Osmolyte-Induced Structural and Thermodynamic Effects on the Protein Folding Landscape

*Ander Francisco Pereira and Leandro Martínez**

Institute of Chemistry and Center for Computing in Engineering & Science,

Universidade Estadual de Campinas (UNICAMP), 13083-861 Campinas, SP, Brazil

---

Corresponding Author:

*Leandro Martínez: lmartine@unicamp.br, Institute of Chemistry, University of Campinas. 13083-970, Campinas, SP. Brazil. http://m3g.iqm.unicamp.br

ABSTRACT

Understanding how osmolytes influence protein folding is fundamental in biology and biotechnology. We developed a computational pipeline to investigate the structural and thermodynamic effects of osmolytes across the protein folding landscape: Using coarse-grained simulations to generate folding ensembles, we then employed atomistic simulations to explore each conformation in aqueous solutions of urea and TMAO. Minimum-distance distribution functions and the Kirkwood-Buff solvation theory provide unprecedented details on the solvation structures and thermodynamics. We studied two model proteins: the β-sheet SH3 domain and the helical B domain of protein A (BdpA). Urea stabilizes most denatured states, while TMAO destabilizes them. However, urea destabilizes partially denatured states of BdpA, which are only marginally affected by TMAO, demonstrating that osmolytes can selectively interfere with denaturation pathways. We show that urea-induced denaturation arises from strengthened non-specific interactions in unfolded states, leading to increased protein dehydration. Cosolvent effects on folding free energies match experimental data within 1 kcal mol^{-1}, validating force-fields and analysis methods. These findings provide a novel perspective on osmolyte-protein interplay, with implications for solvent and biomolecular design.

Introduction

Cells utilize low molecular weight organic molecules, known as osmolytes, to maintain protein function under external stress stimuli such as pressure, temperature, salinity, and pH variations[1,2]. Beyond their natural roles in stress adaptation, osmolytes are of significant interest in the biotechnological and pharmaceutical industries, particularly for stabilizing protein-based drugs and vaccines[3–5]. TMAO, a well-known protective osmolyte found in marine species and the mammalian kidney[6], is preferentially excluded from protein surfaces[7,8], increasing both the melting temperature (T_m) and the free energy of protein unfolding (ΔG_U°)[9–11]. Conversely, urea, a common denaturant, preferentially interacts with proteins, decreasing both T_m and ΔG_U° [11,12].

The effect of osmolytes on protein stability can be analyzed at a first level of approximation from interactions with the protein in its native state: osmolyte exclusion from the protein surface favors compact structures, typically linked to folded forms[13,14]. Conversely, preferential protein-osmolyte interactions promote the exposure of the protein residues and denaturation[15,16]. This rationale is oversimplified for proteins because denaturation alters the overall chemical nature of the protein surface. Therefore, interactions observed in the native state do not fully capture the complex interplay between cosolvents and the unfolded ensemble[17]. Characterizing the unfolded state is difficult both experimentally and computationally, due to its diverse conformational landscape[15]. Furthermore, consistently understanding the solvation structure and thermodynamics across a wide range of unfolded conformations is a major challenge[17,18].

Obtaining preferential interaction parameters for multiple protein conformations (experimentally or computationally) allows for a more comprehensive analysis of the impact of cosolvents on folding equilibrium[16,17,19]. The difference in preferential

solvation among two states is linked to the dependence of their relative population as a function of cosolvent concentration[20]. Furthermore, determining the absolute cosolvent effect on the folding equilibria necessitates either directly measuring equilibrium constants in both pure solvent and cosolvent solutions, or determining the transfer free energies of the native and denatured states[11]. The experimental characterization of the folding equilibrium is challenging, because the conformations of native and unfolded ensembles might differ in protective and denaturant conditions, and experiments depend on a reduced variable to classify the ensembles: the protein activity, the secondary structure content, a specific NMR signal, etc. From a computational perspective, the challenges are equally significant: Proper sampling of protein folding, accurate visualization of protein folding landscapes, reliable methods for identifying the solvation shell around the solute, and achieving convergence or thermodynamic properties, all require methodological precision.

Here, we investigated the effects of urea and TMAO on the energy landscapes of two model proteins with distinct structural motifs: the β-sheet SH3 domain[21] and the α-helical B domain of protein A (BdpA)[22]. We combine structure-based models (SBMs)[23] with atomistic simulations to generate folding ensembles, and utilize Minimum-Distance distribution functions (MDDFs)[24] to characterize solvation structures, preferential interaction coefficients, and transfer free energies of folding intermediates. Our results demonstrate that solvation effects vary across unfolded conformations, influencing the energy landscape and potentially altering unfolding pathways. We further show that urea's denaturing effect stems from its preferential accumulation at sites of dispersive interactions, which are more prevalent in denatured states. The breakdown of the MDDFs into residue contributions revealed unprecedented details regarding the correlation between solvent exposure and denaturation intermediates.

Results

The folding landscape of model proteins

Understanding how cosolvents affect the protein folding free-energy surface requires the characterization of the protein conformational variability. The Energy Landscape (EL) theory[25,26] provides a framework for modeling protein-folding mechanisms, and forms the basis for simulations using Structure-Based Models (SBMs). SBMs are demonstrated to provide coarse-grained representations of the protein folding energy surfaces and structural ensembles of model proteins[27–36]. In this work, SBMs are initially used to sample the conformational landscape of the SH3 (a β-fold model, Figure 1A) and BdpA (a helical domain, Figure 1B)[37]. The folding mechanisms of these models have been extensively characterized by experimental and computational methods, including by SBM simulations[27–31,37].

SBM simulations can be performed at the folding temperature of the model, obtaining multiple folding and unfolding events that allow the reconstruction of the folding free-energy surface. For SH3, the overall properties of the SBM sampling are shown in Figure 1C, and the protocol and characterization of the SBM folding is available in Supplementary Information Figures S1 and S2. For BdpA, the characterization of the folding landscapes is shown in Figure 1D and further details were recently published[37]. Single sharp peaks in the $C_v(T)$ profiles (Figures 1C and 1D) indicate two-state folding mechanisms, but the folding transition is sharper for SH3 than for BdpA. In Figures 1C and 1D, panels iii), the exhaustive sampling of native contacts (Q) over time can be observed, an indicator of the proper sampling of the conformational landscapes.

Visualization of the folding landscape requires techniques for dimensionality reduction. Thus, for each SBM simulation, the sampled folds were projected in a two-dimensional space using the *Energy-Landscape Visualization Method* (ELViM)[30,38]: In

panels i) of Figures 1C and 1D, each dot represents a conformation obtained in the SBM simulation, and the distances of the dots in the projection are optimized to correlate with the dissimilarity of the models[30]. The colors of the dots, from blue to red, represent completely unfolded (Q = 0) to completely folded (Q = 1) conformations. In Figures 1C and 1F labels are assigned to clusters of conformations: N and U represent the native basin and unfolded subsets, with a subscript indicating the average secondary structure content of the cluster, and the superscripts S or B for SH3 or BdpA followed by the classification of the cluster in terms of decreasing Q. For example, N_{80}^{S0} is the native cluster of SH3, which displays on average 80% of the secondary structure of the crystallographic model. Alternatively, U_{37}^{B11} is, in Figure 1F, an unfolded cluster of BdpA with 37% of its native secondary structure content, and the 11th cluster in decreasing amount of native contacts. This labeling will be used along the manuscript to reference the subsets of the folding landscape of the models. Comprehensive structural characterizations of each ensemble are shown in Supplementary Figures S3, S4, and S5, and Table S1.

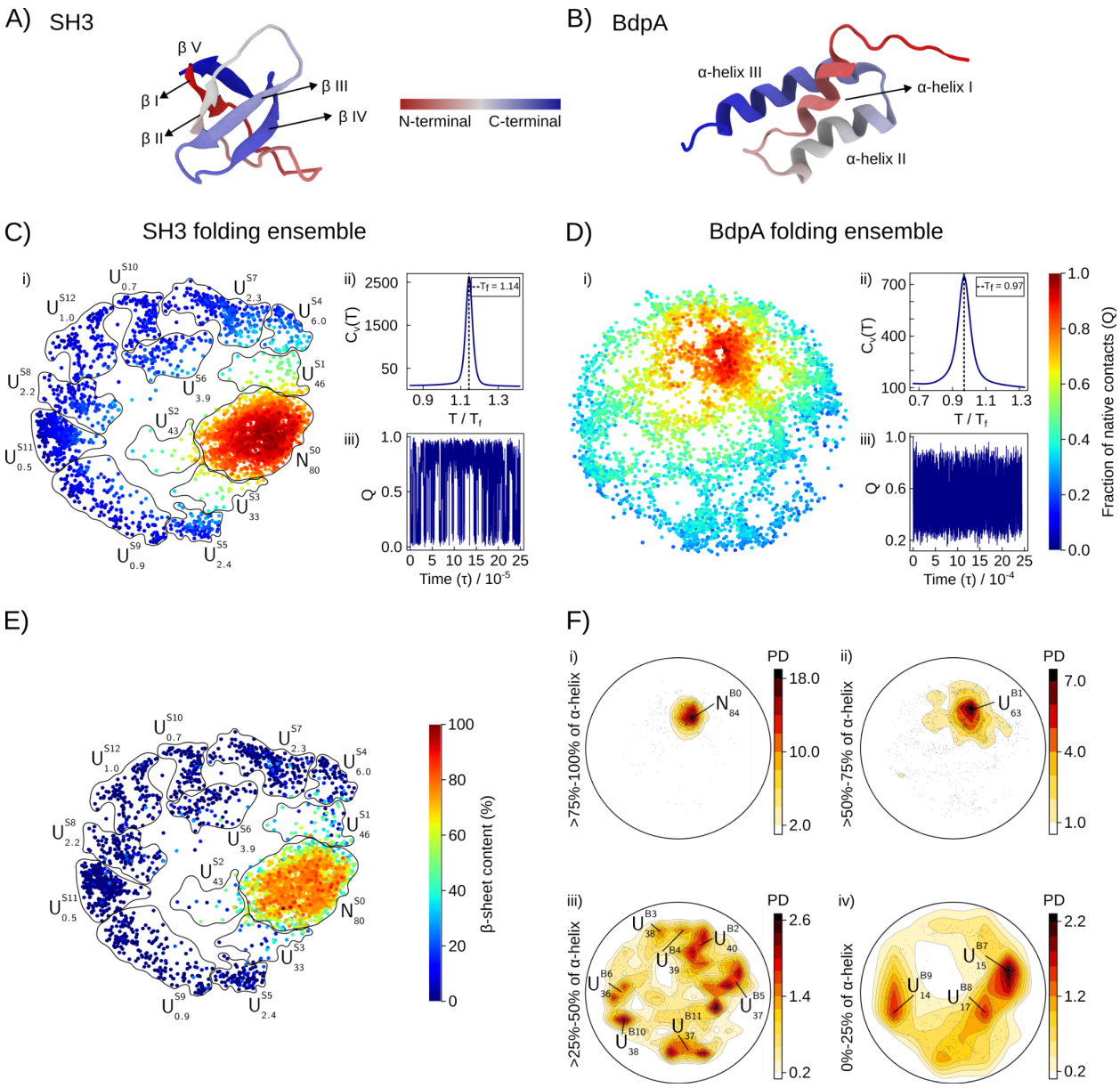

Figure 1. Folding ensembles of SH3 and BpdA. A) and B) Cartoon representation of SH3[21] and BdpA[22] native folds. C) and D) SBM folding ensembles characterized by i) the ELViM projection, ii) the specific heat (C_v) as a function of temperature, and iii) Fraction of native contacts (Q) as a function of time, in reduced units (τ). Each dot in i) and ii) represents one structure in a color scale of Q, ranging from 0 (unfolded state) to 1 (folded state). E) Folding projection of SH3 colored by secondary structure content, and definition of the conformational subsets. There is a high correlation between Q (panel C)

and the secondary structure content. F) Secondary structure dimension of the 2D projection for BpdA, which is required for classifying unfolded states[37]. In Figures E) and F) labels are assigned to clusters of native (N) and unfolded (U) states, with a subscript indicating the average secondary structure content and the superscript the order of decreasing fraction of native contacts.

Figures C and D show that the folded ensembles (in red) are clustered, as expected from the small native-state conformational variability. On the other hand, unfolded states are spread over the projection, because of their structural diversity. The folding variability of SH3 is captured by the 2D projection: there is a clear native-state basin (N_{80}^{S0}), separated from a range of unfolded states ($U_{6.0}^{S4}$ - $U_{1.0}^{S12}$) by only three intermediate sets of conformations (U_{46}^{S1}, U_{43}^{S2}, and U_{33}^{S3}). Furthermore, there is a high correlation between the fraction of native contacts (Figure 1C) and the secondary structure content (Figure 1E).

In contrast, the folding space of BdpA displays multiple pathways connecting the native (*N*) and unfolded (*U*) states via structures with similar Q (Figure 1D). Furthermore, partially unfolded states can be found in the native basin, and the secondary structure content needs to be added as an additional dimension to properly differentiate conformations[37]. Thus, in Figure 1F, projections are split into ranges of protein ellipticity. Panels i), ii), iii), and iv) depict the ensembles within 75-100%, 50-75%, 25-50%, and 0-25% of α-helix content, respectively. The U_{63}^{B1} state, in panel ii) of Figure 1F, is particularly interesting: it is a partially denatured state, characterized by the partial loss of the secondary structure of helix I of BdpA, but with mostly preserved native contacts (Supplementary Table S1 and Figures S3 and S5). The denaturation of helix I is widely recognized to be fundamental for the folding equilibrium of BdpA in

water[37,39,40]. The notable interactions of this U_{63}^{B1} state and its destabilization in the presence of urea will be discussed in detail.

The above characterization of the SH3 and BdpA ensembles is crucial for the primary objective of this work, i.e., to thoroughly investigate urea and TMAO interactions and thermodynamic effects on the protein folding landscapes. With this aim, we obtained all-atom reconstructions of 5000 conformations from the SBM simulations, and each of these atomic models was simulated in aqueous solutions of urea or TMAO at 0.5 mol L^{-1}. After equilibration with constrained protein backbone coordinates, the solvated systems were simulated for 10 ns. For each state, as classified in Figure 1, this implied from 210 to 34130 ns of simulations (21 to 3413 structures in each subset), from which the average structure and thermodynamics of the solvents were obtained.

Osmolyte effects on the folded and unfolded states of SH3

We first focus on the effect of osmolytes urea and TMAO on the folding ensemble of the β-sheet fold of the SH3 domain, which displayed a simpler conformational landscape. Urea is a known denaturant, while TMAO is a stabilizer known to counteract the denaturing effect of urea[41]. The structural characterization of protein-solvent interactions was possible using Minimum-Distance Distribution Functions, and the Kirkwood-Buff theory of solvation. The ComplexMixtures.jl[42] package was developed to perform these analyses. The thermodynamics of protein-solvent interactions was characterized by measuring preferential interaction parameters, the Wyman linkage relation, and transfer free energies[9,16,43].

MDDFs are distribution functions useful for characterizing the structure of solutions of solutes and solvents of complex shapes[24], particularly the unfolded states of proteins[17]. MDDFs are obtained by computing the minimum-distances between solute

atoms and solvent molecule atoms, and thus retain a clear "solvent-shell" interpretation independently of the complexity of the structures involved. We have shown that a proper normalization of these distributions allows the practical computation of KBIs from the minimum-distance counts[24], thus connecting the microscopic picture of solvation to the macroscopic preferential interactions and free-energies of solvation.

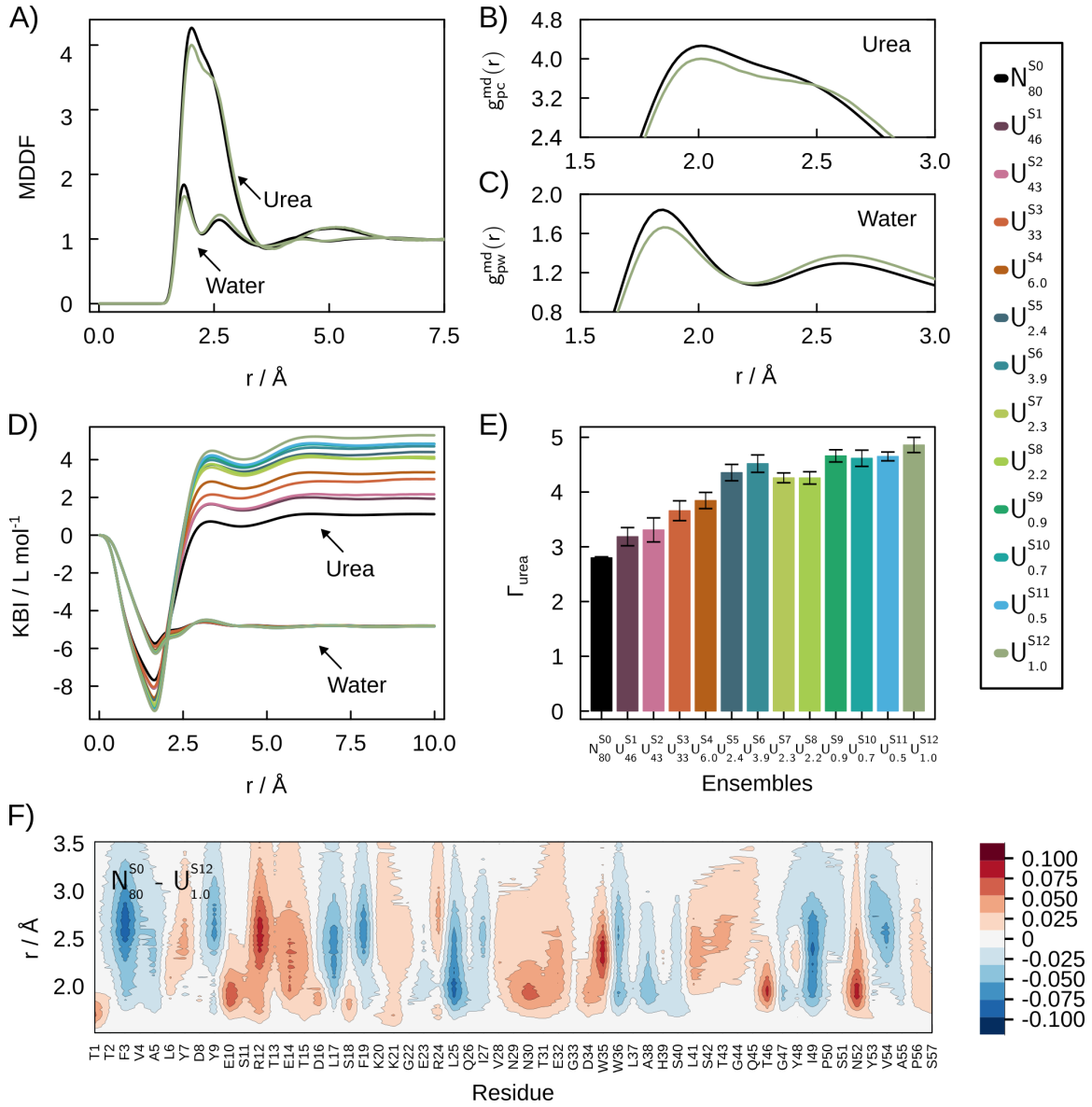

Figure 2. Solvation structures of the folding ensemble of the SH3 domain in 0.5 mol L^{-1} aqueous urea solutions. A) MDDFs of water and urea for the N_{80}^{S0} and $U_{1.0}^{S12}$ ensembles. B)

and C) most prominent peaks in the MDDFs of urea and water. D) KBIs and E) Preferential interaction parameters (Γ). The preferential interaction of urea for unfolded states is greater than for the native (N_{80}^{S0}) state and increase among denatured states. F) Difference in the breakdown of MDDF per residue in the vicinity of N_{80}^{S0} and the unfolded $U_{1.0}^{S12}$ states. Red regions indicate greater urea density near the N_{80}^{S0} state, while blue regions show higher density around the $U_{1.0}^{S12}$ state.

Figures 2A-C display the MDDFs of water and urea relative to the protein in the native fold basin (N_{80}^{S0}) and a selected unfolded state ($U_{1.0}^{S12}$), with 2B and 2C being insets of the most prominent peaks of these distributions. The $U_{1.0}^{S12}$ state was chosen for displaying the greatest preferential interactions with urea. Figure 2A shows clear density augmentation of water and urea in the protein vicinity. Both water and urea form hydrogen bonds (H-bonds) with the protein (as evidenced by the peaks at ~1.9 Å). These H-bond peaks are greater in the native subset. On the other hand, the MDDFs for the denatured state become greater at distances of non-specific interactions (>2.5 Å). The Kirkwood-Buff Integrals (KBIs), which measure the overall solvent accumulation around the protein, provide a quantitative picture of the relevance of the interactions at each distance.

The KBIs for urea, in Figure 2D, are greater than those of water, implying protein dehydration. Indeed, preferential interaction parameters with urea are always positive (Figure 2E). The distance-dependence of the KBIs, furthermore, provides insights on the contributions of interactions at each solvent shell to the build up of the final KBIs: The initial drop (Figure 2D) is associated with the excluded protein-urea volume. At distances smaller than 2.5 Å the KBIs of the denatured states become smaller than those

of the native state, consistently with the observation that specific interactions with the folded state are stronger, as shown in Figures 2A and 2B. The urea KBIs then increase sharply, outgrowing those of water, and assume roughly the final converged value at 5 Å. Therefore, the accumulation of urea at the distances of non-specific interactions is the determining feature leading to protein dehydration. Also, it is responsible for the increased preferential solvation by urea in the denatured states relative to the native state. These observations support the direct mechanism of urea-induced denaturation driven by dispersive interactions[44]. Experimental support for these strengthened dispersive interactions with unfolded states have been found with different denaturants[45].

The preferential interaction parameters (Γ_{urea}) roughly increase monotonically with increasing protein denaturation. For instance, N_{80}^{S0} (the native state) has the lower Γ value compared with the other ensembles: it is 1.74 times lower than the higher Γ observed in the $U_{1.0}^{S12}$ unfolded state. Urea preferential interactions correlate with the average fraction of native contacts (Q), β-sheet content, and solvent-accessible surface area (SASA) (Supplementary Figure S6).

An in-depth molecular picture of urea accumulation at the surface SH3 in different fold states can be obtained by a differential MDDF density map (Figure 2F). This map displays the difference in local density of urea of the N_{80}^{S0} and $U_{1.0}^{S12}$ states around each protein residue. Red indicates a greater urea density in the native state, while blue indicates greater density in the unfolded state. Urea molecules are more densely found in the first shell of the N_{80}^{S0} state around polar residues (threonines T1, T31, and T46, serines S18 and S57, and asparagines N30 and N52), charged residues (glutamate E10, arginine R12, lysine K21, and aspartate D16 and D34), and N- and C-terminal groups. Conversely, urea is found with greater density in the $U_{1.0}^{S12}$ unfolded

state at larger distances, around the hydrophobic residues such as phenylalanine (F3 and F19), tyrosine (Y9), leucine (L17 and L25), tryptophan (W36), isoleucine (I27 and I49), and valine (V54). Similar behavior is observed for water, indicating a competition between water and urea for specific and non-specific interactions in the folded and unfolded structures (Supplementary Figure S7).

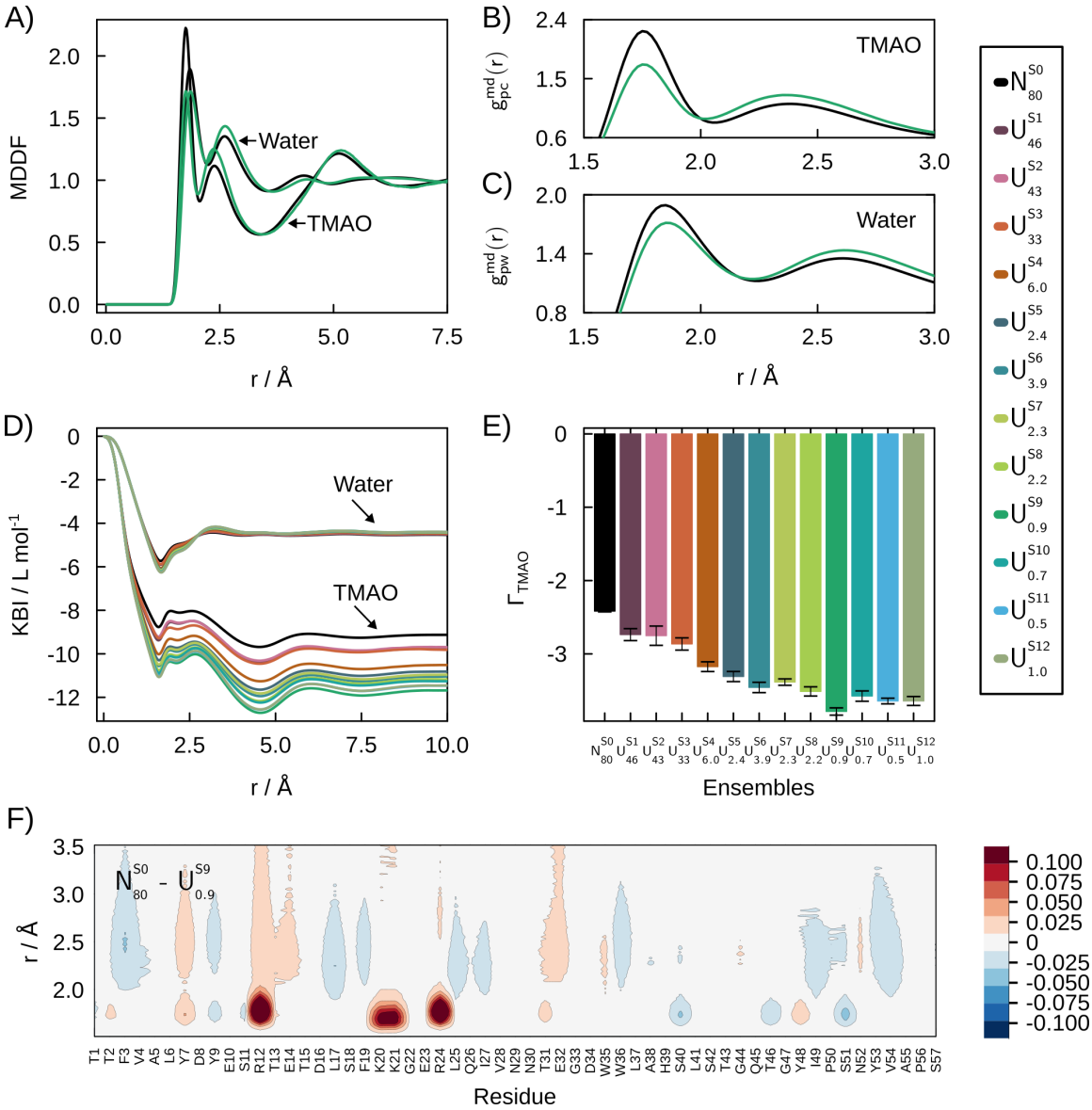

Figure 3. Solvation structures of the folding ensemble of the SH3 domain in 0.5 mol L^{-1} aqueous solutions of TMAO. A) MDDFs of water and TMAO for N_{80}^{S0} and $U_{0.9}^{S9}$

ensembles. B) and C) most prominent peaks in the MDDFs of TMAO and water. D) KBIs and E) Preferential interaction parameters. The exclusion of TMAO for unfolded states is greater than for the native (N_{80}^{S0}) state and increase among denatured states. F) Difference in the MDDF density of the urea in the vicinity of N_{80}^{S0} and of the unfolded $U_{0.9}^{S9}$ state. Red indicates higher TMAO density near the N_{80}^{S0} state, while blue higher density around the $U_{0.9}^{S9}$ state.

Figure 3 shows the analysis of solvation structures of SH3 in aqueous solutions of TMAO. TMAO is a stabilizing osmolyte, a property associated with preferential exclusion (or, equivalently, preferential hydration)[11,46]. In Figure 3A and 3B we see that TMAO and water, nevertheless, form specific interactions with the protein[24] (H-bonds at ~1.75 Å), but the TMAO peaks are much smaller than those of urea at these distances (Figures 2A and 2B). Also, there is a significant dip centered in ~3.5 Å in the MDDFs of TMAO. We compare the MDDFs of the native state (N_{80}^{S0} -black) with the unfolded state with greater TMAO exclusion ($U_{0.9}^{S9}$ - green), and observe that the local density of TMAO in the denatured state is lower at H-bonding distances, but higher at distances associated to non-specific interactions. This behavior is observed in all unfolded ensembles (Supplementary Figure S8). The greater contribution of non-specific interactions in the denatured state is, as with urea, an expected feature associated with the exposure of the protein core. H-bonds with the protein involve the oxygen atom of TMAO (Supplementary Figures S9A and S9C). The second shell of the MDDF of the TMAO arises from its amphiphilic nature: TMAO has three methyl groups, which interact favorably with the hydrophobic regions of unfolded ensembles (Supplementary Figure S9C). Importantly, both specific and non-specific interactions between TMAO and SH3 primarily involve the protein side chains, with negligible interactions with the

backbone (Supplementary Figures S9B and S9D). This is consistent with interactions of TMAO, and other protecting osmolytes, with the protein backbone favoring denaturation[18,47].

Figure 3D displays the KBIs of water and TMAO for all SH3 states, as a function of the distance. Water KBIs are similar for all ensembles, but TMAO KBIs decrease with the increasing unfolding of the protein. TMAO exclusion at the short (< 2.0 Å) and mid-range (~2-5 Å) distances essentially determines the converged values of the KBIs. The smaller KBIs of TMAO relative to water imply that all the ensembles exhibit preferential hydration ($\Gamma_{TMAO} < 0$) (Figure 3E), and TMAO exclusion increases with denaturation. This is consistent with the mechanism where protective osmolytes are excluded from the protein surface[11,46]. An interesting characteristic observed here is that the more unfolded ensembles ($U_{6.0}^{S4}$ to $U_{1.0}^{S12}$) display greater TMAO exclusion, independently of the more hydrophobic character of the exposed protein core.

Figure 3F shows the difference in the MDDF density per residue of TMAO in the vicinity of N_{80}^{S0} and $U_{0.9}^{S9}$. Red and blue indicate higher TMAO density around N_{80}^{S0} and $U_{0.9}^{S9}$, respectively. The density of TMAO around N_{80}^{S0} and unfolded $U_{0.9}^{S9}$ are similar for most residues. Interactions in the first solvation shells of R12, R24, K20 and K21 are responsible for the greater MDDF peak at ~1.75 Å in the native state (Figure 3B). The increase in density of the second solvation shell in the $U_{0.9}^{S9}$ is linked with polar (Y7, R12, and Y53) and nonpolar (F3, L17, L25, W36, I49, and V54) residues. Differential water density maps resemble those observed in the urea solutions (Supplementary Figures S7 and S10). Thus, water accumulates around the same residues independently of the cosolvent, albeit with a higher density in the TMAO solutions.

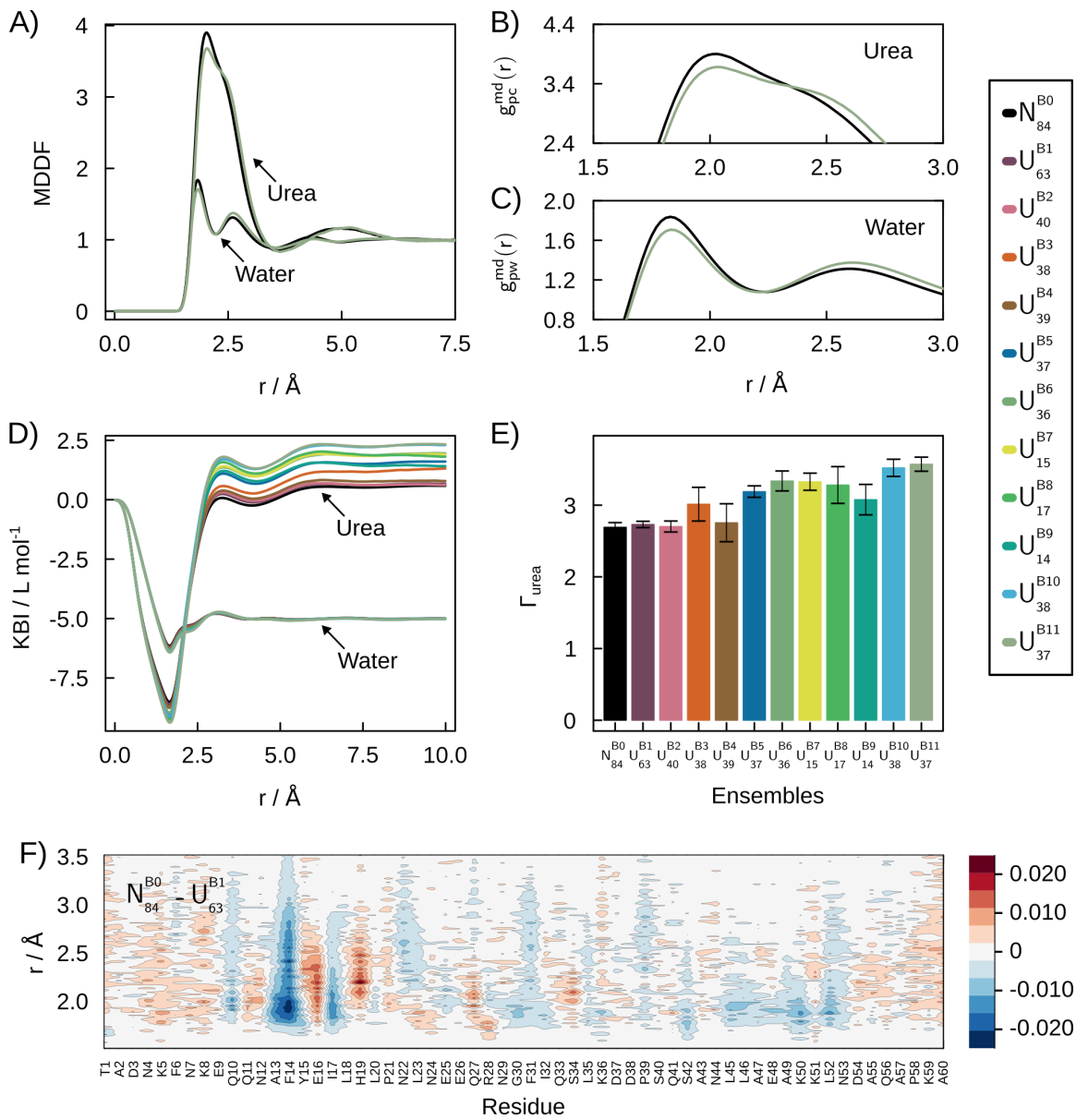

Figure 4. Solvation structures of the folding ensemble of BdpA in 0.5 mol L^{-1} aqueous urea solutions. A) MDDFs of water and urea for N_{84}^{B0} and U_{37}^{B11} ensembles. B) and C) most prominent peaks in the MDDFs of urea and water. D) KBIs and E) Preferential interaction parameters. The preferential interaction of urea for unfolded states is greater than for the native (N_{84}^{B0}) state and increase among denatured states. F) Differential

density map of the MDDF breakdown per residue of N_{84}^{B0} and U_{63}^{B1}. Red indicates greater density around N_{84}^{B0} and blue around U_{63}^{B1}.

Osmolyte effects on the folded and unfolded states of BdpA

In this section we will discuss the solvation structure of BdpA, a model α-helix protein, in the presence of urea and TMAO. On one hand, the overall solvation properties of BdpA are similar to those of SH3: In Figure 4A and 4B, density augmentations of water and urea at short distances from the protein can be observed. Figure 4B shows that the H-bonding peaks are greater in the native state relative to the denatured state, but non-specific interactions are strengthened in the denatured states. Qualitatively similar MDDF profiles are observed for TMAO in Figures 5A and 5B. Urea displays greater KBIs than water (Figure 4D) and solvates preferentially the BdpA domain (Figure 4E). TMAO, on the contrary, is preferentially excluded (Figures 5D and 5E). Urea is a better solvent for SH3 than for BdpA, and TMAO is a poorer one, as judged by the preferential interaction parameters. This is consistent with the observation that urea predominantly unfolds β-sheet structures[12,48,49].

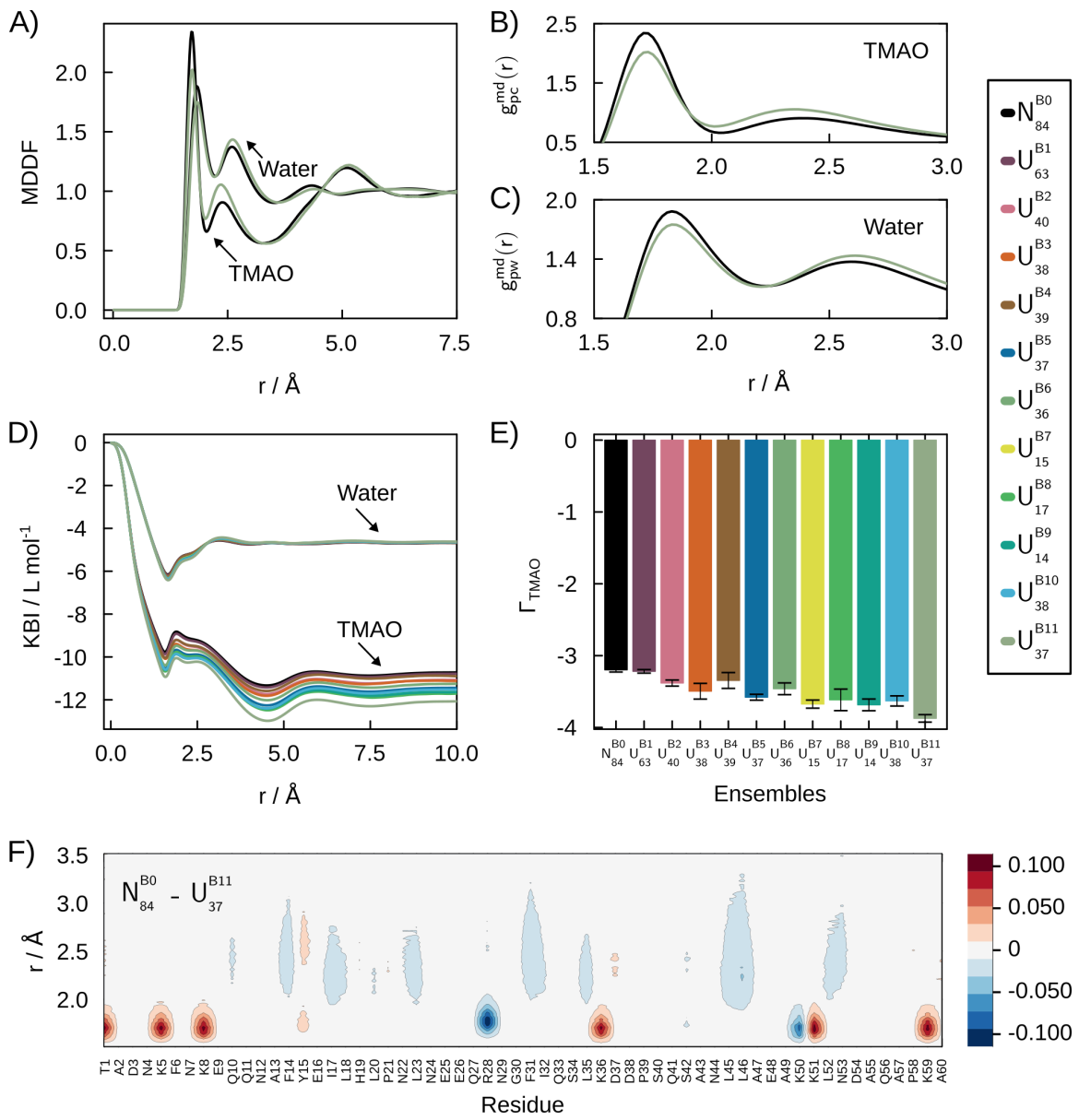

Figure 5. Solvation structures of BdpA in 0.5 mol L^{-1} aqueous solutions of TMAO. A) MDDFs of water and urea for N_{84}^{B0} and U_{37}^{B11} ensembles. B) and C) most prominent peaks in the MDDF of TMAO and water. D) KBIs and E) Preferential interaction parameters. The preferential interaction of TMAO for unfolded states is lower than for the native (N_{84}^{B0}) state and decrease for denatured states. F) Differential density map of the MDDF

breakdown per residue of N_{84}^{B0} and U_{37}^{B11}. Red indicates greater density around N_{84}^{B0} and blue around U_{37}^{B11}.

BdpA first four denatured subsets, U_{63}^{B1}, U_{40}^{B2}, U_{38}^{B3}, and U_{39}^{B4}, display preferential interactions roughly equivalent to those of the native N_{84}^{B0} state within the simulation fluctuations, notably in urea (Figure 4E) but also to a smaller extent in TMAO (Figure 5E). The U_{63}^{B1} state is, effectively, only partially denatured: it displays 63% of the crystallographic helical content and 72% of native contacts, versus 84% and 77% of the N_{84}^{B0} ensemble from the folded basin (Supplementary Table S1). The partial denaturation of U_{63}^{B1} is, for this reason, only perceived when the folding projection is split into secondary-structure compositions (Figure 1F). Nevertheless, U_{63}^{B1} is an important conformational ensemble of BdpA: it is characterized by the partial denaturation of helix I (Supplementary Figure S3). Helix I is known to be labile in BdpA, providing structural variability to the folded ensemble and leading the initial stages of denaturation[37,39,40]. By contrast, the U_{39}^{B4} set, with 39% of helical content and 54% of native contacts, is much further denatured. Yet, its surface accessible surface area, 52.7 nm^2 (Supplementary Table S1) is similar to that of the native set, of 51.2 nm^2, indicating that it has conserved a globular fold despite the rearrangement of the secondary structure and tertiary contacts. These first four BdpA denatured sets have SASAs within 4% of that of the native ensemble, while the remaining unfolded ensembles (U_{37}^{B5} to U_{37}^{B11}) have accessible surface areas that are at least 11% greater than that of N_{84}^{B0}. Therefore, the SASA values determine the resulting preferential interaction parameters with urea.

Figure 4F shows the differential density map of U_{63}^{B1} and N_{84}^{B0} per residue. The exposure of F14 and neighboring residue to urea highlights the link between helix I denaturation and the solvation structure. The pronounced exposure of F14 to urea is a common feature of all denatured ensembles, except of the unfolded but compact U_{39}^{B4} (Supplementary Figure S11), where it remains buried in the protein's hydrophobic core. Variations in the density of TMAO around each residue are smaller due to its preferential exclusion. However, in denatured states, TMAO interacts more with hydrophobic residues, and less with the charged residues, particularly lysines (Figure 5F and Supplementary Figure S12).

In summary, BdpA interactions with urea and TMAO are qualitatively similar to those of SH3, but two sets of ensembles can be distinguished: i) partially denatured states with a slight increase in surface area interact with the solvent similarly to the native state; ii) denatured states with greater surface area are dehydrated by urea and preferentially hydrated in the presence of TMAO. These contrasting behaviors contribute to a multifaceted solvent dependency of the denaturation equilibrium of BdpA, which will be quantitatively discussed below.

Osmolyte induced shifts in denaturation equilibrium constants

The variation of the equilibrium constant, K, of among two folding states U and N, as a function of the osmolyte activity a_c, can be obtained with the Wyman linkage equation[43,50],

$$\left(\frac{\partial \ln K}{\partial \ln a_c}\right)_{m_p} = \Gamma_{pc}^{U} - \Gamma_{pc}^{N} = \Delta\Gamma_{pc} \tag{1}$$

where Γ are the preferential interaction parameters. $\Delta\Gamma_{pc}$ greater than zero implies that an increase in cosolvent concentration favors the denatured state. In qualitative terms, if

the cosolvent solvates more favorably the denatured state than the native state, increasing the cosolvent concentration will shift the equilibrium towards denaturation.

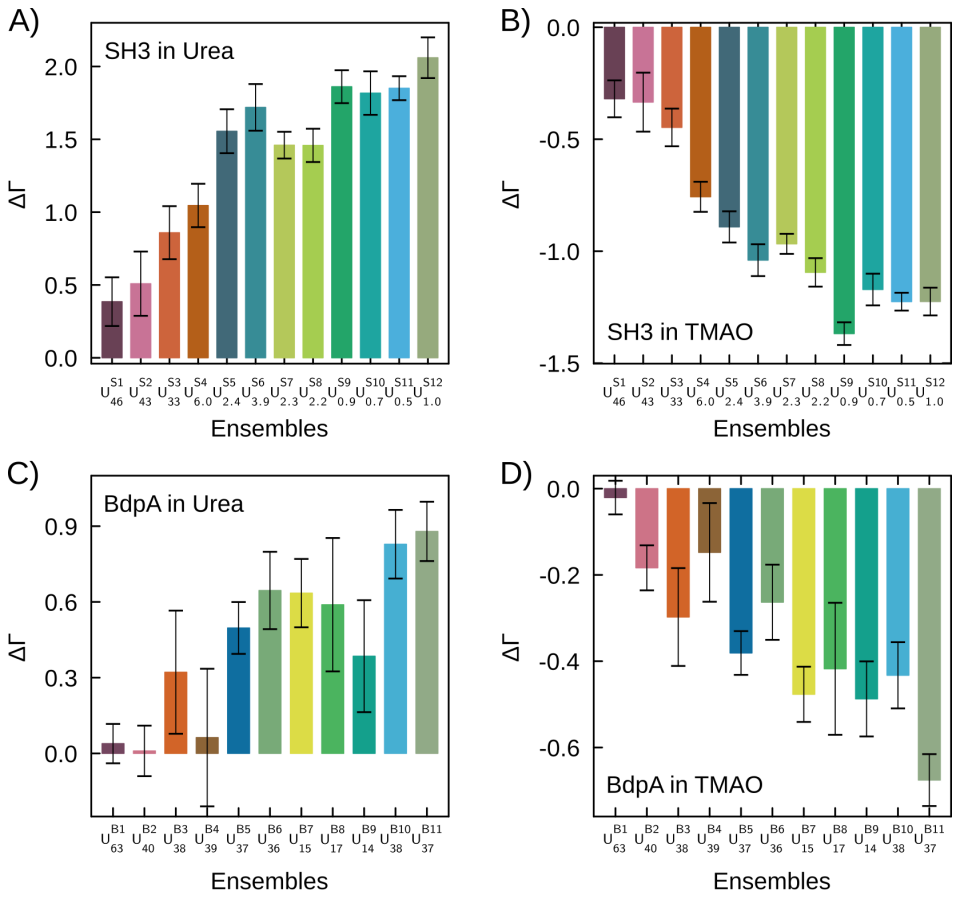

Figure 6. ΔΓ for each folding ensemble of structures of SH3 and BdpA proteins, relative to the native-state clusters. Panels A) and B) show ΔΓ values for each ensemble of SH3 in urea and TMAO 0.5 mol L^{-1}. Similarly, panels C) and D) show ΔΓ values for each ensemble of BdpA under the same cosolvent conditions.

Figure 6 displays ΔΓ for SH3 and BdpA in urea and TMAO at 0.5 mol L^{-1}, considering the equilibrium involving the native states and each of the unfolded states. The denatured states of SH3 are always stabilized by urea addition (Figure 6A - ΔΓ positive) and destabilized by TMAO (Figure 6B). For BdpA, ΔΓ is also always positive

for urea, although within the modeling uncertainties for the first four partially unfolded states. Finally, TMAO favors the native state of BdpA ($\Delta\Gamma$ negative) relative to all unfolded states, with perhaps the exception of U_{63}^{B1}. In all cases, the osmolyte effects are more prominent as the degree of denaturation increases.

Thus, in general, increasing the concentration of urea favors denatured states, while the native state is favored by TMAO, consistently with the overall known roles of these osmolytes on protein stability[51]. Nevertheless, the rule breaks for denatured states displaying similar solvent accessible surface area than the native state. The linkage relation, Equation 1, does not allow, however, obtaining the absolute dependence of the equilibrium constant of denaturation (or, equivalently, on the denaturation free energy) as a function of the cosolvent concentration. It is necessary to integrate the concentration-dependence of the free energy for each state from pure water to the final ternary solution to obtain the net effect of the cosolvent on the folding equilibrium. This computation was performed for selected states, as described below.

Osmolyte dependence of unfolding free energies for selected denatured states

The absolute effect of a cosolvent on the protein folding equilibrium can be determined by the difference of equilibrium constants of folding at each solvent concentration, relative to water. This corresponds to the variation of the free energy of folding upon cosolvent addition. Experimentally, these were measured by NMR with various osmolytes, including urea and TMAO[19], for the SH3 domain from the *Drosophila signal transduction protein* (Drk-SH3), which shares the same fold of the SH3 model here studied[52]. Alternatively, if the folding states can be stabilized in the time-scale of the experiment, transfer free energies of each state from water to the cosolvent solution can be obtained, from which the same concentration dependence of the folding free energy

can be obtained[11]. These two approaches correspond to the reaction pairs of the Tanford Transfer thermodynamic cycle[20,53], shown in Figure 7.

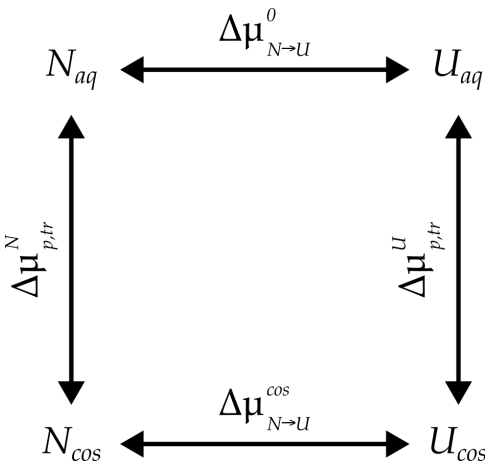

Figure 7. Thermodynamic cycle of the Tanford Transfer Model[20,53]. This cycle illustrates how the folding free energies between the native (N) and unfolded (U) states in the presence of a cosolvent ($N_{cos} \rightleftharpoons U_{cos}$, given by $\Delta\mu^{cos}_{N\to U}$) differs from that in water ($N_{aq} \rightleftharpoons U_{aq}$, given by $\Delta\mu^{0}_{N\to U}$). Direct measures of equilibrium constants with and without cosolvent (horizontal reactions) allow the computation of the cosolvent effect on the folding free energy, $\partial\Delta\mu_{p,tr} = \Delta\mu^{cos}_{N\to U} - \Delta\mu^{0}_{N\to U}$. Alternatively, the transfer free energies for the native ($\Delta\mu^{N}_{p,tr}$) and unfolded ($\Delta\mu^{U}_{p,tr}$) states can be obtained (vertical reactions), and $\partial\Delta\mu_{p,tr} = \Delta\mu^{U}_{p,tr} - \Delta\mu^{N}_{p,tr}$.

Since computational modeling of the folding equilibrium of even small proteins such as SH3 and BdpA is difficult, the effect of the cosolvents on the folding equilibrium will be estimated by computing transfer free energies[9,11,16,54]. The transfer free energy, $\Delta\mu_{p,tr}$, of a structure from pure water to a solution of the cosolvent is

$$\Delta\mu_{p,tr} = \int_0^{m_c} \frac{\partial\mu_p}{\partial m_c} dm_c \quad (2)$$

where $\frac{\partial\mu_p}{\partial m_c}$ is the variation of the chemical potential of the structure with cosolvent concentration, and the integral is performed from pure water to the desired m_c concentration of the cosolvent. This computation mimics the experiment of adding a cosolvent to an aqueous solution of a protein while roughly preserving the protein folding state, either folded or an unfolded construct[11].

Transfer free energies can be obtained by numerical integration of Equation 2, with the integrand computed from preferential interaction parameters and activities obtained at various cosolvents concentrations using[11],

$$\left(\frac{\partial\mu_p}{\partial m_c}\right)_{T,P,m_p} = -(\Gamma)_{T,P,\mu_c}\left(\frac{RTM_p}{M_c}\right)\left(\frac{1}{m_c} + \frac{\partial\ln\gamma_c}{\partial m_c}\right) \quad (3)$$

where γ_c is the cosolvent activity, and M_p and M_c are the molar masses of the protein and the cosolvent, respectively. The activities of urea and TMAO can be obtained from experimental data[11], and the preferential interaction parameters can be obtained by simulating a protein model in different cosolvent concentrations.

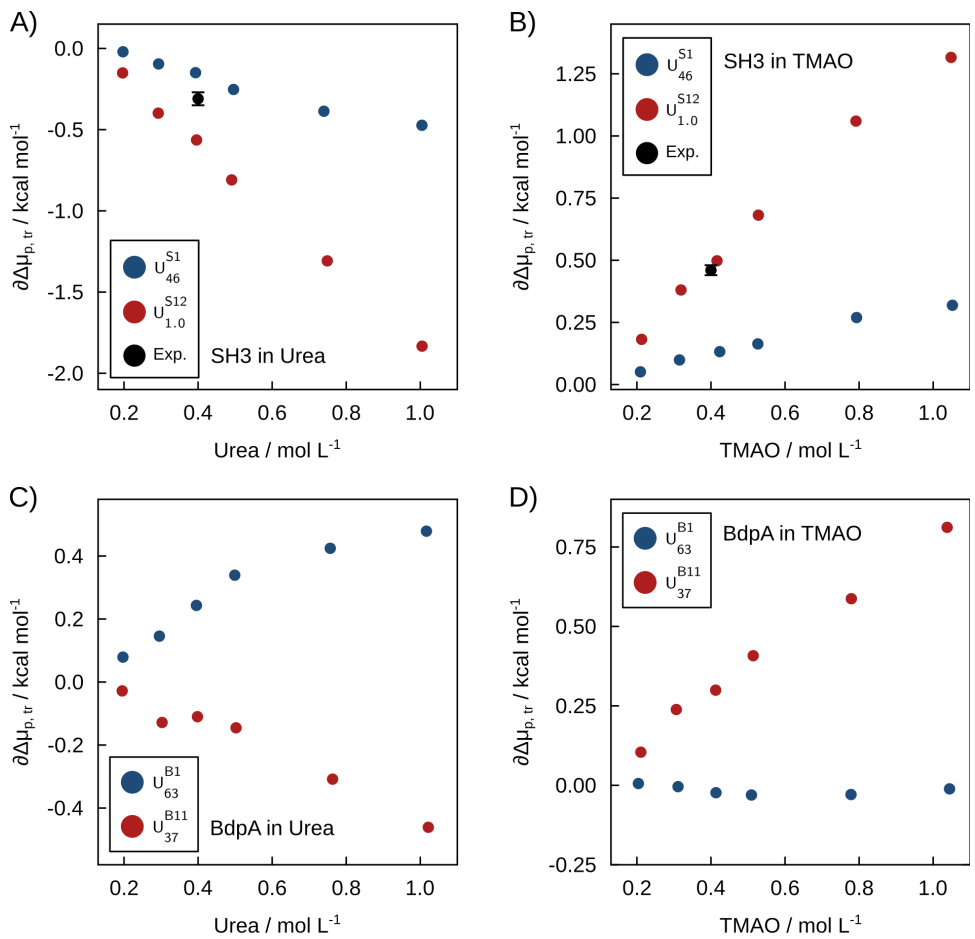

Figure 8. Variation of folding free energies ($\partial\Delta\mu_{p,tr}$) relative to pure water as a function of cosolvent concentration, for selected denatured states. A) and B) $\partial\Delta\mu_{p,tr}$ for SH3 models in urea and TMAO, respectively. This analysis focused on the U_{46}^{S1} (blue) and $U_{1.0}^{S12}$ (red) unfolded ensembles. C) and D) $\partial\Delta\mu_{p,tr}$ for BdpA for the folding equilibrium associated to the U_{63}^{B1} (blue) and U_{37}^{B11} (red) BdpA ensembles.

We focus on three folding states for each protein: the native states (N_{80}^{S0} for SH3 and N_{84}^{B0} for BdpA) and two unfolded states. The unfolded states were chosen from the $\Delta\Gamma$ extremes in urea or TMAO at 0.5 mol L^{-1} (Figure 6). For SH3, we selected U_{46}^{S1} and

$U_{1.0}^{S12}$, while for BdpA, U_{63}^{B1} and U_{37}^{B11} were selected (see Figures 1 and 2). Simulations were performed for the most representative structure of each ensemble in 7 concentrations of urea and TMAO ranging from 0.1 to 1 mol L^{-1}. Solution structures were characterized using MDDFs and the KB theory, as previously described, to obtain $\frac{\partial \mu_p}{\partial m_c}$ using Equation 3. Preferential interactions obtained are shown in Supplementary Figure S15. The transfer free energy of each model ($\Delta\mu_{p,tr}$) was then obtained from the numerical integration of Equation 2. Finally, the effect of the cosolvent in the folding free energies are $\delta\Delta\mu_{p,tr}^{U-N} = \Delta\mu_{p,tr}^{U} - \Delta\mu_{p,tr}^{N}$, the difference in transfer free energies associated with the N ⇌ U transition, for each unfolded state.

Figure 8A shows that the variation in unfolding free energy of SH3 states is negative for all urea concentrations, and thus the denatured states are favored. $U_{1.0}^{S12}$ is a highly unfolded state, with a large surface area, and consistently becomes stabilized by urea to a greater extent than the more compact U_{46}^{S1} model. In the presence of TMAO (Figure 8B), a positive $\partial\Delta\mu_{p,tr}$ implies destabilization SH3 unfolded states, consistently with the "osmophobic effect"[14]. $\partial\Delta\mu_{p,tr}$ for U_{46}^{S1} is smaller than for $U_{1.0}^{S12}$ because TMAO favors compact structures.

Experimentally, the free energy of Drk-SH3 folding equilibrium was shown to be affected by urea, at 0.4 mol L^{-1}, by -0.31 ± 0.02 kcal mol^{-1} [19]. At this concentration, the computed variations of unfolding free energies of SH3 upon urea addition are -0.15 and -0.56 kcal mol^{-1} for the U_{46}^{S1} and $U_{1.0}^{S12}$ states, respectively. Similarly, the experimentally determined variation in folding free energy in 0.4 mol L^{-1} TMAO solutions was 0.46 ± 0.02 kcal mol^{-1}, and here we obtain 0.13 and 0.50 kcal mol^{-1}. Experimental values are, thus, within the lower and upper bounds obtained for the effect of the cosolvents on the folding equilibria. This qualitative agreement in the sub-kcal mol^{-1} energy regime is a

remarkable validation of the folding ensemble obtained by the simulations, the force-fields used for representing solvent-protein interactions, and the methods used for solvent structure analysis. The experimental $\delta\Delta\mu_{p,tr}^{U-N}$ are shown as black dots in Figure 8.

A similar analysis of cosolvent effects on the folding equilibria of BdpA ensembles U_{63}^{B1} and U_{37}^{B11} is shown in Figures 8C and 8D. U_{63}^{B1} has positive $\partial\Delta\mu_{p,tr}^{U-N}$ in urea and virtually zero in the presence of TMAO. For the U_{37}^{B11} state, with greater denaturation and surface area, folding free energies confirm the denaturant and protector effects of urea and TMAO (Figures 8C and 8D - red dots). In the previous section, we demonstrated that the solvation characteristics of the U_{63}^{B1} set are similar to those of the native state of BdpA. Here, we show that this is a particular case where urea exerts a destabilizing effect on a partially unfolded state, while in TMAO the unfolding equilibrium involving this state is essentially independent of the cosolvent concentration. This observation was only partially deducible from the Wyman linkage relations in Figures 6C and 6D, considering the standard error of $\Delta\Gamma$ values. We cannot obtain the experimental population ratio of U_{63}^{B1}-like states vs native states from our simulations, but in our SBM model U_{63}^{B1} was twice as populated as N_{84}^{B0}. This implies a free energy difference of roughly -0.4 kcal mol^{-1} at 300K. The urea effect on the folding equilibrium, computed here, is in the same order of magnitude, but favoring the native state. Thus, it could significantly reduce the population of the partially denatured helix I, redirecting unfolding along new pathways.

Discussion

In this work, we developed a computational approach to characterize the effect of osmolytes on protein folding ensembles. Our pipeline combines SBMs and all-atom simulations to model the folding landscapes of SH3 and BdpA, and the solute-solvent

interactions of every conformation. To elucidate the solvation structures and quantify the stabilization promoted by urea and TMAO, we employed MDDFs, which are crucial for the characterization of solvation of complex molecular shapes. Using these, we applied the Kirkwood-Buff theory to obtain preferential interactions and the osmolyte dependence of folding free energies.

Our results confirm urea's preferential interaction and TMAO's preferential exclusion from protein surfaces, while adding a new dimension to how these interactions vary across the protein folding landscape. Different denatured states exhibit distinct interactions with water and cosolvents, significantly correlating with solvent-accessible surface area. This suggests that globular unfolded states may be only marginally affected by osmolytes, with urea potentially promoting their stabilization. MDDFs reveal that water, urea, and TMAO form both H-bonds and non-specific interactions with the protein surface (Figures 2A-B and 3A-B). Specific interactions are stronger in the native states, whereas the non-specific interactions are more prevalent in the denatured states. These patterns hold across different solvent components, protein types, and folds, reflecting the increased exposure of hydrophobic core residues upon denaturation.

Non-specific interactions between urea and the protein govern the preferential interactions, as shown by the quantitative analysis of KBIs (Figures 2F and 3F). As these interactions strengthen with increasing protein surface exposure, unfolded states become more dehydrated. In contrast, both specific and non-specific interactions of TMAO are weak, leading to preferential hydration, which is further enhanced in denatured states.

Solvent exposure, fraction of native contacts, and the secondary structure elements (β-sheet and α-helix) are key determinants of preferential interactions. In the case of SH3, for which these properties are highly correlated, preferential interactions and folding free energies vary monotonically with increasing denaturation. In the case

of BdpA, on the other side, the secondary structure content does not directly correlate with the solvent accessible surface area or the fraction of native contacts. Thus, denatured states like U_{63}^{B1}, U_{40}^{B2}, and U_{39}^{B4} exhibit solvation properties similar to the native state, despite their lower helical content.

Simulating representative structures of selected folding states in various TMAO and urea concentrations allowed us to compute the absolute osmolyte effect on the folding free energies. Strikingly, the experimentally determined folding free energy dependence on TMAO for an SH3 homologous protein is found within the computed bounds, validating the force-fields, modeling strategy, and solvation analysis methods. At the same time, the solvation of a single denatured state is not sufficient to infer quantitatively the role of cosolvents in the folding equilibrium, consistently with the unfolded ensemble variability playing a key role[55]. For SH3, the concentration dependence of the folding free energies is consistent with urea-induced denaturation and TMAO-protection. In contrast, BdpA exhibits unfolded states that can be either stabilized by urea or unaffected by TMAO, illustrating the complexity of osmolyte effects on proteins with multiple unfolding pathways.

Methods

Structure-Based Models (SBMs) simulations and analyses

Folding ensembles of the SH3 domain (PDB: 1FMK[21]) and BdpA (PDB: 1BDD[22]) were generated from Cα-Structure-Based Models simulations[56,57]. The contact map of the native structures was determined using the *Contact of Structural Units* (CSU) algorithm[58], resulting in 140 and 98 contacts, respectively. The Cα-SBMs were generated using the SMOG web server (https://smog-server.org/)[59] and simulations were performed with Gromacs 4.6.7[60].

The folding ensemble described for BdpA in our previous publication was used in this work[37]. A similar protocol was adopted here to obtain the SH3 folding ensemble. For the SH3 protein, the simulations were conducted in temperatures ranging from 110 to 170 units, with a temperature step of 10 units. Simulations at constant temperatures consisted of 5×10^8 steps with a time step of 0.0005 reduced units. Once the temperature of maximum specific heat, C_v, was roughly identified, a new set of simulations with temperatures varying between 135 and 139 with 1 unit temperature steps was performed, to localize within ~1 unit the folding temperature (T_f). In reduced temperature units, the simulations were performed within 0.91 and 1.41 with 0.083 steps and within 1.12 and 1.15 with 0.0083 temperature unit steps for the first and second sets of simulations. The simulation at the T_f was extended ten times to increase the sampling of conformational transitions.

The temperature dependence of the specific heat, $C_v(T)$, and the potential of mean force as a function of the fraction of native contacts, F(Q), of these simulations were obtained with WHAM[61], as implemented in SMOG2[62]. A contact was classified as native if the distance between the corresponding Cα atoms did not exceed 20% of the value observed in the experimentally determined models. The statistical convergence of the Q-values at the T_f was confirmed by block-averaging using the MolSimToolkit.jl package (Supplementary Figure S2).

Protein folding phase-space

The *Energy Landscape Visualization Method* (ELViM)[30,38,63–67] was used to represent the folding phase space of SH3 and BdpA. The ELViM method uses matrices of internal distances of conformations to establish a reliable structural similarity metric without predefined reaction coordinates. The similarity matrix is projected into a 2D space optimizing the distances between points to be correlated with the dissimilarity between structures. For this projection, we apply the Force Scheme technique[68], as implemented

in ELViM[30]. The secondary structures were computed with the DSSP method,[69,70] using the ProteinSecondaryStructures.jl package.

Atomistic simulations and analysis

All-atom representations of 5,000 SBM models for each protein, extracted from the simulations performed at the T_f were reconstructed using the Pulchra software[71]. We have studied the solvation of each reconstructed structure in solutions of urea 0.5 mol L^{-1} and TMAO 0.5 mol L^{-1}.

Orthorhombic simulation boxes were constructed with a minimum distance of 12.0 Å from the protein extrema using Packmol[72,73]. Details of each system, including the box volume and the number of solvent molecules, are shown in Supplementary Tables S2-S5. Each conformation was constrained by applying harmonic potentials with 25 kcal mol^{-1} force constants on Cα atoms, such that the Cα-SBM topology was preserved while allowing relaxation of the reconstructed atoms and the solvent. The CHARMM36 force field[74] for the protein and the TIP3P water model[75] were used. Urea molecules were modeled with the combination of CGenFF[76,77] and the charges of the KBFF force-field, which was developed to reproduce the KBIs of urea and water molecules[78]. TMAO was described using the Netz force field[79] with the scaling correction proposed by Shea and co-workers[80], designed to accurately reproduce the Kirkwood-Buff integrals (KBI) of TMAO in ternary solutions containing water, TMAO, and ribonuclease T1 (RNase T1)[11].

All atomistic simulations were performed in Gromacs 2021.2[81] at 298.15 K, 1 atm, and with a time step of 2 fs. Initially, the systems were minimized by up to 20,000 Steepest Descent steps and equilibrated for 1 ns in constant-volume and constant-temperature ensemble (NVT) followed by 1 ns of constant-volume and constant-pressure (NPT) simulation. Temperature and pressure were controlled using the modified Berendsen thermostat[82] and Parrinello-Rahman barostat.[83] Finally, 10 ns production simulations were performed in the NPT ensemble for each system, totaling

240 μs of simulation, used to obtain MDDFs, KBIs and preferential interaction parameters of Figures 2 to 5, and the data of Figure 6.

Simulations in urea and TMAO at 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 1.0 mol L^{-1}, of the 6 selected representative structures of BdpA and SH3 ensembles, were performed to determine the transfer free energies used to obtain Figure 8. (see the next section for more details). 20 replicas of 25 ns were performed for each sytem, totalling 42 μs of simulation (Supplementary Tables S4 and S5). Representative structures were selected by the *most_representative_structure* function in the MolSimToolkit.jl package version 1.23.0. The most representative structure has the minimum RMSD relative to the average structure of each set.

Characterization of the solvent structure and thermodynamics

This study employs a notation system for clarity: w (water), p (protein), c (cosolvent), and s (water or cosolvent). We developed the ComplexMixtures.jl package[42] to characterize the solvation structures from Minimum-Distance Distribution Functions (MDDFs)[24] and the Kirkwood-Buff theory of solvation[84]. Specifically, the package computes the distribution functions, Kirkwood-Buff Integrals (KBIs), and atomic, group, and residue decompositions required for a molecular understanding of solvation. Density maps like that of Figure 2F were generated with the *ResidueContributions* function of ComplexMixtures.jl. From the KBIs, preferential interaction parameters (Γ) and the Wyman-linkage relation (Equation 1) can be obtained[43]. The computation of transfer free energies followed the protocol described by Lin and Timasheff[11]. The combination of MDDFs, Γ, ΔΓ, and Δμ parameters comprehensively characterizes the solvation of the folding ensembles.

MDDFs are defined in terms of the average number density of solvent molecules $n_c(r)$ in the simulation, relative to the density of an ideal gas distribution $n_c^*(r)$,

$$MDDF(r) = \frac{n_s(r)}{n_s^*(r)} \quad (4)$$

where r is the minimum-distance between the protein and the cosolvent. Further discussion about MDDFs and their importance in the study of solvent structures are found in your previous publications[18,24,42,85].

With proper normalization, the MDDF is used to compute the Kirkwood-Buff integrals (KBIs) and preferential interaction parameters (Γs). In a finite sub-volume of the system, the (KBIs) is given by Equation 5.

$$G_{ps}(R) = \frac{1}{\rho_s}[N_{ps}(R) - N_{ps}^*(R)] \quad (5)$$

where $N_{ps}(R)$ and $N_{ps}^*(R)$ are the number of minimum distances between the protein and the solvent within R, and the number of equivalent distances within R in an ideal system. The KBIs are the limit of $G_{ps}(R)$ for large R, and are the excess volume occupied by the cosolvent relative to the volume that it would occupy in the absence of solute-solvent interactions.

The thermodynamic property that measures if the cosolvent accumulates or is depleted from the protein domain, compared with the bulk solution, is the preferential interaction parameter (Γ). In the infinitely dilute solute limit, the cosolvent preferential binding to the protein (Γ_{pc}) can be approximated by site counts in a finite volume defined by R, with

$$\Gamma_{pc} \approx \rho_c[G_{pc}(R) - G_{pw}(R)] \quad (6)$$

where R is large enough to encompass the portion of the solution affected by the presence of the solute.

Given Γ_{pc} for two states, the Wyman linkage equation[43] (Equation 1) is used to obtain the direction into which a cosolvent will drive the reaction at a given cosolvent concentration. Transfer free energies are computed using the numerical integration of Equation 2, using Equation 3 as the integrand, from a finite set of simulations of the states considered in different concentrations of the cosolvent. Notebooks with reproducible calculations of the transfer free energies shown in Figure 8 are available at https://github.com/m3g/PereiraMartinez2024.jl. The average of preferential interaction parameters from 20 simulations, calculated across seven cosolvent concentrations, are used to determine $\partial\Delta\mu_{p,tr}$, and are summarized in Supplementary Figure S15. The Plots.jl[86] package was used to create figures.

Author information

Authors and Affiliations

Institute of Chemistry and Center for Computing in Engineering & Science, Universidade Estadual de Campinas (UNICAMP), 13083-861 Campinas, SP, Brazil

Ander Francisco Pereira and Leandro Martínez

Contributions

A. F. P. proposed strategies for, and executed, simulations and data analysis, and wrote the paper. L. M. conceived the project, supervised research, developed software, and wrote the paper.

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary Information

Structural characterization of folding ensembles (Table S1); SBM folds space characterization (Figures S1 and S2); Representative structure of each fold subset (Figure S3); Secondary structure content per residue in each subset (Figures S4 and S5). Correlation between Γ and SASA (Figure S6); Density maps for water (Figure S7 and S10); MDDFs for all ensembles for SH3 and TMAO (Figure S8) and group contributions (Figure S9); Residue contributions differential density maps for all states, urea and TMAO (Figures S11-14). Γ for selected folds at multiple concentrations (Figure S15); Simulation box details (Tables S2 to S5).

Data Availability

Data and codes necessary for the computations here reported and detailed reproduction of the computation of transfer free energies are available at https://github.com/m3g/PereiraMartinez2024.jl.

Acknowledgments

The authors acknowledge the financial support of Fapesp (2010/16947-9, 2018/24293-0, 2013/08293-7, 2018/14274-9, 2019/17874-0, 2020/04549-0), and CAPES 206 -04/092018. Research developed with the help of CENAPAD-SP (National Center for High-Performance Processing in São Paulo), project UNICAMP / FINEP - MCTI, and the Coaraci Supercomputer.

References

1. Sharma, G. S. *et al.* Protecting thermodynamic stability of protein: The basic paradigm against stress and unfolded protein response by osmolytes. *Int J Biol Macromol* **177**, 229–240 (2021).
2. Rani, A. & Venkatesu, P. Correction: Changing relations between proteins and osmolytes: a choice of nature. *Phys Chem Chem Phys* **20**, 23151 (2018).
3. Wang, W. & Ohtake, S. Science and art of protein formulation development. *Int J Pharm* **568**, 118505 (2019).
4. Rabbani, G. & Choi, I. Roles of osmolytes in protein folding and aggregation in cells and their biotechnological applications. *Int J Biol Macromol* **109**, 483–491 (2018).
5. Hasan, T. *et al.* Osmolytes in vaccine production, flocculation and storage: a critical review. *Hum Vaccin Immunother* **15**, 514–525 (2019).
6. Velasquez, M. T., Ramezani, A., Manal, A. & Raj, D. S. Trimethylamine N-Oxide: The Good, the Bad and the Unknown. *Toxins (Basel)* **8**, (2016).
7. Boob, M. M., Sukenik, S., Gruebele, M. & Pogorelov, T. V. TMAO: Protecting proteins from feeling the heat. *Biophys J* **122**, 1414–1422 (2023).
8. Pepelnjak, M. *et al.* In situ analysis of osmolyte mechanisms of proteome thermal stabilization. *Nat Chem Biol* **20**, 1053–1065 (2024).
9. Bolen, D. W. & Rose, G. D. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **77**, 339–362 (2008).
10. Canchi, D. R., Jayasimha, P., Rau, D. C., Makhatadze, G. I. & Garcia, A. E. Molecular mechanism for the preferential exclusion of TMAO from protein surfaces. *J. Phys. Chem. B*

116, 12095–12104 (2012).

11. Lin, T. Y. & Timasheff, S. N. Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry* **33**, 12695–12701 (1994).
12. Bennion, B. J. & Daggett, V. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5142–5147 (2003).
13. Khan, S., Siraj, S., Shahid, M., Haque, M. M. & Islam, A. Osmolytes: Wonder molecules to combat protein misfolding against stress conditions. *Int J Biol Macromol* **234**, 123662 (2023).
14. Bolen, D. W. & Baskakov, I. V. The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.* **310**, 955–963 (2001).
15. Peng, X. *et al.* Prediction and Validation of a Protein's Free Energy Surface Using Hydrogen Exchange and (Importantly) Its Denaturant Dependence. *J Chem Theory Comput* **18**, 550–561 (2022).
16. Timasheff, S. N. Solvent stabilization of protein structure. *Methods Mol. Biol.* **40**, 253–269 (1995).
17. Piccoli, V. & Martínez, L. Ionic liquid solvation of proteins in native and denatured states. *J. Mol. Liq.* **363**, 119953 (2022).
18. Pereira, A. F., Piccoli, V. & Martínez, L. Trifluoroethanol direct interactions with protein backbones destabilize α-helices. *J. Mol. Liq.* **365**, 120209 (2022).
19. Rydeen, A. E., Brustad, E. M. & Pielak, G. J. Osmolytes and protein-protein interactions. *J. Am. Chem. Soc.* **140**, 7441–7444 (2018).
20. Tanford, C. Protein Denaturation. in *Advances in Protein Chemistry* 1–95 (Elsevier, 1970).

21. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).

22. Gouda, H. *et al.* Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. *Biochemistry* **31**, 9665–9672 (1992).

23. Noel, J. K. & Onuchic, J. N. The many faces of structure-based potentials: From protein folding landscapes to structural characterization of complex biomolecules. in *Computational Modeling of Biological Systems* 31–54 (Springer US, Boston, MA, 2012).

24. Martínez, L. & Shimizu, S. Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions. *J. Chem. Theory Comput.* **13**, 6358–6372 (2017).

25. Onuchic, J. N., Nymeyer, H., García, A. E., Chahine, J. & Socci, N. D. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* **53**, 87–152 (2000).

26. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).

27. Lammert, H., Schug, A. & Onuchic, J. N. Robustness and generalization of structure-based models for protein folding and function. *Proteins* **77**, 881–891 (2009).

28. Noel, J. K. *et al.* Mirror images as naturally competing conformations in protein folding. *J. Phys. Chem. B* **116**, 6880–6888 (2012).

29. Whitford, P. C. *et al.* An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430–441 (2009).

30. Oliveira, A. B., Jr, Yang, H., Whitford, P. C. & Leite, V. B. P. Distinguishing Biomolecular Pathways and Metastable States. *J. Chem. Theory Comput.* **15**, 6482–6490 (2019).

31. Noel, J. K., Whitford, P. C. & Onuchic, J. N. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* **116**, 8692–8702 (2012).

32. Demakis, C., Childers, M. C. & Daggett, V. Conserved patterns and interactions in the unfolding transition state across SH3 domain structural homologues. *Protein Sci* **30**, 391–407 (2021).

33. Borreguero, J. M., Ding, F., Buldyrev, S. V., Stanley, H. E. & Dokholyan, N. V. Multiple folding pathways of the SH3 domain. *Biophys J* **87**, 521–533 (2004).

34. Borreguero, J. M., Dokholyan, N. V., Buldyrev, S. V., Shakhnovich, E. I. & Stanley, H. E. Thermodynamics and folding kinetics analysis of the SH3 domain form discrete molecular dynamics. *J Mol Biol* **318**, 863–876 (2002).

35. Gsponer, J. & Caflisch, A. Molecular dynamics simulations of protein folding from the transition state. *Proc Natl Acad Sci U S A* **99**, 6719–6724 (2002).

36. Tsai, J., Levitt, M. & Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J Mol Biol* **291**, 215–225 (1999).

37. Pereira, A. F. & Martínez, L. Helical Content Correlations and Hydration Structures of the Folding Ensemble of the B Domain of Protein A. *J. Chem. Inf. Model.* **64**, 3350–3359 (2024).

38. Viegas, R. G. *et al.* ELViM: Exploring Biomolecular Energy Landscapes through Multidimensional Visualization. *J Chem Inf Model* **64**, 3443–3450 (2024).

39. García, A. E. & Onuchic, J. N. Folding a protein in a computer: an atomic description of the

folding/unfolding of protein A. *Proc Natl Acad Sci U S A* **100**, 13898–13903 (2003).

40. Khalili, M., Liwo, A. & Scheraga, H. A. Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. *J Mol Biol* **355**, 536–547 (2006).
41. Zou, Q., Bennion, B. J., Daggett, V. & Murphy, K. P. The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *J Am Chem Soc* **124**, 1192–1202 (2002).
42. Martínez, L. ComplexMixtures.jl: Investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective. *J. Mol. Liq.* **347**, 117945 (2022).
43. Timasheff, S. N. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc Natl Acad Sci U S A* **99**, 9721–9726 (2002).
44. Hua, L., Zhou, R., Thirumalai, D. & Berne, B. J. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A* **105**, 16928–16933 (2008).
45. Okuno, Y. *et al.* Atomic view of cosolute-induced protein denaturation probed by NMR solvent paramagnetic relaxation enhancement. *Proc Natl Acad Sci U S A* **118**, (2021).
46. Courtenay, E. S., Capp, M. W., Anderson, C. F. & Record, M. T., Jr. 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 (2000).
47. Wang, A. & Bolen, D. W. A naturally occurring protective system in urea-rich cells:

mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* **36**, 9101–9108 (1997).

48. Roy, S. & Bagchi, B. Comparative study of protein unfolding in aqueous urea and dimethyl sulfoxide solutions: surface polarity, solvent specificity, and sequence of secondary structure melting. *J. Phys. Chem. B* **118**, 5691–5697 (2014).
49. Parui, S. & Jana, B. Relative Solvent Exposure of the Alpha-Helix and Beta-Sheet in Water Determines the Initial Stages of Urea and Guanidinium Chloride-Induced Denaturation of Alpha/Beta Proteins. *J. Phys. Chem. B* **123**, 8889–8900 (2019).
50. Wyman, J. Linked Functions and Reciprocal Effects in Hemoglobin: A Second Look. *Adv. Prot. Chem.* **19**, 223–286 (1964).
51. Canchi, D. R. & García, A. E. Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.* **64**, 273–293 (2013).
52. Bezsonova, I., Singer, A., Choy, W.-Y., Tollinger, M. & Forman-Kay, J. D. Structural comparison of the unstable drkN SH3 domain and a stable mutant. *Biochemistry* **44**, 15550–15560 (2005).
53. Tanford, C. Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* **86**, 2050–2059 (1964).
54. Tanford, C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84**, 4240–4247 (1962).
55. Ziv, G. & Haran, G. Protein folding, protein collapse, and tanford's transfer model: lessons from single-molecule FRET. *J Am Chem Soc* **131**, 2942–2947 (2009).
56. Taketomi, H., Ueda, Y. & Gō, N. Studies on protein folding, unfolding and fluctuations by

computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **7**, 445–459 (1975).

57. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
58. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* **15**, 327–332 (1999).
59. Noel, J. K., Whitford, P. C., Sanbonmatsu, K. Y. & Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **38**, W657–61 (2010).
60. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
61. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
62. Noel, J. K. *et al.* SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **12**, e1004794 (2016).
63. Oliveira Junior, A. B. *et al.* Exploring Energy Landscapes of Intrinsically Disordered Proteins: Insights into Functional Mechanisms. *J. Chem. Theory Comput.* **17**, 3178–3187 (2021).

64. Dias, R. V. R. *et al.* Unveiling Metastable Ensembles of GRB2 and the Relevance of Interdomain Communication during Folding. *J. Chem. Inf. Model.* **63**, 6344–6353 (2023).

65. Sanches, M. N. *et al.* Resolving the fine structure in the energy landscapes of repeat proteins. *QRB Discov* **3**, e7 (2022).

66. Viegas, R. G. *et al.* Characterizing the Folding Transition-State Ensembles in the Energy Landscape of an RNA Tetraloop. *J. Chem. Inf. Model.* **63**, 5641–5649 (2023).

67. da Silva, F. B., Simien, J. M., Viegas, R. G., Haglund, E. & Leite, V. B. P. Exploring the folding landscape of leptin: Insights into threading pathways. *J. Struct. Biol.* **216**, 108054 (2023).

68. Hardin, C., Eastwood, M. P., Prentiss, M. C., Luthey-Schulten, Z. & Wolynes, P. G. Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1679–1684 (2003).

69. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

70. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–9 (2011).

71. Rotkiewicz, P. & Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **29**, 1460–1465 (2008).

72. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).

73. Martínez, J. M. & Martínez, L. Packing optimization for automated generation of complex

system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **24**, 819–825 (2003).

74. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
75. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* vol. 79 926–935 Preprint at https://doi.org/10.1063/1.445869 (1983).
76. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* NA–NA Preprint at https://doi.org/10.1002/jcc.21367 (2009).
77. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* vol. 33 2451–2468 Preprint at https://doi.org/10.1002/jcc.23067 (2012).
78. Weerasinghe, S. & Smith, P. E. A Kirkwood−Buff derived force field for mixtures of urea and water. *J. Phys. Chem. B* **107**, 3891–3898 (2003).
79. Schneck, E., Horinek, D. & Netz, R. R. Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations. *J Phys Chem B* **117**, 8310–8321 (2013).
80. Ganguly, P. *et al.* Cosolvent Exclusion Drives Protein Stability in Trimethylamine -Oxide and Betaine Solutions. *J Phys Chem Lett* **13**, 7980–7986 (2022).
81. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

82. 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.* **81**, 3684–3690 (1984).

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

84. Newman, K. E. Kirkwood–Buff solution theory: derivation and applications. *Chem. Soc. Rev.* **23**, 31–40 (1994).

85. Piccoli, V. & Martínez, L. Correlated counterion effects on the solvation of proteins by ionic liquids. *J. Mol. Liq.* **320**, 114347 (2020).

86. Christ, S., Schwabeneder, D., Rackauckas, C., Borregaard, M. K. & Breloff, T. Plots.Jl – A user extendable plotting API for the Julia programming language. *J. Open Res. Softw.* **11**, (2023).