JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

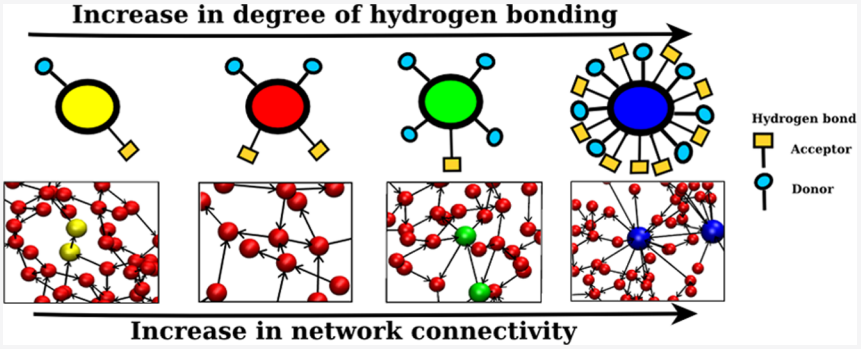
pubs.acs.org/jcim Article

Unraveling the Influence of Osmolytes on Water Hydrogen-Bond Network: From Local Structure to Graph Theory Analysis

Smrithi Sundar, Avilasha A. Sandilya, and M. Hamsa Priya*

Cite This: J. Chem. Inf. Model. 2021, 61, 3927−3944 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Water structure in aqueous osmolyte solutions, deduced from the slight alteration in the water−water radial distribution function, the decrease in water−water hydrogen bonding, and tetrahedral ordering based only on the orientation of nearest water molecules derived from the molecular dynamics simulations, appears to have been perturbed. A careful analysis, however, reveals that the hydrogen bonding and the tetrahedral ordering around a water molecule in binary solutions remain intact as in neat water when the contribution of osmolyte−water interactions is appropriately incorporated. Furthermore, the distribution of the water binding energies and the water excess chemical potential of solvation in solutions are also pretty much the same as in neat water. Osmolytes are, therefore, well integrated into the hydrogen-bond network of water. Indeed, osmolytes tend to preferentially hydrogen bond with water molecules and their interaction energies are strongly correlated to their hydrogen-bonding capability. The graph network analysis, further, illustrates that osmolytes act as hubs in the percolated hydrogen-bond network of solutions. The degree of hydrogen bonding of osmolytes predominantly determines all of the network properties. Osmolytes like ethanol that form fewer hydrogen bonds than a water molecule disrupt the water hydrogen-bond network, while other osmolytes that form more hydrogen bonds effectively increase the connectivity among water molecules. Our observation of minimal variation in the local structure and the vitality of osmolyte−water hydrogen bonds on the solution network properties clearly imply that the direct interaction between protein and osmolytes is solely responsible for the protein stability. Further, the relevance of hydrogen bonds on solution properties suggests that the hydrogen-bonding interaction among protein, water, and osmolyte could be the key determinant of the protein conformation in solutions.

INTRODUCTION

Osmolytes are the small organic compounds that can significantly influence the configurational stability and solubility of proteins by altering their conformational equilibria.[1] For example, sugars like trehalose and polyols like glycerol and sorbitol stabilize the folded conformation of proteins and, therefore, are referred to as protective osmolytes. They are also used as cryoprotectants to preserve and store biomolecules at low temperatures for future use.[2] Other osmolytes like ethanol and urea destabilize a protein by shifting the equilibrium toward unfolded conformations.[3] In fact, many organisms accumulate protective osmolytes to mitigate the deleterious effect of urea and other adverse conditions like high/low temperatures, salinity, and dehydration.[4,5]

The mechanistic action of osmolytes on protein structure has been extensively investigated, yet it is still inconclusive. One of the hypotheses is that addition of any compound to water would modulate the water structure.[6,7] Accordingly, the additives are classified as kosmotropes (structure maker) and chaotropes (structure breaker) based on their ability to enhance and destruct the water structure, respectively. Kosmotropes are expected to stabilize the protein structure,

Received: May 9, 2021
Published: August 11, 2021

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

while chaotropes would denature proteins. Another hypothesis is based on the preferential interaction among protein, water, and osmolyte in solution.[8] The osmolytes that can preferentially interact with protein atoms can enhance the solubility of protein in solution and could eventually denature them.[9] Some osmolytes, however, preferentially exclude from the protein surface and leave the protein preferentially hydrated.[8] These osmolytes could stabilize the folded conformation of a protein, and the osmotic force resulting from the exclusion of osmolytes could also lead to salting-out of the protein; therefore, they can be good crystallizing agents.[10] Additives with larger molecular volumes are expected to exclude from the protein surface as it would be entropically favorable to pack small water molecules around the undulatory protein surface. The interaction of proteins even with macromolecular additives like poly(ethylene glycol) (PEG) is, however, determined to be attractive through static light scattering.[11]

Based on the two postulates, the plausible mechanisms for urea-induced protein denaturation[12−19] are (i) direct mechanism—preferential interaction of urea with protein atoms; or (ii) indirect mechanism—alteration of the water structure by urea enhances the hydrophobic solvation that ultimately leads to dissolution of the protein hydrophobic core. Further research was oriented toward identifying the nature of the interaction (if any), *i.e.*, dispersion[13,17,20] *vs* electrostatic interaction;[12,15] determining the site of interaction, *i.e.*, the protein backbone *vs* the side chain;[13,19,20] and, furthermore, classifying them based on the polarity of the side-chain atoms, *i.e.*, hydrophobic, polar, or charged atoms. Similarly, three mechanisms have been put forth for the stabilizing action of trehalose: (i) mechanical entrapment or vitrification,[21,22] *i.e.*, trehalose vitrifies the solvent forming a glassy matrix that mechanically entraps the protein molecule by retarding its dynamics; (ii) water replacement,[23−25] *i.e.*, trehalose forms hydrogen bonds with protein atoms replacing protein−water hydrogen bonds; and (iii) water entrapment,[26−28] *i.e.*, massive trehalose excludes from the protein surface, so a thin layer of water is trapped between the protein surface and trehalose. Many studies posit that the protein structural stability is attributable to the direct interaction[12−17,19,20,24,25] of osmolytes with protein; however, some studies suggest that the effect of altered structure and dynamics of water by osmolytes cannot be ignored.[14,16,18,20,27]

To understand the influence of osmolytes on the water structure, the binary osmolyte−water solutions were examined by experimental techniques like infrared spectroscopy,[29−32] Raman spectroscopy,[33,34] nuclear magnetic resonance,[33,35] depolarized light scattering,[36] neutron scattering,[37] and neutron diffraction.[35,38,39] For instance, the vibration absorption spectrum of urea solution is found to be insensitive to urea concentration even at 8 M urea concentration.[29] A neutron diffraction study[38] has reported that only the secondary solvation shell of water gets collapsed in the urea solution. IR pump-probe spectroscopy studies, further, assure that the orientational dynamics of water in 8 M urea solution is the same as that in pure water[30,31] with only a very small fraction of water molecules, say one water molecule per urea molecule, exhibiting retarded dynamics possibly because of the multiple hydrogen bonds between those pairs of water and urea molecules. However, the influence of sorbitol on the water structure extends up to the second solvation shell and it increases the vibrational and the orientational relaxation time of water.[31]

The investigation of trehalose solution, through multiple experimental techniques, shows that trehalose perturbs the tetrahedrally hydrogen-bonded structure of water.[33] However, the quasi-electron neutron scattering study reports that trehalose is a kosmotrope, enhancing the water structure.[37] The depolarized light scattering, on the other hand, confirms that only the rotational dynamics of water in the first hydration layer of trehalose gets perturbed.[36] A neutron diffraction study pointed out that glycerol also disrupts the tetrahedrality of water although it compensates for the lost water−water hydrogen bonds.[39] The glycerol−water hydrogen bond is, however, found to progressively become linear with the increase in glycerol concentration.[32] The hydrogen-bond structure in dilute ethanol solution is deduced to be more structured than in pure water as the measured IR spectra for the OH stretch deviates negatively from the corresponding partial molar water spectra.[34]

Numerous computational studies have also investigated the structure and dynamics of water in the presence of osmolytes in binary solutions.[18,31,32,40−46] The water−water radial distribution function (rdf) and the spatial density of water indicate a slight collapse of the second solvation shell of water in 5 M urea solution.[18] The number of hydrogen bonds per water molecule is, however, almost the same as in neat water (NW)[41−43] when urea is also considered as the hydrogen-bond partner for the water molecule, revealing that urea very well substitutes for water in the hydrogen-bond network. Moreover, the orientational relaxation time[18] and the average lifetime of hydrogen bonds[41] also remain the same as in neat water. Contrastingly, the hydrogen-bond correlation time function is observed to decay more slowly in urea solution than in pure water, indicating that the water network becomes rigid in urea solution.[45] The observation of an increase in water−water hydrogen-bond energy with the increase in urea concentration also suggests urea as a structure maker.[42]

Trehalose has been observed to destruct the tetrahedral hydrogen-bond network of water because of the decrease in water−water hydrogen bonds.[40] The distribution of trehalose in the solution is, however, found to be homogeneous because of the higher propensity of the trehalose−water hydrogen bond in contrast to the observation of clustering of sugar molecules in sucrose or maltose solution.[40] The local tetrahedral order parameter computed from the orientation of the nearest water molecules is also found to decrease in urea solution;[44] however, when the orientation of the neighboring urea molecules is also considered, the tetrahedral order parameter remains unaltered from that in neat water.[43] The probability densities of water tetrahedrality in 2 M urea solution, after excluding the first two hydration shells around urea, have been reported to perfectly coincide with the tetrahedrality distribution of pure water at 305 K.[46] In glycerol solution, the average number of hydrogen bonds per water does not change only up to 0.15 mol fraction of glycerol; later, it decreases with the increase in glycerol concentration.[32] The number of water−water hydrogen bonds in sorbitol solutions is also reported to decrease.[31]

All of the inferences on water structure derived from molecular dynamics (MD) studies have been based on the radial distribution function, number of hydrogen bonds, order parameter, hydrogen bond lifetime, and the orientational correlation time. These analyses characterize the local

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

molecular arrangement but do not provide an entire solution-level picture of the hydrogen-bond network. Complex networks in various fields like transportation, social interactions, chemical reaction, and protein−protein interactions are typically analyzed using the graph network theory.[47,48] In this context, graph network analysis would definitely provide a global picture of hydrogen-bond interactions in solution. Graph theory, employed to understand the connectivity between water molecules at room and supercritical conditions,[49] revealed that the hydrogen-bond network of water is a percolated network at room temperature but exhibits a small-world pattern at supercritical conditions. It has also been applied to understand the clustering of ions,[50,51] methanol,[52,53] and osmolytes,[54] *viz.*, urea, trimethyl glycine, and sorbitol.

In this work, we have investigated the influence of several osmolytes like ethanol, glycerol, sorbitol, glucose (both the anomers), trehalose and urea, through molecular dynamics simulation. We have examined the effect of the water model, solution temperature, and osmolyte concentration on the solution hydrogen-bond network from local structural analysis to global network properties. Considering the directionality of the hydrogen bonds, we constructed a directed graph connecting the hydrogen-bond partners. We computed the ensemble averaged graph properties like the distribution of shortest paths between two molecules and the average distance between water molecules in a solution and determined how the presence of osmolytes influences the graph properties. We find that all of the osmolytes are well integrated into the hydrogen-bond network of the solution, complementing water molecules in maintaining the tetrahedral molecular arrangement about a central water molecule. Indeed, osmolytes act as central hubs that interconnect many water molecules in the solution, eventually reducing the average distance between water molecules compared to that in neat water. The observation of a perfect correlation between the average number of hydrogen bonds and the molecular binding energy of water and osmolytes emphasizes the importance of hydrogen bonds in solution thermodynamics.

METHODS

Simulation Details. The molecular dynamics (MD) simulations of neat water (system comprising only water molecules) and aqueous solutions of osmolytes such as ethanol, glycerol, α-glucose, β-glucose, sorbitol, trehalose, and urea were carried out using the GROMACS 2016.4 simulation package. The details of the osmolyte concentrations investigated and the number of osmolyte and water molecules in the simulation box are given in Table 1. In this work, we compare the properties of the osmolyte solution at 20 wt % because it is a moderate concentration within the experimentally relevant concentration range for these osmolytes, neither a very dilute nor a highly concentrated solution exhibiting crowding effect. Furthermore, the total number of hydrogen bonds (∼1000 ± 20) and the number of hydrogen bonds involving osmolytes (∼130 ± 30) are comparable among these systems. In solutions at a given molarity, there would be multifold variation in the number of hydrogen bonds involving osmolytes; hence, the comparison among the solutions may not be appropriate. We have also investigated the effect of osmolyte concentration on solution properties for trehalose and urea.

The initial structure of trehalose was built using GLYCAM Carbohydrate Builder,[55] and others were obtained from the

Table 1. Details of the Systems Studied[a]

system name	osmolyte	concentration		N_{osm}	N_{wat}
		wt %	M		
NW					512
EL	ethanol	20	4.1	50	512
GL	glycerol	20	2.2	25	512
SL	sorbitol	20	1.2	13	525
AG	α-glucose	20	1.2	13	514
BG	β-glucose	20	1.2	13	514
T1	trehalose	10	0.3	3	512
T2	trehalose	20	0.6	7	533
T3	trehalose	40	1.4	19	542
U1	urea	6	1.0	10	514
U2	urea	20	3.6	39	512
U3	urea	43	8.1	120	533

[a]NW, neat water; N_{osm}, number of osmolyte molecules; N_{wat}, number of water molecules.

CHARMM small-molecule library.[56] The osmolytes were modeled using the CHARMM36 force field,[57] one of the popular force fields for protein simulations, because the techniques and inference derived for the binary solutions in this work can be subsequently extended and validated for ternary solutions comprising protein, water, and osmolyte. To investigate the influence of the water model on hydrogen-bonding interactions, we carried out the simulations with the two most widely used three-site water models: the SPC/E water model[58] and the CHARMM modified TIP3P water model.[59,60] The former is the simplest water model known to accurately reproduce the water structure,[61] and the latter is the optimum water model for the CHARMM force field. The starting configuration of every solution was obtained by randomly inserting osmolyte molecules in an equilibrated water box, ensuring that there were no heavy atoms of the other osmolyte molecules or water oxygens within 0.4 nm of every heavy atom of the inserted molecule.

Each system was first energy minimized using the steepest descent algorithm for 50 000 steps with an energy gradient tolerance of 10 kJ mol^{-1} nm^{-1}. The water molecules in the system were first equilibrated for 200 ps at the NVT condition at 298 K using the weak Berendsen thermostat,[62] by restraining the osmolyte molecules with a force constant of 1000 kJ mol^{-1} nm^{-2}. Next, NPT equilibration was carried out in four stages with 300 ps of mdrun in each stage at 298 K and 1 bar by coupling to the Berendsen thermostat and the Berendsen barostat,[62] respectively. The heavy atoms of the osmolyte molecules were restrained in the first three stages with force constants of 1000, 100, and 10 kJ mol^{-1} nm^{-2}. In the fourth stage of NPT equilibration, all of the positional restraints were completely removed. Following this, an mdrun of 5 ns was carried at 298 K and 1 bar pressure by coupling with the Nosé−Hoover thermostat[63,64] and the Parrinello−Rahman barostat,[65] respectively. An integration time step of 2 fs was used for all simulations. The van der Waals interactions were calculated using the force-switch scheme with an inner cutoff radius of 1.0 nm and an outer cutoff radius of 1.2 nm. The particle mesh Ewald (PME)[66] method was used for the electrostatic interactions with a grid spacing of 0.12 nm and a real space cutoff of 1.2 nm. All bonds containing hydrogen atoms were constrained using the SHAKE[67] algorithm. The configurations saved every 1 ps of the last 3 ns of the final NPT run were used for further analysis. The average and the

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

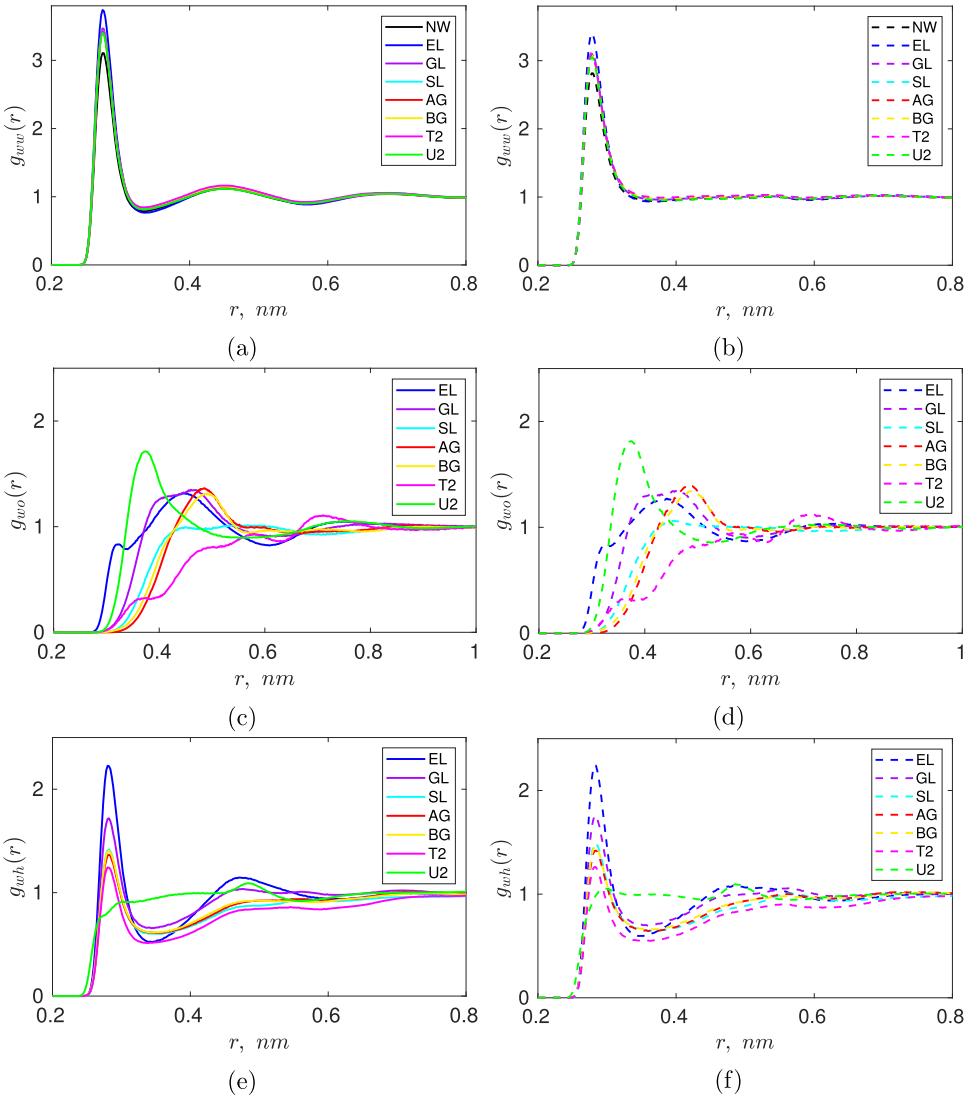


Figure 1. Radial distribution functions (rdfs) about the oxygen of the central water molecule (w) in neat water and in 20% aqueous solutions: (a, b) water oxygen (w); (c, d) osmolyte center of mass (o); and (e, f) heavy atoms of osmolyte (h) that can participate in a hydrogen bond. Legend: black, neat water (NW); blue, ethanol solution (EL); violet, glycerol solution (GL); cyan, sorbitol solution (SL); red, α-glucose solution (AG); yellow, β-glucose solution (BG); magenta, trehalose solution (T2); green, urea solution (U2); solid lines, solutions with SPC/E water; dashed lines, solutions with TIP3P water.

standard deviation of the computed parameters were obtained by splitting the trajectory into three blocks with 1000 configurations each. For some of the systems, we further extended the simulation run for 10 ns. The parameters estimated from the last 10 ns run (using the configurations saved every 10 ps) were quantitatively matching with values obtained in a 1 ns block, ensuring that systems have sufficiently got equilibrated and reported values are indeed well converged.

Coordination Number and Preferential Contact Ratio. The coordination number, n_i, *i.e.*, the number of nearest neighbors around a central water molecule, was determined from the radial distribution function (rdf), $g_{wi}(r)$

$$n_i = 4\pi\rho_{i,b}\int_0^R g_{wi}(r)r^2\,dr \tag{1}$$

where i corresponds to either water (wat) or osmolyte (osm), $\rho_{i,b}$ is the bulk density of the species i, and R is the cutoff distance of 0.35 nm. The water coordination number, n_{wat}, was obtained from water−water rdf (g_{ww}) using oxygen of water as its solvent center. The rdf corresponding to the distribution of potential hydrogen-bond partners of osmolyte molecules (g_{wh}) around a water molecule was used to compute the osmolyte coordination number, n_{osm}. The bulk densities were determined as the ratio of the total number of the corresponding species in the system to the average simulation box volume, V

$$\rho_{wat,b} = \frac{N_{wat} - 1}{V};\ \rho_{osm,b} = \frac{N_{osm} \times N_h}{V}$$

where N_{wat} and N_{osm} are the total numbers of water molecules and osmolyte molecules, respectively, in the simulation box, and N_h is the number of heavy atoms per osmolyte that can form hydrogen bonds. The number 1 in the numerator of the expression for water bulk density ($N_{wat} - 1$) accounts for the fact that one water molecule is considered as the central solute molecule. The preferential contact ratio,[20] γ, of osmolytes was calculated as the ratio of the local osmolyte to water coordination number normalized by the ratio of the total

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

number of osmolyte atoms that can form a hydrogen bond to the total number of water molecules

$$\gamma = \frac{n_{osm}}{n_{wat}} \frac{N_{wat} - 1}{N_{osm} \times N_h} \tag{2}$$

It is to be noted that N_h atoms of the osmolyte may not be able to be involved in hydrogen bonds independent of each other as they are topologically connected, and it is very likely that only one of the heavy atoms of a large osmolyte will hydrogen-bond with the central water molecule, so we define a modified preferential contact ratio, $\tilde{\gamma}$, as

$$\tilde{\gamma} = \frac{n_{osm}}{n_{wat}} \frac{N_{wat} - 1}{N_{osm}} \tag{3}$$

Interaction Energy. The interaction energies were calculated as a sum of dispersion interaction and electrostatic interaction of a chosen molecule with the rest of the system. The dispersion interactions were computed by the force-switch scheme with inner and outer cutoff radii of 1.0 and 1.2 nm, respectively. The electrostatic interactions were computed through the generalized reaction field scheme with appropriate finite box size correction.[68] A cutoff distance of 1.2 nm was used for electrostatic interactions. The intramolecular interactions providing the details of the intrinsic conformational energy of an osmolyte molecule are usually positive and are independent of the solvent environment. Thus, we have considered only the intermolecular interaction energy characterizing the interaction between the molecule of interest and the solvent environment.

Tetrahedral Order Parameter. The local structure around a water molecule can be expressed in terms of the tetrahedral order parameter[69]

$$q_{4,i} = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos \theta_{jik} + \frac{1}{3} \right)^2 \tag{4}$$

where θ_{jik} is the angle formed by neighboring molecules j and k with the central molecule i. The summation involves six such angles accounting for the orientation of the four nearest neighbors of the central molecule. Although the tetrahedral order parameter of a molecule spans the range $-3 \leq q_{4,i} \leq 1$, the normalization factor of 3/8 in eq 4 ensures that the ensemble averaged tetrahedral order parameter, $\langle q_4 \rangle$, varies only within the interval (0,1). For a perfect tetrahedral arrangement of molecules $\langle q_4 \rangle = 1$, on the other hand, $\langle q_4 \rangle = 0$ implies a random mutual arrangement of molecules. We computed $q_{4,i}$ for a water molecule in a binary solution using the coordinates of the proximal heavy atom of its four neighboring residues.

Hydrogen Bond. A hydrogen bond was determined as per the geometric criteria, *i.e.*, the distance between the donor (D) and the acceptor (A) atoms is less than 0.35 nm and the angle between the hydrogen atom, the donor, and the acceptor (∠HDA) is less than or equal to 30°. The closer the distance between the donor and the acceptor atoms, the stronger is the hydrogen bond. Also, a hydrogen bond is strong when all three atoms—the donor, the associated hydrogen, and the acceptor—align in a straight line, *i.e.*, $\theta \to 0 \equiv \cos\theta \to 1$. In short, the strength of a hydrogen bond is inversely proportional to the D−A distance and directly proportional to the cosine of the angle. We characterize the strength of hydrogen bonds through normalized weights, w, varying within the interval (0,1), such that 1 represents a strong hydrogen bond

$$w = \left(\frac{\frac{1}{r_{DA}} - \frac{1}{r_{max}}}{\frac{1}{r_{min}} - \frac{1}{r_{max}}} \right) \left(\frac{\cos\theta - \cos\theta_{max}}{\cos\theta_{min} - \cos\theta_{max}} \right) \tag{5}$$

The closest distance of separation between oxygen atoms of water molecules in a solution is 0.23 nm, as observed from water−water rdf, $g_{ww}(r)$, in Figure 1a,b, because for any distance less than 0.23 nm, $g_{ww}(r)$ is absolutely zero. Hence, we used a r_{min} value of 0.23 nm and the r_{max} of 0.35 nm corresponding to the cutoff distance as per the hydrogen-bond geometric criteria. Similarly, $\theta_{min} = 0°$ and $\theta_{max} = 30°$ were used corresponding to the theoretically most preferred hydrogen-bond orientation and the angular cutoff of the hydrogen-bond geometric criteria, respectively.

Graph Network Analyses. A directed network graph was generated for a given system configuration by considering every residue as a node in the network. The nodes were connected if there exists a hydrogen bond between those residues and the edge direction is from the hydrogen-bond donor to the acceptor residue. All network analyses were carried out using the NetworkX 2.3 module in Python 3.7. As there were instances of two, sometimes three, hydrogen bonds between a pair of nodes in the same direction, we used MultiDigraph of the NetworkX module to generate the graph for the configurations. About 3% of the residues donated multiple hydrogen bonds to the same residue in urea solution, while such an occurrence is found to be only 0.5% in other osmolyte solutions.

The shortest paths, *i.e.*, the minimum number of edges (hydrogen bonds) connecting every pair of nodes, were determined. The closeness centrality of a node, $C_c(u)$, provides the average shortest distance, in terms of the number of hydrogen bonds, to the node from all possible reachable nodes. We used the Wasserman and Faust[70] improved formula to calculate the closeness centrality

$$C_c(u) = \frac{n-1}{N-1} \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)} \tag{6}$$

where n and N are the number of reachable nodes to the node u and the total number of nodes in the system, respectively; and $d(u,v)$ is the shortest distance between the node u and a reachable node v. The inward distance to the node was used in calculating the closeness centrality; therefore, $C_c(u)$ assumes zero either if the corresponding residue does not accept any hydrogen bond or does not participate in hydrogen bonding at all. The betweenness centrality of a node, $C_b(u)$, measures the fraction of the shortest paths in the system that pass through the node u, excluding those paths that are originating from or terminating at the node u

$$C_b(u) = \sum_{\text{all}(s,t)} \frac{np(s,t|u)}{np(s,t)} \tag{7}$$

where $np(s,t)$ is the number of shortest paths between nodes s and t and $np(s,t|u)$ is the number of such paths that pass through the node u. The betweenness centrality was normalized by the maximum number of possible paths in the directed network, *i.e.*, $(N-1)(N-2)$, with N corresponding to the total number of nodes in the network. It is worth noting that a hydrogen-bond network would not be a complete graph.

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

Table 2. Coordination Number around a Water Molecule in 20 wt % Aqueous Solutions[a]

		SPC/E water model					TIP3P water model				
system name	N_h	n_{wat}	n_{osm}	n_{nb}	γ	$\tilde{\gamma}$	n_{wat}	n_{osm}	n_{nb}	γ	$\tilde{\gamma}$
NW	1	5.1		5.1			5.5		5.5		
EL	1	4.4	0.3	4.7	0.7	0.7	4.7	0.3	5.0	0.7	0.7
GL	3	4.6	0.4	5	0.6	1.8	4.9	0.4	5.3	0.6	1.7
SL	6	4.7	0.4	5.1	0.6	3.4	5.1	0.4	5.5	0.5	3.2
AG	6	4.7	0.4	5.1	0.6	3.4	5.1	0.4	5.5	0.5	3.1
BG	6	4.7	0.4	5.1	0.6	3.4	5.1	0.4	5.5	0.5	3.1
T2	10	4.8	0.3	5.1	0.5	4.8	5.1	0.3	5.4	0.5	4.5
U2	3	4.6	0.6	5.2	0.6	1.7	4.9	0.7	5.6	0.6	1.9

[a] N_h, number of heavy atoms of an osmolyte molecule that can participate in a hydrogen bond; n_{wat}, water coordination number; n_{osm}, osmolyte coordination number; n_{nb}, total number of neighboring atoms ($n_{nb} = n_{wat} + n_{osm}$); and γ and $\tilde{\gamma}$, preferential contact ratios defined in eqs 2 and 3, respectively.

For example, theoretically, every water molecule accepts two hydrogen bonds and donates two hydrogen bonds, so in a typical hydrogen-bond network for neat water, a node is connected to only four nodes (two inward edges, two outward edges). Still, we normalized the betweenness centrality by dividing by the term $(N-1)(N-2)$ to eliminate any bias due to the difference in the number of nodes in the various systems studied.

We evaluated the following parameters to compare the network properties in binary solutions to that in neat water. The relative mean path length was calculated as the difference in the mean path lengths in a solution and in neat water

$$\Delta l_p = \langle l_{p,soln} \rangle - \langle l_{p,NW} \rangle \tag{8}$$

The effective betweenness/closeness centrality measure of osmolytes was computed as the ratio of the average betweenness/closeness centrality of an osmolyte in solution to the average betweenness/closeness centrality of water in neat water

$$\eta_{osm} = \frac{\langle C_b^{osm} \rangle}{\langle C_b^{NW} \rangle}; \ \chi_{osm} = \frac{\langle C_c^{osm} \rangle}{\langle C_c^{NW} \rangle} \tag{9}$$

The effective closeness (ξ) of water in solution, *i.e.*, the reduction in the average distance between water molecules, was determined from the average closeness centrality of water when osmolytes were included in, $\langle C_{c,wat} \rangle$, and excluded from, $\langle \tilde{C}_{c,wat} \rangle$, the network

$$\xi = \frac{\langle C_{c,wat} \rangle - \langle \tilde{C}_{c,wat} \rangle}{\langle C_{c,wat} \rangle} \tag{10}$$

RESULTS AND DISCUSSION

Preferential Interaction in Osmolyte Solutions. Water−water rdf, $g_{ww}(r)$, values in neat water and in aqueous osmolyte solutions, considering oxygen of water molecule as the solvent center, are shown in Figure 1a,b for the simulations using SPC/E and TIP3P water models, respectively. The rdfs of the two water models are quite different[61]—the rdf of SPC/E water is structured with a primary peak at 0.28 nm and the secondary peak at ∼0.45 nm, whereas the rdf of neat TIP3P water shows only a primary peak at 0.28 nm. In the presence of osmolytes, the position(s) of the peak(s) in $g_{ww}(r)$, of both water models, remain unaltered; however, the magnitude of the peak height gets slightly enhanced. In ethanol solution, the magnitude of the peak increases by 0.6 units, while in other solutions, it increases by 0.3 units. We can observe that the first minimum is a little deeper in ethanol solution compared to that in neat water or other 20% solutions.

Table 2 lists the number of water molecules in the first solvation shell—water coordination number—computed from $g_{ww}(r)$ (eq 1). The water coordination number is found to be less in binary solutions than that in neat water for both the water models—0.4 water molecules less in sugar and sorbitol solutions, 0.6 water molecules less in urea and glycerol solutions, and 0.8 water molecules less in ethanol solution. Surprisingly, the water coordination number is the lowest in ethanol solution even though the magnitude of the primary peak of $g_{ww}(r)$ is the highest among all solutions. This is mainly attributable to the deep first minimum in ethanol solution as $g_{ww}(r)$ gets weighted by the r^2 term in the integral in eq 1. In a similar context, the water coordination number in TIP3P solutions is higher than that in SPC/E solutions because of the well-defined first minimum in the SPC/E water model.

Water−osmolyte rdfs, g_{wo}, considering the center of mass of an osmolyte molecule as its solvent center, in 20 wt % aqueous solutions are shown in Figure 1c,d for the simulations using SPC/E and TIP3P water models, respectively. The choice of a water model has a little effect on these rdfs. The distribution exhibits multiple peaks of varying heights as it is a convolution of multiple site−site rdfs of various functional groups of the osmolytes. Ethanol exhibits a small peak of about 0.32 nm followed by a prominent peak of about 0.43 nm. Urea shows a prominent primary peak at about 0.37 nm and a small secondary peak at 0.75 nm. Glycerol has two nearby peaks at 0.39 and 0.47 nm. Both α and β anomers of glucose have a prominent primary peak at about 0.49 nm, and sorbitol does not have a distinct peak but the g_{wo} reaches unity at about 0.43 nm. Trehalose exhibits a steplike distribution with peaks at 0.36, 0.46, 0.58, and 0.72 nm, with only the magnitude of the last peak being above unity. The radial distance at which g_{wo} crosses unity (for the first time) is in the order almost corresponding to the molecular size of the osmolyte: urea < ethanol ∼ glycerol < sorbitol ∼ α-glucose ∼ β-glucose < trehalose. As the molecular size of the osmolytes is larger than a water molecule, their centers of mass are found to be at a large distance; indeed, as per g_{wo}, the osmolyte molecules are not found at a distance less than 0.3 nm.

To identify any possibility of close interactions, especially hydrogen bonding, between osmolytes and water, we looked at the rdf of potential hydrogen-bonding partners of the osmolytes, g_{wh}, around a central water molecule (Figure 1e,f) for both the water models. The choice of the water model,

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

Table 3. Atomwise Listing of Average Number of Hydrogen Bonds per Osmolyte Molecule[a]

osmolyte	atomwise averages											molecular average
						Water Model: SPC/E						
EL	2.4											2.4
GL	2.0	1.9	2.0									5.9
SL	2.0	1.8	1.5	1.6	1.8	2						10.7
AG	1.6	1.7	1.8	1.8	0.4	2.1						9.4
BG	2.0	1.8	1.8	1.8	0.5	2.0						9.9
T2	0.0	1.8	1.9	1.8	0.3	2.1	1.8	1.9	1.8	0.3	2.1	15.8
U2	1.5	2.3	1.5									5.3
						Water Model: TIP3P						
EL	2.4											2.4
GL	2.0	1.9	2.0									5.9
SL	2.0	1.8	1.5	1.6	1.8	2						10.7
AG	1.7	1.7	1.9	1.8	0.4	2.1						9.5
BG	2.0	1.8	1.8	1.8	0.5	2.1						10.0
T2	0.0	1.8	1.9	1.8	0.3	2.1	1.8	1.9	1.8	0.3	2.1	15.8
U2	1.6	2.3	1.6									5.5

[a]List of atom names for every osmolyte molecule: ethanol (EL), (O); glycerol (GL), (O1, O2, O3); sorbitol (SL), (O1, O2, O3, O4, O5, O6); α-glucose (AG), (O1, O2, O3, O4, O5, O6); β-glucose (BG), (O1, O2, O3, O4, O5, O6); trehalose (T2), (O11, O21, O31, O41, O51, O61, O22, O32, O42, O52, O62); urea (U2), (N1, O2, N3).

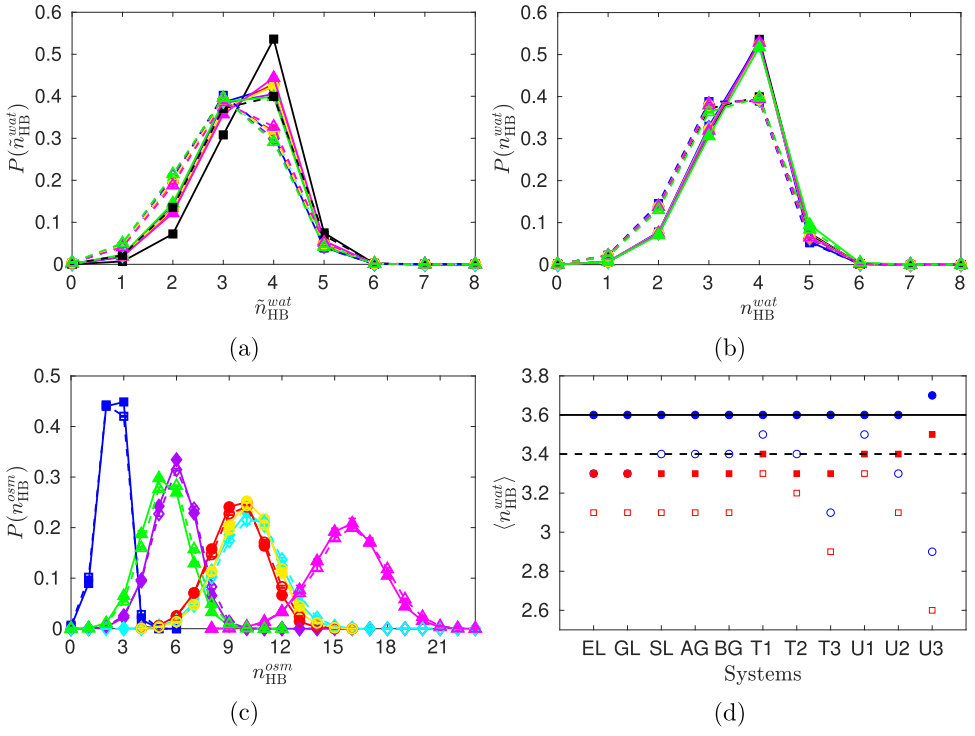


Figure 2. Probability distribution of the number of hydrogen bonds per water molecule in neat water and in 20% aqueous binary solution (a) when only water−water hydrogen bonds are considered ($\tilde{n}_{HB}^{wat}$) and (b) when both water−water and water−osmolyte hydrogen bonds are considered (n_{HB}^{wat}). (c) Probability distribution of the number of intermolecular hydrogen bonds per osmolyte (n_{HB}^{osm}) molecule in 20% aqueous solutions. The associated error bars are smaller than the symbol size. Legend: black square, neat water; blue square, ethanol solution; violet diamond, glycerol solution; cyan diamond, sorbitol solution; red circle, α-glucose solution; yellow circle, β-glucose solution; magenta triangle, trehalose solution; green triangle, urea solution; closed symbols and solid lines, solutions with SPC/E water; open symbols and dashed lines, solutions with TIP3P water. Lines are drawn connecting the data points for visual clarity. (d) Average number of hydrogen bonds per water molecule ($\langle n_{HB}^{wat} \rangle$) in all of the aqueous solutions investigated. Legend: blue circles, solutions with SPC/E water; red squares, solutions with TIP3P water; open symbols, considering water−water hydrogen bonds only; closed symbols, considering both water−water and water−osmolyte hydrogen bonds; solid line, $\langle n_{HB}^{wat} \rangle$ in neat SPC/E water; dashed line, $\langle n_{HB}^{wat} \rangle$ in neat TIP3P water.

however, does not seem to have any significant effect on the g_{wh} of osmolytes. The list of heavy atoms of osmolytes that can potentially form a hydrogen bond with a water molecule is provided in the footnote of Table 3. As all osmolytes, except urea, comprise hydroxyl groups, their g_{wh} values are qualitatively similar with a primary peak at 0.28 nm, the first minimum at 0.34−0.35 nm, and the secondary peak at 0.48 nm. The peak heights (both primary and secondary peaks) of g_{wh} in ethanol solution are the highest, followed by glycerol solution. The primary peak height in sorbitol and sugar

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

solution is almost the same; the small variation in their rdfs is observed in the magnitude of the first minimum and the secondary peak. Contrastingly, the g_{wh} of urea does not show any prominent peak with the rdf oscillating about 1.0 from 0.3 nm, indicating a spherical symmetry of the molecule with respect to its hydrogen-bonding capability as both the amine groups and the carbonyl group of urea can hydrogen bond with a water molecule.[38] Moreover, it has been reported that urea forms a nearly ideal solution with water.[41]

The number of osmolyte hydrogen-bond partners in the first solvation shell of a water molecule—osmolyte coordination number—determined from g_{wh} through eq 1 is given in Table 2. The osmolyte coordination number is found to be less than 1 in all binary solutions. Interestingly, the total number of molecules in the first solvation shell, computed from the sum of water and osmolyte coordination number, is almost constant and is equivalent to the coordination number in neat water for most of the binary solutions, except for ethanol and glycerol solutions. The coordination number clearly demonstrates that water molecules are predominantly surrounded by water molecules with only 6−13% of molecules being osmolytes. Any preferential interaction between water and osmolytes can be characterized by the preferential interaction parameter.[9] The Kirkwood−Buff theory[71] provides a neat theoretical framework to estimate the preferential interaction parameter from rdfs; however, it is practically difficult to exactly compute the Kirkwood−Buff integrals.[72] To obtain an accurate unbiased estimate, one has to adopt special strategies like smearing the rdfs[72] or adopting the quasi-chemical theory approach.[73] Since we are keen in the extent of hydrogen bonding in this study, we simply calculated the preferential contact ratio[20] based on the coordination number in the first solvation shell.

The preferential contact ratio (eq 2) for all binary solutions is determined to be 0.5−0.7, indicating that the water molecules get preferentially hydrated in the solutions—more in trehalose solution and less in ethanol solution. Equation 2, however, assumes that the N_h heavy atoms of an osmolyte can simultaneously form multiple hydrogen bonds with a given water molecule ignoring the topological connectivity between them. It has been found through MD simulations that trehalose molecules, comprising several hydrogen-bonding atoms, mostly form only one hydrogen bond with a protein.[28] It is very unlikely for a water molecule, whose solvation shell is significantly smaller than that of a protein, to form multiple hydrogen bonds with the same osmolyte molecule. Urea molecules can form multiple hydrogen bonds with the same water molecule, but such a population is very small.[30] When we simply normalize using the total number of osmolytes and water molecules in the simulation box as in eq 3, we obtain the preferential contact ratio, $\tilde{\gamma}$, greater than 1 in all osmolyte solutions, except for ethanol solution. We could deduce that there exists a preferential interaction between osmolyte and water; the extent of the trehalose−water preferential interaction is the highest among all of the investigated osmolytes. We will discuss shortly that osmolytes are found to preferentially hydrogen bond with water.

Water Hydrogen Bonding Remains Unaltered in Binary Solutions. Figure 2a shows the probability distribution of the number of water−water hydrogen bonds in neat water and in 20 wt % binary solutions. In neat SPC/E water, about 55% of water molecules form four hydrogen bonds with only 32% water molecules forming three hydrogen bonds; on the other hand, in neat TIP3P water, the fractional population forming four and three hydrogen bonds is equivalent (∼40%). This difference in the hydrogen-bonding capability between the water models can be ascribed to the difference in the water structure observed in their water−water rdfs (Figure 1a,b). In binary solutions, for both water models, the distribution of the number of water−water hydrogen bond shifts to the right—the population of water molecules forming four hydrogen bonds is seen to decrease by 10% and correspondingly the probability of forming only two or three hydrogen bonds increases. The water molecules in 20 wt % binary solutions form 0.2−0.3 less number of water−water hydrogen bonds than that in neat water; with an increase in osmolyte concentration (T3 or U3 systems), the number of water−water hydrogen bonds decreases further (Figure 2d). Many other MD studies have reported the decrease in water−water hydrogen bonds in osmolyte solutions and, hence, classified them as structure breakers.[31,32,40]

Although the number of water−water hydrogen bonds decreases in urea solution, the number of hydrogen bonds per water molecule has been reported to be matching with the number of hydrogen bonds in pure water upon inclusion of urea−water hydrogen bonds.[41−43] Following this, when we considered both water−water and water−osmolyte hydrogen bonds to compute the number of hydrogen bonds formed by a water molecule, we find the corresponding distribution (Figure 2b) to be coinciding onto a single curve for all osmolyte solutions. The average number of hydrogen bonds per water molecule in SPC/E solutions exactly matches that of neat water (3.6); only in the U3 system, the average value is slightly higher. In TIP3P solutions, the average number of hydrogen bonds formed by a water molecule is within 0.1 of the average value in neat water (3.4). This clearly demonstrates that the hydrogen-bonding capability of the water molecules is not compromised in binary solutions, and osmolytes readily compensate for lost water−water hydrogen bonds.

Osmolytes Prefer to Hydrogen Bond with Water. The probability distribution of the number of intermolecular hydrogen bonds formed by an osmolyte molecule (Figure 2c) is found to be independent of the water model. The number of intramolecular hydrogen bonds in an osmolyte has been determined to be minimal, almost negligible (data not shown). The average number of intermolecular hydrogen bonds increases with the increase in molecular size of the osmolyte in the following order: ethanol < urea < glycerol < sorbitol ∼ α-glucose ∼ β-glucose < trehalose (Table 3). The values of 15.8 for trehalose and 5.9 for glycerol are comparable to the experimentally determined hydration number of 16−18 for trehalose[74] and 5.58 for glycerol.[75] We find a urea molecule on average to form 5.3−5.5 hydrogen bonds, in close agreement with the value of 5.7 estimated from the empirical potential structure refinement (EPSR) simulation characterizing neutron scattering data[38] of urea solutions.

One can notice that the average number of hydrogen bonds for an osmolyte, except for urea, is roughly proportional to twice the number of hydroxyl (−OH) groups present in the molecule. The number of hydrogen bonds observed to be formed by every non-carbon heavy atom of the osmolytes molecules is provided in Table 3. The terminal hydroxyl groups in glycerol and sorbitol form two hydrogen bonds; however, the number of hydrogen bonds formed by the inner hydroxyl groups shows a steady decrease because of the steric restrictions. The hydroxyl groups attached to the middle carbons (C3 and C4) of sorbitol form only 1.5−1.6 hydrogen

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

bonds. In sugars, the hydroxyl groups form about 1.8 hydrogen bonds, but the contribution from the ring oxygen is only 0.3−0.5. The orientation of the hydroxyl group at the anomeric carbon is responsible for the small difference in the average number of hydrogen bonds between α- and β-glucose;[76] the equatorial orientation of the hydroxyl group at C1 in β-glucose increases the probability of the hydrogen bonding for the O1 group. The carbonyl oxygen of urea forms (accepts) 2.3 hydrogen bonds and the amine groups of urea are involved in only 1.6 hydrogen bonds each, in consensus with the EPSR simulation.[38]

The intermolecular hydrogen bonds are further classified into four classes based on the donor → acceptor pairs: osmolyte → osmolyte, osmolyte → water, water → osmolyte, and water → water. The observed population of each of the classes is listed in Table 4. The water → water hydrogen bonds

Table 4. Percentage Population of Various Intermolecular Donor → Acceptor Pairs of Hydrogen Bonds in Solutions[a]

system name	SPC/E water model				TIP3P water model			
	p_{oo}	p_{ow}	p_{wo}	p_{ww}	p_{oo}	p_{ow}	p_{wo}	p_{ww}
EL	0.5	4.0	7.4	88.1	0.5	4.4	7.8	87.4
GL	0.7	5.5	8.0	85.8	0.7	6.0	8.6	84.7
SL	0.8	4.8	6.6	87.8	0.6	5.5	7.3	86.6
AG	0.6	4.9	6.4	88.1	0.5	5.4	7.0	87.1
BG	0.7	4.9	6.8	87.6	0.5	5.5	7.5	86.5
T1	0.1	2.1	2.8	95.0	0.1	2.3	3.0	94.6
T2	0.5	4.2	5.6	89.6	0.5	4.5	6.1	88.9
T3	2.5	9.2	12.4	75.9	2.1	10.3	13.6	74.0
U1	0.1	2.7	2.5	94.7	0.1	3.0	2.8	94.1
U2	1.6	8.8	8.0	81.6	1.6	9.7	8.9	79.8
U3	8.5	17.5	15.9	58.1	8.5	19.2	17.4	54.9

[a] p_{ij} is the percentage population of intermolecular hydrogen bonds donated by species i to the acceptor species j. The indices o and w correspond to osmolyte and water, respectively. The associated standard error is less than 0.1−0.2.

are, of course, predominant in all solutions corresponding to 80−90% of the total hydrogen bonds in 20 wt % solutions owing to the higher number of water molecules in the system. The variation in the hydrogen-bond population because of the water model used is found to be less than 1−3%, with a maximal variation observed in the 8 M urea solution. Interestingly, the osmolyte → osmolyte hydrogen bond is almost negligible (less than 1%) in many solutions investigated; only in the highest urea concentration (8 M), about 8.5% of urea−urea hydrogen bonds was observed, still considerably (about 3-fold) lower than the population of hydrogen bonds between a water and an osmolyte molecule. Thus, we can confidently deduce that osmolytes prefer to hydrogen bond with a water molecule.

In alcohol and sugar solutions, the population of water → osmolyte hydrogen bonds is slightly higher than the osmolyte → water hydrogen bonds, revealing that these osmolytes readily accept hydrogen bonds from water than the extent they donate to water. The presence of two hydrogens per water molecule (compare it to one hydrogen in the −OH group of the alcohols or sugars) enables a higher probability of water donating hydrogen bond to the osmolyte, and also water molecules being comparatively small in size can easily orient their hydrogen atom to facilitate hydrogen bonding with these massive osmolytes. Exceptionally, the number of hydrogen bonds donated by urea to water is slightly larger than the number of hydrogen bonds urea accepts from water. Once again, the availability of four hydrogens in a urea molecule can be ascribed to its higher propensity to donate hydrogen bonds than a water molecule with two hydrogens.

Strength of Water−Water Hydrogen Bonds Is the Strongest and Remains Unaltered in the Presence of Osmolytes. The strength of the hydrogen bonds is characterized by the normalized weights as per eq 5, based on the distance of separation of donor−acceptor atoms and the involved angle in the hydrogen-bond definition. A higher weight implies a closer donor−acceptor distance and the linearity of the hydrogen bond and, therefore, a stronger hydrogen bond.

The probability distributions of the normalized weights of the hydrogen bonds in neat water and 20 wt % binary solutions are shown in Figure S1. The weights of the hydrogen bonds vary from 0.025 to 0.8, with a mean of 0.4. The low weight for a hydrogen bond (<0.1) is mainly due to either one or both of the geometric conditions being close to the upper threshold limit. The most striking observation is that the distribution for urea is very different from the other osmolytes because in other osmolytes the hydrogen-bond partners are similar chemically (hydroxyl group), whereas in the case of urea the hydrogen-bond acceptor is the carbonyl oxygen and the donors are the amine groups. It is worth noting that the strength of water−water hydrogen is higher in the SPC/E water model than that in the TIP3P water model. The presence of osmolytes does not alter the strength of the hydrogen bonds between water molecules as depicted by the exactly same distribution of weights for water−water hydrogen bonds in all of the systems.

The relative average weights of the hydrogen bonds for various donor → acceptor pairs, normalized with respect to the average weight of water → water hydrogen bonds in neat water, are listed in Table 5. In general, water → water hydrogen bonds are the strongest, while osmolyte → osmolyte hydrogen bonds are the weakest. In urea solutions, irrespective of the water model, the urea → water hydrogen bond is

Table 5. Normalized Average Weights, Characterizing the Strength of Hydrogen Bonds of Various Donor → Acceptor Pairs, with Respect to That in Neat Water[a]

system name	SPC/E water model				TIP3P water model			
	w_{oo}	w_{ow}	w_{wo}	w_{ww}	w_{oo}	w_{ow}	w_{wo}	w_{ww}
NW				1.00				1.00
EL	0.76	0.83	0.88	1.01	0.89	0.99	0.92	1.02
GL	0.72	0.80	0.87	1.01	0.85	0.97	0.89	1.01
SL	0.69	0.80	0.86	1.01	0.83	0.95	0.88	1.01
AG	0.72	0.83	0.83	1.01	0.83	1.00	0.85	1.01
BG	0.71	0.84	0.82	1.01	0.84	1.01	0.84	1.01
T1	0.70	0.84	0.85	1.00	0.87	1.01	0.87	1.00
T2	0.76	0.84	0.86	1.01	0.88	1.01	0.88	1.01
T3	0.75	0.85	0.87	1.02	0.89	1.02	0.89	1.02
U1	0.54	0.50	1.05	1.00	0.66	0.61	1.06	1.00
U2	0.55	0.51	1.07	1.01	0.67	0.62	1.08	1.01
U3	0.57	0.52	1.10	1.03	0.69	0.63	1.10	1.03

[a] w_{ij} is the normalized average weight for the hydrogen bonds from the donor species i to the acceptor species j. The indices o and w correspond to osmolyte and water, respectively.

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

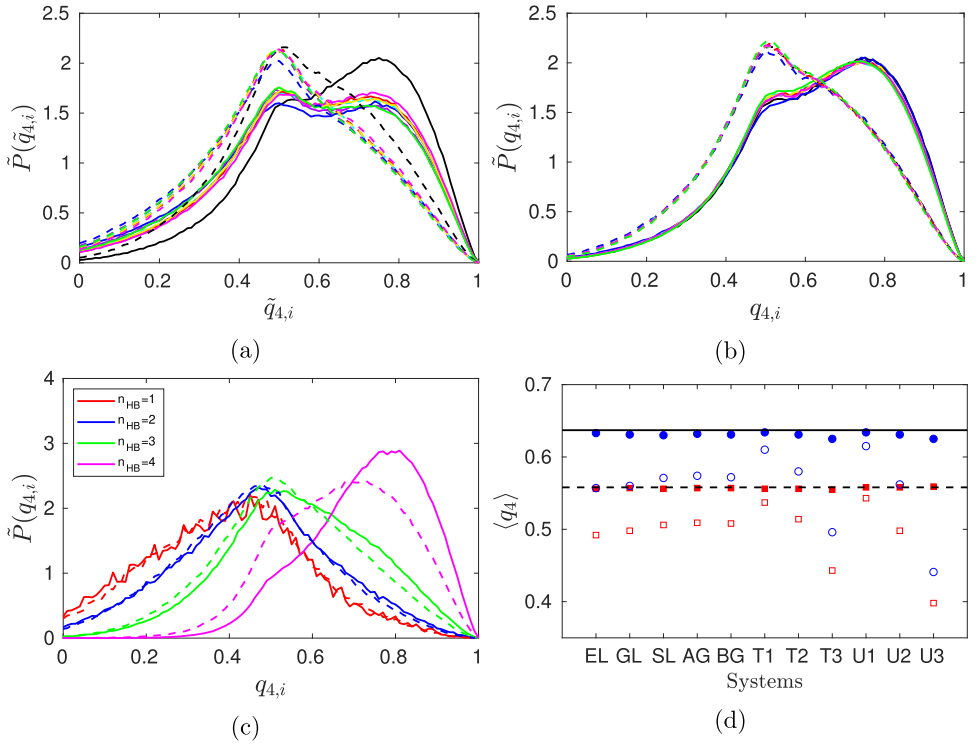


Figure 3. Distribution of probability density of the tetrahedral order parameter of water in neat water and in 20% osmolyte solution (a) considering only neighboring water molecules ($\tilde{q}_{4,i}$) and (b) including both the surrounding water and osmolytes molecules as neighbors ($q_{4,i}$). Legend: black, neat water; blue, ethanol solution; violet, glycerol solution; cyan, sorbitol solution; red, α-glucose solution; yellow, β-glucose solution; magenta, trehalose solution; green, urea solution; solid lines, solutions with SPC/E water; dashed lines, solutions with TIP3P water. (c) Probability density distribution in neat water conditioned on the number of hydrogen bonds with the nearest neighbors. Legend: solid lines, SPC/E water; dashed lines, TIP3P water. (d) Ensemble averaged tetrahedral order parameter ($\langle q_4 \rangle$) of water molecules in all of the systems studied. Legend: blue circles, solutions with SPC/E water; red squares, solutions with TIP3P water; open symbols, considering only neighboring water molecules; closed symbols, including both the surrounding water and osmolytes molecules as the neighbors; solid line, $\langle q_4 \rangle$ in neat SPC/E water; dashed line, $\langle q_4 \rangle$ in neat TIP3P water.

comparable to the urea → urea hydrogen bond and is almost half of the strength of the water → urea hydrogen bond. The trend and the relative magnitude of the weights in urea solution are proportional to the reported hydrogen-bond energies:[42] water−water (27.6 kJ/mol) > water−urea (24.1 kJ/mol) ≫ water−urea ≈ urea−urea (17 kJ/mol). This suggests that urea strongly accepts a hydrogen bond from water than it donates to water although the population of urea donating hydrogen bonds to water is higher (Table 4); however, there are instances of multiple hydrogen bonds between the same urea−water pair as suggested by the IR pump-probe spectroscopy study.[30] Furthermore, w_{wo} is slightly higher than w_{ww} in urea solution, indicating that carbonyl oxygen is a strong hydrogen-bond acceptor than a water molecule itself. In other binary solutions with SPC/E water, the strength of the hydrogen bonds between osmolyte and water is equivalent, irrespective of the direction of the hydrogen bond. Contrastingly, in TIP3P solutions, the strength of the hydrogen bonds with water as an acceptor is stronger than the case when water donates a hydrogen bond to osmolytes—$w_{wo} < w_{ow} \sim w_{ww}$.

Osmolytes Do Not Alter the Local Structure around a Water Molecule. The local molecular structure can be characterized by the tetrahedral order parameter based on the relative orientation of four nearest neighbors around a molecule of interest.[69] In the case of a perfect tetrahedral molecular arrangement with all nearest neighbors of the molecule subtending an angle of 109°, the molecular order parameter $q_{4,i}$ will assume a value of unity. The probability density distributions of the order parameter in neat water and in 20 wt % binary solutions using both SPC/E and TIP3P water models are shown in Figure 3a. The distribution in neat SPC/E water shows two peaks with a major peak at about 0.8 and a secondary peak at about 0.5, in agreement with other studies,[46,77] whereas the distribution of TIP3P water exhibits only one peak at 0.5. The observed difference in the distribution between the water models can be potentially attributed to their difference in the hydrogen-bonding capacity.

To ascertain the hypothesis, we looked at the conditional distribution of the molecular tetrahedral parameter in neat water based on the number of hydrogen bonds the central water molecule makes with the nearest neighbors (Figure 3c). When the central molecule forms less than four hydrogen bonds with its nearest neighbors, in both the water models, the conditional density distribution of the order parameter shows a single peak at about 0.5. Only the distribution for the water molecules that form hydrogen bonds with all their four closest neighbors shows a dominant peak at 0.8; still, a shoulder at 0.5 is observed as some water molecules form more than four hydrogen bonds (Figure 2a). Around 48% of SPC/E water molecules form three or less hydrogen bonds with their first four neighbors, while 68% of TIP3P water molecules form three or less hydrogen bonds with their first four neighbors. Because of the dominance of three or less hydrogen bonds of TIP3P water molecules with their nearest neighbors, the peak

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

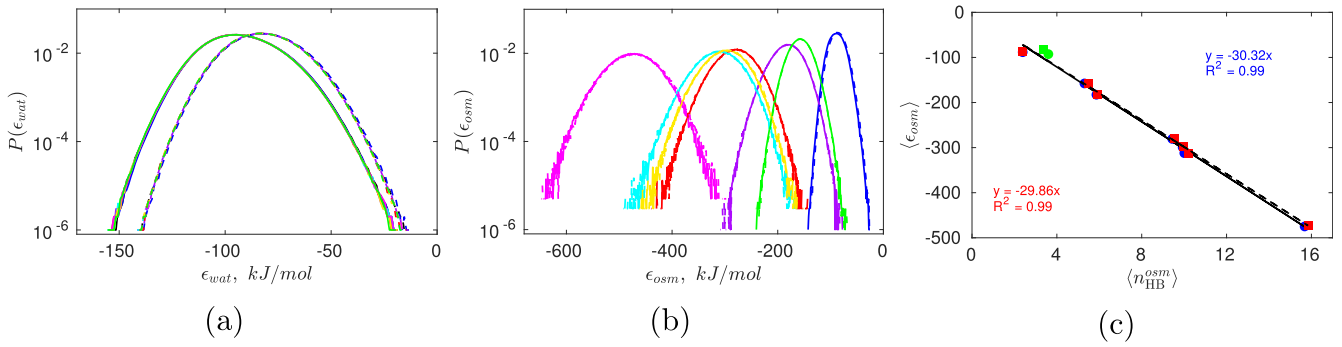


Figure 4. Probability distribution of binding energies of (a) water molecules (ϵ_{wat}) in neat water and in 20% aqueous solutions, (b) osmolytes (ϵ_{osm}) in the solutions. Legend: black, neat water; blue, ethanol solution; violet, glycerol solution; cyan, sorbitol solution; red, α-glucose solution; yellow, β-glucose solution; magenta, trehalose solution; green, urea solution; solid lines, solutions with SPC/E water; dashed lines, solutions with TIP3P water. (c) Correlation between the average binding energy ($\langle \epsilon \rangle$) and the average number of hydrogen bonds ($\langle n_{HB} \rangle$) formed by the molecule in solution. Legend: green circle, SPC/E neat water; green square, TIP3P neat water; blue circles, osmolytes in solution with SPC/E water; red squares, osmolytes in solution with TIP3P water; solid and dashed lines, linear fit for the systems with SPC/E and TIP3P water, respectively.

at 0.8 in the tetrahedral order parameter distribution is lost in TIP3P water.

In binary solutions, the distribution of the tetrahedral order parameter determined from the orientation of the four nearest water molecules (Figure 3a) shifted to lower $\tilde{q}_{4,i}$ values than that in neat water. Urea has been reported to be a structure breaker as it reduces the tetrahedral order parameter of water.[44] Another MD study[43] clarified that the tetrahedral arrangement seems to be lost when one considers the orientation of the four nearest water molecules as some of the four nearest water molecules might be part of the secondary hydration shell and do not participate in hydrogen bonding with the central water molecule. Further, they showed that the tetrahedral order parameter computed from the orientation of the nearest four neighbors, either urea or water molecules, was equivalent to that in pure water even at 9.3 M urea concentration. Similarly, when we consider the four nearest neighbors irrespective of their chemical identity, we find that the order parameter probability density distribution of all of the solutions coincides with the distribution of the neat water for the given water model (Figure 3b) as all of the nearest neighbors belong to the first solvation shell. The ensemble averaged tetrahedral parameter, $\langle q_4 \rangle$, shown in Figure 3d clearly demonstrates that upon inclusion of osmolyte molecules as the nearest neighbor, the local structure around a water molecule in solution remains intact. Furthermore, the dependence of the tetrahedral order parameter on the hydrogen bond re-emphasizes that osmolytes absolutely compensate for the loss in water−water hydrogen bonds in solutions.

Hydrogen Bond Determines the Thermodynamic Properties of the Solution. Similar to the distribution of the number of hydrogen bonds and the tetrahedral parameter, distribution of binding energies of water molecules (Figure 4a) also remains invariant in 20 wt % osmolyte solutions; only the variation due to the water model is observed. According to the potential distribution theory,[78] the excess chemical potential of solvation, μ^{ex}, depends on the binding energy distribution

$$\exp(\beta\mu^{ex}) = \int P(\epsilon) \exp(\beta\epsilon)\, d\epsilon$$

The coinciding energy distribution profile of the binary solutions, therefore, implies that the excess chemical potential of solvation of water in these solutions is the same in the dilute solutions investigated. Using the histogram overlap technique,[79] we estimated the excess solvation chemical potential as -27 ± 1 kJ/mol for TIP3P water and -30 ± 2 kJ/mol for SPC/E water; both are comparable to the experimental value of -26.5 kJ/mol in water[80] and other simulation studies of the corresponding water models within the simulation error.[81−83]

Figure 4b shows the distribution of the intermolecular energies of the osmolytes in 20 wt % solutions. The energy distribution is found to be independent of the water model. The mean interaction energy follows the following order: trehalose < sorbitol ∼ β-glucose ∼ α-glucose < glycerol ∼ urea < ethanol. The order is inverse of the order obtained for the number of hydrogen bonds formed by the osmolytes. Indeed, Figure 4c shows that the average interaction energy and the average number of hydrogen bonds of the osmolytes are perfectly inversely correlated—the more the number of hydrogen bonds an osmolyte can participate in the solution, the lower is its interaction energy in the solution. It has been recently reported that the binding energy of a water molecule in a confined environment is strongly correlated to the number of hydrogen bonds the water molecule forms with its neighbors.[84] The slope of -30 kJ/mol is a little higher than the hydrogen-bond energy of -20.6 kJ/mol[85] for a water dimer; however, it has been reported that cooperativity in associative liquids influences the strength of hydrogen bonds in water with an average energy of -27 kJ/mol in a trimer and -42 kJ/mol in a hexamer.[85] The correlation clearly demonstrates that hydrogen bonding determines the thermodynamic properties of the solution.

Osmolytes Are Central to the Percolated Hydrogen-Bond Network of Solutions. The analyses like rdf, tetrahedral order parameter, and number of hydrogen bonds per molecule provide insights into the local structure around a molecule. Since all thermodynamic properties of a solution are dominantly determined by hydrogen bonds, it is necessary to analyze the hydrogen-bond network at the solution level. Graph theory, a technique commonly used to probe the social and chemical networks,[47,48] is an apt tool to investigate the hydrogen-bond network for the whole solution. It will provide information on the role of osmolytes in the entire solution network. In this regard, we have transformed the molecular-level hydrogen bonding into a directed graph as depicted in

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

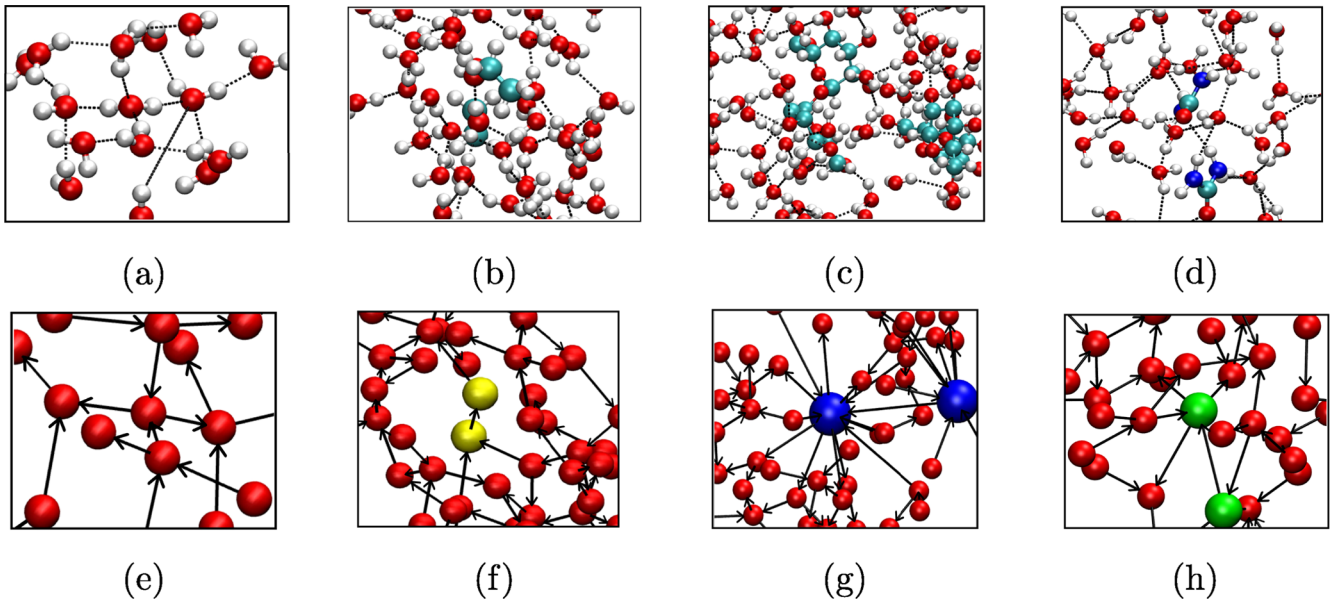


Figure 5. Top panel: representative VMD snapshots showing hydrogen bonds (dashed lines) between molecules (CPK representation). Color code: red, oxygen atoms; white, hydrogen atoms; cyan, carbon atoms; blue, nitrogen atoms. Bottom panel: directed hydrogen-bond network for the shown molecular configuration. Color code: red, water; yellow, ethanol; blue, trehalose; green, urea. Systems: (a, e) neat water; (b, f) 20% ethanol solution; (c, g) 20% trehalose solution; and (d, h) 20% urea solution.

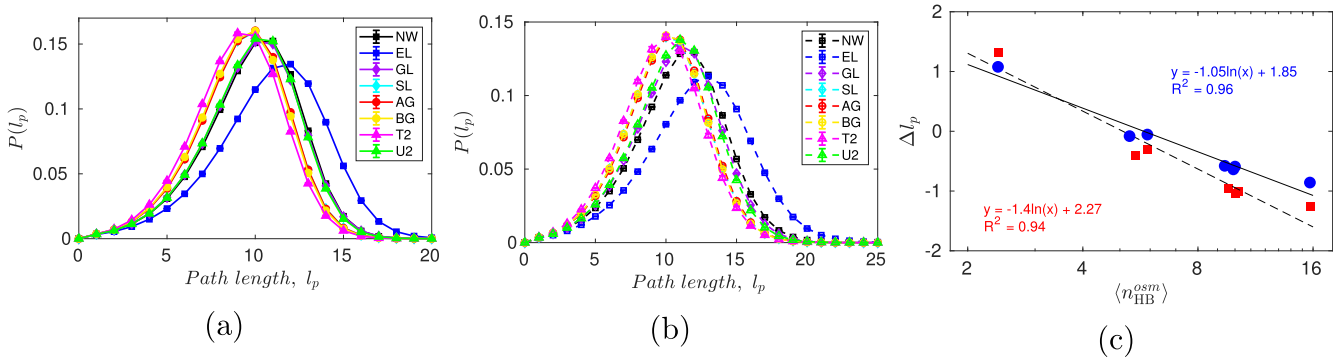


Figure 6. Probability distribution of the shortest path lengths (l_p) between every pair of the connected nodes in the network of solutions with (a) SPC/E water and (b) TIP3P water. Legend: black square, neat water (NW); blue square, ethanol solution (EL); violet diamond, glycerol solution (GL); cyan diamond, sorbitol solution (SL); red circle, α-glucose solution (AG); yellow circle, β-glucose solution (BG); magenta triangle, trehalose solution (T2); green triangle, urea solution (U2); closed symbols and solid lines, solutions with SPC/E water; open symbols and dashed lines, solutions with TIP3P water. (c) Power law dependence of the relative mean path length in solutions relative to that in neat water (Δl_p) on the average number of hydrogen bonds per osmolyte molecule ($\langle n_{HB}^{osm} \rangle$) in the solution. Legend: blue circles, systems with SPC/E water; red squares, systems with TIP3P water; solid and dashed lines are the fits to the data corresponding to the solutions with SPC/E and TIP3P water, respectively. The fit appears linear when we plot $\langle n_{HB}^{osm} \rangle$ in the log scale.

Figure 5. An earlier attempt to employ the graph theory on urea and sorbitol solutions[54] considered unweighted and undirected network of water and osmolytes separately to understand the clustering of osmolytes through spectral analysis. In this study, we incorporate all of the hydrogen bonds in solution in the graph and compare its network properties against a graph representing only water molecules.

We computed the shortest path, *i.e.*, the path comprising the minimum number of edges (hydrogen bonds), connecting every pair of nodes in the graph. The distributions of path lengths in solutions are shown in Figure 6a,b for the SPC/E and TIP3P water models, respectively. The path length of l_p implies that a pair of nodes is connected by l_p hydrogen bonds in the shortest path; the lower the l_p value, the closer the molecules, corresponding to the nodes in the graph, are connected to each other. All of our analyses are based on three sets of 1 ns of production run. In many cases, we found the associated error bar to be smaller than the symbol size in the graphs. To further ensure the system indeed samples all possible molecular configurations, we have carried out a longer simulation of 10 ns simulation for some of the systems and analyzed the configuration saved every 10 ps. Figure S2 compares the distribution of graph properties computed from 1 and 10 ns simulations. Since the distribution obtained from both simulations matched exactly, we are convinced that our systems are well equilibrated.

A graph of N nodes will exhibit small-world properties if its mean path length is less than a threshold value of $\ln N/\ln k$, where k is the average degree of connectivity between the nodes.[49] The threshold value for a graph of 512 nodes with an average degree of connectivity of 3.6 (average number of hydrogen bonds per water molecule) is 4.8. The mean path

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

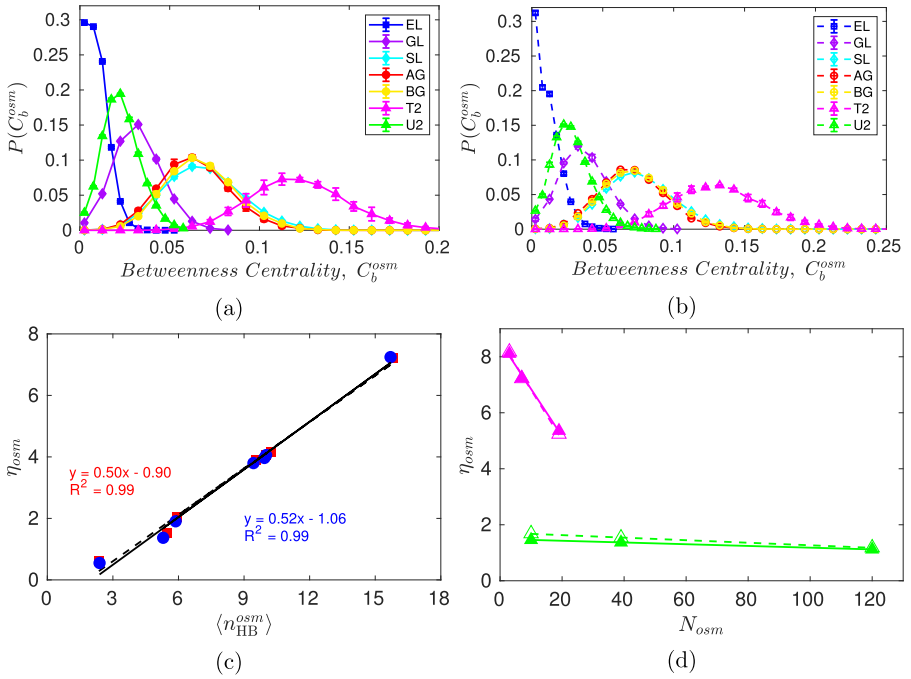


Figure 7. Probability distribution of betweenness centrality (C_b^{osm}) of the nodes representing osmolyte molecules in the solution network for simulations using (a) SPC/E and (b) TIP3P water. Legend: black square, neat water (NW); blue square, ethanol solution (EL); violet diamond, glycerol solution (GL); cyan diamond, sorbitol solution (SL); red circle, α-glucose solution (AG); yellow circle, β-glucose solution (BG); magenta triangle, trehalose solution (T2); green triangle, urea solution (U2); closed symbols and solid lines, solutions with SPC/E water; open symbols and dashed lines, solutions with TIP3P water. (c) Correlation between the effective betweenness (η_{osm}) of osmolyte in 20% aqueous solution and the average number of hydrogen bonds ($\langle n_{HB}^{osm} \rangle$) formed by the osmolyte molecule. Legend: blue circles, systems with SPC/E water; red squares, systems with TIP3P water; solid and dashed lines are the linear fits to the data corresponding to the solutions with SPC/E and TIP3P water, respectively. (d) Effect of osmolyte concentration (the number of osmolyte molecules) on their effective betweenness (η_{osm}). Legend: magenta triangles, trehalose solutions; green triangles, urea solutions; closed symbols and solid lines, solutions with SPC/E water; open symbols and dashed lines, solutions with TIP3P water.

length of all of the solutions investigated is greater than 9, indicating that the hydrogen network of solutions is a percolated network. It is known that at room temperature the hydrogen-bond network of water is a percolated network, and only at supercritical condition it exhibits small-world properties.[49] The path length distribution in urea and glycerol solutions is very similar to that in neat water, and in ethanol solution, the distribution is right-shifted to higher values. The trend is found to be the same in both the water models. In all other solutions, the distribution is left-shifted, implying that the molecules in those solutions are closely connected to each other as the shortest path connecting a pair of nodes includes less number of edges than that in neat water. The relative mean path length in solution, Δl_p, defined in eq 8 with respect to the mean path length in neat water is negative in all solutions, except in ethanol solution, and it exhibits a power law dependence on the average number of hydrogen bonds of the osmolytes (Figure 6c). If we ignore the hydrogen bonding between water and osmolyte molecules, we found that the mean path length between the water nodes is greater than that in neat water (data not shown). Osmolytes are, therefore, well integrated into the solution hydrogen-bond network, increasing the connectivity between the molecules.

To elucidate how well the osmolytes are embedded in the hydrogen-bond network of the solution, we computed the betweenness centrality of the osmolytes. The betweenness centrality of a node provides a measure of how many shortest paths are en routed through it, excluding those paths originating from or terminating at the node; the more the shortest paths through a node, the higher is its betweenness centrality. Figure 7a,b shows the distribution of the betweenness centrality of the osmolytes. We can find the mean of the distribution to follow the order trehalose > α-glucose ~ β-glucose ~ sorbitol > glycerol > urea > ethanol. The order remains the same in both the water models and is similar to the order observed for the number of hydrogen bonds the osmolyte can participate in. Except for the ethanol solution, the average betweenness centrality of the osmolytes is found to be higher than a node in neat water. The effective betweenness centrality of osmolytes, η_{osm} (eq 9), is 7 for trehalose, 4 for α-glucose, β-glucose and sorbitol, 2 for glycerol, and ~1.5 for urea; however, for ethanol, it is only half of that in neat water. A perfect direct correlation between the effective betweenness centrality of the osmolytes and their average number of hydrogen bonds is observed (Figure 7c) in 20 wt % solutions. The more the number of hydrogen bonds an osmolyte could participate in, the higher is the number of shortest paths passing through the corresponding node and the higher is its betweenness centrality. Thus, osmolyte molecules act as hubs in the solution hydrogen-bond network. With the increase in osmolyte concentration, however, the betweenness centrality of the osmolyte decreases (Figure 7d), considerably in trehalose solution and minimally in urea solution, because with the availability of more hubs in the network, the traffic through each hub reduces.

Extent of Osmolyte−Water Hydrogen Bonds Determines the Closeness of the Solution Hydrogen-Bond Network. Another commonly used centrality measure in a

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

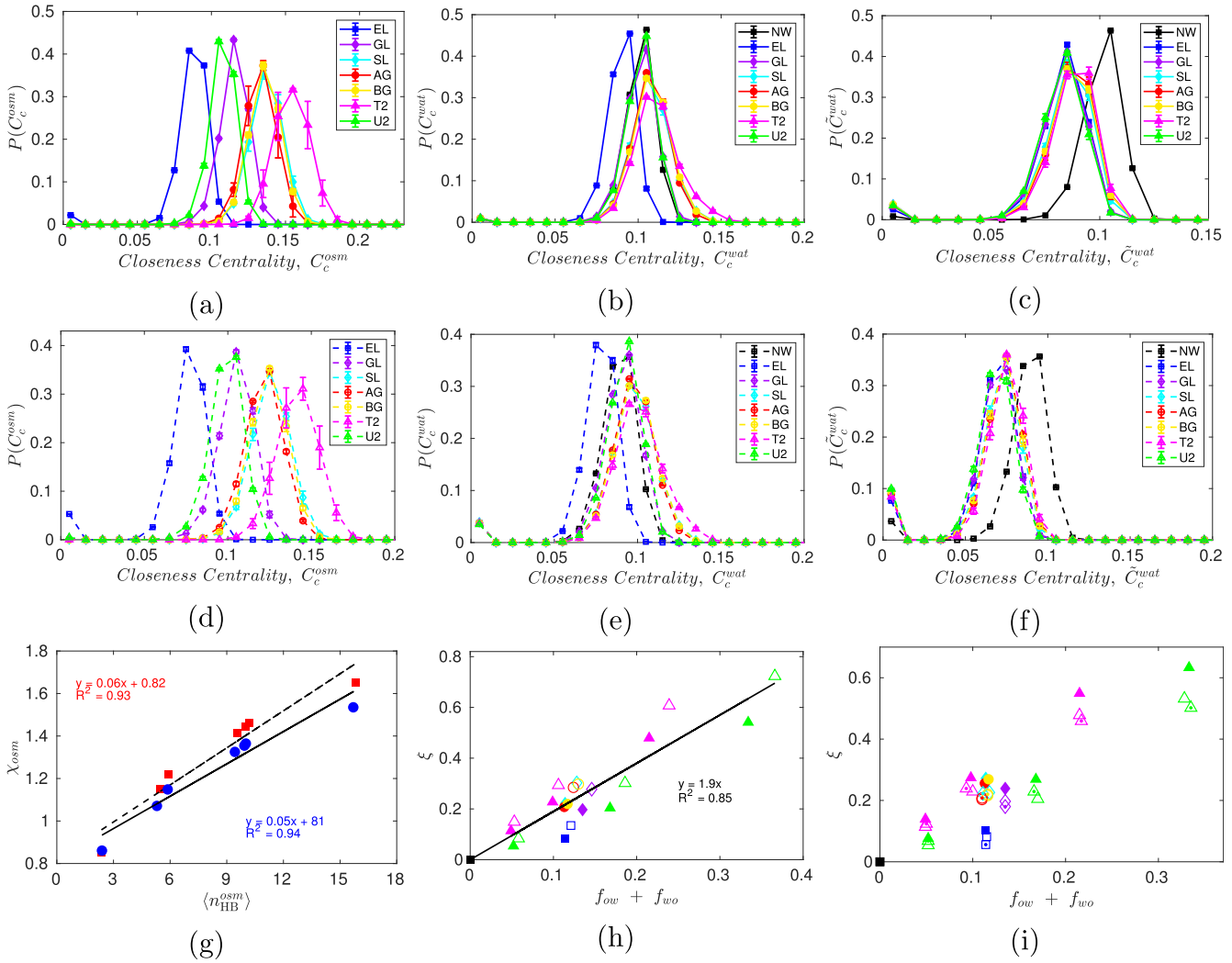


Figure 8. Probability distribution of closeness centrality of the nodes representing (a, d) osmolytes (C_c^{osm}) and (b, e) water (C_c^{wat}) molecules in the solution network for simulations using SPC/E (a−c) and TIP3P (d−f) water models. (c, f) Probability distribution of closeness centrality of water in water-only hydrogen-bond network ($\tilde{C}_c^{wat}$). Legend: black square, neat water (NW); blue square, ethanol solution (EL); violet diamond, glycerol solution (GL); cyan diamond, sorbitol solution (SL); red circle, α-glucose solution (AG); yellow circle, β-glucose solution (BG); magenta triangle, trehalose solution (T2); green triangle, urea solution (U2); closed symbols and solid lines, solutions with SPC/E water; open symbols and dashed lines, solutions with TIP3P water. (g) Correlation between the effective closeness centrality, χ_{osm}, of osmolyte and their average number of hydrogen bonds. Legend: blue circles, systems with SPC/E water; red squares, systems with TIP3P water; solid and dashed lines are the linear fits to the data corresponding to the solutions with SPC/E and TIP3P water, respectively. (h) Correlation between the effective closeness (ξ) of water in solutions and the fractional population of hydrogen bonds between water and osmolytes ($f_{ow} + f_{wo}$). Legend as in (a). (i) Compares the effective closeness of water in SPC/E solutions for directed-weighted graph (closed symbols), directed-unweighted graph (open symbols), and undirected-weighted graph (open symbols with a central dot).

graph theory is the closeness centrality, providing a measure of how closely a node is connected to the other nodes in the network. It is inversely related to the average distance between a node and the rest of the nodes in the graph, *i.e.*, the shorter the average distance, the higher is its closeness centrality. Figure 8a,d shows the probability distribution of closeness centrality of osmolyte in binary solutions with SPC/E and TIP3P water models, respectively. The trend is the same in both the water models. The osmolyte closeness centrality follows the same order as the order for the number of hydrogen bonds formed by an osmolyte: trehalose > α-glucose ∼ β-glucose ∼ sorbitol > glycerol ∼ urea > ethanol. Like the effective betweenness centrality η_{osm}, the effective closeness centrality, χ_{osm}, (eq 9) of osmolytes is also strongly correlated to their average number of hydrogen bonds (Figure 8g).

The distribution of closeness centrality of water in binary solutions is shown in Figure 8b,e for SPC/E and TIP3P water models, respectively. The corresponding distribution for a network graph comprising only water−water hydrogen bonds is shown in Figure 8c,f. We can observe that the latter distributions are left-shifted to lower centrality values compared to the distribution in neat water, indicating that the average distance between the water molecules based on the water-alone network is far apart in solutions than in neat water. When the osmolytes are included as the integral part of the graph network, however, the mean closeness centrality of water, except for in ethanol solution, is found to be higher than the neat water value. This reveals that the osmolytes being the hubs increase the connectivity among water molecules. The effective closeness, ξ, of water in solution (eq 10) measures the

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

reduction in the average distance between water molecules in solution when the osmolytes are considered as nodes in the graph network compared to the graph network comprising nodes corresponding to only water molecules. In binary solutions, the water molecules are closely connected ($\xi > 0$) than in neat water. Figure 8h depicts that the effective closeness of water in solution is reasonably proportional to the fractional population of hydrogen bonds between osmolyte and water molecules, *i.e.*, the sum of fraction of hydrogen bonds from water to osmolyte (f_{wo}) and *vice versa* (f_{ow}). It can be noted that the deviation from linearity is the highest for ethanol, followed by trehalose. Ethanol with only one hydroxyl group and two hydrophobic groups (methyl and methylene groups) needs to orient properly to establish hydrogen bonds with water molecules; this orientational restriction can be ascribed to the low effective closeness in ethanol solutions. In trehalose solutions, on the other hand, the effective closeness slightly deviates positively from the linear trend owing to the larger number of hydroxyl groups of trehalose that can hydrogen bond with water molecules. Even when we consider edge weights (small weights for stronger hydrogen bonds) in the directed graph (directed-weighted graph) or ignore the directionality of the hydrogen bond but include the edge weights (undirected-weighted graph), the qualitative trend is found to be the same (Figure 8i). The graph properties of various osmolyte solutions are mainly dependent on the number of hydrogen bonds formed by the osmolytes, especially with water, clearly suggesting that the osmolytes cannot be classified as a stabilizing osmolyte or a denaturant just based on the hydrogen-bond network of the binary solution. We could notice that the graph properties of glycerol (a protein stabilizer) and urea (a protein denaturant) have been very similar and even equivalent because of their equivalent degree of hydrogen bonding. Furthermore, the effect of temperature on the network properties is found to be only marginal (see Figure S3); hence, it cannot explain the cryoprotective characteristics of the osmolytes.

SUMMARY AND CONCLUSIONS

The water−water rdf in binary solutions reveals a little perturbation of the water structure; although the positions of the peaks and minima remain the same as in neat water rdf, their magnitudes are slightly altered. Further, the number of water−water hydrogen bonds and the tetrahedral order parameter determined from the orientation of the nearest waters molecules have been found to be less in binary solutions. But when we properly account for water−osmolyte hydrogen bonds, the distribution of the number of hydrogen bonds and the average number of hydrogen bonds formed by a water molecule are in good agreement with that in neat water. Similarly, the estimation of the tetrahedral order parameter including the closest osmolyte molecules as the nearest neighbor of the central water molecule exactly matches the neat water value. This clearly demonstrates that osmolytes readily replace some water molecules in the neighborhood of a water molecule, compensating for the lost water−water interactions in solutions, and retaining the local structure intact.

In fact, osmolytes tend to preferentially hydrogen bond with water molecules. Such a preferential interaction cannot be identified when one looks at the distribution of the osmolyte, with the center of mass of osmolytes as its reference center, around a water molecule. Owing to their larger molecular size, it would appear that osmolytes are excluded from water. The thermodynamic properties of water in the investigated binary solutions are not altered by osmolytes as the distribution of the binding energies of water molecules in the solutions is found to be exactly matching with the corresponding distribution in neat water. Interestingly, the intermolecular interaction energy of the osmolytes is determined to be perfectly inversely correlated to the average number of hydrogen bonds formed by an osmolyte molecule. The observation vehemently confirms the significance of hydrogen-bonding interactions in determining the solution thermodynamic properties.

Further, we used the graph theory to understand the influence of the osmolytes on the hydrogen-bond network of the solution as a whole by transforming the molecular configurations to the directed graph with every residue as a node. The solution hydrogen-bond network is determined to be a percolated network as the mean path length of the solution is greater than a threshold value corresponding to a small-world network. The reduction in the mean path length in solution, except in ethanol solution, from that in neat water demonstrates that molecules in solutions are well-connected than in neat water. The estimation of the betweenness and closeness centrality of osmolytes, further, reveals that osmolytes are hubs in the solution hydrogen-bond network. The computed graph properties are strongly dependent on the number of hydrogen bonds formed by the osmolyte molecules.

The network analysis points out that of all of the osmolytes investigated in this study, ethanol is an exception owing to the fewer number of hydrogen bonds formed by ethanol molecules compared to a water molecule. The osmolytes that can form more hydrogen bonds than water enhance the connectivity and effectively decrease the average distance between water molecules in the solution. The effect of temperature on the graph properties, relative to the corresponding property in neat water, is marginal. But the decrease in the translational motion of molecules with a decrease in temperature makes the hydrogen-bond network stiff.

The observation of similar local structural and graph properties in urea and glycerol solution conveys that it is not possible to classify osmolytes as a denaturant or stabilizer based on the hydrogen-bond analysis of binary solutions. This implies that only the direct interaction between protein and osmolyte is responsible for the protein conformational changes. From the observation of the strong correlation between the osmolyte binding energy and their average number of hydrogen bonds and the relevance of hydrogen bonding in the protein structure, we posit that the extent of hydrogen bonding between the protein and osmolyte could be attributed to the alteration in the protein conformational equilibria. It has been recently insisted that hydrophilic interactions are the key driving force for protein folding as the temperature effect on hydrophilic interactions can explain both the cold and heat denaturation of proteins.[86] A network-based analysis of protein solutions would provide a clear picture of the osmolyte influence on protein structure.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00527.

Figures showing the probability distribution of weights characterizing the strength of hydrogen bonds, compar-

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

ison of network properties computed from simulations of different lengths, effect of temperature on the graph properties of osmolyte solutions (PDF)

AUTHOR INFORMATION

Corresponding Author

M. Hamsa Priya − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India;* orcid.org/0000-0003-3049-3531; Phone: +91 044 2257 4132; Email: hamsa@iitm.ac.in

Authors

Smrithi Sundar − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India*

Avilasha A. Sandilya − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00527

Funding

This work was supported by the grant BT/PR8353/BID/7/454/2013 to M.H.P. from the Department of Biotechnology, Ministry of Science and Technology, India.

Notes

The authors declare no competing financial interest.

All simulations were performed with GROMACS 2016.4, which is freely available to download (https://manual.gromacs.org/documentation/2016.4/download.html). The parameters for the molecules considered are readily available in charmm36 parameter files (zipped file charmm36-jul2017.ff.tgz on http://mackerell.umaryland.edu/charmm_ff.shtml). Network analyses were done using the Python module NetworkX 2.3 that can be freely downloaded and installed (https://networkx.org/documentation/networkx-2.3/install.html).

ACKNOWLEDGMENTS

We thank the High Performance Computing Environment (HPCE), IIT Madras, for providing the computational resources.

REFERENCES

(1) Timasheff, S. N. The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes? *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67−97.

(2) Jang, T. H.; Park, S. C.; Yang, J. H.; Kim, J. Y.; Seok, J. H.; Park, U. S.; Choi, C. W.; Lee, S. R.; Han, J. Cryopreservation and its Clinical Applications. *Integr. Med. Res.* **2017**, *6*, 12−18.

(3) Tanford, C. Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. *J. Am. Chem. Soc.* **1962**, *84*, 4240−4247.

(4) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. Living with Water Stress: Evolution of Osmolyte Systems. *Science* **1982**, *217*, 1214−1222.

(5) Robinson, C. H. Cold Adaptation in Arctic and Antarctic Fungi. *New Phytol.* **2001**, *151*, 341−353.

(6) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* **1945**, *13*, 507−532.

(7) Ben-Naim, A. On the Origin of the Stabilization of the Structure of Water by Nonelectrolytes. *J. Phys. Chem. A.* **1965**, *69*, 1922−1927.

(8) Timasheff, S. N. Control of Protein Stability and Reactions by Weakly Interacting Cosolvents: The Simplicity of the Complicated. *Adv. Protein Chem.* **1998**, *51*, 355−432.

(9) Timasheff, S. N.; Xie, G. Preferential Interactions of Urea with Lysozyme and their Linkage to Protein Denaturation. *Biophys. Chem.* **2003**, *105*, 421−448.

(10) Sousa, R. Use of Glycerol, Polyols and Other Protein Structure Stabilizing Agents in Protein Crystallization. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 271−277.

(11) Priya, M. H.; Pratt, L. R.; Paulaitis, M. E. Effect of PEG End-Group Hydrophobicity on Lysozyme Interactions in Solution Characterized by Light Scattering. *Langmuir* **2011**, *27*, 13713−13718.

(12) Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. Molecular Dynamics Simulations of the Unfolding of Barnase in Water and 8 M Aqueous Urea. *Biochemistry* **1997**, *36*, 7313−7329.

(13) Caflisch, A.; Karplus, M. Structural Details of Urea Binding to Barnase: A Molecular Dynamics Analysis. *Structure* **1999**, *7*, 477−488.

(14) Bennion, B. J.; Daggett, V. The Molecular Basis for the Chemical Denaturation of Proteins by Urea. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142−5147.

(15) Tobi, D.; Elber, R.; Thirumalai, D. The Dominant Interaction between Peptide and Urea is Electrostatic in Nature: A Molecular Dynamics Simulation Study. *Biopolymers* **2003**, *68*, 359−369.

(16) Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. Effect of Urea on Peptide Conformation in Water: Molecular Dynamics and Experimental Characterization. *Biophys. J.* **2005**, *89*, 842−857.

(17) 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.* **2008**, *105*, 16928−16933.

(18) Wei, H.; Fan, Y.; Gao, Y. Q. Effects of Urea, Tetramethyl Urea, and Trimethylamine N-Oxide on Aqueous Solution Structure and Solvation of Protein Backbones: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2010**, *114*, 557−568.

(19) Canchi, D. R.; García, A. E. Backbone and Side-Chain Contributions in Protein Denaturation by Urea. *Biophys. J.* **2011**, *100*, 1526−1533.

(20) Stumpe, M. C.; Grubmüller, H. Interaction of Urea with Amino Acids: Implications for Urea-Induced Protein Denaturation. *J. Am. Chem. Soc.* **2007**, *129*, 16126−16131.

(21) Green, J. L.; Angell, C. A. Phase Relations and Vitrification in Saccharide-Water Solutions and the Trehalose Anomaly. *J. Phys. Chem. B.* **1989**, *93*, 2880−2882.

(22) Lins, R. D.; Pereira, C. S.; Hünenberger, P. H. Trehalose−Protein Interaction in Aqueous Solution. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 177−186.

(23) Carpenter, J. F.; Crowe, J. H. An Infrared Spectroscopic Study of the Interactions of Carbohydrates with Dried Proteins. *Biochemistry* **1989**, *28*, 3916−3922.

(24) Fedorov, M. V.; Goodman, J. M.; Nerukh, D.; Schumm, S. Self-assembly of Trehalose Molecules on a Lysozyme Surface: The Broken Glass Hypothesis. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2294−2299.

(25) Lerbret, A.; Affouard, F.; Hédoux, A.; Krenzlin, S.; Siepmann, J.; Bellissent-Funel, M.-C.; Descamps, M. How Strongly Does Trehalose Interact with Lysozyme in the Solid State? Insights from Molecular Dynamics Simulation and Inelastic Neutron Scattering. *J. Phys. Chem. B* **2012**, *116*, 11103−11116.

(26) Belton, P. S.; Gil, A. M. IR and Raman Spectroscopic Studies of the Interaction of Trehalose with Hen Egg White Lysozyme. *Biopolymers* **1994**, *34*, 957−961.

(27) Corradini, D.; Strekalova, E. G.; Stanley, H. E.; Gallo, P. Microscopic Mechanism of Protein Cryopreservation in an Aqueous Solution with Trehalose. *Sci. Rep.* **2013**, *3*, No. 1218.

(28) Cottone, G.; Ciccotti, G.; Cordone, L. Protein−Trehalose−Water Structures in Trehalose Coated Carboxy-Myoglobin. *J. Chem. Phys.* **2002**, *117*, 9862−9866.

(29) Carr, J. K.; Buchanan, L. E.; Schmidt, J. R.; Zanni, M. T.; Skinner, J. L. Structure and Dynamics of Urea/Water Mixtures Investigated by Vibrational Spectroscopy and Molecular Dynamics Simulation. *J. Phys. Chem. B* **2013**, *117*, 13291−13300.
(30) Rezus, Y. L. A.; Bakker, H. J. Effect of Urea on the Structural Dynamics of Water. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18417−18420.
(31) Verma, P. K.; Lee, H.; Park, J. Y.; Lim, J. H.; Maj, M.; Choi, J. H.; Kwak, K. W.; Cho, M. Modulation of the Hydrogen Bonding Structure of Water by Renal Osmolytes. *J. Phys. Chem. Lett.* **2015**, *6*, 2773−2779.
(32) Dashnau, J. L.; Nucci, N. V.; Sharp, K. A.; Vanderkooi, J. M. Hydrogen Bonding and the Cryoprotective Properties of Glycerol/Water Mixtures. *J. Phys. Chem. B* **2006**, *110*, 13670−13677.
(33) Branca, C.; Magazù, S.; Maisano, G.; Migliardo, P.; Tettamanti, E. On the Bioprotective Effectiveness of Trehalose: Ultrasonic Technique, Raman Scattering and NMR Investigations. *J. Mol. Struct.* **1999**, *480*−*481*, 133−140.
(34) Stehle, S.; Braeuer, A. S. Hydrogen Bond Networks in Binary Mixtures of Water and Organic Solvents. *J. Phys. Chem. B* **2019**, *123*, 4425−4433.
(35) Laurent, H.; Baker, D. L.; Soper, A. K.; Ries, M. E.; Dougan, L. Solute Specific Perturbations to Water Structure and Dynamics in Tertiary Aqueous Solution. *J. Phys. Chem. B* **2020**, *124*, 10983−10993.
(36) Gallina, M. E.; Comez, L.; Morresi, A.; Paolantoni, M.; Perticaroli, S.; Sassi, P.; Fioretto, D. Rotational Dynamics of Trehalose in Aqueous Solutions Studied by Depolarized Light Scattering. *J. Chem. Phys.* **2010**, *132*, No. 214508.
(37) Magazù, S.; Migliardo, F.; Telling, M. T. F. α,α-Trehalose-Water Solutions. VIII. Study of the Diffusive Dynamics of Water by High-Resolution Quasi Elastic Neutron Scattering. *J. Phys. Chem. B* **2006**, *110*, 1020−1025.
(38) Soper, A. K.; Castner, E. W.; Luzar, A. Impact of Urea on Water Structure: A Clue to its Properties as a Denaturant? *Biophys. Chem.* **2003**, *105*, 649−666.
(39) Towey, J. J.; Soper, A. K.; Dougan, L. Preference for Isolated Water Molecules in a Concentrated Glycerol-Water Mixture. *J. Phys. Chem. B* **2011**, *115*, 7799−7807.
(40) Lerbret, A.; Bordat, P.; Affouard, F.; Descamps, M.; Migliardo, F. How Homogeneous Are the Trehalose, Maltose, and Sucrose Water Solutions? An Insight from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2005**, *109*, 11046−11057.
(41) Kokubo, H.; Pettitt, M. B. Preferential Solvation in Urea Solutions at Different Concentrations: Properties from Simulation Studies. *J. Phys. Chem. B* **2007**, *111*, 5233−5242.
(42) Stumpe, M. C.; Grubmüller, H. Aqueous Urea Solutions: Structure, Energetics, and Urea Aggregation. *J. Phys. Chem. B* **2007**, *111*, 6220−6228.
(43) Bandyopadhyay, D.; Mohan, S.; Ghosh, S. K.; Choudhury, N. Molecular Dynamics Simulation of Aqueous Urea Solution: Is Urea a Structure Breaker? *J. Phys. Chem. B* **2014**, *118*, 11757−11768.
(44) Idrissi, A.; Gerard, M.; Damay, P.; Kiselev, M.; Puhovsky, Y.; Cinar, E.; Lagant, P.; Vergoten, G. The Effect of Urea on the Structure of Water: A Molecular Dynamics Simulation. *J. Phys. Chem. B* **2010**, *114*, 4731−4738.
(45) Lovrinčević, B.; Požar, M.; Balić, M. Dynamics of Urea-Water Mixtures Studied by Molecular Dynamics Simulation. *J. Mol. Liq.* **2020**, *300*, No. 112268.
(46) Zeman, J.; Holm, C.; Smiatek, J. The Effect of Small Organic Cosolutes on Water Structure and Dynamics. *J. Chem. Eng. Data* **2020**, *65*, 1197−1210.
(47) Albert, R.; Barabási, A.-L. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **2002**, *74*, 47−97.
(48) García-Domenech, R.; Gálvez, J.; de Julián-Ortiz, J. V.; Pogliani, L. Some New Trends in Chemical Graph Theory. *Chem. Rev.* **2008**, *108*, 1127−1169.
(49) Marques Leite dos Santos, V.; Brady Moreira, F. G.; Longo, R. L. Topology of the Hydrogen Bond Networks in Liquid Water at Room and Supercritical Conditions: A Small-World Structure. *Chem. Phys. Lett.* **2004**, *390*, 157−161.
(50) Choi, J.-H.; Cho, M. Ion Aggregation in High Salt Solutions. II. Spectral Graph Analysis of Water Hydrogen-Bonding Network and Ion Aggregate Structures. *J. Chem. Phys.* **2014**, *141*, No. 154502.
(51) Choi, J.-H.; Lee, H.; Choi, H. R.; Cho, M. Graph Theory and Ion and Molecular Aggregation in Aqueous Solutions. *Annu. Rev. Phys. Chem.* **2018**, *69*, 125−149.
(52) Bakó, I.; Megyes, T.; Bálint, S.; Grósz, T.; Chihaia, V. Water-Methanol Mixtures: Topology of Hydrogen Bonded Network. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5004−5011.
(53) da Silva, J. A. B.; Moreira, F. G. B.; dos Santos, V. M. L.; Longo, R. L. Hydrogen Bond Networks in Water and Methanol with Varying Interaction Strengths. *Phys. Chem. Chem. Phys.* **2011**, *13*, 593−603.
(54) Lee, H.; Choi, J.-H.; Verma, P. K.; Cho, M. Spectral Graph Analyses of Water Hydrogen-Bonding Network and Osmolyte Aggregate Structures in Osmolyte-Water Solutions. *J. Phys. Chem. B* **2015**, *119*, 14402−14412.
(55) Woods Group. *GLYCAM Web*; Complex Carbohydrate Research Center, University of Georgia: Athens, GA, 2019. http://glycam.org (accessed Jan 03, 2019).
(56) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(57) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(58) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem. C.* **1987**, *91*, 6269−6271.
(59) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(61) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.
(62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(63) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(64) Hoover, W. G. Canonical dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(65) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(66) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(67) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(68) Hummer, G.; Pratt, L. R.; Garcia, A. E. Free Energy of Ionic Hydration. *J. Phys. Chem. D.* **1996**, *100*, 1206−1215.
(69) Errington, J.; Debenedetti, P. Relationship between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, *409*, 318−321.
(70) Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press, 1994.
(71) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.* **1951**, *19*, 774−777.
(72) Priya, M. H.; Ashbaugh, H. S.; Paulaitis, M. E. Cosolvent Preferential Molecular Interactions in Aqueous Solutions. *J. Phys. Chem. B* **2011**, *115*, 13633−13642.

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944

(73) Priya, M. H.; Merchant, S.; Asthagiri, D.; Paulaitis, M. E. Quasi-Chemical Theory of Cosolvent Hydrophobic Preferential Interactions. *J. Phys. Chem. B* **2012**, *116*, 6506−6513.
(74) Paolantoni, M.; Comez, L.; Gallina, M. E.; Sassi, P.; Scarponi, F.; Fioretto, D.; Morresi, A. Light Scattering Spectra of Water in Trehalose Aqueous Solutions: Evidence for Two Different Solvent Relaxation Processes. *J. Phys. Chem. B* **2009**, *113*, 7874−7878.
(75) Charkhesht, A.; Lou, D.; Sindle, B.; Wen, C.; Cheng, S.; Vinh, N. Q. Insights into Hydration Dynamics and Cooperative Interactions in Glycerol-Water Mixtures by Terahertz Dielectric Spectroscopy. *J. Phys. Chem. B* **2019**, *123*, 8791−8799.
(76) Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. Carbohydrate Intramolecular Hydrogen Bonding Cooperativity and Its Effect on Water Structure. *J. Phys. Chem. B* **2005**, *109*, 24152−24159.
(77) Bandyopadhyay, D.; Mohan, S.; Ghosh, S. K.; Choudhury, N. Correlation of Structural Order, Anomalous Density, and Hydrogen Bonding Network of Liquid Water. *J. Phys. Chem. B* **2013**, *117*, 8831−8843.
(78) Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem. E.* **1982**, *86*, 869−872.
(79) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(80) BenNaim, A.; Marcus, Y. Solubility and Thermodynamics of Solution of Xenon in Liquid nalkanes. *J. Chem. Phys.* **1984**, *80*, 4438−4440.
(81) Paliwal, A.; Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S.; Paulaitis, M. E. An Analysis of Molecular Packing and Chemical Association in Liquid Water using Quasichemical Theory. *J. Chem. Phys.* **2006**, *124*, No. 224502.
(82) Kristóf, T.; Rutkai, G. Chemical Potential Calculations by Thermodynamic Integration with Separation Shifting in Adaptive Sampling Monte Carlo Simulations. *Chem. Phys. Lett.* **2007**, *445*, 74−78.
(83) Heidari, M.; Kremer, K.; Cortes-Huerto, R.; Potestio, R. Spatially Resolved Thermodynamic Integration: An Efficient Method to Compute Chemical Potentials of Dense Fluids. *J. Chem. Theory Comput.* **2018**, *14*, 3409−3417.
(84) Sandilya, A. A.; Natarajan, U.; Priya, M. H. Molecular View into the Cyclodextrin Cavity: Structure and Hydration. *ACS Omega* **2020**, *5*, 25655−25667.
(85) Wendler, K.; Thar, J.; Zahn, S.; Kirchner, B. Estimating the Hydrogen Bond Energy. *J. Phys. Chem. A* **2010**, *114*, 9529−9536.
(86) Ben-Naim, A. *The Protein Folding Problem and Its Solution*; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2013.

https://doi.org/10.1021/acs.jcim.1c00527
J. Chem. Inf. Model. 2021, 61, 3927−3944