# Backbone and Side-Chain Contributions in Protein Denaturation by Urea

Deepak R. Canchi[†] and Angel E. García[‡*]
[†]Department of Chemical and Biological Engineering and [‡]Department of Physics, Applied Physics and Astronomy, and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York

ABSTRACT Urea is a commonly used protein denaturant, and it is of great interest to determine its interaction with various protein groups to elucidate the molecular basis of its effect on protein stability. Using the Trp-cage miniprotein as a model system, we report what we believe to be the first computation of changes in the preferential interaction coefficient of the protein upon urea denaturation from molecular-dynamics simulations and examine the contributions from the backbone and the side-chain groups. The preferential interaction is obtained from reversible folding/unfolding replica exchange molecular-dynamics simulations of Trp-cage in presence of urea, over a wide range of urea concentration. The increase in preferential interaction upon unfolding is dominated by the side-chain contribution, rather than the backbone. Similar trends are observed in simulations using two different force fields, Amber94 and Amber99sb, for the protein. The magnitudes of the side-chain and backbone contributions differ in the two force fields, despite containing identical protein-solvent interaction terms. The differences arise from the unfolded ensembles sampled, with Amber99sb favoring conformations with larger surface area and lower helical content. These results emphasize the importance of the side-chain interactions with urea in protein denaturation, and highlight the dependence of the computed driving forces on the unfolded ensemble sampled.

## INTRODUCTION

The stability of proteins is marginal, and the conformational equilibrium Folded (F) $\rightleftharpoons$ Unfolded (U) can be modulated by addition of compounds known as cosolvents to the solution (1). Cosolvents such as trimethylamine *n*-oxide, glycerol, and sugars that push the equilibrium toward F are known as "protecting osmolytes" and play a crucial role in maintaining the function of intracellular proteins in extreme environmental conditions (2). On the other hand, urea is a protein denaturant, i.e., it shifts the equilibrium toward the unfolded ensemble.

Urea is a widely used denaturant in protein folding studies (3,4), and its molecular mechanism has been actively debated. The indirect mechanism postulates that the solvation properties of the protein in urea solutions is altered due to changes in the water structure (5–8), and this view has been questioned by many studies (9–12). In the direct mechanism, the denaturing ability of urea is attributed to its favorable interaction with the protein surface (13–18). The direct interaction model raises further questions on the nature and strength of the interactions of urea with the polar/apolar side chains and the peptide backbone, and the role of urea-backbone hydrogen bonding in denaturation.

Recent studies have provided evidence for the direct mechanism wherein urea denatures proteins due to favorable interactions with various protein groups, including the backbone as well as the side chains (15–18). However, there is disagreement on the contribution of the various protein groups to the free energy of unfolding (19–21). We seek to examine the relative contribution of the side chains and the backbone in the process of denaturation by urea using the framework of preferential interaction coefficients, which are computed here using equilibrium folding/unfolding replica exchange molecular dynamics (REMD) simulations.

The preferential interaction coefficient for a protein, $\Gamma$, is defined as (22)

$$\Gamma = -\left(\frac{\partial \mu_2}{\partial \mu_3}\right)_{m_2,T,P} = \left(\frac{\partial m_3}{\partial m_2}\right)_{\mu_3,T,P}, \quad (1)$$

where $\mu$ is the chemical potential, $m$ is the concentration, and the subscripts 1, 2, and 3 indicate water, protein, and the cosolvent, respectively. This quantity is experimentally measured using the technique of dialysis (23) or vapor pressure osmometry (24). It is a measure of the change in the chemical potential of the protein in response to the cosolvent. Equivalently, it is the change in the cosolvent concentration to maintain constant chemical potential when the protein is added to the solution. The latter description has been interpreted, using a two-domain model, as the difference between the cosolute concentration in the local domain of the protein and the bulk solution (25). Using the two-domain model, $\Gamma$ can be expressed as

$$\Gamma = \left\langle N_3^{local} - \left(\frac{N_3^{bulk}}{N_1^{bulk}}\right) N_1^{local} \right\rangle, \quad (2)$$

where $N$ denotes the number of molecules. A positive value of $\Gamma$ implies an accumulation of the cosolvent in the vicinity of the protein, showing a net favorable interaction, whereas negative values imply exclusion of the cosolvent from the local domain. With the above expression, $\Gamma$ can be obtained

*Submitted November 11, 2010, and accepted for publication January 3, 2011.*

*Correspondence: angel@rpi.edu

Editor: Bertrand Garcia-Moreno.



0006-3495/11/03/1526/8 $2.00

doi: 10.1016/j.bpj.2011.01.028

directly, using all-atom MD simulations, as a function of distance from the protein by varying the boundary of the local domain (26,27). The quantity, $\Delta\Gamma = \Gamma^{unfold} - \Gamma^{fold}$ is related to the derivative of the free energy of unfolding with respect to the cosolvent concentration (22).

The quantity $\Gamma$ has been computed from MD simulations for many systems in the literature (28–31), but no calculations of $\Delta\Gamma$ for a protein have been reported yet. To calculate $\Gamma$ and $\Delta\Gamma$ for a protein, we sample the reversible folding/unfolding equilibrium of the Trp-cage miniprotein in aqueous urea, over a broad range of urea concentration, using all-atom, explicit solvent REMD simulations. Trp-cage is a designed 20-residue protein with a nontrivial fold (32) that shows the thermodynamic features observed for globular proteins, such as the temperature dependence of the unfolding free energy and enthalpy (33–35). Its small size and fast folding kinetics make computation of its folding/unfolding equilibrium tractable. We simulate the protein using two different force fields—Amber94 (36) and Amber99sb (37)—with the following motivation. We are interested in obtaining the driving forces in the process of denaturation, and for any property, $X$, the driving force is computed as the change upon unfolding, i.e., $\Delta X = X^{unfold} - X^{fold}$. This implies that any driving force computed depends on the unfolded ensemble sampled in the simulation, because $X^{unfold}$ is obtained as an ensemble average over all conformations populated in the unfolded ensemble—which further depends on the force field employed. The Amber94 force field introduces a helical bias in the unfolded ensemble, and predicts a melting point for Trp-cage higher than the experimental value (34). In a recent study by Day et al. (38), it was shown that the newly developed Amber99sb force field shows an excellent agreement with the experimental thermodynamics of Trp-cage (33), while avoiding the helical bias of Amber94.

In this study, we first explore the differences in the behavior of the system between the two force fields. We then analyze the two datasets using the framework of preferential interaction coefficient, and provide insight into the interaction of urea with various protein moieties such as the backbone and the side chains and their role in the process of denaturation.

## METHODS

### Amber94 simulations

We use REMD simulations (39) to study the unbiased folding/unfolding equilibrium of the Trp-cage miniprotein at different concentrations of urea, using Amber94 force field for the protein (36), TIP3P model for water (40), and the Kirkwood-Buff model for urea (41). We simulated systems at three different concentrations of urea, i.e., 1.9 M, 3.8 M, 5.8 M, 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, amounting to a total sampling of 32.5 $\mu$s. The system details are summarized in Table 1. The last half of the simulations is used for analysis, after allowing for convergence.

**TABLE 1 System details**

| Concentration | Number of urea | Number of water | Replica range | Box length |
|---|---|---|---|---|
| 1.9 M | 105 | 2637 | 280–579 K | 4.50904 nm |
| 3.8 M | 233 | 2637 | 280–557 K | 4.67515 nm |
| 5.8 M | 391 | 2637 | 280–537 K | 4.83007 nm |
| 0 M | 0 | 2637 | 280–590 K | 4.35826 nm |

The Trp-cage sequence was generated in an initially all-PP2 conformation using the LEAP program distributed with Amber 6.0. This structure was simulated in 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, obtained thus, are energy-minimized and equilibrated 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 equilibration runs at all concentrations. The respective structures are used to start the REMD simulations at various concentrations.

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

$$P_{acc} = \min\left[1, \exp\left[\left(\beta_i - \beta_j\right)\left(U\left(\vec{r}_i^{\,N}\right) - U\left(\vec{r}_j^{\,N}\right)\right)\right]\right], \quad (3)$$

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 to ensure sufficient overlap of energy distributions, and to set the acceptance probability to 0.2 (42). REMD simulations were carried out using GROMACS4 (43), with a time step of 2 fs and exchange moves were attempted between all adjacent replicas every 4 ps. The simulations were carried out in the canonical ensemble, using the Nosé-Hoover thermostat with a coupling constant of 0.5 ps (44,45). Solvent constraints were solved using SETTLE (46), and other bond constraints were imposed using LINCS (47). The electrostatic interactions were treated by smooth-particle mesh Ewald summation (48). Appropriate Lennard-Jones long-range correction for energy and pressure were taken into account. The folded and the unfolded ensembles are distinguished using a $C_\alpha$ RMSD cutoff of 0.23 nm from the native structure (PDB code:1L2Y). Analysis based on this data set was reported in our recent work (17).

### Amber99sb systems

We picked a folded configuration at random from the converged part of the Amber94 simulations at 300 K for the 1.9 M and 3.8 M urea systems and changed the protein force-field parameters to Amber99sb (37). These configurations were used to start the REMD simulations, with all the other simulation settings unchanged. Both the systems using Amber99sb force field were simulated for 500 ns per replica, amounting to a total sampling of 50 $\mu$s. We have used the last 200-ns segment of the replicas for the analysis presented here. We did not simulate the 5.8 M urea system in Amber99sb, as we expect the fraction of folded configurations to be too low to compute quantities for the folded ensemble with any statistical significance. The data for the 0 M urea system is obtained from Day et al. (38).

## RESULTS AND DISCUSSION

The folding/unfolding equilibrium of Trp-cage is studied at different urea concentrations, namely, 0 M, 1.9 M, 3.8 M, and 5.8 M, using the Amber94 and Amber99sb force fields

for the protein. The two protein force fields differ only in the backbone dihedral angle potentials, whereas the intermolecular protein-solvent interactions are identical (37). However, this modification leads to significant differences in the stability of the protein as a function of urea concentration, as shown in Fig. 1, *A* and *B*. The protein is much more destabilized by urea in Amber99sb than Amber94, as can be seen by the sharper decrease in the fraction folded, $x_f$, with increasing urea concentration at a given temperature. The fraction folded at a given temperature is related to the free energy of unfolding through $\Delta G_u = -RT \ln [(1 - x_f)/x_f]$. It is experimentally observed that the stability of proteins decreases linearly with increasing urea concentration, i.e., $\Delta G^{Urea} = \Delta G^{Water} - m[C]$, where the slope or the $m$ value measures the response of the protein stability to the addition of urea. This relationship, known as the linear extrapolation model (LEM), is used to estimate the stability of proteins in water (4). The $m$ values obtained from Amber94 and Amber99sb simulations are 0.41 kJ mol$^{-1}$ M$^{-1}$ and 1.81 kJ mol$^{-1}$ M$^{-1}$, respectively, whereas the experimentally measured $m$ value is 1.3 kJ mol$^{-1}$ M$^{-1}$ (35). The $m$ values calculated from the simulations are in the same order of magnitude as the experimental value, and the variation is reasonable for a protein of its size.

The Amber94 and Amber99sb simulations differ significantly in the unfolded ensemble sampled as shown in Fig. 2. From Fig. 2, *A* and *B*, we see that both the force fields show similar behavior for solvent-accessible surface area (SAS) of the unfolded ensemble, as a function of temperature and urea concentration. At any temperature, the unfolded ensembles in various concentrations of urea in either force field can be distinguished based on their average SAS. However, it is seen that the SAS of the 1.9 M system in Amber99sb at 280 K is comparable to the SAS of the 5.8 M system at 500 K. This increase in the SAS of the unfolded ensemble in Amber99sb simulations is due to the decrease in the helical content when compared to Amber94. We define helical content as the average number of residues that occupy the helical basin at a given temperature. The

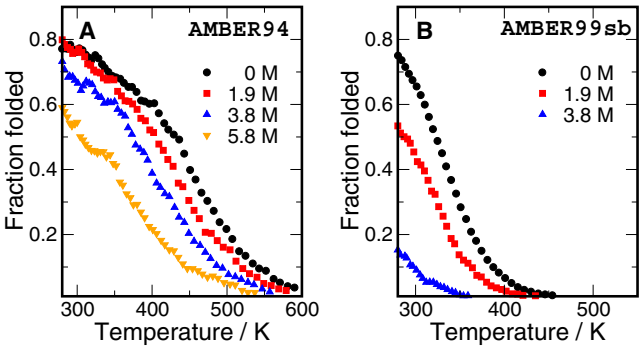


FIGURE 1 Effect of urea on the folding equilibrium. (*A* and *B*) Fraction of folded states as a function of temperature for the various urea concentrations studied, shown for both Amber94 and Amber99sb systems.

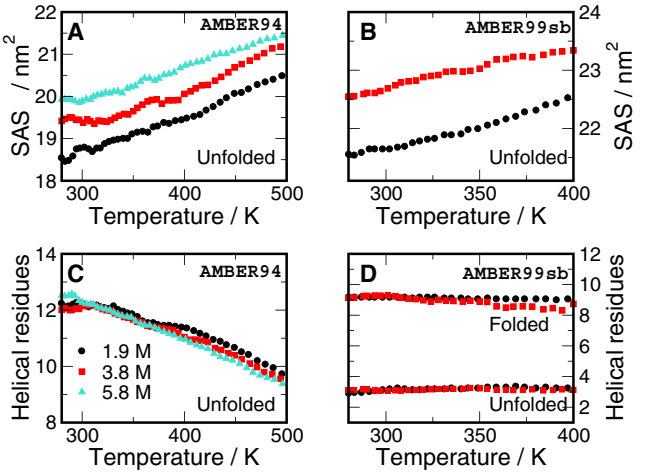


FIGURE 2 Characterization of the unfolded ensembles. (*A* and *B*) Average solvent-accessible surface (SAS) area of the unfolded ensemble as a function of temperature for the various urea concentrations, shown for both Amber94 and Amber99sb systems. (*C* and *D*) Helical content as a function of temperature, shown for Amber94 and Amber99sb systems.

Amber94 force field exhibits a helical bias, and the unfolded ensemble has significant helical content as shown in Fig. 2 *C*. From Fig. 2 *D*, it is seen that the folded and the unfolded ensembles in Amber99sb are well separated by helical content. Therefore, the unfolded ensembles sampled in Amber99sb simulations are populated by conformations that are less helical and have larger SAS than the corresponding Amber94 ensemble.

In this study, we compute the preferential interaction of Trp-cage with urea at 300 K, separately for the folded and the unfolded ensembles, for all the urea concentrations studied. The calculation scheme is illustrated in Fig. 3, using the 3.8 M urea system as an example. Fig. 3, *A* and *B*, shows the proximal radial distribution function (49) for both urea and water around the protein for the two force fields employed. For both the solvents, the proximal radial distribution functions computed for the folded and the unfolded ensembles are indistinguishable. In Fig. 3, *C* and *D*, we compute the preferential interaction for the protein as a function of distance, $\Gamma(r)$, using Eq. 2 and find that the solvent distribution around the folded and the unfolded ensembles can be distinguished with this metric, i.e., by using the information for both the solvents together. In Fig. 3, *E* and *F*, we dissect the $\Gamma(r)$ obtained for the protein (P) into side chains (SC) and backbone (BB) contributions. Out of 158 nonhydrogen atoms in the protein, we classify 84 atoms belonging to the repeating unit of -C-O-N-$C_\alpha$- as the backbone and the rest of the protein atoms as side chains. The decomposition is done by using the proximity criterion (50)—assign a solvent molecule to a particular protein group if it is closest to the group—thereby ensuring additivity i.e., $\Gamma^P = \Gamma^{SC} + \Gamma^{BB}$.

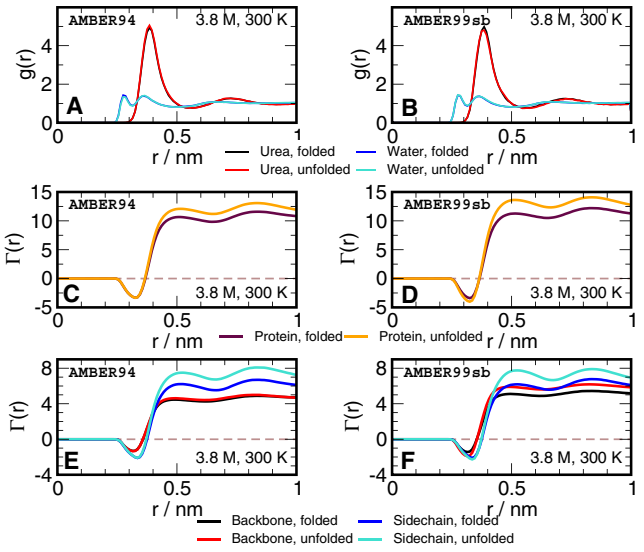


FIGURE 3 Local solvent distribution around the protein for 3.8 M urea system at 300 K. (*A* and *B*) Proximal radial distribution for urea and water around the folded and the unfolded ensembles of the protein, shown for both Amber94 and Amber99sb. (*C* and *D*) Preferential interaction, as a function of distance from the protein, for the folded and the unfolded ensembles. (*E* and *F*) Decomposition of preferential interaction into backbone and side-chain contributions.

Solvent molecules interact with the entire protein as dictated by the governing Hamiltonian, and we use assignment of a solvent molecule to a particular group based on the proximal distance only as a method to account for the local density. Therefore, Γ is more aptly interpreted as the preferential spatial arrangement of the cosolvent around the protein (and its components), which arises due to underlying favorable energetic and/or entropic considerations. The errors in the calculation of $\Gamma(r)$ are determined using the method of block averaging. The converged part of the simulation at 300 K is divided into blocks of lengths 10 ns and 20 ns in Amber94 and Amber99sb, respectively. The folded and unfolded configurations are separated to estimate the average value of $\Gamma^{fold}$ and $\Gamma^{unfold}$ in each block, and the errors are obtained from the variation in the block averages. The error in determining Γ increases with increasing distance from the protein and we use a cutoff of 0.65 nm to define the local domain, beyond which the difference in the Γ between the folded and the unfolded ensembles is not well resolved. Although the Γ-values reported in the subsequent figures correspond to the above cutoff, the conclusions are qualitatively similar for any cutoff larger than 0.5 nm.

In Fig. 4, we compare the contribution of the side chains and the backbone to the preferential interaction of the protein in a given ensemble, either folded or unfolded. Firstly, we note that all the values of Γ are positive, irrespective of side-chain or backbone classification, which shows preferential interaction of both these groups with urea in the folded as well as the unfolded ensembles. Fig. 4, *A* and *B*, shows the data from Amber94 for the folded and the unfolded ensembles, respectively, whereas Fig. 4, *C* and *D*, are from Amber99sb. It can be clearly seen that the preferential interaction of the side chains is larger than that of the backbone in a given ensemble, for all the urea concentrations studied.

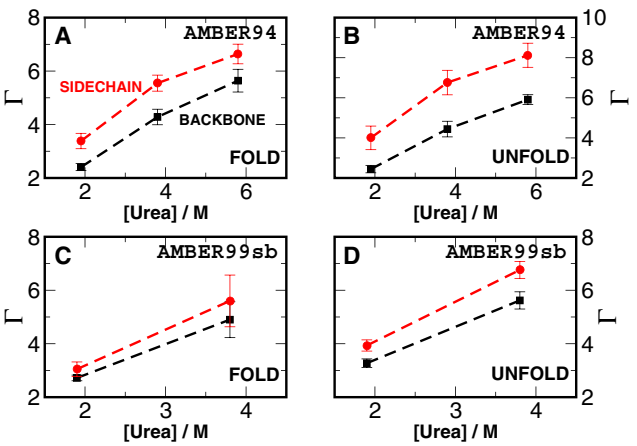


FIGURE 4 Preferential interaction contributions in a given ensemble. (*A*) Folded ensemble in Amber94. (*B*) Unfolded ensemble in Amber94. (*C*) Folded ensemble in Amber99sb. (*D*) Unfolded ensemble in Amber99sb. (*Solid squares* and *circles*) Backbone and side-chain contribution, respectively.

In Fig. 5, *A* and *D*, we compare the preferential interaction for the protein between the folded and the unfolded ensembles from Amber94 and Amber99sb simulations, respectively. As can be anticipated from the previous figure, the preferential interaction for both the folded and the unfolded ensembles is positive. The unfolded ensemble has a larger preferential interaction with urea than the folded, thereby providing a thermodynamic driving force

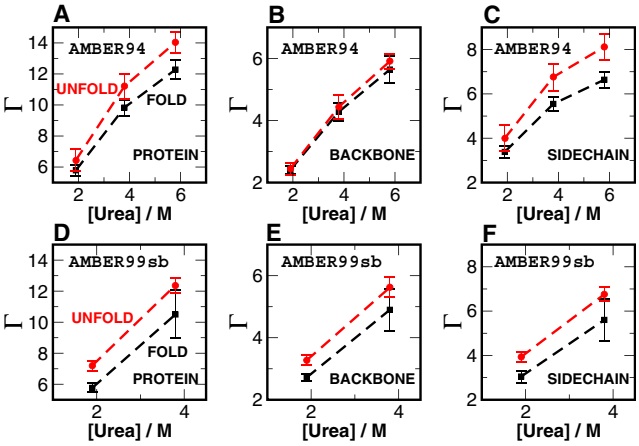


FIGURE 5 Comparing the preferential interaction between folded and unfolded ensembles. (*A*) $\Gamma^{P}$ in Amber94. (*B*) $\Gamma^{BB}$ in Amber94. (*C*) $\Gamma^{SC}$ in Amber94. (*D*) $\Gamma^{P}$ in Amber99sb. (*E*) $\Gamma^{BB}$ in Amber99sb. (*F*) $\Gamma^{SC}$ in Amber99sb. (*Solid squares* and *circles*) Folded and unfolded ensembles, respectively.

for unfolding. We observe $\Delta\Gamma^P$ to increase with urea concentration in both Amber94 and Amber99sb simulations, with the increase being linear in Amber94. We are aware of only one direct experimental measurement of $\Delta\Gamma$ for a protein in urea solution (51), and our data is in qualitative agreement with the results reported there.

An estimate of the $m$ value can be obtained can be obtained from the measurement of $\Delta\Gamma$. Starting from the Wyman linkage relation (22) and assuming the LEM for $\Delta G$, it can shown that

$$\tilde{m} \simeq \frac{RT\Delta\Gamma}{[C]}, \tag{4}$$

where $\tilde{m}$ is the estimate of the $m$ value from $\Delta\Gamma$, and $[C]$ is the urea concentration. The $\Delta\Gamma$-values in Amber99sb are larger than Amber94, implying a larger $\tilde{m}$ for Amber99sb simulations. By averaging over all urea concentrations studied in a given force field, we estimate $\tilde{m}$ to be 0.83 kJ mol$^{-1}$ M$^{-1}$ and 1.56 kJ mol$^{-1}$ M$^{-1}$ for Amber94 and Amber99sb, respectively. The agreement of $\tilde{m}$ with the $m$ values obtained from the LEM analysis of data in Fig. 1 is quite good, considering the fact that these are two different measurements.

We now examine the contributions to $\Delta\Gamma^P$ from the backbone and the side-chain groups. Fig. 5, *B* and *C*, indicates that the increase in $\Gamma^P$ upon unfolding in Amber94 simulations can be entirely accounted by increase in $\Gamma^{SC}$ i.e., the contribution of the backbone is negligible. From Fig. 5, *E* and *F*, we see that the increase in $\Gamma^P$ upon unfolding in Amber99sb simulations has nonzero contributions from both the backbone and the side chains. The contribution of the side chains is estimated to be ~60% for both 1.9 M and 3.8 M urea systems in Amber99sb. The differences in the backbone contribution for the two force fields arises due to the differences in helical content present in the unfolded ensemble. The unfolded ensemble in Amber99sb has very little helical content, and this leads to a greater exposure of the backbone and consequently a larger contribution from the backbone to $\Delta\Gamma^P$.

The preceding analysis shows that not only is $\Gamma^{SC} > \Gamma^{BB}$ in a given ensemble, the difference upon unfolding, $\Delta\Gamma^P$, is dominated by $\Delta\Gamma^{SC}$ in both the force fields studied. This is despite fewer atoms in the protein being classified as side chain than backbone. Both the force fields are consistent with experimental thermodynamics, i.e., linearity of unfolding free energy with urea concentration. The variation of the $m$ value in the two force fields correlates well with change in solvent-exposed surface upon unfolding (52). We expect the data from the Amber99sb simulations to provide a better estimate for the backbone and the side-chain contributions. Even if not realistic in the unfolded ensemble sampled, the Amber94 scenario is remarkable—it shows that the preferential interaction of side chains with urea alone is sufficient to capture the experimental trend.

We now study how the differences in backbone and side-chain contributions to $\Delta\Gamma^P$ manifest in the direct interaction and the hydrogen-bonding data. We calculate the nonbonded interaction energy, $E^{PU}$, of the protein with all urea molecules in its first solvation shell, i.e., within 0.5 nm from the protein surface. We then calculate the difference upon unfolding, $\Delta E^{PU}$, and examine the relative contribution of the Coulomb and Lennard-Jones (LJ) interactions. Negative values of $\Delta E$ indicate a favorable driving force toward unfolding. The results obtained are weakly dependent on the distance cutoff used to define the first solvation shell. The calculations, obtained by averaging over the temperature range 280–310 K, are shown in Fig. 6, *A* and *B*. In Amber94, we find that the contribution of LJ is larger than the Coulomb. On the other hand, the Coulomb interaction dominates in the Amber99sb simulations. The scenario in Amber99sb is qualitatively similar to the data reported for Amber94 for the temperature range 410–460 K (17), but the effect is stronger. We find more favorable Coulomb interactions to correlate well with reduced helical content for urea as well as water (data not shown). We also note that the change in the direct interaction, $\Delta E^{PU}$, is larger in Amber99sb when compared to Amber94.

The effect of decreased helical content is also reflected in the hydrogen-bonding data. Using a acceptor-hydrogen distance of 0.26 nm and donor-hydrogen-acceptor angle $> 90°$ as the existence criterion for a hydrogen bond (53), we calculate the hydrogen-bond statistics between the protein backbone and the solvent, separately for the folded and the unfolded ensembles and plot the difference upon unfolding, $\Delta N_u$, for the two force fields in Fig. 6, *C* and *D*. We can see a significant increase in $\Delta N_u$ for both urea and water in Amber99sb simulations, when compared

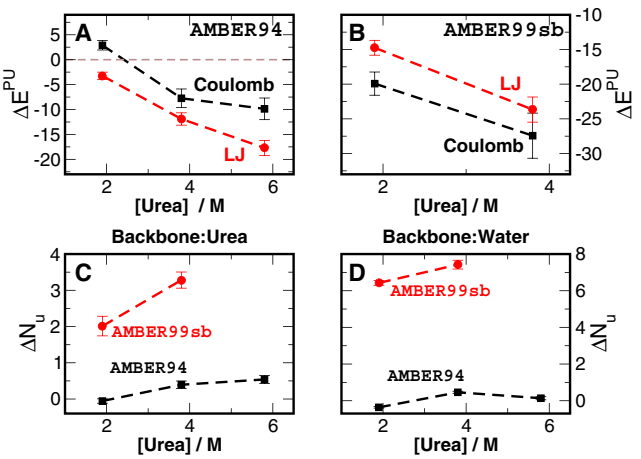


FIGURE 6 Comparison of the direct interaction and hydrogen bonding. Coulomb (*squares*) and LJ (*circles*) contributions to $\Delta E^{PU}$ are shown (*A*) forAmber94 and (*B*) for Amber99sb. Comparison of Amber94 (*squares*) and Amber99sb (*circles*) is shown (*C*) for backbone-urea hydrogen bonding and (*D*) for backbone-water hydrogen bonding.

to Amber94, and this can be ascribed to the greater exposure of the backbone due to reduced helical content in the unfolded ensemble sampled in Amber99sb. These results show that the driving forces computed from the two force fields exhibit different behavior entirely due to the differences in the unfolded ensembles sampled, without requiring any change in the protein-solvent interaction parameters.

## CONCLUSIONS

It is widely accepted that urea denatures proteins through favorable direct interaction with the protein (12). Within the direct interaction model, there exists debate on the interaction of the side chains and the backbone with urea, and their contribution in the process of denaturation. One can deduce the propensity of various protein groups to interact with urea by studying the interaction of model compounds—amino acids, side-chain/backbone analogs—with urea. Such studies have been carried out, in experiment as well as simulation (15,18,54), and these show that most of the amino acids have favorable interaction with urea. However, model compounds do not undergo a folding transition and the interaction measured in such experiments can be considered bare, i.e., not coupled to the context provided by the protein sequence and conformation. The contributions of the various protein groups to the free energy of unfolding cannot be directly determined without using further assumptions or models.

One of the popular models to analyze the experimental data is Tanford's transfer model, where folding/unfolding of the protein and the process of transfer of folded and unfolded states from water to aqueous urea solution are tied together in a thermodynamic cycle (19). In the transfer model, the free energy of unfolding of a protein in 1 M urea solution ($\Delta G^{Urea}{}_{1M}$) is related to the free energy of unfolding in water ($\Delta G^{Water}$) through

$$\Delta G^{Urea}_{1\ \mathrm{M}} = \Delta G^{Water} + \sum_i n_i \alpha_i \delta g^{tr}_i. \quad (5)$$

Here $\alpha_i$ is the fractional change in solvent-exposed surface area of group $i$ upon unfolding, $\delta g^{tr}{}_i$ is the experimentally measured free energy of transfer of group $i$ from water to 1 M urea solution, and $n_i$ is the number of groups of type $i$ present in the protein (54). The underlying assumption of the transfer model is that the interaction of various groups with urea contribute in an additive manner in unfolding. Calculation of $\alpha_i$ requires a knowledge of the exposed surface area of the residue $i$ in the unfolded ensemble, which is obtained by using polymer models for the unfolded ensemble (19). The contribution of each side-chain group is obtained by subtracting the transfer free energy of glycine from the transfer free energy of the corresponding amino acid. Invoking the linear extrapolation model, it can be seen that

$$\sum_i n_i \alpha_i \delta g^{tr}_i = -m$$

and correct prediction of the $m$ values is seen as the validity of the model. Based on this approach, Auton and Bolen (19) and Auton et al. (20) have proposed that the dominant contribution to the free energy of unfolding in urea denaturation comes from the peptide backbone, and the total contribution of the side chains may even turn out to be unfavorable.

By studying folding equilibrium of Trp-cage as a function of urea concentration using REMD simulations, we have shown that both the force fields employed—Amber94 and Amber99sb—are consistent with experimental thermodynamics of urea denaturation, i.e., the calculated $m$ values are in reasonable agreement with the experimental value. The preferential interaction for the protein, $\Gamma^P$ and $\Delta\Gamma^P$, was calculated from our simulation data and it was shown that the behavior of these quantities are in qualitative agreement with the experimental expectation. The decomposition of $\Gamma^P$ into $\Gamma^{BB}$ and $\Gamma^{SC}$ using the proximity criterion is additive in a trivial manner, because $\Gamma(r)$ for the protein is itself calculated using the proximal distance. The assumption in calculation of $\Gamma^X$ is that the solvent molecule belonging to group $X$ (in the proximal sense) has preferential or strongest interaction with $X$, given the context of protein sequence and conformation. Preferential interaction for a protein is not a measure of bare interaction, either in the folded or the unfolded ensemble, and there is no explicit separation of geometry and chemistry in these calculations. The underlying thermodynamics of the system dictates the protein and the solvent configurations sampled, and we stress that $\Gamma$ is a combined measure of the protein configuration and the solvent distribution around it. Though the assignment of solvent molecules to a particular group can be done for distances up to half the simulation box length, the interpretation in terms of preferential interaction is valid only for reasonably short distances ($<1$ nm) from the protein.

Our analysis, based on simulations with two protein force fields (Amber94 and Amber99sb) provides a consistent picture for the behavior of $\Gamma^P$, $\Delta\Gamma^P$, $\Gamma^{BB}$, and $\Gamma^{SC}$. Urea has favorable interaction with the protein backbone and the side chains, in both the folded and the unfolded ensembles. The increase in preferential interaction upon unfolding, $\Delta\Gamma^P$, provides the thermodynamic driving force for unfolding. Though the two force fields differ in terms of the relative contribution from the backbone and the side chains to $\Delta\Gamma^P$, the side-chain contribution is found to be larger than the backbone in simulations using both the force fields. We expect the Amber99sb data to provide a better estimate for the individual contributions, i.e., ~60% contribution from the side chains. The differences between the two force fields are ascribed to the unfolded ensembles sampled and the microscopic manifestation of these differences were highlighted in Fig. 6. Our study shows the

importance of obtaining a good description of the unfolded ensemble to evaluate the driving forces for unfolding. In contrast to the results obtained from the transfer model studies (19,20), the side-chain contribution to the preferential interaction, $\Delta\Gamma^{SC}$, and the LJ contribution to the protein-urea interaction, $\Delta E^{PU}$, taken together for both the force fields emphasize the importance of the interaction of urea with the side chains in the process of denaturation. Our study strongly supports the direct mechanism, in which urea has favorable interactions with the side chains as well as the protein backbone.

This work was funded by the National Science Foundation (MCB-1050966 and MCB-0543769).

## REFERENCES


1. Timasheff, S. N. 1993. The control of protein stability and association by weak interactions with water: how do solvents affect these processes? *Annu. Rev. Biophys. Biomol. Struct.* 22:67–97.
2. Yancey, P. H., M. E. Clark, …, G. N. Somero. 1982. Living with water stress: evolution of osmolyte systems. *Science.* 217:1214–1222.
3. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.
4. Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131:266–280.
5. Frank, H. S., and F. Franks. 1968. Structural approach to solvent power of water for hydrocarbons—urea as a structure breaker. *J. Chem. Phys.* 48:4746–4757.
6. Das, A., and C. Mukhopadhyay. 2009. Urea-mediated protein denaturation: a consensus view. *J. Phys. Chem. B.* 113:12816–12824.
7. Wei, H. Y., Y. B. Fan, and Y. Q. Gao. 2010. Effects of urea, tetramethyl urea, and trimethylamine N-oxide on aqueous solution structure and solvation of protein backbones: a molecular dynamics simulation study. *J. Phys. Chem. B.* 114:557–568.
8. Wei, H. Y., L. J. Yang, and Y. Q. Gao. 2010. Mutation of charged residues to neutral ones accelerates urea denaturation of HP-35. *J. Phys. Chem. B.* 114:11820–11826.
9. Soper, A. K., E. W. Castner, and A. Luzar. 2003. Impact of urea on water structure: a clue to its properties as a denaturant? *Biophys. Chem.* 105:649–666.
10. Rezus, Y. L. A., and H. J. Bakker. 2006. Effect of urea on the structural dynamics of water. *Proc. Natl. Acad. Sci. USA.* 103:18417–18420.
11. Kokubo, H., and B. M. Pettitt. 2007. Preferential solvation in urea solutions at different concentrations: properties from simulation studies. *J. Phys. Chem. B.* 111:5233–5242.
12. Rossky, P. J. 2008. Protein denaturation by urea: slash and bond. *Proc. Natl. Acad. Sci. USA.* 105:16825–16826.
13. Robinson, D. R., and W. P. Jencks. 1965. Effect of compounds of urea-guanidinium class on activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J. Am. Chem. Soc.* 87:2462–2470.
14. Makhatadze, G. I., and P. L. Privalov. 1992. Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* 226:491–505.
15. Stumpe, M. C., and H. Grubmüller. 2007. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *J. Am. Chem. Soc.* 129:16126–16131.
16. Hua, L., R. H. Zhou, …, B. J. Berne. 2008. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. USA.* 105:16928–16933.
17. Canchi, D. R., D. Paschek, and A. E. García. 2010. Equilibrium study of protein denaturation by urea. *J. Am. Chem. Soc.* 132:2338–2344.
18. Lee, S., Y. L. Shek, and T. V. Chalikian. 2010. Urea interactions with protein groups: a volumetric study. *Biopolymers.* 93:866–879.
19. Auton, M., and D. W. Bolen. 2005. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. USA.* 102:15065–15068.
20. Auton, M., L. M. F. Holthauzen, and D. W. Bolen. 2007. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. USA.* 104:15317–15322.
21. Cannon, J. G., C. F. Anderson, and M. T. Record, Jr. 2007. Urea-amide preferential interactions in water: quantitative comparison of model compound data with biopolymer results using water accessible surface areas. *J. Phys. Chem. B.* 111:9675–9685.
22. Timasheff, S. N. 2002. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc. Natl. Acad. Sci. USA.* 99:9721–9726.
23. Arakawa, T., and S. N. Timasheff. 1982. Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry.* 21:6545–6552.
24. Courtenay, E. S., M. W. Capp, …, M. T. Record, Jr. 2000. Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro. *Biochemistry.* 39:4455–4471.
25. Record, Jr., M. T., and C. F. Anderson. 1995. Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a two-domain model. *Biophys. J.* 68:786–794.
26. Baynes, B. M., and B. L. Trout. 2003. Proteins in mixed solvents: a molecular-level perspective. *J. Phys. Chem. B.* 107:14058–14067.
27. Shukla, D., C. Shinde, and B. L. Trout. 2009. Molecular computations of preferential interaction coefficients of proteins. *J. Phys. Chem. B.* 113:12546–12554.
28. Athawale, M. V., J. S. Dordick, and S. Garde. 2005. Osmolyte trimethylamine-N-oxide does not affect the strength of hydrophobic interactions: origin of osmolyte compatibility. *Biophys. J.* 89:858–866.
29. Hu, C. Y., G. C. Lynch, …, B. M. Pettitt. 2010. Trimethylamine N-oxide influence on the backbone of proteins: an oligoglycine model. *Proteins.* 78:695–704.
30. Ma, L., L. Pegram, …, Q. Cui. 2010. Preferential interactions between small solutes and the protein backbone: a computational analysis. *Biochemistry.* 49:1954–1962.
31. Shukla, D., and B. L. Trout. 2010. Interaction of arginine with proteins and the mechanism by which it inhibits aggregation. *J. Phys. Chem. B.* 114:13426–13438.
32. Neidigh, J. W., R. M. Fesinmeyer, and N. H. Andersen. 2002. Designing a 20-residue protein. *Nat. Struct. Biol.* 9:425–430.
33. Streicher, W. W., and G. I. Makhatadze. 2007. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. *Biochemistry.* 46:2876–2880.
34. Paschek, D., S. Hempel, and A. E. García. 2008. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. USA.* 105:17754–17759.
35. Wafer, L. N. R., W. W. Streicher, and G. I. Makhatadze. 2010. Thermodynamics of the Trp-cage miniprotein unfolding in urea. *Proteins.* 78:1376–1381.
36. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* 117:5179–5197.
37. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
38. Day, R., D. Paschek, and A. E. Garcia. 2010. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins.* 78:1889–1899.
39. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

40. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

41. Weerasinghe, S., and P. E. Smith. 2003. A Kirkwood-Buff derived force field for mixtures of urea and water. *J. Phys. Chem. B.* 107:3891–3898.

42. Garcia, A., H. Herce, and D. Paschek. 2006. Simulations of temperature and pressure unfolding of peptides and proteins with replica exchange molecular dynamics. *Annu. Rep. Comput. Chem.* 2:83–96.

43. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

44. Nose, S. 1984. A molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268.

45. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

46. Miyamoto, S., and P. A. Kollman. 1992. SETTLE—an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.

47. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

48. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

49. Ashbaugh, H. S., and M. E. Paulaitis. 1996. Entropy of hydrophobic hydration: extension to hydrophobic chains. *J. Phys. Chem.* 100:1900–1913.

50. Mehrotra, P. K., and D. L. Beveridge. 1980. Structural-analysis of molecular solutions based on quasi-component distribution-functions—application to $[H_2CO]_{aq}$ at 25°C. *J. Am. Chem. Soc.* 102:4287–4294.

51. Timasheff, S. N., and G. Xie. 2003. Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys. Chem.* 105:421–448.

52. Myers, J. K., C. N. Pace, and J. M. Scholtz. 1995. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* 4:2138–2148.

53. Gnanakaran, S., R. M. Hochstrasser, and A. E. García. 2004. Nature of structural inhomogeneities on folding a helix and their influence on spectral measurements. *Proc. Natl. Acad. Sci. USA.* 101:9229–9234.

54. Wang, A., and D. W. Bolen. 1997. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry.* 36:9101–9108.