JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

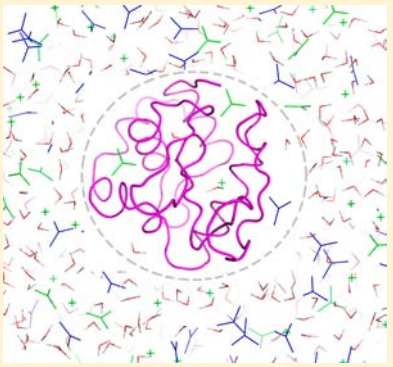
Collapse of Unfolded Proteins in a Mixture of Denaturants

Zhen Xia,[†,§] Payel Das,[†] Eugene I. Shakhnovich,*[,‡] and Ruhong Zhou*[,†,⊥]

[†]Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, United States
[‡]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States
[§]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States
[⊥]Department of Chemistry, Columbia University, New York, New York 10027, United States

Supporting Information

ABSTRACT: Both urea and guanidinium chloride (GdmCl) are frequently used as protein denaturants. Given that proteins generally adopt extended or unfolded conformations in either aqueous urea or GdmCl, one might expect that the unfolded protein chains will remain or become further extended due to the addition of another denaturant. However, a collapse of denatured proteins is revealed using atomistic molecular dynamics simulations when a mixture of denaturants is used. Both hen egg-white lysozyme and protein L are found to undergo collapse in the denaturant mixture. The collapse of the protein conformational ensembles is accompanied by a decreased solubility and increased non-native self-interactions of hydrophobic residues in the urea/GdmCl mixture. The increase of non-native interactions rather than the native contacts indicates that the proteins experience a simple collapse transition from the fully denatured states. During the protein collapse, the relatively stronger denaturant GdmCl displays a higher tendency to be absorbed onto the protein surface due to their stronger electrostatic interactions with proteins. At the same time, urea molecules also accumulate near the protein surface, resulting in an enhanced "local crowding" for the protein near its first solvation shell. This rearrangement of denaturants near the protein surface and crowded local environment induce the protein collapse, mainly by burying their hydrophobic residues. These findings from molecular simulations are then further explained by a simple analytical model based on statistical mechanics.



INTRODUCTION

Urea and guanidinium chloride (GdmCl) are both commonly used as protein denaturants in various studies.[1−15] The molecular denaturation mechanism continues to be controversial in the past decades. In general, two different mechanisms have been proposed: an "indirect mechanism" where denaturants disrupt the structure of water and thus enhance the solubility of hydrophobic groups of proteins,[16−21] and a "direct mechanism" where denaturants directly interact with proteins via electrostatic or van der Waals forces.[22−27] Our recent study on the denaturation of hen egg-white lysozyme in 8 M urea solution strongly supports this "direct mechanism", in which urea interacts with the protein backbone and side chains via stronger dispersion interactions than water.[27] A two-stage kinetic model has been proposed, which begins with a "dry-globule" transient state, followed by a global unfolding of the protein.[27] Later experiments have further confirmed the "direct mechanism" by observing urea molecules directly forming hydrogen bonds to the backbone of dialanine (*N*-acetyl-L-alanine *N*′-methylamide).[28] Similar mechanisms have been proposed for GdmCl molecules, which directly interact with proteins.[25,29] In addition, the planar, charged Gdm^+ is found to interact with aromatic side chains more strongly by "stacking" in both a model helical peptide and protein L[30,31] (which is also seen in our simulations, more below). Therefore, Gdm^+ is often considered to be approximately twice as effective as urea in its ability to denature proteins.[28]

On the other hand, the behavior of proteins immersed in an aqueous mixture of two denaturants has not been systematically studied. It is of great interest to see what happens to the protein conformations solvated in one denaturant (either urea or GdmCl only) when another denaturant (GdmCl or urea) is also added (i.e., in a mixture of both urea and GdmCl denaturants). One might expect that the unfolded protein chain will remain or be further extended due to the addition of another denaturant. However, Shakhnovich and co-workers previously proposed that a mixture of cosolvents may trigger a collapse of a polymer in a broad range of conditions.[32−34] DeGennes and Brochard have also suggested a possibility of a chain collapse in a mixed solvent near the critical mixing point of the solvent.[35] Considering that two denaturants compete with each other in interacting with proteins when they are mixed, the competition might lead to unexpected complex behavior of protein energetics and dynamics. Meanwhile, proteins solvated in mixed solvents are considered to be a more realistic representation of a cellular environment. Thus, it is of fundamental importance to investigate the conformational diversity of a protein in mixed solvents, which may play a

Received: April 2, 2012
Published: October 11, 2012

   dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

crucial role in modulating its function in various environments. In addition, how the interactions among the stronger denaturant (such as GdmCl), the relatively weaker denaturant (such as urea), water, and the protein interplay in the complex solution requires more detailed investigation, because the delicate balance of those interactions will influence the kinetics and thermodynamics of the protein denaturation reaction.

To address these questions, we performed extensive molecular dynamics simulations of proteins in urea/GdmCl mixture with different concentrations. Two independent protein systems, hen egg-white lysozyme and protein L, were simulated as the representative model proteins. We found a collapse of denatured protein conformations in urea/GdmCl mixtures for both lysozyme and protein L, compared with their respective conformations in the single denaturant solution. The word "collapse" here means that the protein populates more compact structures (but not necessarily with more native contacts or secondary structures). This contraction of the protein chain in mixed denaturant suggests that mixing two denaturants, that is, good solvents, results in a poor solvent for proteins. Further analysis reveals that GdmCl is preferentially absorbed onto the charged protein residues due to its stronger electrostatic interactions, whereas urea is preferentially adsorbed onto the hydrophobic and polar residues. This residue-specific preferential adsorption of a particular denaturant results in a "urea cloud" near the first solvation shell of the noncharged residues, which constitute majority of the protein. This "local urea cloud" around particular protein residues results in an effective residue−residue interaction, inducing the protein collapse mainly by burying the hydrophobic residues. In the main text, as well as the Appendix in Supporting Information, we provide a more complete analytical theory than the previous works[32] to explain this seemingly surprising phenomenon with a model polymer in a mixture of water, urea, and GdmCl.

METHODS

Preparation of the Denatured Proteins. The denatured structures of two proteins, hen egg-white lysozyme (PDB entry 193L, with single mutation W62G)[36] and wild-type of protein L (PDB entry 2PTL),[37] were prepared by the following procedure with three steps: (1) Solvating each protein with its native structure in a pre-equilibrated 8 M urea water box. The size of the 8 M urea solution box was the same as our previous study[38] and the same for both proteins, 73.1 Å × 73.1 Å × 73.1 Å, which contained 1920 urea and 8192 water molecules, with a density of 1.12 g/cm^3. (2) Running molecular dynamics simulations for at least 500 ns with *NPT* ensemble (1 atm and 310 K) until the proteins were fully unfolded. We defined the structures as the unfolded states when the radius of gyration (Rg) of the proteins does not increase for at least 50 ns in the simulation. (3) For each protein, a representative snapshot from the final 50 ns was chosen as the starting denatured structure for the current denaturant mixture study. The Rg were 24.8 and 16.3 Å for denatured lysozyme and protein L, respectively.

For the lysozyme protein, the full length of the protein was used in the simulations (residue index from 1 to 129); for protein L, a fragment of residue 18−78 was used, because the N terminal 17-residue fragment is a long, straight, and Glu-rich loop that was separated from other parts of protein L; removal of the N terminal residues was often used by previous studies in order to effectively distinguish denatured states from the native structure,[39] particularly when measuring Rg.

Proteins in the Mixed Denaturants. In this step, the denatured proteins were then immersed in a mixed-denaturant water box with different combinations of concentrations of urea and guanidinium chloride (GdmCl) for both lysozyme and protein L. Four different urea/GdmCl combinations were studied: (1) 8 M urea + 0 M GdmCl, (2) 6 M urea + 2 M GdmCl (replacing one-quarter of the initial urea molecules in 8 M urea by GdmCl), (3) 4 M urea + 4 M GdmCl (replacing half of the initial urea molecules in 8 M urea by GdmCl), and (4) 0 M urea + 6 M GdmCl. For each system, we generated three independent trajectories, with each about 120−300 ns long in *NPT* ensemble (at 310 K and 1 atm, with Berendsen thermostat and barostat). All simulations were performed using the NAMD2 molecular dynamics program.[40] Molecular dynamics simulations have been widely used to complement experiments,[27,38,41−54] which can provide atomic details that are often inaccessible in experiments due to resolution limits, even with the currently available sophisticated experimental techniques. The CHARMM force field (c32b1 parameter set)[55−57] is used in the current study for lysozyme, protein L, and the denaturants urea and GdmCl. A modified TIP3P water model was used for water with its bond lengths constrained with SHAKE/RATTLE.[58] The long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME)[59,60] method (updated every time step), and a typical 12 Å cutoff was used for the van der Waals interactions. The time step for all production runs was 1.5 fs.

RESULTS AND DISCUSSION

Statistical Mechanical Theory Predicts Possibility of Chain Collapse in Mixed Solvent. The statistical-mechanical model for polymer in mixed solvent is illustrated in Figure 1.

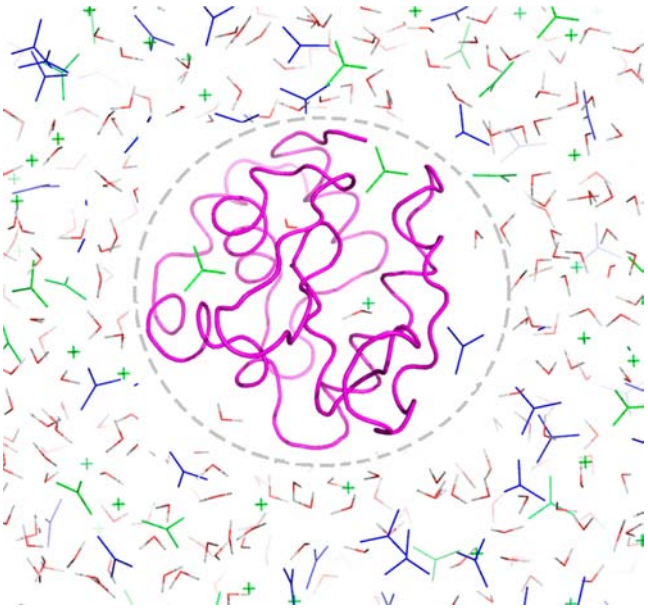


Figure 1. An illustration of a polymer chain forming a "wet" globule with much space available for penetrating solvents inside. The solvent density inside the globule is lower than that outside. The polymer is represented as cartoon and colored in magenta. The water molecules (red) and other two solvents (green and blue) are shown as sticks.

Solvent molecules exchange between the area occupied by the polymer molecule, where it interacts with monomers of the polymer and part of the solvent free of polymers. Interaction between polymer and solvent and solvent molecules themselves results in redistribution of solvent molecules and shift of the solvent composition inside the polymer compared with that free of solvent. Specifically, we consider the area occupied by the polymer (volume V) and an outside solution, which consists of water (dominant solvent) and cosolvents A and B. The number of A and B molecules adsorbed inside the polymer area is N_{in}^{A} and N_{in}^{B}, respectively. Water + solvent + polymer are at constant volume and pressure (see Figure 1).

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

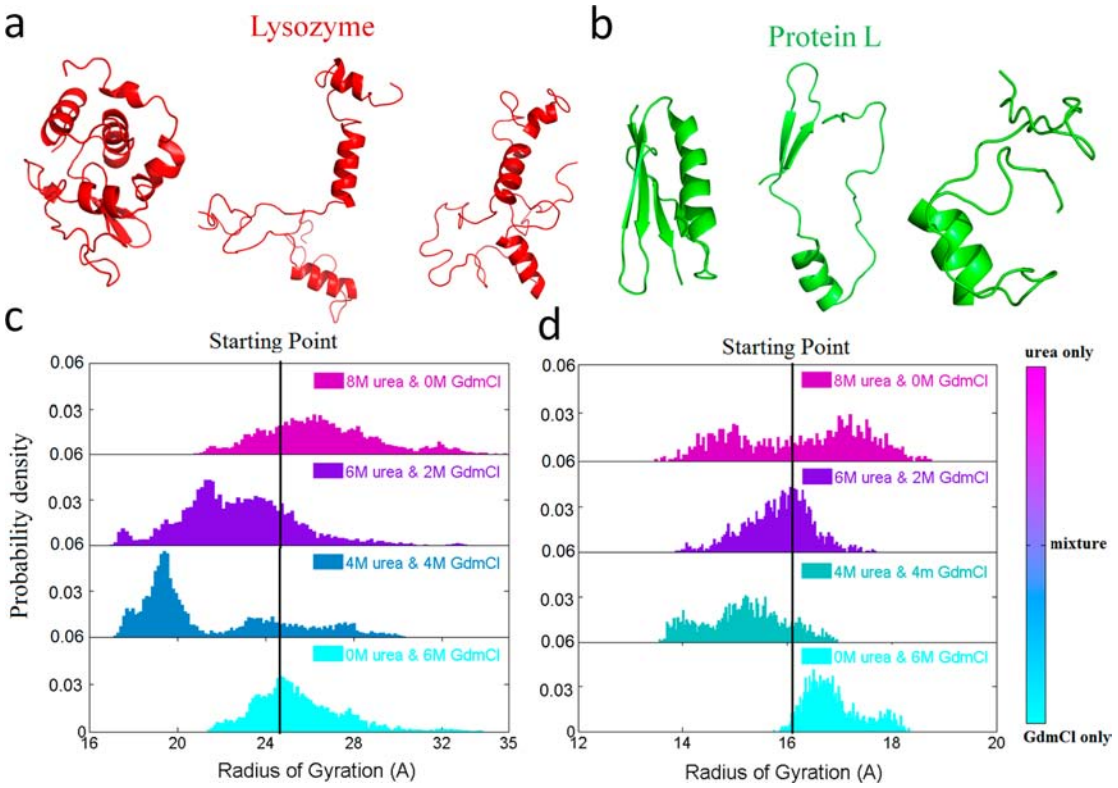


Figure 2. Protein collapse in urea/GdmCl mixture. (a) The structure of lysozyme protein. The protein is presented in cartoon view and colored in red. The native structure is shown on the left, and the denatured structure is in the middle, which is used as the starting point for the simulations with different denaturant combinations. The collapsed structure is shown on the right. (b) The structure of protein L. The protein is colored in green, and the structures are shown with similar representations and orders as lysozyme. (c, d) The distribution of the radius of gyration (Rg) of proteins under different concentrations of urea and GdmCl for lysozyme and protein L systems, respectively. The black line is the starting point for all the simulations. Smaller Rg are seen in GdmCl/urea mixtures for both lysozyme and protein L.

For simplicity, we consider water density as fixed. The Gibbs free energy in this case is

$$\begin{aligned} G &= G_p(V, N_p, N_{in}^A, N_{in}^B) + G_s(N_0^A - N_{in}^A, N_0^B - N_{in}^B, V_0 - V) \\ &= G_p(V, N_p, N_{in}^A, N_{in}^B) - \mu_0^A N_{in}^A - \mu_0^B N_{in}^B + \text{const} \end{aligned} \quad (1)$$

where G_p is free energy of the polymer + inside solvent, which includes all interactions and entropy of solvent molecules inside the area occupied by the polymer, G_s is free energy of outside solvent, which contains fewer molecules (e.g., $N_0^A - N_{in}^A$) because a number of molecules (N_{in}^A) migrated from free solvent into the interior of the globule. $\mu_0^{A,B}$ is the chemical potential of solvent components.

The distribution of cosolvents inside the globule is defined by the relation of equilibrium that minimizes the Gibbs free energy with respect to the amount of molecules that penetrated inside.

$$\begin{aligned} \frac{\partial G_p(V, N_p, N_{in}^A, N_{in}^B)}{\partial N_{in}^A} &= \mu_0^A \\ \frac{\partial G_p(V, N_p, N_{in}^A, N_{in}^B)}{\partial N_{in}^B} &= \mu_0^B \end{aligned} \quad (2)$$

Excluding N_{in}^A and N_{in}^B from G_p using eq 2 allows us to obtain a closed form expression for free energy of polymer only. The complete analysis presented in Supporting Information indeed shows that effective interaction between monomers of the polymer (after solvent molecules are "integrated out") depends nonmonotonically on solvent composition. The physical reason for such "counterintuitive" behavior is that solvent composition inside the polymer globule shifts to increase concentration of a solvent component whose interactions with polymer are favorable (or less unfavorable) and deplete that of the solvent component whose interactions with polymer are less favorable or more unfavorable. Since such redistribution, as shown in our later molecular simulations, results in a more favorable net energetics of polymer−solvent interactions, there appears a free energy force to make the polymer more compact, because higher polymer density gives rise to greater concentration redistribution of solvent inside the polymer globule that leads to favorable energetics. The opposing factor that limits the degree of compositional shift inside the polymer is mixing entropy.

Protein Conformation Collapse in Urea and Guanidinium Chloride Mixture. Denatured protein conformations were used as the starting structures for our current simulations. Here, we used the single mutant (W62G) lysozyme for illustration, because it unfolded faster and more globally in 8 M urea than the wild-type lysozyme. The selected denatured structure of lysozyme has a radius of gyration (Rg) of 24.8 Å, which is significantly larger than that of its native state (Rg = 16.0 Å) (Figure 2a). A similar unfolding simulation in 8 M urea was carried out to obtain the starting denatured structure of protein L, where the Rg for the denatured state and native state were 16.3 and 11.0 Å, respectively (Figure 2b).

For lysozyme, we investigated four different denaturant combinations, with different concentrations of urea and

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

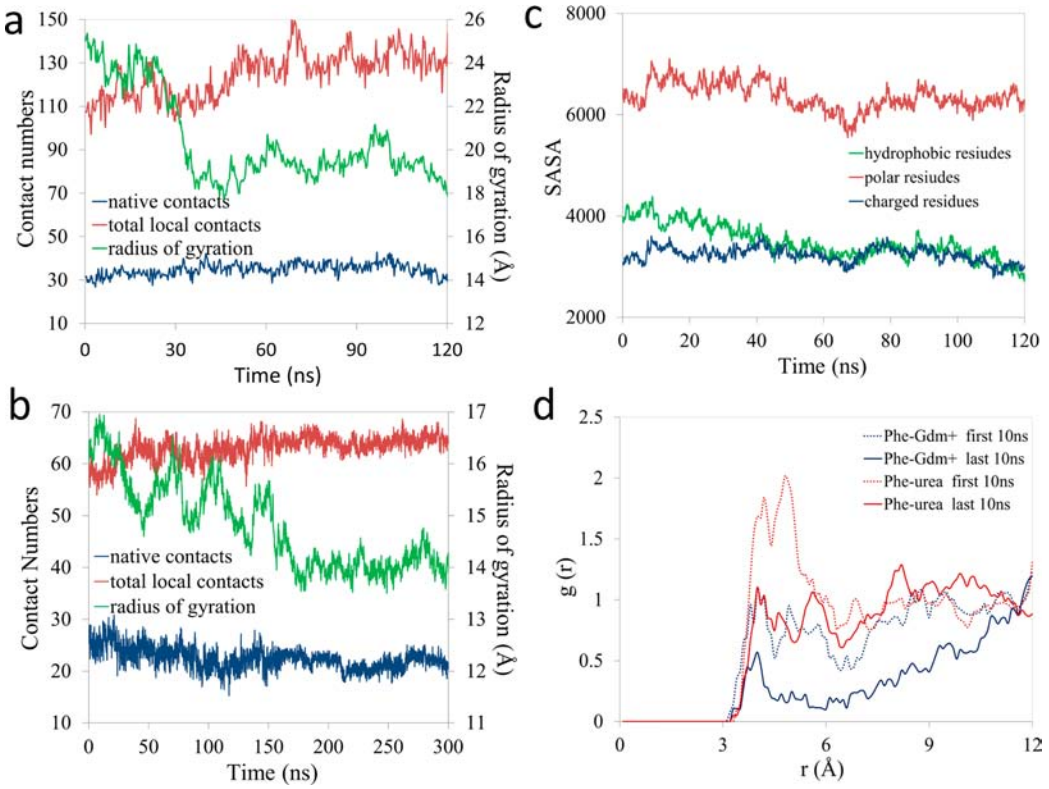


Figure 3. The time dependent of total local contacts (red), native contacts (blue), and radius of gyration (green) in "4 M urea + 4 M GdmCl" mixture for lysozyme (a) and protein L (b). The residues are considered in contact when their Cα−Cα distance is less than 6.5 Å. (c) Protein solvent-accessible surface area (SASA) of different types of residues in "4 M urea + 4 M GdmCl" mixture for lysozyme during the simulation time. (d) Time-dependent pair radial distribution function $g(r)$ between side chain of Phe38 and the carbon atom of urea/Gdm^+. The geometric center of the benzene ring at Phe38 is used to represent the position of the side chain. All of the $g(r)$ functions are averaged over the first 10 ns (dash line) and the last 10 ns (solid line) of the total 120 ns trajectory.

guanidinium chloride (GdmCl). We replaced part of the urea with the corresponding number of GdmCl molecules in the 8 M urea system to generate different combinations of the urea/GdmCl mixture (see Methods section for details). In pure 6 M GdmCl solution, the major peak of the Rg distribution shifts to ∼25 Å (25.5 Å on average), whereas the same peak in pure 8 M urea is at ∼26 Å (Rg = 26.5 Å on average) (Figure 2c). Here we use CHARMM force field parameters for GdmCl, and both lysozyme and protein L (see below) do denature in 6 M GdmCl, as shown in Figure 2; however, some other recent study has shown that the current fixed-charge parameters for GdmCl might not be sufficient to denature proteins in Amber ff99SB force field, indicating some compatibility complexity with force fields used.[61] To our surprise, the distributions of Rg clearly show that the denatured lysozyme collapsed in all urea/GdmCl mixtures (Figure 2c). In "4 M urea + 4 M GdmCl" mixed denaturants (replacing half of the initial urea molecules in 8 M urea by GdmCl), the major distribution peak of Rg was shifted to ∼19.4 Å (with average Rg = 21.4 Å), which was significantly smaller than the typical peak of ∼24−26 Å in single denaturant solutions, as shown in Figure 2c. This decrease in Rg can also be seen from the time evolution of Rg in the equimolar mixture of denaturants (Figure 3a). And another two independent simulations gave very similar results (see Figure S1, Supporting Information). The Rg distribution in the "6 M urea + 2 M GdmCl" mixture (replacing one-quarter of the initial urea molecules in 8 M urea by GdmCl) also shows similar behavior, with its major peak shifted to 21−22 Å (with average Rg = 22.8 Å), between the corresponding peaks of the "4 M urea + 4 M GdmCl" mixture and the pure denaturant solutions (either 8 M urea or 6 M GdmCl, see Figure 2c).

This collapse of the denatured protein conformations in the mixed denaturants was then further confirmed by simulating another protein, protein L. There, we found a similar denaturant mixture-induced collapse. In the "4 M urea + 4 M GdmCl" mixture, the average Rg value was shifted from 16.2 Å in pure 8 M urea and 16.9 Å in pure 6 M GdmCl to a lower value of 14.5 Å (Figure 2d). Similarly, the time evolution of Rg shows the same decreasing trend when the mixture of denaturants was used (Figure 3b).

Denaturant Mixture Triggers a Decrease in Solvent Exposure of Protein Hydrophobic Residues. The collapse of denatured proteins in mixed denaturants can also be seen from the lowering of protein solvent-accessible surface areas (SASA) during the simulation. Upon immersion of lysozyme in the "4 M urea + 4 M GdmCl" mixture, the overall protein SASA decreased from initial value of ∼13 300 $Å^2$ (in pure 8 M urea) to ∼11 500 $Å^2$. A similar trend was also found for protein L, where the corresponding SASA dropped from ∼6600 $Å^2$ in pure 8 M urea to 6100 $Å^2$ in the mixture. We further decomposed the total SASA into different amino acid types: hydrophobic, polar, and charged residues. We found that the reduction in SASA was mainly contributed by the hydrophobic residues (Figure 3c and Figure S2, Supporting Information). For example, the SASA dropped by ∼1100 $Å^2$ in lysozyme for hydrophobic residues, which contributed ∼65% total loss in SASA, with the remaining 35% loss was from the polar and charged residues. In other words, the solubility of hydrophobic

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

residues decreased significantly in the mixture. Similar SASA decrease could be seen for protein L, in which hydrophobic residues contributed ~60% of total loss (the SASA drop from 1622 Å^2 at the beginning to 1347 ± 32 Å^2 for the last 10 ns). Meanwhile, a non-native hydrophobic core was formed during the protein collapse. For further analysis, we calculated the radial distribution function (rdf) of urea and Gdm^+ around the side chain of one representative hydrophobic residue Phe38 in lysozyme. A significant decrease in the solvent exposure can be seen for residue Phe38 after the protein collapse (Figure 3d). Other hydrophobic residues near the protein surface, particularly aromatic ones, show similar behavior. In addition, the number of backbone−backbone hydrogen bonds increased 70% for the hydrophobic residues (from 3.07 ± 0.83 at the beginning 10 ns to 5.26 ± 1.10 for the last 10 ns, see Figure S3, Supporting Information), which further confirms enhanced self-interactions among them. In contrast, the SASA remained mostly unchanged for charged residues (due to their strong electrostatic interactions with GdmCl) and only decreased slightly for polar residues during the same collapsing process for both protein systems. To further confirm the solubility decrease of hydrophobic residues in mixtures, but not in GdmCl alone, we also calculated the SASA of proteins in a pure 6 M GdmCl system ("0 M urea + 6 M GdmCl"). No obvious decrease of SASA was observed for all three amino acid types (See Figures S4 and S5, Supporting Information). In summary, our simulation results revealed a reduced solubility and increased self-interactions of hydrophobic residues in mixed denaturants, indicating the hydrophobic collapse as the underlying main factor for the consequent collapse of the entire protein.

The Increased Contacts during the Collapse Are Mostly Non-native. The numbers of native contacts and total local contacts (including both native and non-native) were calculated for both lysozyme and protein L in the "4 M urea + 4 M GdmCl" mixture during the collapse process (Figure 3a,b). Two residues are considered to be in contact if their Cα−Cα distance is less than 6.5 Å. For the lysozyme system, we noticed a significant increase in the number of total local contacts after ~40 ns of the simulation time. The number of total local contacts increased 21.5%, from the initial 107 to the final 130 with an average number of 125 ± 10, in the "4 M urea + 4 M GdmCl" mixture (Figure 3a). Coincidently, we found that the radius of gyration, Rg, dropped from ~25 to ~19 Å after ~40 ns (Figure 3a), which was correlated with the increase in the number of local contacts. However, the native contact number remained roughly constant during this process, fluctuating around 34 with a standard deviation of 3.0. For the protein L mixture system, very similar trends were found (Figure 3b). The number of local contacts increased by 16%, from the initial 56 to the final 65, with an average of 62 ± 2.4, while the final number of native contacts even decreased slightly, from the initial 26 to the final 23 with an average number of 23 ± 2.3. We further calculated the secondary structure content (as a percentage of residues in α-helices or β-sheets) for lysozyme and protein L in both pure denaturant systems and their mixtures. We found very similar α-helical and β-sheet structure contents for the "4 M urea + 4 M GdmCl" mixture or pure denaturants (8 M urea or 6 M GdmCl only) (Figure S6, Supporting Information). For example, lysozyme, which displays the helical structures mainly in the denatured state, has an average helical content of 26.8% ± 3.0%, 25.1% ± 2.1%, and 25.2% ± 2.0% in "8 M urea + 0 M GdmCl" solvent, "4 M urea + 4 M GdmCl" mixture, and "0 M urea + 6 M GdmCl" solvent, respectively (Figure S6a, Supporting Information). Therefore, it seems that both lysozyme and protein L collapsed toward non-native structures, indicating the proteins were still in their denatured states, albeit more compact structures. In other words, the induced collapse of the denaturant mixture is not a refolding process, but a mere collapse with more non-native contacts formed during our 200+ ns simulations. It should be noted that some recent simulations of protein unfolded states suggest that microsecond-or-longer trajectories might be needed to sufficiently model equilibrated unfolded ensembles.[39,62]

Rearrangement of Denaturants near Protein Surface and Enhanced Local Crowding Induce the Protein Collapse. The driving force toward the protein collapse was then investigated from an energetic perspective. The total interaction energy distributions between the individual solvent molecules (guanidinium, urea, or water) and protein were calculated for lysozyme in the "4 M urea + 4 M GdmCl" mixture system. Figure 4a shows the comparison of the

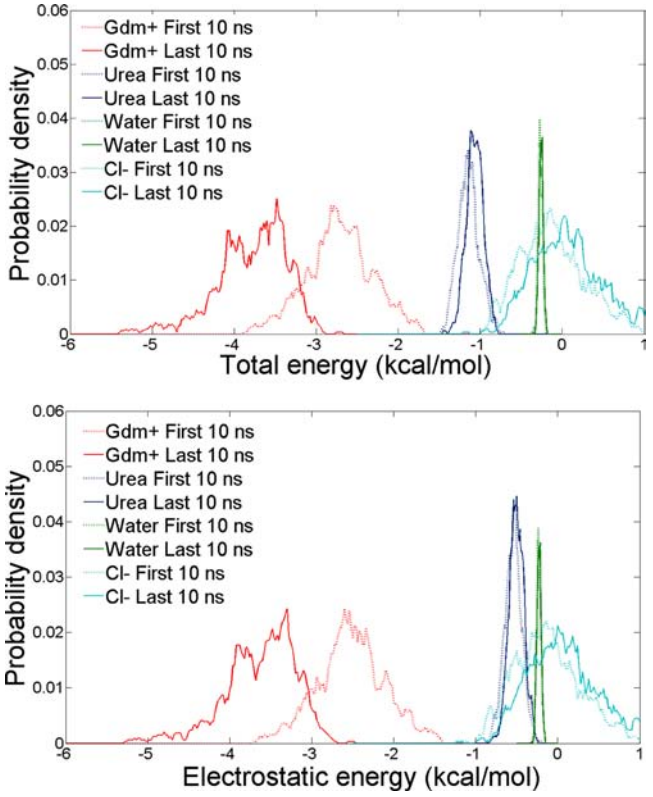


Figure 4. The distributions of total interaction energy (normalized per molecule) (a) and electrostatic component energy (b) between solvents (guanidinium, urea, water, and chloride ion) and protein for "4 M urea + 4 M GdmCl" mixture lysozyme system. The calculations were performed for the first 10 ns (dash line) and the last 10 ns (solid line) in the trajectory, which represent the fully unfolded state and the collapsed state of protein, respectively.

interaction energy distributions (normalized to be per molecule) in the first and last 10 ns of the trajectory, which represent the initial unfolded state and the final collapsed state of protein, respectively. It is apparent from this figure that GdmCl overall has a significantly more favorable interaction with the protein than the urea, confirming that GdmCl is a relatively stronger denaturant. For the interaction between water and protein, we found the distribution of the total

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

interaction energies almost the same before and after the collapse (Figure 4a). For the interaction between urea and protein, we observed a slightly higher energy distribution (i.e., less favorable) after the protein collapse, with the interaction energy weakened by ∼0.1 kcal/mol per urea molecule on average, indicating that some of the initial nearby urea molecules were replaced by the stronger denaturant GdmCl. The largest changes were observed in the interaction between guanidinium and protein. Interaction energy between guanidinium and protein was much enhanced by the collapse, with the average interaction energy enhanced by 1.07 kcal/mol per guanidinium (−2.72 kcal/mol in the beginning and −3.79 kcal/mol after the collapse). In order to better understand whether the primary driving force for this process was electrostatic or van der Waals (vdW), the total interaction energy was further decomposed into electrostatic and vdW contributions. We found most of the interaction energy enhancements (1.00 kcal/mol out of 1.07 kcal/mol total) stemmed from the electrostatic interaction between guanidinium and protein (Figure 4b), with the average electrostatic energy changed from −2.53 kcal/mol in the beginning to −3.63 kcal/mol after the protein collapse. Meanwhile, the interaction energy between chloride and protein was not changed significantly, with only a 0.18 kcal/mol decease. Therefore, the protein collapse is favored by an enhancement in the electrostatic interaction energy between guanidinium and protein. In the meantime, the total density of denaturants consistently increases near the first solvation shell (FSS) of protein lysozyme from 10 to 30 ns until the protein collapses, as evident from the Figure 5. Similar trend of denaturant density before the collapse can also be seen for other repeated runs and protein L (see Figure S7 and S8, Supporting Information). This increase in total denaturant density near the protein stems from two different factors: (1) protein residue-specific affinity for a particular denaturant, which originates from the chemical composition of an amino acid. For example, the acidic residues interact with guanidinium with much higher affinity than with urea. On the other hand, hydrophobic residues have a preference for urea over guanidinium. (2) Self-aggregation tendency of a particular denaturant. It is known that urea can form large clusters in solution, whereas guanidinium prefers to stay as a homodimer. Thus, the preferential adsorption of urea onto the noncharged amino acids, which are the major constituent of the protein chain, creates a "urea cloud" around those residues, as evident from Figure 6. Brochard and de Gennes suggested that near the critical temperature of the binary solvent mixture, the solvent density fluctuations become more and more correlated, resulting in a solvent correlation length ξ that is comparable to the size of the polymer, leading to its collapse.[35] In the urea + GdmCl mixtures studied here, the urea cloud near hydrophobic residues results in similar indirect long-range attractions, which shield the excluded-volume interactions. As a result, the protein collapses by burying the hydrophobic residues. In this line, Cho et al. proposed a similar solvent-induced crowding mechanism to explain the role of TMAO in stabilizing proteins.[63] The statistical mechanical theory explaining the protein collapse in a mixture of denaturants is provided in the appendix (Supporting Information) for a model polymer in water and two codenaturants A (GdmCl) and B (urea). In this case, we show that the polymer collapse provides an essential attractive correction to the self-interaction of the polymer due to an effective redistribution of A and B components (see Appendix, Supporting Information, for

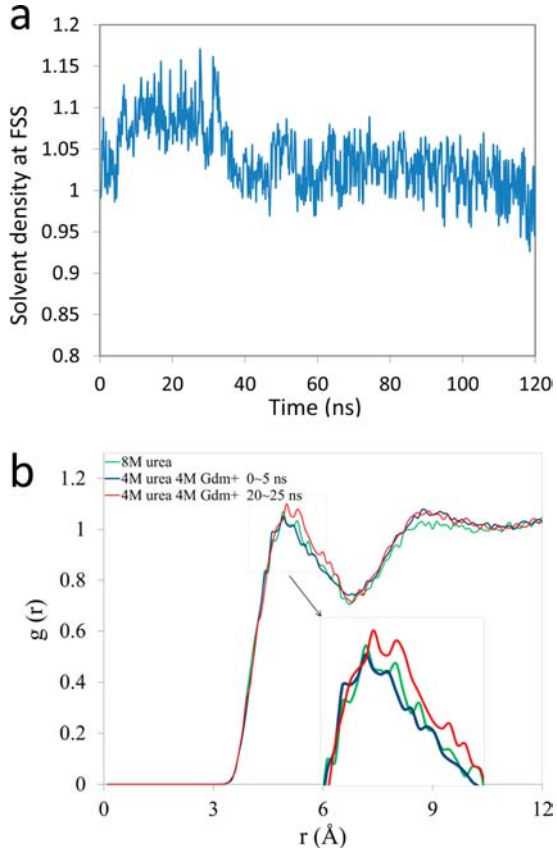


Figure 5. The local crowding effect at the surface of protein lysozyme. (a) The time-dependent density of urea and guanidinium molecules at the first solvation shell of protein. The density is calculated from the total number of urea and guanidinium molecules and then normalized by the solvent-accessible surface area of the protein. An enhanced local crowding environment with more denaturant molecules is seen from 10 to 30 ns before the protein collapse. (b) Time-dependent pair radial distribution function, $g(r)$, between the α carbon atoms of the protein backbone and the carbon atoms of urea (and guanidinium if any) in "8 M urea + 0 M GdmCl" mixture (green) and "4 M urea + 4 M GdmCl" mixture (at t = 0−5 ns in blue as reference and 20−25 ns in red). It is clear that a higher density of urea and guanidinium appears at the first solvation shell from 20 to 25 ns in "4 M urea + 4 M GdmCl" mixture.

derivations). Meanwhile, this collapsed protein conformation is also accompanied by the most energetically favorable environment for both guanidinium and urea surrounding the protein, as shown by the interaction energy analysis.

The protein−solvent interaction was further investigated by calculating the ratio of GdmCl to urea molecules ($\rho_{gdm/urea}$) at the first solvation shell (FSS) of each protein residue. Any water, urea, or GdmCl molecule is considered to be in the FSS if it is within 5 Å of any protein atom. For both lysozyme and protein L solvated in the "4 M urea + 4 M GdmCl" mixtures, we found that the $\rho_{gdm/urea}$ increased for most of the protein residues when half of the original urea molecules in 8 M urea were replaced by GdmCl. Figure 6a,b shows the comparison of $\rho_{gdm/urea}$ in the first and last 10 ns, which indicates that Gdm^+ replaced urea in the FSS of proteins due to its stronger electrostatic interactions with the protein, as discussed above. Interestingly, we also noticed that $\rho_{gdm/urea}$ dropped at the locations of some residues in protein L during the simulation, which are found to be mainly lysine residues (marked with * in

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

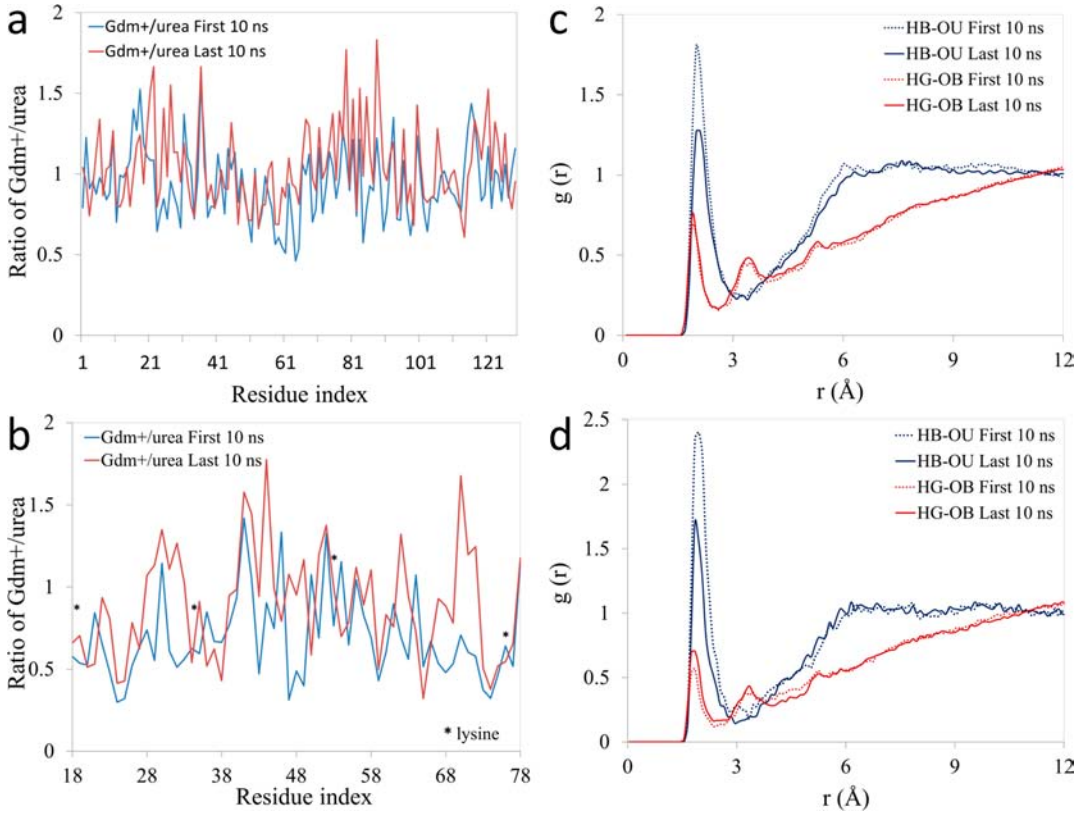


Figure 6. (a, b) The ratio of GdmCl to urea molecules ($\rho_{gdm/urea}$) at the first solvation shell for each protein residue in "4 M urea + 4 M GdmCl" mixture for lysozyme and protein L, respectively. Amino acid lysine is labeled as * for protein L. The first solvation shell is defined as within 5.0 Å of any protein atoms. (c, d) Time-dependent pair radial distribution function, $g(r)$, between backbone amide hydrogen HB and urea oxygen OU (blue), as well as between backbone carbonyl oxygen OB and Gdm^+ hydrogen HG (red) residue in "4 M urea + 4 M GdmCl" mixture for lysozyme and protein L, respectively. All of the $g(r)$ functions are averaged over the first 10 ns (dash line) and the last 10 ns (solid line) of the total 120 ns trajectory.

Figure 6b). Considering the unfavorable electrostatic interactions between two positively charged groups, Gdm^+ and $-NH_3^+$ of lysine, this seems reasonable (see more data and discussions below with Figure S9, Supporting Information, as well as Figure S10 for Arg).

This can also be seen from the time evolution of detailed atomic radial distribution functions (rdf). Figure 6c shows the rdf between the oxygen of urea (OU) and the backbone amide hydrogen (HB) [$g_{OU-HB}(r)$], which experiences a noticeable reduction after the protein collapse, indicating the loss of overall urea−backbone interactions during this process due to the burying of hydrophobic residues. On the other hand, a slight increase in the first peak was observed for the corresponding $g_{HG-OB}(r)$ between the amide hydrogen (HG) of GdmCl and the carbonyl oxygen (OB) of the protein backbone (Figure 6d). Therefore, the stronger denaturant GdmCl replaces part of the weaker denaturant urea and enhances its overall interaction with the protein during this collapsing process. Our previous denaturation studies on urea-induced lysozyme unfolding have suggested a "direct interaction mechanism", in which urea has stronger interactions with protein than water.[27,64−69] In the current urea/GdmCl mixtures, GdmCl has even stronger interactions with protein than urea, indicating that GdmCl acts as the leading player among water, urea, and GdmCl in terms of their direct interactions with proteins. Therefore, it appears that guandiniums are attracted by the protein to maximize the total number of strongly interacting guanidiniums on the protein surface. Interestingly, the protein also collapses somewhat by burying its own hydrophobic residues to accommodate more denaturant molecules near its surface, which is the most energetically favored state for the protein and denaturant mixture, confirming the predictions from the analytical model (see Appendix, Supporting Information).

Finally, since guanidinium is positively charged, we further investigated the role of charged protein residues during this collapse. We analyzed the solvation of both glutamic acid (E) and lysine (K) side chains as an example (see Figure S9, Supporting Information). The negatively charged side-chain oxygens (OE) of glutamic acid were highly solvated by Gdm^+ for both protein systems, which is 20−30 fold higher than that of urea. Following the protein collapse, guanidinium replaced urea molecules originally in contact with glutamic acid side chains. In contrast, the positively charged side chain ($-NH_3^+$) of lysine was mainly solvated by urea rather than Gdm^+, with more urea and less Gdm^+ near $-NH_3^+$ at the end of the simulations for both lysozyme and protein L (Figure S9, Supporting Information). We noticed that the solvation of another positively charged amino acid, arginine, mostly remained the same during the protein collapse (Figure S10, Supporting Information). A possible explanation is that the unfavorable electrostatic interaction was compensated by the favorable stacking interaction between Gdm^+ and the guanidinium group from the arginine side chain. These detailed results with charged residues further support that the electrostatic interactions play the dominant role in guanidi-

dx.doi.org/10.1021/ja3031505 | J. Am. Chem. Soc. 2012, 134, 18266−18274

nium's interaction with proteins. It should be noted that some early studies also indicate that a stable β-hairpin (CLN025) cannot be unfolded with a high concentration of GdmCl, partially due to stacking of GdmCl molecules in water.[15] Therefore, the pre-equilibrated structure of GdmCl−urea mixture in the absence of protein may be important for the follow-up protein collapse. To address this question, we have performed additional simulations for the lysozyme in 4 M urea + 4 M GdmCl system, after the denaturant mixture has been pre-equilibrated for 50 ns (i.e., lysozyme is solvated in the pre-equilibrated co-denaturant solution). As expected, the same local denaturant crowding and protein collapse have been observed, as shown in Figure S11 in Supporting Information.

CONCLUSION

In general, guanidinium chloride (GdmCl) is considered to be a stronger denaturant than urea for proteins; therefore, adding more GdmCl to the solution or replacing part of the urea with GdmCl would presumably cause the protein to unfold further to a more stretched state (or at least remain at the current stretched state) if a simple additive denaturation effect is in action. However, a counterintuitive phenomenon was observed in our molecular dynamics simulations with both hen egg-white lysozyme and protein L, where the unfolded proteins collapsed in urea/GdmCl mixture compared with the single denaturant (either GdmCl or urea). We then found that the collapse was accompanied by a burying of hydrophobic residues at the protein surface and an increase of local non-native contacts, indicating that it was not a refolding process but rather a simple collapse of the denatured state.

Detailed energetic and structural analyses then showed that GdmCl molecules replaced some urea molecules in the FSS of proteins through their stronger electrostatic interactions with protein backbones. Meanwhile, the urea molecules, though some were replaced by the stronger denaturant GdmCl, still accumulate near the protein surface, creating a more crowded local environment for the protein. This rearrangement of denaturants near the protein surface and the crowded local environment induce the protein collapse, mainly by burying the hydrophobic residues, resulting in an enhanced self-interaction among the protein residues as seen in its increased non-native local contacts. These findings from detailed molecular simulations not only confirm the predictions from analytical statistical mechanics models but also provide us with a deeper molecular picture of this denaturation process: when the two denaturants compete with each other to interact with the protein, the stronger denaturant (GdmCl) has a greater tendency to move closer to the protein surface than the weaker one (urea); in order to accommodate more GdmCl near the protein surface, the protein itself will collapse somewhat in order to accommodate more GdmCl molecules. The redistribution of denaturants near the protein surface and crowded local environment induce the protein hydrophobic collapse and result in an overall minimized free energy state, as predicted by the statistical mechanics model.

Mixed solvents are considered to be a more realistic environment for proteins in nature. It is thus of fundamental importance to investigate the conformational diversity of a protein, which plays a crucial role in modulating its function in various environments. Our study indicates that the dynamic structure of a protein in mixed solvents is more complicated than we have expected, and such a detailed study of the protein solvated in different denaturants may have provided new insights into the mechanisms of protein folding and unfolding.

ASSOCIATED CONTENT

Supporting Information

Appendix providing a more complete analytical theory to explain the protein collapse phenomenon with a model polymer in a mixture of water, urea, and GdmCl, time dependence of radius of gyration in "6 M urea + 2 M GdmCl" and "4 M urea + 4 M GdmCl" mixture for another two independent simulations of lysozyme, protein solvent-accessible surface area of different types of residues in "4 M urea + 4 M GdmCl" mixture for protein L, time dependence of the backbone−backbone hydrogen bonds formed by the residue pairs of hydrophobic−hydrophobic, hydrophilic−hydrophilic, and hydrophobic−hydrophilic in "4 M urea + 4 M GdmCl" mixture for lysozyme, protein solvent-accessible surface area of different types of residues in "0 M urea + 6 M GdmCl" single denaturant system for lysozyme, protein solvent-accessible surface area of different types of residues in "0 M urea + 6 M GdmCl single denaturant system for protein L, secondary structure content (as a percentage of residues in α-helices and β-sheets) for lysozyme and protein L, time-dependent density of urea and guanidinium molecules at the first solvation shell of lysozyme for another independent trajectory, time-dependent density of urea and guanidinium molecules at the first solvation shell of protein L, time-dependent radial distribution functions of urea and GdmCl to charged side chains in "4 M urea + 4 M GdmCl" mixture, time-dependent pair radial distribution function between the hydrogen atom (HR) at the side chain of arginine and the oxygen atom of urea (OU) or the nitrogen atom (NG) of Gdm^+, and protein lysozyme collapse in pre-equilibrated "4 M urea + 4 M GdmCl" mixture. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

shakhnovich@chemistry.harvard.edu; ruhongz@us.ibm.com

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Tobin Sosnick, Bruce Berne, Pengyu Ren, and Seung-gu Kang for many helpful discussions. We also thank the IBM Blue Gene Science Program for the financial support.

REFERENCES

(1) Tanford, C. *Adv. Protein Chem.* **1970**, *24*, 1.
(2) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.
(3) Alonso, D. O.; Dill, K. A. *Biochemistry* **1991**, *30*, 5974.
(4) Auton, M.; Holthauzen, L. M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15317.
(5) Courtenay, E. S.; Capp, M. W.; Record, M. T., Jr. *Protein Sci.* **2001**, *10*, 2485.
(6) Schellman, J. A. *C. R. Trav. Lab. Carlsberg* **1955**, *29*, 230.
(7) Scholtz, J. M.; Barrick, D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 185.
(8) Kuharski, R. A.; Rossky, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 5786.
(9) Kuharski, R. A.; Rossky, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 5794.
(10) TiradoRives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997**, *36*, 7313.
(11) Caflisch, A.; Karplus, M. *Struct. Folding Des.* **1999**, *7*, 477.
(12) Tobi, D.; Elber, R.; Thirumalai, D. *Biopolymers* **2003**, *68*, 359.

(13) Kokubo, H.; Pettitt, B. M. *J. Phys. Chem. B* **2007**, *111*, 5233.
(14) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2008**, *130*, 7380.
(15) Hatfield, M. P.; Murphy, R. F.; Lovas, S. *J. Phys. Chem. B* **2011**, *115*, 4971.
(16) Wetlaufer, D; Malik, S; Stoller, L; R, C. *J. Am. Chem. Soc.* **1964**, *86*, 508.
(17) Frank, H; F, F. *J. Chem. Phys.* **1968**, *48*, 4746.
(18) Barone, G.; Rizzo, E.; Vitaglia., V *J. Phys. Chem.* **1970**, *74*, 2230.
(19) Finer, E. G.; Franks, F.; Tait, M. J. *J. Am. Chem. Soc.* **1972**, *94*, 4424.
(20) Hammes, G. G.; Schimmel, P. R. *J. Am. Chem. Soc.* **1967**, *89*, 442.
(21) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142.
(22) Robinson, D. R.; Jencks, W. P. *J. Am. Chem. Soc.* **1965**, *87*, 2462.
(23) Wallqvist, A.; Covell, D. G.; Thirumalai, D. *J. Am. Chem. Soc.* **1998**, *120*, 427.
(24) Klimov, D. K.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14760.
(25) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346.
(26) Stumpe, M. C.; Grubmuller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126.
(27) Hua, L.; Zhou, R. H.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16928.
(28) Lim, W. K.; Rosgen, J.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2595.
(29) Godawat, R.; Jamadagni, S. N.; Garde, S. *J. Phys. Chem. B* **2010**, *114*, 2246.
(30) Camilloni, C.; Rocco, A. G.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. *Biophys. J.* **2008**, *94*, 4654.
(31) Mason, P. E.; Brady, J. W.; Neilson, G. W.; Dempsey, C. E. *Biophys. J.* **2007**, *93*, L4.
(32) Grosberg, A. Y.; Erukhimovitch, I. Y.; Shakhnovitch, E. I. *Biopolymers* **1982**, *21*, 2413.
(33) Finkelstein, A. V.; Shakhnovich, E. I. *Biopolymers* **1989**, *28*, 1681.
(34) Grosberg, A. Y.; Erukhimovich, I. Y.; Skahnovich, E. I. *Biofizika* **1981**, *26*, 415.
(35) Brochard, F.; Degennes, P. G. *Fer* **1980**, *30*, 33.
(36) Vaney, M. C.; Maignan, S.; Ries-Kautt, M.; Ducriux, A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1996**, *D52*, 505.
(37) Wikstrom, M.; Drakenberg, T.; Forsen, S.; Sjobring, U.; Bjorck, L. *Biochemistry* **1994**, *33*, 14011.
(38) Zhou, R. H.; Eleftheriou, M.; Royyuru, A. K.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5824.
(39) Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 4702.
(40) Kumar, S.; Huang, C.; Zheng, G.; Bohm, E.; Bhatele, A.; Phillips, J. C.; Yu, H.; Kale, L. V. *IBM J. Res. Dev.* **2008**, *52*, 177.
(41) Brooks, C. L.; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11037.
(42) Brooks, C. L.; Onuchic, J. N.; Wales, D. J. *Science* **2001**, *293*, 612.
(43) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.
(44) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.
(45) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898.
(46) Kokubo, H.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *383*, 397.
(47) Zhou, R. H.; Huang, X. H.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605.
(48) Liu, P.; Huang, X. H.; Zhou, R. H.; Berne, B. J. *Nature* **2005**, *437*, 159.
(49) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(50) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931.
(51) Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280.
(52) Li, J.; Gong, X.; Lu, H.; Li, D.; Fang, H.; Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 3687.
(53) Tu, Y.; Xiu, P.; Wan, R.; Hu, J.; Zhou, R. H.; Fang, H. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18120.
(54) Ge, C.; Du, J.; Zhao, L.; Wang, L.; Liu, Y.; Li, D.; Yang, Y.; Zhou, R. H.; Zhao, Y.; Chai, Z.; Chen, C. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16968.
(55) MacKerell, A. D.; Bashford, D.; Bellott, M.; Evanseck, R. L. D. J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(56) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(57) Brooks, B. R.; Bruccoeri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(59) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678.
(60) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
(61) Beauchamp, K. A.; Ensign, D. L.; Das, R.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12734.
(62) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787.
(63) Cho, S. S.; Reddy, G.; Straub, J. E.; Thirumalai, D. *J Phys Chem B* **2011**, *115*, 13401.
(64) Zhou, R. H.; Li, J. Y.; Hua, L.; Yang, Z. X.; Berne, B. J. *J. Phys. Chem. B* **2011**, *115*, 1323.
(65) Xiu, P.; Yang, Z. X.; Zhou, B.; Das, P.; Fang, H. P.; Zhou, R. H. *J. Phys. Chem. B* **2011**, *115*, 2988.
(66) Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2011**, *115*, 1327.
(67) Gao, M.; She, Z. S.; Zhou, R. H. *J. Phys. Chem. B* **2010**, *114*, 15687.
(68) Das, P.; Zhou, R. H. *J. Phys. Chem. B* **2010**, *114*, 5427.
(69) Zangi, R.; Zhou, R. H.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131*, 1535.