Computational and Structural Biotechnology Journal 23 (2024) 4077–4087

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

Osmolyte-induced protein stability changes explained by graph theory

Mattia Miotto [a,*,1], Nina Warner [b,1], Giancarlo Ruocco [a,c], Gian Gaetano Tartaglia [d,a], Oren A. Scherman [b,2], Edoardo Milanetti [c,a,2]

[a] Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy
[b] Melville Laboratory for Polymer Synthesis, Yusuf Hamied Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[c] Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00185, Rome, Italy
[d] Department of Biology, Sapienza University, Piazzale Aldo Moro 5, 00185, Rome, Italy

ABSTRACT

Enhanced stabilization of protein structures via the presence of inert osmolytes is a key mechanism adopted both by physiological systems and in biotechnological applications. While the intrinsic stability of proteins is ultimately fixed by their amino acid composition and organization, the interactions between osmolytes and proteins together with their concentrations introduce an additional layer of complexity and in turn, a method of modulating protein stability. Here, we combined experimental measurements with molecular dynamics simulations and graph-theory-based analyses to predict the stabilizing/destabilizing effects of different kinds of osmolytes on proteins during heat-mediated denaturation. We found that (i) proteins in solution with stability-enhancing osmolytes tend to have more compact interaction networks than those assumed in the presence of destabilizing osmolytes; (ii) a strong negative correlation (R = -0.85) characterizes the relationship between the melting temperature T_m and the preferential interaction coefficient defined by the radial distribution functions of osmolytes and water around the protein and (iii) a positive correlation exists between osmolyte-osmolyte clustering and the extent of preferential exclusion from the local domain of the protein, suggesting that exclusion may be driven by enhanced steric hindrance of aggregated osmolytes.

1. Introduction

Rigorous understanding of osmolyte-induced protein stability changes is a long-standing aim across both applied and fundamental biology [11,26]. From an applied perspective, understanding the stabilizing potential of therapeutically inert additives is key to developing sophisticated formulations for labile biologics; from a fundamental standpoint, these interactions offer insight into the remarkable ability of life itself to persist in extreme environments by osmolyte-related mechanisms [27,32,35,44].

Efforts to explain osmolyte-induced protein stability changes have centered around several inter-related mechanisms which can be described generally by the preferential exclusion theory [17,22,40]. This theory, originally proposed by Serge Timasheff in the 1980s, attributes osmolyte-derived stability to the preference for these molecules—largely on account of an unfavorable interaction with the peptide backbone— to be excluded from the protein surface [3,19,50–53]. The thermodynamic drive to minimize solvent-exposed surface area results both in compaction of the native state and enhancement of the energetic barrier to unfolding, manifesting as an increase in protein denaturation temperature, T_m. On the contrary, destabilizing osmolytes such as urea are believed to behave in an opposite manner, inducing a more diffuse structure and lower barrier to unfolding (suppressed T_m) via preferential interaction with the protein [4,11,21].

Despite general acceptance, much of the molecular detail pertaining to the preferential exclusion theory remains poorly understood or hotly debated [11]. For example, the nature of the primary repulsive forces between the protein and solute remains contentious as does the role of osmolyte-mediated perturbations to the solvent network and concomitant perturbations to solvent-protein interactions [7]. Modern computational techniques have helped to answer some of these questions, whilst occasionally raising more. Molecular dynamics (MD) simulations have produced mounting evidence attributing urea-induced denaturation to preferential interaction with the protein, with variable degrees of importance given to secondary mechanisms involving indirect interactions through structural perturbations of the solvent [11,15,28,30,47,48]. Simulations have also shed light on the driving forces behind these interactions, providing evidence for steric and enthalpically motivated preferential exclusion and interaction, respectively [1]. Moreover, molecular understanding of osmolyte-induced stability changes

* Corresponding author.
E-mail address: mattia.miotto@roma1.infn.it (M. Miotto).
[1] These authors contributed equally to the present work.
[2] These authors contributed equally to the present work.

https://doi.org/10.1016/j.csbj.2024.10.014
Received 14 August 2024; Received in revised form 7 October 2024; Accepted 8 October 2024
Available online 28 October 2024

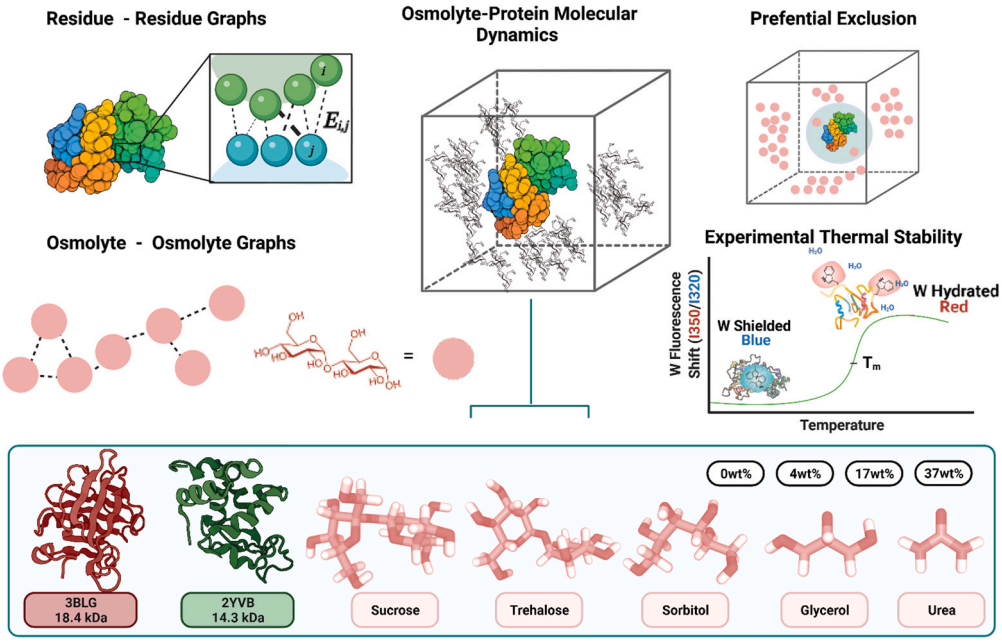

Fig. 1. Study overview and scope. MD simulations were performed for thirty-two systems in total; two model proteins β-Lactoglobulin, BLG, (PDB: 3BLG) and Lysozyme, LYS, (PDB: 2YVB) were studied in aqueous solutions of four common protecting osmolytes: sucrose, trehalose, sorbitol, glycerol, and one denaturing osmolyte: urea. All species were simulated at four concentrations: 0, 4, 17, and 37wt%. Protein thermodynamic stability changes in the presence of osmolytes were measured experimentally by intrinsic tryptophan fluorescence (ITF) shifting. Simulations were analyzed from a graph-theoretical lens with respect to both protein residue-residue interactions and osmolyte-osmolyte interactions as well as by residue dynamics, and osmolyte preferential exclusion coefficients. For more information, see Methods.

has been greatly enriched through MD-based study of water destructuring, hydrogen bonding network perturbations (water-peptide backbone, water-side chain), water-protein/water-water/osmolyte-protein radial distribution functions (RDFs), protein hydration fraction, residue mean square fluctuations (RMSFs), and protein/solvent relaxation times among other parameters [31,33–36,42].

Interestingly, many of these studies have observed self-aggregated clustering behavior amongst stabilizing osmolytes [33,34,36,42]. These clusters have been studied in the absence of protein on various occasions. Lee et al. showed that applying spectral graph analysis to three binary water-osmolyte simulations (water-urea, water-sorbitol, and water-trimethylglycine) could reveal distinct morphological differences between the graph networks formed by the protecting (sorbitol and TMG) and denaturing (urea) osmolyte classes. In the former case, the authors observed the formation of extended networks, descriptively similar to that of water, whereas even at high concentrations, urea failed to form a continuous network and instead assembled only into small, segregated clusters [29]. Similarly, persistent homology-based topological characterization of urea and TMAO aggregates by Mu and colleagues revealed morphological differences in aggregates formed by the chaotropic and protective osmolyte species [2,56]. Seo et al. [45] estimated the microheterogeneity of osmolyte-water mixtures via molecular dynamics simulation, graph theory, and spatial distribution analysis in four osmolyte solutions of trimethylamine-N-oxide (TMAO), tetramethylurea (TMU), dimethyl sulfoxide, and urea. A work by Sundar et al. presented an exhaustive graphical analysis of seven aqueous osmolyte solutions and found no significant perturbations to the water network when properly accounting for the contribution of water-osmolyte interactions. From their analysis, the authors attributed osmolyte-induced protein stability changes to direct interaction between the protein and osmolyte. Unfortunately, however, the absence of protein in the simulated solutions precluded direct assessment of this hypothesis [49]. Over the past two decades, graph representations of proteins have proved powerful models for unearthing molecular origins of stability. Analyses of residue interaction networks (RINs)— or energy-weighted networks of non-covalent interactions between non-adjacent amino acids (nodes)— have shown success in identifying stability-linked residues, structural elements, and network descriptors [8,16,20,39,43,55]. In addition to providing mechanistic insight into stability, RINs have been used extensively in a predictive context, particularly relating to the effect of mutations on protein stability; indeed, the first example of ab initio structure-based stability predictions were reported in 2019 using a RIN framework [13,37,38].

Encouraged by both the success of RINs in elucidating protein stability at a molecular level and the distinct morphological differences in graphical networks of protecting and denaturing osmolytes, a series of protein-osmolyte systems were investigated from a graph-theoretical perspective [29]. As far as the authors are aware, this is the first graph theoretical study of protein-osmolyte solutions. Herein, the thermal stability of two model proteins in a series of industrially and biologically relevant osmolyte solutions is described. Each system is probed by a combined experimental-computational approach encompassing intrinsic tryptophan fluorescence, molecular dynamics (MD) simulation, and graph analysis of both protein residue interactions and osmolyte interactions. These results are discussed in reference to modern osmolyte theory; a graph-theoretical descriptor is introduced to predict osmolyte-induced stability changes.

2. Results and discussion

To investigate the effects of osmolyte-induced protein thermal stability with generalizability, two model proteins were studied— a mainly alpha species, Lysozyme (LYS, PDB:2YVB), and a mainly beta species, β-Lactoglobulin (BLG, PDB:3BLG). Both model proteins were relatively small (LYS: 14.3 kDa, BLG: 18.4 kDa) in the interest of computational efficiency. LYS and BLG were studied by experiment and MD simulation (100 ns) in the presence of both protecting (sucrose, trehalose, sorbitol, glycerol) and denaturing (urea) osmolyte solutions of 0, 4, 17, and 37 wt%. An overview of the study discussed herein is provided in Fig. 1.

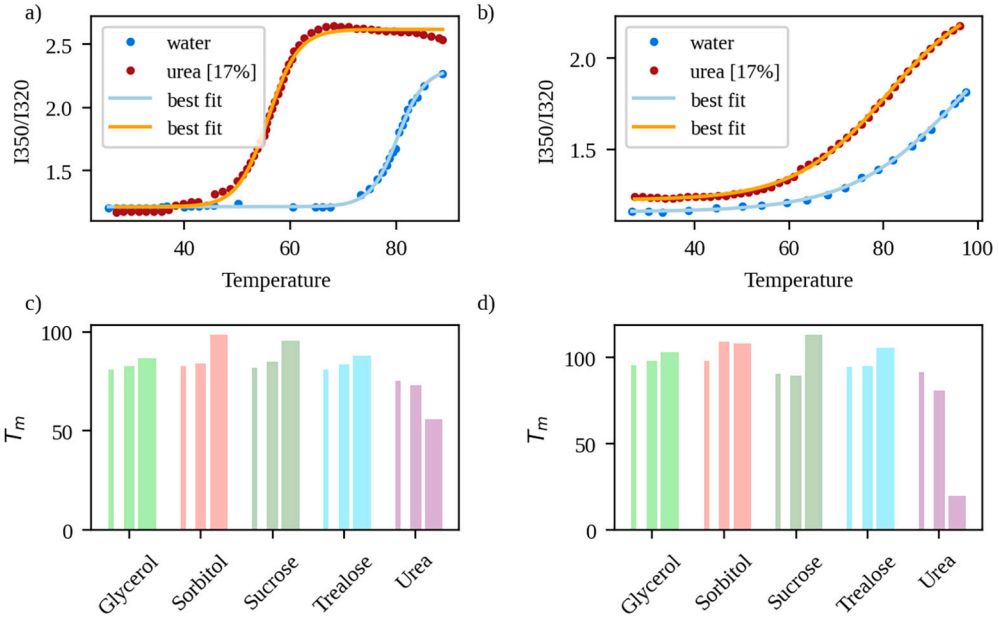

Fig. 2. Experimental stability data for LYS and BLG in the presence of osmolytes. a) Example of experimental ITF unfolding curve for lysozyme in water and in the presence of urea. Dots represent experimental data while continuous lines represent the best-fit solution of Eq. (1). **b)** Same as in panel a, but for BLG ITF unfolding curves. **c)** Measured melting temperatures for lysozyme in the presence of different osmolyte concentrations. **d)** Same as in panel c) but for BLG's T_m. See the Methods section for details on measurement and T_m calculation.

2.1. Characterization of protein thermal stability

Osmolyte-induced thermodynamic protein stability changes were studied by ITF. (see Fig. 2). In particular, Fig. 2a-b report measured ITF for LYS (panel a) and BLG (panel b) in the absence of osmolytes (red dots) and the presence of urea (blue dots) at a relative concentration of 17% with respect to water. Curves exhibit sigmoidal behaviors as a function of the temperature, serving as a proxy of the progressive tryptophane exposure due to protein unfolding. Defining the melting temperature as the one corresponding to the flex of the sigmoid (see Methods and in particular Eq. (1) for details), we found that the melting temperature of lysozyme and βlactoglobulin in the absence of osmolytes are 80.2 ± 0.1 and 96 ± 2 ^{o}C, respectively. In the presence of urea, the melting temperatures decreased at 55.8 ± 0.1 and 80.8 ± 0.4 ^{o}C, respectively (in accordance with literature).

In general, as expected, as the concentrations of protecting osmolytes were increased, protein T_m followed; addition of urea, in contrast, induced deleterious effects on the thermal stability of both proteins as shown in Fig. 2c,d) again in a concentration-dependent manner. This was more pronounced in the case of BLG, for which high urea concentrations (37 wt%) induced denaturation in the absence of heating (see Fig. 2d). For this sample, T_m was approximated as 20 ^{o}C (room temperature).

Interestingly, the relative efficacy of protecting osmolytes in raising protein T_m was found to depend on concentration. In the case of LYS, sucrose conferred the best stabilization at 4 and 17 wt%, however at the highest concentration studied (37 wt%), the effects of sucrose, sorbitol, and trehalose on LYS' T_m fell within error ($\Delta_{T_m} = \pm 1$ ^{o}C) of one another. In contrast, in the case of BLG, sorbitol conferred the best stability across all three concentrations. For both BLG and LYS, glycerol exhibited the poorest stabilization of all protecting osmolytes studied. Furthermore, in both cases, variance in T_m amongst all protecting osmolyte systems was roughly 10 ^{o}C, indicating comparable magnitudes of stabilization.

Interestingly, room temperature ITF measurements— indicative of the extent of structural perturbation induced by the osmolyte in the absence of heat— do not provide direct clues on the effect of the different osmolytes of the protein melting temperature changes (see Figure S4). This was particularly evident in formulations containing sorbitol, for which relatively high red shifting at room temperature was observed together with high T_m. This may suggest that sorbitol stabilizes proteins in a manner that is mechanistically distinct from the other osmolytes.

2.2. Protein dynamics in osmolyte solutions

To predict the effect of osmolytes on protein thermal stability, we performed a set of full-atom short molecular dynamics simulations of proteins in mixed solutions of water and osmolytes at various concentrations at room temperature. As described in detail in the Method sections, we carried out simulations for lysozyme (LYS) and beta-lactoglobulin (BLG) in the presence of the five osmolytes used in experiments at four concentrations, i.e. 0, 4, 17, and 37% of osmolytes over water solvent. Note that we opted to run 100 ns-long MD simulations to reduce as much as possible the computational cost of the simulations as our ultimate goal is to reach a predictive method to select stabilizing osmolytes for a given protein.

In this framework, we first checked that both protein and osmolytes reached equilibrium within the simulated time window monitoring the Root Mean Squared Deviation (RMSD) of the system as a function of the simulation time (see Figure S2). The RMSD trends show that neither LYS nor BLG showed a significant change in structure (secondary or tertiary) when modeled in the presence of osmolytes at room temperature.

Surprisingly, this was even the case when proteins were simulated for a microsecond in 37 wt% urea, despite BLG being fully denatured under these conditions when studied by experiment (see Fig. 2). Similar observations of unperturbed LYS and BLG native structure simulated in the presence of concentrated urea at room temperature have been reported by both Biwas et al. and Eberini et al. [6,15]. This may reflect a limit of the force fields in reproducing the experimental systems' global dynamics. To monitor the local motion of the protein residues, we measured the Residue Mean Square Fluctuation (RMSF), a common metric for assessing the relative stability of simulated proteins below the threshold of detectable secondary or tertiary structural change [14].

We studied the relationship between protein mean RMSF and T_m to discover if room temperature protein dynamics– unlike structural changes– could predict the thermodynamic stability of proteins in the presence of osmolytes. For systems containing protecting osmolytes,

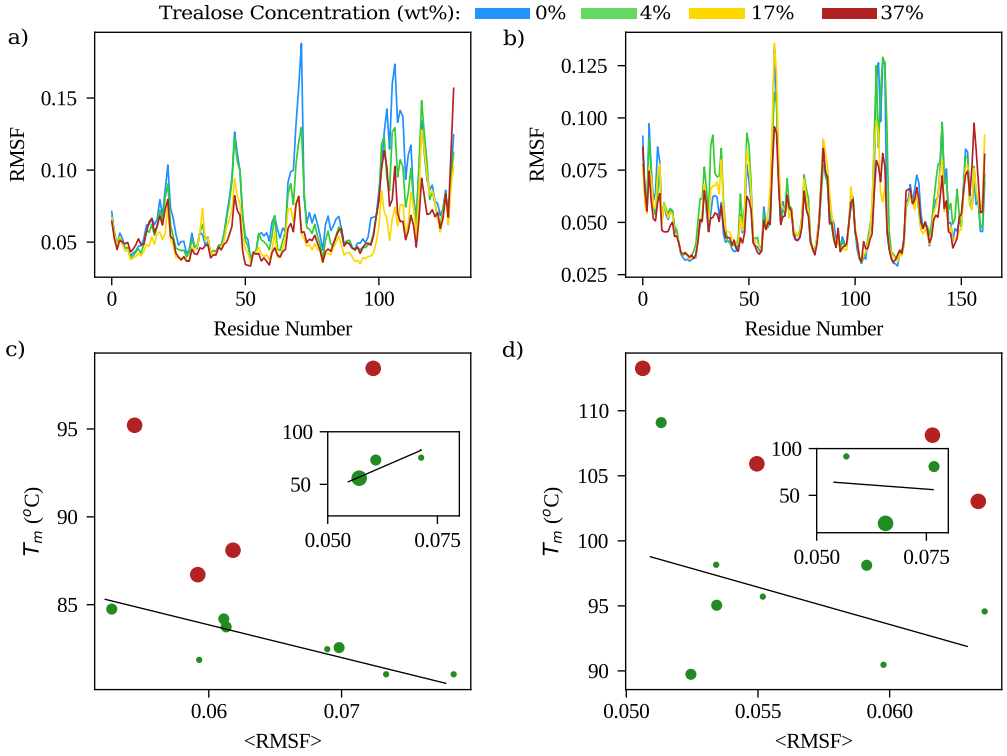

Fig. 3. Mean RMSF v. T_m for proteins in the presence of osmolytes. Example of raw RMSF data at various trehalose concentrations for **a)** LYS and **b)** BLG. Mean RMSF v. T_m considering protecting and destabilizing (inset) osmolytes for **c)** LYS and **d)** BLG. Red dots in the plots (respectively, green) represent systems with osmolyte concentrations higher (resp. lower) than 18%. Similarly, the sizes of the dots increase with system osmolyte concentrations.

negative correlations (R = -0.81 and -0.37 for LYS and BLG, respectively, with p-values of 0.02 and 0.4 for 8 points) were found to relate mean RMSF and T_m for systems at low osmolyte concentration (4 and 17 wt% data are shown as green dot in Fig. 3c,d]). This relationship, however, broke down at high concentrations of protecting osmolyte (37 wt%); indeed, when data points corresponding to this concentration were included in the model (red data points in Figs. 3a and c), the correlations between mean RMSF and T_m for both proteins were considerably weakened (R = -0.20, p-value: 0.5 for 12 points and -0.19, p-value: 0.5 for 12 points for LYS and BGL, respectively).

Furthermore, when studied in the context of urea, the relationship between RMSF and T_m failed to extrapolate. In fact, while the correlation between mean RMSF and T_m for BLG was not significant (R = -0.07, p-value: 0.93, points: 4), an opposite positive correlation was observed between protein dynamics and T_m for LYS/urea systems (R = 0.77, p-value: 0.23, points: 4) (see insets in Fig. 3c,d).

This may be related to bonding between LYS surface residues and urea, leading to reduced local flexibility of the protein. Independent of origin, however, the inconsistent relationship between mean RMSF and T_m confounds the interpretation of RMSF in osmolyte systems with respect to protein stability. Thus, we sought a finer description of the protein structural organization able to capture the effect of the osmolytes.

2.3. Osmolyte-induced perturbations to protein graphs

Residue interaction networks (RINs) have recently been shown to predict protein thermostability *ab initio* [13,38,39]. Thus, we hypothesized that analyzing the RINs of βL and LYS in the presence of different osmolytes may afford a superior method for probing stability than conventional methods (tertiary/secondary structure or mean RMSF analysis).

The networks of non-covalent interactions amongst protein residues were modeled as graphs wherein nodes were defined by the Cα atoms of amino acids, covalent interactions were ignored, links between (non-adjacent) nodes existed if the distance between them was $\leq$ 9 Å, and the weight of each link was defined by either the van der Waals or Coulombic interaction energy. See the Methods Section for more details.

The RIN mean betweenness centrality (BC)— a measure of the frequency with which a node appears amongst the shortest paths connecting all other possible node pairs in the network (see Fig. 4a) — was found to be more predictive of osmolyte-induced protein stability changes than either RMSF or conventional structure measures. As shown in Fig. 4b-c, left panels, a strong, positive correlation is present between mean RIN BC and T_m, suggesting that osmolyte-induced protein stability enhancement can be largely attributed to increased BC amongst protein residues, a proxy for structural compaction. Like RMSF, however, in the case of LYS, this trend was found to somewhat collapse at high osmolyte concentration; the inclusion of data points corresponding to 37 wt% osmolytes led to a dampening of the correlation coefficient from R = 0.82 (p-value 0.01 of for 8 points) to R = 0.07 (p-value of 0.82 for 12 points). A strong positive correlation of 0.80 (p-value of 0.002) is shown by the BC of the BLG systems.

Despite this, unlike RMSF, mean RIN BC demonstrated good predictability of T_m across all osmolyte concentrations for BGL. In addition, a concord trend is observed for the denaturing agent (urea) for which a consistently strong positive correlation was observed for both proteins (R = 0.83, p-value of 0.37, and R = 0.98, p-value of 0.12, for the three urea concentrations fo r LYS and BLG, respectively, as shown in Fig. 4b, c, right panels).

These results suggest that mean RIN BC offers an alternative descriptor for monitoring osmolyte-induced room temperature perturbations to protein thermodynamic stability often too subtle to detect by conventional metrics.

2.4. Protein-osmolytes interaction network

To check whether the stabilizing/destabilizing effect of osmolytes is protein-dependent, we moved to consider the interaction between proteins and osmolytes more directly. To this aim, we first evaluate the disposition of the osmolyte around the protein structure.

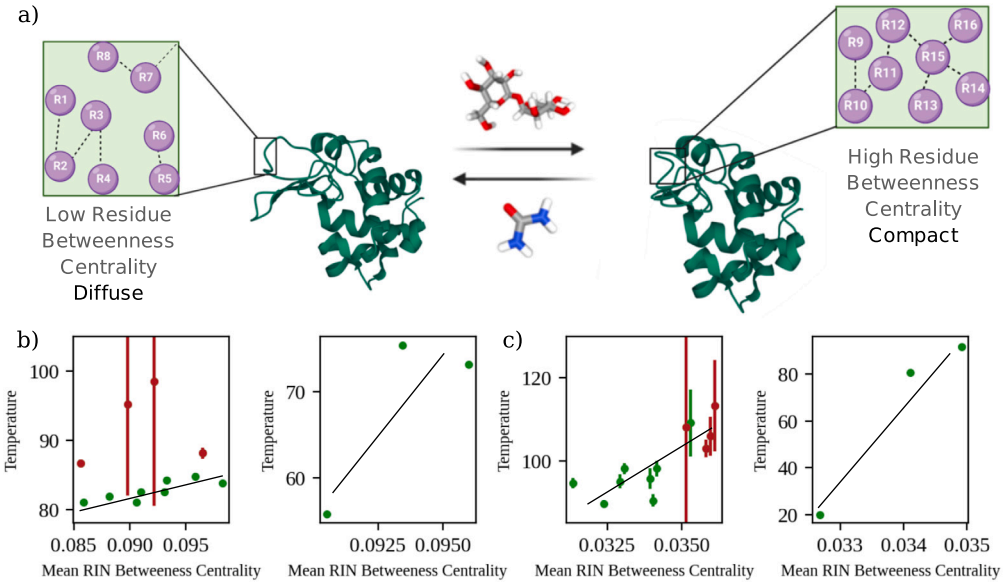

Fig. 4. Mean RIN BC v. T_m for proteins in the presence of protecting and denaturing osmolytes. a) Illustration of RIN BC for protein stability monitoring mechanism. **b)** Mean RIN BC v. T_m for LYS in solutions of protecting osmolytes (left panel) and urea (right panel). **c)** Mean RIN BC v. T_m for BLG in solutions of protecting osmolytes (left panel) and urea (right panel).

Fig. 5a,b depicts an example of the radial distribution functions (g(r))— the relative density of osmolyte molecules as a function of distance from the protein center— for sucrose, urea, and water around LYS. See Supplementary Information for all studied systems. Comparative inspection of these graphs reveals several key distinctions among osmolyte-protein interactions. First, in general, it is evident that glycerol and urea present well-defined peaks with high proximity to the protein core relative to other osmolytes; the intensity and position of these peaks, however, appear to be protein-dependent. Furthermore, poor conservation of the shape of g(r) across various concentrations of a single osmolyte suggests a concentration dependence of protein-osmolyte interactions; this is least apparent for urea, in the case of which minimal peak shifting or emergence is observed. Finally, it is interesting to note that in disaccharide (trehalose, sucrose) systems, several well-defined peaks are evident at a distance ≥ 15 Å from the protein. These suggest the presence of ordered osmolyte structures around the protein surface.

Protein-osmolyte preferential interaction coefficients were calculated from the protein-osmolyte and protein-water radial distribution functions as described in the Methods Section. Positive preferential interaction coefficients indicate an excess of the osmolyte in the local domain of the protein, whilst a negative preferential interaction coefficient indicates exclusion of the osmolyte from the direct surroundings of the protein. Preferential interaction coefficients for all protein-osmolyte systems are presented in Fig. 5c-d.

Consistent with previous findings [11,46], urea presented a positive preferential interaction coefficient with both proteins at all concentrations (Fig. 5). The preferential interaction coefficient characterizing urea/BLG was higher than that characterizing urea/LYS across all concentrations; this difference was most dramatic at 37 wt%, however, reflecting the relative magnitude of T_m depression observed amongst the two systems (BLG: Δ_{T_m} = - 65, LYS: Δ_{T_m} = -26). The origin of the differential extent of preferential interaction amongst urea/BLG and urea/LYS (as well as differences amongst coefficients characterizing interactions between the two proteins and protecting osmolytes) was not the primary focus of this study and was thus not investigated. Nonetheless, a systematic investigation into the relationship between protein size, degree of backbone solvation, surface residue composition, or three-dimensional shape/topology, and preferential interaction coefficients could be a fascinating area of future research.

Amongst all protecting osmolytes studied, the degree of exclusion increased with concentration. Generally, sucrose and trehalose were most strongly excluded from the local environment of the protein; interestingly, the protein-osmolyte radial distribution functions of these species also showed the greatest degree of structural organization at distances ≥ 15 Å from the protein (Figure S2). Such an observation might suggest the exclusion of relatively ordered disaccharide assemblies; this hypothesis is investigated in detail in the following. It is also noted that glycerol— the least effective protecting osmolyte by concentration-dependent melting temperature elevation— is the only protecting species not preferentially excluded at high concentrations (37 wt%).

To more quantitatively assess the relationship between osmolyte-induced thermal stability changes and preferential exclusion, the preferential interaction coefficients of all osmolyte-protein combinations were plotted against protein melting temperatures. A strong negative correlation (R = -0.76, p-value < 0.001 on 30 points) was found to characterize the relationship between T_m and preferential interaction coefficient of both proteins, consistent with preferential exclusion theory [11,19,53]. Dividing the two systems, we obtained a correlation of -0.88 (p-value < 0.001 on 15 points) and -0.80 (p-value < 0.001 on 15 points) for LYS and BLG, respectively.

Correlations between osmolyte-protein preferential interaction coefficients and T_m were further disaggregated by osmolyte/protein pair; Table 1 reports R values for the relationships between T_m and preferential interaction coefficients across all ten protein/osmolyte combinations.

In all cases, the relationship between T_m and preferential interaction was consistent (negative R-value). Interestingly, the osmolyte-protein pairs for which the largest (by absolute value) preferential interaction coefficients were calculated (BLG/urea, BLG/trehalose, BLG/sucrose, and LYS/sucrose) were also characterized by the strongest correlations between T_m and preferential interaction. This suggests that when a significant difference in osmolyte preference for the local (protein) and bulk domains is absent, secondary factors likely become the predominant predictors of T_m.

2.5. *Preferential exclusion of osmolyte clusters*

To better understand the relationship between osmolyte clustering and preferential exclusion, osmolyte-osmolyte networks were analyzed

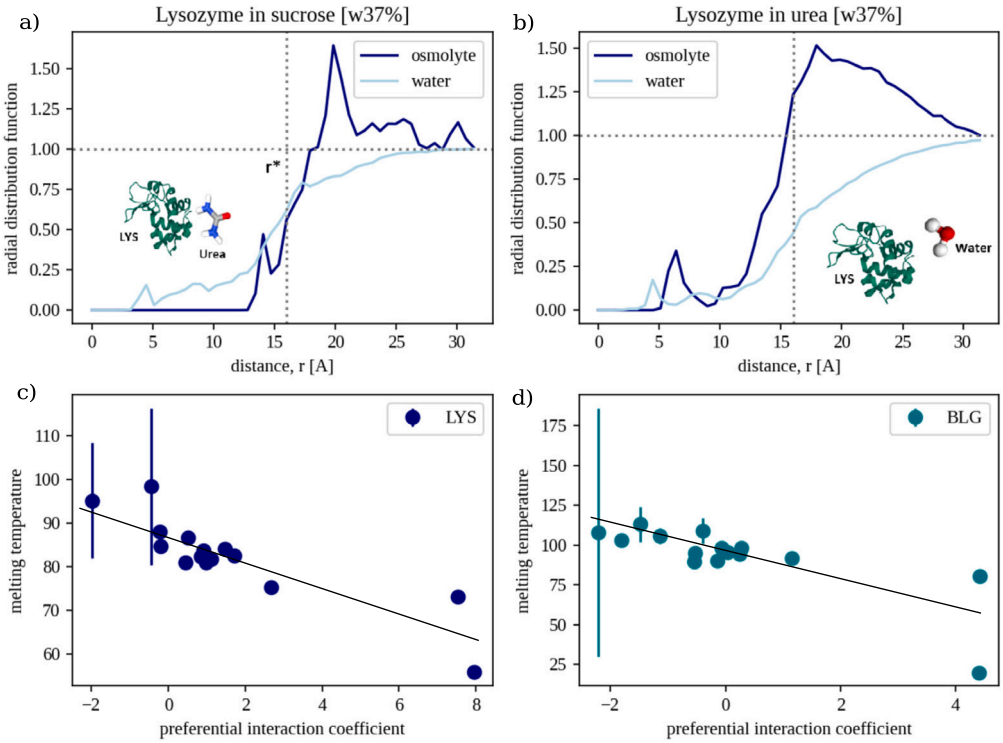

Fig. 5. Analysis on osmolyte-protein preferential interaction. a) Protein-osmolyte (dark blue) and protein-water (light blue) g(r) of an exemplary system (LYS in 37 wt% sucrose). Dotted vertical line marks the local (defined by $r = [0, r^{\star}]$) and distant domains ($[r^{\star}, r_{box}]$). The cutoff distance between the local and bulk domains, $r^{\star}$, is selected as described in the Methods. **b)** Same as in panel a) but for lysozyme in the presence of urea. **c)** Osmolyte-protein preferential interaction coefficients ($\Gamma_{o,p}$) for all 15 simulated lysozime-osmolyte systems. **c)** Same as in panel c) but for BLG.

Table 1
T_m v. Preferential Interaction Correlation Coefficients (R) by Osmolyte-Protein Pair.

	LYS	BLG
Glycerol	-0.59	-0.51
Sorbitol	-0.75	-0.72
Sucrose	-0.99	-0.98
Trehalose	-0.77	-0.95
Urea	-0.58	-0.80

Table 2
Mean Clique Volume v. Preferential Interaction Coefficient Correlation (R) by Concentration.

Osmolyte Concentration (wt%)	LYS	βL
4	-0.67	-0.61
17	-0.71	-0.73
37	-0.66	-0.61

as unweighted graphs since osmolytes are of the same kind in each considered system (see Methods). To quantify the organization of the network, we evaluate the cliques, i.e. subsets of vertices of the undirected graph that have the property that any two distinct vertices in the clique are adjacent. Measuring the mean clique size provides a proxy of the average size of the clusters in the network. Interestingly, the mean clique volume (i.e. the volume of fully connected subgraphs or clusters) was found to be predictive of the protein thermal stability (Fig. 6) This relationship was characterized by positive correlation coefficients for both proteins of R = 0.42 (p-value of 0.02 for 30 points) and 0.32 (p-value of 0.08 for 30 points) for LYS and BLG, respectively.

Interestingly, the mean clique volume (i.e. the volume of fully connected subgraphs or clusters) was found to be predictive of the osmolyte-protein preferential interaction coefficient (see Fig. 6c-d). This relationship was characterized by negative correlation coefficients for both proteins (R = -0.67 and -0.49 for LYS and βL, respectively), suggesting that preferential exclusion may be driven to a significant extent by steric exclusion of osmolyte clusters— rather than discrete molecules— from the local domain of the protein. Furthermore, the relationship between preferential interaction coefficients and mean osmolyte clique volume was consistent across all concentrations of protecting osmolytes, however failed to linearly extrapolate to systems containing high urea concentrations (17 and 37 wt%).

We hypothesized that this inconsistency in trend observed at high urea concentrations could be accounted for by increasing node density, and in turn, node-to-node proximity. Thus, to disentangle the observed correlation from the effect of concentration (and concomitant effect of changes to global node density), the correlation between mean clique volume and preferential interaction was analyzed across concentration normalized data (Table 2). Indeed, a moderately strong, negative correlation (-0.61 ≤ R ≤ -0.71) was observed amongst all concentrations.

Additionally, the correlation between preferential exclusion and osmolyte clustering could be visually comprehended. Fig. 6e depicts MD snapshots of systems (37 wt% osmolyte) exhibiting preferential interaction (left) and preferential exclusion (right). In general, preferentially interacting osmolytes did not aggregate; instead, they appeared to diffuse across the entirety of the simulation space with little preference for one region of the bulk domain over another. In contrast, strongly excluded osmolytes (sucrose, trehalose) formed extended structural networks around the protein, failing to diffuse into the whole simulation space. In particular, looking at the cluster size distributions, one clearly sees that these osmolytes form many, smaller clusters, which dispose around the protein structure (see Table 3 for

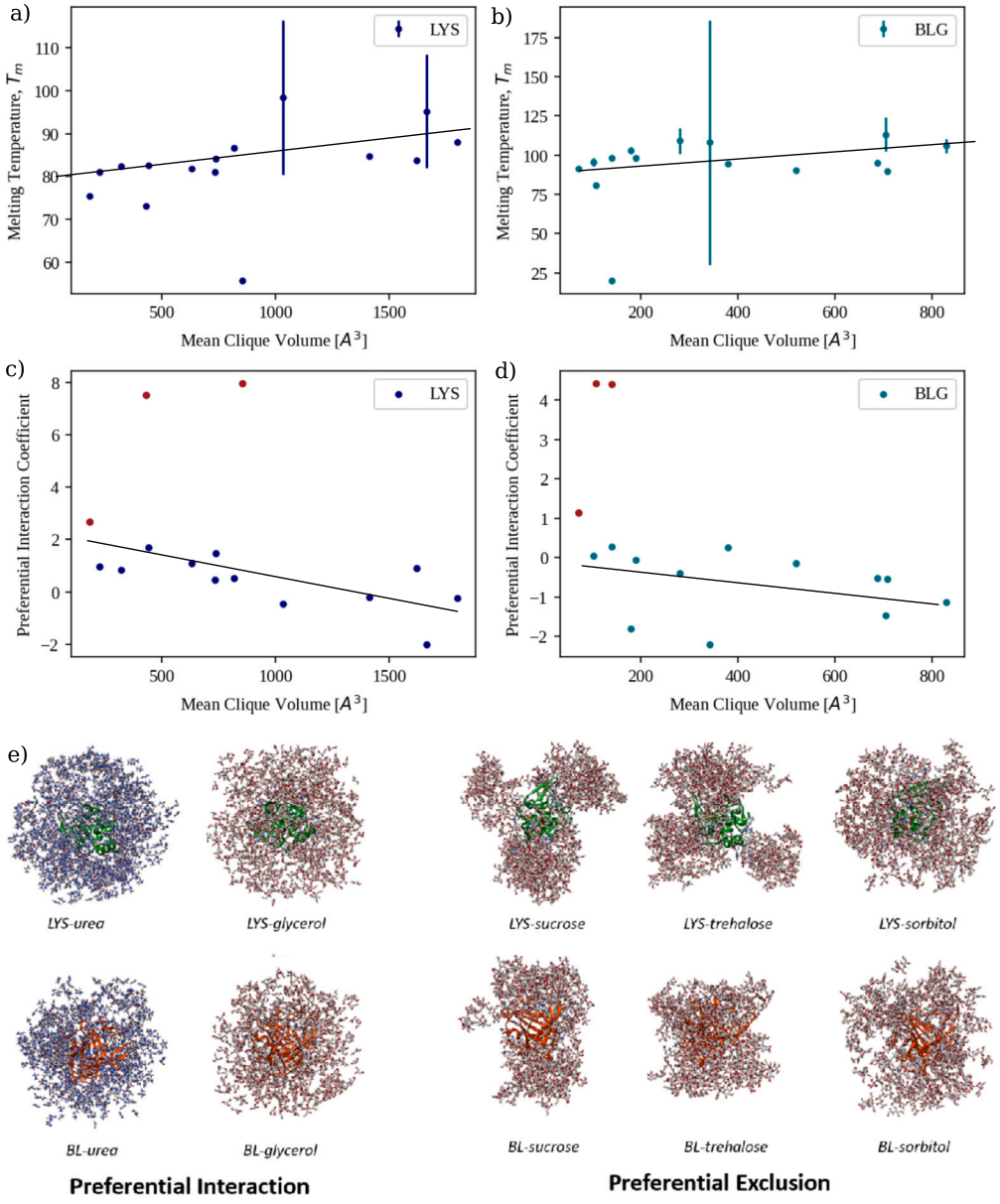

Fig. 6. Osmolyte clique volume v. preferential interaction coefficient. a) Mean osmolyte clique volume v. melting temperatures for LYS. **b)** Same as in a) but for BLG. **c)** Mean osmolyte clique volume v. Preferential Interaction Coefficient for LYS. **d)** Same as in c) but for BLG. **e)** MD snapshots of osmolytes with positive (glycerol, urea) and negative (sorbitol, sucrose, trehalose) preferential interaction coefficients at 37 wt%. Preferentially excluded osmolytes tend to form osmolyte-osmolyte clusters whilst preferentially interacting species lack ordered domains within the bulk phase, appearing to exhibit no preference for one region of the bulk domain over another. Furthermore, unlinked osmolytes (unimers and dimers) are visually apparent in preferentially interacting systems, while preferentially excluded osmolytes appear as subunits of extended networks.

the average number of formed clusters and Supplementary Information).

Sorbitol, weakly excluded from both proteins, in contrast, fell somewhere between these two morphological scenarios, reflective of the osmolyte's aberrant experimental behavior.

Closer inspection of the snapshots in Fig. 6e reveals yet more granular distinctions between the two classes of osmolytes. Examination of the osmolyte-protein system boundaries reveals the presence of isolated unimers and dimers in preferentially interacting species, whilst those characterized by preferential exclusion appear linked to a larger network. Lastly, it is interesting to note the strong dependence of the morphology of aggregates formed by strongly excluded osmolytes (sucrose, trehalose) on the protein identity. This could suggest that the protein itself governs osmolyte assembly, perhaps by indirectly acting as a scaffold via the first hydration layer.

3. Conclusions and outlook

Two major conclusions can be drawn from this work; the first is that RIN BC is a high-performing metric for monitoring subtle osmolyte-induced, room-temperature structural perturbations by MD. RIN BC shows a consistent, moderately strong negative correlation with experimental thermal stability across systems containing protecting or denaturing osmolytes; in contrast, conventional methods such as mean RMSF and secondary/tertiary structure analysis fail to predict T_m with generalizability across both osmolyte classes. Future investigations into the mechanistic relevance/predictive power of RIN BC in protein systems lacking osmolytes (i.e. proteins at high temperature or in the presence of other classes of stabilizers, etc.) would be necessary to elucidate the generalizability of this metric and to test whether, alongside offering predictive utility, analysis of RIN BC may also offer insight into the

Table 3
Mean values of the distribution of max clique size and of the number of clusters for the considered protein-osmolyte systems.

		2YVB		3BLG	
osmolyte	conc[%]	$< c_{max} >$	$< s >$	$< c >$	$< s >$
sorbitol	4	3.1	31.1	2.8	23.2
	17	4.5	44.8	4.3	34.3
	37	5.3	18.3	5.1	14.5
sucrose	4	3.0	15.4	3.2	10.9
	17	3.9	17.4	3.7	12.7
	37	4.4	15.7	4.3	12.7
trehalose	4	2.6	12.4	3.1	10.5
	17	3.3	13.5	3.8	8.4
	37	4.4	8.8	4.4	9.7
glycerol	4	3.3	57.1	3.1	42.8
	17	5.1	44.8	5.3	39.2
	37	6.4	3.3	6.4	3.4
urea	4	4.1	67.2	4.3	50.2
	17	8.3	20.3	8.2	16.5
	37	10.6	2.9	10.3	2.5

mechanistic origins of osmolyte-induced protein stability— structural compaction— consistent with preferential exclusion theory [11,52].

Second, it has been shown that a positive correlation exists between an osmolyte's clustering and the extent to which it is preferentially excluded from the local domain of the protein. This correlation is further shown to be dependent upon cluster size, with clusters of larger volume being more strongly excluded from the protein.

These results suggest that the preferential exclusion phenomenon— to which thermodynamic protein stabilization by protecting osmolytes is attributed— is likely, at least in part, motivated by the steric exclusion of molecular networks rather than unimeric species.

Finally, it is noted that the morphology of clusters formed by excluded osmolytes appears to be protein-dependent. This is intriguing, as— by nature of being preferentially excluded— these species do not significantly interact. Thus, the relationship between the structure of osmolyte aggregates and protein identity is likely rooted in indirect relations, *e.g.* via the first hydration layer. Such an explanation would suggest that the shape of the protein and in turn, the architecture of the primary hydration layer, governs the morphology of osmolyte clusters via a scaffolding effect.

Finally, we want to stress that while the set of studied descriptors has the potential to be used as effective predictors of the effect of osmolyte kinds/concentrations on the protein melting temperatures, the link to the stabilizing/destabilizing mechanisms could be affected by the specifics of the molecular dynamics simulations we chose. In fact, previous works pointed out a force-field dependence in the aggregation properties of osmolytes in solution [18]. More studies will be required to investigate whether the observed mechanism is supported by experimental evidence.

An investigation into this hypothesis could serve as a fascinating premise for future study.

4. Materials and methods

4.1. Materials

All materials were purchased from Sigma Aldrich unless otherwise noted.

4.2. Measurement of protein thermal stability

Protein conformational changes were monitored in duplicate by tryptophan fluorescence using a Cary Eclipse Fluorimeter. Samples were excited at 295 nm (slit width = 2 nm) to avoid excitation of tyrosine. An emission scan was collected from 300 to 400 nm and the ratio of Emission Intensities at 350 nm (I350) to 320 nm (I320) was plotted against temperature. In the case of lysozyme, the data was then fit to a sigmoidal distribution using the Boltzmann function, and T_m was obtained from x_o, the inflection point of the function,

$$y = \frac{y_{max} - y_{min}}{1 + e^{-(T-T_m)/\delta}} + y_{min} \tag{1}$$

For β-Lactoglobulin, incomplete convergence of the I350/I320 plot over the experimentally feasible temperature range leads to poor sigmoidal fits. As such, thermal stability was assessed by $T_{threshold}$, the temperature at which a pre-determined extent of unfolding, as captured by the ratio of I350/I320 was reached. $T_{threshold}$ was chosen based on the midpoint of the thermal transition in the absence of osmolyte.

4.3. Simulated systems

Computational analyses were performed on the same systems studied by experiment. Crystallographic structures of Hen egg white lysozyme (PDB ID: 2YVB) and bovine β-lactoglobulin (PDB ID: 3BLG) were retrieved from the Protein Data Bank [5]. In addition, five osmolyte molecules were considered, i.e. trehalose (PubChem ID: 7427), sucrose (PubChem ID: 5988), sorbitol (PubChem ID 5780), glycerol (PubChem ID: 753), and urea (PubChem ID: 1176). Starting systems were built with one protein in an osmolyte-water environment, with osmolyte concentrations at either 0%, 4%, 17%, or 37% by weight.

To fix concentration we constructed a dodecahedral box so that each atom of the protein structures was at a distance of at least 1.1 nm from the nearest box face. We filled one copy of the box with only water molecules and another copy with only molecules of an osmolyte; we call w the ratio between the number of water molecules (WM) and the number of osmolyte molecules (CM), that is $r = WM/CM$. Since the molecular masses of water (m) and osmolytes (M) are known, we have that the concentration is

$$w = \frac{N \cdot M}{N \cdot M + n \cdot m} \tag{2}$$

where w is the concentration, N is the final number of osmolyte molecules and n is the final number of water molecules.

We can write a system of two equations, where the first equation is (2), while the second equation is obtained by considering that in the absence of an osmolyte, we have WM water molecules and that each molecule replaces r water molecules, therefore

$$\begin{cases} N = \dfrac{m}{M} \cdot \dfrac{w}{1-w} \cdot n \\ \\ N + n = t - rN \end{cases} \tag{3}$$

4.4. Molecular dynamics simulations

All simulations were performed using GROMACS [54]. Topologies of the system were built using the CHARMM-27 force field [9]. The protein was placed in a dodecahedral simulative box, with periodic boundary conditions, filled with TIP3P water molecules [24]. The 2YVB and 3BLG systems were neutralized with 8 Cl atoms and 9 Na atoms respectively. We chose the CHARMM27 force field and TIP3P water model; we obtained the force field parameters of the osmolytes from SwissParam [57]. Each dynamic extends for 100 ns (see next section for details).

The number of molecules used for each simulation is given in Table 4. For all simulated systems, we checked that each atom of the proteins was at least at a distance of 1.1 nm from the box borders. Each system was then minimized with the steepest descent algorithm. Next, a relaxation of water molecules and thermalization of the system was run in NVT and NPT environments each for 0.1 ns at 2 fs time-step. The temperature was kept constant at 300 K with v-rescale thermostat [10]; the final pressure was fixed at 1 bar with the Parrinello-Rahman barostat [41].

Table 4
MD simulation details. N_{OS} = Total number of osmolyte molecules. N_{WATER} = Total number of water molecules.

		LYS		BLG	
Osmolyte Concentration (wt%)	Osmolyte	N_{OS}	N_{WATER}	N_{OS}	N_{WATER}
4	glycerol	79	9060	56	6874
17		299	7754	231	5800
37		645	5582	482	4159
4	sorbitol	40	9078	29	6901
17		155	7808	116	5905
37		342	5412	241	4325
4	sucrose	21	9111	15	6908
17		80	8034	61	6006
37		170	5911	125	4574
4	trehalose	21	9134	15	6943
17		79	8037	62	5959
37		169	5901	126	4228
4	urea	122	9032	86	6891
17		467	7877	361	5880
37		1026	5720	769	4257

LINCS algorithm [23] was used to constrain bonds involving hydrogen atoms. A cut-off of 12 Å was imposed for the evaluation of short-range non-bonded interactions and the Particle Mesh Ewald method [12] for the long-range electrostatic interactions. The described procedure was used for all the performed simulations.

Interaction energy calculation

Intra-molecular interaction energies were computed using the parameters obtained from the CHARMM force field. In particular, given two atoms a_l and a_m holding partial charges q_l and q_m, the Coulombic interaction between them can be computed as:

$$E_{lm}^{C} = \frac{1}{4\pi\epsilon_0}\frac{q_l q_m}{r_{lm}} \tag{4}$$

where r_{lm} is the distance between the two atoms, and ϵ_0 is the vacuum permittivity. Van der Waals interactions can instead be calculated as a 12-6 Lennard-Jones potential:

$$E_{lm}^{LJ} = \sqrt{\epsilon_l \epsilon_m}\left[\left(\frac{R_{min}^{l} + R_{min}^{m}}{r_{lm}}\right)^{12} - 2\left(\frac{R_{min}^{l} + R_{min}^{m}}{r_{lm}}\right)^{6}\right] \tag{5}$$

where ϵ_l and ϵ_m are the depths of the potential wells of a_l and a_m respectively, R_{min}^{l} and R_{min}^{m} are the distances at which the potentials reach their minima.

The total interaction energy between each couple of residues is defined as:

$$E_{AA_{ij}}^{X} = \sum_{l=1}^{N_{atom}^{i}}\sum_{m=1}^{N_{atom}^{j}} E_{lm}^{X} \tag{6}$$

where $E_{AA_{ij}}^{X}$ is the energy between two amino acids i and j, obtained as the sum of the interactions between each atom of the two residues (N_{atom}^{i}, N_{atom}^{j}); X stands for the kind of interaction considered, either Coulombic ($X = C$) or Lennard-Jones ($X = LJ$).

As for the distance between a pair of residues, this was assessed by selecting the minimum distance between the atoms composing them.

4.5. Hydrogen bond calculation

For each of the dynamics, we extracted one frame every 1 ns for a total of 101 frames. For each frame, using the Chimera software, we counted the number of hydrogen bonds that each residue forms with other residues, with water molecules, and with osmolytes.

4.6. Calculation of radial distribution functions

Radial distribution functions (g(r)) for protein-osmolyte and protein-water interactions were calculated by the equation below:

$$g(r)_{a,b} = \frac{1}{\rho_{bulk,b}}\frac{\delta N_b(r)}{\delta V(r)} \tag{7}$$

wherein δV(r) and δN_b(r) indicate the volume and the number of b particles corresponding to the bin defined by (r, dr), respectively; r is the distance from reference particle a and dr is the bin size.

4.7. Calculation of preferential interaction parameters

Osmolyte-protein preferential interaction parameters ($\Gamma_{os.,prot}$) were calculated using the equation,

$$\Gamma_{os.,prot.} = n_{os.}^{close} - \frac{n_{os.}^{distant}}{n_{water}^{distant}}(n_{water}^{close}) \tag{8}$$

wherein $n_{os.}^{close}$ and $n_{os.}^{distant}$ are the number of osmolyte molecules found in close proximity to the protein or distant from its molecular surface, respectively, and n_{water}^{close} and $n_{water}^{distant}$ are the number of water molecules less or more distant from a radius $r^{\star}$ from the center of the protein (adapted from [46]). The cutoff distance marking the boundary between bulk and local domains, $r^{\star}$, was set to the distance comparable with the average radius of the proteins, which was found to be $\sim$ 16.0 Å. The number of osmolyte or water molecules was calculated via integration of the respective protein-osmolyte or protein-water radial distribution function:

$$n_{(r_a,r_b)} = 4\pi * \rho * \int_{r_a}^{r_b} g(r) r^2 dr \tag{9}$$

4.8. Graph analysis

Perturbations to protein structure and osmolyte clustering behavior were studied by graph analysis. Protein intermolecular bonds were modeled as residue interaction networks (RINs) wherein residue Cα atoms were approximated as nodes and two nodes were considered to be in contact if their distance was less than 12 Å. The weight for each link was determined by interaction energy (van der Waals or Coulombic). For each RIN, average metrics were calculated considering only key residues selected based on the threshold criteria ($E < -10$, degree > 25).

Osmolyte-osmolyte clustering was studied via graph analysis. Osmolyte systems were modeled as unweighted graphs using the NetworkX package in Python. Discrete molecules were represented as nodes by their center of mass. A link existed between two nodes when the distance between them was less than 9 Å [25].

Cliques were defined as complete subgraphs wherein a link existed between all possible node pairs. The average clique size was calculated by NetworkX and volume-normalized using RDKit-generated molecular volumes for each species. All metrics were calculated as the time averages over 101 frames extracted every 1 ns of simulation time.

Normalized betweenness centrality metrics for residue interaction and osmolyte networks were calculated according to the equation,

$$BC(v) = \frac{\sum_{s\neq v\neq t}\frac{\sigma_{st}(v)}{\sigma_{st}}}{(n-1)(n-2)/2} \tag{10}$$

wherein the $BC(v)$ is the betweenness centrality of node v, n is the number of nodes in the graph, (s,t) represents each pair of vertices, σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the total number of shortest paths that pass through the non-terminal node v. Betweenness centrality is a measure used in graph theory to quantify the centrality or importance of a node within a network. It reflects the extent to which a node lies on the shortest paths between other pairs

of nodes in the graph. Nodes with high betweenness centrality are crucial for maintaining efficient communication and connectivity within the network. Mathematically, the betweenness centrality of a node is calculated based on the number of shortest paths passing through that node compared to the total number of shortest paths between all pairs of nodes in the graph. In practical terms, a high betweenness centrality indicates that a node is strategically positioned in the network, serving as a bridge or intermediary between different parts of the graph. Nodes with high betweenness centrality often play a critical role in maintaining the overall connectivity of the network.

CRediT authorship contribution statement

Mattia Miotto: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nina Warner:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Giancarlo Ruocco:** Supervision, Resources. **Gian Gaetano Tartaglia:** Supervision, Conceptualization. **Oren A. Scherman:** Supervision, Resources, Conceptualization. **Edoardo Milanetti:** Supervision, Conceptualization.

Declaration of competing interest

We declare no conflict of interests for the present manuscript.

Acknowledgement

The authors thank Leonardo Bo' for technical support during the early phases of the work. M.M., G.R., and G.G.T. acknowledge support by the European Research Council through its Synergy grant programme, project ASTRA (grant agreement No 855923), and by European Innovation Council through its Pathfinder Open Programme, project ivBM-4PAP (grant agreement No 101098989). N.W. and O.A.S thank AB Agri for financial support.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csbj.2024.10.014.

References

[1] Adamczak Beata, Kogut Mateusz, Czub Jacek. Effect of osmolytes on the thermal stability of proteins: replica exchange simulations of Trp-cage in urea and betaine solutions. Phys Chem Chem Phys 2018;20(16):11174–82.
[2] Anand D Vijay, Meng Zhenyu, Xia Kelin, Mu Yuguang. Weighted persistent homology for osmolyte molecular aggregation and hydrogen-bonding network analysis. Sci Rep December 2020;10(1):9685.
[3] Arakawa T, Timasheff SN. The stabilization of proteins by osmolytes. Biophys J March 1985;47(3):411–4.
[4] Auton Matthew, Holthauzen Luis Marcelo F, Bolen D Wayne. Anatomy of energetic changes accompanying urea-induced protein denaturation. Proc Natl Acad Sci September 2007;104(39):15317–22.
[5] Berman Helen M, Westbrook John, Feng Zukang, Gilliland Gary, Bhat TN, Weissig Helge, et al. The Protein Data Bank. Nucleic Acids Res January 2000;28(1):235–42.
[6] Biswas Biswajit, Muttathukattil Aswathy N, Reddy Govardhan, Singh Prashant Chandra. Contrasting effects of guanidinium chloride and urea on the activity and unfolding of lysozyme. ACS Omega October 2018;3(10):14119–26.
[7] Bolen D Wayne, Rose George D. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu Rev Biochem June 2008;77(1):339–62.
[8] Brinda KV, Vishveshwara Saraswathi. A network representation of protein structures: implications for protein stability. Biophys J December 2005;89(6):4159–70.
[9] Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, et al. The biomolecular simulation program. J Comput Chem 2009;30(10):1545–614.
[10] Bussi Giovanni, Donadio Davide, Parrinello Michele. Canonical sampling through velocity-rescaling. J Chem Phys January 2007;126(1):014101.
[11] Canchi Deepak R, García Angel E. Cosolvent effects on protein stability. Annu Rev Phys Chem April 2013;64(1):273–93.
[12] Cheatham III TE, Miller JL, Fox T, Darden TA, Kollman PA. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. J Am Chem Soc April 1995;117(14):4193–4.
[13] Desantis Fausta, Miotto Mattia, Di Rienzo Lorenzo, Milanetti Edoardo, Ruocco Giancarlo. Spatial organization of hydrophobic and charged residues affects protein thermal stability and binding affinity. Sci Rep July 2022;12(1).
[14] Dong Yun-wei, Liao Ming-ling, Meng Xian-liang, Somero George N. Structural flexibility and protein adaptation to temperature: molecular dynamics analysis of malate dehydrogenases of marine molluscs. Proc Natl Acad Sci USA February 2018;115(6):1274–9.
[15] Eberini Ivano, Emerson Andrew, Sensi Cristina, Ragona Laura, Ricchiuto Piero, Pedretti Alessandro, et al. Simulation of urea-induced protein unfolding: a lesson from bovine β-lactoglobulin. J Mol Graph Model September 2011;30:24–30.
[16] Fernández Michael, Caballero Julio, Leyden Fernández, Abreu Jose Ignacio, Acosta Gianco. Classification of conformational stability of protein mutants from 3D pseudo-folding graph representation of protein sequences using support vector machines. Proteins, Struct Funct Bioinform 2008;70(1):167–75.
[17] Ferreira Luisa A, Fan Xiao, Madeira Pedro P, Kurgan Lukasz, Uversky Vladimir N, Zaslavsky Boris Y. Analyzing the effects of protecting osmolytes on solute–water interactions by solvatochromic comparison method: II. Globular proteins. RSC Adv 2015;5(73):59780–91.
[18] Ganguly Pritam, Boserman Pablo, van der Vegt Nico FA, Shea Joan-Emma. Trimethylamine n-oxide counteracts urea denaturation by inhibiting protein–urea preferential interaction. J Am Chem Soc December 2017;140(1):483–92.
[19] Gekko Kunihiko, Timasheff Serge N. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. Biochemistry August 1981;20(16):4667–76.
[20] Giollo Manuel, Martin Alberto JM, Walsh Ian, Ferrari Carlo, Tosatto Silvio CE. NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. BMC Genomics May 2014;15(4):S7.
[21] Guinn Emily J, Pegram Laurel M, Capp Michael W, Pollock Michelle N, Record M Thomas. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. Proc Natl Acad Sci September 2011;108(41):16932–7.
[22] Harries Daniel, Rösgen Jörg. A practical guide on how osmolytes modulate macromolecular properties. 2008. p. 679–735.
[23] Hess Berk, Bekker Henk, Berendsen Herman JC, Fraaije Johannes GEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem September 1997;18(12):1463–72.
[24] Jorgensen William L, Chandrasekhar Jayaraman, Madura Jeffry D, Impey Roger W, Klein Michael L. Comparison of simple potential functions for simulating liquid water. J Chem Phys July 1983;79(2):926–35.
[25] Kalia Munishikha, Miotto Mattia, Ness Deborah, Opie-Martin Sarah, Spargo Thomas P, Di Rienzo Lorenzo, et al. Molecular dynamics analysis of superoxide dismutase 1 mutations suggests decoupling between mechanisms underlying als onset and progression. Comput Struct Biotechnol J 2023;21:5296–308.
[26] Kamerzell Tim J, Esfandiary Reza, Joshi Sangeeta B, Middaugh C Russell, Volkin David B. Protein–excipient interactions: mechanisms and biophysical characterization applied to protein formulation development. Adv Drug Deliv Rev October 2011;63(13):1118–59.
[27] Khan Shagufta H, Ahmad Nihal, Ahmad Faizan, Kumar Raj. Naturally occurring organic osmolytes: from cell physiology to disease prevention. IUBMB Life 2010;62(12):891–5.
[28] Kuffel Anna, Zielkiewicz Jan. The hydrogen bond network structure within the hydration shell around simple osmolytes: urea, tetramethylurea, and trimethylamine-N-oxide, investigated using both a fixed charge and a polarizable water model. J Chem Phys July 2010;133(3):035102.
[29] Lee Hochan, Choi Jun-Ho, Verma Pramod Kumar, Cho Minhaeng. Spectral graph analyses of water hydrogen-bonding network and osmolyte aggregate structures in osmolyte–water solutions. J Phys Chem B November 2015;119(45):14402–12.
[30] Lee Juneyoung, Ko Jeong Hoon, Lin En-Wei, Wallace Peter, Ruch Frank, Maynard Heather D. Trehalose hydrogels for stabilization of enzymes to heat. Polym Chem May 2015;6(18):3443–8.
[31] Lerbret A, Bordat P, Affouard F, Hédoux A, Guinet Y, Descamps M. How do trehalose, maltose, and sucrose influence some structural and dynamical properties of lysozyme? Insight from molecular dynamics simulations. J Phys Chem B August 2007;111(31):9410–20.
[32] Lin Tiao-Yin, Timasheff Serge N. Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. Biochemistry October 1994;33(42):12695–701.
[33] Lins Roberto D, Pereira Cristina S, Hünenberger Philippe H. Trehalose–protein interaction in aqueous solution. Proteins, Struct Funct Bioinform 2004;55(1):177–86.
[34] Liu Fu-Feng, Ji Luo, Zhang Lin, Dong Xiao-Yan, Sun Yan. Molecular basis for polyol-induced protein stability revealed by molecular dynamics simulations. J Chem Phys June 2010;132(22):225103.
[35] Lushchekina Sofya V, Inidjel Gaetan, Martinez Nicolas, Masson Patrick, Trovaslet-Leroy Marie, Nachon Florian, et al. Impact of sucrose as osmolyte on molecular dynamics of mouse acetylcholinesterase. Biomolecules December 2020;10(12):1664.
[36] Mehrnejad Faramarz, Ghahremanpour Mohammad Mehdi, Khadem-Maaref Mahmoud, Doustdar Farahnoosh. Effects of osmolytes on the helical conforma-

tion of model peptide: molecular dynamics simulation. J Chem Phys January 2011;134(3):035104.
[37] Miotto Mattia, Armaos Alexandros, Di Rienzo Lorenzo, Ruocco Giancarlo, Milanetti Edoardo, Tartaglia Gian Gaetano. Thermometer: a webserver to predict protein thermal stability. Bioinformatics January 2022;38(7):2060–1.
[38] Miotto Mattia, Olimpieri Pier Paolo, Di Rienzo Lorenzo, Ambrosetti Francesco, Corsi Pietro, Lepore Rosalba, et al. Insights on protein thermal stability: a graph representation of molecular interactions. Bioinformatics August 2019;35(15):2569–77.
[39] Miotto Mattia, Di Rienzo Lorenzo, Corsi Pietro, Ruocco Giancarlo, Raimondo Domenico, Milanetti Edoardo. Simulated epidemics in 3d protein structures to detect functional properties. J Chem Inf Model 2020;60(3):1884–91.
[40] Ohtake Satoshi, Kita Yoshiko, Arakawa Tsutomu. Interactions of formulation excipients with proteins in solution and in the dried state. Adv Drug Deliv Rev 2011;63(13):1053–73.
[41] Parrinello M, Rahman A. Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett October 1980;45(14):1196–9.
[42] Pazhang Mohammad, Mehrnejad Faramarz, Pazhang Yaghub, Falahati Hanieh, Chaparzadeh Nader. Effect of sorbitol and glycerol on the stability of trypsin and difference between their stabilization effects in the various solvents. Biotechnol Appl Biochem 2016;63(2):206–13.
[43] Pires Douglas EV, Ascher David B, Tom L. Blundell. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics February 2014;30(3):335–42.
[44] Rydeen Amy E, Brustad Eric M, Pielak Gary J. Osmolytes and protein–protein interactions. J Am Chem Soc June 2018;140(24):7441–4.
[45] Seo Jiwon, Singh Ravi, Ryu Jonghyuk, Choi Jun-Ho. Molecular aggregation behavior and microscopic heterogeneity in binary osmolyte–water solutions. J Chem Inf Model November 2023;64(1):138–49.
[46] Shukla Diwakar, Shinde Chetan, Trout Bernhardt L. Molecular computations of preferential interaction coefficients of proteins. J Phys Chem B September 2009;113(37):12546–54.
[47] Smolin Nikolai, Voloshin Vladimir P, Anikeenko Alexey V, Geiger Alfons, Winter Roland, Medvedev Nikolai N. TMAO and urea in the hydration shell of the protein SNase. Phys Chem Chem Phys 2017;19(9):6345–57.
[48] Soper AK, Castner EW, Luzar Alenka. Impact of urea on water structure: a clue to its properties as a denaturant? Biophys Chem September 2003;105(2–3):649–66.
[49] Sundar Smrithi, Sandilya Avilasha A, Priya M Hamsa. Unraveling the influence of osmolytes on water hydrogen-bond network: from local structure to graph theory analysis. J Chem Inf Model August 2021;61(8):3927–44.
[50] Timasheff SN. Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. Adv Protein Chem 1998;51:355–432.
[51] Timasheff Serge N. Water as ligand: preferential binding and exclusion of denaturants in protein unfolding. Biochemistry October 1992;31(41):9857–64.
[52] Timasheff Serge N. Protein hydration, thermodynamic binding, and preferential hydration. Biochemistry November 2002;41(46):13473–82.
[53] Timasheff Serge N. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. Proc Natl Acad Sci USA July 2002;99(15):9721–6.
[54] Van Der Spoel David, Lindahl Erik, Hess Berk, Groenhof Gerrit, Mark Alan E, Berendsen Herman JC. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701–18.
[55] Vishveshwara Saraswathi, Brinda KV, Kannan N. Protein structure: insights from graph theory. J Theor Comput Chem July 2002;01(01):187–211.
[56] Xia Kelin, Anand D Vijay, Shikhar Saxena, Mu Yuguang. Persistent homology analysis of osmolyte molecular aggregation and their hydrogen-bonding networks. Phys Chem Chem Phys September 2019;21(37):21038–48.
[57] Zoete Vincent, Cuendet Michel A, Grosdidier Aurélien, Michielin Olivier. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 2011;32(11):2359–68.