A Transfer Free Energy Based Implicit Solvent Model for Protein Simulations in Solvent Mixtures: Urea-Induced Denaturation as a Case Study

Andrea Arsiccio,[†] Pritam Ganguly,[†] and Joan-Emma Shea[*,†,‡]

[†]*Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States*

[‡]*Department of Physics, University of California, Santa Barbara, California 93106, United States*

E-mail: shea@ucsb.edu

Abstract

We developed a method for implicit solvent molecular dynamics simulations of proteins in solvent mixtures (Model with Implicit Solvation Thermodynamics, MIST). The MIST method introduces experimental group transfer free energies to the Generalized Born formulation for generating molecular trajectories without the need for developing rigorous explicit-solvent force fields for multicomponent solutions. As a test case, we studied the urea-induced denaturation of Trp-cage miniprotein in water. We demonstrate that our method allows efficient exploration of the conformational space of the protein in only a few hundreds of nanoseconds of all-atom unbiased simulations. Furthermore, selective implementation of the transfer free energies of specific peptide groups, backbone, and side chains enables us to decouple their specific energetic contributions to the conformational changes of the protein. The approach herein developed can readily be extended to the investigation of complex matrices, as well as to the characterization of protein aggregation. The MIST method is implemented in Plumed (https://www.plumed.org/, from version 2.8) as a separate module called SASA.

Introduction

The cellular medium in which biomolecules reside consists of a complex aqueous solution of bioinorganic and organic cosolutes such as salts, sugars, amines or polyols. Understanding the effects of these molecules on protein structures is critical in order to understand protein stability and function in the cell. Molecular dynamics (MD) simulations offer an extremely powerful and effective tool for probing the structural stability of proteins in the cellular matrix. Incorporating an explicit description of the solvent environment in the simulation has the advantage of directly accounting for solvent entropy, however, the computational power required for a meaningful sampling of biologically-relevant phenomena in these conditions is often too high, making this approach impractical in many situations. In addition, creating mixed solutions with water and cosolutes requires the challenging task of parame-

terizing multicomponent systems.[1–5] In this regard, implicit solvent simulations have several advantages. First, the number of interactions to be described is reduced, making the calculations faster. Second, the absence of viscosity in implicit solvent simulations accelerates the exploration of the conformational phase space allowing, for instance, a faster observation of folding-unfolding processes. Finally, the removal of the slow degrees of freedom involving the solvent dynamics further accelerates the sampling, making this approach extremely competitive in terms of computational cost.[6,7]

We already worked on the inclusion of temperature[8] and pressure[9] effects in implicit solvent simulations of proteins. We here develop and test a novel implicit solvent method for the description of proteins behavior in common biological matrices. Specifically, we study the behavior of proteins in binary water-osmolyte mixtures as case study, although the approach could in principle be extended to more complex solvent environments.

Organic osmolytes, also known as cosolutes, are commonly accumulated in diverse taxa.[10,11] They can be classified into four main categories: (1) polyols and sugars, such as sorbitol, glycerol, sucrose, trehalose; (2) amino acids, including proline, glutamate, alanine and derivatives like taurine; (3) methylammonium and methylsulfonium solutes, such as trimethylamine-N-oxide (TMAO); and, finally, (4) urea. A common feature of all of these molecules is their polarity that makes them highly soluble in water. With the notable exception of urea, they are 'compatible solutes', in the sense that do not show perturbing effects on cellular macromolecules.[12] Some stabilizing osmolytes can even counteract the denaturing effect of urea.[12] These features make cosolutes of interest for applications in biotechnology, agriculture, pharmacy and medicine.[13] Their interaction with proteins, and potential ability to prevent or minimize human diseases related with misfolding, unfolding or aggregation, is also a subject of intense investigation.

To describe the effect of osmolytes on proteins within an implicit solvent model, we start from the seminal work by Bolen and coworkers, who introduced an additive approach based on group transfer free energies (GTFEs). This approach allowed them to evaluate the free

energy change upon unfolding/refolding in terms of the separate backbone and sidechains contributions, weighed by their respective solvent accessible surface area (SASA).[14–16]

Specifically, this approach is based on the evaluation of the free energies associated with the transfer of the separate sidechains ($g^{tr}_{k,sc}$), or of the backbone group (g^{tr}_{bb}), from water to a given osmolyte solution (at concentration c). Once these values have been determined, the global free energy cost required for transferring a whole protein molecule containing n_r residues between water and the osmolyte solution ($G^{tr,0\rightarrow c}$) can be computed by summing over all the surface exposed groups in the protein structure,[14,17]

$$G^{tr,0\rightarrow c} = \sum_{k=1}^{n_r} g^{tr}_{k,sc}\alpha_{k,sc} + g^{tr}_{bb}\sum_{k=1}^{n_r}\alpha_{k,bb} \tag{1}$$

Each contribution is weighed by the fractional solvent accessible surface area $SASA_k$ of residue k,

$$\alpha_k = \frac{SASA_k}{SASA_{k,Gly-X-Gly}} \tag{2}$$

where $SASA_{k,Gly-X-Gly}$ is the solvent accessibility of amino acid X in the tripeptide Gly-X-Gly, and X is the amino acid residue type k.

The difference in transfer free energy between the native N ($G^{tr,0\rightarrow c}_N$) and unfolded U ($G^{tr,0\rightarrow c}_U$) protein can then be related to the change in the free energy of unfolding ($G^{N\rightarrow U}_c$ - $G^{N\rightarrow U}_0$) associated with the presence of the osmolyte, according to the transfer model shown in Figure 1.

The additive construction proposed by Bolen and coworkers has been utilized in molecular dynamics simulations by Thirumalai and collaborators in the framework of the Molecular Transfer Model (MTM).[18–22] The MTM framework, in conjunction with coarse-grained or all-atom Go-models for proteins, has successfully been applied to mimic structural and thermodynamical properties of both globular and disordered proteins at varying solution conditions such as cosolute concentration, pH and temperature.[23] The MTM procedure has

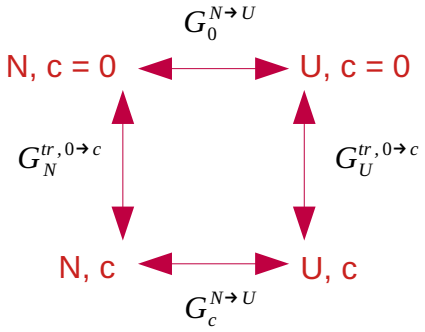

Figure 1: Schematic representation of the transfer model.

been shown to be formally exact if the conformations at the desired level of perturbation are exhaustively sampled,[21] but it relies on significant overlapping between the protein conformational spaces explored in pure water and in solutions of the cosolute. However, even using enhanced sampling methods, it is extremely challenging to explore the entire protein conformational space for finite cosolute concentrations from simulations in pure water without the cosolutes.[24,25] It is also important to note that the native folded conformations may be different in water and in osmolyte solutions, as observed for instance in the case of TMAO for Staphylococcal Nuclease fragments using Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR) spectroscopy[26] and for cyclic dipeptides using Dynamic Light Scattering (DLS) measurements.[27] This poor overlap between the protein conformational spaces in water and cosolute solutions, observed experimentally for folded proteins, may be even more pronounced in the case of disordered peptides.

Here, we address this issue by introducing a novel computational scheme to study the effect of osmolytes on protein stability, which allows the direct generation of trajectories at a given cosolute concentration. For this purpose, we make use of the additive principle proposed by Bolen and coworkers and couple experimentally determined GTFEs into an implicit solvent framework. We combine the GTFE-based term for the osmolytes with a Generalized-Born (GB) method to model electrostatic interactions,[28] and the ACE(analytical continuum electrostatic) approximation[29] to describe the non-polar solvation energy in water.

To the best of our knowledge, this represents the first attempt to describe binary osmolyte-water mixtures in implicit solvent simulations.

The proposed approach, that we refer to as Model with Implicit Solvation Thermodynamics (MIST), represents therefore a novel strategy, that considerably expands the possibilities of implicit solvation in describing biologically-relevant systems and has several advantages. First, the influence of osmolytes on protein behavior is completely described, in our simulations, by means of experimentally-derived free energies of transfer. This eliminates the need to derive force fields for the cosolutes to be studied, which would require a huge number of parameters to be adjusted to properly fit the experimental data. In addition, the additive nature of the GTFEs approach herein used makes it easy to separately investigate the different contributions involved in protein-osmolyte interaction, a critical aspect in order to understand the mechanism of action of common cosolutes. For instance, both the backbone[30,31] and the sidechains contributions[32–34] have been suggested to be key in driving osmolyte-induced conformational transitions of proteins, and the implicit solvation approach herein described makes it possible to clarify this point.

We focus our discussion here on the denaturant urea, which is perhaps experimentally the best studied osmolyte. Urea can disrupt noncovalent bonds in proteins, and for this reason it is used extensively to assess protein stability, as well as the effects of mutations on stability.[35] The urea-induced denaturation of proteins has been the subject of intense investigation,[25,36–38] but the exact mechanism behind its action is still under debate.

We emphasize that the method presented in this paper can be extended to any cosolute for which experimental transfer free energy data is available. Moreover, it may also be straightforwardly extended to more complex mixtures, provided that the free energy of transfer between water and such mixtures is known.

In the case of urea, we consider two different sets of transfer free energies, an original set proposed by Auton, Holthauzen and Bolen[30] (urea - B) and a modified set introduced by Moeser and Horinek[39] (urea - H). The two data sets differ for two reasons. First, Bolen and

coworkers assigned the backbone contribution g_{bb}^{tr} to the backbone accessible surface area of each amino acid type X in the respective Gly-X-Gly sequence. In contrast, Moeser and Horinek suggested a universal backbone approach, where the backbone transfer free energy g_{bb}^{tr} is assigned, for each amino acid type, to the accessible surface area of the central glycine in the tripeptide Gly-Gly-Gly. Second, Auton, Holthauzen and Bolen took the activity coefficient of glycine in water and urea into account when deriving their transfer free energies for the side chains, and this translates into mostly positive values of $g_{k,sc}^{tr}$ in the urea - B description. However, a mistake was made in the conversion of activity coefficient data between concentration scales. Moeser and Horinek corrected for this mistake in their new set of transfer free energy values, and, as a result, most side chain contributions $g_{k,sc}^{tr}$ are negative in the urea - H description.

We will show that the mechanism of urea-induced unfolding of Trp-cage changes depending on the set of transfer free energies considered, with the Horinek set predicting a comparable contribution of backbone and sidechain groups, which results in a more realistic description of the unfolding process. This last point is particularly important in order to understand the mechanism of action of common cosolutes. Indeed, by looking at the two sets of GTFEs for urea, we can understand how different interactions between the protein sidechains/backbone groups and the denaturing osmolyte result in different denaturation pathways, and these pathways find an automatic correspondence in the trajectories generated within our implicit solvent simulations.

We will first provide some theoretical background information, and introduce the MIST approach herein proposed. The computational details of the simulations performed will then be described. Finally, we will show an example of application, focusing for this purpose on the α-helical Trp-cage (NLYIQWLKDGGPSSGRPPPS, pdb 1L2Y[40]). Trp-cage has been extensively characterized both computationally and experimentally, shows secondary structure despite being small, and is known to be a fast folder.[25,41–48] The effects of the denaturing osmolyte urea on Trp-cage conformation will then be investigated.

Materials and Methods

Theoretical Background

The free energy of a hydrated molecule can be written as:

$$G^{tot} = E^{vac} + G^{np} + G^{el} \tag{3}$$

where E^{vac} is the molecule's energy in vacuum, which is the sum of internal contributions (bond and angle stretching, dihedral angles interactions) and van der Waals energy terms. G^{np} is the non-polar solvation contribution, i.e., the free energy of solvation for a molecule from which all charges have been removed. G^{el} is the electrostatic part, calculated as the free energy for turning on the partial charges in solution.

We imagine a solute composed of n atoms, treated as spheres of radius R_i and charge q_i. The interior of the atoms is assumed to be filled with material having a dielectric constant equal to 1, while a dielectric constant ε is assigned to the surrounding solvent medium. G^{el} is therefore calculated from the generalized Born equation,[49]

$$G^{el} = \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{n} \sum_{j>1}^{n} \frac{q_i q_j}{\sqrt{r_{ij}^2 + b_i b_j \exp\left(-\frac{r_{ij}^2}{4 b_i b_j}\right)}} \tag{4}$$

where r_{ij} is the distance between atoms i and j, and b_i, b_j are the Born radii of atoms i and j. The Born radius depends on the degree of the atom burial inside the solvated molecule;[50,51] b_i is similar to the van der Waals radius R_i for a surface atom, while b_i is on the order of the solvated molecule radius for a deeply buried atom. The OBC(II) model[28] was used in the present work for calculating the Born radii.

The dielectric constant of the solution was approximated to that of pure water, i.e., we assumed the presence of the osmolyte to have a negligible effect. This assumption is justified by the form of the prefactor $(1 - 1/\varepsilon)$ in Eq. 4, that dampens changes in ε.

The non-polar solvation contribution G^{np} can also be computed from the Born radius of

each atom, using the ACE type approximation,[29]

$$G^{np} = 4\pi\sigma \sum_i (R_i + R_s)^2 \frac{R_i}{b_i} \tag{5}$$

where R_s is the radius of a water probe sphere (0.14 nm), and σ is a surface tension. A single solvation parameter σ was here used for all atom types (5.4 cal mol^{-1} Å^{-2}).

Our objective is to add the effect of an osmolyte to the implicit description outlined in Eq. 3. For this purpose, the concept of free energy of transfer G^{tr} [14,15,17,52] is used. G^{tr} represents the free energy change observed upon transferring a protein from pure water to the osmolyte solution, and its use makes it possible to generalize Eq. 3 as follows,

$$G^{tot} = E^{vac} + G^{np} + G^{el} + G^{tr} \tag{6}$$

where G^{tr} is described according to Eq. 1.

The LCPO (linear combination of pairwise overlaps) algorithm[53] was used in this work to compute the solvent accessible surface area. More details on LCPO can be found in the Supporting Information. The script for the computation of the fractional solvent accessible surface area, and the energetic term G^{tr}, is currently part of the 2.8 version of Plumed[54] as a separate module called SASA. Plumed has been selected for the implementation of the MIST method because it can be readily combined with several molecular simulation codes, including Amber,[55] CP2K,[56] Espresso,[57] LAMMPS,[58] Gromacs[59] and NAMD.[60]

Simulation Details

Trp-cage (pdb 1L2Y[40]) was capped by an acetyl group and an amide moiety at the N and C termini, respectively. It was simulated at pH 7 (+1 charge), using the Amber 99SB-ILDN force field.[61] We chose the Amber 99SB-ILDN force field because previous work confirmed its suitability to describe the conformational landscape of Trp-cage.[25,44]

For all simulations, Gromacs 5.1.4,[59] patched with Plumed 2.4.7[54] plug-in, was used. A

list of all simulations performed is shown in Table 1.

Table 1: List of the simulations performed in this work. bb: backbone, sc: sidechains

Sim. #	Protein	Solution	Duration ns	Temperature K
1	Trp-cage	water	300	298
2	Trp-cage	1M urea - B	300	298
3	Trp-cage	2M urea - B	300	298
4	Trp-cage	3M urea - B	300	298
5	Trp-cage	4M urea - B	300	298
6	Trp-cage	1M urea - H	300	298
7	Trp-cage	2M urea - H	300	298
8	Trp-cage	3M urea - H	300	298
9	Trp-cage	4M urea - H	300	298
10	Trp-cage	8M urea - B	300	298
11	Trp-cage	8M urea - B (only bb)	300	298
12	Trp-cage	8M urea - B (only sc)	300	298
13	Trp-cage	8M urea - B (only bb + polar sc)	300	298
14	Trp-cage	8M urea - B (only bb + apolar sc)	300	298
15	Trp-cage	4M urea - H (only bb)	300	298
16	Trp-cage	4M urea - H (only sc)	300	298
17	Trp-cage	4M urea - H (only bb + polar sc)	300	298
18	Trp-cage	4M urea - H (only bb + apolar sc)	300	298
19	Trp-cage	water	300	320
20	Trp-cage	water	300	340

Equilibration in explicit solvent

Trp-cage was first equilibrated in explicit TIP3P water.[62] A cubic box with 5.4 nm side length and containing one peptide molecule was energy minimized using the steepest descent algorithm, and then equilibrated for 1 ns at 1 bar and 300 K, using Berendsen pressure (3 ps relaxation time) and temperature (0.5 ps relaxation time) coupling.[63] A second equilibration for 5 ns was then performed at 1 bar and 300 K using the Nosé-Hoover thermostat[64,65] (0.5 ps relaxation time) and Parrinello-Rahman barostat[66] (3 ps relaxation time). For this equilibration in explicit solvent, periodic boundary conditions were used, and the cut-off radius

for both Coulombic (calculated using the PME method[67]) and Lennard-Jones interactions was set to 1.2 nm. An equilibrated protein structure (Figure 2) was extracted from this 5-ns run, and used as starting configuration for the implicit solvent trajectories in Table 1. We note that a first equilibration in explicit solvent was performed in this work, but the equilibration could have also been directly performed in implicit solvent.

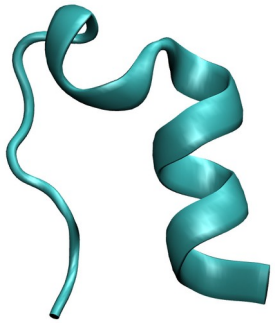

Figure 2: Starting configuration of Trp-cage used in simulations 1-20.

Implicit solvent simulations

In simulations 1-18, the water or water-osmolyte solution was described implicitly, as detailed in the Theoretical Background section. The Gromacs implicit solvation tool was used to simulate water (Eq. 3), while the free energy of transfer term G^{tr} (Eq. 1) was added as an external bias using Plumed, so that the system was eventually described by Eq. 6.

The $g^{tr}_{k,sc}$ and g^{tr}_{bb} values (to be inserted into Eq. 1) for urea were taken from the experimental work by Bolen and coworkers[30] or Moeser and Horinek,[39] and are listed in Table S1. In particular, the values for 2M, 3M, 4M and 8M urea were obtained multiplying by 2, 3, 4 or 8 the experimental data for 1M urea. This means that we assumed linearity of the free energy of transfer with the osmolyte concentration, which is a common assumption[15].

For two simulations (11 and 15 in Table 1) the values of $g^{tr}_{k,sc}$ were set to zero, i.e., only the backbone contribution of the osmolyte was considered. Similarly, in simulations 12 and 16 only the sidechain contributions were used, while g^{tr}_{bb} was set to zero. In simulations 13

and 17, only the backbone and polar sidechains (Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Lys, Ser and Thr) contributions were considered. The opposite was done in simulations 14 and 18, where the $g^{tr}_{k,sc}$ values of Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Lys, Ser and Thr were set to zero. This is equivalent to considering the contribution of the backbone and of the apolar amino acids (Ala, Ile, Leu, Met, Phe, Pro, Trp, Tyr and Val) only.

Two additional simulations (19 and 20) were performed in water at 320 and 340 K, with the objective of extracting the temperature dependence of Trp-cage stability.

For all the simulations in implicit solvent (1-20 in Table 1), no cut-off was used for the Coulombic and Lennard-Jones interactions. The production run lasted 300 ns; the first 100 ns were considered as an equilibration, and only the last 200 ns were used for the subsequent analyses. The Nosé-Hoover thermostat[64,65] (0.5 ps relaxation time) was used to control temperature, and the Lincs algorithm was employed for constraining all bonds.[68] The systems were simulated without periodic boundary conditions, and center of mass translational and rotational motions were removed every 100 steps. A 2 fs timestep was used, and configurations were saved every 2 ps.

300 ns were enough to observe conformational transitions of Trp-cage, thanks to the faster equilibration allowed by the implicit treatment of the solvent.[6] Convergence graphs are shown in Figure S1, where it is evident that the backbone root mean square deviation (RMSD), α-helix content and free energy G^{tot} of Trp-cage in water or 8M urea - B fluctuate around equilibrium values during the last 200 ns of the trajectory.

Analysis of the Trajectories

Cluster analysis

The peptide conformations during the last 200 ns of the trajectories were grouped together by performing a cluster analysis based on the Daura algorithm.[69] The conformations were grouped together if the root mean square deviations (RMSD) of the N-C_α-C atoms were less than 0.14 nm compared to each other. The most probable conformations were subsequently

visualized using VMD (Visual Molecular Dynamics).[70]

α-helix content

The α-helix (α) content of Trp-cage used in the remainder of this work is defined as the number of 6 residue sections of the peptide having an α-helical configuration,[71]

$$\alpha = \sum_{\mu} g[r_{dist}(\{R_i\}_{i\in\Omega_\mu}, \{R^0\})] \tag{7}$$

The summation runs over all possible segments involved in the α-helix, while $\{R_i\}_{i\in\Omega_\mu}$ are the atomic coordinates of a set Ω_μ of 6 residues of the protein, and $g(r_{dist})$ is the following switching function,

$$g(r_{dist}) = \frac{1 - \left(\frac{r_{dist}}{r_0}\right)^8}{1 - \left(\frac{r_{dist}}{r_0}\right)^{12}} \tag{8}$$

A cutoff distance of $r_0 = 0.08$ nm was used, and r_{dist} is the distance RMSD with respect to a reference α-helix structure $\{R^0\}$.

Folded fraction

The folded fraction of Trp-cage during the equilibrated trajectories (last 200 ns) was extracted from the backbone root mean square deviation (RMSD) compared to the most likely cluster in water at 298 K (sim. 1 in Table 1). Trp-cage was deemed to be folded when the backbone RMSD was less than 0.14 nm.

Results and Discussion

Trp-cage Unfolds in Urea in MIST Simulations

Nine simulations (1-9 in Table 1) were performed with the objective to explore the effect of the modelling approach herein proposed (Model with Implicit Solvation Thermodynamics, MIST) on the conformational stability of Trp-cage. A potent denaturant (urea) was considered for these simulations, and compared to the case of water only. As previously mentioned, two different descriptions of urea were tested, as proposed by either Auton, Holthauzen and Bolen[30] (urea - B) or Moeser and Horinek[39] (urea - H). In the description by Bolen and coworkers the backbone contribution g_{bb}^{tr} is negative, and represents the main factor driving unfolding. In contrast, Moeser and Horinek suggested a second set of transfer free energy values (urea - H), where most side chain contributions $g_{k,sc}^{tr}$ are also negative, thus contributing to the unfolding process.

Figures 3 and S2, S3 summarize the results obtained from these initial simulations. The folded fraction was high in pure water (0.80 ± 0.05), and urea concentrations of 1M and 2M were not sufficient to promote substantial denaturation in our simulations. Instead, the folded fraction decreased in 3M or 4M urea (Figure 3A). The Horinek description of urea (urea - H) resulted in more pronounced unfolding compared to the Bolen set (urea - B), with a folded fraction value as low as 0.2 ± 0.1 in the 4M urea - H simulation.

A relation exists between the free energy of unfolding ΔG, for the folded-to-unfolded conversion, and the folded fraction f,

$$\Delta G = -\mathrm{R}T \ln(\frac{1-f}{f}) \tag{9}$$

where R is the universal gas constant, and T is temperature. Exploting this relation, we could compute the free energy of unfolding for Trp-cage in the different systems simulated, and compared the results obtained to the experimental values by Wafer et al.[72] (Figure 3A). We observed that Trp-cage was generally slightly stabler in the implicit solvent simulations

herein performed than in experiments. This may be related to the force field (Amber 99SB-ILDN) employed, which tends to slightly overestimate the Trp-cage free energy of unfolding, also in pure water.

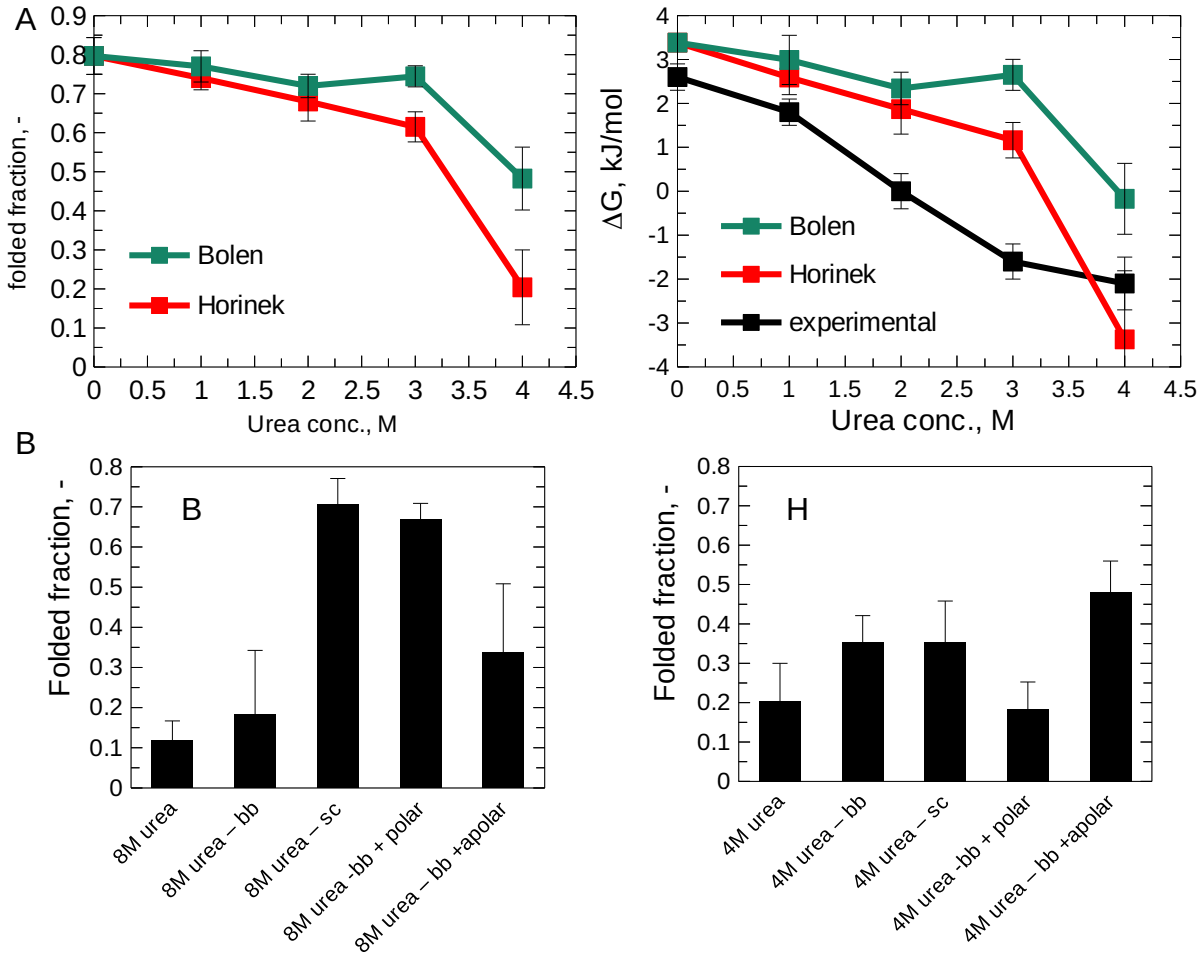

Figure 3: (A) Folded fraction of Trp-cage as function of urea concentration, and corresponding free energy of unfolding values ΔG. (B) Folded fraction of Trp-cage during simulations 11-18, for either the urea - B (left) or urea - H (right) description. Errors were estimated by block averaging. Briefly, the equilibrated trajectories were divided into 4 blocks, and the standard deviation computed over the average values of the folded fraction in each of the blocks.

According to the linear extrapolation model[73] the free energy of unfolding in a urea solution at concentration c should be related to the free energy of unfolding in water ($\Delta G(0)$) by the following linear equation, where the slope is the so-called m-value,

$$\Delta G(c) = \Delta G(0) + mc \tag{10}$$

We extracted from Figure 3A approximate m-values. Linear regression yields m-values of 0.75 kJ mol^{-1} M^{-1} and 1.49 kJ mol^{-1} M^{-1} for the Bolen and Horinek descriptions, respectively. The value obtained with the Horinek description is closer to the experimental result of 1.36 kJ mol^{-1} M^{-1} by Wafer et al.[72]

The most probable configuration in water (Figure S2) was characterized by α-helix content $\alpha \approx 6.5$, radius of gyration $R_g \approx 0.77$ nm, and solvent accessible surface area $SASA \approx 19$ nm^2. When the G^{tr} contribution corresponding to 3M or 4M urea was added, a progressive loss in α-helix content was observed (Figure S3), especially during the 4M urea - H simulation. The residues that were mostly involved in the unfolding process included the fragment Tyr3-Leu7. The α-helix which characterizes this protein sequence in water was partially replaced by a turn, and a 3_{10}-helix in urea. Also residues Pro12-Ser14 displayed a change in conformation, with an increased probability of forming a 3_{10}-helix, and reduced probability of being observed in a α-helix. Correspondingly, residue Trp6, which is buried within the protein core in the native conformation, became more accessible to the solvent in urea.

The free energy of unfolding was negative in 4M urea - H (-3.37 $\pm$ 1.6 kJ mol^{-1} K^{-1}), while it was still close to zero in 4M urea - B (-0.17 $\pm$ 0.81 kJ mol^{-1} K^{-1}). We therefore also investigated a higher urea concentration with the Bolen description (8M urea - B, simulation 10) to identify a condition where unfolding was sufficiently pronounced. At this higher concentration, the folded fraction was 0.11 $\pm$ 0.05, corresponding to a free energy of unfolding $\Delta G = -4.98 \pm 1.2$ kJ mol^{-1} K^{-1}.

When the G^{tr} contribution corresponding to 8M urea - B was added (Figure S4), the α-helix content was partially lost, even though this did not translate into a marked increase in radius of gyration or solvent accessible surface area. A large ensemble of protein conformations could be sampled during this simulation, as confirmed by application of the Daura

algorithm (right panel in Figure S4).

Two additional simulations (19 and 20 in Table 1) were performed with the objective of extracting the temperature dependence of Trp-cage stability. The folded fraction of Trp-cage was extracted from these simulations, and used to compute the free energy of unfolding according to Eq. 9. The free energy of unfolding as function of temperature T was fit to the following equation,

$$\Delta G(T) = [(1 - T/T_m)\Delta H(T_m)] + (T - T_m)\Delta C_p - \ln(T/T_m)T\Delta C_p \tag{11}$$

where T_m is the melting temperature and $\Delta H(T_m)$ and ΔC_p are the enthalpy change and change in heat capacity upon unfolding. The fitting procedure is shown in Figure S5, and yielded the thermodynamic parameters listed in Table 2, where also the experimental data by Wafer et al.[72] are shown for comparison. Overall, the agreement between experiments and simulations was quite good, as shown in Figure S5.

Table 2: List of thermodynamic parameters for Trp-cage stability in water.

Parameter	Experimental[1]	Simulations
$\Delta H(T_m)$, kJ mol^{-1}	56 ± 2	47
T_m,°C	43.9 ± 0.8	48.5
ΔC_p, kJ mol^{-1} K^{-1}	0.3 ± 0.1	0.08

[1] From Wafer et al.[72]

Overall, partial unfolding of Trp-cage was observed in urea at 298 K, with both the Bolen and Horinek set of transfer free energies. These first results confirm that the simulation approach herein proposed can describe the effect of osmolytes, like the denaturant urea, on protein stability. In the following, further simulation outputs will be presented, with the objective to highlight the differences between the Bolen and the Horinek set of transfer free energies.

The Backbone Drives Urea-Induced Unfolding of Trp-cage Within the Bolen Model, While the Backbone and Sidechains Contribute Equally in the Horinek Description

As a second objective of the present work, the mechanism of unfolding observed in urea in Figures 3A, S3 and S4 was further investigated, for both the Bolen and Horinek descriptions. For this purpose, the additivity of Eq. 1 was exploited. This additivity makes it easy to decouple some terms from the expression of G^{tr}. In simulations 11-18 of Table 1 we tried to dissect the different contributions to the free energy of transfer in 8M urea - B or 4M urea - H, considering only the backbone (sim. 11, 15), only the sidechains (sim. 12, 16), or a combination of the backbone with either the polar (sim. 13, 17) or the apolar (sim. 14, 18) sidechains.

Considering the 8M urea - B case first, we observed unfolding when only the backbone contribution was considered (Figures 3B and 4A), and the loss in secondary structure was even more pronounced than when all the backbone and sidechains contributions were included (Figure S4). The folded fraction was only 0.18 $\pm$ 0.2, and also in this case the fragments Tyr3-Leu7 and Pro12-Ser14 were mostly involved in the unfolding process. This suggests that the favorable urea-backbone interaction ($g^{tr}_{bb} = -1.305$ kJ mol^{-1} in 8M urea - B) within the Bolen description is sufficient to explain the denaturing action of urea, in line with what was already suggested in the literature.[30]

In agreement with this hypothesis, Trp-cage retained its secondary structure when the sidechains only were biased (Figures 3B and 4B). The folded fraction in these conditions was high (0.71 $\pm$ 0.06), and the most probable protein conformations (right panel of Figure 4B) were native-like, without any appreciable change in the degree of compactness or secondary structure of the native state observed in water.

When both the backbone and the polar sidechains contributions were considered (Figure 4C), the loss of secondary structure was again very limited. This means that the

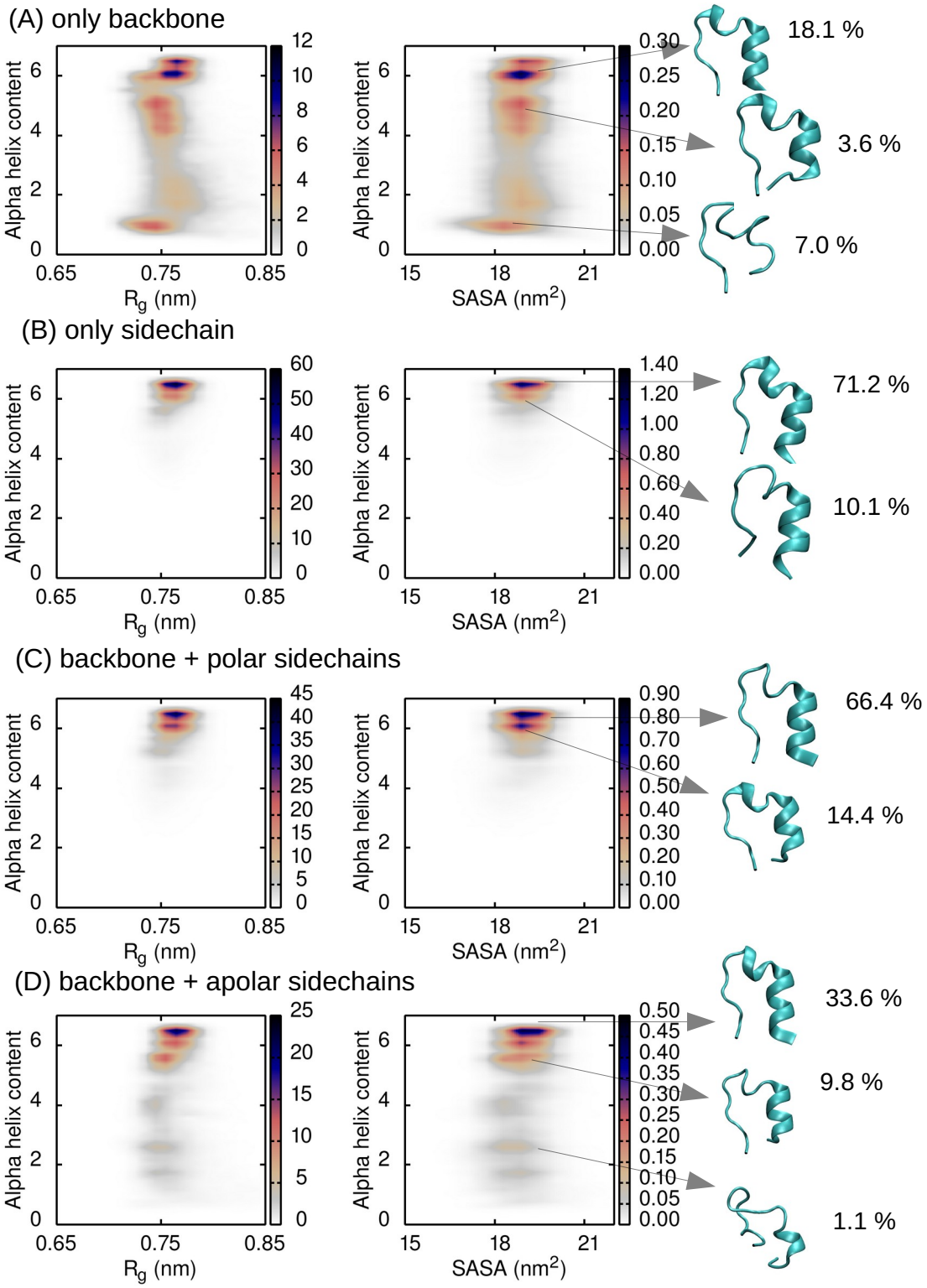

Figure 4: Distribution of α-helix content as function of radius of gyration (left panel) or solvent accessible surface area (middle panel), and most probable conformations (right panel) for Trp-cage in 8M urea - B at 298 K. The following contributions were considered: (A) only the backbone, (B) only the sidechains, (C) backbone + polar sidechains, (D) backbone + apolar sidechains.

predominantly unfavorable urea-polar sidechains interaction within the Bolen description ($\sum_{polar} g^{tr}_{k,sc}\alpha_{k,sc} \approx 2.96$ kJ mol^{-1}) could effectively counteract the negative backbone contribution ($g^{tr}_{bb}\sum_{k=1}^{n_r}\alpha_{k,bb} \approx -12.47$ kJ mol^{-1}), leading to mostly native-like conformations of Trp-cage. The results obtained also suggest that a perfect compensation of the very favorable urea-backbone interactions (≈ -12.47 kJ mol^{-1}) is not necessary to provide stabilization, as the much smaller unfavorable contribution of the polar sidechains (≈ 2.96 kJ mol^{-1}) was enough to almost completely prevent denaturation.

The apolar sidechains contributions ($\sum_{apolar} g^{tr}_{k,sc}\alpha_{k,sc} \approx -0.29$ kJ mol^{-1}) are slightly favorable in the Bolen description, although close to zero, but nevertheless did not increase the unfolding capability of the backbone-urea interactions (the folded fraction in these conditions was 0.34 ± 0.17). As can be observed in Figure 4D, the structures obtained when both the backbone and the apolar sidechains were biased preserved most of their secondary structure. The R_g vs. α and $SASA$ vs. α distributions in these conditions (Figure 4D) were restricted to more native-like structures compared to the case of sim. 11 (Figure 4A), or sim. 10 (Figure S4).

The same type of approach was further used for the case of 4M urea - H (Figures 3B and 5). In this case, the backbone and sidechains contributions were equally effective in promoting Trp-cage unfolding (the folded fractions were extremely similar, 0.35 ± 0.07 when only the backbone was considered, and 0.35 ± 0.10 when only the sidechains transfer free energies were included). A partial loss in α-helix content was observed in these conditions (Figures 5A,B), again at the expenses of residues Tyr3-Leu7 and Pro12-Ser14. The mostly favorable urea-sidechains interaction within the Horinek description ($\sum g^{tr}_{k,sc}\alpha_{k,sc} \approx -2.91$ kJ mol^{-1}) could effectively denature Trp-cage, in line with what was already suggested by Moeser and Horinek.[39]

When both the backbone and the polar sidechains contributions were considered (Figure 5C), the loss of secondary structure was particularly pronounced (the folded fraction was only 0.18 ± 0.07 in these conditions, see Figure 3B). This means that the predominantly favorable

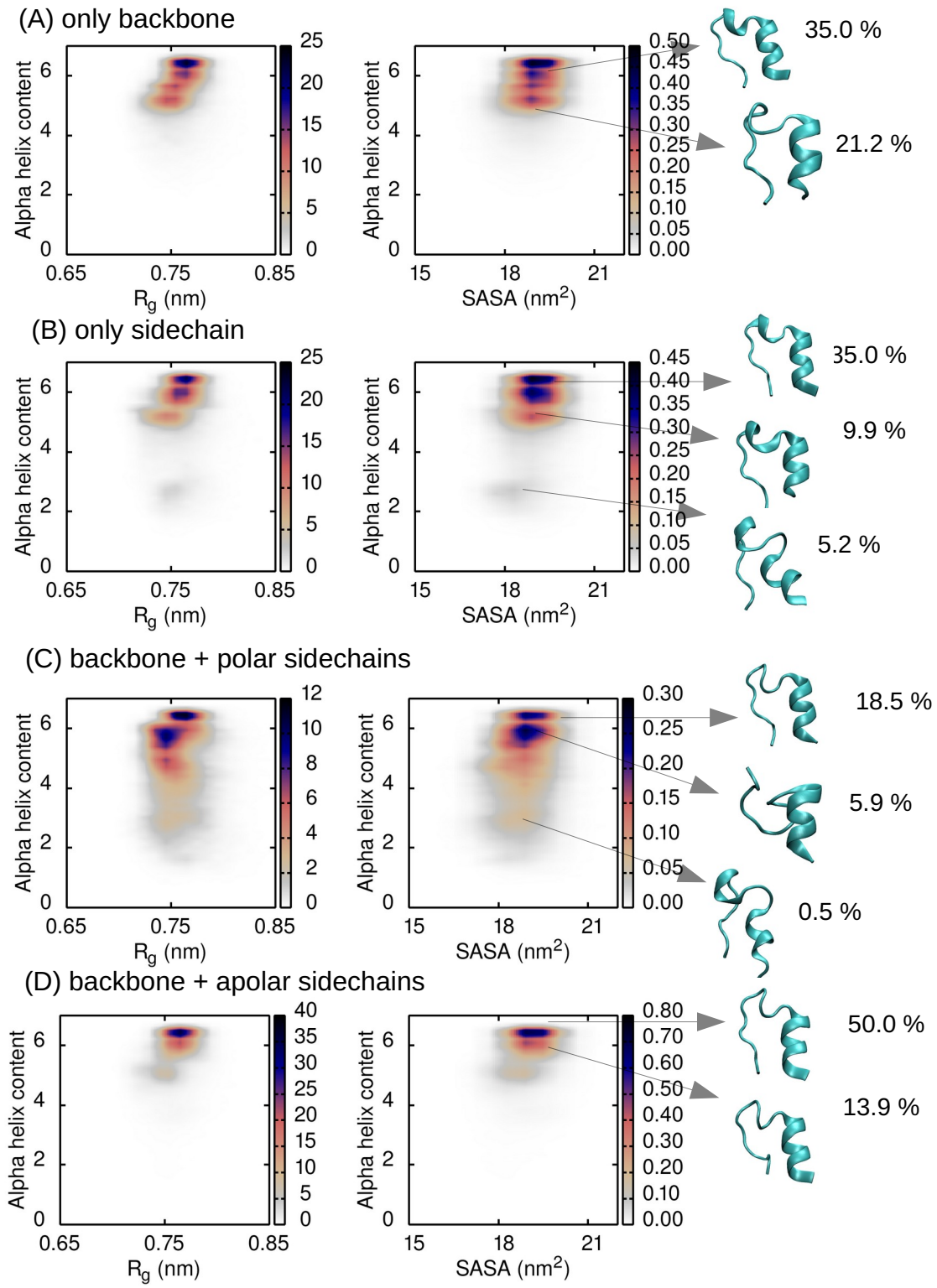

Figure 5: Distribution of α-helix content as function of radius of gyration (left panel) or solvent accessible surface area (middle panel), and most probable conformations (right panel) for Trp-cage in 4M urea - H at 298 K. The following contributions were considered: (A) only the backbone, (B) only the sidechains, (C) backbone + polar sidechains, (D) backbone + apolar sidechains.

urea-polar sidechains interaction within the Horinek description ($\sum_{polar} g^{tr}_{k,sc}\alpha_{k,sc} \approx -1.03$ kJ mol^{-1}) could effectively support the negative backbone contribution ($g^{tr}_{bb}\sum_{k=1}^{n_r}\alpha_{k,bb} \approx -4.05$ kJ mol^{-1}), leading to mostly unfolded conformations of Trp-cage.

The apolar sidechains contributions could not increase the unfolding capability of the backbone-urea interactions (the folded fraction in these conditions was 0.48 $\pm$ 0.08, see Figure 3B). As a result, the structures obtained when both the backbone and the apolar sidechains were biased preserved most of their secondary structure (Figure 5D).

Overall, the following conclusions can be drawn from the simulations just presented for the case of 8M urea - B and 4M urea - H. The favorable backbone-urea interaction within the Bolen description seems to be the main driving force for its denaturing action, as already suggested in the literature by authors using this same model,[30] although other studies suggest that the interaction with the backbone is not dominant.[74] The polar sidechains contributions can effectively counteract the unfolding process, and, surprisingly, the apolar sidechains also seem to oppose conformational changes, although less markedly. This picture, however, probably does not correspond to reality. Indeed, as previously discussed, Bolen and coworkers overestimated the contribution of the backbone to urea-induced unfolding, and underestimated the sidechains contribution due to a mistake in the conversion of activity coefficient data. These two errors compensate each other when describing the overall unfolding profile (Figure 3A), but become apparent when the backbone and sidechains contributions are considered separately, as allowed by our MIST approach.

The Horinek description corrects for these two errors, thus recovering the correct mechanism of urea-induced unfolding. In this case, the backbone and sidechains transfer free energies contribute equally to the denaturation of Trp-cage in 4M urea, in line with the conclusions by Moeser and Horinek.[39] Based on our simulation results, the polar sidechains contribution seems to be particularly effective, at least in the case of Trp-cage, to promote denaturation.

Conclusions

A new approach (Model with Implicit Solvation Thermodynamics, MIST) for the implicit simulation of protein-osmolyte solutions, based on transfer free energies, has been here proposed. This method has been applied to the urea-induced denaturation of Trp-cage as model system. We have described the theoretical background of this tranfer free energy-based method, and shown that it can succesfully describe the effect of osmolytes on protein conformational stability. Moreover, the additive construction of the transfer free energy term makes it easy to decouple the different contributions involved in the protein-osmolyte interaction. This fact has been here exploited to separately investigate the backbone and sidechains terms for the urea-induced unfolding of Trp-cage.

Two sets of transfer free energies have been investigated for urea, as proposed by Bolen and coworkers[30] or by Horinek and coworkers.[39] We have observed that in both cases urea leads to unfolding of Trp-cage. This occurs because of the favorable urea-backbone interaction in the Bolen model, while the backbone and sidechains contributions are equally important in the framework of the Horinek description. The Horinek description is likely to be more accurate, as it corrects for two errors that were made in the derivation of the Bolen model. The mechanism predicted by the MIST approach directly reflects the set of transfer free energy values employed. Bolen and coworkers overestimated the contribution of the backbone to urea-induced unfolding, and underestimated the sidechains contribution, and this clearly emerges from the output of our simulations. Looking at both the Bolen and Horinek description was therefore very informative, as it allowed us to test the sensitivity of the MIST approach to differences in the set of transfer free energies considered.

Overall, the approach proposed in this work gives promising results for the simulation of aqueous osmolyte solutions. For the first time, we show how free energies of transfer can be fit into an implicit solvent simulation to reproduce the effect of osmolytes on protein behavior. The advantage is that the description we achieve using this method is completely and directly based upon experimental quantities, thus eliminating the need for a complex

and time-consuming development of force field parameters.

We have shown that this transfer free energy approach can be used to model the effect of urea on protein conformational stability. However, this approach can straightforwardly be extended to other osmolytes, and also to the investigation of protein-protein interactions and protein aggregation in complex mixtures. Furthermore, the method could directly be extended to more complex matrices, provided that the free energies of transfer between water and such matrices are known.

Supporting Information

Details on the LCPO algorithm, free energy of transfer values for urea, convergence plots, Trp-cage behavior in water, 3M or 4M urea and 8M urea - B at 298 K, free energy of unfolding in water as function of temperature.

Acknowledgement

The authors acknowledge support from the Center for Scientific Computing at the California Nanosystems Institute (CNSI, NSF grant CNS-1725797) for the availability of high performance computing resources and support. This work used the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation grant number ACI-1548562 (MCA05S027). The authors aknowledge support from the NSF (MCB-1716956 and CHE-1800352) and the NIH (R01-GM118560-01A).

References

(1) Ganguly, P.; Hajari, T.; Shea, J.-E.; van der Vegt, N. F. A. Mutual Exclusion of Urea and Trimethylamine N-Oxide from Amino Acids in Mixed Solvent Environment. *J. Phys. Chem. Lett.* **2015**, *6*, 581–585.

(2) Ganguly, P.; Boserman, P.; van der Vegt, N. F. A.; Shea, J.-E. Trimethylamine N-oxide Counteracts Urea Denaturation by Inhibiting Protein–Urea Preferential Interaction. *J. Am. Chem. Soc.* **2018**, *140*, 483–492.

(3) Arsiccio, A.; Ganguly, P.; La Cortiglia, L.; Shea, J.-E.; Pisano, R. ADD Force Field for Sugars and Polyols: Predicting the Additivity of Protein-Osmolyte Interaction. *J. Phys. Chem. B* **2020**, *124*, 7779–7790.

(4) Smith, L. J.; Berendsen, H. J. C.; van Gunsteren, W. F. Computer Simulation of Urea-Water Mixtures: A Test of Force Field Parameters for Use in Biomolecular Simulation. *J. Phys. Chem. B* **2004**, *108*, 1065–1071.

(5) Cloutier, T.; Sudrik, C.; Sathish, H. A.; Trout, B. L. Kirkwood–Buff-derived alcohol parameters for aqueous carbohydrates and their application to preferential interaction coefficient calculations of proteins. *J. Phys. Chem. B* **2018**, *122*, 9350–9360.

(6) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 1153–1164.

(7) Malaspina, D. C.; Pérez-Fuentes, L.; Drummond, C.; Bastos-González, D.; Faraudo, J. Protein-surface interactions at the nanoscale: Atomistic simulations with implicit solvent models. *Curr. Opin. Colloid Interface Sci.* **2019**, *41*, 40–49.

(8) Arsiccio, A.; Shea, J.-E. Protein Cold Denaturation in Implicit Solvent Simulations: A Transfer Free Energy Approach. *J. Phys. Chem. B* **2021**, *125*, 5222–5232.

(9) Arsiccio, A.; Shea, J.-E. Pressure Unfolding of Proteins: New Insights into the Role of Bound Water. *J. Phys. Chem. B* **2021**, *125*, 8431–8442.

(10) Yancey, P.; Clark, M.; Hand, S.; Bowlus, R.; Somero, G. Living with water stress: evolution of osmolyte systems. *Science* **1982**, *217*, 1214–1222.

(11) Hochachka, P. W.; Somero, G. N. *Biochemical Adaptation: Mechanism and process in physiological evolution*; Oxford University Press: Oxford, 2002; pp 217 – 289.

(12) Yancey, P. H. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* **2005**, *208*, 2819–2830.

(13) Yancey, P. H. Water Stress, Osmolytes and Proteins. *Am. Zool.* **2001**, *41*, 699 – 709.

(14) Auton, M.; Bolen, D. W. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15065–15068.

(15) Auton, M.; Bolen, D. W.; Rösgen, J. Structural thermodynamics of protein preferential solvation: Osmolyte solvation of proteins, aminoacids, and peptides. *Proteins* **2008**, *73*, 802–813.

(16) Auton, M.; Rösgen, J.; Sinev, M.; Holthauzen, L. M. F.; Bolen, D. W. Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophys. Chem.* **2011**, *159*, 90 – 99.

(17) Tanford, C. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **1964**, *86*, 2050–2059.

(18) Qin, M.; Denesyuk, N.; Liu, Z.; Wang, W.; Thirumalai, D. Temperature and Guanidine Hydrochloride Effects on the Folding Thermodynamics of WW Domain and Variants. *J. Phys. Chem. B* **2021**, *125*, 11386–11391.

(19) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci.* **2008**, *105*, 13403–13408.

(20) Liu, Z.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc. Natl. Acad. Sci.* **2011**, *108*, 7787–7792.

(21) Liu, Z.; Reddy, G.; Thirumalai, D. Theory of the Molecular Transfer Model for Proteins with Applications to the Folding of the src-SH3 Domain. *J. Phys. Chem. B* **2012**, *116*, 6707–6716.

(22) Liu, Z.; Reddy, G.; Thirumalai, D. Folding PDZ2 Domain Using the Molecular Transfer Model. *J. Phys. Chem. B* **2016**, *120*, 8090–8101.

(23) Mugnai, M. L.; Thirumalai, D. Molecular Transfer Model for pH Effects on Intrinsically Disordered Proteins: Theory and Applications. *J. Chem. Theory Comput.* **2021**, *17*, 1944–1954.

(24) Mishra, P.; Jha, S. K. The native state conformational heterogeneity in the energy landscape of protein folding. *Biophys. Chem.* **2022**, *283*, 106761.

(25) Ganguly, P.; Shea, J.-E. Distinct and Nonadditive Effects of Urea and Guanidinium Chloride on Peptide Solvation. *J. Phys. Chem. Lett.* **2019**, *10*, 7406–7413.

(26) Xie, T.; Liu, D.; Feng, Y.; Shan, L.; Wang, J. Folding Stability and Cooperativity of the Three Forms of 1–110 Residues Fragment of Staphylococcal Nuclease. *Biophys. J.* **2007**, *92*, 2090–2107.

(27) Attri, P.; Venkatesu, P. Refolding of urea-induced denaturation of model proteins by trimethylamine N-oxide. *Thermochim. Acta* **2011**, *526*, 143–150.

(28) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.

(29) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model1 1Edited by A. F. Fersht. *J. Mol. Biol.* **1998**, *284*, 835 – 848.

(30) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci.* **2007**, *104*, 15317–15322.

(31) Hong, J.; Capp, M. W.; Saecker, R. M.; Record, M. T. Use of Urea and Glycine Betaine To Quantify Coupled Folding and Probe the Burial of DNA Phosphates in Lac Repressor-Lac Operator Binding. *Biochemistry* **2005**, *44*, 16896–16911.

(32) Nozaki, Y.; Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **1963**, *238*, 4074–4081.

(33) Tanford, C. In *Protein Denaturation: Part C*; Anfinsen, C., Edsall, J. T., Richards, F. M., Eds.; Adv. Protein Chem.; Academic Press: Cambridge, Massachusetts, United States, 1970; Vol. 24; pp 1 – 95.

(34) Watlaufer, D. B.; Malik, S. K.; Stoller, L.; Coffin, R. L. Nonpolar Group Participation in the Denaturation of Proteins by Urea and Guanidinium Salts. Model Compound Studies. *J. Am. Chem. Soc.* **1964**, *86*, 508–514.

(35) Pace, C. *Methods in enzymology*; Elsevier: Cambridge, Massachusetts, United States, 1986; Vol. 131; pp 266–280.

(36) Das, A.; Mukhopadhyay, C. Urea-Mediated Protein Denaturation: A Consensus View. *J. Phys. Chem. B* **2009**, *113*, 12816–12824.

(37) Rossky, P. J. Protein denaturation by urea: Slash and bond. *Proc. Natl. Acad. Sci.* **2008**, *105*, 16825–16826.

(38) Camilloni, C.; Rocco, G.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. Urea and Guanidinium Chloride Denature Protein L in Different Ways in Molecular Dynamics Simulations. *Biophys. J.* **2008**, *94*, 4654–4661.

(39) Moeser, B.; Horinek, D. Unified Description of Urea Denaturation: Backbone and Side Chains Contribute Equally in the Transfer Model. *J. Phys. Chem. B* **2014**, *118*, 107–114.

(40) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430.

(41) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.

(42) Snow, C. D.; Zagrovic, B.; Pande, V. S. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549.

(43) Pitera, J. W.; Swope, W. Understanding folding and design: Replica-exchange simulations of ''Trp-cage'' miniproteins. *Proc. Natl. Acad. Sci.* **2003**, *100*, 7587–7592.

(44) Levine, Z. A.; Fischer, S. A.; Shea, J.-E.; Pfaendtner, J. Trp-Cage Folding on Organic Surfaces. *J. Phys. Chem. B* **2015**, *119*, 10417–10425.

(45) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins* **2010**, *78*, 1889–1899.

(46) Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13280–13285.

(47) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLOS Comput. Biol.* **2009**, *5*, 1–18.

(48) Juraszek, J.; Bolhuis, P. G. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci.* **2006**, *103*, 15859–15864.

(49) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(50) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(51) Schaefer, M.; Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578–1599.

(52) Auton, M.; Bolen, D. W. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* **2004**, *43*, 1329–1342.

(53) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.

(54) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604 – 613.

(55) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1 – 41.

(56) Kuhne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schutt, O.; Schiffmann, F. et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(57) Weik, F.; Weeber, R.; Szuttor, K.; Breitsprecher, K.; de Graaf, J.; Kuron, M.; Landsgesell, J.; Menke, H.; Sean, D.; Holm, C. ESPResSo 4.0 - an extensible software package for simulating soft matter systems. *Eur. Phys. J. Spec. Top.* **2019**, *227*, 1789–1816.

(58) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D. et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.* **2022**, *271*, 108171.

(59) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19 – 25.

(60) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Henin, J.; Jiang, W. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.

(61) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(63) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(64) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(65) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(66) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(68) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(69) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.

(70) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33 – 38.

(71) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197–2201.

(72) Wafer, L. N. R.; Streicher, W. W.; Makhatadze, G. I. Thermodynamics of the Trp-cage miniprotein unfolding in urea. *Proteins* **2010**, *78*, 1376–1381.

(73) Bolen, D. W.; Santoro, M. M. Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of .DELTA.G.degree.N-U values in a thermodynamic cycle. *Biochemistry* **1988**, *27*, 8069–8074.

(74) Canchi, D. R.; Paschek, D.; García, A. E. Equilibrium Study of Protein Denaturation by Urea. *J. Am. Chem. Soc.* **2010**, *132*, 2338–2344.

Graphical TOC Entry