17, 1992

ABSTRACT: In the study of protein folding, much attention has focused on the characterization of folding intermediates. We report here molecular dynamics simulations in which the initial stages of the thermal denaturation of hen egg white lysozyme in aqueous solution are examined in detail. It is found that lysozyme unfolds in a two-stage process with the initial formation a quasi-stable state in which significant rearrangement of the secondary structure takes place. No evidence for distinct folding domains was found. The simulations suggest that the formation of well-defined secondary structure occurs after the initial collapse of the peptide chain and thus tend against the framework model of protein folding.

A growing body of evidence suggests that the transient formation of a compact intermediate state, characterized by nativelike secondary structure but a slowly fluctuating tertiary structure, is a general feature in the folding pathway of globular proteins (Dobson, 1991; Matouschek et al., 1990; Ewbank & Creighton, 1991; Christensen & Pain, 1991; Kim & Baldwin, 1991; Ptitsyn et al., 1990). The study of such transient intermediates at atomic resolution is, however, extremely difficult. Crystallization is impossible. Analysis of data from spectroscopic methods, primarily nuclear magnetic resonance (NMR) and circular dichroism (CD), is complicated not only by the short lifetime of the intermediate but also by the fact that the system is not characterized by a single state. Fortunately, computer simulation techniques provide one avenue for gaining further insight. Molecular dynamics simulations combine detailed information at an atomic level with fine resolution in time to permit analysis of the sequence of changes in atomic interactions during the unfolding process. Thus, simulations can provide unique insight into the underlying mechanism.

Under equilibrium conditions the folding–unfolding transition of hen egg white lysozyme appears a highly cooperative two-state process (Tanford et al., 1973). This is in marked contrast to the homologous protein α-lactalbumin which forms stable partially folded states under mild denaturing conditions (Baum et al., 1989; Kuwajima et al., 1985). Indirect evidence, nevertheless, suggests that lysozyme and α-lactalbumin fold via a similar structural intermediate (Kuwajima et al., 1985; Miranker et al., 1991). On the basis of differences in amide proton exchange rates during the folding of lysozyme, the presence of two distinct folding domains has recently been suggested (Miranker et al., 1991).

In the present study we have used molecular dynamics simulation techniques to analyze the initial stages of the unfolding of an individual lysozyme molecule in water at atomic resolution. It is found that lysozyme unfolds in a two-stage process. There is an initial formation of a quasi-stable state in which significant rearrangement of the secondary structure takes place. The simulations do not indicate the presence of distinct folding domains. The simulations do suggest that the formation of well-defined secondary structure occurs after the initial collapse of the peptide chain.

METHODS

The process of protein folding and unfolding is driven by the interplay between protein–protein and protein–water interactions; for this reason water was explicitly included in all simulations. To demonstrate that lysozyme is stable within the GROMOS (van Gunsteren & Berendsen, 1987) force field and for comparison with later simulations at higher temperature, an initial simulation of 550 ps at 300 K was performed. A second simulation was used to follow the denaturation process. This simulation was branched from the first after 50

* To whom correspondence should be addressed.

0006-2960/92/0431-7745$03.00/0

ps. Lysozyme thermally denatures at about 350 K at pH 7.0 (Sophianopoulos & Weiss, 1964). However, unfolding was initiated by heating the system to 500 K in order to dramatically increase the speed of the unfolding process which is too slow to be simulated at 350 K. All simulations were performed at constant volume using rectangular periodic boundary conditions. The temperature was maintained by weak coupling (τ = 0.1 ps) to an external bath (Berendsen et al., 1984). The starting structure was taken from the crystal structure of triclinic hen egg white lysozyme (Ramanadham et al., 1987), entry 2LZT of the Brookhaven Protein Data Bank (Bernstein et al., 1977). The dimensions of the periodic box, x = 4.88 nm, y = 5.31 nm, and z = 6.81 nm, were chosen such that no protein atom in the starting structure lay within 0.95 nm of the box wall. Treating nonpolar hydrogens using a united atom approach leads to a total system size of 17 299 atoms including 5345 simple point charge (SPC) water molecules (Berendsen et al., 1981). Bond lengths were constrained using the SHAKE procedure (Ryckaert et al., 1977). Nonbonded interactions were treated using a twin-range method (van Gunsteren & Berendsen, 1990). Within the short-range cutoff of 0.8 nm all interactions were determined every time step. Longer range electrostatic interactions within a cutoff range of 1.4 nm were updated only every 10 fs. A time step of 1 fs was used for simulations at 500 K. At 300 and 320 K a 2-fs time step was used. Simulations and analysis were performed using the GROMOS package of programs (van Gunsteren & Berendsen, 1987). Additional analysis was performed using the secondary structure analysis program DSSP (Kabsch & Sander, 1983). The calculations were performed in parallel on a Silicon Graphics 380 and on an IBM 3090-60 apparatus. The calculations required the equivalent of 1470 cpu h on a Cray Y-MP computer.

RESULTS AND DISCUSSION

The 550-ps simulation of hen egg white lysozyme at 300 K used in this study as a control clearly showed that the overall fold of the protein and the nature of the secondary structure elements of the protein are stable in the force field at 300 K. The average structure over the final 250 ps of the simulation showed a root mean square (rms) deviation of C^α atoms from the starting X-ray structure of 0.22 nm. The most significant structural changes observed during the simulation were shifts in the relative position of the two main structural domains associated with the hinge bending mode of hen egg white lysozyme (McCammon et al., 1976). Details of this and other features of the simulation will be presented elsewhere.

The process of unfolding was monitored by following the radius of gyration, the rms deviation of C^α atoms from the starting X-ray structure, and changes in the hydrogen-bonding pattern within the protein backbone as a function of simulation time. At 300 K the radius of gyration stabilized at approximately 1.46 nm after 175 ps, as compared to a radius of gyration of 1.43 nm for the starting X-ray structure, and remained effectively constant thereafter. The initial increase in the radius of gyration can be primarily attributed to surface residues and loops which pack against the body of the protein in the crystal extending into solution during the simulation. At 500 K the radius of gyration first increased steadily for 90 ps and then remained effectively constant for 40 ps at approximately 1.6 nm before again rising, this time more rapidly than before (Figure 1a). The magnitude of the rms deviation of C^α atoms from the starting X-ray structure asymptotically approached a value of 0.35 nm before rising linearly after 110 ps (Figure 1b). Figure 1 clearly indicates

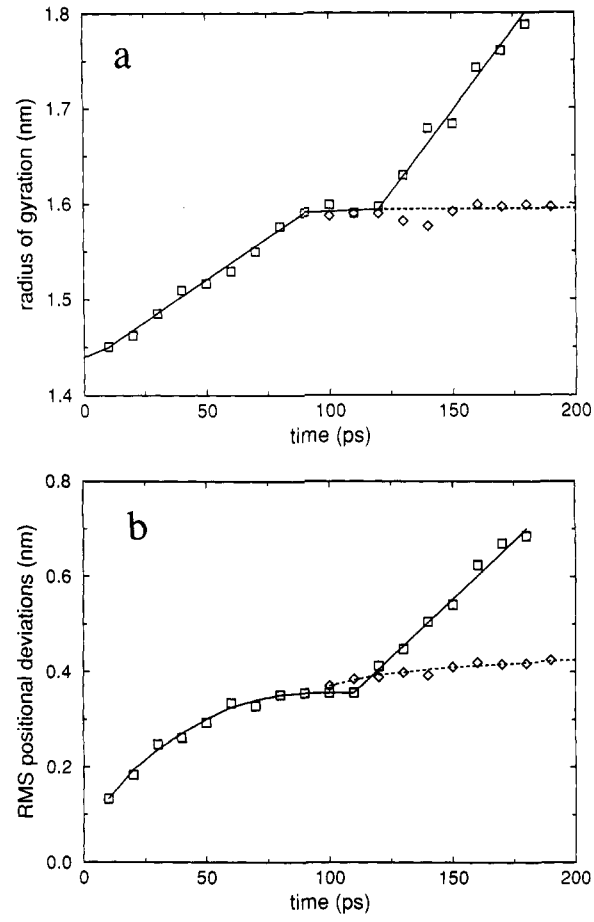

FIGURE 1: Radius of gyration (a) and root mean square deviation of C^α atoms from the X-ray structure (b) as a function of simulation time for hen egg white lysozyme. Values correspond to structures averaged over 10-ps windows of the trajectory. The open squares (□) relate to simulations at 500 K after 50-ps equilibration at 300 K. The tilted squares (◇) relate to simulations at 320 K branched from the simulation at 500 K after 85 ps. The solid and broken lines reflect an attempt to fit the data.

that, for lysozyme, unfolding is a two-stage process. There is an initial formation of an intermediate state which after a short period of time, 30–40 ps at 500 K, becomes unstable and begins to fully unfold. It was possible to stabilize the intermediate state by cooling the system to 320 K over 5 ps at the start of the plateau region after 85 ps at 500 K. The simulation was continued for 190 ps at 320 K, permitting analysis of the structural properties of the partially unfolded intermediate. During the 190-ps simulation at 320 K the radius of gyration remained effectively constant, fluctuating within 0.025 nm of an average value of 1.6 nm. Values for the first 100 ps are given in Figure 1a.

The state stabilized in the simulation shows many characteristics of folding intermediates observed experimentally. The radius of gyration is the same as that measured for partially folded states of the highly homologous protein α-lactalbumin (Kuwajima et al., 1985). There is a marked increase in solvent penetration. The solvent-accessible surface area (Kabsch & Sander, 1983) has increased from 66 nm^2 for the crystal conformation to $\approx$100 nm^2. The structure, nevertheless, retains a high degree of native-like secondary structure (Figure 2). Sections of α-helix and the overall fold of the molecule are clearly visible in the partially unfolded state. Averaged over the entire 190 ps at 320 K the hydrogen-bonding pattern of the partially unfolded state contains each of the secondary structure elements found in the crystal structure. Changes in hydrogen-bonding patterns as a function of time, however,

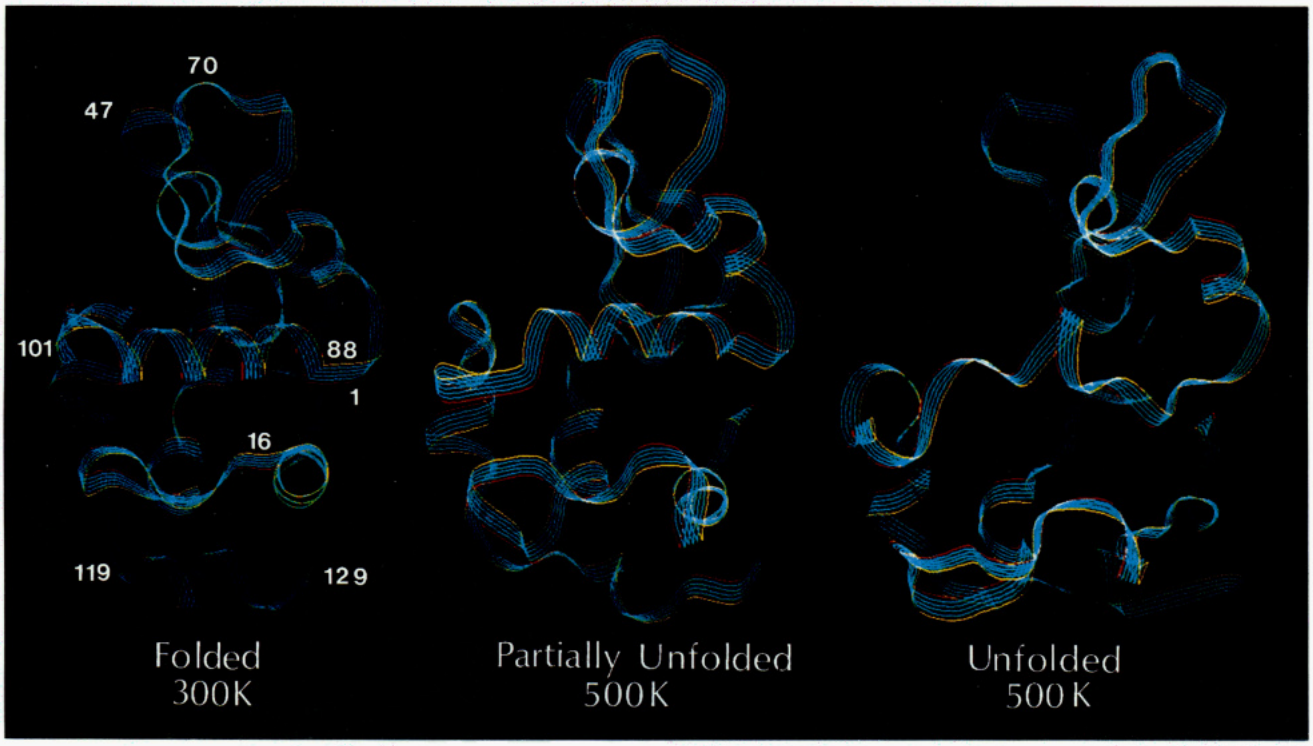

FIGURE 2: Ribbon representations of the protein backbone of HEW lysozyme at 300 and 500 K. Structures correspond to trajectory averages (⟨...⟩) over the ranges indicated. Panels: (a, left) ⟨350–550 ps⟩ at 300 K (folded); (b, center) ⟨110–120 ps⟩ at 500 K (partially unfolded); (c, right) ⟨170–180 ps⟩ at 500 K (unfolded).

indicate a high degree of conformational flexibility. There is a general reduction in the average lifetime of individual hydrogen bonds compared with the simulation at 300 K. The most stable secondary structure elements in the partially unfolded state are two sections of α-helix residues 6–12 and 88–98. No evidence for the presence of two distinct folding domains, as had been inferred from NMR studies (Miranker et al., 1991), was found. Two large loop regions, residues 44–52 and 62–80, are still clearly visible in the later stages of unfolding. These two loops are the remnants of the section of β-sheet, residues 41–60, and a large loop, residues 61–84, present in the native structure. Changes observed in amide proton exchange rates during folding more likely reflect differences in solvent accessibility rather than the order of formation of secondary structure elements. The largest deviations from the X-ray structure in the partially unfolded state were observed for residues 66–74. These residues, which are part of a large loop in the native structure, move away from the body of the protein during the simulation, thereby increasing the solvent exposure of the triple-stranded β-sheet, residues 41–60 in the native structure.

At 500 K the major loss of secondary structure coincides with the plateau region in the radius of gyration (Figure 3). Expressed in terms of the number of hydrogen bonds per hundred residues, it can be seen that there is a sharp decline in $i \rightarrow i + 4$ (α-helical) type hydrogen bonding in favor of $i \rightarrow i + 2$ interactions after 90 ps at 500 K. Loss of $i \rightarrow i + 4$ type bonding begins at both ends of each helix. There is rapid interconversion between $i \rightarrow i + 4$, $i \rightarrow i + 3$, and $i \rightarrow i + 5$ hydrogen bond pairing. From Figure 2c it can be seen that small sections of helix still remain present in the latter stages of unfolding. The positions of these short regions of helix change rapidly during the simulation and do not appear to represent fixed nucleation sites.

The change in the hydrogen-bonding pattern from predominantly $i \rightarrow i + 4$ to $i \rightarrow i + 2$ occurs while the protein is still highly compact. The total number in intramolecular

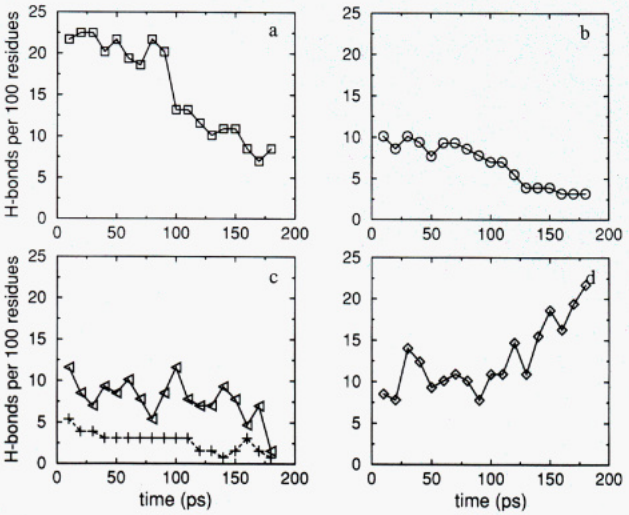

FIGURE 3: Hydrogen-bonding type expressed as the number of hydrogen bonds per 100 residues. Values were determined for structures averaged over 10-ps windows of the trajectory using the secondary structure analysis program DSSP (Kabsch & Sander, 1983), where i denotes the residue sequence number: (a) (□) $i \rightarrow i + 4$ (α-helix); (b) (○) β-sheet; (c) (△) $i \rightarrow i + 3$ (3_{10} helix) and $+i \rightarrow i + 5$ (π-helix); (d) (◇) $i \rightarrow i + 2$.

hydrogen bonds is only slightly decreased. Little recognizable secondary structure remains after 180 ps at 500 K. Nevertheless, 40% of all residues continue to form hydrogen bonds within the protein backbone. This compares with an average of ≈50% for the simulation of the native structure at 300 K. The simulations suggest that during the unfolding process the loss of secondary structure elements occurs while the peptide chain is still highly compact. If we assume that the folding and unfolding processes follow the same basic pathway, this would indicate that upon folding the formation of well-defined secondary structure occurs after the initial collapse of the peptide chain to form a compact intermediate. The initial collapse of the chain may possibly be driven by the formation

of $i \rightarrow i + 2$ hydrogen bonds. The simulations thus tend against a framework model for protein folding involving first the formation of sections of secondary structure which then coalesce to form the final tertiary structure.

The examination of protein folding and unfolding using molecular dynamics simulations is clearly limited by the time scale of the processes involved. The computationally demanding nature of this work means that the results presented are necessarily those of a single simulation. A temperature of 500 K has been used in order to dramatically increase the rate of the unfolding process. The exact effect of this temperature on the dynamics of the different stages of the unfolding process is unclear. The mechanism of (un)folding at a given temperature will depend on the relative rate of secondary structure formation (or loss) compared to the rate of the collapse of the peptide chain or the diffusion of the elements of secondary structure toward (or away from) each other. To verify the results we have presented, the simulations will need to be repeated from different starting configurations and at different temperatures. Nevertheless, this work has shown that by the application of time-saving techniques (van Gunsteren, 1991) molecular dynamics can be a powerful tool in providing insight into the process of protein folding at an atomic level. Without doubt, as access to longer time scales becomes available, simulation techniques will make an ever more important contribution to the understanding of protein folding.

ACKNOWLEDGMENT

It is a pleasure to acknowledge Chris Dobson's challenge to unfold lysozyme by computer.

REFERENCES

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

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981) in *Intermolecular Forces* (Pullman, B., Ed.) pp 331–342, Reidel, Dordrecht.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684.

Bernstein, F. C., Koetzle, T. F., Williams, G. B., Meyer, G. F., Price, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 122*, 535.

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

Dobson, C. M. (1991) *Curr. Opin. Struct. Biol. 1*, 22.

Ewbank, J. J., & Creighton, T. E. (1991) *Nature 350*, 518.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577.

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

Kuwajima, K., Hiraoka, Y., Ikeguchi, M., & Sugai, S. (1985) *Biochemistry 24*, 874.

Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., & Fersht, A. R. (1990) *Nature 346*, 440.

McCammon, J. A., Gelin, B. R., Karplus, M., & Wolynes, P. G. (1976) *Nature 262*, 325.

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

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20.

Ramanadham, M., Sieker, L. C., & Jensen, L. H. (1987) *Acta Crystallogr., Sect. A (Suppl.) 43*, 13.

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

Sophianopoulos, A. J., & Weiss, B. J. (1964) *Biochemistry 3*, 1920.

Tanford, C., Aune, K. C., & Ikai, A. (1973) *J. Mol. Biol. 73*, 185.

van Gunsteren, W. F. (1991) in *Advances in Biomolecular Simulations* (Lavery, R., Rival, J.-L., & Smith, J., Eds.) pp 131–146, American Institute of Physics Conference Proceedings, No. 239, American Institute of Physics, Woodbury, NY.

van Gunsteren, W. F., & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

van Gunsteren, W. F., & Berendsen, H. J. C. (1990) *Angew. Chem., Int. Ed. Engl. 29*, 992.