

# Quantitative Characterization of Local Protein Solvation To Predict Solvent Effects on Protein Structure

Vincent Vagenende<sup>†\*</sup> and Bernhardt L. Trout<sup>‡</sup>

<sup>†</sup>Bioprocessing Technology Institute, Agency for Science, Technology and Research (A\*STAR), Centros, Singapore; and <sup>‡</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

**ABSTRACT** Characterization of solvent preferences of proteins is essential to the understanding of solvent effects on protein structure and stability. Although it is generally believed that solvent preferences at distinct loci of a protein surface may differ, quantitative characterization of local protein solvation has remained elusive. In this study, we show that local solvation preferences can be quantified over the entire protein surface from extended molecular dynamics simulations. By subjecting microsecond trajectories of two proteins (lysozyme and antibody fragment D1.3) in 4 M glycerol to rigorous statistical analyses, solvent preferences of individual protein residues are quantified by local preferential interaction coefficients. Local solvent preferences for glycerol vary widely from residue to residue and may change as a result of protein side-chain motions that are slower than the longest intrinsic solvation timescale of ~10 ns. Differences of local solvent preferences between distinct protein side-chain conformations predict solvent effects on local protein structure in good agreement with experiment. This study extends the application scope of preferential interaction theory and enables molecular understanding of solvent effects on protein structure through comprehensive characterization of local protein solvation.

## INTRODUCTION

Cosolvents such as denaturants, salts, amino acids, polyols, and sugars play an important role in many protein processes involving protein folding, stabilization, and association (1–3). This is because the addition of cosolvents to aqueous protein solutions commonly alters the equilibrium between protein conformations (4–6). Over the past decades, a rigorous thermodynamic framework has been developed that relates cosolvent effects on protein conformations with solvent preferences of the protein surface (4,7–19). This framework—which is often referred to as “preferential interaction theory”—stipulates that adding cosolvent to a protein solution will shift the protein toward conformations with a greater degree of preferential solvation by the cosolvent. Preferential solvation of a protein is quantified by the preferential interaction coefficient  $\Gamma_{XP}$  (4,20). Because every solvent molecule at the protein surface contributes to  $\Gamma_{XP}$  (7,11,21), detailed characterization of protein solvation is required for molecular understanding of solvent effects on protein conformations.

Preferential interactions reflect relative preferences of the protein surface for either cosolvent or water, and are manifested in local concentration ratios of cosolvent and water that are either greater (preferential solvation), smaller (preferential hydration), or equal (neutral solvation) with respect to the bulk solvent (4). Because protein surfaces comprise physically and chemically distinct surface loci, protein solvation in a mixed solvent can be conceived of as an ensemble of preferentially hydrated, solvated, and neutral solvent regions near the protein surface (4,22). Just like

a protein displays a mosaic of heterogeneous surface loci to the solvent, the solvent surrounding a protein is expected to form a mosaic of solvent regions with varying degrees of preferential interactions.

Unfortunately, characterization of the protein solvation mosaic has remained elusive to this day. Although various spectroscopy- and NMR-based techniques have revealed many aspects of protein solvation, these techniques do not allow quantifying local solvation preferences (23–25). On the other hand, equilibrium techniques such as vapor pressure osmometry and dialysis-densitometry allow the measurement of the preferential interaction coefficient  $\Gamma_{XP}$  of the ensemble-average of all protein conformations in a solvent mixture (4,26,27). However, equilibrium techniques cannot quantify differences of  $\Gamma_{XP}$  between distinct protein conformations and are unable to resolve local preferential interactions at distinct protein surface loci.

One approach to understand solvent effects on protein conformations was pioneered by Tanford (22), who quantified thermodynamic solvent effects on smaller constituent groups of a protein molecule and hypothesized the additivity of individual contributions of the constituent groups. Group additivity has proven to be a useful assumption for quantifying the effects of a number of cosolvents on protein folding (28–30); however, other studies reported solvent effects on proteins for which group additivity is not valid (31–33). Hence, more research is needed to determine the range of cosolvents and the extent of conformational changes for which group additivity holds.

Solvent effects on protein (un)folding events have been directly observed from extended molecular-dynamics (MD) simulations (34–38), and solvent effects on the free energy landscape of peptides and mini-proteins have been

Submitted May 28, 2012, and accepted for publication August 3, 2012.

\*Correspondence: vincent\_vagenende@bti.a-star.edu.sg

Editor: Nathan Baker.

© 2012 by the Biophysical Society  
0006-3495/12/09/1354/9 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2012.08.011>

quantified through advanced sampling techniques (39–41). These techniques have provided a wealth of information on the folding mechanisms of proteins, but they have also generated widely differing views on the molecular mechanisms by which cosolvents alter protein conformations. Perhaps the most prevalent reason for this disagreement is the difficulty to differentiate cause and effect between concurrent changes of protein structure and protein solvation. Another reason for the longstanding disagreement on the molecular mechanisms of cosolvent effects on protein conformations could be that these studies provide no or limited insight in local solvent preferences at distinct loci of the protein surface.

An alternative approach for obtaining detailed molecular understanding of solvent effects on protein conformations is by quantifying differences of  $\Gamma_{XP}$  values between specific protein conformations.  $\Gamma_{XP}$  of a specific protein conformation can be calculated from MD simulations of a single protein conformation with well-defined atom coordinates or an ensemble of protein conformations corresponding to a particular protein state (21,42–46). The solvent-induced equilibrium shift between two protein conformations can then be quantified from the difference of  $\Gamma_{XP}$  values between the respective protein conformations (42,46). Because  $\Gamma_{XP}$  values are calculated by summing up solvent molecules at all loci of the protein surface (21), this approach could, in principle, elucidate the role of individual solvent molecules and protein surface loci in solvent-induced conformational changes. However, even though previous studies have shown that  $\Gamma_{XP}$  values calculated from MD simulations—when run for tens of nanoseconds—are generally in good agreement with experiment (21,42–45), simulation times in this range turned out to be insufficient to quantify local preferential interaction coefficients (47).

Convergence of local preferential interaction coefficients from MD simulations is perceived as a formidable challenge because  $\Gamma_{XP}$  is intrinsically “a fluctuation difference and is prone to large levels of noise” (48) and because “dissection of the influence of different protein group contributions is not straightforward” (49). Key issues that hinder convergence are limitations in computational power and lacking insight on pertinent timescales of protein solvation. To this day, these issues continue to curtail the potential of computational studies of solvent effects on biomolecules, and molecular insight has been mostly limited to inferences from nonconverged protein simulations and simulations of smaller model compounds such as amino acids and peptides (49–53).

In this study, key issues that have curtailed the potential of computational studies of solvent effects on proteins are addressed. We demonstrate that the protein solvation mosaic can be characterized over the entire protein surface from MD simulations that are sufficiently longer than the longest intrinsic solvation timescale, which is  $\sim 10$  ns. Local preferential interaction coefficients are quantified for all protein residues of two proteins (lysozyme and antibody fragment

D1.3) in 4 M glycerol, and cosolvent effects on local protein conformations are predicted from changes in local preferential interactions between distinct protein side-chain conformations.

## METHODS

### Molecular-dynamics simulations

All-atom molecular simulations are performed for two proteins, Hen Egg-white Lysozyme (HEL) and the antibody fragment Fv D1.3 (54), which are explicitly solvated in a box of water with 4 M glycerol (see Table S1 in the Supporting Material). The CHARMM22 parameter set (55) is used to model protein atoms, water is modeled by the TIP3-model (56), and force-field parameters for glycerol are based on the carbohydrate hydrate parameters developed by Ha et al. (57). Crystallographic structures for HEL and D1.3 are taken from PDB:1VFB (54) and the setup of each simulation is carried out with CHARMM version c32b2 (58). Simulations of the solvated proteins are run for 1.3  $\mu$ s at constant pressure and temperature (1 atm and 298 K) with NAMD version 2.7 (59), as described previously in Vagenende et al. (47).

In a previous study, we pointed out that conformational changes of the backbone generally occur within nanoseconds along divergent trajectories and result in protein conformations with significantly different preferential interaction coefficients (42). Because this study aims to determine converged characteristics of the protein solvation mosaic, coordinates of the protein backbone are constrained with respect to the crystal structure. To investigate effects of side-chain motions, an additional simulation whereby all protein atom coordinates are constrained is performed for HEL (see Table S1). Unless indicated otherwise, the results in this study are based on the simulations with constrained backbone coordinates and free side-chain motions.

### Computation of preferential interaction coefficients

The global preferential interaction coefficient  $\Gamma_{XP}$  can be calculated from MD simulations by counting water and cosolvent molecules within a distance  $R$  from the protein surface (21,47):

$$\Gamma_{XP}(R) = \left\langle n_{XP}(r < R) - \frac{n_X - n_{XP}(r < R)}{n_W - n_{WP}(r < R)} n_{WP}(r < R) \right\rangle_{\tau_{run}} \quad (1)$$

In the above equation, brackets  $\langle \cdot \rangle_{\tau_{run}}$  refer to the time average over the entire simulation time  $\tau_{run}$ ;  $n_X$  and  $n_W$  are the total number of cosolvent and water molecules in the simulation box; and  $n_{XP}(r < R)$  and  $n_{WP}(r < R)$  are the number of cosolvent and water molecules for which the center of mass falls within a radial distance  $R$  from the protein van der Waals surface. The global preferential interaction coefficient  $\Gamma_{XP}(R)$  reaches a plateau from 5 Å onwards (see Fig. S1 in the Supporting Material). This indicates that preferential interactions of a protein in a mixture of water and glycerol are confined to the solvent region within 5 Å from the protein van der Waals surface.

Residue-based preferential interaction coefficients  $\Gamma_{XP}^i$  are calculated by assigning each solvent molecule within 5 Å from the protein surface to its closest protein residue  $i$  (21):

$$\Gamma_{XP}^i = \left\langle n_{XP}^i(r < 5 \text{ \AA}) - \frac{n_X - n_{XP}(r < 5 \text{ \AA})}{n_W - n_{WP}(r < 5 \text{ \AA})} n_{WP}^i(r < 5 \text{ \AA}) \right\rangle_{\tau_{run}} \quad (2)$$

In the above equation,  $n_{XP}^i$  ( $r < 5 \text{ \AA}$ ) and  $n_{WP}^i$  ( $r < 5 \text{ \AA}$ ) are the number of cosolvent and water molecules within  $5 \text{ \AA}$  from the protein surface that are closer to residue  $i$  than to any other residue of the protein. Residues with, on average, less than 1 water molecule and less than 0.1 glycerol molecule are considered to have limited solvent accessibility. Because each solvent molecule is assigned to only one residue, residue-based preferential interaction coefficients are additive and the following equation is automatically met:

$$\sum_i \Gamma_{XP}^i = \Gamma_{XP} \quad (3)$$

The additivity of residue-based preferential interaction coefficients implies that an independent change of  $\Gamma_{XP}^i$  for any of the protein residues results in an identical change of the global preferential interaction coefficient  $\Gamma_{XP}$  and therefore codetermines the effect of cosolvent on the chemical potential of the protein.

To quantify local solvation preferences of groups of residues, we calculate regional preferential interaction coefficients by assigning solvent molecules to a particular residue group if the residue that is closest to the solvent molecule belongs to that residue group. For a residue group consisting of residues  $[a, b, \dots]$ , we get:

$$\Gamma_{XP}^{[a,b,\dots]} = \left\langle \sum_{i \in [a,b,\dots]} n_{XP}^i(r < 5 \text{ \AA}) - \frac{n_X - n_{XP}(r < 5 \text{ \AA})}{n_W - n_{WP}(r < 5 \text{ \AA})} \times \sum_{i \in [a,b,\dots]} n_{WP}^i(r < 5 \text{ \AA}) \right\rangle_{\tau_{run}} \quad (4)$$

The prime symbol ( $'$ ) is used to refer to all protein residues excluding residues  $[a, b, \dots]$ , and we get

$$\Gamma_{XP}^{[a,b,\dots]} + \Gamma_{XP}^{[a,b,\dots]'} = \Gamma_{XP} \quad (5)$$

Although regional preferential interaction coefficients of a residue group are simply the sum of the residue-based preferential interaction coefficients of the residues of that group, they can be determined with higher precision because solvent fluctuations near neighboring residues are generally negatively correlated.

## Statistical analysis

Statistical analysis of any property  $A$  is performed based on plots of  $p(\tau_B)$  (47,60):

$$p(\tau_B) = \frac{\tau_B}{\tau_{rec}} \frac{\sigma^2 \langle A \rangle_{\tau_B}}{\sigma^2 \langle A \rangle_{\tau_{rec}}} \quad (6)$$

In the above equation,  $\tau_{rec}$  and  $\tau_B$  are the trajectory recording period and the variable block time, respectively, and  $\sigma^2 \langle A \rangle_{\tau_{rec}}$  and  $\sigma^2 \langle A \rangle_{\tau_B}$  are the variances calculated for the respective times. The block time  $\tau_B$  for which  $p(\tau_B)$  reaches a plateau indicates the correlation time beyond which block time averages  $\langle A \rangle_{\tau_B}$  become statistically independent (60). If  $p(\tau_B)$  remains at the plateau value for all block times, the standard error  $\sigma \langle A \rangle$  is estimated from the average plateau value. If this is not the case, an upper limit of  $\sigma \langle A \rangle$  is obtained from the average of  $\sigma \langle A \rangle_{\tau_B}$  for the four largest block times after it has been verified that  $\sigma \langle A \rangle_{\tau_B}$  continually decreases for increasing block times.

## Local concentration maps and characteristic residence times

Local concentrations are calculated based on the solvent occupancy of a three-dimensional grid with grid size  $1.5 \text{ \AA}$  and visualized with the software VMD 1.9 (61), as described previously in Vagenende et al. (47). Cutoff values are selected such that local concentration maps depict all solvent regions with local solvent concentrations  $c_\alpha(\vec{r})$  greater than the respective bulk solvent concentrations  $c_{\alpha, bulk}$ .

Residence times of solvent molecules near the protein surface ( $r < 5 \text{ \AA}$ ) and survival functions for glycerol and water, i.e.,  $N_{XP}(t)$  and  $N_{WP}(t)$ , are calculated as described previously (47). Characteristic residence times are obtained by fitting  $N_{XP}(t)$  and  $N_{WP}(t)$  to the following function:

$$N_{ZP}(t) \cong n_1 e^{-\frac{t}{\tau_1}} + n_2 e^{-\frac{t}{\tau_2}} + n_3 e^{-\frac{t}{\tau_3}} + c \quad (7)$$

Equation 7 allows for good fitting of  $N_{XP}(t)$  and  $N_{WP}(t)$  for both HEL and D1.3 (see Fig. S2).

To estimate the surface density of specific surface loci, protein solvent-accessible surface areas are calculated with VMD 1.9 (61) using a probe radius of  $1.4 \text{ \AA}$ . The average solvent-accessible surface areas of HEL and D1.3 are  $7248 \text{ \AA}^2$  and  $11,118 \text{ \AA}^2$ , respectively.

## RESULTS

### Global protein solvation

Molecular-dynamics simulations of two proteins, HEL and D1.3, in a mixture of water and 4 M glycerol are run for  $1.3 \mu\text{s}$  and subjected to extensive statistical analyses. For both proteins, the global preferential interaction coefficient  $\Gamma_{XP}$  has a distinctive correlation time of  $\sim 10 \text{ ns}$  (Fig. 1). This time corresponds with the longest characteristic residence time of glycerol (Table 1). Glycerol molecules with characteristic residence times of  $\sim 10 \text{ ns}$  reside at specific protein surface loci that form multiple hydrogen bonds with glycerol (47), such as near HEL-residues Lys<sup>13</sup> and Asp<sup>18</sup> (Fig. 2). This surface locus is occupied by glycerol for approximately half of the time, and, within a time period

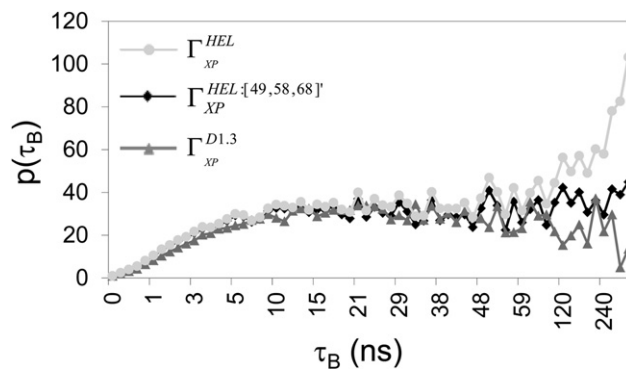


FIGURE 1 Statistical analysis of the global preferential interaction coefficient for HEL and D1.3 and the regional preferential interaction coefficient for all HEL residues except residue 49, 58, and 68. Values of  $p(\tau_B)$  reach a plateau at  $\sim 10 \text{ ns}$ , which indicates the correlation time of  $\Gamma_{XP}$ . For  $\Gamma_{XP}^{HEL}$  (but not for  $\Gamma_{XP}^{HEL-[49,58,68]'}$ ), values of  $p(\tau_B)$  increase from 100 ns onwards, indicating that  $\Gamma_{XP}$  for HEL is affected by local solvation changes near residues 48, 58, and 68 on timescales  $> 100 \text{ ns}$ .

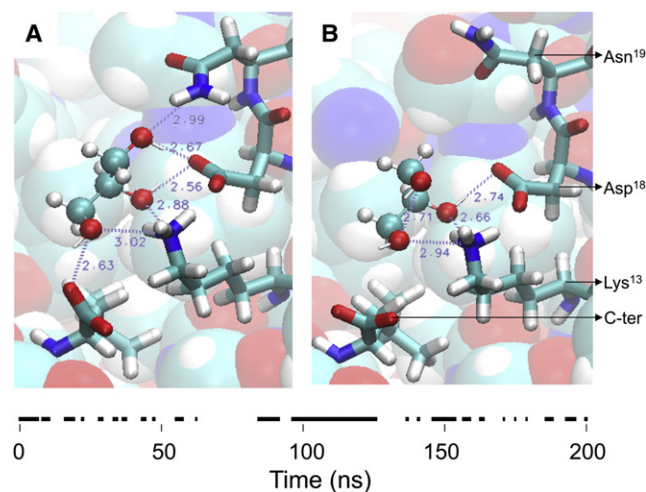
**TABLE 1** Characteristic residence times of water and glycerol molecules at the protein surface

	Protein	Constraint	$n_1$	$\tau_1$	$n_2$	$\tau_2$	$n_3$	$\tau_3$	$c$
			[–]	[ns]	[–]	[ns]	[–]	[ns]	[–]
Glycerol	D1.3	Backbone	21.5	0.3	91.7	1.4	10.3	8.6	0.8
	HEL	Backbone	15.1	0.3	66.1	1.3	2.3	10.3	0.9
	HEL	All	14.4	0.3	62.5	1.4	3.1	11.3	0.3
Water	D1.3	Backbone	1200.9	0.4	67.7	2.0	9.2	22.5	2.0
	HEL	Backbone	831.6	0.3	16.0	2.0	3.4	17.1	2.4
	HEL	All	781.5	0.3	30.6	2.0	5.3	8.7	3.5

Parameters obtained by fitting  $N_{XP}(t)$  and  $N_{WP}(t)$  according to Eq. 7.

of 200 ns, there are three instances whereby glycerol stays longer than 5 ns at this locus (Fig. 2 and see Movie S1 in the Supporting Material). Because each glycerol molecule increases  $\Gamma_{XP}$  by nearly one unit (Eq. 1), temporal changes in glycerol occupancy of such loci cause proportional fluctuations of  $\Gamma_{XP}$  on a timescale of  $\sim 10$  ns. We conclude therefore that the distinctive correlation time of  $\Gamma_{XP}$  at  $\sim 10$  ns (Fig. 1) results from temporal changes of solvent occupancy of protein surface loci that form multiple hydrogen bonds with glycerol.

If the slowest timescale of protein solvation is 10 ns, preferential interaction coefficients could be quantified with statistical significance based on MD simulations that are at least 10 times longer, i.e., 100 ns (44). Although this is the case for D1.3, block time averages of  $\Gamma_{XP}$  for HEL are correlated over time intervals exceeding 100 ns (Fig. 1). Long correlation times of  $\Gamma_{XP}$  for HEL are also reflected in the increasing number of glycerol molecules near the protein surface at simulation times  $>700$  ns (see Table S2 and Fig. S3). Nevertheless, variances  $\sigma\langle\Gamma_{XP}\rangle_{\tau_B}$  for HEL continue to decrease for larger block times (see Fig. S4) and an upper limit of the standard deviation of  $\Gamma_{XP}$  can be determined (see Table S1).

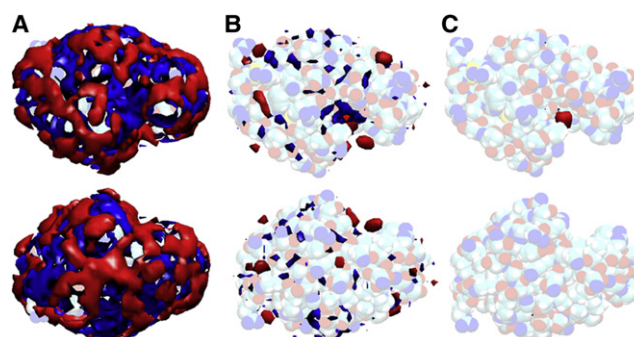


**FIGURE 2** Glycerol at the protein surface of HEL near Lys<sup>13</sup> and Asp<sup>18</sup> at 107 ns (A) and at 118 ns (B). (Bottom line) Times when this surface locus is occupied by glycerol.

## Local protein solvation

The protein solvation mosaic—which is the ensemble of preferentially solvated, preferentially hydrated and neutral solvent regions at the protein surface—can be visualized by local concentration maps. Unlike previous studies using shorter simulations (47), convergence of local concentration maps is obtained from 100 ns trajectories (see Fig. S5). Local concentration maps from microsecond simulations reveal that most solvent regions near the protein surface of HEL are either preferentially hydrated or preferentially solvated, and very few solvent regions do not have any preference for either solvent (Fig. 3 A). The majority of the mapped solvent regions rapidly vanish for increasing cutoff-values (Fig. 3, B and C), which indicates that solvent preferences of most protein surface loci are weak. A limited number of solvent regions have a significantly higher degree of preferential interactions (Fig. 3 B) and the highest degree of preferential interactions is found for glycerol in the catalytic binding pocket of HEL (Fig. 3 C). Similar observations are made for local concentration maps of D1.3 (see Fig. S6).

Local protein solvation is further characterized by quantifying local preferential interaction coefficients  $\Gamma_{XP}^i$  for all residues of HEL (Fig. 4 and see Fig. S7) and D1.3 (see Fig. S8). Local preferential interaction coefficients  $\Gamma_{XP}^i$  of solvent-accessible protein residues differ widely and range from significantly negative values (i.e., strong preferential hydration) to significantly positive values (i.e., strong preferential solvation by glycerol). For most protein residues, correlation times of  $\Gamma_{XP}^i$  are smaller than 100 ns, but for  $\sim 10\%$  of the residues, correlation times are significantly longer (Fig. 4 and see Fig. S8). Interestingly, slow convergence of  $\Gamma_{XP}$  for HEL is caused by local solvation changes near three of the 13 residues with slow convergence, i.e., Gly<sup>49</sup>, Ile<sup>58</sup>, and Arg<sup>68</sup> (Fig. 1). Moreover, slow local solvation changes disappear and  $\Gamma_{XP}$  converges in 10 ns when protein side-chain motions are constrained (see Fig. S9). We conclude, therefore, that slow solvation changes affecting  $\Gamma_{XP}$  of HEL on timescales exceeding 100 ns are related to protein side-chain motions near Gly<sup>49</sup>, Ile<sup>58</sup>, and Arg<sup>68</sup>.



**FIGURE 3** Local concentration maps for glycerol (red) and water (blue) at the surface of HEL (top and bottom view) for increasing cutoff values:  $c_\alpha(\vec{r}) > c_{\alpha, \text{bulk}}$  (A),  $c_\alpha(\vec{r}) > 1.5 c_{\alpha, \text{bulk}}$  (B), and  $c_\alpha(\vec{r}) > 3 c_{\alpha, \text{bulk}}$  (C).



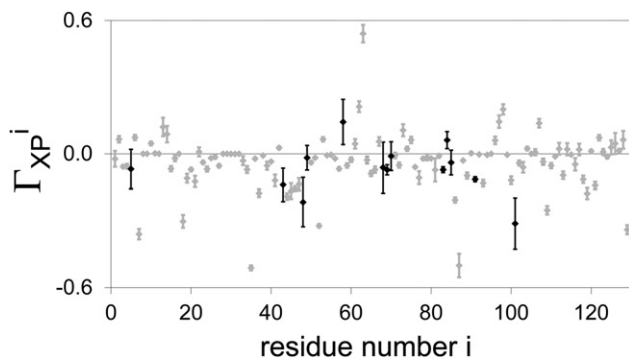


FIGURE 4 Residue-based preferential interaction coefficients  $\Gamma_{XP}^i$  for HEL. Data points for residues with correlation times  $<100$  ns (gray) and  $>100$  ns (black).

### Interdependence of protein solvation and side-chain motions

A single instance whereby side-chain motions affect protein solvation is observed near Gly<sup>49</sup> and Arg<sup>68</sup> of HEL. At the start of the simulation, the side chain of Arg<sup>68</sup> quickly changes from conformation A with its guanidinium-group contacting the carboxyl group of Gly<sup>49</sup> (Fig. 5 A) to conformation B with its guanidinium-group separated from Gly<sup>49</sup> by Arg<sup>45</sup> (Fig. 5 B). At 860 ns, Arg<sup>68</sup> switches again from conformation B to conformation A. This conformational change coincides with an increase of  $\Gamma_{XP}^i$  for both Arg<sup>68</sup> and Gly<sup>49</sup> (Fig. 5 C). Changes in protein solvation near these residues are also evident from local concentration maps: when Arg<sup>68</sup> adopts conformation A, solvent region 1 is preferentially hydrated and solvent region 2 is preferentially solvated by glycerol (Fig. 5 D), but when Arg<sup>68</sup> adopts conformation B, both solvent regions 1 and 2 are preferentially hydrated near the protein surface and preferentially solvated by glycerol further from the protein surface (Fig. 5 E). Thus, local preferential interaction coefficients and concentration maps concordantly confirm that side-chain motions of Arg<sup>68</sup> affect local protein solvation.

The sum of  $\Gamma_{XP}^i$  for Arg<sup>68</sup> and Gly<sup>49</sup> is significantly larger for Arg<sup>68</sup> conformation A ( $0.17 \pm 0.06$ ) than for conformation B ( $-0.23 \pm 0.07$ ). Based on the thermodynamic principles of preferential interactions applied to aqueous glycerol solutions (42), a difference of  $\Gamma_{XP}^i$  between conformation A and B of 0.4 results in a glycerol-induced shift of the relative free energies of conformation A and B in favor of conformation A by  $\sim 1$  kJ/mol. This prediction is in good agreement with the range of free energy shifts between local protein conformations in aqueous glycerol solutions derived from hydrogen-exchange experiments (62).

Similar to Gly<sup>49</sup> and Arg<sup>68</sup>, local preferential interaction coefficients of 10 other residues of HEL have long correlation times ( $>100$  ns) because of slow side-chain motions (Fig. 4). However, unlike Gly<sup>48</sup> and Arg<sup>68</sup>, side-chain motions near these residues only affect the proximity to distinct solvent regions without affecting local protein

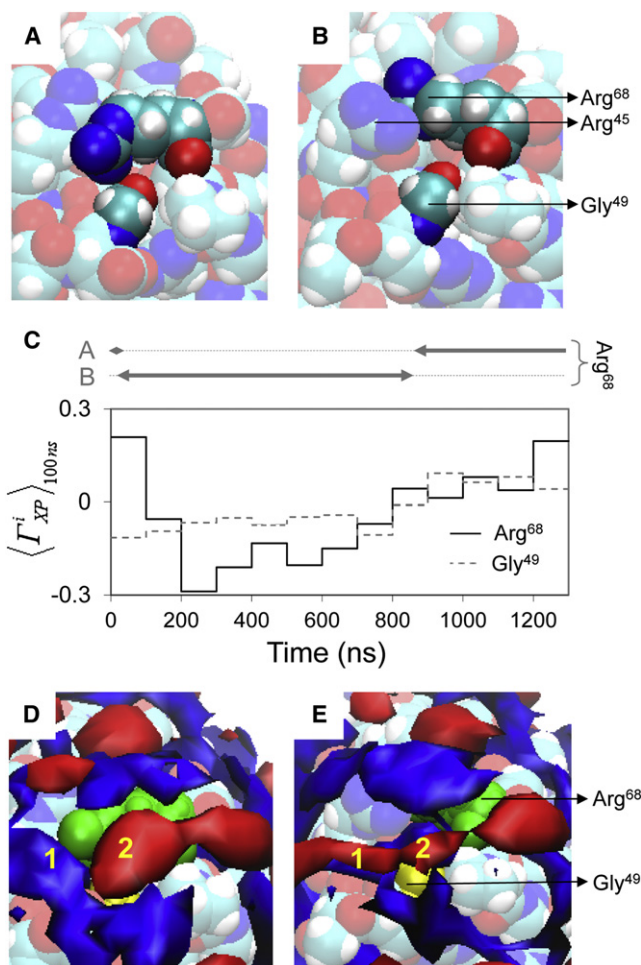


FIGURE 5 Arg<sup>68</sup> side-chain conformation A and B (A and B, respectively), local preferential interaction coefficients for Gly<sup>49</sup> and Arg<sup>68</sup> (C), and local concentration maps for glycerol (red) and water (blue) near Arg<sup>68</sup> conformation A and B (D and E, respectively). In panels D and E, distinct solvent regions near Gly<sup>49</sup> (yellow) and Arg<sup>68</sup> (green) are numbered.

solvation. The corresponding changes of  $\Gamma_{XP}^i$  values of these residues are consequently such that their overall contributions to the global preferential interaction coefficient  $\Gamma_{XP}$  cancel out (Eq. 3). That most side-chain motions do not considerably affect protein solvation is also evidenced by the limited effects of constraining side-chain motions on 1), characteristic residence times of solvent at the protein surface (Table 1); 2), the global preferential interaction coefficient (see Table S1); and 3), local concentration maps (see Fig. S10).

### Solvation of a protein-binding pocket

A unique case of local protein solvation is observed in the catalytic binding pocket of HEL near Ile<sup>58</sup>. When protein side chains are constrained, this binding pocket contains only water. However, for the simulation of HEL with unconstrained side-chain coordinates, a glycerol molecule

enters the binding pocket at 80 ns and it remains there for  $\sim 500$  ns before leaving the binding pocket (Fig. 6). Consecutively, several other glycerol molecules populate the binding pocket with residence times of  $\sim 100$  ns, and the binding pocket is occupied by glycerol for  $>90\%$  of the simulation time (Fig. 6 C, and see Movie S2). Glycerol in the binding pocket adopts two major binding orientations: orientation A is such that its O-atoms are in contact with the protein and its C-atoms point toward the solvent (Fig. 6 A), whereas for orientation B the direction of O- and C-atoms is reversed (Fig. 6 B). The switch from glycerol orientation A to orientation B around the middle of the simulation coincides with the increase of residue-based preferential interaction coefficient  $\Gamma_{XP}^i$  for Ile<sup>58</sup> (Fig. 6 C). This indicates that local solvation preferences of a protein can be affected by slow ( $>100$  ns) orientational changes of cosolvent molecules at unique binding pockets.

## DISCUSSION

In this study we have characterized the protein solvation mosaic—i.e., the ensemble of preferentially hydrated, preferentially solvated, and neutral solvent regions—over the entire surface of a specific protein conformation in a mixed solvent. This was achieved by performing extended (1.3  $\mu$ s) classical all-atom MD simulations of a protein in a mixed solvent with constrained protein coordinates. We found that correlation times of preferential interaction coefficients are governed by the longest characteristic residence time of glycerol of  $\sim 10$  ns, which arises from glycerol molecules

forming multiple hydrogen bonds with specific protein surface loci. Solvation timescales of  $\sim 10$  ns were also reported for proteins in mixtures of water and other cosolvents such as urea (63) and arginine (43). This suggests that the slowest intrinsic timescale for protein solvation in mixed solvents is generally  $\sim 10$  ns. As a corollary, simulations need to be sufficiently longer than 10 ns to obtain multiple statistically independent block-time averages of  $\Gamma_{XP}$ , and we find that a minimum simulation time of  $\sim 100$  ns is needed to quantify global and local preferential interactions of a specific protein conformation with statistical significance.

Interestingly, Ma et al. (51) found that the minimum simulation time to obtain convergence of  $\Gamma_{XP}$  for a triglycine peptide in aqueous urea solutions was also  $\sim 100$  ns. These authors used multiple 2–3 ns simulations, whereas we used a single extended simulation per protein. Although both simulation schemes appear equally effective for quantifying local preferential interaction coefficients and equally efficient in their use of computational resources, multiple shorter simulation times can arguably be completed in a shorter time. This apparent advantage loses much of its significance when one considers that a 100 ns simulation of a medium-size protein ( $\sim 25$  kDa) in a solvent box only takes  $\sim 5$  days on a standard high-performance computing cluster with  $\sim 100$  CPUs. However, an important difference between both simulation schemes is that only continuous extended simulations allow characterizing residence of solvent molecules at the protein surface. This is a considerable advantage because characteristic residence times determine convergence times of preferential interactions and reveal pertinent protein solvation characteristics (42,47).

Specific protein-surface loci forming multiple hydrogen bonds with the cosolvent have a surprisingly low affinity for cosolvent. For example, a specific protein surface locus that forms up to six hydrogen bonds with a single glycerol molecule is only occupied by glycerol for half of the time (Fig. 2). Assuming a Langmuir binding model, this surface locus binds glycerol with a dissociation constant  $K_d$  of  $\sim 4$  M. Thus, multiple hydrogen-bonding at this surface-locus results in specific glycerol orientations without pronounced increases of the binding affinity. This is in agreement with empirical observations that hydrogen bonds in molecular recognition processes primarily convey specificity rather than affinity (33). Taking into account the average number of glycerol molecules with long characteristic residence times (i.e.,  $n_3$  in Table 1), the solvent-accessible surface areas of the simulated proteins, and an average locus occupancy by glycerol of 50%, we find that the surface density of specific loci that form multiple hydrogen bonds with glycerol is typically in the range  $0.5\text{--}2.0 \times 10^{-3}/\text{\AA}^2$ . This corresponds with several specific surface loci for smaller proteins and tens of specific surface loci for larger proteins. Specific surface loci forming multiple hydrogen bonds with a cosolvent molecule will therefore significantly contribute to  $\Gamma_{XP}$ , and local

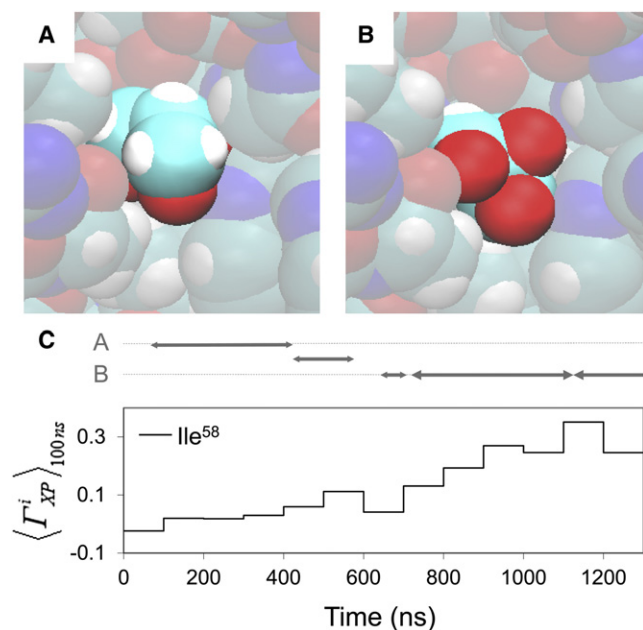


FIGURE 6 Major glycerol orientations in the catalytic pocket of HEL (A and B) and corresponding local preferential interaction coefficients for Ile<sup>58</sup> (C).

solvation preferences at such loci are expected to affect overall cosolvent effects on protein processes.

Most protein surface loci have a similar affinity for glycerol as for water (Fig. 3). However, several protein surface loci, including loci forming multiple hydrogen bonds with glycerol, have a higher affinity for glycerol with dissociation constants  $K_d$  as low as 2 M. By far the greatest binding affinity for glycerol is found in the catalytic binding pocket of HEL, which is occupied by glycerol for ~90% of the time (Fig. 6 C). The high occupancy of the binding pocket by glycerol, which is in good agreement with crystal structures of HEL that resolve glycerol in the binding pocket (64), corresponds with a dissociation constant  $K_d$  of ~0.4 M. The clear distinction in timescales and binding affinities between the catalytic binding pocket of HEL and the rest of the protein surface prompts us to differentiate between general protein solvation, which is determined by low binding affinities ( $K_d > 2$  M) and characteristic solvent residence times  $< 20$  ns, versus high-affinity protein solvation, which is determined by substantially higher binding affinities ( $K_d < 0.5$  M) and residence times  $> 100$  ns.

Despite the high binding-affinity of the catalytic binding pocket of HEL, glycerol only enters this pocket for the first time after 80 ns. This clearly demonstrates that shorter simulation times and simulations that fail to demonstrate convergence behavior of local preferential interactions may lead to erroneous conclusions on local solvation preferences of proteins. Hence, careful evaluation is needed of local protein solvation data from shorter simulations or simulations during which the protein undergoes large conformation changes. This is especially critical when using classical MD simulations to identify binding pockets and estimate the affinity of druglike molecules (65).

After entering the binding pocket, glycerol predominantly adapts two distinct binding orientations and repeatedly undergoes orientational changes on timescales in the order of ~100 ns. Such slow orientational changes obviously do not converge within microsecond classical molecular dynamics simulations and other computational methods such as 3D-RISM based methods (66) could be more appropriate to characterize solvent orientations in high-affinity binding pockets. By constraining cosolvent molecules at high-affinity sites with respect to predominant binding orientations, local solvation preferences of the binding pocket could then be characterized by classical MD simulations. In this manner, the structural and dynamical properties of the protein solvation mosaic could be determined over the entire protein surface, even near high-affinity pockets. Alternatively, if the protein process of interest does not involve the binding pocket, high-affinity binding pockets could be excluded by considering regional preferential interaction coefficients of all protein residues outside of the binding pockets.

Local preferential interaction coefficients may differ for protein residues of the same amino-acid type. For example,

Lys<sup>13</sup> of HEL is preferentially solvated by glycerol ( $\Gamma_{XP}^i = 0.12 \pm 0.04$ ) whereas Lys<sup>116</sup> of HEL is preferentially hydrated ( $\Gamma_{XP}^i = -0.05 \pm 0.03$ ) (Fig. 4). Local preferential interaction coefficients of these two residues differ because of adjacent residues: Lys<sup>13</sup> and its adjacent residues cooperatively bind glycerol (Fig. 2), whereas protein residues near Lys<sup>116</sup> do not cooperatively interact with the solvent. Cooperative interactions of adjacent protein groups with solvent molecules depend on the size and nature of the constituent groups that partition the protein surface. Although smaller constituent groups allow characterization of local protein solvation at higher resolution, local preferential interaction coefficients of smaller protein constituent groups become increasingly dependent on solvent interactions with adjacent groups and information on solvent interactions of individual constituent groups becomes more convoluted.

Results from our simulations of two proteins (lysozyme and antibody fragment D1.3) in 4 M glycerol agree well with available experimental data on 1), characteristic residence times of solvent molecules at the protein surface; 2), the high occupancy of the catalytic cleft of HEL by glycerol; and 3), the glycerol-induced shift of the free energy between local protein conformations. Previously, we also demonstrated quantitative agreement of  $\Gamma_{XP}$  values for lysozyme in aqueous glycerol between simulation and experiment (42). This indicates that force fields used in this study are appropriate for characterizing local protein solvation in aqueous glycerol solutions.

## CONCLUSIONS

We have shown that local solvation preferences can be quantified over the entire protein surface from extended MD simulations of specific protein conformations in mixed solvents. This has been achieved by combining analysis methods with gradually increasing resolution, starting with the statistical analysis of the global preferential interaction coefficient  $\Gamma_{XP}$  and the identification of protein solvation timescales, followed by the characterization of local concentration maps and residue-base preferential interaction coefficients, and complemented with the inspection of the trajectories of specific solvent molecules and protein surface loci. This methodology allows quantifying the contribution of individual solvent molecules and protein residues to  $\Gamma_{XP}$  and therefore expands the scope of preferential interaction theory by directly linking atom-level details of local protein solvation with thermodynamic solvent effects on protein processes.

From the statistical analysis of solvent fluctuations at the protein surface, we derived that the slowest intrinsic solvation timescale of a protein in a mixture of water and cosolvent is determined by long residence times of cosolvent molecules at specific protein surface loci that enable multiple hydrogen-bonding with a cosolvent molecule. The slowest intrinsic solvation timescale is generally ~10 ns, and simulation times



to quantify local solvation preferences of a constrained protein conformation need to be longer than 100 ns. Most protein side-chain motions do not significantly affect protein solvation, although slow (>100 ns) side-chain motions of a single protein residue can lead to significant changes of local solvent preferences. Another singular event that affects local solvation preferences is the slow reorientation of cosolvent molecules at protein surface loci with high cosolvent affinity, such as observed for glycerol in the catalytic binding pocket of lysozyme.

Because convergence of local solvent preferences of proteins in mixed solvents can be achieved on standard high-performance clusters, computational characterization of local protein solvation emerges as an accessible high-resolution technique for studying solvent effects on proteins. Granted the availability of accurate force fields, this technique enables direct identification of physicochemical properties that determine solvent preferences without assuming group additivity. Even for a relatively simple cosolvent like glycerol, solvent preferences are remarkably heterogeneous at distinct protein residues. Detailed characterization of local protein solvation, therefore, appears indispensable to further understanding of the molecular mechanisms by which solvents affect protein structure.

## SUPPORTING MATERIAL

Two tables, 10 figures, and two movies are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(12\)00872-7](http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00872-7).

This research was supported in part by the National Science Foundation through TeraGrid resources provided by Texas Advanced Computing Centre under grant No. TG-MCB100058, and by the Biomedical Research Council of A\*STAR, Singapore.

## REFERENCES

- Cohen, F. E., and J. W. Kelly. 2003. Therapeutic approaches to protein-misfolding diseases. *Nature*. 426:905–909.
- Kamerzell, T. J., R. Esfandiary, ..., D. B. Volkin. 2011. Protein-excipient interactions: mechanisms and biophysical characterization applied to protein formulation development. *Adv. Drug Deliv. Rev.* 63:1118–1159.
- Ohtake, S., Y. Kita, and T. Arakawa. 2011. Interactions of formulation excipients with proteins in solution and in the dried state. *Adv. Drug Deliv. Rev.* 63:1053–1073.
- Timasheff, S. N. 1998. Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. *Adv. Protein Chem.* 51:355–432.
- López, C. J., M. R. Fleissner, ..., W. L. Hubbell. 2009. Osmolyte perturbation reveals conformational equilibria in spin-labeled proteins. *Protein Sci.* 18:1637–1652.
- Flores Jiménez, R. H., M.-A. Do Cao, ..., D. S. Cafiso. 2010. Osmolytes modulate conformational exchange in solvent-exposed regions of membrane proteins. *Protein Sci.* 19:269–278.
- Kirkwood, J. G., and F. P. Buff. 1951. The statistical mechanical theory of solutions. I. *J. Chem. Phys.* 19:774–777.
- Casassa, E. F., and H. Eisenberg. 1964. Thermodynamic analysis of multicomponent solutions. *Adv. Protein Chem.* 19:287–395.
- Wyman, Jr., J. 1964. Linked functions and reciprocal effect in hemoglobin: a second look. *Adv. Protein Chem.* 19:223–286.
- Tanford, C. 1969. Extension of the theory of linked functions to incorporate the effects of protein hydration. *J. Mol. Biol.* 39:539–544.
- Schellman, J. A. 1978. Solvent denaturation. *Biopolymers.* 17:1305–1322.
- Smith, P. E. 2004. Cosolvent interactions with biomolecules: relating computer simulation data to experimental thermodynamic data. *J. Phys. Chem. B.* 108:18716–18724.
- Shimizu, S., and C. L. Boon. 2004. The Kirkwood-Buff theory and the effect of cosolvents on biochemical reactions. *J. Chem. Phys.* 121: 9147–9155.
- Schurr, J. M., D. P. Rangel, and S. R. Aragon. 2005. A contribution to the theory of preferential interaction coefficients. *Biophys. J.* 89:2258–2276.
- Shulgin, I. L., and E. Ruckenstein. 2005. A protein molecule in an aqueous mixed solvent: fluctuation theory outlook. *J. Chem. Phys.* 123:054909.
- Shimizu, S., and N. Matubayasi. 2006. Preferential hydration of proteins: a Kirkwood-Buff approach. *Chem. Phys. Lett.* 420:518–522.
- Smith, P. E. 2006. Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophys. J.* 91:849–856.
- Kang, M., and P. E. Smith. 2007. Preferential interaction parameters in biological systems by Kirkwood-Buff theory and computer simulation. *Fluid Phase Equilib.* 256:14–19.
- Jiao, Y., and P. E. Smith. 2011. Fluctuation theory of molecular association and conformational equilibria. *J. Chem. Phys.* 135:014502.
- Record, Jr., M. T., W. T. Zhang, and C. F. Anderson. 1998. 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. Protein Chem.* 51:281–353.
- Baynes, B. M., and B. L. Trout. 2003. Proteins in mixed solvents: a molecular-level perspective. *J. Phys. Chem. B.* 107:14058–14067.
- Tanford, C. 1964. Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* 86:2050–2059.
- Chen, X., L. B. Sagle, and P. S. Cremer. 2007. Urea orientation at protein surfaces. *J. Am. Chem. Soc.* 129:15104–15105.
- Harries, D., and J. Rosgen. 2008. A practical guide on how osmolytes modulate macromolecular properties. In *Biophysical Tools for Biologists*, Vol. 1: In Vitro Techniques. Methods in Cell Biology Series, Vol. 84. Academic Press, Elsevier, London, Amsterdam, Oxford, Burlington, San Diego. 679–735.
- Hilser, V. J. 2011. Structural biology: finding the wet spots. *Nature*. 469:166–167.
- 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.
- Schneider, C. P., and B. L. Trout. 2009. Investigation of cosolute-protein preferential interaction coefficients: new insight into the mechanism by which arginine inhibits aggregation. *J. Phys. Chem. B.* 113: 2050–2058.
- Auton, M., and D. W. Bolen. 2005. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. USA.* 102:15065–15068.
- O’Brien, E. P., G. Ziv, ..., D. Thirumalai. 2008. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. USA.* 105:13403–13408.
- Guinn, E. J., L. M. Pegram, ..., M. T. Record, Jr. 2011. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc. Natl. Acad. Sci. USA.* 108:16932–16937.



31. 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.
32. Roseman, M. A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* 200: 513–522.
33. Bissantz, C., B. Kuhn, and M. Stahl. 2010. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 53:5061–5084.
34. Bennion, B. J., and V. Daggett. 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA.* 100:5142–5147.
35. Das, A., and C. Mukhopadhyay. 2008. Atomistic mechanism of protein denaturation by urea. *J. Phys. Chem. B.* 112:7903–7908.
36. Stumpe, M. C., and H. Grubmüller. 2008. Polar or apolar—the role of polarity for urea-induced protein denaturation. *PLOS Comput. Biol.* 4:e1000221.
37. Stumpe, M. C., and H. Grubmüller. 2009. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophys. J.* 96:3744–3752.
38. 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.
39. Canchi, D. R., D. Paschek, and A. E. García. 2010. Equilibrium study of protein denaturation by urea. *J. Am. Chem. Soc.* 132:2338–2344.
40. Saladino, G., S. Pieraccini, ..., M. Sironi. 2011. Metadynamics study of a  $\beta$ -hairpin stability in mixed solvents. *J. Am. Chem. Soc.* 133:2897–2903.
41. Berteotti, A., A. Barducci, and M. Parrinello. 2011. Effect of urea on the  $\beta$ -hairpin conformational ensemble and protein denaturation mechanism. *J. Am. Chem. Soc.* 133:17200–17206.
42. Vagenende, V., M. G. S. Yap, and B. L. Trout. 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry.* 48:11084–11096.
43. Shukla, D., and B. L. Trout. 2011. Preferential interaction coefficients of proteins in aqueous arginine solutions and their molecular origins. *J. Phys. Chem. B.* 115:1243–1253.
44. Shukla, D., C. P. Schneider, and B. L. Trout. 2011. Complex interactions between molecular ions in solution and their effect on protein stability. *J. Am. Chem. Soc.* 133:18713–18718.
45. Schneider, C. P., D. Shukla, and B. L. Trout. 2011. Effects of solute-solute interactions on protein stability studied using various counterions and dendrimers. *PLoS One.* 6:e27665.
46. Canchi, D. R., and A. E. García. 2011. Backbone and side-chain contributions in protein denaturation by urea. *Biophys. J.* 100:1526–1533.
47. Vagenende, V., M. G. S. Yap, and B. L. Trout. 2009. Molecular anatomy of preferential interaction coefficients by elucidating protein solvation in mixed solvents: methodology and application for lysozyme in aqueous glycerol. *J. Phys. Chem. B.* 113:11743–11753.
48. Hu, C. Y., H. Kokubo, ..., B. M. Pettitt. 2010. Backbone additivity in the transfer model of protein solvation. *Protein Sci.* 19:1011–1022.
49. Horinek, D., and R. R. Netz. 2011. Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. *J. Phys. Chem. A.* 115: 6125–6136.
50. Roccatano, D. 2008. Computer simulations study of biomolecules in non-aqueous or cosolvent/water mixture solutions. *Curr. Protein Pept. Sci.* 9:407–426.
51. Ma, L., L. Pegram, ..., Q. Cui. 2010. Preferential interactions between small solutes and the protein backbone: a computational analysis. *Biochemistry.* 49:1954–1962.
52. Kokubo, H., C. Y. Hu, and B. M. Pettitt. 2011. Peptide conformational preferences in osmolyte solutions: transfer free energies of decaalanine. *J. Am. Chem. Soc.* 133:1849–1858.
53. 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.
54. Bhat, T. N., G. A. Bentley, ..., R. J. Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA.* 91:1089–1093.
55. MacKerell, A. D., D. Bashford, ..., M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
56. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
57. Ha, S. N., A. Giammona, ..., J. W. Brady. 1988. A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr. Res.* 180:207–221.
58. Brooks, B. R., R. E. Bruccoleri, ..., M. Karplus. 1983. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
59. Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
60. Allen, M. P., and D. J. Tildesley. 1987. *Computer Simulation of Liquids.* Clarendon Press, Oxford, UK.
61. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
62. Gregory, R. B. 1988. The influence of glycerol on hydrogen isotope exchange in lysozyme. *Biopolymers.* 27:1699–1709.
63. Lindgren, M., T. Sparrman, and P.-O. Westlund. 2010. A combined molecular dynamic simulation and urea  $^{14}\text{N}$  NMR relaxation study of the urea-lysozyme system. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 75:953–959.
64. Ueno, T., S. Abe, ..., Y. Watanabe. 2010. Elucidation of metal-ion accumulation induced by hydrogen bonds on protein surfaces by using porous lysozyme crystals containing Rh(III) ions as the model surfaces. *Chemistry.* 16:2730–2740.
65. Seco, J., F. J. Luque, and X. Barril. 2009. Binding site detection and druggability index from first principles. *J. Med. Chem.* 52:2363–2371.
66. Imai, T., K. Oda, ..., A. Kidera. 2009. Ligand mapping on protein surfaces by the 3D-RISM theory: toward computational fragment-based drug design. *J. Am. Chem. Soc.* 131:12430–12440.