Equilibrium Study of Protein Denaturation by Urea

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · FEBRUARY 2010

Impact Factor: 12.11 · DOI: 10.1021/ja909348c · Source: PubMed

3 AUTHORS, INCLUDING:

Deepak R Canchi
Rensselaer Polytechnic Institute
4 PUBLICATIONS **263** CITATIONS
SEE PROFILE

Angel Garcia
Rensselaer Polytechnic Institute
209 PUBLICATIONS **10,168** CITATIONS
SEE PROFILE

J|A|C|S
ARTICLES

Equilibrium Study of Protein Denaturation by Urea

Deepak R. Canchi,[†] Dietmar Paschek,[‡] and Angel E. García*[,‡,§]

Department of Chemical and Biological Engineering, Department of Physics, Applied Physics and Astronomy, and Center for Biotechnology and Interdisciplinary studies, Rensselaer Polytechnic Institute, Troy, New York 12180

Received November 4, 2009; E-mail: angel@rpi.edu

Abstract: Though urea is commonly used to denature proteins, the molecular mechanism of its denaturing ability is still a subject of considerable debate. Previous molecular dynamics simulation studies have sought to elucidate the mechanism of urea denaturation by focusing on the pathway of denaturation rather than examining the effect of urea on the folding/unfolding equilibrium, which is commonly measured in experiment. Here we report the reversible folding/unfolding equilibrium of Trp-cage miniprotein in the presence of urea, over a broad range of urea concentrations, using all-atom Replica exchange MD simulations. The simulations capture the experimentally observed linear dependence of unfolding free energy on urea concentration. We find that the denaturation is driven by favorable direct interaction of urea with the protein through both electrostatic and van der Waals forces and quantify their contribution. Though the magnitude of direct electrostatic interaction of urea is larger than van der Waals, the difference between unfolded and folded ensembles is dominated by the van der Waals interaction. We also find that hydrogen bonding of urea to the peptide backbone does not play a dominant role in denaturation. The unfolded ensemble sampled depends on urea concentration, with greater urea concentration favoring conformations with greater solvent exposure. The *m*-value is predicted to increase with temperature and more strongly so with pressure.

Introduction

Proteins exhibit marginal stability, determined by the balance of many competing effects. This stability can be perturbed by changes in temperature, pressure, and solvent conditions.[1−4] Osmolytes are small organic compounds that modulate the conformational equilibrium, folded (F) ⇌ unfolded (U), of proteins as cosolvents. Protecting osmolytes such as trimethylamine *N*-oxide (TMAO), glycerol, and sugars that push the equilibrium toward F play a crucial role in maintaining the function of intracellular proteins in extreme environmental conditions.[5] Urea is a denaturing osmolyte that shifts the equilibrium toward U.

The ability of urea to denature proteins has been long known, and it is widely used in protein folding studies.[6−8] The molecular mechanism of urea denaturation has attracted active debate, and a consensus is still elusive. There are two basic lines of reasoning proposed for the denaturing effect of urea. The "indirect mechanism" postulates that urea promotes ready solvation of hydrophobic groups by altering water structure. This view emerged from transfer experiments that showed hydrocarbons to be more soluble in aqueous urea[9] and has been questioned by recent studies.[10−14] The other view attributes the denaturation to the "direct" interactions of the protein with urea.[14−20] Within the direct interaction model, there is debate on the nature and strength of the interactions of urea with the polar/apolar side chains and the peptide backbone, and on the role of hydrogen bonding.

Theoretical models based on the transfer free energies of side chains and peptide backbone predict that the driving force for denaturation stems from favorable interaction of urea with the peptide backbone.[21,22] MD simulations of small hydrophobic solutes in aqueous urea have led to conflicting results on the

[†] Department of Chemical and Biological Engineering.
[‡] Department of Physics, Applied Physics and Astronomy.
[§] Center for Biotechnology and Interdisciplinary Studies.

(1) Privalov, P. *Adv. Protein Chem.* **1979**, *33*, 167–241.
(2) Privalov, P. L. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 281–305.
(3) Heremans, K. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 1–21.
(4) Timasheff, S. N. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67–97.
(5) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214–1222.
(6) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1–63.
(7) Greene, R. F.; Pace, C. N. *J. Biol. Chem.* **1974**, *249*, 5388–5393.
(8) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266–280.
(9) Frank, H. S.; Franks, F. *J. Chem. Phys.* **1968**, *48*, 4746–4757.
(10) Soper, A. K.; Castner, E. W.; Luzar, A. *Biophys. Chem.* **2003**, *105*, 649–666.
(11) Batchelor, J. D.; Olteanu, A.; Tripathy, A.; Pielak, G. J. *J. Am. Chem. Soc.* **2004**, *126*, 1958–1961.
(12) Rezus, Y. L. A.; Bakker, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18417–18420.
(13) Kokubo, H.; Pettitt, B. M. *J. Phys. Chem. B* **2007**, *111*, 5233–5242.
(14) Hua, L.; Zhou, R. H.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16928–16933.
(15) Robinson, D. R.; Jencks, W. P. *J. Am. Chem. Soc.* **1965**, *87*, 2462–2470.
(16) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1992**, *226*, 491–505.
(17) Tobi, D.; Elber, R.; Thirumalai, D. *Biopolymers* **2003**, *68*, 359–369.
(18) Klimov, D. K.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14760–14765.
(19) Stumpe, M. C.; Grubmuller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126–16131.
(20) Zangi, R.; Zhou, R. H.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131*, 1535–1541.
(21) Auton, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15065–15068.
(22) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15317–15322.

effect of urea on hydrophobic interactions.[23−27] On the other hand, urea is found to interact favorably with charged solutes through hydrogen bonding and disrupt ion pairs.[23] Previously reported simulation studies of proteins in aqueous urea have sought to understand the phenomenon mechanistically, i.e., unfold the protein starting from the native state.[14,27−32] Some studies attributed the denaturation to both "direct" and "indirect" mechanisms,[28,29] while others emphasized the role of hydrogen bonding between urea and protein.[17,18,27] The importance of direct van der Waals interaction of urea has been recognized in recent studies of model systems, amino acids, and proteins.[14,19,20,33−35]

Though the different modes of interaction of urea with proteins have been identified, the balance of the driving forces is not yet clear. At the molecular level, is the driving "direct" interaction electrostatic or van der Waals? At the thermodynamic level, is the dominant driving force enthalpic or entropic? We seek to answer these questions by providing a thermodynamic description of urea denaturation of a protein, the Trp-cage miniprotein, using all-atom Replica exchange molecular dynamics (REMD) simulations.

Obtaining the folding/unfolding equilibrium of a biomolecule by molecular simulation presents a major computational challenge, even for small systems. Trp-cage miniprotein is a model protein designed to understand protein folding pathways and stability.[36] It is a 20-residue protein that is well characterized and folds co-operatively to a tertiary structure.[37] Its structure has been determined by NMR (Protein Data Bank code 1L2Y), and its melting behavior has been characterized by various methods.[38−41] Its small size and fast folding kinetics make it a computationally feasible target, and many studies using both implicit and explicit solvent models have been reported.[42−51]

Table 1. System details

concn (M)	no. of urea	no. of water	replica range (K)	box length (nm)
1.9	105	2637	280−579	4.50904
3.8	233	2637	280−557	4.67515
5.8	391	2637	280−537	4.83007
0	0	2637	280−590	4.35826

Trp-cage shows the thermodynamic features observed for globular proteins, such as the temperature dependence of unfolding free energy and enthalpy.[51] Here, we present unbiased REMD simulations, in all atomic detail, of Trp-cage in aqueous urea. REMD enables a study of equilibrium over a range of temperatures, allowing calculations of entropic effects on protein stability. We also study the dependence of the folding/unfolding equilibrium on urea concentration, which allows us to model the free energy of unfolding as $\Delta G_u(P,T,[C])$. We calculate the *m*-value from our simulations, and examine its temperature and pressure dependence. We also obtain structural and energetic insights into the process of urea denaturation.

Methods

REMD simulations[52] are used to study the unbiased equilibrium folding/unfolding of Trp-cage miniprotein at different concentrations of urea, using the Amber (ff94) forcefield,[53] the TIP3P model for water,[54] and the Kirkwood-Buff model for urea.[55] We simulated systems at three different concentrations of urea, i.e., 1.9, 3.8, 5.8, and 0 M (water). All systems consisted of 50 replicas and were simulated for 150 ns per replica, except for 0 M, which was simulated for 200 ns per replica. This amounts to a total sampling of 32.5 μs. The system details are summarized in Table 1. The analysis presented here was carried out on the last half of the simulations, after allowing for convergence.

The Trp-cage sequence was generated in an initially all-PP2 conformation using LEAP program distributed with AMBER 6.0. The peptide consists of 313 atoms. This structure was simulated in the gas phase at 300 K to obtain an unstructured conformation, which is solvated in a cubic box of urea, and then water to obtain the desired concentrations. The simulation boxes thus obtained are energy-minimized and equilibriated for volume in a constant-pressure simulation for 2 ns at 300 K and 1 atm. The final box lengths are given in Table 1. The peptide is completely unfolded and lacks secondary structure elements in the final configuration of the equilibriation runs at all concentrations. The respective structures are used to start the REMD simulations at various concentrations.

(23) Wallqvist, A.; Covell, D. G.; Thirumalai, D. *J. Am. Chem. Soc.* **1998**, *120*, 427–428.
(24) Shimizu, S.; Chan, H. S. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 560–566.
(25) Oostenbrink, C.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2005**, *7*, 53–58.
(26) Lee, M. E.; van der Vegt, N. F. A. *J. Am. Chem. Soc.* **2006**, *128*, 4948–4949.
(27) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346–7353.
(28) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142–5147.
(29) Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. *Biophys. J.* **2005**, *89*, 842–857.
(30) Stumpe, M. C.; Grubmuller, H. *PLoS Comput. Biol.* **2008**, *4* (11), e1000221.
(31) Smith, L. J.; Jones, R. M.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2005**, *58*, 439–449.
(32) Gattin, Z.; Riniker, S.; Hore, P. J.; Mok, K. H.; van Gunsteren, W. F. *Protein Sci.* **2009**, *18*, 2090–2099.
(33) Graziano, G. *J. Phys. Chem. B* **2001**, *105*, 2632–2637.
(34) Trzesniak, D.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2004**, *6*, 697–702.
(35) England, J. L.; Pande, V. S.; Haran, G. *J. Am. Chem. Soc.* **2008**, *130*, 11854–11855.
(36) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
(37) Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Biol.* **2002**, *9*, 408–410.
(38) Qiu, L. L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
(39) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943–10950.
(40) Neuweiler, H.; Doose, S.; Sauer, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16650–16655.
(41) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876–2880.
(42) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.
(43) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549.
(44) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.
(45) Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711–717.
(46) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147–155.
(47) Ota, M.; Ikeguchi, M.; Kidera, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17658–17663.
(48) Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
(49) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.
(50) Paschek, D.; Nymeyer, H.; Garcia, A. E. *J. Struct. Biol.* **2007**, *157*, 524–533.
(51) Paschek, D.; Hempel, S.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–17759.
(52) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(53) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(55) Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B.* **2003**, *107*, 3891–3898.

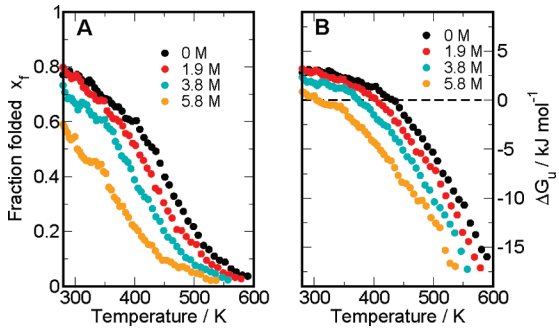

Figure 1. Dependence of folding/unfolding equilibrium on urea concentration. (A) Fraction of folded states as a function of temperature. (B) Free energy of unfolding as a function of temperature.

REMD is an enhanced sampling technique where multiple copies (replicas) of identical systems are simulated at different temperatures in parallel. State-exchange moves are attempted periodically, where two neighboring replicas exchange their temperature. The acceptance rule for each state-exchange move between adjacent states i and j is given by

$$P_{\mathrm{acc}} = \min[1, \exp[(\beta_i - \beta_j)(U(\vec{r}_i^N) - U(\vec{r}_j^N))]] \qquad (1)$$

where $\beta = 1/k_B T$ and $U(\vec{r}_i^N)$ represents the configurational energy of the system in state i. The temperature spacing was chosen so that the energy distributions overlap sufficiently, and state exchange attempts are accepted with a probability of 0.2.[56] To obtain the temperature spacings, energy distributions at select temperatures were obtained from NVT simulations of the starting structure of the REMD simulations. REMD simulations were carried out using GROMACS 4,[57] with a time step of 2 fs and state exchange moves attempted, between all adjacent replicas, every 4 ps. With this exchange protocol, the approach to equilibrium was found to be similar to that of previous studies.[50,51] The REMD simulations were carried out in NVT ensemble, using a Nosé-Hoover thermostat with a coupling constant of 0.5 ps.[58,59] Solvent constraints were solved using SETTLE,[60] and other bond constraints were imposed using LINCS.[61] The electrostatic interactions were treated by smooth-particle mesh Ewald summation.[62] Appropriate Lennard-Jones long-range correction for energy and pressure were taken into account.

Results and Discussion

All-atom REMD simulations were performed, starting from an unfolded configuration to avoid any bias, to sample the folding/unfolding equilibrium of the Trp-cage miniprotein at different urea concentrations, viz., 1.9, 3.8, 5.8, and 0 M (water). On the basis of the results obtained from the simulation of Trp-cage in water,[51] we use the C_α rmsd from the NMR structure (1L2Y, frame 1) as a criterion to distinguish between folded and unfolded configurations, with a cutoff value of 0.23 nm.

The advantage of REMD lies in the direct computation of free energies and free energy landscapes over a range of temperatures. Figure 1A shows the fraction of folded states as a function of temperature for the three different concentrations of urea, along with the result for the simulation in water. Figure 1B shows the Gibbs free energy of unfolding, which can be computed from the knowledge of fraction folded using the relation $\Delta G_u = -RT \ln[(1 - x_f)/x_f]$, assuming volume changes are small. We clearly observe the effect of urea on the folding/unfolding equilibrium, with increasing urea concentration shifting the equilibrium further toward the unfolded ensemble. This alteration in stability is reflected in the decrease of the melting temperature from 430 K for 0 M to 304 K for 5.8 M. In Figure 2, we compare the free energy landscapes of Trp-cage, as a function of rmsd and temperature, obtained for 0 and 5.8 M systems. The folded basin is diminished and the unfolded basin is expanded in 5.8 M urea when compared to 0 M. At lower temperatures, the unfolded ensemble is more diverse in urea than water. Thus, addition of urea has the effect of flattening the free energy landscape.

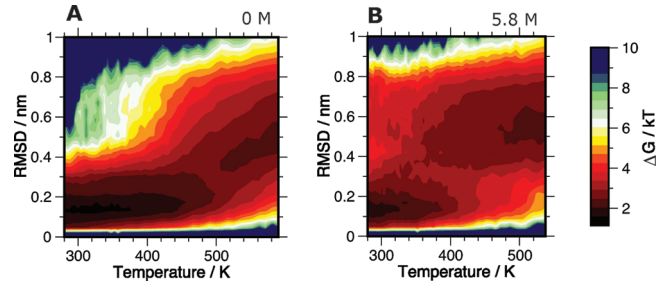

Figure 2. Free energy landscape of Trp-cage, as a function of temperature and rmsd in (A) 0 M urea, i.e., water and (B) 5.8 M urea. The free energy is shown in units of kT.

Changing urea concentration not only shifts the equilibrium toward U but also leads to the sampling of qualitatively different unfolded ensembles, as shown in Figure 3. Since the folded state is defined by a geometic criterion, the distributions of solvent accessible surface area (SAS) for the folded ensemble at different concentrations of urea at 300 K are nearly identical and is indicated by the dashed line in Figure 3A and B. At 300 K, the unfolded ensemble shows a bimodal distribution in SAS, with the first peak resembling folded-like states and the second peak representing states with a greater degree of unfolding (Figure 3A). The relative population is altered by urea concentration, with increasing concentration favoring greater solvent exposure. As the temperature is increased to 450 K, we sample broad distributions of SAS that show further shift toward states with greater solvent exposure (Figure 3B). The radius of gyration (RG) of the unfolded ensemble is also dependent on the urea concentration (Figure 3C). The unfolded ensembles can be distinguished on the basis of the average values of SAS and RG over the entire concentration and temperature range studied. While this is shown in Figure 3D for the case of RG, the behavior for SAS is similar. The RG appears to decrease with temperature before increasing again; however, the errors are large at lower temperatures, and further investigation is required to confirm if there is any "re-entrant" behavior. The dependence of the unfolded ensemble on denaturant concentration has been anticipated theoretically[63] and shown in experiment and simulation.[64,65]

(56) Garcia, A.; Herce, H.; Paschek, D. In *Annual Reports in Computational Chemistry*; Spellmeyer, D., Ed.; Elsevier: Amsterdam, 2006; Vol. 2, pp 83−97.
(57) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(58) Nose, S. *Mol. Phys.* **1984**, *52*, 255–268.
(59) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(60) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
(61) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
(62) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(63) Alonso, D. O. V.; Dill, K. A. *Biochemistry* **1991**, *30*, 5974–5985.
(64) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743–747.
(65) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528–1533.

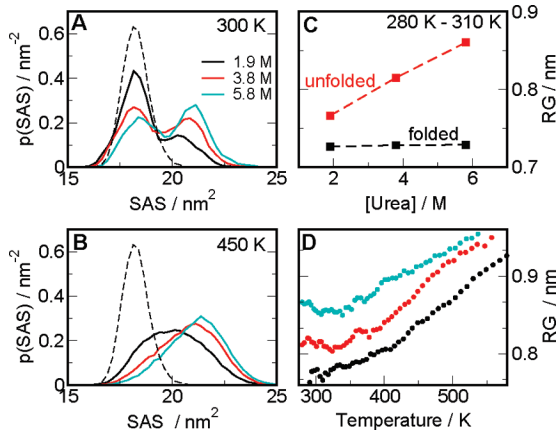

Figure 3. Distinguishing between unfolded ensembles at different concentrations of urea. (A) Distributions of SAS for the unfolded ensemble at 300 K. (B) Distributions of SAS for unfolded ensemble at 450 K. (C) RG for folded and unfolded ensembles, averaged over 280−310 K. (D) RG for the unfolded ensembles as a function of temperature. The dashed line in A and B is representative of the SAS distribution for the folded ensemble at 300 K.

It is experimentally observed that the free energy of unfolding of proteins varies linearly with urea concentration.[7] This observation is widely used to estimate the stability of proteins in water via the linear extrapolation model, i.e., $\Delta G^{\text{urea}} = \Delta G^{\text{water}} - m[\text{C}]$.[8] The slope of the linear fit, known as the m-value, then measures the response of protein stability to the addition of urea. The m-value can be interpreted as the difference in the transfer free energies of denatured and native states going from water to 1 M urea solution[21] and correlates well with the size of protein and change in solvent exposed surface upon unfolding.[66]

To obtain thermodynamic information about the denaturation process, we model the unfolding free energy as a function of temperature, pressure, and concentration, $\Delta G_u(P, T,[\text{C}])$, and expand it to second order around a reference state as given by eq 2 to obtain a Hawley-type free energy surface.[67] Additional terms have been added to include the effect of urea on free energy (m_1), entropy (m_2), volume (m_3), and specific heat (m_4). We do not include $\Delta\beta$ in our model.

$$\begin{aligned}\Delta G_u(P,T,[\text{C}]) = {} & \Delta G_0 + m_1[\text{C}] \\ & + (-\Delta S_0 + m_2[\text{C}])(T - T_0) \\ & + (\Delta V_0 + m_3[\text{C}])(P - P_0) \\ & + \Delta\alpha(P - P_0)(T - T_0) \\ & - (\Delta C_p + m_4[\text{C}])[T(\ln(T/T_0) - 1) + T_0]\end{aligned} \quad (2)$$

The fraction folded data, ΔE_u (potential energy differences obtained from the simulations) data for all temperatures and concentrations studied, are fit simultaneously to the thermodynamic model above. The fit did not require a dependence of expansivity on urea concentration. The obtained fit parameters are listed in Table 2.

In Figure 4A, we show the free energy obtained from fraction folded data by averaging in the temperature range 280−310 K along with the fit obtained at the reference state, $T = T_0 = 300$ K and $P = P_0 = 0.1$ MPa. The free energy is linear with urea

Table 2. Fit Parameters for the Thermodynamic Model[a]

ΔG_0	3.26 kJ mol^{-1}
ΔV_0	4.54 mL mol^{-1}
ΔS_0	4.13×10^{-3} kJ mol^{-1} K^{-1}
ΔC_p	1.14×10^{-1} kJ mol^{-1} K^{-1}
$\Delta\alpha$	-3.53×10^{-2} ml mol^{-1} K^{-1}
m_1	-4.12×10^{-1} kJ mol^{-1} K^{-1}
m_2	-9.80×10^{-6} kJ mol^{-1} K^{-1} M^{-1}
m_3	-1.89×10^{-3} ml mol^{-1} M^{-1}
m_4	4.51×10^{-3} kJ mol^{-1} K^{-1} M^{-1}

[a] The reference state is 300 K and 0.1 MPa.

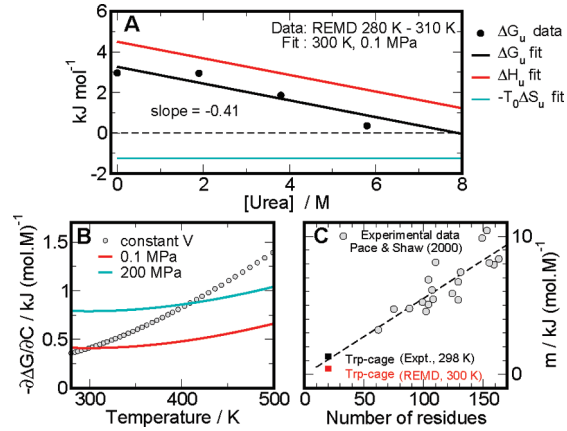

Figure 4. Thermodynamic features of urea denaturation. (A) Free energy of unfolding as function of urea concentration, along with enthalpic and entropic contributions. (B) $-\partial\Delta G_u/\partial C$ as a function of temperature along two isobars and a isochore (C) Comparison of m-values from simulation and experiment. Red and black squares are m-values for Trp-cage from simulation and experiment,[70] respectively. The dashed line is an extrapolation of experimental m-values for larger proteins taken from Pace and Shaw.[71]

concentration, which shows that the all-atom model employed is able to capture the experimental trend. From the thermodynamic model, we can decompose the free energy into enthalpic and entropic contributions. In Figure 4A, we see that the sign of ΔG_u as well as its behavior with urea concentration is dictated by the enthalpy term. The entropic term favors unfolding as expected and is very weakly concentration-dependent. It has been argued that that preferential binding of urea to the protein is entropically favorable, as it involves displacement of water by larger urea molecules.[19,68,69] In view of this argument, the weak concentration dependence of entropy is intriguing. There could be compensatory effects at play, e.g., loss of orientational entropy of urea molecules.

The quantity $-\partial\Delta G_u/\partial C$ can be thought of as a generalized m-value, dependent on temperature and pressure, which reduces to the traditional m-value at the reference state, $T = T_0$ and $P = P_0$. The dependence on various terms is shown in eq 3.

$$-\frac{\partial\Delta G_u}{\partial C} = -m_1 - m_2(T - T_0) - m_3(P - P_0) - m_4[T(\ln(T/T_0) - 1) + T_0] \quad (3)$$

(66) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138–2148.
(67) Hawley, S. A. *Biochemistry* **1971**, *10*, 2436–2442.
(68) Rossky, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16825–16826.
(69) Sagle, L. B.; Zhang, Y. J.; Litosh, V. A.; Chen, X.; Cho, Y.; Cremer, P. S. *J. Am. Chem. Soc.* **2009**, *131*, 9304–9310.
(70) Wafer, L. N. R.; Streicher, W. W.; Makhatadze, G. I. *Proteins* **2010**, DOI 10.1002/prot.22681.
(71) Pace, C. N.; Shaw, K. L. *Proteins: Struct., Funct., Genet.* **2000**, *4*, 1–7.

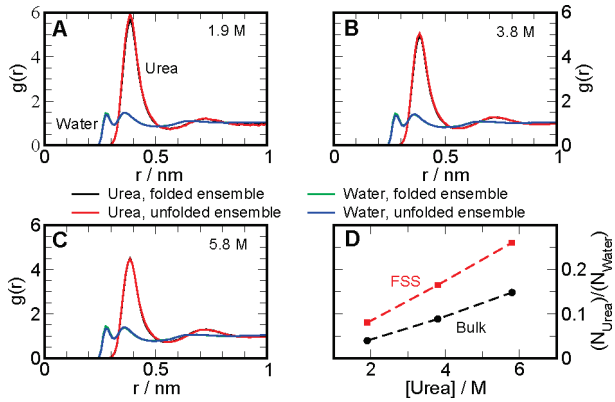

Figure 5. Local solvent structure around the protein. (A−C) Proximal radial distribution functions at 300 K shown for water and urea, separately for folded and unfolded ensembles, at each urea concentration. (D) Solvent composition in FSS and in bulk at 300 K.

In Figure 4B, we plot $-\partial\Delta G_u/\partial C$ as a function of temperature at two different pressures and along an isochore. Increase in this quantity with temperature and pressure indicates that denaturation by urea is aided by both of these factors, with the effect of pressure being markedly stronger. The *m*-value calculated from the model at 300 K and 0.1 MPa is 0.41 ± 0.03 kJ mol^{-1} M^{-1}. At 300 K and 200 MPa, the *m*-value is calculated to be 0.79 kJ mol^{-1} M^{-1}, almost double the value at ambient conditions. The strong pressure dependence arises from the term $m_3(P - P_0)$, whose contribution to volume changes (i.e., $\partial\Delta G_u/\partial P$) is negligible. The temperature dependence is weaker compared to pressure, giving a *m*-value of 0.66 kJ mol^{-1} M^{-1} at 500 K and 0.1 MPa. The contribution of the m_2 term to the temperature dependence of *m*-value is negligible. In Figure 4C, we plot experimental *m*-values for larger proteins obtained from Pace and Shaw[71] as a function of number of residues and the linear extrapolation of this data as a dashed line. The experimental *m*-value for Trp-cage is 1.3 kJ mol^{-1} M^{-1},[70] which agrees very well with the extrapolation, while the *m*-value obtained from simulation is in the same order of magnitude.

To probe the local solvent density around the protein, we calculate the proximal radial distribution function (RDF).[51,72,73] The calculation employs heavy atoms of the protein, the water oxygen, and the urea carbon as the solute sites. In Figure 5A−C, the proximal RDF has been calculated for both water and urea, separately for the folded and the unfolded ensembles at 300 K. At all concentrations, we find a peak for urea at ∼0.4 nm from the protein surface indicating an accumulation of urea in the first solvation shell. A priori, one would expect an enrichment of urea and a depletion of water for the unfolded ensemble vis-a-vis the folded ensemble in a preferential interaction scenario. However, the calculated proximal RDFs are almost indistinguishable, i.e., Trp-cage is able to fold and unfold without significant changes in the local solvent configuration. In other words, the increase in number of solvent molecules in the vicinity of the protein upon unfolding is exactly compensated by the increase in its surface area. A visual illustration is provided by Figure 6, which shows representative configurations of folded and unfolded ensembles from the 5.8 M system at 300 K. This behavior, also observed at higher temperatures,

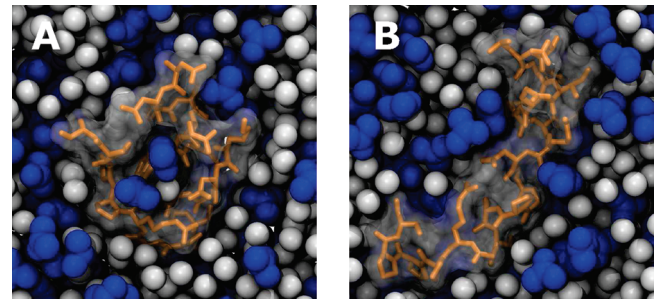

Figure 6. Representative configurations of Trp-cage for the 5.8 M system at 300 K from (A) folded ensemble and (B) unfolded ensemble. Urea is shown in blue and water in white. Hydrogen atoms are not shown for clarity.

could be attributed to the small size of the protein and the absence of a large hydrophobic core sequestered from the solvent. The two-peak feature for the water RDFs is due to polar and apolar hydration.[51] In Figure 5D, we have calculated the ratio (N_{urea}/N_{water}) in the first solvation shell (FSS, defined as 0.5 nm from the protein heavy atoms) in comparison to bulk. The differences in the ratio for the FSS upon unfolding, $\Delta^{FSS}(N_{urea}/N_{water})$, are modest. We define the coordination number of water around urea as the number of water molecules located within 0.5 nm of a urea molecule. This quantity, larger for urea molecules in the bulk than for those in the FSS, is also found to be insensitive to the folding-unfolding transition (data not shown).

We can obtain further insight into the process by examining the energy differences that accompany the unfolding process. We observe greater concentration of urea in the FSS relative to bulk, which prompts an examination of the "direct interaction" of urea with the protein. To do this, we calculate the nonbonded interaction energy, E^{FSS}, of the protein with its local solvent environment, i.e., with both urea (E^{PU}) and water (E^{PW}) molecules located in the FSS. We then calculate the difference upon unfolding, ΔE^{FSS}, and examine the relative contribution of the Coulomb (CB) and Lennard-Jones (LJ) interactions. The calculation scheme is described in eqs 4 and 5. Negative values of ΔE indicate a favorable driving force toward unfolding.

$$E^{FSS} = E^{PU} + E^{PW} = E^{PU}_{LJ} + E^{PU}_{CB} + E^{PW}_{LJ} + E^{PW}_{CB} \quad (4)$$

$$\Delta E^{FSS} = E^{FSS}_{unfolded} - E^{FSS}_{folded} \quad (5)$$

Figure 7A and B show the result for urea and water, respectively, obtained by averaging over configurations in temperature range 280−310 K, while Figure 7C and D are obtained for the temperature range 410−460 K. The temperature ranges are chosen to correspond to different signs of ΔG_u. The results obtained are only weakly dependent on the distance cutoff used in defining the FSS. We find that urea interacts favorably with the unfolded ensemble at both temperature ranges considered, as indicated by negative values for the components of ΔE^{PU}. Figure 7A and C show that both the Coulomb and Lennard-Jones interactions of the protein with urea get stronger with increasing urea concentration at both the temperature ranges. Lennard-Jones is the dominant interaction at the lower temperature range, about 1.8 times more favorable than Coulomb at 5.8 M. At higher temperature, ΔE^{PU}_{CB} is slightly more favorable than ΔE^{PU}_{LJ}. Note that the protein−urea Coulomb interaction for 5.8 M at higher temperature is more than twice the interaction at lower temperature. In Figure 7B and D, we observe the ΔE^{PW}_{CB}

(72) Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1980**, *102*, 4287–4294.
(73) Ashbaugh, H. S.; Paulaitis, M. E. *J. Phys. Chem.* **1996**, *100*, 1900–1913.

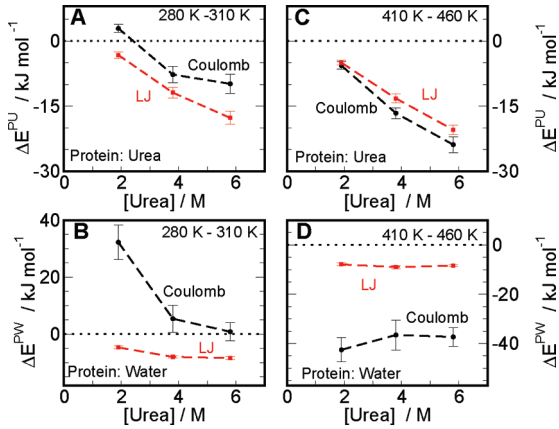

Figure 7. Interaction energy changes upon unfolding of the protein with its local solvent environment, in two different temperature ranges. (A) Protein−urea for 280−310 K. (B) Protein−water for 280−310 K. (C) Protein−urea for 410−460 K. (D) Protein−water for 410−460 K. LJ interactions are denoted by the red curve and the Coulomb by black.

to show very different behavior across the two temperature ranges, while the ΔE_{LJ}^{PW} is relatively unchanged. At the lower temperature range, ΔE_{CB}^{PW} is unfavorable and shows a sharp concentration dependence whereas at higher temperatures the interaction is very favorable and varies slightly with concentration. The origin of this behavior may lie in the unfolded ensemble sampled in the two temperature ranges. At the lower temperatures, the configurations in the unfolded ensemble are rather compact, and favorable interaction of urea could come at a cost of replacing waters in close vicinity. At higher temperatures, the protein samples extended configurations with low helical content, and this provides ample opportunity for both water and urea to interact with the different moieties of the protein.

It has been speculated that denaturation by urea is driven by hydrogen bonding to the backbone,[17,18,27] and recently it was demonstrated experimentally that urea participates in hydrogen bonding with the peptide group.[74] In the framework of a classical MD forcefield, a hydrogen bond is defined by a geometric criterion, and it is not possible to assign an energy for the bond formation that is independent of Coulomb and Lennard-Jones interactions. We use an acceptor−hydrogen distance of 0.26 nm and donor−hydrogen−acceptor angle >90° as the existence criterion for a hydrogen bond.[75] We calculate the hydrogen bond statistics between the protein backbone and the solvent, separately for the folded and the unfolded ensembles, and present the difference ΔN_u upon unfolding at two temperature ranges in Figure 8. Note that N corresponds to the number of bonds between solvent and the backbone, rather than the number of molecules participating in hydrogen bonding with the backbone. Averaging over both the folded and unfolded ensembles, the number of urea-backbone hydrogen bonds was found to be ∼4, 7, and 10 for 1.9, 3.8, and 5.8 M, respectively in the temperature range 280−310 K. This indicates the propensity of urea to form hydrogen bonds with the backbone, with *both* the folded and unfolded ensembles. Figure 8A shows only a slight difference between the unfolded and folded ensembles for urea, with only the 5.8 M system showing a difference greater than one hydrogen bond at higher temperatures, while the behavior for water in Figure 8B shows greater

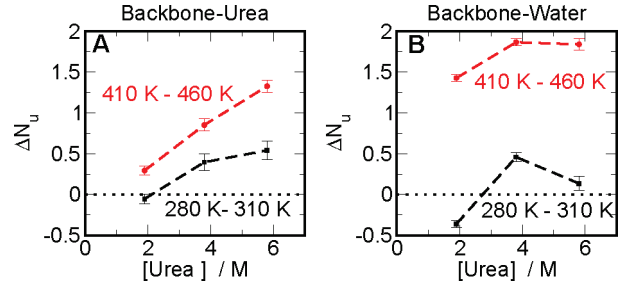

Figure 8. Change in number of backbone-solvent hydrogen bonds upon unfolding. (A) Urea. (B) Water. Low temperature data is shown in black and high temperature in red.

difference across the temperature ranges studied. The temperature dependence arises because the protein samples more extended and less helical configurations at higher temperatures. We find the hydrogen bond data to correlate well with the protein-urea and protein-water Coulomb interactions presented in Figure 7, with the increase in ΔN_u reflected by more negative values for the respective ΔE_{CB}, both as function of temperature and concentration. A noteworthy aspect is that a positive value of ΔN_u for urea is not accompanied by corresponding negative value for water, implying that urea is not competing with water for backbone hydrogen bonds. It is also seen that the total number of hydrogen bonds formed by the backbone in the unfolded ensemble with the solvent, i.e., urea and water taken together, increases with urea concentration over the entire temperature range studied. These results can be easily explained, given the concentration and temperature behavior of RG and SAS for the unfolded ensemble (see Figure 3D). The propensity of urea to form hydrogen bonds with the backbone is observed; however, the energy data of Figure 7A as well as the small differences in number of backbone-urea hydrogen bonds between folded and unfolded ensembles suggests that hydrogen bonding does not play the dominant role in the denaturation process. It would be interesting to carry out a similar analysis at equilibrium to ascertain the action of urea in the destabilization of RNA, where a recent study reported stacking interactions and multiple hydrogen bonding of urea with the nucleic acid bases.[76]

Conclusion

In order to obtain an equilibrium view of the denaturation of proteins by urea, we have sampled the folding-unfolding free energy of Trp-cage at various concentrations of urea. This enabled us to make comparisons between the folded and unfolded ensembles over a range of urea concentration. The effect of changing urea concentration is seen on the folding/unfolding equilibrium, the unfolded ensemble sampled, and the energetics of protein−solvent interactions. We have also computed the *m*-value from our simulations and predicted its behavior with temperature and pressure, which may be tested experimentally.

There is growing recognition of the role of van der Waals interaction in the process of urea denaturation. Zangi et al. demonstrated that a hydrophobic polymer unfolds in urea because of stronger intermolecular van der Waals interaction compared with intramolecular,[20] while England et al. showed

(74) Lim, W. K.; Rosgen, J.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2595–2600.
(75) Gnanakaran, S.; Hochstrasser, R.; Garcia, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9229–9234.
(76) Priyakumar, U. D.; Hyeon, C.; Thirumalai, D.; MacKerrel, A. D., Jr. *J. Am. Chem. Soc.* **2009**, *131*, 17759–17761.

that the attraction between hydrophobic plates is reduced in urea medium when compared to water.[35] Recent studies by Stumpe and Grubmüller, based on contact preferences of amino acids to urea, have shown that apolar residues interact preferentially with urea, i.e., the interaction is of van der Waals character.[19,30] The extensive simulations of Hua et al. showed that stronger dispersion interaction of urea with the protein led to the intrusion of urea in the protein interior.[14] However, the relative importance of van der Waals and electrostatic interaction was not known.

Decomposing the protein-urea interaction energies for unfolding ΔE^{PU} into LJ and Coulomb components shows that both interactions favor unfolding. The LJ interaction is the dominant contribution in the temperature range 280−310 K for our system, whereas the Coulomb interaction is slightly more favorable at higher temperatures. While the details are specific to our system and the forcefield employed, we can speculate how Figure 7A changes qualitatively in different scenarios. For a larger protein in urea, we still expect the LJ to be more dominant than Coulomb. For denaturants stronger than urea, we expect the total direct interaction, ΔE^{FSS}, to be more favorable than urea and with different relative contributions to ΔE^{FSS}. For 1,3-dimethyl urea, a stronger denaturant than urea,[77] the LJ curve is expected to be shifted to more negative values than urea and the Coulomb shifted to more positve values since it is a larger apolar molecule. Again in the case of the stronger GdmCl, we expect a similar LJ curve as for urea (4 van der Waal interaction sites each), but the Coulomb curve is expected to be shifted to much larger negative values than urea.

The relative dominance of LJ interaction and the distribution of unfolded ensemble SAS of Figure 3, taken together with the fact that large concentration of urea is usually required to denature a protein, points to the importance of weak, nonspecific interactions in the denaturation process. On the basis of our data, we cannot entirely attribute the denaturation to the specific interaction of hydrogen bonding with the backbone, the contribution of which is subsumed in the Coulomb interaction. A similar conclusion on the hydrogen bonding model was reached recently Sagle et al. from their study of PNIPAM, a protein mimetic.[69] Their results, ascribed to entropic considerations, can also be interpreted in terms of more favorable LJ interactions.

We now have a picture for how urea denatures proteins. Urea is a small organic molecule, which is soluble in water in large concentrations and incorporates itself well in the hydrogen-bond network of water. It has a higher van der Waal site density than water and can accept and donate hydrogen bonds. Urea accumulates in the vicinity of the protein due to favorable "direct" interaction with the protein. The unfolded ensemble has more favorable interaction with urea than the folded ensemble, which provides an enthalpic driving force for unfolding.

Using present simulation techniques and parallel computing power, we have characterized the protein−solvent interactions in a two-solvent system and studied the effects on protein conformation. This study can be easily extended to other denaturants to obtain insights into their mechanism. A challenging extension would be to study the compensating effect on protein stability in a mixed-osmolyte system for, e.g., urea-TMAO-water. Our study also underlines the utility of developing forcefields that are parametrized to reproduce experimental thermodynamic data.[55]

Acknowledgment. The authors thank G. I. Makhatadze for valuable discussions. This work has been supported by the National Science Foundation (Grant MCB-0543769).

JA909348C

(77) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063–8068.