Molecular Dynamics Simulations of the Unfolding of Barnase in Water and 8 M Aqueous Urea †

ARTICLE *in* BIOCHEMISTRY · JULY 1997

Impact Factor: 3.02 · DOI: 10.1021/bi970096i · Source: PubMed

3 AUTHORS, INCLUDING:

Julian Tirado-Rives
Yale University
88 PUBLICATIONS **14,031** CITATIONS

SEE PROFILE

Molecular Dynamics Simulations of the Unfolding of Barnase in Water and 8 M Aqueous Urea[†]

Julian Tirado-Rives,[‡] Modesto Orozco,[§] and William L. Jorgensen*,[||]

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, and Department de Bioquímica, Facultat de Química, Universitat de Barcelona, Barcelona 08028, Spain

Received January 16, 1997; Revised Manuscript Received March 25, 1997[⊗]

ABSTRACT: Molecular dynamics simulations of barnase have been conducted both in water and in 8 M urea solution for 500 ps at 25 °C and for 2000 ps at 85 °C. The final structure of the aqueous simulation at room temperature matches closely the structure obtained by NMR and the experimentally observed protections from isotopic exchange. The comparison of the structures generated by the aqueous simulation at 85 °C reveals a trajectory composed of groups of geometrically related structures separated by narrow regions of rapid change in structure. The first of these regions displays changes in backbone rmsd to the crystal structure and solvent-accessible area suggestive of a transition state, while the properties observed during the final 300 ps of the simulation are consistent with a stable intermediate. These assignments were confirmed by calculation of the "progress along the reaction coordinate" ϕ-values using an empirical equation based on a linear response method. The pathway of unfolding defined in this fashion agrees well with the experimental results of site-directed mutagenesis in terms of secondary structure content of the transition state and the intermediate and reproduces the relative stability of the different elements of secondary structure. The results of the simulations in urea suggest a mechanism at the molecular level for its well-known enhancement of the denaturation of proteins. The analysis of radial distribution functions shows that the first solvation shell of the protein is enriched in urea relative to the bulk solvent. The displacement of water molecules allows greater exposure of hydrophobic side chains, as witnessed particularly in the analysis of solvent-accessible surface areas at the higher temperature. Almost all urea molecules in the first shell form at least one hydrogen bond with the protein. They provide a more favorable environment for accommodation of the remaining water molecules, and they facilitate the separation of secondary structure elements by acting as a bridge between groups previously forming intraprotein hydrogen bonds.

Proteins display an exquisitely specific but widely varied range of distinct three-dimensional structures and topologies based on the repetition of a limited set of components (Richardson, 1981). Ever since the seminal experiments of Anfinsen showed that ribonuclease A folds to its native structure *in vitro* after denaturation (Anfinsen, 1959), it has been generally accepted that globular, single domain proteins encode in their sequence enough information to fold into their native structure (Baldwin, 1989; Jaenicke, 1987). The details and mechanism of this process are not completely understood, and the many attempts to formulate a model that explains this phenomenon have given rise to the "protein folding problem" (King, 1989). Experimental studies in this area have been challenging since most folding and unfolding transitions are highly cooperative and often fit a simple two-state model in which only the native and unfolded forms are populated (Tanford, 1968; Privalov, 1979).

In spite of the patent difficulties, great strides have been made in the experimental study of the folding pathway of proteins. Intermediates have now been detected for the unfolding of several proteins under mild conditions, and although their complete 3-D structures are not known, some of their physical properties have been measured and earned them the name of "molten globules" (Ptitsyn, 1987; Kuwajima, 1989). Intermediates have also been found in kinetic folding pathways through CD[1] 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 during refolding, presumably due to formation of secondary structure or to inaccessibility to the solvent. Some protected sites match the locations of secondary structure 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*., 1990a,b; Matouschek *et al*., 1992), and lysozyme (Miranker *et al*., 1991; Radford *et al*., 1992; Lu & Dahlquist, 1992). The capability of preparing significant quantities of mutants in which a single residue can be exchanged for another in a controlled fashion through protein engineering techniques has contributed greatly to the field. The folding pathways of barnase and chymotrypsin inhibitor 2 (CI2) have been extensively studied through a combination of protein engineering with kinetic

[†] This research was supported by NIH Grant GM32136 and a fellowship from the Fullbright Foundation (M.O.).

[‡] Yale University. E-mail: julian@matrix.chem.yale.edu.
[§] Universitat de Barcelona. E-mail: modesto@luz.bq.ub.es.
[||] Yale University. E-mail: bill@adrik.chem.yale.edu.

[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1997.

[1] Abbreviations: MD, molecular dynamics; MC, Monte Carlo; NPT, NVT simulations at constant pressure and volume, respectively; rmsd, root-mean-square deviation; CD, circular dichroism; SPT, statistical perturbation theory; SAA, solvent-accessible area; R_g, radius of gyration; CI2, chymotrypsin inhibitor 2; rdf, radial distribution function.

S0006-2960(97)00096-2 CCC: $14.00

and thermodynamic studies, which allowed characterization of the transition states for both proteins and a folding intermediate of barnase (Fersht, 1993; Otzen *et al*., 1994). It should be noted that the direct results of these experiments are obtained as ratios or differences in rate or equilibrium constants and therefore characterize the intermediates or transition states in terms of the effects of the mutations on their relative free energies.

The theoretical study of protein folding and unfolding can also be a valuable tool 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. Applications of molecular dynamics to the folding problem using atomic representations of both solute and solvent were initially concentrated on smaller peptides. Early examples include the unfolding of an analog of the S-peptide of ribonuclease A at 358 K (Tirado-Rives & Jorgensen, 1991) and the unfolding of an 18-residue peptide from the α-helix-H of myoglobin (Soman *et al*., 1991). The free energy changes on the interconversion between the α- and 3_{10}-helical forms have also been reported for undecaalanine using SPT and Monte Carlo techniques (Tirado-Rives *et al*., 1993) and for undecaaminoisobutyric acid by a SPT-MD combination (Smythe *et al*., 1993). Two more recent studies of peptides were focused on small fragments corresponding to regions of secondary structure of barnase, the β-hairpin composed of the third and fourth strands of the β-sheet (Pugliese *et al*., 1995) and a fragment containing the main α-helix and first loop (Braxenthaler *et al*., 1995).

It is only in the last five years that MD simulations of unfolding of complete proteins in water have appeared in the literature. Among the reported simulations are the unfolding of both the native and reduced forms of bovine pancreatic trypsin inhibitor at 423 K for 550 ps (Daggett & Levitt, 1992), hen egg white lysozyme at 500 K for 190 ps (Mark & van Gunsteren, 1992), apomyoglobin at 358 K for 500 ps (Tirado-Rives & Jorgensen, 1993), β-lactamase for 200 ps at 600 K (Vijayakumar *et al*., 1993), barnase at 600 K for 120 and 230 ps (Caflish & Karplus, 1994), and CI2 for 2 ns at 498 K (Li & Daggett, 1994). The latter study was later followed by three shorter 1 ns simulations that showed qualitative agreement to the previous results (Li & Daggett, 1996). It is important to note that most of these studies utilized highly elevated temperatures at constant volume to decrease the computation time needed to observe the unfolding, and the rather unnatural conditions employed may have influenced the unfolding pathways.

Other techniques have also been used to increase the probability of observing unfolding transitions in molecular dynamics studies. A set of three simulations of hen egg white lysozyme for approximately 200 ps each explored the effect of using elevated pressure (10 kbar) at 342 K, an additional radial force, or a kinetic energy gradient that provided additional outward velocity (Hünenberger *et al*., 1995). An additional unfolding simulation of barnase, in which Asp, Glu, and His side chains were protonated to simulate the effect of acidic conditions for a total simulation time of 1050 ps, was conducted at a lower temperature, 360 K (Caflish & Karplus, 1995).

The need to enhance the rate of unfolding is not unique to theoretical simulations. Experimental studies are often aided by the use of low pH or chaotropic agents such as guanidinium chloride or urea. Although the fact that addition of urea to protein solutions causes denaturation has been known for a long time [see, for instance, Tanford (1968)], the mechanism of action at the molecular level is largely unknown. This is a rather difficult problem to address experimentally, since direct probes in a protein−denaturant−water system are hard to formulate. Model studies on the solubility of hydrocarbons and peptides support either a bulk solvent effect by which the chaotropes make water a "more hydrophobic" medium or direct interactions between the denaturants and polypeptide, which favor the unfolded state (Roseman & Jencks, 1975). Support for a mechanism involving direct interactions has been mounting slowly through CD (Tiffany & Krimm, 1973), surface tension (Breslow & Guo, 1990), and calorimetric measures (Makhatadze & Privalov, 1992) and computational studies of model compounds (Duffy *et al*., 1993a).

The current level of knowledge on the mechanism of action of urea is well summarized by the statement that "at best, it may be qualitatively concluded that denaturants act by solvating more equally all portions of the unfolded polypeptide, increasing the aqueous solubility of the hydrophobic portions while maintaining the hydrogen-bonding capability of the aqueous solvent" (Creighton, 1985). Given the lack of experimentally decisive probes, the study of chaotrope-induced denaturation of proteins *via* MD simulations could be invaluable in the interpretation of the experimental results and greatly enhance the understanding of this effect. To the best of our knowledge there have not been any theoretical studies of proteins in the presence of urea or guanidinium chloride solutions.

The present simulations were undertaken to study the thermal and urea-induced unfolding of barnase using full atomic representations for both the protein and the solvent water under experimentally accessible conditions. Barnase is a small, extracellular ribonuclease with 110 residues and $M_r = 12\,382$ (Nishimura & Nomura, 1958). It is a single-chain $\alpha + \beta$ protein that contains no disulfide bonds and undergoes reversible unfolding by heat ($T_m \sim 50$ °C) or urea ($C_m \sim 4.2$ M). The structure of barnase has been solved both in the crystal (Mauguen *et al*., 1982; Baudet & Janin, 1991) and in solution (Bycroft *et al*., 1990a,b, 1991). And, very importantly for this study, the pathway it follows on folding and unfolding has been experimentally elucidated through a combination of kinetic and thermodynamic studies of over 120 mutants [for a review see Fersht (1993)]. This wealth of experimental data has made barnase a very attractive subject of study in other fields as well. In addition to the studies mentioned above, free energy perturbations have been performed to calculate the effect of mutating I^{96} into alanine (Prévost *et al*., 1991; Sun *et al*., 1996) and to gain insights into the causes for the raised pK_a of H^{18} (Prévost, 1996).

The crystal structure, shown as a "ribbon" representation using the Molscript program (Kraulis, 1991) in Figure 1, features three α-helices (H1, T^6-H^{18}; H2, $T^{26}-G^{34}$; and H3, $N^{41}-A^{46}$), a five-stranded antiparallel β-sheet (S1, $I^{51}-I^{55}$; S2, $T^{70}-I^{76}$; S3, $S^{85}-S^{91}$; S4, $W^{94}-T^{99}$; and S5, $F^{106}-K^{108}$), and three hydrophobic cores. The main hydrophobic core (C1) is formed by the packing of the longest helix against the β-sheet. The smallest core (C2) comprises the residues from the second and third helices that form hydrophobic contacts among them, the first and second loops and the bottom of the β-sheet, while the third and last core (C3),

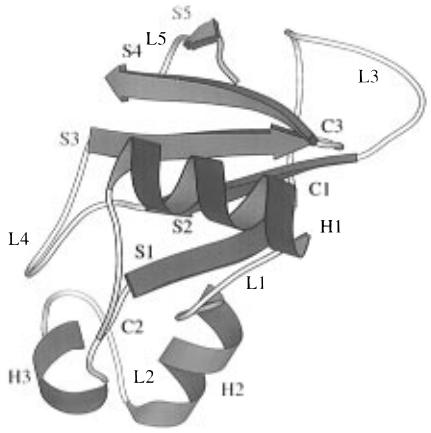

FIGURE 1: Ribbon diagram representation of the crystal structure of barnase prepared with the Molscript program (Kraulis, 1991). The color scheme is used to mark the regions of secondary structure in the crystal structure and is maintained in all subsequent figures.

which contains the guanine-specific binding site, is formed by the packing of the third loop against the back of the β-sheet.

COMPUTATIONAL METHODS

All simulations were conducted using the AMBER 4.0 program (Pearlman *et al*., 1991). The protein was represented by the OPLS nonbonded parameters (Jorgensen & Tirado-Rives, 1988) including the all-atom parameters for aromatic side chains (Jorgensen & Severance, 1990) combined with the angle bending and torsional parameters from the AMBER force field (Weiner *et al.*, 1984), and the solvent was represented by the TIP3P model for water (Jorgensen *et al.*, 1983) and the OPLS parameters for urea (Duffy *et al*., 1993b). The dielectric constant was fixed at 1.0, and the scaling factors were 8.0 and 2.0 for the 1,4 Lennard-Jones and 1,4 electrostatic interactions, respectively. All bond lengths and the H−H distances in water were fixed using the SHAKE algorithm (Ryckaert *et al*., 1977), which allowed a time step of 2 fs. A residue-based cutoff of 9 Å was used to generate the nonbonded list. All the solution calculations utilized periodic boundary conditions to avoid edge effects. The pressure was fixed at 1 bar (0.987 atm) in all NPT simulations. Structures and energies were saved every 0.2 ps and were analyzed separately using software developed in the authors' laboratories. The secondary structure assignments and solvent-accessible areas were obtained with the DSSP program (Kabsch & Sander, 1983) from these saved structures.

The simulations were initiated by adding the necessary hydrogen atoms to the crystal structure, entry 1BNJ in the Brookhaven Protein Data Bank (Bernstein *et al*., 1977), using the EDIT module of AMBER 4.0. Both histidines were protonated since most of the experiments were conducted at a pH range of 5.0−6.3, and the pK_a's have been determined to be 7.9 and 6.3 for H^{18} and H^{102}, respectively (Sali *et al*., 1988). The initial steric clashes were then relieved by steepest descent minimization *in vacuo* using a distance-dependent dielectric of $4R_{ij}$. This minimization converged in 510 steps and produced only modest changes in structure, the backbone rmsd being only 0.215 Å, while decreasing the total energy from an initial value of +1200 to −1000 kcal mol^{-1}. Since in this protonation state the total charge of barnase was +4, chloride counterions were added to provide a neutral simulation cell. In order to locate these ions in a relatively objective fashion, the protein was placed in the center of a box of water that extended at least 8 Å beyond the protein in each Cartesian and had a density close to 1.0 g·cm^{-3}. The initial steric clashes were first relieved by 100 steps of steepest descent minimization followed by 13 ps of constant volume MD, during which only solvent molecules were allowed to move and the temperature was raised from 100 to 300 K during the first 4 ps. At the end of this time a subset of the 30 water molecules within 9 Å of the protein and with the most negative electrostatic energies was selected, and from this subset four water molecules near the surface of the protein and far from each other were graphically selected and replaced by chloride ions. The resulting system was then equilibrated through 30 ps of NVT-MD at 300 K in which only the protein atoms were kept fixed. The positions and velocities of the solvent and the counterions and the positions of the protein atoms were taken from this final configuration, while the velocities of the protein were assigned from a separate 30 ps MD simulation at constant volume in which both the counterions and water molecules were kept fixed and the temperature was raised from 100 to 300 K. The initial setup was then finished by a short, 1 ps NVT-MD run at 300 K to yield the initial configuration W1 that contains the protein, four chloride ions, and 4706 water molecules in a box of dimensions 59.3 × 55.2 × 48.0 Å.

This initial configuration was used as the starting point for two different simulations in pure water. A control simulation at 300 K (CW) was continued for 500 ps at NPT. A second simulation at 360 K (HW), *ca*. 35 °C above the T_m, was also started from the same configuration by raising the temperature over 30 ps and then continued to achieve a total simulation time of 2000 ps at 360 K. Since under unfolding conditions the shape of the protein can change, the periodic distances were closely monitored during the simulation in order to avoid artifacts caused by proximity of portions of the protein with periodic images of itself. When the minimum periodic distance fell below 11.0 Å, the box was reorganized by removing a small layer of solvent from the Cartesian direction in which the periodic distance was the largest and adding the same number of solvent molecules, by replication of their periodic images, to the edge of the box in which the distances were the shortest. The positions and velocities of the outermost shell of water were then reinitialized from separate 10 ps NVT-MD simulations in which only the modified solvent and an additional layer of 3.0 Å width were allowed to move, and the temperature was raised from 100 to 360 K over the first 5 ps. The remainder of the system, including the protein, counterions, and all of the unmodified solvent maintained their positions and velocities. This procedure minimizes the disruption to the trajectory while it avoids the need to resolvate the system and repeat the simulation. For the HW simulation only two such reorganizations were needed at 600 and 1860 ps.

The simulations in 8 M urea were started from the same initial configuration W1 by first creating a box of the same size containing only urea by replication of a previously equilibrated box and superimposing the boxes. Next all urea molecules that overlapped with the protein or the chlorides were deleted, and from the remainder the 759 urea molecules with the most favorable interaction energies with the protein and chlorides were kept. The interaction energy between

Table 1: Average Properties Calculated over the Last 100 ps of the Simulations[a]

property	simulation				
	X-ray	CW	HW	CU	HU
temperature (K)		300	360	300	360
[urea] (mol L^{-1})		0	0	8	8
time (ps)		500	2000	500	2000
E_{pot} (kcal mol^{-1})		−53049 (99)	−49001 (133)	−65951 (115)	−62575 (129)
rmsd (Å)[b]		2.11 (0.13)	5.45 (0.16)	2.28 (0.18)	6.74 (0.18)
density (g cm^{-3})		1.031 (0.003)	0.968 (0.003)	1.161 (0.003)	1.108 (0.003)
R_g (Å)[c]	13.55	13.78 (0.08)	14.60 (0.11)	13.97 (0.07)	15.71 (0.09)
SAA ($Å^2$)[d]	6064	6659 (84)	7508 (144)	6778 (59)	8124 (151)
helix 1	T^6−H^{18}	F^7−T^{16}	A^{11}−Q^{15}	F^7−Q^{15}	F^7−Q^{15}
helix 2	K^{27}−G^{34}	K^{27}−A^{30}	K^{27}−A^{30}	K^{27}−Q^{31}	S^{28}−A^{32}
helix 3	L^{42}−A^{46}	L^{42}−V^{45}	L^{42}−V^{45}	L^{42}−V^{45}	L^{42}−V^{45}
residues in helices	26	18	13	18	18
sheet 1	S^{50}−I^{55}	S^{50}−F^{56}	S^{50}−I^{55}	S^{50}−F^{56}	S^{50}
sheet 2	W^{71}−D^{75}	W^{71}−D^{75}	W^{71}−D^{75}	W^{71}−D^{75}	W^{71}−E^{73}
sheet 3	R^{87}−S^{91}	R^{87}−S^{91}	I^{88}−S^{91}	R^{87}−S^{91}	L^{89}−S^{91}
sheet 4	I^{96}−T^{99}	I^{96}−T^{99}	L^{95}−Y^{97}	I^{96}−T^{99}	I^{96}−Y^{97}
sheet 5	T^{107}−K^{108}	K^{108}	K^{108}	T^{107}−K^{108}	
residues in sheets	22	22	19	23	9
sec structure	48	40	32	41	27
% sec structure (X-ray)	100	83	67	85	56
protein Hb[e]	111	109 (5)	108 (6)	108 (4)	96 (4)
solvent Hb[e]		347 (10)	344 (12)	376 (11)	398 (11)
SAA charged [f]	2451	2751	2991	2731	3152
SAA polar [f]	2434	2415	2489	2532	2715
SAA hydrophobic [f]	1139	1493	2028	1516	2233

[a] Secondary structure assignments were made with the DSSP program (Kabsch & Sander, 1983). [b] Root-mean-square deviation of the backbone heavy atoms. [c] Radius of gyration. [d] Solvent-accessible area as calculated by DSSP. [e] Hydrogen bonds were defined as having a H−acceptor distance less than or equal to 2.5 Å and a donor−H−acceptor angle greater than or equal to 120°. [f] Partition of the solvent-accessible area for charged, polar, and hydrophobic residues.

the urea and the water molecules was then evaluated, and the 2023 water molecules with the highest nonbonded energies were deleted. This initial box was then equilibrated by 141 steps of steepest descent minimization followed by 25 ps of NVT-MD in which the protein was kept fixed and the temperature raised from 100 to 300 K over the first 4 ps. The positions of the protein atoms were then minimized (200 steps of steepest descent), and their velocities were assigned from a 10 ps NVT-MD run in which only the protein was allowed to move and the temperature was raised from 100 to 300 K over the initial 4 ps. The resulting box was then subjected to 1 ps of NVT-MD at 300 K in which all the atoms were allowed to move. The final system (U1) contains the protein, 4 chloride ions, 759 urea molecules, and 2683 water molecules in a box of dimensions 59.3 × 55.2 × 48.0 Å corresponding to an 8.0 M urea concentration and a density of 1.124 g·cm^{-3}.

This configuration was used as the initial point for an NPT simulation in urea at 300 K (CU) for 500 ps. After 200 ps a second NPT run (HU) was started during which the temperature was raised to 360 K during 6 ps and then continued at this temperature to yield a total trajectory of 2000 ps. Monitoring of periodic distances was also done during the simulation, and reorganizations of the box in the same manner as described were done at 631 and 1136 ps.

RESULTS

CW. This simulation, primarily intended as a control, was run in water at room temperature for 500 ps. As expected, the structure remained very similar to the crystal structure. The average values over the last 100 ps of some of the properties followed during the run are given in Table 1. From the plots of the potential energy (E_{pot}) and density in Figure 2 it can be seen that convergence has been achieved. The root-mean-square deviation of the complete backbone stabilized after 200 ps at its final value of 2.1 ± 0.1 Å as shown in Figure 3. The final radius of gyration is 13.78 Å, only

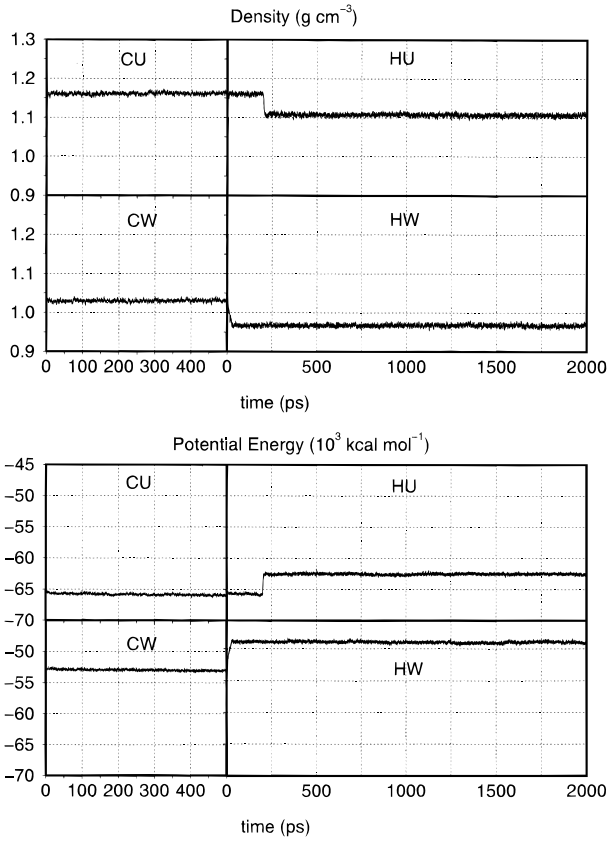

FIGURE 2: Plots of the density and potential energy *vs* time followed during all four simulations.

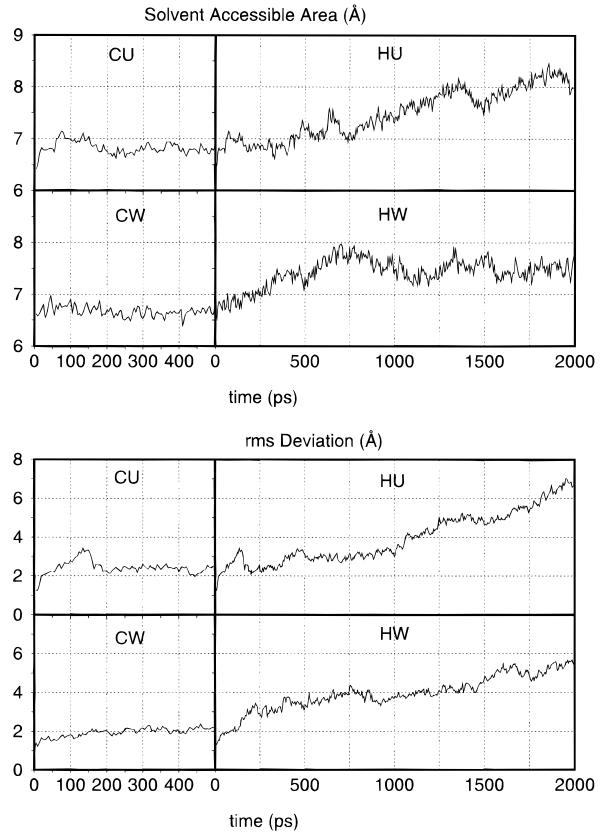

FIGURE 3: Plots of the solvent-accessible area and rms deviation *vs* time followed during all four simulations.

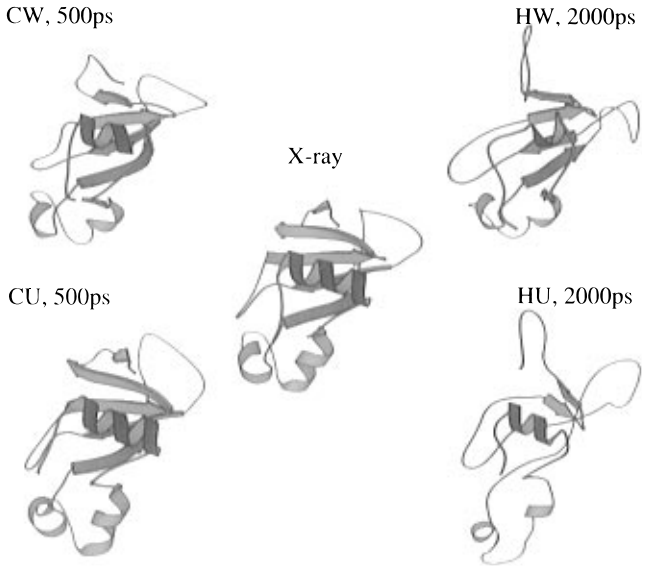

FIGURE 4: Comparison of the final structures of the simulations with the crystal structure.

2% larger than for the crystal structure, while the solvent-accessible area is 6659 Å^2, a 10% increase from the X-ray structure. This increase is mainly caused by the side chains of the surface residues extending farther into the solvent. The similarity indicated by these measures can be clearly seen in the comparisons of the final structures from the simulations with the crystal structure given in Figure 4.

HW. the next simulation was also run in water but at a higher temperature, 360 K, *ca*. 35 °C above the T_m, in order to induce unfolding. From the plots given in Figures 2 and 3 it can be seen that although the global properties, E_{pot} and density, have converged, the increase in the backbone rmsd indicates that the structure is still evolving. The structural changes in the last 750 ps of the simulation are not a large-scale motion, since both the radius of gyration and the solvent-accessible area reached their final values, 14.6 Å (8% larger than the crystal structure) and 7508 Å^2 (24% larger than the crystal structure) respectively, by 1260 ps. Instead, the structural variations in the last portion of the simulation are due to smaller localized rearrangements. The rmsd averaged over the last 100 ps reached a final value of 5.5 ± 0.2 Å, indicative of extensive unfolding. This can be seen clearly in the final structure of the simulation displayed in Figure 4, which shows that the helices and sheet are shortened from both termini and loops L2 and L4 are bent away from their positions in the crystal structure.

CU. The first simulation in 8 M aqueous urea was run at 300 K for 500 ps. For the most part the trajectory is very similar to the run in pure water (CW), with the exception of the transient increase in rmsd at 100−200 ps (Figure 3). The final value for the rmsd (2.3 ± 0.2 Å) was reached by 250 ps, while the radius of gyration and solvent-accessible area converged at 300 ps to their final values of 14.0 and 6778 Å^2, slightly higher than the corresponding values in pure water. The transient increase in rmsd seems to be correlated with a temporary separation of loops of L2 and L4 caused by a disruption of their interaction by urea molecules. Overall, the protein remained stable in its folded form in the presence of urea at room temperature, at least for the duration of the 500 ps run.

HU. The last simulation was conducted at the higher temperature (360 K) in 8 M urea in order to study the effect of urea on the unfolding transition. This run was started from the structure at 200 ps from the CU simulation. Analogous to the results of HW, the global properties (E_{pot} and density) stabilize, but the structure is still evolving as evidenced by the rmsd, R_g, and SAA. The total rmsd shows two distinct plateaus from 400 to 800 ps and from 1250 to 1600 ps, after which it increases until the end of the simulation at 2000 ps (Figure 3). Not surprisingly, the largest deviation from the X-ray structure (6.7 Å) is obtained in this simulation. As seen in the values in Table 1 and the final structure in Figure 4, there is considerable decrease in the secondary structure content, most notably in the β-sheet. The final values for the radius of gyration and the solvent-accessible area are 15.7 and 8124 Å^2, 16% and 34% larger than the values for the crystal structure. This indicates that considerable disruption of the internal packing of the protein has occurred. The decomposition of the accessible area by residue type in Table 1 shows that the hydrophobic residues account for the largest contribution to this increase in accordance with the accepted role of urea in stabilizing hydrophobic solutes in aqueous solution.

ANALYSIS AND DISCUSSION

CW, the Native Structure in Solution. In order to study the unfolding of a protein from simulations, the methodology employed must be capable of reproducing first the experimentally observed structure under conditions that favor the folded state. This also establishes a baseline for comparison of the simulations under more extreme conditions. As mentioned above, the simulation in water at room temperature produced a conformation very similar to that of the crystal structure. Even the detailed comparison of each

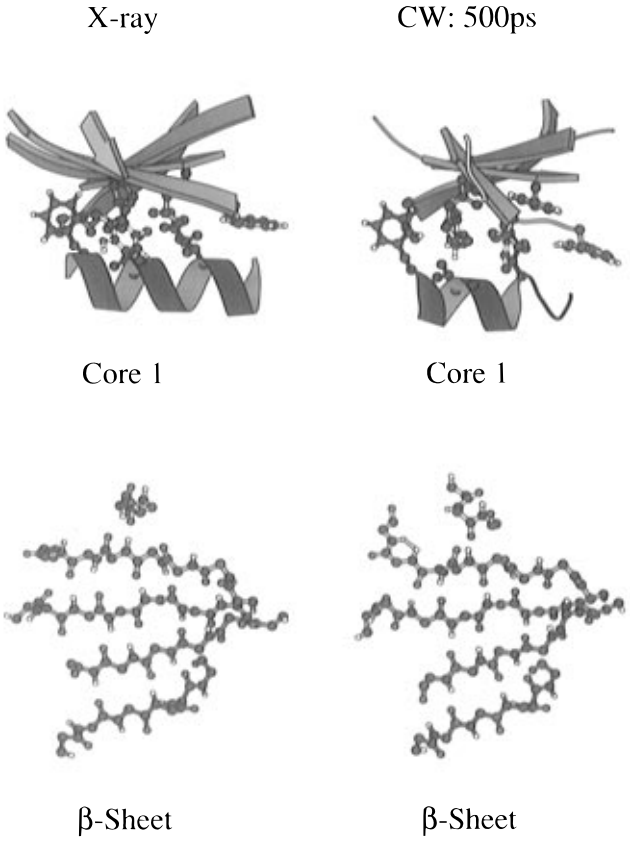

FIGURE 5: Comparison of the main hydrophobic core (top) and the backbone of the β-sheet (bottom) of the crystal structure (left) and the final structure of simulation CW (right).

element of secondary structure does not reveal substantial changes. Two such comparisons, for the major hydrophobic core (C1) and the β-sheet, are shown in Figure 5. The most notable changes from the crystal structure are the fraying at the termini of H1, the C-terminal Y^{17} and H^{18} showing significant population (55% and 10% over the last 100 ps, respectively) of π-helix, and the loss of the last two residues at the C-terminus of H2 and the N-terminus of H3.

The comparison of the final state of the simulation with the structure in solution as determined by NMR shows good qualitative agreement. Most importantly, the backbone rmsd is stable from about 200 ps at its final value of 2.1 ± 0.1 Å, a value clearly comparable to the 1.8 Å deviation obtained between the average NMR and the crystal structures (Bycroft *et al*., 1991). The backbone angles are similar to the crystal, with the largest differences in the loops; the rmsd of the angles of residues in secondary structures is 21°. Also, just as in the crystal and the NMR structures, residues H^{18}, N^{58}, N^{77}, W^{94}, H^{102}, and Y^{103} have positive values of the backbone dihedral ϕ. Additional support for this structure can be gathered from the protection to isotopic exchange. Forty-four of the 45 amide protons whose protection to H/D exchange was measured (Clarke *et al*., 1993; Perrett *et al*., 1995) show high populations (at least 70%) of intramolecular hydrogen bonds. The sole exception being T^{26} which was partially inaccessible to solvent during the simulation, its solvent-accessible area is only 37% of the value for an extended ATA peptide. The side chain of this residue is commonly hydrogen bonded to the carbonyl oxygen of G^{52} for 50% of the time during the last 100 ps of the simulation bringing the N-termini of S1 and H2 together. This agreement with available experimental data confirms the adequacy of the simulation protocols and the force field and allows us to discuss the results of the remaining simulations with confidence.

HW, the Pathway of Thermal Unfolding. The global structure of barnase changes in several discrete stages during the simulation. In the backbone rmsd, an initial sharp increase from 25 to 130 ps is followed by a more gradual increase that continues until it reaches a maximum at 750 ps, where a decrease actually ensues until 930 ps. This is followed by another gradual increase that lasts until 1700 ps followed by a decrease during 50 ps and then increases until the end of the simulation at 2000 ps. The solvent-accessible area has a very sharp increase from 0 to 750 ps, follow by a decrease until 1250 ps, and two different plateaus from 1250 to 1750 ps and from 1750 to 2000 ps. These fluctuations seem to be correlated to motions in the β-sheet.

Some of the structural changes can be followed in the instantaneous structures shown in Figure 6, particularly the separation between L2 and L4 at the left end of each structure which starts early in the simulation, near 150 ps. The unraveling of H1 starts near 200 ps from the C-terminus and 300 ps for the N-terminus. The C-terminus re-forms rapidly at 400 ps, while the N-terminus is permanently lost by 600 ps. A similar trend is observed for H2 where both the N- and C-termini are seen broken at 400 ps, but the C-terminus re-forms by 800 ps and remains until 1220 ps. After this time residues A^{32} and L^{33} form stable ($i \rightarrow i - 3$) hydrogen bonds to E^{29} and A^{30}, respectively, that remain formed until the end of the simulation at 2000 ps but their backbone angles are distorted enough that they are not classified as helical by the DSSP program. The central residues of the helix ($K^{27}-A^{30}$) remain helical until the end of the simulation. L3, the loop containing the binding site remains, albeit somewhat distorted, wrapped around the edge of the β-sheet until 1400 ps when it finally separates, opening up the third core to solvent.

As seen in Figures 6 and 7, the β-sheet shows clear signs of fraying at the ends, particularly in the regions of interaction between S3 and S4 and between S4 and S5. By the end of the simulation only the hydrogen bond between the K^{98} amide hydrogen and the K^{107} carbonyl is left from the latter. The S1−S2 contact is diminished from early in the simulation due to the sporadic interaction formed between the first two residues of S1 (S^{50}, I^{51}) with a small portion of non-native β-sheet formed just before the N-terminus of H2 (Y^{24}, I^{25}). This interaction is firmly established by 1600 ps. The hydrophobic cores also evolve considerably as the simulation progresses. Figure 7 shows several snapshots of residues forming the main hydrophobic core. The left-hand portion of C1 lost most of its integrity, with F^{7} and A^{11} drifting away from the rest of the core by 200 ps followed by V^{10} at 800 ps. These three residues are either at the N-terminus of the major α-helix or in its second turn and become exposed to solvent when its N-terminus unfolds. V^{10} and A^{11} return closer to the rest of the core at 1100 ps, while its central residues (W^{94}, Y^{90}, I^{88}, and I^{76}) remain in relatively constant orientations throughout the simulation. This rearrangement of the major hydrophobic core is the cause for the large rise in solvent-accessible surface area observed around 750 ps. Figure 8 shows that the residues at the top of hydrophobic core C2 (I^{25} and I^{51}) drift away from the remainder of the core rapidly; their positions stabilized by 125 ps. Early in the simulation L^{33} and W^{35} separate from the core synchro-

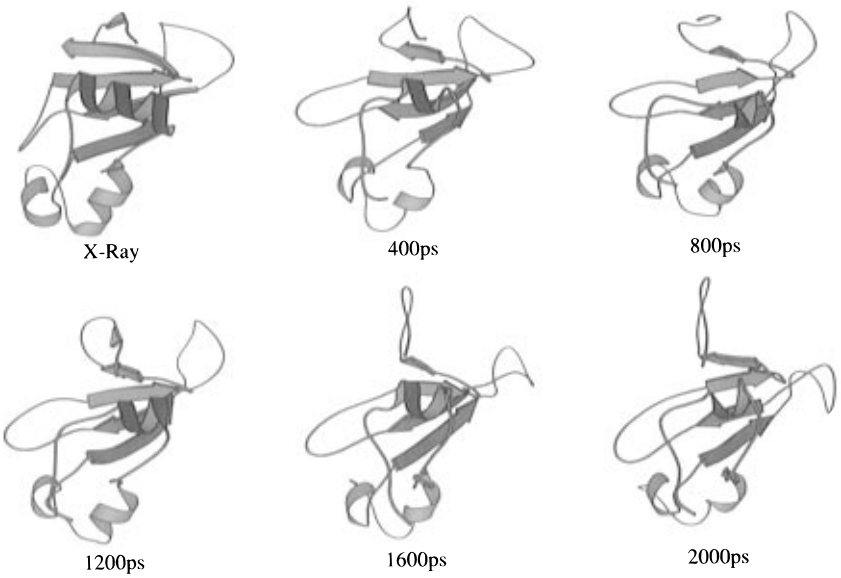

FIGURE 6: Ribbon representations of some instantaneous structures along the HW simulation.

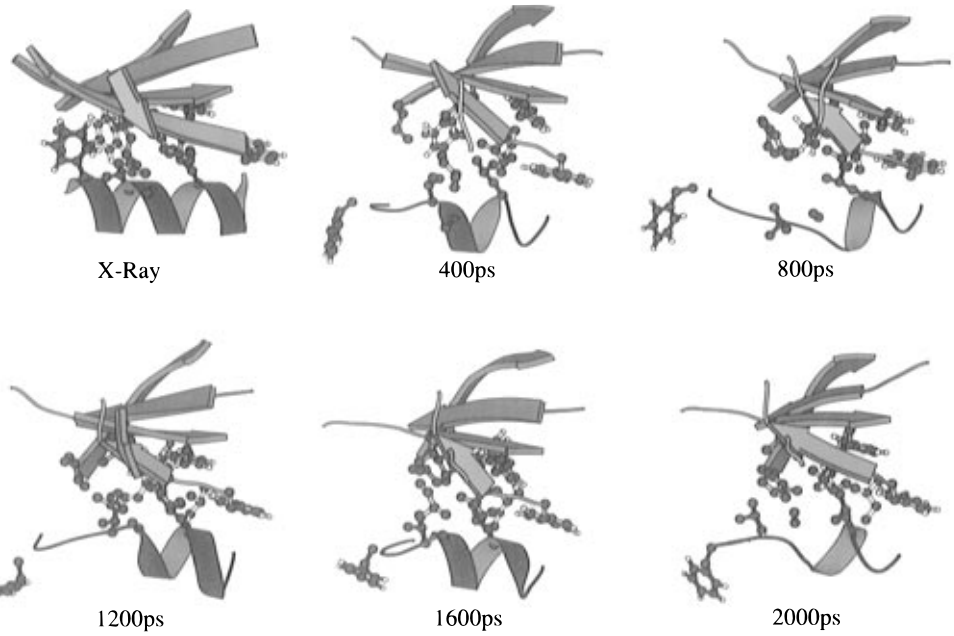

FIGURE 7: Changes of the first hydrophobic core along the HW simulation.

nously to the unfolding of the N-terminus of H2, and although they return to positions closer to the remainder of the core at 800 ps, W^{35} maintains considerable exposure to solvent. The center of this core, formed by A^{30}, L^{42}, and V^{45}, seems to stay in a relatively constant position. C3, on the other hand, is affected the least in the initial part of the simulation as seen in Figure 9 and retains most of its integrity until Y^{103} becomes exposed to solvent at 920 ps, followed by L^{63} and F^{106} at 1200 ps. Although the residues at the center of the core, F^{56}, W^{71}, L^{89}, and Y^{97}, remain closely associated until the end of the simulation, the movement of the other residues in the core increases their exposure to the solvent.

The progress of the simulation can be compared to the experimental results of the site-directed mutagenesis method of Fersht and co-workers [for a review see Fersht (1993)]. The data of over 120 mutants are represented in the bar chart in Figure 10 in which the sequence is separated into the different regions of secondary structure along the abscissa, while the height of each bar is proportional to an average value of ϕ^X, the "progress along the reaction coordinate". The different bars for each segment represent its degree of unfolding for the native (N), transition state (TS), intermediate (I), or unfolded (U) states from the back to the front of the graph. The experiments measure directly the effect of each mutation on the rate and the equilibrium constants of folding or unfolding and are converted to the relative changes in free energies of the transition state or intermediate. The free energy changes are then combined into the ϕ^X values according to eq 1 (Fersht *et al*., 1992). In this equation the

$$\phi^X = \frac{\Delta\Delta G^{X-U}}{\Delta\Delta G^{F-U}} = \frac{\Delta G^X - \Delta G^U}{\Delta G^F - \Delta G^U} \qquad (1)$$

superscript X refers to either the transition state or the intermediate, while F and U designate the folded and unfolded states, respectively.

The observed trends for ϕ^X indicate that, in progressing to the transition state, the largest changes occur for the loops L1, L2, and L4, the N-termini of H1 and H2, the edges of

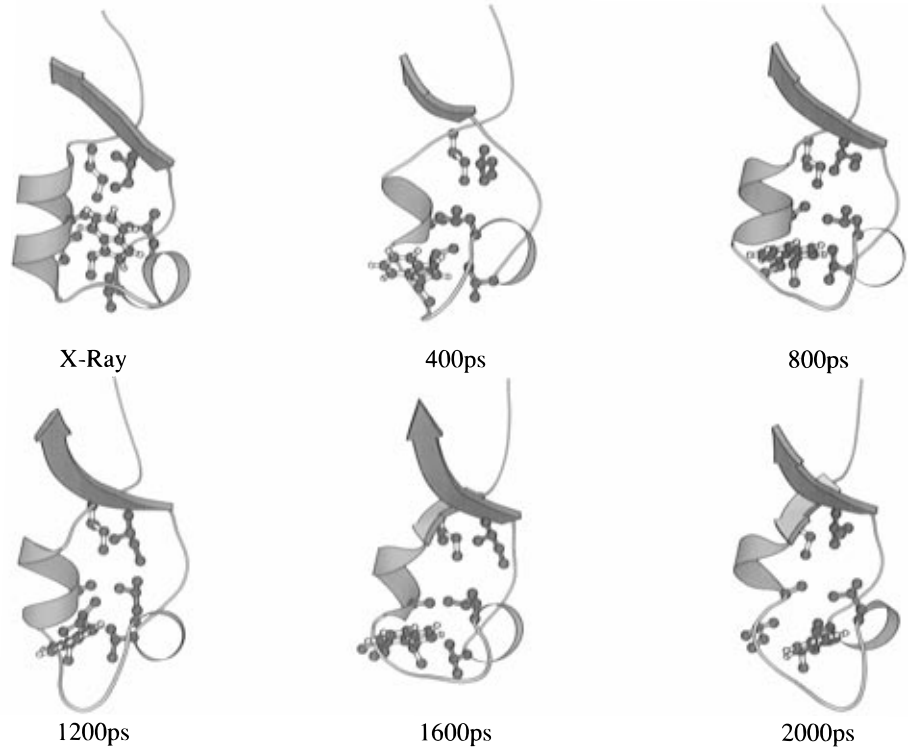

FIGURE 8: Changes of the second hydrophobic core along the HW simulation.

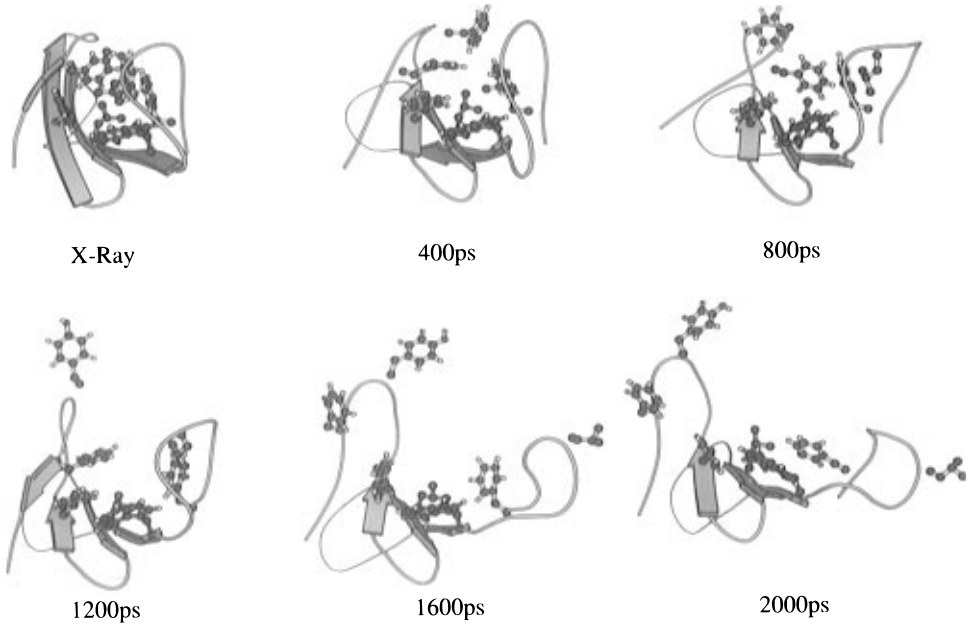

FIGURE 9: Changes of the third hydrophobic core along the HW simulation.

the β-sheet, and the edges of the hydrophobic cores. These changes continue in approximately the same direction until the intermediate is reached. The comparison to the simulation can only be qualitative as the experimental values represent the ratio of the effect of the mutation on the changes in free energy in the transition state or the intermediate to the change of free energy of the folded structure, while the simulation results directly reflect the structure. The agreement in patterns nonetheless is quite good, and it suggests that the final structure in the high temperature, aqueous simulation HW has passed through the transition state and is consistent with the intermediate stage. Even though the rmsd is still changing, the solvent-accessible area and other properties seem to have reached a plateau.

A more precise location of the transition state was sought from the simulation. A simple search along all the structures prior to the intermediate for a match of the degree of unfolding as suggested by the experimental values of ϕ^{TS} has several drawbacks. This comparison can be only qualitative since the experimental values reflect changes in energy rather than structure. Furthermore, given that the trajectory generated a continuum of structures that span from the initial folded state to the intermediate, it seems unlikely that a single, unique structure or family of structures would contain the desired degrees of unfolding. A less ambiguous, and more objective, method of locating the transition state should be based on its properties. A transition state is formally a free energy maximum along the reaction path.

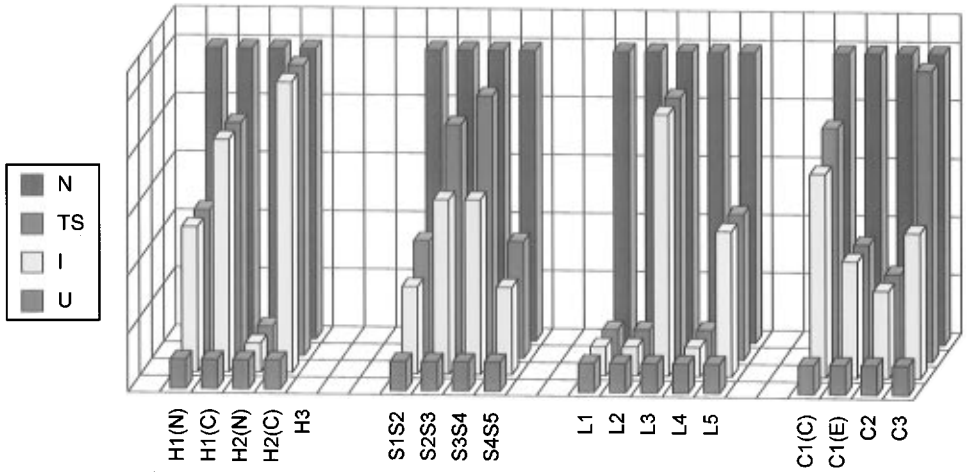

FIGURE 10: Brief summary of the experimentally measured degrees of folding as obtained by Fersht's protein engineering method. The different sections of secondary structure in the abscissa are assigned bars of height proportional to their value of ϕ in the native state, transition state, or unfolded state.

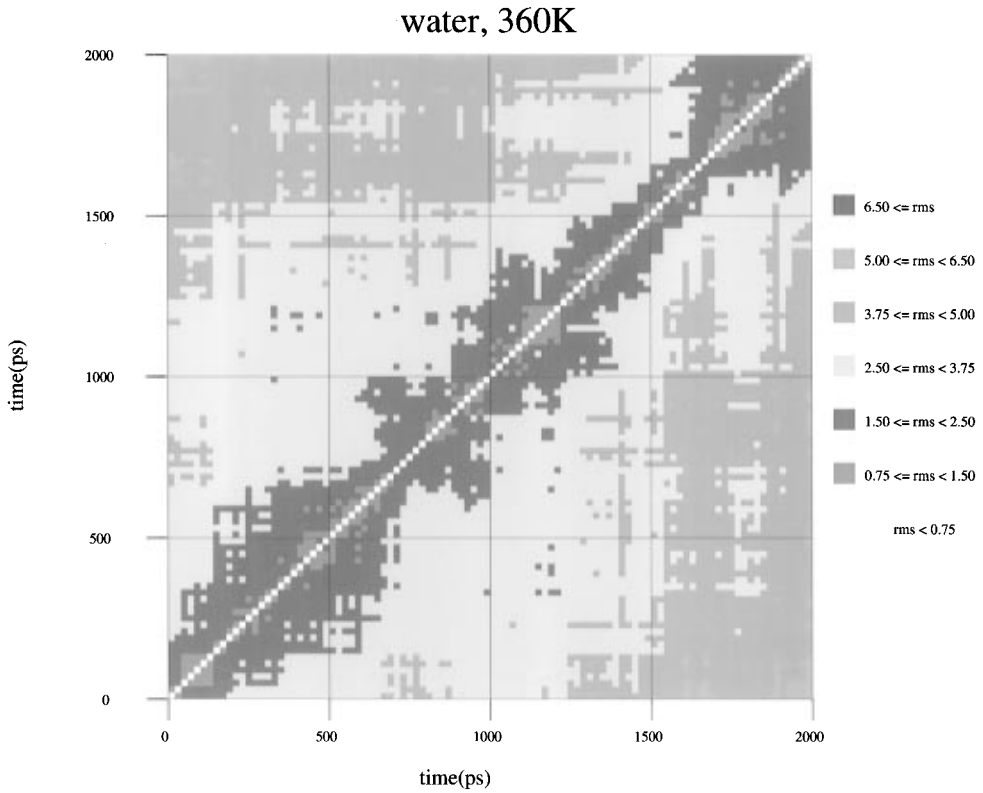

FIGURE 11: Two-dimensional map for the simulation in water at 360 K, HW. Each structure along the simulation is compared to every other and its rmsd is represented as a color-coded block.

The energies calculated in the course of the simulation are not useful in this regard since they are not free energies and are large numbers with very significant fluctuations dominated by the nonbonded solvent−solvent interactions. On the other hand, being an energy maximum the population should be the lowest at this state. It would be logical then to expect that its structure differs the most from those around it. A comparison among all the structures from the simulation is shown as a 2-D rmsd map in Figure 11. Several regions in which entire families of structures around the diagonal are similar, separated by narrow constrictions, can be distinguished. The first major constriction, and a likely candidate for the transition state, can be found at 750 ps. Further support for this assignment can be gathered from the behavior of the backbone rmsd and the solvent-accessible area, which increase and reach maxima at this stage, followed by decreases. Such a rapid change in properties could be expected after a transition state is reached. Following a similar logic, a stable intermediate should appear in the 2-D rmsd map as a wide area of closely related structures. The region between 1680 and 2000 ps in Figure 11 fits these characteristics.

Once the transition state and the intermediate have been located, quantitative approaches to estimate the observed ϕ^X values were sought. Experimentally these are measured as the effects of the mutations on the different equilibrium or rate constants of folding or unfolding and are related to the changes in free energies and degrees of unfolding using eq 1. These free energy differences could, in principle, be calculated through statistical perturbation theory in the

context of either Monte Carlo or molecular dynamics simulations. In practice, however, the computational demands of such calculations for even a small set of mutations are very large, and an additional complication is the lack of a precise stable geometry for the transition state. A simplified approach can be developed on the basis of the linear response method first introduced by Åqvist to calculate free energies of binding (Åqvist *et al*., 1994). Carlson and Jorgensen (1995) extended the method to compute free energies of hydration of organic molecules. According to the latter formulation the difference in free energies of hydration for two molecules can be approximated by the equation:

$$\Delta G_{\mathrm{solv}} = \alpha \Delta E_{\mathrm{LJ}} + \beta \Delta E_{\mathrm{El}} + \gamma \Delta \mathrm{SAA} \qquad (2)$$

where ΔSAA is the difference in the solvent-accessible area between the two species and ΔE_{LJ} and ΔE_{El} are the differences in the solute−solvent Lennard-Jones nonbonded and electrostatic energies, respectively. The α, β, and γ coefficients are adjustable parameters that are determined on a case by case basis. In a strict application of this equation the energy differences are calculated through separate MC or MD simulations for both species. In the present context, a similar equation could be used in place of the free energy change due to a mutation in either the transition state, the intermediate, or the folded state. Combination of eqs 1 and 2 suggests eq 3, where α, β, and γ are

$$\phi^{\mathrm{X}} = \frac{\alpha \Delta E^{\mathrm{X}}_{\mathrm{LJ}} + \beta \Delta E^{\mathrm{X}}_{\mathrm{El}} + \gamma \mathrm{SAA}^{\mathrm{X}} + G_0}{\alpha \Delta E^{\mathrm{F}}_{\mathrm{LJ}} + \beta \Delta E^{\mathrm{F}}_{\mathrm{El}} + \gamma \mathrm{SAA}^{\mathrm{F}} + G_1} \qquad (3)$$

still adjustable parameters. The energy differences ΔE_{LJ} and ΔE_{El} are the differences between the corresponding interaction energies of the wild-type and mutant forms of the residue being examined with the rest of the protein and the solvent, and SAA is the solvent-accessible area of the wild-type form of the same residue. The latter value was utilized instead of the difference in areas since the structures generated in the simulations and used in these calculations did not allow relaxation of the solvent and the rest of the protein after substituting the mutant form of each residue.

Most of the numerical values for the different terms needed were evaluated from the structures saved during the high temperature aqueous simulation (HW) as follows. The transition state was represented by the segment of the trajectory between 725 and 775 ps while the structures from 1950 and 2000 ps were used for the intermediate, as defined above. For each of the structures the electrostatic and Lennard-Jones interaction energies between a given residue and the rest of the system were evaluated first in its wild-type and then in the mutant form by substituting the appropriate nonbonded parameters q, σ, and ϵ without doing any changes in geometry. That is, the coordinates for the backbone atoms and C^{β} were utilized directly to represent the mutant structure (alanine in all cases). The remaining side-chain atoms were kept at their positions but were effectively annihilated since their interactions energies are null when the parameters are set to zero. The corresponding solvent-accessible areas were also evaluated at this time. The values thus obtained were averaged over each trajectory fragment separately. Since barnase is not stable under the conditions of the HW simulation, the values needed for the folded state were obtained in a similar fashion by averaging over the last 50 ps of the aqueous simulation at room temperature (CW). The difference in the conditions of the simulations is the origin of the need to have different values for G_0 and G_1 in eq 3, although they are constrained to be constant for mutations of the same residue type.

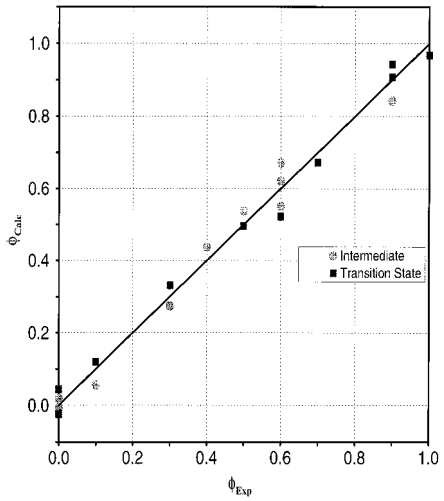

FIGURE 12: Comparison of experimental and calculated values for ϕ. The calculated values were obtained from eq 3 with the adjustable parameters in Table 2.

Table 2: Parameters for Eq 3 Obtained by Fitting to Experimental ϕ^{X}

	N/A	S/A	T/A	Y/A
α		0.480		
β		0.163		
γ		0.0367		
G_0	−7.108	1.061	2.382	−3.596
G_1	−5.201	7.717	3.812	4.857

The parameters α, β, γ, G_0, and G_1 were then fitted to a subset of the experimentally measured ϕ's. All the data selected were the result of mutations from polar residues (N, T, S, and Y) into alanines in order to simplify the choice of geometry for the final states. Mutations with charged groups were not included since the simplified calculations done here would not be able to correctly represent the significant rearrangements that typically occur upon changes of charge states. The results of the correlation to the experimental values of ϕ^{int} and ϕ^{TS} for the mutations N5A, N23A, N58A, N77A, T6A, T16A, S91A, S92A, Y13A, and Y17A (Matouschek *et al.*, 1990; Horovitz *et al*., 1991; Matouschek *et al*., 1992) can be seen in Figure 12, and the values obtained for the adjustable parameters are listed in Table 2. Although the results can only be considered qualitative due to the limited number of points represented, the agreement shown by the rmsd of 0.04 kcal mol^{-1} supports the assignments of the transition state and intermediate structures done above. It should be noted that the parameters α, β, and γ were kept the same for all the mutations selected, but G_0 and G_1 are unique to each residue type. The statistical significance of the (G_0, G_1) pairs obtained in this fashion, particularly those for mutations of Ser, Thr, and Tyr, is much lower than for the α, β, and γ parameters which depend on all 20 data points.

CU, Effect of Urea on the Folded Structure. Unfolding by urea at room temperature was not achieved in this simulation, presumably due to the short total time (500 ps)

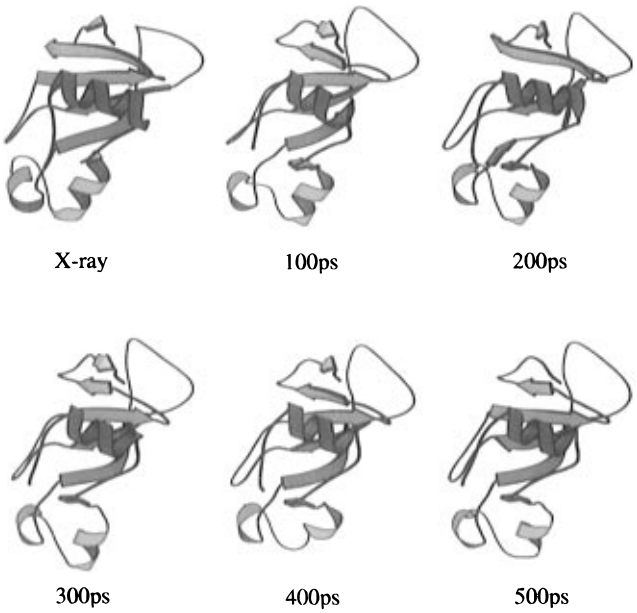

FIGURE 13: Instantaneous structures along the room temperature simulation in urea, CU.

of the simulation. From the pseudo-first-order rate constant measured under these conditions (Horovitz & Fersht, 1992), a half-life of 1.1 s can be calculated. Although there is not necessarily a direct correlation of an experimentally measured rate to the time course of a single MD trajectory, the computer time needed for simulation of unfolding at this temperature is clearly much longer than presently possible. Nevertheless, it is interesting to observe the changes in the structure as an indication of the initial stages of unfolding in urea. The most noticeable effects, evident in Figures 4 and 13, are the transient separation of L2 and L4 and the deformation of L3 which opens up the binding site to solvent. Another effect that is not readily apparent in the ribbon diagrams is the large proportion of 3_{10}-helical conformation in the last 100 ps for H2; three of the four remaining hydrogen bonds are between i and $i - 3$ residues.

HU, Thermal Unfolding in the Presence of Urea. The behavior of the elements of secondary structure is quite different in the presence of hot urea and can be followed in the ribbon diagrams in Figure 14 and the schematic representations of Figure 15. In the major α-helix, H1, the terminal residues T^6 and H^{18} were not found to be helical by the DSSP program from the very beginning of the simulation, even when the $V^{10} \rightarrow T^6$ hydrogen bond was present. This hydrogen bond was finally lost at 1620 ps, at which point F^7 becomes frayed. At the C-terminus, the hydrogen bond between $Y^{17}$and Y^{13} disappears earlier at 70 ps. At this time T^{16} becomes increasingly frayed and after 900 ps forms part of the helix only occasionally. The N-terminus of H2 follows a similar unfolding path. The first residue, T^{26}, forms a stable hydrogen bond with A^{30} but is distorted enough not to be classified as helical. This hydrogen bond is lost at 375 ps, and at this point K^{27} is not considered helical by DSSP although it forms a stable hydrogen bond with Q^{31} until the end of the simulation. The central part of this helix, formed by S^{28}, E^{29}, A^{30}, Q^{31}, and A^{32}, undergoes a series of cooperative transitions, first from α- to 3_{10}-helical at 905 ps. The helix is then broken at 1025 ps, after which it is re-formed again as a short 3_{10}-helix at 1100 ps. It then undergoes a transition to re-form the α-helix

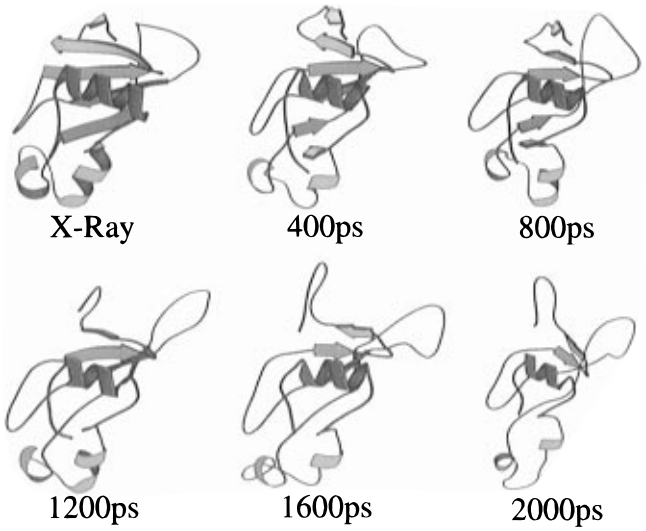

FIGURE 14: Instantaneous structures along the high temperature simulation in urea, HU.

at 1180 ps. This intermediacy of the 3_{10} conformation was also observed in previous simulations of unfolding of the S-peptide of ribonuclease A and apomyoglobin (Tirado-Rives & Jorgensen, 1991, 1993). The center of H2 is stable for the rest of the simulation, other than for a short period from 1325 to 1490 ps where the conformation of the helix is 3_{10} rather than α. The third helix is stable during the entire simulation with the exception of the very first and last residue in both termini (N^{41} and A^{46}) not being considered helical by DSSP despite having the requisite ($i \rightarrow i - 4$) hydrogen bonds.

The behavior of the β-sheet is affected by urea to a much larger extent. The N-terminus of S1 starts from the very beginning of the simulation to form contacts with a non-native short strand of β-sheet formed by $Y^{24}$and I^{25} preceding the N-terminus of H2. The ($G^{53} \rightarrow E^{73}$) hydrogen bond breaks shortly thereafter at 175 ps. The residues at the C-terminus follow suit; the ($E^{73} \rightarrow D^{54}$) hydrogen bond starts oscillating at approximately 410 ps and ultimately breaks at 520 ps. These DSSP assignments do not classify these residues as β-sheet at 470 ps, after which time two "bridge" residues (S^{50} and F^{56}) are all that remain of this strand with the exception of a short non-native strand formed by F^{56}, S^{57}, and N^{58} that appears sporadically from 1095 to 1270, 1395 to 1420, and 1620 to 1800 ps. The final hydrogen bond of this strand ($D^{75} \rightarrow I^{51}$) disappears at 1950 ps.

The second and third strands display a very interesting phenomenon which is better illustrated in Figure 16. The first third of the simulation appears not to affect these two strands except for some slight fraying at the termini of both strands. After 700 ps the residues T^{79}, S^{80}, G^{81}, and F^{82} at the center of the loop connecting these strands (L4) form a β-turn. The residues at the arms connecting the turn to the two strands of β-sheet start interacting, sporadically at first and then more frequently at 1080 ps to form rather extended, non-native strands S2 and S3. This extended β-sheet starts weakening at 1315 ps until 1770 ps when the central β-turn breaks, followed by a rapid collapse of the extended S2 and S3 from the turn outward. By the end of the simulation only a single bridge residue (W^{71}, which interacts with S1) remains of S2 and the three residues at the C-terminus of S3, L^{89}, Y^{90}, and S^{91} which interact with S4. This behavior is more logical when considered in the opposite, folding sense, that

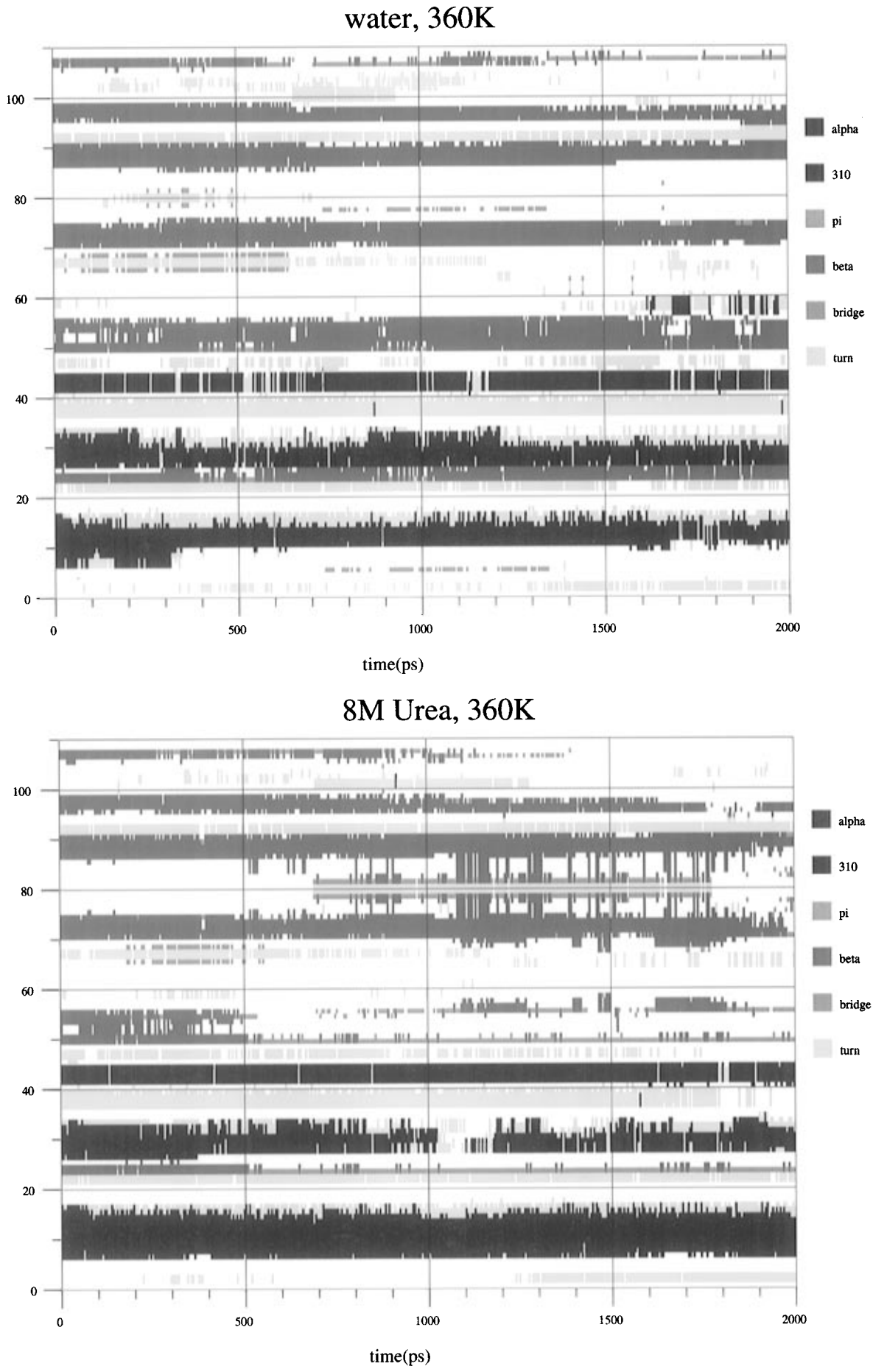

FIGURE 15: Schematic representation of secondary structure elements during the high temperature simulations, HW (top) and HU (bottom). The assignments were made with the DSSP program.

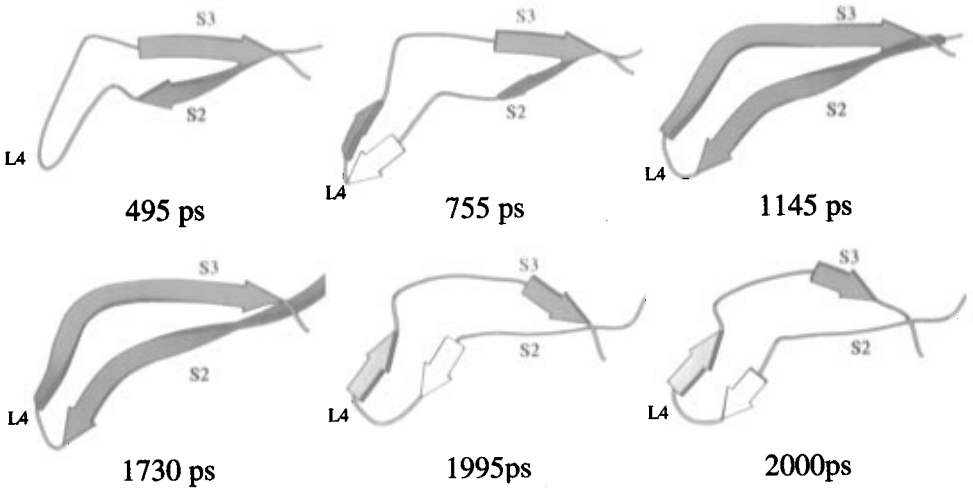

FIGURE 16: Some instantaneous structures of the S2−L4−S3 region along the high temperature unfolding in urea, HU.

is, as a mechanism for the formation of two strands of β-sheet separated by a loop. From this perspective the central β-turn serves as a nucleation site for a long, but unstable β-sheet that acts as a scaffold to hold together two long, unstable strands until the more stable contacts at the end farther from the turn are established. The unneeded scaffold then melts into a more stable "loop" conformation.

The last two strands of the β-sheet seem to follow a more straightforward pattern of unfolding. The residues at the termini of both strands (W^{94}, L^{95}, T^{100}, and D^{101} in S4 and F^{106} in S5) are again not detected as β-sheet by DSSP. The hydrogen bonds between T^{99} and R^{87} start to fray at 625 ps and finally disappear by 1550 ps. The interaction between Y^{97} in S4 and L^{89} in S3 is stable except for the period between 1870 and 1950 ps. The remaining interaction with S3, the single hydrogen bond from L^{95} to S^{91}, is well maintained. The interaction between S4 and S5, comprised mostly of the double hydrogen bond between $T^{107}$and K^{98}, disappears between 1100 and 1400 ps.

The behavior for the hydrophobic cores parallels the changes for the secondary structure elements that comprise them. The main core, C1, is formed by the packing of H1 onto the β-sheet, and although the edges of the sheet unfold, the residues involved in contacts with the helix are still in similar positions. The largest changes occur when L^{20} becomes accessible to the solvent at 780 ps, followed by I^{96} at 930 and finally I^{88} at 1600 ps. The central part of this core remains closely associated in this simulation. C2, formed by the two smaller helices H2 and H3 packing against S1, seems to be mostly unfolded. Most of the residues that comprise it gain accessible area as they fragment into three smaller local associations: W^{35} (in L2) with L^{33} (in H2), L^{42} with V^{45} (both of them in H3), and I^{51} from S1 with I^{25} at L1. The hydrophobic core containing the binding site, C3, is formed by the docking of L3 on the back of the β-sheet and also seems to follow these elements. Initially L^{63} (in L3) separates slightly from the rest of the core becoming accessible to the solvent, followed by the slow drift of Y^{103} (in L5) and F^{106} (in S5) that starts at 900 ps and seems to consolidate a separate hydrophobic contact with L^{63} at 1050 ps. At this time L3 separates from the sheet and L^{63} moves away and is replaced by F^{56}. All three of these aromatic residues then drift away from each other as the contact between S4 and S5 disappears.

Using the same criteria as for the thermal unfolding above, the transition state for this simulation can be located at approximately 1300 ps. Interestingly enough, even when it appears later in time than in the high temperature aqueous simulation it contains more secondary structure, 20 residues in α-helices and 17 in β-sheets (approximately 77% of the crystal structure), as compared to 12 residues in α-helices and 21 residues in β-sheets (70% of the crystal) for the aqueous simulation. The transition state is then earlier along the reaction coordinate, as could be expected for a perturbation that destabilizes the native structure to bring it closer in energy to the transition state. This behavior, in agreement with the Hammond postulate, has been experimentally observed in the *m*-values (the slope of the logarithmic plot of the rate of unfolding log k_u *vs* denaturant concentration) for the urea unfolding of barnase (Matouschek & Fersht, 1993).

Overall, although the structure obtained at the end of this simulation is more unfolded than that obtained in the thermal unfolding simulation (HW), it still contains appreciable amounts of structure and cannot be considered to be completely unfolded. It is unclear how long the simulation would have to proceed before reaching such a state. Still, much information can be learned from these simulations, particularly from the comparison with the simulations in the absence of urea. The β-sheet is affected by urea to a much larger extent than the α-helices. This is particularly noticeable in the major α-helix H1, which is less unfolded in the 8 M urea simulation. Such a differential effect is not altogether unexpected. Experimentally, poly(L-glutamic acid) has been found to be largely helical at this concentration (Hermans, 1966), and block copolymers of poly(L,D-glutamic acid) and poly(D-alanine), leucine, or phenylalanine showed only a moderate decrease of helicity from 92% to 83% upon treatment with 8 M urea (Gratzer & Doty, 1963; Auer & Doty, 1966). More recent NMR studies by Kemp and coworkers have found that urea in concentrations up to 8 M has only a small effect on the helical content of alanine-based polypeptides (Allen, 1993).

The correlations with experiment for this simulation are far fewer, since the intermediate state is not stable at high urea concentrations, but NMR studies of the urea unfolded state show ^1H chemical shift deviations from random coil values in the regions from $G^9$to Y^{17} (H1), R^{69} to I^{76} (S2),

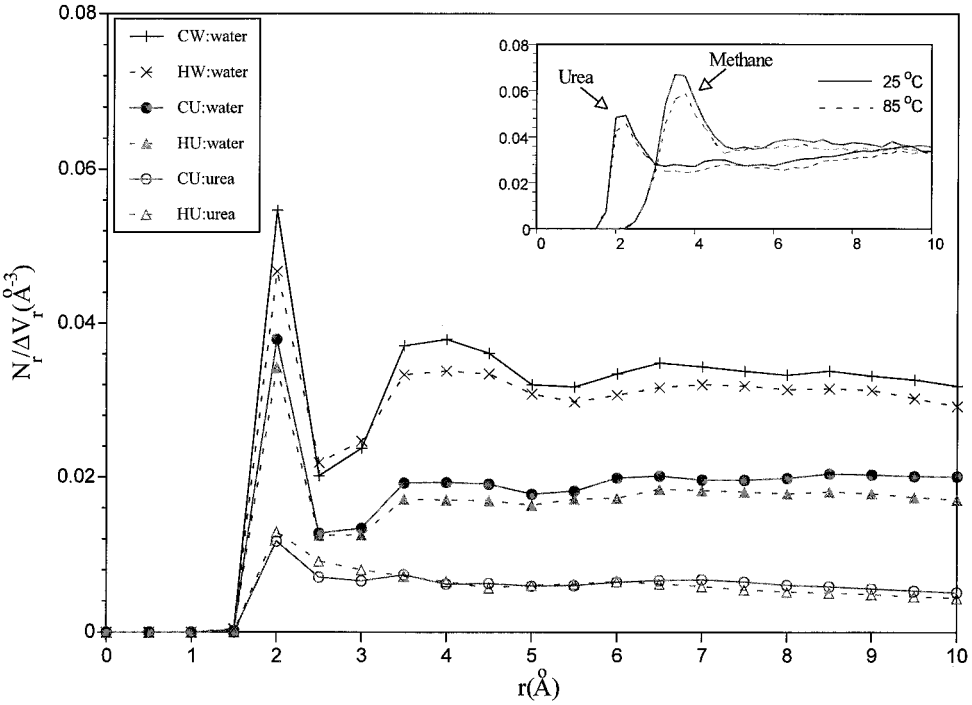

FIGURE 17: Protein−water and protein−urea radial distribution functions for all four barnase simulations. The inset shows the urea−water and methane−water rdfs from separate simulations of a single urea or methane molecule in water for comparison.

and R^{87} to I^{96} (S3 and S4) (Arcus *et al.*, 1995). As can be seen in Table 1 and Figure 15, all these regions have residual secondary structure at the end of the simulation. The chemical shift deviations were considered to be weak evidence for structured conformations since the corresponding effects in the ^{13}C and ^{15}N NMR could not be unequivocally located. The acid unfolded form, on the other hand, seems to contain more structure. In particular, the positive deviations in the $^{13}C^{\alpha}$ chemical shift for the region from S^{92} to I^{96} seem to point to non-native α-helix.

Evidence from optical spectroscopy is also consistent with the results presented here. Using the perturbation of DMSO in the spectra, Pace determined that the solvent-accessibility of tryptophan and tyrosine residues doubles upon unfolding in 8 M urea (Pace *et al.*, 1992). The average of the sum of the solvent-accessible areas of these residues over the last 100 ps of the simulations yields a result of 821 $Å^2$ for HU as compared to 394 $Å^2$ for CW, a value very close to the 384 $Å^2$ of the crystal structure.

The Mode of Action of Urea. The comparison of all four simulations can shed light on the mechanism of protein denaturation by urea. In agreement with conventional wisdom, the results of the simulations indicate that urea stabilizes hydrophobic residues that become exposed to solvent upon unfolding. The partition of solvent-accessible areas into contributions from charged, polar, and hydrophobic residues for the crystal structure and the averages over the last 100 ps of each simulation is given in Table 1. This shows that the simulations in urea have noticeably more solvent-exposed hydrophobic area than their aqueous counterparts.

More insight into the role of the solvent can be obtained by examining the distribution of solvent molecules around the solute protein. This is normally done in simulations through the use of radial distribution functions, $g(r)$, which represent the distribution of atom pairs, normalized for the available volume of the spherical shell around the atoms, as a function of their distance. For a molecule of the size of barnase (in the OPLS representation it contains 1160 atoms) this is not practical. Instead, a molecular radial distribution function (rdf) was calculated in which the minimum distance between any pair of atoms in each molecule pair is used as the basis for counting the interactions. The size of the protein also precludes the normalization by the available volume as a function of the distance through a spherical shell approximation, so it must be calculated by numerical integration. The results of these calculations over the last 100 ps of all simulations are given in Figure 17. Note that the values plotted along the ordinate are the number densities calculated directly from the simulations and are not normalized by the numbers of urea or water molecules. The equivalent results from simulations of a single urea or methane molecule in water are also shown for comparison. A united-atom model was used for methane as for the aliphatic CH_n groups of the protein.

The first thing to be noticed is that the curves obtained in pure water show a sharp maximum at about 2 Å followed by a broad, wide second peak at 3.5−4.0 Å, while the methane−water rdf has a single broad peak at 3.5 Å. The solvation of the protein by water follows the pattern characteristic of a polar species, which is expected given the distribution of charged groups on the surface. The short (1.5−2.5 Å) contacts reflect protein−water hydrogen bonds. The first solvation shell, as defined by adding the number of solvent molecules up to 3 Å, a common point close to the first minimum in all of the curves, contains 338 water molecules in the aqueous simulation at room temperature, CW, and 347 in the higher temperature run, HW. In the urea runs, the first solvation shell contains 91 urea and 220 water molecules for CU, while in the higher temperature run, HU, it contains 120 ureas and 223 waters. The number of urea molecules in direct contact with the protein is larger than could be statistically expected from the composition of the system. The entire box contains 759 urea molecules and

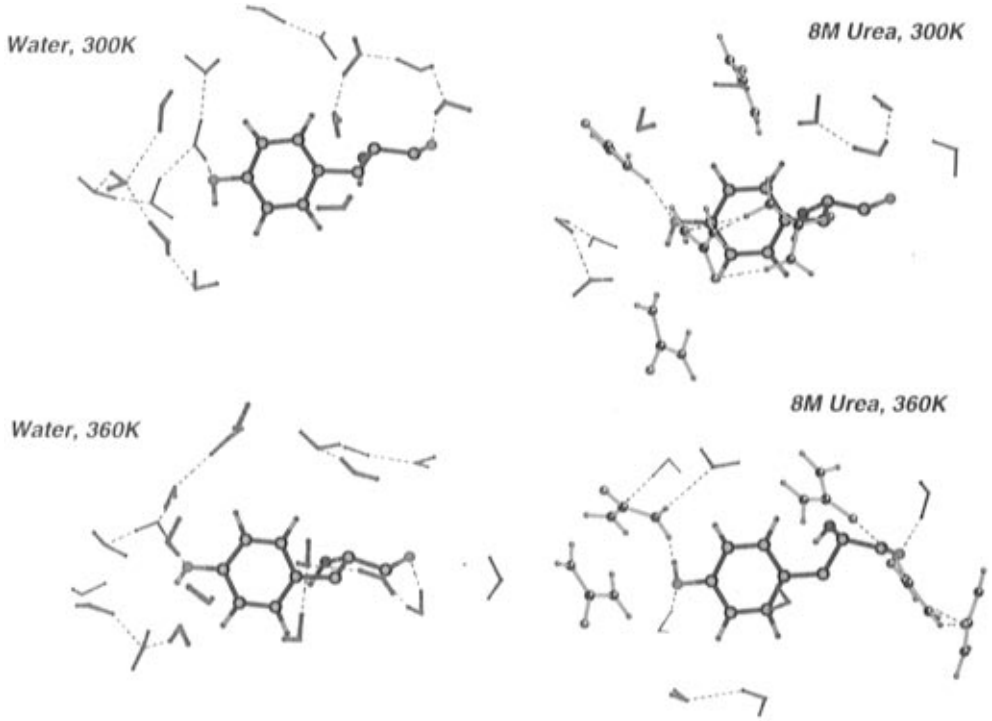

FIGURE 18: Solvation spheres around Y^{17} in the final structures of all four simulations. Different representations have been used for the protein residue (dark ball and stick), urea (lighter ball and stick), and water (depth-cued wire frame).

2683 waters, to give a ratio of urea:water of 0.28, while the corresponding ratio in the first shell is 0.41 at room temperature and 0.54 at 360 K. The mode of interaction is made clear by analysis of the hydrogen bonding. A hydrogen bond was defined by any X−H···Y distance below 2.5 Å for X, Y = N or O. For the simulation at room temperature, 47 urea molecules have at least one hydrogen bond with the protein and 86 more have lower frequencies. After correcting for the frequencies, a total of 87 urea molecules are on average hydrogen bonded to barnase. A slightly larger number of urea molecules, 56, have at least one hydrogen bond with the protein at 360 K, but a marked increase to 103 in the number of ureas fractionally bound is observed. Under these conditions a total of 108 molecules are hydrogen bonded on average to the protein after correcting for the frequencies. Thus, 87 of the 91 first-shell ureas are hydrogen bonded to the protein at 298 K, and these numbers increase to 108 of the 120 ureas at 360 K. Overall, these results clearly indicate that (1) the first solvation shell of the protein is substantially enriched in urea and (2) almost all of the first-shell ureas form at least one hydrogen bond with the protein. Of course, the enrichment of urea in the first shell allows reduction in the number of water molecules in proximity to hydrophobic side chains and an increase in the exposed hydrophobic surface area, as observed (Table 1). This pattern does not decrease at the higher temperature, but the average frequency of protein−urea hydrogen bonds decreases somewhat.

Although it is difficult to measure experimentally these types of effects, there are some observations that support them. It has been reported that the fluorescence of 1-anilino-8-naphthalenesulfonic acid (ANS) binding to cardiotoxin III is greatly diminished by urea. Since this protein is not unfolded by urea, the implication is that urea binds to the surface of the protein and displaces ANS (Kumar *et al*., 1996). Also, NMR experiments with the small proteins BPTI and PEC-60, which are not unfolded under the conditions studied, show that amide protons undergo the largest changes in ^1H chemical shifts in going from 0 to 8 M urea. The effect also decreases on raising the temperature up to 36 °C (Liepinsh & Otting, 1994).

Urea has additional advantages as a cosolvent for proteins due to its structure. Like water, it is both a hydrogen bond donor and acceptor. Once an association with the polar groups of the exposed protein surface is established, urea can provide anchoring for water molecules. The solvation spheres around Y^{17} and K^{108} in Figures 18 and 19 show that in the aqueous simulations the water molecules arrange around the solute in a very specific manner by forming hydrogen bonds to the polar groups of the solute and maximizing hydrogen bonds among themselves. In the urea simulations, the cosolvent displaces some of the waters molecules forming hydrogen bonds to the solute; the remaining water molecules then interact with the urea. Furthermore, the urea molecule is larger than water. This allows it to form longer bridges between portions of the protein that were interacting directly and hence to pry apart contacts between regions of secondary structure. Figure 20 shows two urea molecules in the active site which bridge between L3 and the β-sheet in the CU simulation at room temperature. The effect is even more pronounced in the high temperature simulation HU. Figure 21 shows urea molecules between the N-terminus and L1, the C-terminus of H1 and the β-sheet, and the N-terminus of H2 and the β-sheet.

Overall, the results of these simulations show encouraging agreement with available experimental data. This is especially important since each of the simulations described here follows a single molecule along a unique path or a single unfolding event, while experimental results are produced by the average of innumerable molecules following individual trajectories. The level of agreement obtained supports the adequacy of the methods and models utilized in these studies to represent the actual structures and the processes followed by their interconversion. The structure obtained from the

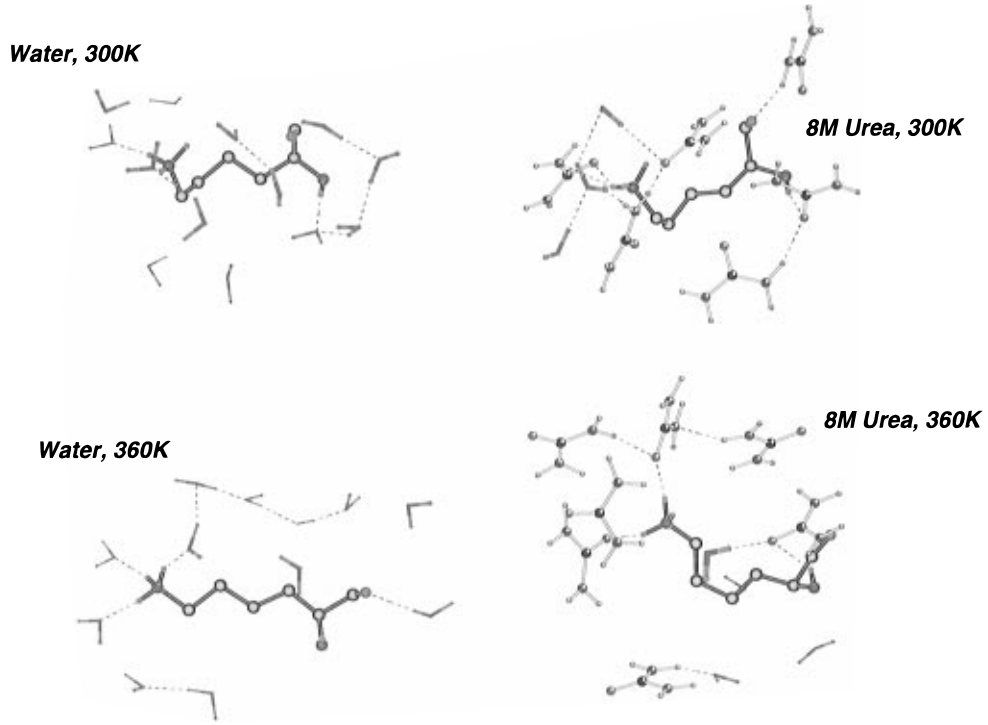

FIGURE 19: Solvation spheres around K^{108} in the final structures of the simulations. The same representations as in Figure 17 were used.

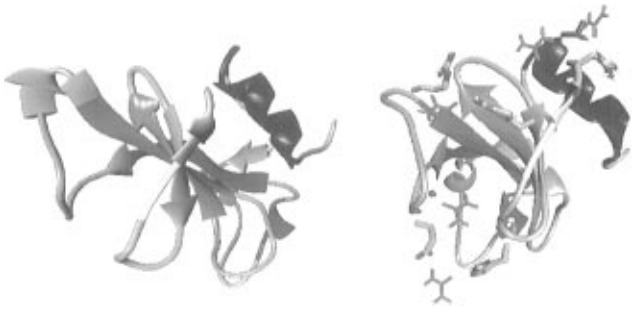

FIGURE 20: Ribbon representation of the final structures of the room temperature CW and CU simulations with a few urea molecules. The point of view is approximately orthogonal to previous illustrations, rotated toward the top of the β-sheet.

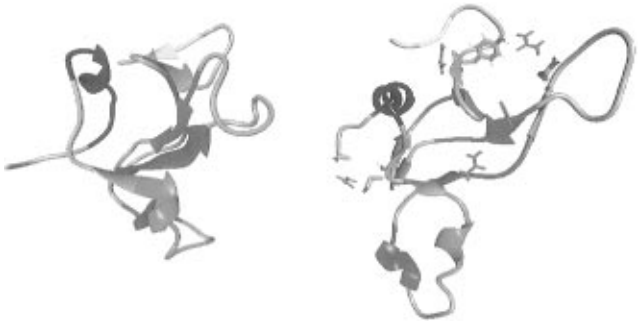

FIGURE 21: Ribbon representation of the final structures of the high temperature HW and HU simulations. The point of view is approximately orthogonal to previous illustrations, rotated toward the N-terminus of the major α-helix.

simulation under native conditions agrees well with NMR data and matches the pattern of protection from isotopic exchange, and the unfolding pathway obtained in water at 85 °C follows closely the energetic trends observed by protein engineering. The results of simulations both at room temperature and at 85 °C in 8 M urea also match the experimental data, of which there is far less available. More importantly, these results suggest a plausible model for the effect of urea in protein denaturation. Urea molecules accumulate in excess near the protein surface and form normal hydrogen bonds with the surface-exposed polar groups. The displacement of water molecules from the first solvation shell leads to greater exposure of nonpolar side chains. The urea molecules also act both as a scaffold for the remaining water molecules, thereby further minimizing their contact with the hydrophobic groups, and as a wedge to separate groups previously forming intramolecular hydrogen bonds. This mode of interaction is consistent with a model in which urea promotes protein unfolding through the stabilization of the unfolded form rather than destabilizing the native state.

ACKNOWLEDGMENT

Gratitude is expressed to Prof. A. R. Fersht for supplying the X-ray coordinates used in this study.

REFERENCES

Allen, T. J. (1993) Ph.D. Thesis, Massachusetts Institute of Technology.

Åqvist, J., Medina, C., & Samuelsson, J. E. (1994) *Protein Eng. 7*, 385.

Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., & Fersht, A. R. (1995) *J. Mol. Biol. 254*, 305.

Auer, H. E., & Doty, P. (1966) *Biochemistry 5*, 1716.

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

Baudet, S., & Janin, J. (1991) *J. Mol. Biol. 219*, 123.

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.

Braxenthaler, M., Avbelj, F., & Moult, J. (1995) *J. Mol. Biol. 250*, 239.

Breslow, R., & Guo, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 167.

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

Bycroft, M., Sheppard, R. N., Lau, F. T.-K., & Fersht, A. R. (1990b) *Biochemistry 29*, 7425.
Bycroft, M., Ludvigsen, S., Fersht, A. R., & Poulsen, F. M. (1991) *Biochemistry 30*, 8697.
Caflish, A., & Karplus, M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1746.
Caflish, A., & Karplus, M. (1995) *J. Mol. Biol. 252*, 672.
Carlson, H. A., & Jorgensen, W. L. (1995) *J. Phys. Chem. 99*, 10667.
Creighton, T. E. (1985) *J. Phys. Chem. 89*, 2452.
Daggett, V., & Levitt, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5142.
Duffy, E. M., Kowalczyk, P. J., & Jorgensen, W. L. (1993a) *J. Am. Chem. Soc. 115*, 9271.
Duffy, E. M., Severance, D. L., & Jorgensen, W. L. (1993b) *Isr. J. Chem. 33*, 323.
Fersht, A. R. (1993) *FEBS Lett. 325*, 5.
Fersht, A. R., Matouscheck, A., & Serrano, L. (1992) *J. Mol. Biol. 224*, 771.
Gratzer, W. B., & Doty, P. (1963) *J. Am. Chem. Soc. 85*, 1193.
Hermans, J., Jr. (1966) *J. Am. Chem. Soc. 88*, 2418.
Horovitz, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 733.
Horovitz, A., Serrano, L., & Fersht, A. R. (1991) *J. Mol. Biol. 219*, 5.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544.
Hünenberger, P. H., Mark, A. E., & van Gunsteren, W. (1995) *Proteins 21*, 196.
Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol. 49*, 117.
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.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2757.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631.
King, J (1989) *Chem. Eng. News 67*, 32.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946.
Kumar, T. K. S., Jayaraman, G., Lin, W.-Y., & Yu, C. (1996) *Biochim. Biophys. Acta 1294*, 103.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 56*, 87.
Li, A. J., & Daggett, V. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10430.
Li, A. J., & Daggett, V. (1996) *J. Mol. Biol. 257*, 412.
Liepinsh, E., & Otting, G. (1994) *J. Am. Chem. Soc. 116*, 9670.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749.
Makhatadze, G. I., & Privalov, P. L (1992) *J. Mol. Biol. 224*, 491.
Mark, A. E., & van Gunsteren, W. F. (1992) *Biochemistry 31*, 7745.
Matouschek, A., & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., & Fersht, A. R. (1990) *Nature 346*, 440.
Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 837.
Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C., & Jack, A. (1982) *Nature 297*, 162.
Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633.
Nishimura, S., & Nomura, M. (1958) *Biochim. Biophys. Acta 30*, 430.
Otzen, D. E., Itzhaki, L. S., ElMasry, N. F., Jackson, S. E., & Fersht, A. R, (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422.
Pace, C. N., Laurents, D. V., & Erickson, R. E. (1992) *Biochemistry 31*, 2728.
Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P., & Kollman, P. A. (1991) *AMBER 4.0*, University of California, San Francisco.
Perrett, S., Clarke, J., Hounslow, A. M., & Fersht, A. R. (1995) *Biochemistry 34*, 9288.
Prévost, M. (1996) *J. Mol. Biol. 260*, 99.
Prévost, M., Wodak, S. J., Tidor, B., & Karplus, M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10880.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167.
Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 273.
Pugliese, L., Prévost, M., & Wodak, S. J. (1995) *J. Mol. Biol. 251*, 432.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700.
Roseman, M., & Jencks, W. P. (1975) *J. Am. Chem. Soc. 97*, 631.
Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327.
Sali, D., Bycroft, M., & Fersht, A. R. (1988) *Nature 335*, 740.
Smythe, M. L., Huston, S. E., & Marshall, G. R. (1993) *J. Am. Chem. Soc. 115*, 11594.
Soman, K. V., Karimi, A., & Case, D. A. (1991) *Biopolymers 31*, 1351.
Sun, Y.-C., Veenstra, D. L., & Kollman, P. A. (1996) *Protein Eng. 9*, 273.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121.
Tiffany, M. L., & Krimm, S. (1973) *Biopolymers 32*, 575.
Tirado-Rives, J., & Jorgensen, W. L. (1991) *Biochemistry 30*, 3864.
Tirado-Rives, J., & Jorgensen, W. L. (1993) *Biochemistry 32*, 4175.
Tirado-Rives, J., Maxwell, D. S., & Jorgensen, W. L. (1993) *J. Am. Chem. Soc. 115*, 11590.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694.
Vijayakumar, S., Vishveshwara, S., Ravishanker, G., & Beveridge, D. L. (1993) *Biophys. J. 65*, 2304.
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.
White, F. H., Jr., & Anfinsen, C. B (1959) *Ann. N.Y. Acad. Sci. 81*, 515.

BI970096I