

Published in final edited form as:

*Biochemistry*. 2010 March 9; 49(9): 1954–1962. doi:10.1021/bi9020082.

## Preferential interactions between small solutes and the protein backbone: A computational analysis

Liang Ma<sup>1</sup>, Laurel Pegram<sup>2</sup>, M. T. Record Jr.<sup>1,2</sup>, and Qiang Cui<sup>\*1</sup>

<sup>1</sup>Graduate Program in Biophysics and Department of Chemistry University of Wisconsin, Madison 1101 University Avenue, Madison, WI 53706

<sup>2</sup>Department of Biochemistry University of Wisconsin, Madison 433 Babcock Drive, Madison, WI 53706

### Abstract

To help better understand the effects of small solutes on protein stability, we carry out atomistic simulations to quantitatively characterize the interactions between two broadly used small solutes, urea and glycine betaine (GB), with a tri-glycine peptide, which is a good model for protein backbone. Multiple solute concentrations are analyzed and each solute-peptide-water ternary system is studied with ~200–300 ns of molecular dynamics simulations with the CHARMM force field. The comparison between calculated preferential interaction coefficients ( $\Gamma_{23}$ ) and experimentally measured values suggests that semi-quantitative agreement with experiments can be obtained if care is exercised to balance interactions among the solute, protein and water. On the other hand, qualitatively incorrect (i.e., wrong sign in  $\Gamma_{23}$ ) result can be obtained if a solute model is constructed by directly taking parameters for chemically similar groups from existing force field. Such sensitivity suggests that small solute thermodynamic data can be valuable in the development of accurate force field models of biomolecules. Further decomposition of  $\Gamma_{23}$  into group contributions leads to additional insights regarding the effects of small solutes on protein stability. For example, use of the CHARMM force field predicts that urea preferentially interacts with not only amide groups in the peptide backbone but also aliphatic groups, suggesting a role for these interactions in urea-induced protein denaturation; quantitatively, however, it is likely that the CHARMM force field overestimates the interaction between urea and aliphatic groups. The results on GB support a simple thermodynamic model that assumes additivity of preferential interaction between GB and various biomolecular surfaces.

Biomolecules exist in a heterogeneous cellular environment and it is well known that small solutes can play important roles in regulating the stability, structure and therefore function of biomolecules [1–3]. In the context of protein stability, small solutes can be classified into denaturants and protectants according to their effect of destabilizing or stabilizing the native state of proteins [3, 4]. Urea and guanidinium are two well-known protein denaturants [5, 6] and widely used in protein denaturation experiments *in vitro*; glycine betaine, which is rich in mammalian kidney and some vascular plants [7–9], and trimethylamine N-oxide (TMAO), which is accumulated in cartilaginous fish and the coelacanth [9–11], act as protein protectants. Although much is known about the interaction between small solutes and biomolecules [2, 12], the precise denaturing/protecting mechanisms of these small solutes at the molecular level remain controversial. For example, multiple proposals have been put forward regarding how urea denatures proteins, and they range from modifying

\*To whom the correspondence should be addressed: cui@chem.wisc.edu. Phone: 608-262-9801; Fax: 608-262-9918.

**Supporting Information Available** Detailed results (charges and interaction energies) are available for the development of GB parameters.

water structure [13–16] to exhibiting strong interactions with protein backbone through either polar [17–19] or dispersive forces [17,20–22].

One important avenue for probing the mechanism of small solute effects is to carry out systematic thermodynamic analysis. A key quantity in this context is the “ $m$ -value”, defined as: [2]

$$m - \text{value} = \frac{\partial \Delta G_{obs}^o}{\partial m_3}, \quad (1)$$

in which  $\Delta G_{obs}^o$  is the standard free energy change for the unfolding process, and  $m_3$  is the solute concentration on the molar scale; throughout the paper, we use the following conventions to label the components in a ternary system: component 1 is water, component 2 is the protein and component 3 is the small solute. The  $m$ -value is closely related to the difference of interactions between the solute with the unfolded and native states. For example, if the solute interacts more strongly with the unfolded state than with the native state, increasing the solute concentration will shift the equilibrium towards the unfolded state, leading to protein denaturation.

Alternatively, the  $m$ -value is also be given as  $\Delta\mu_{23}$ , the change of  $\mu_{23}$  for the unfolding process, where

$$\mu_{23} = \frac{\partial \mu_2}{\partial m_3} |_{m_2, T, P} \quad (2)$$

is the derivative of the chemical potential of component 2,  $\mu_2$ , with respect to  $m_3$ , at constant protein concentration on the molar scale ( $m_2$ ), constant temperature and pressure.  $\mu_{23}$  refers to change of the protein's chemical potential in response to change of the solute concentration; i.e., it quantifies the stabilizing/distabilizing effect of the solute on a specific (native or unfolded) state of the protein.

Another thermodynamic quantity closely related to  $\mu_{23}$  is the preferential interaction coefficient [2],  $\Gamma_{23}$ , between the solute and the protein:

$$\Gamma_{23} = -\frac{\mu_{23}}{\mu_{33}}, \quad (3)$$

in which  $\mu_{33}$  is the counterpart of  $\mu_{23}$  in the solute-water binary system. The physical meaning of  $\Gamma_{23}$  can be understood based on the two-domain model [23,24] (see Ref. [25] and references therein for discussions of  $\Gamma_{23}$  in different ensembles). Due to the presence of the protein (component 2), the distributions of water (component 1) and the solute (component 3) in the local domain II (close to the protein) are different from those in the bulk domain I (far from the protein).  $\Gamma_{23}$  can be cast into the following form [23]:

$$\Gamma_{23} = \left\langle n_3^{\text{II}} - n_1^{\text{II}} \frac{n_3^{\text{I}}}{n_1^{\text{I}}} \right\rangle, \quad (4)$$

which clearly characterizes the extent of accumulation or exclusion of the solute from the protein surface relative to the bulk. In Eq. 4, the brackets indicate ensemble average;  $n_3^{\text{II}}$  and  $n_1^{\text{II}}$  are the numbers of solute and water molecules in the local domain II, while  $n_3^{\text{I}}$  and  $n_1^{\text{I}}$  are the corresponding values in the bulk domain I.  $\Gamma_{23}$  is positive when the solute accumulates around the protein relative to the bulk and negative when the solute is excluded from the protein. Therefore, depending on the degree of preferential association to the native vs.

unfolded states, a solute may stabilize or destabilize the native state of a protein. By systematically probing how  $\Gamma_{23}$  or  $\mu_{23}$  depends on the types (e.g., polar/non-polar/anionic) of protein surfaces, rather quantitative relations can be established regarding the effect of small solutes on the stability of proteins [25–27].

In principle, atomistic simulations are ideally suited for in-depth analysis of small solute effects and a significant number of simulation studies have indeed been carried out over the years [16, 28–31]. In practice, however, the reliability of atomistic simulations using existing force fields is not straightforward to evaluate considering the relatively subtle nature of solute effects. Therefore, it is important to explicitly benchmark atomistic simulations using accurate experimental data for model systems. For urea, for example, careful studies have been carried out by Smith and co-workers [30], who have shown that popular force fields (e.g., OPLS) give qualitatively correct but quantitatively inaccurate Kirkwood-Bu integrals [31, 32], which are closely related to  $\Gamma_{23}$ . Using the Kirkwood-Bu integrals as guidance, Smith and co-workers have developed atomic force fields for a few small molecules [33–37] such as urea [33] and peptide backbone analogue (N-methyl acetamide) [37]; these parameters have been shown to be transferable to protein-urea-water ternary systems.

In this work, we carry out simulation study on the preferential interactions of urea and glycine betaine (GB) with tri-glycine (TGLY) peptide, which is a good model for protein backbone. There are several motivations for this study. First, Record *et al.* [25, 27] have recently carried out highly accurate measurements of  $\mu_{23}$  and  $\Gamma_{23}$  between urea/GB and small peptides (e.g. di-glycine, tri-glycine) using vapor pressure osmometry techniques; these data have provided an unprecedented opportunity to test current force fields regarding intermolecular interactions. Second, although urea/protein interactions have been extensively studied using molecular simulations, no thorough simulation studies have been carried out so far for GB. Finally, we note that the small solute thermodynamic data directly reflect the interaction between functional groups in proteins *relative* to their interactions with solvent, the underlying driving force behind the assembly of amino acids in proteins. The molecular (as opposed to atomic) nature of these data makes them particularly suited for guiding the development of a protein model at the coarse-grained level, which is an active area of research. Therefore, another goal of this study is to explore to what degree the computed  $\Gamma_{23}$  value and its approximate group decomposition are sensitive to parameters in the force field, which will determine the feasibility of using  $\Gamma_{23}$  for coarse-grained model development.

The computational details are summarized in Sect.1, the results are presented and discussed in Sect.2. Finally, we draw a few conclusions in Sect.3.

## 1 Computational Methods

The basic computational methods are similar to those reported in the literature [29]. We pay particular attention to the amount of sampling for properly converging calculated  $\Gamma_{23}$ . Moreover, we find that although using CHARMM27 parameters [38] for urea leads to the correct sign in  $\Gamma_{23}$  and the dependence of  $\Gamma_{23}$  on urea concentration, the computed value for GB around TGLY has the wrong sign if GB is described by simply taking force field parameters from CHARMM27 for functional groups. A simple procedure is adopted to adjust the force field parameters, which we also briefly describe below.

### 1.1 Molecular dynamics (MD) simulations

As summarized in Table 1, multiple solute concentrations have been studied for each small solute to better compare with experiments. For urea-TGLY, three urea concentrations are

studied: 1.21m, 2.26m and 3.61m. For GB-TGLY, two GB concentrations are studied: 1.35m and 2.04m. To minimize the size of the simulation system, only one copy of TGLY is included with the proper number of urea/GB and water molecules. The initial structure of TGLY is constructed as the ideal geometry in CHARMM 27 force field [38] using the IC BUILD module in CHARMM. [39] The peptide is first solvated with a pre-equilibrated rhombic dodecahedron (RHDO) solvent box with a box length of  $\sim 35 \text{ \AA}$ . The proper number of solutes (urea or GB) are then randomly inserted into the simulation box according to the solute concentration; water molecules that overlap with the solutes are removed. For representative snapshots, see Fig.1.

TGLY and urea are described using the CHARMM 27 force field [38] and the modified TIP3P model [40] is adopted for water molecules; there are no existing parameters for GB in the CHARMM 27 force field, thus a simple procedure is used to develop a model (see Sect. 1.4). A switching scheme [41] for inter-atomic distances between 10 and 12  $\text{\AA}$  is used for van der Waals interactions. For electrostatic interactions, the particle mesh Ewald scheme (PME) [42] is used. All simulations are carried out under the NPT ensemble: temperature is controlled using the Nose-Hoover algorithm [43,44], with a mass of  $250 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{ps}^2$  for the thermostat and a reference temperature of 300K; pressure is controlled with the Andersen algorithm [45,46], using a mass of 500 amu for the pressure piston, a reference pressure of 1 atm, a Langevin piston collision frequency of  $10 \text{ ps}^{-1}$ , and a Langevin piston bath temperature of 300K.

To ensure convergence in the calculated  $\Gamma_{23}$  values, for each ternary system,  $\sim 100$  independent trajectories of 2–3 ns long are carried out. These independent trajectories start with the same equilibrated structure ( $\sim 5 \text{ ns}$ ) but with different initial velocities at 300K. Test calculations have shown that this is an effective scheme (in part because of the highly parallel nature) for converging  $\Gamma_{23}$ . Throughout the simulations, the SHAKE algorithm [47] is used to allow an integration time step of 2 fs. All simulations have been carried out using the CHARMM [39] program.

To supplement the TGLY-urea-water simulations and better understand the preferential interaction between urea and aliphatic ( $-\text{CH}_2-$ ) groups, a set of propane-urea-water simulations are also performed at  $[\text{urea}] = 2.26\text{m}$ . Specifically, 40 independent trajectories of 2ns long are carried out following the same MD simulation protocol used for the TGLY-urea-water systems; the propane is described with the standard CHARMM27 [38] force field.

## 1.2 Calculations of solute/water radial distribution function $g(r)$ with respect to peptide surface

To characterize the distributions of small solutes and water around the peptide, the radial distribution functions,  $g(r)$ , of solute/water relative to the peptide surface are calculated.  $g(r)$  are computed numerically in a similar way to our recent studies of ion distributions around proteins [48,49]. The shortest distances  $r$  between the center of mass (COM) of each solute or water molecule to the van der Waals (vdW) surface of the peptide are collected from MD trajectories and binned to generate  $\delta N(r)$  for each bin ( $r, dr$ ). The volume factor ( $\delta V(r)$ ) corresponding to the bin ( $r, dr$ ) is estimated using the volume function in CHARMM (COOR VOLU) with all peptide atoms having a radius of  $r_{vdW} + r \pm dr/2$  and  $\delta V(r)$  is calculated as:  $\delta V(r) = V(r_{vdW} + r + dr/2) - V(r_{vdW} + r - dr/2)$ ;  $r_{vdW}$  is the vdW radius of each atom. The bin size  $dr$  is set to be  $0.1 \text{ \AA}$  for all  $g(r)$  calculations. The solute/water-peptide surface radial distribution function  $g(r)$  is then calculated as:

$$g(r) = \frac{1}{\rho_{bulk}} \frac{\delta N(r)}{\delta V(r)} \frac{1}{N_{frame}} \quad (5)$$

where  $N_{frame}$  is the number of simulation snapshots used for calculations;  $\rho_{bulk}$  is the bulk number density of solute or water and is estimated by dividing the number of solute or water molecules by the average volume of the simulation box throughout all MD trajectories.

### 1.3 Calculations of $\Gamma_{23}$

$\Gamma_{23}$  for each ternary system is calculated based on a two-domain model [23, 24] according to Eq.4. MD simulations generate an ensemble of snapshots and for each snapshot, the number of solute/water molecules in the local and bulk domains are counted based on the shortest distance ( $r$ ) from the COM of each solute/water molecule to the peptide vdW surface; i.e.,  $r$  is the same as that in the calculations of  $g(r)$ . To define the local and bulk domains, a boundary  $r^*$  has to be specified, which is determined based on the radial distribution functions  $g(r)$ . As shown in Figs.2 and 5, both solute and water reach their bulk concentrations at around 7–8 Å, thus  $r^* = 8$  Å is chosen as the boundary radius. Note that the calculated  $\Gamma_{23}$  is not sensitive to the precise value of  $r^*$ ; this relative insensitivity of computed  $\Gamma_{23}$  to the precise value of  $r^*$  (as far as it is larger than ~4 Å) was also found in a recent computational study by Trout and co-workers [50].

### 1.4 Force field parameters for GB

Since there is no existing parameters for GB in CHARMM27 force field [38], we start building a model for GB by selecting parameters for similar chemical groups. Specifically, the head group of the lipid molecule POPC (Palmitoyl oleoyl phosphatidyl choline) is used to describe the -N(Me)<sub>3</sub> group, and the CTER patch is used to model the carboxylate group. This simple scheme is adopted with an attempt to maximize the compatibility between the GB model and other components (TGLY and water) in the system. Test calculations indicate, however, that this “raw” model leads to qualitatively incorrect  $\Gamma_{23}$  values. At the GB concentration of 1.35m, for example, the calculated value is +0.40, which differs in sign from the experimental value of -0.16 [27]; i.e., this simple model leads to GB accumulation near TGLY rather than exclusion as quantified by the experimental data.

As a simple way to improve the model (so as to maintain the compatibility with the CHARMM force field in general), the partial charges on GB are scaled down by a uniform factor to achieve a proper balance of interactions between GB, TGLY and water. As shown in the Supporting Information, we consider six gas phase molecular complexes formed between GB, TGLY and water in different relative orientations. The internal geometry of the monomers is held fixed to that optimized at the HF/6–31G(d) level [51] in the gas phase using Gaussian03 [52]; this particular level of theory is chosen because it is consistent with the parameterization procedure of the CHARMM27 force field. With the relative orientation also held fixed, the distance between monomers in each complex is then varied to determine the optimal separation at the HF/6–31G(d) level in the gas phase. The interaction energy between the monomers in each complex at the optimal separation is then multiplied by a factor of 1.16 to account for the lack of electronic polarization in the CHARMM27 force field; interaction energies for the six molecular complexes are used to determine a single scaling factor for the raw GB partial charges. As shown in the Supporting Information, the scaled partial charges for functional groups are in fact rather close to the Natural Bonding Orbital charges in GB.

## 2 Results and Discussions

In the following, we first present results for urea/TGLY, then results for GB/TGLY interactions. Finally, we discuss these data in the context of relevant computational and experimental studies in the literature.

### 2.1 Preferential interaction between urea and TGLY

**2.1.1  $\Gamma_{23}$  between urea and TGLY**—As shown in Fig.2, the radial distribution functions  $g(r)$  for urea and water around TGLY vary only slightly among three urea concentrations. The peak height of  $g(r)$  is similar for urea and water, although the width of the first peak is substantially broader for urea. Qualitatively, these results highlight a higher degree of local accumulation of urea around TGLY (relative to bulk) than water.

The convergence behaviors of computed  $\Gamma_{23}$  between urea and TGLY for the three urea concentrations are shown in Fig. 3. The red dashed lines represent the converged  $\Gamma_{23}$  plus/minus the estimated standard deviation using block average [53]. Evidently, it takes  $\sim 100$ ns for  $\Gamma_{23}$  to converge, which is substantially (and surprisingly) longer than the time scale of a few nanoseconds reported in previous studies [29, 31, 50]. This difference is probably due to the small protein (peptide) size, and therefore the low protein surface area, in our study although it also highlights the importance of carrying out extensive simulations for a quantitative computation of  $\Gamma_{23}$ .

The calculated  $\Gamma_{23}$  values between urea and TGLY are compared to the experimental results [25] in Table 2. In qualitative agreement with experiments, the calculated  $\Gamma_{23}$  are positive, which indicate that urea accumulates around TGLY relative to the bulk; as the urea concentration increases, both calculated and measured  $\Gamma_{23}$  increase in a linear fashion. The linear trend can be rationalized by noting that  $\mu_{23}$  and  $\mu_{33}$  in Eq.3 can be expanded into polynomials of  $m_2$  and  $m_3$  [25]:

$$\mu_{23} = RT (a_{11} + a_{21}m_2 + a_{12}m_3), \quad (6)$$

$$\mu_{33} = RT \left( \frac{1}{m_3} + 2a_{02} + a_{12}m_2 + 3a_{03}m_3 \right), \quad (7)$$

where  $a_{11}$ ,  $a_{21}$ , etc. are constants specific for a given ternary system. Since  $m_2$  is much smaller than  $m_3$  and  $a_{03}$  is much smaller than  $a_{12}$  for the TGLY-urea-water ternary system, the major contribution of  $m_3$  to  $\Gamma_{23}$  comes from  $a_{12}m_3$ , which gives rise to the linear dependence of  $\Gamma_{23}$  on  $m_3$ .

At the quantitative level, however, the computed values are too large by a factor of  $\sim 2$ ; for example, at the urea concentration of 1.21m, the calculated  $\Gamma_{23}$  is 0.33 while the measured value is 0.12. These results indicate that the standard CHARMM27 force field can be further improved to balance interactions among urea, protein backbone (TGLY does not have sidechain) and water (see below).

#### 2.1.2 Group decomposition of urea-protein backbone preferential interaction

—To gain additional insights into the molecular details of urea-protein backbone interactions, we decompose the overall  $\Gamma_{23}$  into group contributions from TGLY. We define seven groups in TGLY: the N-terminus  $-\text{NH}_3^+$ , the C-terminus carboxylate, two amide groups and three  $-\text{CH}_2-$  groups. The group-specific preferential interaction,  $\Gamma_{23}^i$ , is defined similarly to the overall  $\Gamma_{23}$  as:

$$\Gamma_{23}^i = \left\langle n_3^{II,i} - n_1^{II,i} \frac{n_3^I}{n_1^I} \right\rangle \quad (8)$$

where  $n_3^{II,i}$  and  $n_1^{II,i}$  are the number of solute/water molecules in the local domain II that are assigned to group  $i$ . A solute/water molecule is assigned to a specific group based on the shortest surface-to-surface distance between a solute/water and different groups in TGLY. In this way, the sum of  $\Gamma_{23}^i$  over all seven groups is identical to the overall  $\Gamma_{23}$ .

Such calculated  $\Gamma_{23}^i$  values between urea and TGLY are summarized in Table 3. The same trends are observed at three urea concentrations:  $\Gamma_{23}^i$  is negligible at two termini ( $-\text{NH}_3^+$  and  $-\text{COO}^-$ ) and positive around both amide groups and the aliphatic  $-\text{CH}_2-$  groups. In other words, CHARMM simulations predict that urea accumulates at not only amide groups, as deduced from previous experimental [4, 19, 25, 54] and simulation [16, 18] studies, but also at aliphatic  $-\text{CH}_2-$  groups, consistent with recent simulation results [21,22,28,55].

As emphasized in previous studies [56], a group based decomposition of  $\Gamma_{23}$  is not necessarily straightforward due to potential co-operative effects and the complex nature of peptide surfaces. Therefore, to probe the preferential interaction between urea and aliphatic ( $-\text{CH}_2-$ ) groups directly, a set of propane-urea-water simulations (80 ns in total) are also carried out with the CHARMM27 [38] force field; it's worth noting that while the non-polar hydrogen atoms in propane and TGLY bear the same partial charge of +0.09 (a convention adopted in CHARMM22/27 force fields), the net charge of the  $-\text{CH}_2$  group in TGLY is +0.16 while the  $-\text{CH}_{2,3}$  groups in propane are all charge-neutral).  $\Gamma_{23}$  between urea and propane is then calculated from these simulations following the same procedure as for urea and TGLY. As shown in Fig.4, the calculated  $\Gamma_{23}$  between urea and propane is essentially converged within 30ns. This faster convergence compared to urea/TGLY is presumably due to the more homogeneous chemical nature of propane compared to TGLY. The calculated  $\Gamma_{23}$  between urea and propane is  $0.15 \pm 0.07$  at [urea]=2.26m, which is qualitatively consistent with the sum of contributions from three  $-\text{CH}_2-$  groups in Table 3 for TGLY ( $0.25 \pm 0.04$ ). Therefore, it is confirmed that, with the CHARMM27 force field, urea preferentially interacts with the aliphatic  $-\text{CH}_2-$  groups independent of the peptide backbone groups (see below for additional discussions).

## 2.2 Preferential interaction between GB and TGLY

As mentioned in Sect.1, the distribution of GB near TGLY is qualitatively incorrect without scaling the partial charges of “raw” CHARMM27 parameters. Therefore, we focus on results obtained with the scaled partial charges (summarized in Supporting Information).

**2.2.1  $\Gamma_{23}$  between GB and TGLY**—The radial distribution functions  $g(r)$  of GB and water around TGLY for two GB concentrations are shown in Fig. 5. Similar to the urea case,  $g(r)$  exhibits very small variation with respect to GB concentration. The  $g(r)$  of GB has a much lower peak height than water, which suggests that GB is locally more excluded from TGLY than water. The convergence behaviors of calculated  $\Gamma_{23}$  between GB and TGLY for two GB concentrations are shown in Fig. 6. Once again, it takes  $\sim 100$ ns to obtain converged  $\Gamma_{23}$  values, similar to the urea case.

As shown in Table 2, with the scaled partial charges, calculated  $\Gamma_{23}$  values agree rather nicely with experimental data for both GB concentrations. For example, the calculated  $\Gamma_{23}$  at [GB]=1.35m is  $\sim -0.24 \pm 0.04$ , compared to the experimental value of  $\sim -0.16$ . The negative value indicates that GB is excluded from the TGLY surface (protein backbone), which is

consistent with the fact that GB acts as a protein protectant. Moreover, the concentration dependence of  $\Gamma_{23}$  (i.e.  $\Gamma_{23}$  decreases as [GB] increases) is also qualitatively reproduced by MD simulations, which is encouraging considering the simple scheme used to construct a force field model for GB.

### 2.2.2 Group decomposition of GB-protein backbone preferential interaction—

Following the same group definitions used for characterizing urea-TGLY interactions, group decomposition is performed for the GB-TGLY  $\Gamma_{23}$ . As shown in Table 3, the main exclusion region for GB is the carboxylate group of the peptide, which is consistent with experimental results for GB and anionic (protein carboxylate or DNA phosphate) biopolymer surface [7]; to a much reduced level, GB is excluded from the aliphatic  $-\text{CH}_2-$  groups. By contrast, GB weakly accumulates around the  $-\text{NH}_3^+$  group and, to an even smaller degree, around amide groups.

In a recent set of analyses [27], Record *et al.* have measured  $\mu_{23}$  between GB and a broad series of small solutes and proteins. Based on global fitting of these data, they obtained the  $\mu_{23}/\text{ASA}$  values between GB and different types of surface by assuming that  $\mu_{23}$  for different surfaces are additive and proportional to ASA; here, ASA stands for (solvent) accessible surface area. They found that this model can predict  $\mu_{23}$  values for a set of small solutes and proteins in an overall good agreement with experiments. Motivated by these findings, we perform a second type of group decomposition for the GB-TGLY  $\Gamma_{23}$ , in which TGLY is divided into five instead of seven groups based on surface type: carboxylate O, amide O, cationic N, amide N and C. The hydrogen atoms are assigned to the heavy atoms which they are bonded to and the surface C includes all carbon atoms. The group decomposition of  $\Gamma_{23}$  with this scheme are shown in Table 5 for two GB concentrations, along with the predicted ones based on  $\mu_{23}/\text{ASA}$  values fitted to experiments. Overall, the calculated  $\Gamma_{23}^i$  from MD simulations agree well with predictions based on  $\mu_{23}/\text{ASA}$ . The only exception is for amide O, for which the predicted  $\Gamma_{23}^i$  is negative while the calculated value based on MD is nearly zero. Nevertheless, the key feature that GB is significantly excluded from carboxylate O is well captured by MD simulations. Moreover, we note that the total  $\Gamma_{23}$  between GB and TGLY predicted based on  $\mu_{23}/\text{ASA}$  doesn't agree perfectly with experimental measurement. For example, the predicted  $\Gamma_{23}$  at [GB]=1.35m is  $\sim -0.35$ , which is further away from the experimental value of  $-0.16$  compared to MD simulations ( $-0.24$ ).

## 2.3 Implications

The present study has important mechanistic and technical implications regarding small solute effects on protein stability.

Regarding the thermodynamic basis of protein denaturation by urea, the group decomposition of CHARMM-predicted values of  $\Gamma_{23}$  (Table 3) indicates that urea interacts favorably relative to water with both the amide and the aliphatic  $-\text{CH}_2-$  groups. The favorable urea-amide preferential interaction has been observed in simulations [16,18] and has been quantified by recent analyses of experimental data [19, 25, 54]. For example, from an analysis of experimental osmometric (isopiestic distillation) values of  $\Gamma_{23}$  for interactions of urea with small peptides (e.g. GlyGly, TGLY, GlyAla), Cannon *et al.* [25] concluded that the dominant preferential interaction of urea with the protein backbone is a favorable interaction with amide groups and were able to quantify this interaction. In a recent NMR study [19], Lim *et al.* concluded that urea forms hydrogen bonds with the amide group based on the dependence of hydrogen exchange rate on urea concentration. The urea-aliphatic group preferential interaction observed here is qualitatively consistent with limited data from solubility experiments, which show that aliphatic hydrocarbons (longer than ethane) are

more soluble in 7 M urea than in water [57,58]. Additionally, the authors of an osmotic stress/X-ray scattering study of the effect of urea on forces between hydroxypropylcellulose polymers concluded that urea interacts favorably (but very weakly) with the mostly aliphatic surface of these polymers. Moreover, several recent simulation studies [21, 22, 28, 55] also proposed that the dispersion interaction between urea and aliphatic groups could be a driving force for denaturation. The most direct computational evidence is the observation that urea weakens the hydrophobic interactions between pure hydrophobic particles and unfolds pure hydrophobic polymers [22]. Only our study, however, has explicitly and quantitatively compared simulation result with experimental thermodynamic data, which is essential for evaluating the findings from simulations. For example, with the CHARMM27 force field, the calculated preferential coefficient of urea near propane is  $0.15 \pm 0.07$  at  $[\text{urea}] = 2.26\text{m}$ , which is rather large considering the experimental observation that the butane solubility as a function of urea concentration [57] indicates a very small positive  $\Gamma_{23}$ . Overestimation of preferential interaction between urea-aliphatic groups with the CHARMM27 force field is consistent with the overestimated  $\Gamma_{23}$  for urea near TGLY (e.g., at  $[\text{urea}]$  of  $1.21\text{m}$ , the calculated  $\Gamma_{23}$  is 0.33 while the measured value is 0.12); as noted above, the  $-\text{CH}_2$  groups in propane and TGLY have fairly different overall charges although polar hydrogens have the same partial charges. To further validate this conjecture, comparing computed and measured group decomposition of  $\Gamma_{23}$  for urea/peptide (E. Guinn and MTR, work in progress) as we have done for GB/peptide is valuable, and the results can provide guidance to the improvement of the CHARMM27 force field. In this context, we note that even a small positive preferential interaction between urea and aliphatic groups still may indicate a significant contribution of such effects in protein unfolding since the amount of exposed aliphatic hydrocarbon surface upon unfolding is very large.

For the stabilizing mechanism of GB, the present study reproduces the overall and group-specific  $\Gamma_{23}$  values between GB and TGLY reasonably well compared to experimental analyses. The major exclusion region for GB (see Table 3 and 5) is the carboxylate group, which is consistent with the experimental evidence [7] that GB is excluded from anionic biopolymer surfaces. GB is also slightly excluded from aliphatic groups ( $-\text{CH}_2$  groups in Table 3 and aliphatic C surfaces in Table 5), which is in line with findings from recent X-ray scattering experiments, which showed that GB is excluded from the nonpolar surface of hydroxypropylcellulose [59]. The semi-quantitative agreement between the surface-specific  $\Gamma_{23}$  and the prediction of Record et al. based on  $\mu_{23}/ASA$  (Table 5) supports the simple thermodynamic model that assumes additivity of preferential interactions between GB and various biomolecular surfaces.

On the technical side, the present study serves as a benchmark for CHARMM [38], a popular classical, non-polarizable force field, in terms of describing interactions between small solutes (urea and GB), protein backbone (TGLY peptide) and water. It is encouraging that the CHARMM27 force field is observed to reproduce  $\Gamma_{23}$  between urea and TGLY on a semi-quantitative level, although there is clearly room for further improvements. For GB, however, the computed  $\Gamma_{23}$  has even the wrong sign if parameters are not carefully chosen to balance the interactions among GB, TGLY and water. This result clearly highlights that care needs to be exercised when constructing force field models based on existing parameters for chemically similar groups, which is commonly done in the literature based on the consideration of force field compatibility.

The sensitivity of  $\Gamma_{23}$  to force field parameters suggests that small solute thermodynamic data can be valuable in guiding the development of accurate molecular models, which has already been recognized by several authors [33, 37] but yet to be broadly appreciated. In force field developments, emphasis is often put on solvation free energies, which measure the interaction between functional groups and water. By contrast, small solute data measure

interactions between functional groups *relative* to their interactions with water, which are precisely the driving force behind many biological processes such as ligand binding and protein folding/association. Therefore, small solute thermodynamic data nicely complement solvation free energies in force field developments. Considering the molecular (as opposed to atomic) nature of small solute data (e.g., as illustrated by the group decompositions), this is particularly true for model development at the coarse-grained level, which we plan to pursue in the near future with joint experimental/computational efforts. Along this line, we note that to reliably describe the thermodynamics of processes that involve a significant change in the local environment of amino acids (e.g., protein unfolding), it might be important to include electronic polarization so that the effective interaction between groups can adequately adjust to the local environment.

### 3 Conclusions

Motivated by recent experimental efforts in understanding small solute effects on protein stability, we have carried out extensive MD simulations to characterize the interaction between two widely used small solutes, urea and glycine betaine (GB), and a tri-glycine peptide that represents protein backbone. The key quantity of interest is the preferential interaction coefficient,  $\Gamma_{23}$ , which characterizes the degree of solute accumulation near or exclusion from the protein surface relative to the bulk. To ensure convergence and a thorough comparison with experiments,  $\Gamma_{23}$  is calculated for multiple solute concentrations and each solute-peptide-water ternary system is studied with 200–300 ns of atomistic simulations using the CHARMM force field. The results show that good agreement with experiments can be obtained for both solutes if care is exercised to balance the interactions among water, solute and protein. On the other hand, a model based on directly taking parameters for chemically similar groups from existing force field leads to qualitatively incorrect results (i.e., wrong sign in  $\Gamma_{23}$ ), which highlights the value of small solute thermodynamic data in guiding the development of accurate force fields for biomolecules.

Despite potential caveats noted in previous study [56], we find that decomposing the calculated  $\Gamma_{23}$  values into group contributions allows us to gain new insights regarding the mechanism of small solute effects. Use of the CHARMM force field predicts that urea preferentially interacts with both amide and aliphatic hydrocarbon groups ( $\text{CH}_2$  groups in the tri-glycine peptide); this supports the conclusion from several recent simulation studies [21, 22, 55] that interaction between urea and aliphatic groups also plays an important role in urea-induced denaturation. Quantitatively, however, comparison to experimental solubility [57, 58] and X-ray scattering [59] data indicates that the CHARMM force field and perhaps others likely overestimate the interaction between urea and aliphatic groups. Regarding GB, the present study finds that it is excluded mainly from the carboxylate groups and weakly from nonpolar groups, results which are consistent with recent experimental findings [7, 59]. The group decomposition of  $\Gamma_{23}$  also supports an additive thermodynamic model for preferential interactions between GB and various biomolecular surfaces [27].

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

The authors thank Dr. I. A. Shkel for stimulating discussions. Computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

**Funding:** The research has been supported from the National Institutes of Health (R01-GM071428 to QC and R01-GM047022 to T. Record).

## Abbreviations

<b>GB</b>	Glycine Betaine
<b>TGLY</b>	Tri-glycine
<b>COM</b>	center of mass
<b>vdW</b>	van der Waals
<b>ASA</b>	Accessible Surface Area

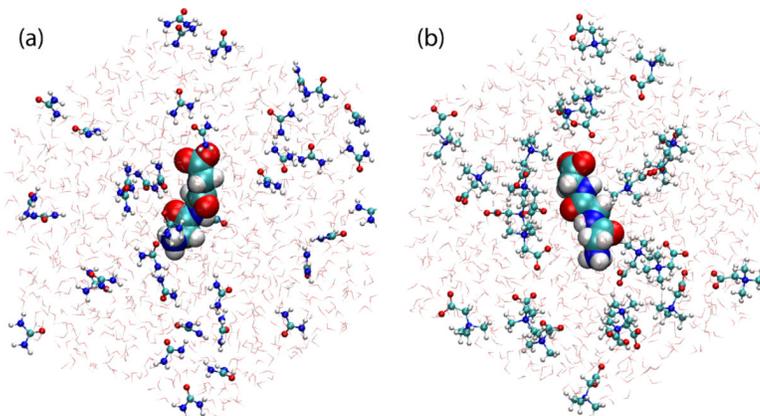
## References

- [1]. Wyman J. Linked functions and reciprocal effects in hemoglobin: A second look. *Adv. Prot. Chem.* 1964; 19:223–286.
- [2]. Record MT, Zhang WT, Anderson CF. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, hofmeister effects, and osmotic effects of salts. *Adv. Prot. Chem.* 1998; 51:281–353.
- [3]. Timasheff SN. Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of the complicated. *Adv. Prot. Chem.* 1998; 51:355–432.
- [4]. Courtenay ES, Capp MW, Anderson CF, Record MT. 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.* 2000; 39:4455–4471. [PubMed: 10757995]
- [5]. Hopkins FG. Denaturation of proteins by urea and related substances. *Nature.* 1930; 126:328–330. 383–384.
- [6]. Tanford C. Protein denaturation. *Adv. Prot. Chem.* 1968; 23:121–282.
- [7]. Felitsky DJ, Cannon JG, Capp MW, Hong J, Van Wynsberghe AW, Anderson CF, Record MT. The exclusion of glycine betaine from anionic biopolymer surface: Why glycine betaine is an effective osmoprotectant but also a compatible solute. *Biochemistry.* 2004; 43:14732–14743. [PubMed: 15544344]
- [8]. Storey R, Jones RGW. Betaine and choline levels in plants and their relationship to nacl stress. *Plant Sci. Lett.* 1975; 4:161–168.
- [9]. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water-stress - evolution of osmolyte systems. *Science.* 1982; 217:1214–1222. [PubMed: 7112124]
- [10]. Lin TY, Timasheff SN. Why do some organisms use a urea-methylamine mixture as osmolyte - thermodynamic compensation of urea and trimethylamine n-oxide interactions with protein. *Biochemistry.* 1994; 33:12695–12701. [PubMed: 7918496]
- [11]. Yancey PH, Somero GN. Methylamine osmoregulatory solutes of elasmobranch fishes counteract urea inhibition of enzymes. *J. Exp. Zool.* 1980; 212:205–213.
- [12]. Volker J, Breslauer KJ. Communication between noncontacting macromolecules. *Annu. Rev. Biophys. Biomol. Struct.* 2005; 34:21–42. [PubMed: 15869382]
- [13]. Frank HS, Franks F. Structural approach to solvent power of water for hydrocarbons - urea as a structure breaker. *J. Chem. Phys.* 1968; 48:4746–&.
- [14]. Vanzi F, Madan B, Sharp K. Effect of the protein denaturants urea and guanidinium on water structure: A structural and thermodynamic study. *J. Am. Chem. Soc.* 1998; 120:10748–10753.
- [15]. Idrissi A, Sokolic F, Perera A. A molecular dynamics study of the urea/water mixture. *J. Chem. Phys.* 2000; 112:9479–9488.
- [16]. Bennion BJ, Daggett V. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Nat. Acad. Sci. USA.* 2003; 100:5142–5147. [PubMed: 12702764]

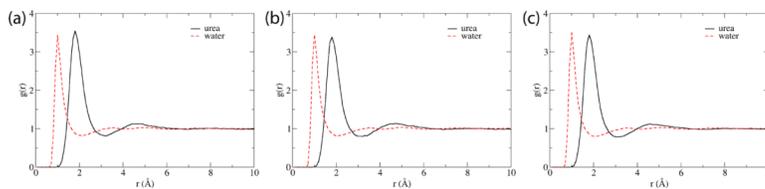
- [17]. Zou Q, Habermann-Rottinghaus SM, Murphy KP. Urea effects on protein stability: Hydrogen bonding and the hydrophobic effect, *Proteins Struct. Funct. Genet.* 1998; 31:107–115.
- [18]. Tobi D, Elber R, Thirumalai D. The dominant interaction between peptide and urea is electrostatic in nature: A molecular dynamics simulation study. *Biopolymers.* 2003; 68:359–369. [PubMed: 12601795]
- [19]. Lim WK, Rosgen J, Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc. Nat. Acad. Sci. USA.* 2009; 106:2595–2600. [PubMed: 19196963]
- [20]. Paul S, Patey GN. The influence of urea and trimethylamine-n-oxide on hydrophobic interactions. *J. Phys. Chem. B.* 2007; 111:7932–7933. [PubMed: 17580863]
- [21]. Hua L, Zhou RH, Thirumalai D, Berne BJ. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Nat. Acad. Sci. USA.* 2008; 105:16928–16933. [PubMed: 18957546]
- [22]. Zangi R, Zhou RH, Berne BJ. Urea's action on hydrophobic interactions. *J. Am. Chem. Soc.* 2009; 131:1535–1541. [PubMed: 19123816]
- [23]. Shearwin KE, Timasheff SN. Linkage between ligand-binding and control of tubulin conformation. *Biochemistry.* 1992; 31:8080–8089. [PubMed: 1510990]
- [24]. Record MT, Anderson CF. Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a 2-domain model. *Biophys. J.* 1995; 68:786–794. [PubMed: 7756545]
- [25]. Cannon JG, Anderson CF, Record MT. Urea-amide preferential interactions in water: Quantitative comparison of model compound data with biopolymer results using water accessible surface areas. *J. Phys. Chem. B.* 2007; 111:9675–9685. [PubMed: 17658791]
- [26]. Hong J, Capp MW, Saecker RM, Record MT. Use of urea and glycine betaine to quantify coupled folding and probe the burial of dna phosphates in lac repressor - lac operator bindings. *Biochemistry.* 2005; 44:16896–16911. [PubMed: 16363803]
- [27]. Capp MW, Pegram LM, Saecker RM, Kratz M, Riccardi D, Wendorff T, Cannon JG, Record MT Jr. Interactions of glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation and prediction of *m*-values. *Biochem.* 2009; 48:10372–10379. [PubMed: 19757837]
- [28]. Caflisch A, Karplus M. Structural details of urea binding to barnase: a molecular dynamics analysis, *Structure Fold & Design.* 1999; 7:477–488.
- [29]. Baynes BM, Trout BL. Proteins in mixed solvents: A molecular-level perspective. *J. Phys. Chem. B.* 2003; 107:14058–14067.
- [30]. Smith PE. Cosolvent interactions with biomolecules: Relating computer simulation data to experimental thermodynamic data. *J. Phys. Chem. B.* 2004; 108:18716–18724.
- [31]. Pierce V, Kang M, Aburi M, Weerasinghe S, Smith PE. Recent applications of kirkwood-buff theory to biological systems. *Cell Biochem. Biophys.* 2008; 50:1–22. [PubMed: 18043873]
- [32]. Kang M, Smith PE. Preferential interaction parameters in biological systems by kirkwood-buff theory and computer simulation. *Fluid Phase Equilibria.* 2007; 256:14–19.
- [33]. Weerasinghe S, Smith PE. A kirkwood-buff derived force field for sodium chloride in water. *J. Chem. Phys.* 2003; 119:11342–11349.
- [34]. Weerasinghe S, Smith PE. Kirkwood-buff derived force field for mixtures of acetone and water. *J. Chem. Phys.* 2003; 118:10663–10670.
- [35]. Weerasinghe S, Smith PE. A kirkwood-buff derived force field for the simulation of aqueous guanidinium chloride solutions. *J. Chem. Phys.* 2004; 121:2180–2186. [PubMed: 15260772]
- [36]. Weerasinghe S, Smith PE. A kirkwood-buff derived force field for methanol and aqueous methanol solutions. *J. Phys. Chem. B.* 2005; 109:15080–15086. [PubMed: 16852908]
- [37]. Kang M, Smith PE. A kirkwood-buff derived force field for amides. *J. Comput. Chem.* 2006; 27:1477–1485. [PubMed: 16823811]
- [38]. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J,

- Watanabe M, Wiorcikiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 1998; 102:3586–3616.
- [39]. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1983; 4:187–217. <http://www.charmm.org>.
- [40]. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 79:926–935.
- [41]. Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* 1994; 15:667–683.
- [42]. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh ewald method. *J. Chem. Phys.* 1995; 103:8577–8593.
- [43]. Nose S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* 1984; 81:511–519.
- [44]. Hoover WG. Canonical dynamics - equilibrium phase-space distributions. *Phys. Rev. A.* 1985; 31:1695–1697. [PubMed: 9895674]
- [45]. Andersen HC. Molecular-dynamics simulations at constant pressure and-or temperature. *J. Chem. Phys.* 1980; 72:2384–2393.
- [46]. Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant-pressure molecular-dynamics simulation - the langevin piston method. *J. Chem. Phys.* 1995; 103:4613–4621.
- [47]. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *J. Comput. Phys.* 1977; 23:327–341.
- [48]. Formanek MS, Ma L, Cui Q. Effects of temperature and salt concentration on the structural stability of human lymphotactin: Insights from molecular simulations. *J. Am. Chem. Soc.* 2006; 128:9506–9517. [PubMed: 16848488]
- [49]. Ma L, Cui Q. The temperature dependence of salt-protein association is sequence specific. *Biochemistry.* 2006; 45:14466–14472. [PubMed: 17128985]
- [50]. Shukla D, Shinde C, Trout BL. Molecular computations of preferential interaction coefficients of proteins. *J. Phys. Chem. B.* 2009; 113:12546–12554. [PubMed: 19697945]
- [51]. Harihara PC, Pople JA. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theoretica Chimica Acta.* 1973; 28:213–222.
- [52]. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr.; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Revision C.02. Gaussian, Inc.; Wallingford, CT: 2004. Gaussian 03.
- [53]. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press; San Diego, London: 2002.
- [54]. Courtenay ES, Capp MW, Record MT Jr. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. *Prot. Sci.* 2001; 10:2485–2497.
- [55]. Stumpe MC, Grubmuller H. Interaction of urea with amino acids: Implications for urea-induced protein denaturation. *J. Am. Chem. Soc.* 2007; 129:16126–16131. [PubMed: 18047342]
- [56]. Aburi M, Smith PE. A combined simulation and Kirkwood-Buff approach to quantify co-solvent effects on the conformational preferences of peptides in solution. *J. Phys. Chem. B.* 2004; 108:7382–7388.

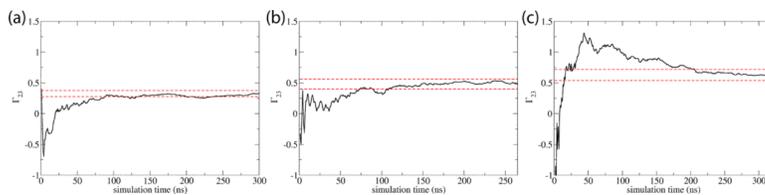
- [57]. Wetlaufer DB, Malik SK, Stoller L, Coffin RL. Nonpolar group participation in the denaturation of proteins by urea and guanidinium salts. model compound studies. *J. Am. Chem. Soc.* 1964; 86:508–514.
- [58]. Graziano G. On the solubility of aliphatic hydrocarbons in 7 m aqueous urea. *J. Phys. Chem. B.* 2001; 105:2632–2637.
- [59]. Stanley C, Rau DC. Assessing the interaction of urea and protein-stabilizing osmolytes with the nonpolar surface of hydroxypropylcellulose. *Biochemistry.* 2008; 47:6711–6718. [PubMed: 18512956]
- [60]. Humphrey W, Dalke A, Schulten K. Vmd - visual molecular dynamics. *J. Mol. Graph.* 1996; 14:33–38. [PubMed: 8744570]



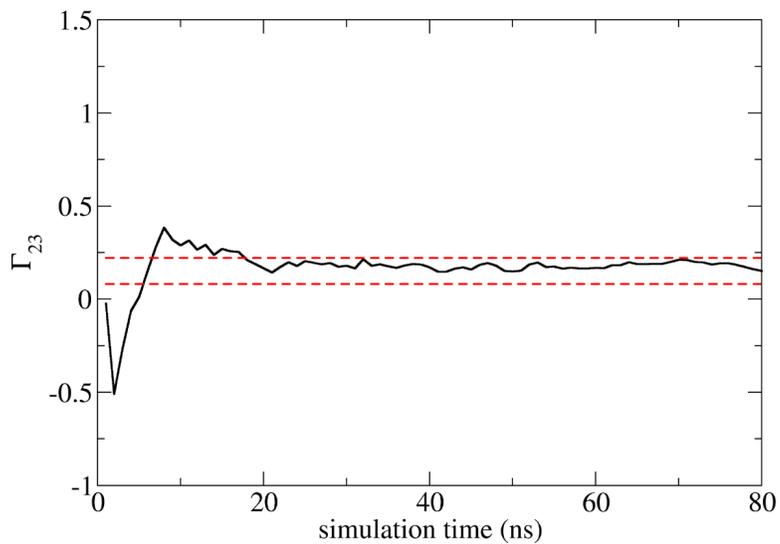
**Figure 1.** Snapshots from MD trajectories for (a) TGLY-urea-water at [urea]=2.26m and (b) TGLY-GB-water at [GB]=1.35m. TGLY and water molecules are shown in van der Waals and line scheme, respectively. Urea and GB molecules are shown in CPK scheme. Figures are made using VMD. [60]



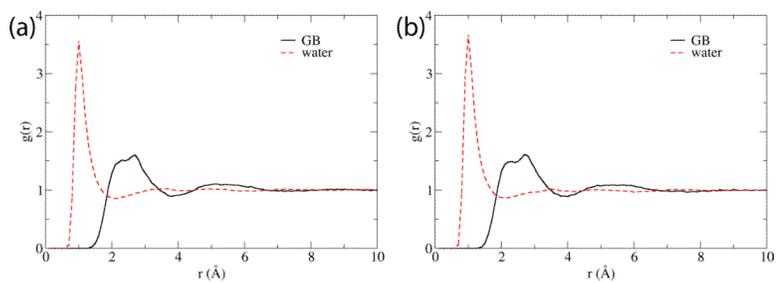
**Figure 2.** Radial distribution function,  $g(r)$ , of urea and water molecules around TGLY at different urea concentrations: (a) [urea]=1.21m, (b) [urea]=2.26m, and (c) [urea]=3.61m. The distance  $r$  is the shortest distance between the center of mass of a urea or water molecule to the van der Waals surface of TGLY.



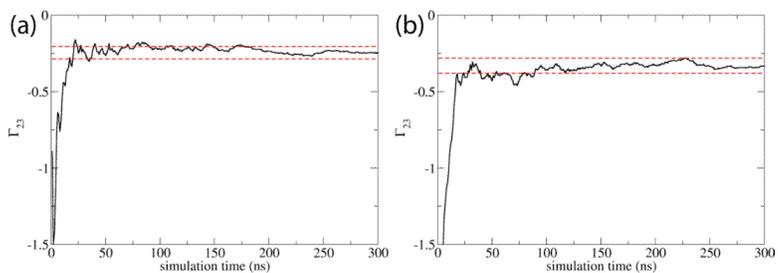
**Figure 3.** The convergence behaviors of  $\Gamma_{23}$  with respect to accumulated simulation time for TGLY-urea-water systems at different urea concentrations: (a) [urea]=1.21m, (b) [urea]=2.26m and (c) [urea]=3.61m. The red dashed lines indicate the converged  $\Gamma_{23}$  value plus/minus the estimated standard deviation.



**Figure 4.** The convergence behavior of  $\Gamma_{23}$  with respect to accumulated simulation time for the propane-urea-water system at [urea]=2.26m. The red dashed lines indicate the converged  $\Gamma_{23}$  value plus/minus the estimated standard deviation.



**Figure 5.** Radial distribution function,  $g(r)$ , of GB and water molecules around TGLY at different GB concentrations: (a)  $[GB]=1.35m$  and (b)  $[GB]=2.04m$ . The distance  $r$  is the shortest distance between the center of mass of a GB or water molecule to the van der Waals surface of TGLY.



**Figure 6.** The convergence behaviors of  $\Gamma_{23}$  with respect to accumulated simulation time for TGLY-GB-water systems at different GB concentrations: (a) [GB]=1.35m and (b) [GB]=2.04m. The red dashed lines indicate the converged  $\Gamma_{23}$  value plus/minus the estimated standard deviation.

**Table 1**Simulation details of TGLY-urea-water and TGLY-GB-water systems<sup>a</sup>

<b>solute</b>	<b>conc. (m)</b>	<b>n<sub>3</sub></b>	<b>n<sub>1</sub></b>	<b>simulation scheme<sup>b</sup></b>	<b>box length<sup>c</sup> (Å)</b>
urea	1.21	23	1052	100 × 3ns	36.0
urea	2.26	39	957	133 × 2ns	35.4
urea	3.61	56	862	160 × 2ns	34.8
GB	1.35	23	943	100 × 3ns	35.4
GB	2.04	32	873	100 × 3ns	35.4

<sup>a</sup>Throughout the manuscript, component 1 is water, component 2 is protein/peptide and component 3 is the small solute;

<sup>b</sup>Given in the format of number of independent trajectories (e.g., 100) × length of each trajectory (e.g., 3 ns);

<sup>c</sup>Average box length with the rhombic dodecahedron periodic boundary condition.

**Table 2**Calculated and experimentally measured [25]  $\Gamma_{23}$  values for all simulated systems<sup>a</sup>

System	urea 1.21m	urea 2.26m	urea 3.61m	GB 1.35m	GB 2.04m
$\Gamma_{23}^{cal}$	$0.33 \pm 0.05$	$0.48 \pm 0.08$	$0.63 \pm 0.09$	$-0.24 \pm 0.04$	$-0.33 \pm 0.05$
$\Gamma_{23}^{exp}$	0.12	0.18	0.20	-0.16	-0.21

<sup>a</sup>The error bar for experimentally determined  $\Gamma_{23}$  is less than 0.01.

**Table 3**

Group-specific  $\Gamma_{23}$  values for all simulated systems following the first decomposition scheme that partitions TGLY into seven groups

System	urea 1.21m	urea 2.26m	urea 3.61m	GB 1.35m	GB 2.04m
$\Gamma_{23} \text{NH}_3^+$	0.01 (0.01) <sup>b</sup>	0.01 (0.02)	-0.03 (0.02)	0.09 (0.01)	0.12 (0.02)
$\Gamma_{23} \text{CH}_2(1)^a$	0.04 (0.01)	0.05 (0.02)	0.07 (0.02)	-0.06 (0.01)	-0.08 (0.01)
$\Gamma_{23} \text{CH}_2(2)$	0.08 (0.01)	0.10 (0.01)	0.17 (0.01)	-0.04 (0.01)	-0.09 (0.01)
$\Gamma_{23} \text{CH}_2(3)$	0.05 (0.01)	0.10 (0.01)	0.13 (0.01)	-0.01 (0.01)	0 (0.01)
$\Gamma_{23} \text{CONH}(1)$	0.06 (0.01)	0.11 (0.02)	0.16 (0.01)	0.01 (0.01)	0.04 (0.01)
$\Gamma_{23} \text{CONH}(2)$	0.07 (0.01)	0.11 (0.02)	0.15 (0.01)	0.02 (0.01)	0.02 (0.01)
$\Gamma_{23} \text{COO}^-$	0.03 (0.02)	0.01 (0.03)	-0.01 (0.03)	-0.25 (0.02)	-0.34 (0.02)

<sup>a</sup>CH<sub>2</sub>(1) refers to the first -CH<sub>2</sub>- group close to the N-terminus of TGLY;

<sup>b</sup> numbers with parentheses are standard deviations.

**Table 4**

Group-specific  $\Gamma_{23}$  values for TGLY-urea-water with the second decomposition scheme that partitions TGLY into five types of surfaces

System	urea 1.21m	urea 2.26m	urea 3.61m
$\Gamma_{23}$ carboxylate O	0.01 (0.02)	-0.01 (0.03)	-0.05 (0.03)
$\Gamma_{23}$ amide O	0.08 (0.01)	0.10 (0.02)	0.18 (0.02)
$\Gamma_{23}$ cationic N	0.01 (0.01)	~0 (0.02)	-0.03 (0.02)
$\Gamma_{23}$ amide N	0.03 (~0)	0.06 (~0)	0.07 (~0)
$\Gamma_{23}$ C	0.20 (0.02)	0.34 (0.04)	0.46 (0.03)
$\Gamma_{23}$ total	0.33 (0.05)	0.48 (0.08)	0.63 (0.09)

**Table 5**

Group-specific  $\Gamma_{23}$  values for TGLY-GB-water with the second decomposition scheme that partitions TGLY into five types of surfaces

System	GB 1.35m (sim) <sup>a</sup>	GB 1.35m (pred) <sup>b</sup>	GB 2.04m (sim)	GB 2.04m (pred)
$\Gamma_{23}$ carboxylate O	-0.25 (0.02)	-0.29	-0.34 (0.02)	-0.38
$\Gamma_{23}$ amide O	0 (0.01)	-0.15	0 (0.01)	-0.20
$\Gamma_{23}$ cationic N	0.09 (0.01)	0.04	0.12 (0.02)	0.05
$\Gamma_{23}$ amide N	0.02 (0.003)	0.11	0.04 (0.004)	0.15
$\Gamma_{23}$ C	-0.10 (0.02)	-0.05	-0.16 (0.02)	-0.06
$\Gamma_{23}$ total	-0.24 (0.04)	-0.35	-0.33 (0.05)	-0.43

<sup>a</sup>Group-specific  $\Gamma_{23}$  values from MD simulations, those with parentheses are standard deviations;

<sup>b</sup>Group-specific  $\Gamma_{23}$  values predicted from a model based on  $\mu_{23}/ASA$  fitted to experimental data (see text for details).