



**HAL**  
open science

## Protein preferential solvation in water:glycerol mixtures

Nicolas Chéron, Margaux Naepels, Eva Pluhařová, Damien Laage

► **To cite this version:**

Nicolas Chéron, Margaux Naepels, Eva Pluhařová, Damien Laage. Protein preferential solvation in water:glycerol mixtures. *Journal of Physical Chemistry B*, 2020, 124 (8), pp.1424-1437. 10.1021/acs.jpbc.9b11190 . hal-02409375v2

**HAL Id: hal-02409375**

**<https://hal.science/hal-02409375v2>**

Submitted on 27 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Protein Preferential Solvation in Water:Glycerol Mixtures

Nicolas Chéron,<sup>†</sup> Margaux Naepels,<sup>†</sup> Eva Pluhařová,<sup>†,‡</sup> and Damien Laage<sup>\*,†</sup>

<sup>†</sup>*PASTEUR, Département de chimie, École Normale Supérieure, PSL University, Sorbonne  
Université, CNRS, 75005 Paris, France*

<sup>‡</sup>*Present address: J. Heyrovský Institute of Physical Chemistry, Czech Academy of  
Sciences, Dolejškova 2155/3, 182 23 Prague, Czech Republic*

E-mail: [damien.laage@ens.fr](mailto:damien.laage@ens.fr)

Phone: +33 (0)144322418

## Abstract

For proteins in solvent mixtures, the relative abundances of each solvent in their solvation shell have a critical impact on their properties. Preferential solvation of a series of proteins in water-glycerol mixtures is studied here over a broad range of solvent compositions via classical molecular dynamics simulations. Our simulation results reveal that the differences between shell and bulk compositions exhibit dramatic changes with solvent composition, temperature and protein nature. In contrast with the simple and widely used picture where glycerol is completely excluded from the protein interface, we show that for aqueous solutions with less than 50% glycerol in volume, protein solvation shells have approximately the same composition as the bulk solvent and proteins are in direct contact with glycerol. We further demonstrate that at high glycerol concentration, glycerol depletion from the solvation shell is due to an entropic factor arising from the reduced accessibility of bulky glycerol molecules in protein cavities. The resulting molecular picture is important to understand protein activity and cryopreservation in mixed aqueous solvents.

## Introduction

Adding cosolvents to aqueous protein solutions can dramatically change the structural stability<sup>1,2</sup> and biochemical activity<sup>3</sup> of proteins. While some cosolvents denature proteins, others help to preserve their structure. Typical denaturants include for example urea, while protectants include e.g. polyols like glycerol and sugars like trehalose, which are widely used for lyoprotection and cryopreservation.<sup>1,4,5</sup> A broad range of water-cosolvent binary mixtures have been studied,<sup>6</sup> and their structure, dynamics and phase diagram have been characterized (see e.g. refs 7–14). It was for example shown that concentrated water:glycerol mixtures form a homogeneous glass at low temperatures, thus avoiding cell-damaging crystallization.<sup>13</sup> However, the study of ternary mixtures including water, cosolvent and protein is more challenging. Despite recent advances (see e.g.<sup>15</sup>), their properties thus remain com-

parably less well understood and a comprehensive molecular picture of protein (and more generally, polymer) (de-)stabilization by cosolvents has so far remained elusive.

In order to elucidate how proteins are affected by cosolvents, a central question is to determine the cosolvent abundance at the protein exposed interface. Depending on the competition between protein-cosolvent and protein-water interactions, the protein solvation layer may be either enriched or depleted in cosolvent with respect to the bulk solvent. A number of experimental techniques – including e.g. densimetry,<sup>16</sup> osmometry,<sup>17</sup> calorimetry,<sup>18,19</sup> neutron scattering<sup>20,21</sup> and dielectric relaxation<sup>22</sup> – have thus been employed for a range of cosolvents in order to determine whether proteins are preferentially solvated by water or by the cosolvent. However, connecting these measurements to a microscopic picture of the solvation layer is not straightforward, and often relies on approximate models.<sup>23</sup> In addition, the key protein and cosolvent features which govern preferential solvation, including e.g. cosolvent size and protein-cosolvent interaction,<sup>24</sup> remain to be clearly identified.

In this work we focus on protein preferential solvation in water/glycerol mixtures, due to their great importance in cryopreservation.<sup>4,5</sup> The broad range of prior experimental and theoretical studies in these mixtures have concluded that proteins are preferentially solvated by water, and suggested that glycerol is disfavored or even excluded from the protein vicinity.<sup>20,21,25–27</sup> While steric exclusion arguments due to the large glycerol molecular size have been commonly used to justify this exclusion, experiments suggested that an additional, chemically specific interaction is also involved.<sup>16</sup> Both factors should dramatically vary over the protein surface and the solvation layer composition is expected to exhibit a pronounced heterogeneity reflecting the great variety of shapes and chemical natures of protein exposed sites.<sup>28,29</sup>

In order to gain a molecular picture of the protein solvation layer in aqueous solvent mixtures, we use molecular dynamics (MD) simulations, which were shown to be a powerful and incisive tool to study preferential solvation. Significant advances in the understanding of protein/water/glycerol mixtures were made in a series of MD simulations.<sup>28,30–32</sup> These stud-



ies showed that distinct protein sites have different affinities for each solvent, and suggested a connection between preferential hydration and the orientation of glycerol molecules at the protein interface. However, a number of key questions remain unresolved, including e.g. how far the protein affects the solvent composition, how much preferential solvation varies among typical globular proteins, how much it changes with the bulk solvent composition and with temperature. These questions are critical to obtain a global molecular picture of preferential hydration, and we address them in the present work.

The outline of the remainder of this paper is the following. We first introduce the preferential hydration coefficient definition, describe its connection to the microscopic solution structure accessible via molecular dynamics simulations and suggest several extensions to overcome some of its limitations. We then describe our simulation methodology and discuss the critical points of convergence and force-field sensitivity for these simulations. The next sections examine the popular solvent-cosolvent exchange model, widely employed to interpret experiments, and compare our simulation results with available experimental data to validate our simulation methodology. We then successively study how preferential hydration is affected by the glycerol concentration, by the protein nature and by temperature, in order to establish a molecular picture of the key features determining the protein solvation shell composition in solvent mixtures. We finally offer some concluding remarks.

## **Preferential interaction coefficient**

In this section, we provide a brief introduction to the preferential interaction coefficient, its thermodynamic definition, its connection with microscopic protein solvation structure and its calculation from MD simulations. We then discuss some of its limitations, suggest several extensions, and illustrate their application to a typical protein/water/glycerol system.

## Thermodynamic definition

The preferential interaction coefficient provides a measure of a protein’s relative affinities with water and a cosolvent. Its most widely used thermodynamic definition probes the change in protein chemical potential caused by adding cosolvent to an aqueous solution. For a protein (P) in a mixture of water (W) and co-solvent (here glycerol, G), the preferential interaction coefficient is defined as<sup>2,33–37</sup>

$$\Gamma_{GP} = - \left. \frac{\partial \mu_P}{\partial \mu_G} \right|_{T, m_W, m_G} = \left. \frac{\partial m_G}{\partial m_P} \right|_{T, \mu_W, \mu_G}, \quad (1)$$

where  $m_i$  and  $\mu_i$  are respectively the molality and chemical potential of component  $i$ . While other definitions involving e.g. molar concentrations and molar fractions have been used, the present molality-based definition offers more intuitive interpretations<sup>37,38</sup> and more direct connections to experiments.<sup>33,36</sup>

This thermodynamic definition can be connected to the local solvent structure surrounding the protein solute in the limit of an infinitely dilute protein,<sup>37–41</sup>

$$\Gamma_{GP}^\circ = \lim_{m_P \rightarrow 0} \left. \frac{\partial m_G}{\partial m_P} \right|_{T, \mu_W, \mu_G} = \rho_G (G_{PG} - G_{PW}) \quad (2)$$

where  $\rho_G$  is the glycerol number density and  $G_{ij}$  are the Kirkwood-Buff integrals<sup>42</sup>

$$G_{ij} = \int_0^\infty (g_{ij}(r) - 1) 4\pi r^2 dr \quad (3)$$

with  $g_{ij}(r)$  the radial distribution function between molecules  $i$  and  $j$ .

## Microscopic picture

An intuitive interpretation of  $\Gamma_{GP}$  was suggested with a two-domain model,<sup>43</sup> where the solvent is separated in two regions: i) the local region surrounding the protein solute and whose composition is affected by the solute, and ii) the more remote region with the bulk

composition.  $\Gamma_{GP}$  quantifies the excess number of cosolvent molecules within the protein solvation shell with respect to the expected number if the shell had the same composition as the bulk.  $\Gamma_{GP}$  can be equivalently expressed as<sup>32,34</sup>

$$\Gamma_{GP} = \langle n_G^{\text{shell}} \rangle - \frac{\langle n_G^{\text{bulk}} \rangle}{\langle n_W^{\text{bulk}} \rangle} * \langle n_W^{\text{shell}} \rangle \quad (4)$$

where  $n_{G,W}^{\text{shell,bulk}}$  are the numbers of glycerol (G) and water (W) molecules in the protein local solvation domain and in the bulk, and  $\langle \dots \rangle$  are ensemble averages.  $\Gamma_{GP}$  is positive (resp. negative) when the protein solvation domain is enriched (resp. depleted) in glycerol compared to the bulk. Typical denaturants usually exhibit positive preferential interaction coefficient values, while protectants lead to negative values, corresponding to a preferential hydration.<sup>2</sup>

The  $\Gamma_{GP}$  formulation in Eq 4 is particularly amenable to MD studies, since it involves terms which can be directly determined from simulations. A large number of MD studies have thus used an approach based on a variant of this formulation<sup>28,31,32,44-49</sup> to investigate preferential solvation.

## Assessment of typical approximations

The practical implementation of  $\Gamma_{GP}$  calculations in molecular dynamics simulations usually involves a number of approximations. We now systematically review the validity of the three most critical ones.

First, a key point lies in the shell and bulk domain definitions. In the thermodynamic formulation Eqs 2-3, the Kirkwood-Buff integrals run over solvent locations up to infinitely large distances from the protein. These integrals converge when the solvent reaches the average bulk composition, and the shell boundary  $r^{\text{shell}}$  in the two-domain model is thus the distance where the protein's influence on solvent composition disappears. We stress that because the volume element in Eq 3 increases with increasing distance, the solvation

shell thickness may be larger than simply estimated by determining where the distribution function  $g_{ij}(r)$  seems to reach a plateau. Prior simulation studies proposed to use a shell thickness corresponding to the end of the protein second solvation layer, and a range of values between 5 and 8 Å have been used.<sup>28,30–32,44–49</sup> However, as we will show below, these distances are often too short for the solvent to reach the bulk composition. To compare our simulated  $\Gamma_{GP}$  values with experimental measurements, we will employ the largest possible radius given our simulation box size; we will use a 14Å distance, where the solvent has typically reached the bulk composition, and which leads to similar numbers of shell and bulk solvent molecules and thus offers the best precision on the composition of both regions.

A second related point pertains to the spatial resolution when studying the solvent composition at increasing distances from the protein surface. Typical MD studies consider the solvent molecules’ centers of mass when determining the composition within a given radius from the protein. It follows that the resolution cannot exceed the size of the largest solvent molecules and misleading results can be obtained for mixtures with large cosolvent molecules. For example, as shown in Fig 1, for large glycerol cosolvent molecules in direct contact with the protein surface, their centers of mass cannot reach the short distances found for the smaller water molecules, and this definition incorrectly overestimates the amount of water within the first layer. To circumvent this difficulty, in the following we count the numbers of solvent heavy atoms (see SI); adequate normalization by the number of heavy atoms in each molecule is used and for distances larger than the molecular sizes, one recovers the same limit as when centers of mass are considered (a similar approach was recently reported for the calculation of Kirkwood-Buff integrals<sup>50</sup>).

Finally, we note that prior simulations approximated the  $\Gamma_{GP}$  definition in Eq 4 by

$$\Gamma_{GP} \simeq \left\langle n_G^{\text{shell}} - \frac{n_G^{\text{bulk}}}{n_W^{\text{bulk}}} * n_W^{\text{shell}} \right\rangle, \quad (5)$$

where one considers the average of a product and ratio of instantaneous numbers of solvent

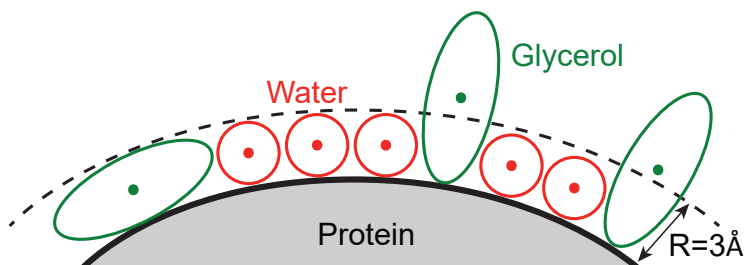


Figure 1: Schematic representation of a protein’s first solvation shell, with water and glycerol solvent molecules. When considering the solvent molecules’ centers of mass (represented by dots) within a typical 3 Å first-layer radius, several glycerol molecules would be ignored.

particles in the two domains, instead of the product and ratio of average numbers. The latter expression is clearly not correct at very low water concentrations where the instantaneous number of water molecules in the local protein solvation domain can be zero, thus leading to an unphysical divergence that does not exist in the original Eq 4 definition. In addition, it implicitly assumes that the numbers of glycerol and water molecules in the two domains are not correlated. An assessment of this approximation is thus necessary. We report in Fig S1 (see SI) a systematic comparison for a typical system. Although this approximate formulation should be avoided in very concentrated cosolvent solutions, our results show that it remains acceptable for intermediate concentrations. However, in the following, we will use the exact definition Eq 4.

## Molar fraction difference

Although  $\Gamma_{GP}$  has several attractive features, including being experimentally accessible and thermodynamically well defined, connecting its value to intuitive molecular aspects remains difficult. One of the reasons is that  $\Gamma_{GP}$  is an extensive quantity, which implies that cosolvent affinities of proteins with different sizes cannot be easily compared: for example, similar  $\Gamma_{GP}$  values can be obtained for small, strongly preferentially hydrated proteins and for large,

weakly hydrated ones. Equation 4 can be rearranged as (see SI-1)

$$\Gamma_{GP} = \frac{\langle n_G^{\text{shell}} + n_W^{\text{shell}} \rangle * (x_G^{\text{shell}} - x_G^{\text{bulk}})}{1 - x_G^{\text{bulk}}}, \quad (6)$$

where  $x_G^{\text{shell,bulk}} = \langle n_G^{\text{shell,bulk}} \rangle / \langle n_G^{\text{shell,bulk}} + n_W^{\text{shell,bulk}} \rangle$  are the shell and bulk glycerol molar fractions (similarly, in the following,  $x_G$  will designate the overall glycerol fraction in the solution). Equation 6 shows that  $\Gamma_{GP}$  increases proportionally to the overall number of solvent molecules in the shell  $\langle n_G^{\text{shell}} + n_W^{\text{shell}} \rangle$ . Prior quantitative analyses of preferential interaction coefficients have thus for example considered a normalization by the protein solvent accessible surface area.<sup>17</sup>

In order to obtain a more intuitive description of the solvent shell composition, in the following we will complement the  $\Gamma_{GP}$  calculations with the determination of the difference in glycerol molar fractions in the shell and in the bulk,

$$\Delta x_G^{\text{shell}} = x_G^{\text{shell}} - x_G^{\text{bulk}}. \quad (7)$$

While both  $\Gamma_{GP}$  and  $\Delta x_G^{\text{shell}}$  report on the solvent shell composition, a key advantage of  $\Delta x_G^{\text{shell}}$  is that it allows comparisons of the shell and bulk compositions independently of the solvation shell size. As for  $\Gamma_{GP}$ ,  $\Delta x_G^{\text{shell}}$  depends on the chosen shell thickness  $r^{\text{shell}}$ . In what follows we will adopt two different  $r^{\text{shell}}$  values for these two quantities. For the experimentally-accessible thermodynamic  $\Gamma_{GP}$  which includes long-range effects induced by the protein on the solvent composition we will use  $r^{\text{shell}}=14\text{\AA}$ , as mentioned above. In contrast, to provide a molecular characterization of the first solvent shell in contact with the protein, we will consider  $\Delta x_G^{\text{shell}}$  with  $r^{\text{shell}}=6\text{\AA}$ , corresponding approximately to the first layer of glycerol molecules (a similar 5.9 Å value was obtained by analyzing neutron scattering results in ref<sup>21</sup>).

## Paradigm case of lysozyme in water:glycerol

We now illustrate these different considerations for a typical lysozyme protein in a water:glycerol  $x_G = 0.20$  mixture and show how the preferential interaction coefficient is determined by the local solvent structure and composition (simulation details will be provided in the next section). Figure 2-(a) shows the local excess in glycerol molar fraction  $\Delta x_G^{local}(r)$  in thin successive layers at increasing distances  $r$  from the protein surface. As one moves away from the protein, one first sees an excess of water molecules at  $\approx 2 \text{ \AA}$  due to their small molecular size, followed by a local excess in glycerol at approximately  $4 \text{ \AA}$  where glycerol molecules in contact with the protein are more likely to reside; finally, a moderate glycerol depletion is observed at  $6 \text{ \AA}$  before the local solvent composition reaches the bulk value at  $11 \text{ \AA}$ .

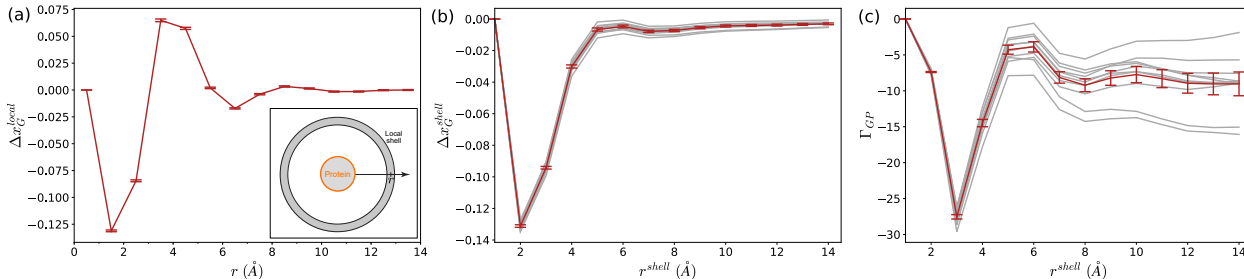


Figure 2: (a) Difference in glycerol molar fractions between  $1 \text{ \AA}$ -thick layers at distance  $r$  from the protein surface and the bulk for a lysozyme protein in a water:glycerol mixture at  $x_G^{bulk} = 0.20$ . (b) Difference of molar fractions and (c) preferential interaction coefficient between lysozyme and glycerol for increasing solvation shell thickness  $r^{shell}$  in a  $x_G = 0.20$  aqueous solution. Results from 10 independent 100 ns replicas are shown in gray and the average value is shown in red. Error bars represent the Student 95% confidence interval, calculated from averages over 2 ns independent blocks from all trajectories.

The protein solvation shell average composition echoes these changes in the local solvent composition.  $\Delta x_G^{shell}(r^{shell})$  is the excess glycerol molar fraction in a protein solvation shell of thickness  $r^{shell}$  from the protein. Its changes with increasing  $r^{shell}$  are shown in Fig 2-(b) and reflect the layered structure in Fig 2-(a). The apparent large depletion in glycerol at very short  $\simeq 2 \text{ \AA}$  distances is strongly enhanced by the different molecular sizes of water and

glycerol. While our present approach based on solvent heavy atoms leads to a moderately deep first minimum of  $\Delta x_G^{\text{shell}}(2\text{\AA}) = -0.13$ , the usual definition involving solvent molecules' centers of mass leads to an enhanced first minimum  $\Delta x_G^{\text{shell}}(2\text{\AA}) = -x_G^{\text{bulk}} = -0.20$ . This is caused by the vanishing fraction of glycerol centers of mass at such short distances. However, we found that on average 18.9 glycerol heavy atoms and 41.8 water heavy atoms are within  $2\text{\AA}$  of the protein surface, i.e. 31% of the shell heavy atoms belong to glycerol molecules (to be contrasted with the overall  $x_G = 0.20$  water:glycerol composition where 60% of the heavy atoms belong to glycerol molecules). Thus, this shows that our heavy-atom based approach provides a more faithful picture of the local solvent composition, and that simplified representations where the protein surface is completely depleted in glycerol are seriously misleading.

We now turn to the preferential interaction coefficient  $\Gamma_{GP}$  reported in Fig 2-(c). Its changes with the shell thickness  $r^{\text{shell}}$  follow those in the shell composition  $\Delta x_G^{\text{shell}}$ .  $\Gamma_{GP}$  displays a local minimum at  $3\text{\AA}$  and a local maximum at  $6\text{\AA}$ , in agreement with prior simulations of proteins in water:glycerol mixtures.<sup>30,32</sup> The shift in the locations of these extrema with respect to those discussed for  $\Delta x_G^{\text{shell}}$  in Fig 2-(b) arises from the growing solvation shell size with increasing  $r^{\text{shell}}$ . Both  $\Delta x_G^{\text{shell}}$  and  $\Gamma_{GP}$  are found to converge for  $r^{\text{shell}} > 12\text{\AA}$ , which justifies our choice to use  $r^{\text{shell}} = 14\text{\AA}$  for comparisons of  $\Gamma_{GP}$  with experimental measurements. We already note that the large dispersion between  $\Gamma_{GP}$  results obtained from different independent trajectories shown in Fig 2-(c) stresses the need for careful convergence studies, that will be described further.

## Simulation methodology

### Systems

We performed MD simulations of dilute protein solutions in a series of water:glycerol mixtures with different proportions. The list of simulated systems and their compositions are given in



Table 1. Proteins were described with the Charmm27 force-field<sup>51</sup> and water with the TIP3P potential.<sup>52</sup> Glycerol was described with a force-field successfully used in prior simulation studies<sup>28,30-32</sup> of proteins in water:glycerol mixtures, that was derived from a carbohydrate potential<sup>53</sup> and made consistent with Charmm27.<sup>54</sup> For DHFR, the folate substrate and the NADPH cofactor were described with parameters from the literature.<sup>55,56</sup> Additional simulations were performed to determine the protein and glycerol force-field impact on the results, and the Amber99SB<sup>57</sup> protein force-field and the GAFF<sup>58</sup> glycerol potential were used. In contrast with prior MD studies,<sup>28,31,32</sup> no constraints were applied to the proteins during simulations.

**Table 1: List of simulated systems with overall glycerol molar fraction  $x_G$ , total numbers of water and glycerol molecules  $N_{W/G}$ , water:glycerol volume fraction (estimated from molar fractions and experimental molar volumes<sup>59</sup>), overall glycerol concentration in solution, and protein concentration.**

Protein	$x_G$	$N_W$	$N_G$	water:glycerol (v/v)	[Gly] (mol/L)	[protein] (g/L)
Lysozyme	0.01	7828	80	96:4	0.5	90
Lysozyme	0.05	6425	340	82:18	2.4	93
Lysozyme	0.10	5231	590	69:31	4.3	94
Lysozyme	0.20	3722	930	49:51	6.9	94
Lysozyme	0.30	2738	1180	36:64	8.8	93
Lysozyme	0.40	2080	1380	27:73	10.1	90
Lysozyme	0.50	1575	1570	20:80	11.1	87
Lysozyme	0.60	1172	1770	14:86	11.8	82
Lysozyme	0.75	681	2095	8:92	12.6	75
Lysozyme	0.88	343	2500	3:97	13.3	65
RNase A	0.20	4522	1150	49:51	6.9	74
DHFR	0.20	3135	790	49:51	6.9	135

## Simulation procedure

All MD simulations were performed with Gromacs 5.1.2.<sup>60-63</sup> Crystallographic structures were used as starting protein configurations (2LYM for hen egg-white lysozyme, 1FS3 for ribonuclease A, 1RX2 for dihydrofolate reductase), keeping the crystallographic water molecules.

Simulations employed periodic boundary conditions with a Particle Mesh Ewald (PME)<sup>64</sup> treatment of long range electrostatic interactions. A cut-off was used for real space electrostatic interactions (8.5 Å with Charmm27 and 8.0 Å with Amber99SB) and Lennard-Jones interactions (smooth switching between 7.5 and 8.5 Å for Charmm27 and truncation at 8.0 Å for Amber99SB). Covalent bonds containing a hydrogen atom were constrained with the LINCS algorithm.<sup>65,66</sup> The systems were solvated in rhombic dodecahedron boxes with a distance of at least 9.0 Å between protein and box edges. The number of glycerol molecules to be included in each system was determined iteratively to obtain the desired glycerol molar fraction ( $x_G$ ). To facilitate the placement of glycerol and water molecules at the largest  $x_G$  value (0.88), the distance between the protein and the box edges was set to at least 12.0 Å and the pre-determined number of glycerol and water molecules were inserted in the box. The box size was then allowed to decrease during an extra 50 ns NPT equilibration. The minimal number of ions was finally added to neutralize the system (8 chloride ions for lysozyme, 4 chloride ions for RNase A and 12 sodium ions for DHFR).

The equilibration of each system included an initial energy minimization, followed by a progressive temperature increase from 100 to 300 K in 200 ps, then a 50 ps trajectory at 300 K under NPT conditions with the velocity-rescale thermostat<sup>67</sup> and Berendsen barostat<sup>68</sup> and a final NPT equilibration (without any constraint on the protein atoms) with the same thermostat and the Parrinello-Rahman barostat.<sup>69</sup> The length of this final equilibration changed with the molar fraction, ranging from 50 ns for  $x_G \leq 0.60$  to 200 ns for  $x_G = 0.75$  and  $x_G = 0.88$  (see below and SI). Production runs of 100 ns were then performed with the velocity-rescale thermostat<sup>67</sup> and the Parrinello-Rahman barostat,<sup>69</sup> saving data every 10 ps. Convergence considerations are discussed further. For each system composition, 10 independent trajectories were generated from 10 different initial configurations and velocities. For lysozyme with a 10% glycerol molar fraction, an additional set of 10 independent trajectories was generated, together with two independent 1000 ns trajectories, in order to examine the convergence of our results. In total, more than 30  $\mu$ s of simulations were performed.

## Convergence

Typical MD simulations of such mixtures converge slowly, mostly due to two factors. First, the viscosity of water:glycerol solutions increases dramatically with the glycerol fraction,<sup>70</sup> thus slowing down the structural relaxation. Slow relaxation from the initial, random and out-of-equilibrium distribution of solvent molecules leads to a monotonic drift in the calculated  $\Gamma_{GP}$  for increasing trajectory lengths.<sup>32</sup> An additional factor arises at very low (co-)solvent molar fractions, because of the small number of (co-)solvent molecules sampling the simulation box. Prior simulation studies<sup>28</sup> suggested that convergence is reached after 100 ns for a single continuous trajectory at intermediate glycerol molar fractions (4 M,  $x_G = 0.09$ ). However, Fig 2-(c) shows that the computed  $\Gamma_{GP}$  value remains very dependent on the starting conformation for such trajectory lengths ( $\Gamma_{GP}(14\text{\AA})$  ranges from -2 to -16 for individual simulations).

In the present work, we adopted the following strategy to obtain converged values. First, for each system, 10 independent starting points were prepared, thus providing adequate configurational sampling independently of the slow structural relaxation time. Second, each trajectory was propagated for at least 100 ns, and stopped either when confidence intervals on two successive 50 ns-blocks overlap (Figure 3,  $x_G = 0.20$ ), or when the average  $\Gamma_{GP}$  exhibits non-monotonic variations on three successive 50 ns blocks (Figure 3,  $x_G = 0.05$ ). In each case, only the last 100 ns were considered for the analysis. Thus, in the following all averages are performed over 10 independent 100 ns simulations. We verified that such a trajectory length is much larger than the typical solvation shell composition relaxation time (see SI). Reported error bars are the Student 95% confidence interval, calculated from 2 ns independent blocks. Comparisons with a single long simulation are provided in SI.

## Impact of force-field

We now examine the sensitivity of our preferential hydration results on the chosen empirical potentials. We compared two different popular potentials for the lysozyme protein and

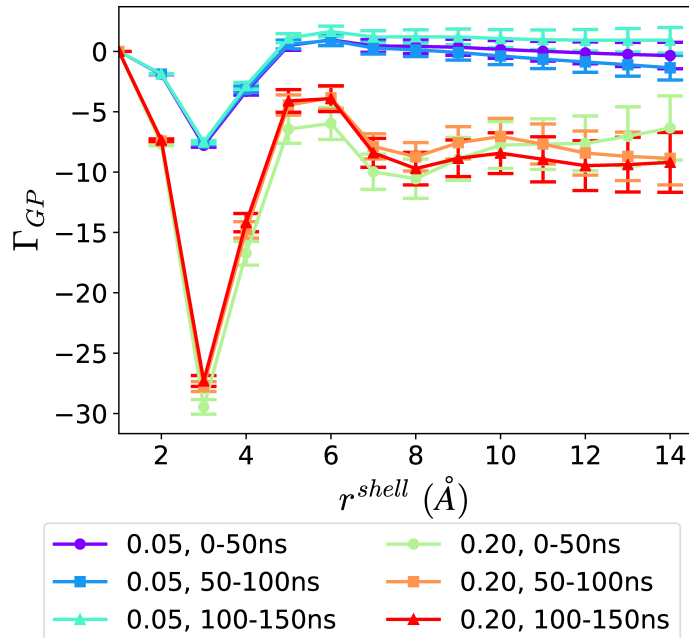


Figure 3: Comparisons of  $\Gamma_{GP}(R)$  calculated on successive time blocks for two solvent compositions ( $x_G = 0.05$  and  $x_G = 0.20$ ).

the glycerol molecules, while water was consistently described with the TIP3P parameters. For the protein, we contrasted the Charmm27 potential<sup>51</sup> used throughout our study with Amber99SB;<sup>57</sup> for glycerol, we compared the Charmm-derived force field used in our study with the default GAFF force field which was used in several preferential hydration simulations in water/glycerol mixtures.<sup>71,72</sup> This led to the following three force-field combinations for the protein/glycerol/water components: setup 1 with Charmm27/Charmm/TIP3P, setup 2 with Amber99SB/Charmm/TIP3P and setup 3 with Amber99SB/GAFF/TIP3P.

The  $\Delta x_G^{\text{shell}}$  and  $\Gamma_{GP}$  values for the three combinations are presented in Figure 4 and reveal two major effects. First, the preferential hydration behavior is found to be robust vis-a-vis a change of protein force field, even when Charmm and Amber force fields are combined for the protein and glycerol components. The protein behavior is almost identical with the two force fields. The protein RMSD, radius of gyration and solvent accessible surface area differ by 0.7%, -0.3% and 0.7% between setups 1 and 2; while the  $\Delta x_G^{\text{shell}}$  values are practically the same, the slightly more pronounced difference among the  $\Gamma_{GP}$  values probably arises from

the slight protein volume increase which also causes the solvent accessible surface area small +0.7% increase. Second, and in contrast to the protein potential, the glycerol force field is critical. Our results show that while the Charmm-derived glycerol force field indicates a protein preferential hydration, the GAFF potential leads to a preferential protein solvation by glycerol, in contradiction with the general consensus obtained from experiments.<sup>16,17,20–22</sup> This dramatic difference probably arises from the different Lennard-Jones non-bonding parameters of the glycerol carbons:  $\epsilon$  is 0.08 kJ/mol with the Charmm-derived force field and 0.46 kJ/mol with GAFF. Thus, GAFF glycerol is more attracted towards the protein. This leads to a net increase of the solvent accessible surface area (+3.1%), whereas RMSD and radius of gyration are not affected by the change of solvent force field (+0.1% and +0.2% respectively). We note that a Amber-based glycerol potential<sup>73</sup> was shown<sup>74</sup> to provide a good description of the thermodynamical and dynamical properties in water-glycerol binary mixtures; however, it uses the same  $\epsilon$  value as GAFF for glycerol carbons, and in the presence of a protein, one would expect it to lead to the same problematic behavior as GAFF. Our results thus call for a careful examination of protein/glycerol/water simulation results obtained with the GAFF glycerol force field. The sensitivity of simulation results to the cosolvent force field has already been observed in other studies.<sup>74–76</sup> We will show further that the chosen Charmm-derived force field combination yields results in good agreement with experiments and is thus appropriate for our present purposes.

## Test of solvent-cosolvent exchange model

Before we present a comparison between our simulation results and available experimental measurements, we pause to examine the popular solvent-cosolvent exchange model,<sup>23</sup> which is very often used to connect experimental results<sup>21</sup> to a molecular interpretation of the protein solvation shell composition. The central assumption in this model is that water and cosolvent molecules undergo exchanges between the shell and bulk domains in a 1:1

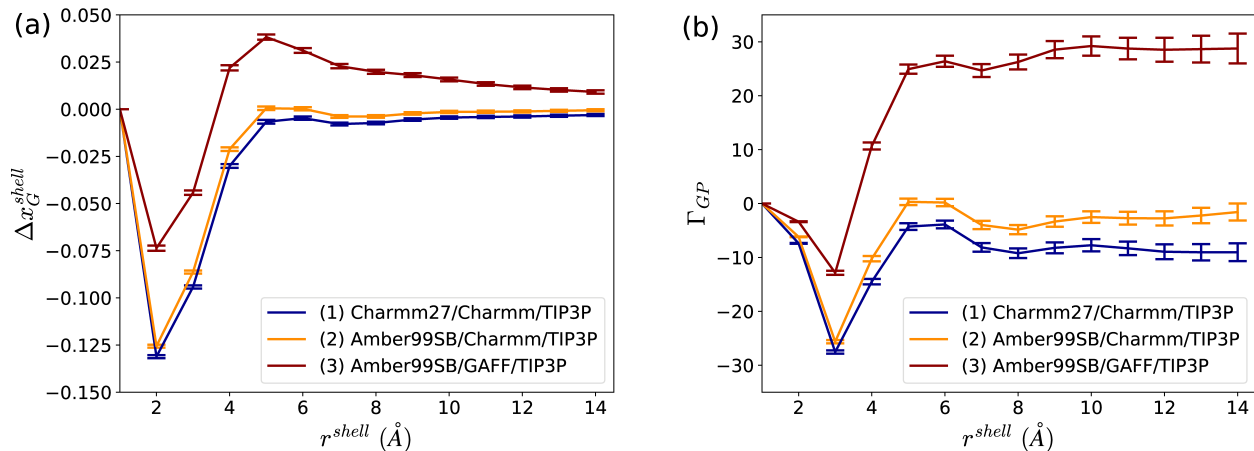


Figure 4: Comparisons of (a)  $\Delta x_G^{\text{shell}}$  and (b)  $\Gamma_{GP}$  for lysozyme in a water/glycerol mixture with  $x_G = 0.20$  for different protein/glycerol/water force field combinations.

stoichiometric ratio, i.e.  $P \bullet G_{\text{shell}} + W_{\text{bulk}} \rightleftharpoons G_{\text{bulk}} + P \bullet W_{\text{shell}}$ . This assumption would be valid if the number of solvent binding sites at the protein surface was constant and all sites had the same binding constant. These assumptions lead to a great simplification of the preferential hydration problem, however their validity can be questioned, especially when the solvent and cosolvent molecules have very different sizes, as for water and glycerol whose molar volumes differ by a factor of  $\approx 4$ .

We therefore computed the average number of solvent molecules within 6 Å from the lysozyme protein for different solvent compositions. The results presented in Figure 5-(a) clearly show that the main approximation of the solvent exchange model is not satisfied. First, the number of solvent molecules within this fixed solvation shell changes dramatically with solvent composition, ranging from 280 at  $x_G = 0.88$  to 1050 at  $x_G = 0.01$  (we note that the ratio between these numbers is close to that of the solvent and cosolvent molar volumes). Second, for a given solvent composition, the number of solvent molecules within the shell exhibits large fluctuations (3% average standard deviation, i.e.  $\approx 25$  molecules at  $x_G^{\text{bulk}} \leq 0.10$ ). These results further support the need to consider  $\Delta x_G^{\text{shell}}$  to analyze the shell composition, and not only  $\Gamma_{GP}$  which fluctuates with the number of solvent molecules in the shell.

However, our simulations reveal that while the number of solvent molecules in the shell exhibits large fluctuations, the volume of the shell changes very little. Figure 5-(b) shows the shell volume calculated as described in Equation S14 (see SI). Fluctuations are smaller, with 2% average standard deviation (e.g.  $\approx 435$  mL/mol at  $x_G^{\text{bulk}} = 0.10$ ). Our results thus show that the extension of the solvent exchange model based on the volume fraction<sup>24</sup> is preferable to the original constant-number model, and call for a reinterpretation of experimental results previously analyzed with the simple solvent-cosolvent exchange model.

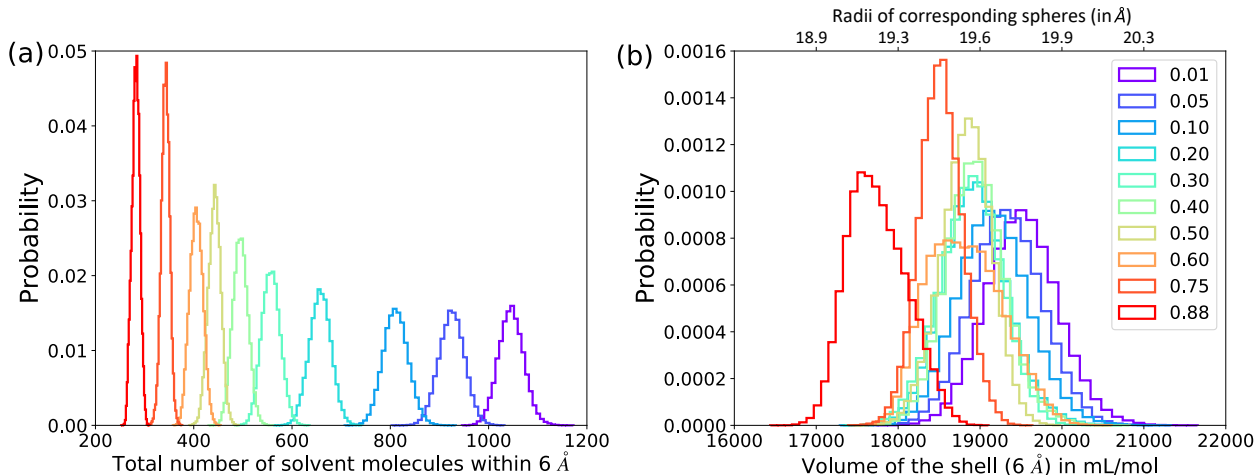


Figure 5: Test of the assumptions in the solvent exchange model, (a) Number of molecules in the first shell, (b) Volume of the first shell.

## Comparison with available experimental results

We now provide a detailed comparison between our simulation results and a series of experimental measurements of protein preferential hydration in water/glycerol mixtures (since we showed that the solvent exchange model approximations are not satisfied, we will not include experimental results relying on this model). We start with the thermodynamic  $\Gamma_{GP}$  value. Figure 6-(a) shows that for three typical systems, involving different proteins and different solvent compositions, respectively lysozyme in a  $x_G^{\text{bulk}} = 0.06$  solution,<sup>77</sup> lysozyme in a  $x_G^{\text{bulk}}=0.50$  solution,<sup>21</sup> and RNaseA in a  $x_G^{\text{bulk}} = 0.14$  solution,<sup>16</sup> the simulation results

exhibit a very good correlation with the experimental values (our simulated solvent compositions were very similar to the experimental ones with respectively  $x_G^{\text{bulk}}=0.06, 0.50$  and  $0.20$ ). While simulations slightly underestimate  $\Gamma_{GP}$ , the trend is correctly reproduced over a broad range of  $\Gamma_{GP}$  values.

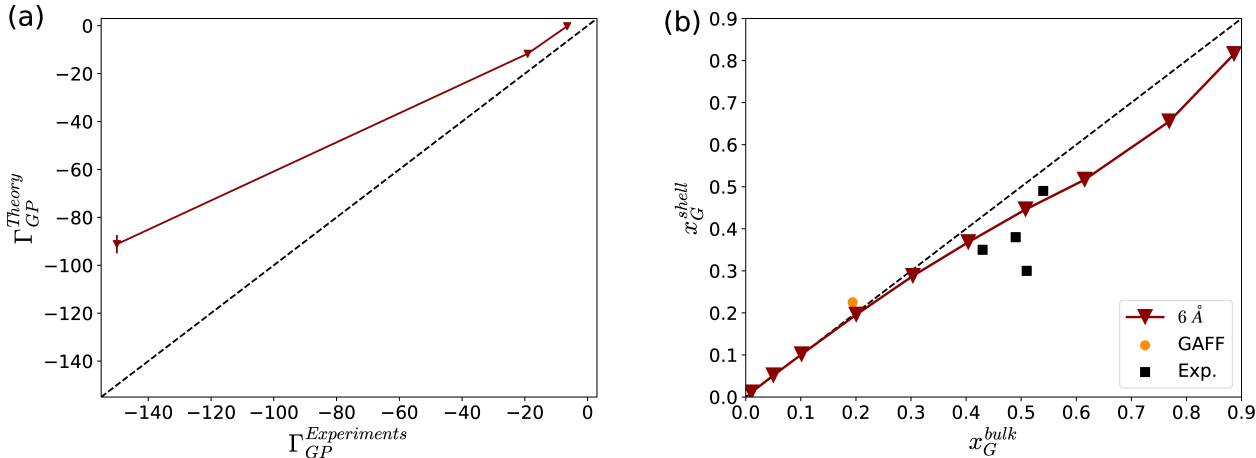


Figure 6: (a) Comparison between experimental<sup>16,21,77</sup> and simulated values of  $\Gamma_{GP}$ . (b) Shell composition  $x_G^{\text{shell}}$  for different bulk solvent compositions from our simulations (triangles) and from experiments<sup>21</sup> (squares), assuming  $r^{\text{shell}} = 6 \text{ \AA}$ . Our simulation result with the glycerol GAFF force field (dot) is shown to incorrectly predict a glycerol enrichment in the solvation shell. In both panels, the dotted line provides the  $y=x$  reference.

Comparing the solvation shell composition obtained in the simulations with experimental determinations is not straightforward, since many experimental techniques do not directly probe the solvent composition and simplifying assumptions are required to obtain a molecular interpretation. An extensive characterization of lysozyme solvation in water:glycerol mixtures was obtained *via* neutron scattering experiments.<sup>21</sup> We thus compare our simulation results of  $x_G^{\text{shell}}$  and  $x_G^{\text{bulk}}$  with their determination, using similar shell thickness values ( $5.9 \text{ \AA}$  in ref 21 and  $6.0 \text{ \AA}$  in our analysis) and the same protein concentration (typically 90 g/L in our simulations, see Table 1, and in the experiments). Figure 6-(b) shows that our simulated values of  $x_G$  are in very good agreement with the experimental data. (In contrast, we note that the GAFF glycerol force field leads to a marked preferential protein solvation by glycerol at  $x_G^{\text{bulk}}=0.2$ , and thus to a behavior opposite to the preferential hydration



determined both experimentally and with our reference set of simulations at  $x_G^{\text{bulk}}=0.4$ .)

A recent study based on wide-angle X-ray scattering and small-angle neutron scattering<sup>27</sup> suggested that a change in solvation behavior occurs at a glycerol concentration of 40% v/v ( $x_G = 0.14$ ). Our simulations do not show such change in this range of compositions. The contrast may come from the models used to analyze the experimental data, especially the assumption of glycerol non-penetration into the protein solvation shell at low  $x_G$ , while our simulations show that glycerol is always present within the first solvation shell.

## Impacts of protein nature and solvent composition on preferential hydration

Now that the validity of our simulation approach has been confirmed, we analyze our simulation results to propose a molecular picture of protein preferential hydration and of its changes with solvent composition and protein nature.

### Glycerol concentration

We first examine how preferential hydration changes with glycerol concentration. The dependence of  $\Gamma_{GP}$  on the solvent composition is known experimentally but has so far not been investigated in simulations.  $\Gamma_{GP}$  and  $\Delta x_G^{\text{shell}}$  are reported in Fig 7 for lysozyme in aqueous glycerol solutions ranging from dilute  $x_G=0.01$  to concentrated  $x_G=0.88$ . Protein preferential hydration is shown to dramatically change with solvent composition. While  $\Gamma_{GP}$  is usually considered to be proportional to the glycerol molality,<sup>17,30</sup> our results reveal that this is only verified on a narrow range of concentrations and that the situation is far more complex. Three main composition regimes can be identified.

First, for  $0 \leq x_G \leq 0.2$  (i.e. less than 50% glycerol in volume), both  $\Gamma_{GP}$  and  $\Delta x_G^{\text{shell}}$  remain very close to 0, indicating that there is no preferential solvation by either component of the mixture ( $|\Delta x_G^{\text{shell}}(6\text{\AA})| < 0.005$ ). In this composition range, glycerol is not repelled

from the protein interface. At very low glycerol fractions ( $x_G=0.01$ ), our results even suggest a slight protein preferential solvation by glycerol ( $\Gamma_{GP}(14\text{\AA}) = +0.46 \pm 0.33$ ), which is in stark contrast with the traditional picture of glycerol depletion.

Next, for  $0.2 \leq x_G \leq 0.75$ , both  $\Gamma_{GP}$  and  $\Delta x_G^{\text{shell}}$  exhibit a strong decrease with increasing glycerol fraction. As the glycerol concentration increases, glycerol molecules are increasingly disfavored in the protein solvation shell. This can be qualitatively explained by the different molecular sizes of water and glycerol and by the progressive saturation of the protein solvation shell in bulky glycerol molecules, while smaller water molecules can still access tighter binding sites.

Finally,  $\Gamma_{GP}$  and  $\Delta x_G^{\text{shell}}$  reach an extremum at  $x_G = 0.75$  and increase (decrease in absolute value) for increasing  $x_G$  values. By definition,  $\Delta x_G$  decays to 0 in pure glycerol at  $x_G^{\text{bulk}} = 1$ , which explains the presence of the extremum. The location of the extremum in our simulations is consistent with neutron scattering experiments<sup>21</sup> which suggested that it occurs for  $x_G > 0.6$ .

## Protein nature

We now study how much the solvent shell composition is affected by the protein nature. We therefore compare the protein solvation shell compositions for three different globular proteins – lysozyme, RNaseA and DHFR – at the same glycerol molar fraction  $x_G = 0.20$  (this solvent composition was for example recently used in experimental studies of enzymatic reactivity in water:glycerol mixtures<sup>3</sup>). Figure 8-(a) reveals that the preferential interaction coefficient always exhibits similar variations with increasing shell thickness  $r^{\text{shell}}$  for all 3 proteins, but that the  $\Gamma_{GP}$  magnitude strongly changes with the protein nature. As shown above,  $\Gamma_{GP}$  scales with the solvation shell size, which is here very different for the three proteins (657 solvent molecules for lysozyme, 707 for RNaseA and 812 for DHFR at 6 Å of the protein). However, the  $\Gamma_{GP}$  differences between the series of proteins cannot be solely explained by their different sizes. In a complementary analysis, we examine the difference

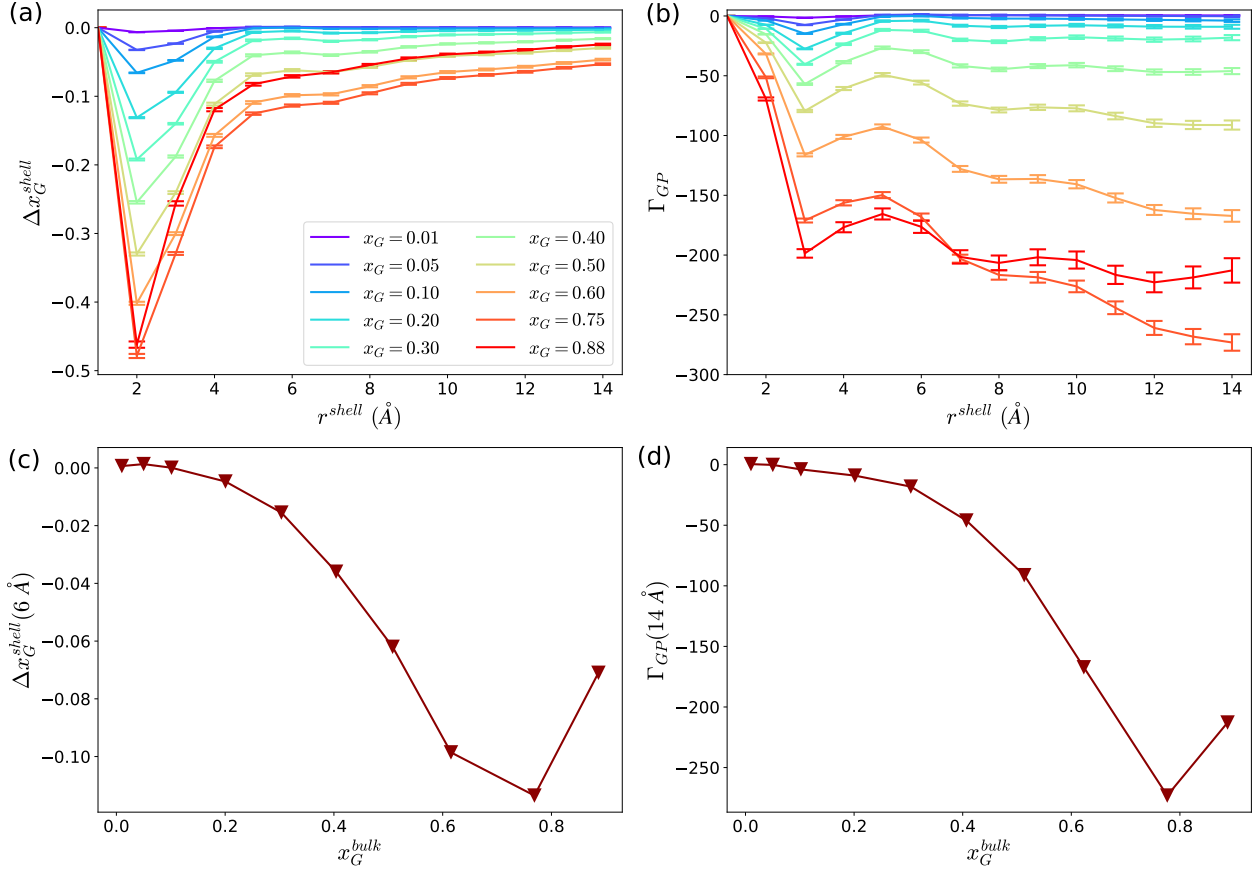


Figure 7: Influence of bulk glycerol molar fraction on the difference of molar fractions  $\Delta x_G^{shell}$  ((a) and (c)) and on the preferential interaction coefficient  $\Gamma_{GP}$  ((b) and (d)).

in glycerol and water molar fractions in the protein vicinity, in order to focus on the differences in solvation shell compositions independently of the shell sizes. The extent of glycerol depletion in the protein solvation shell is found to be different for the 3 selected proteins (Figure 8-(b)). This shows that preferential interaction depends on characteristic protein features, possibly involving their shape and the nature of the solvent exposed groups. Prior suggestions included hydrophobicity and hydrophilicity considerations<sup>30,31</sup> and side-chain motions.<sup>28</sup> However, the reported  $\Gamma_{GP}$  and  $\Delta x_G^{shell}$  values are averaged over the protein solvation shell, while the protein exposed surfaces are extremely heterogeneous, both in their topography and in their local polarities, and a local analysis is required. To gain a better molecular understanding of preferential solvation, the following section will thus focus on the correlation between the protein exposed sites' properties and local preferential solvation.

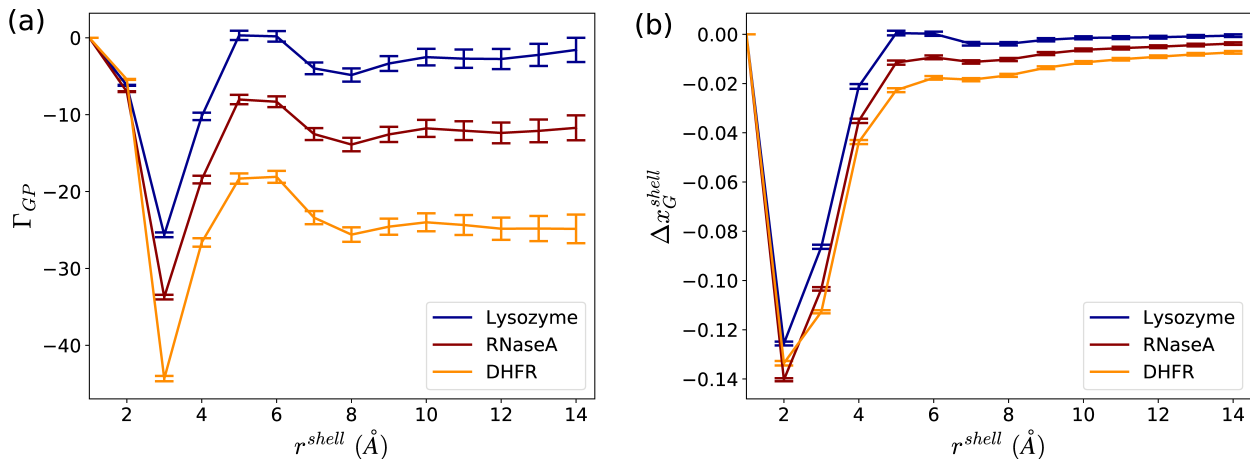


Figure 8: (a) Preferential interaction coefficient  $\Gamma_{GP}$  for different proteins at  $x_G = 0.20$ , (b) difference in molar fractions  $\Delta x_G^{shell}$ . All proteins were described with the Amber99SB force field.

## Local preferential solvation: charge and accessibility

To determine the impact of each protein exposed group on the local solvent composition, we have determined the local  $\Gamma_{GP}$  and  $\Delta x_G^{shell}$  values next to each solvent-exposed protein site, for the three selected protein systems.

We first examine the role of the residue’s charge. Figure 9-(a) shows that no significant differences are observed between the distributions of  $\Gamma_{GP}$  for neutral, positively charged and negatively charged residues. In contrast,  $\Delta x_G^{shell}$  values (Figure 9-(b)) suggest that glycerol depletion is less pronounced next to positively and negatively charged residues than next to neutral residues (however, the small fraction of charged residues causes a large uncertainty).

We now turn to protein local topography’s effects on preferential solvation. Figure 10 reveals that regions with the most negative  $\Delta x_G^{shell}$  values (i.e. with the largest glycerol depletion) are found in cavities, while protuberances typically yield  $\Delta x_G^{shell}$  values closer to 0 or positive. This suggests that the protein exposed surface topography plays an important role in the biomolecule’s preferential solvation. Glycerol being bulkier than water, its penetration in tight cavities is hindered.

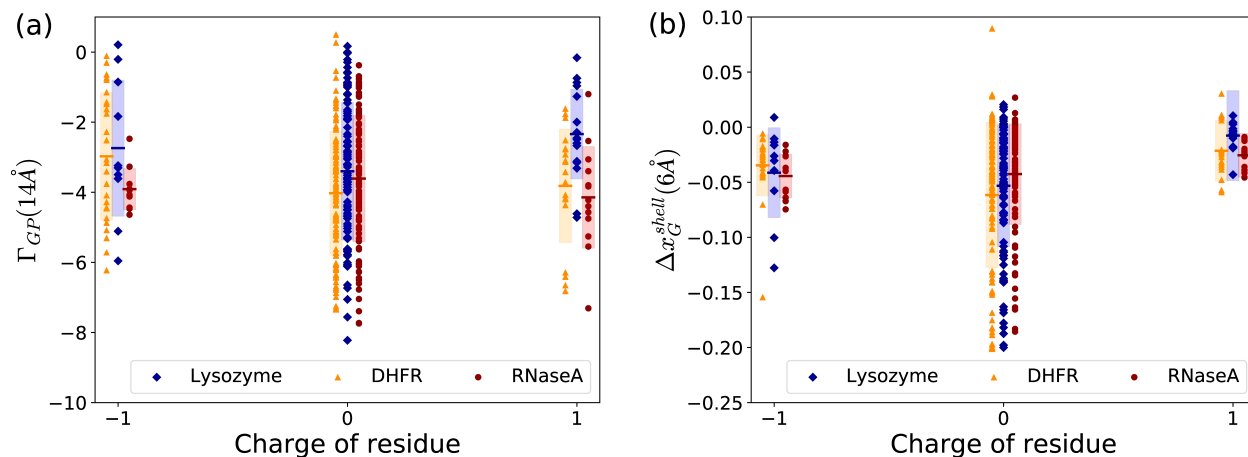


Figure 9: Local  $\Gamma_{GP}$  (a) and  $\Delta x_G^{\text{shell}}$  (b) next to each individual residue, classified according to the residue charge, for  $x_G = 0.20$ . Average values are represented by horizontal lines and standard deviation by shaded boxes.

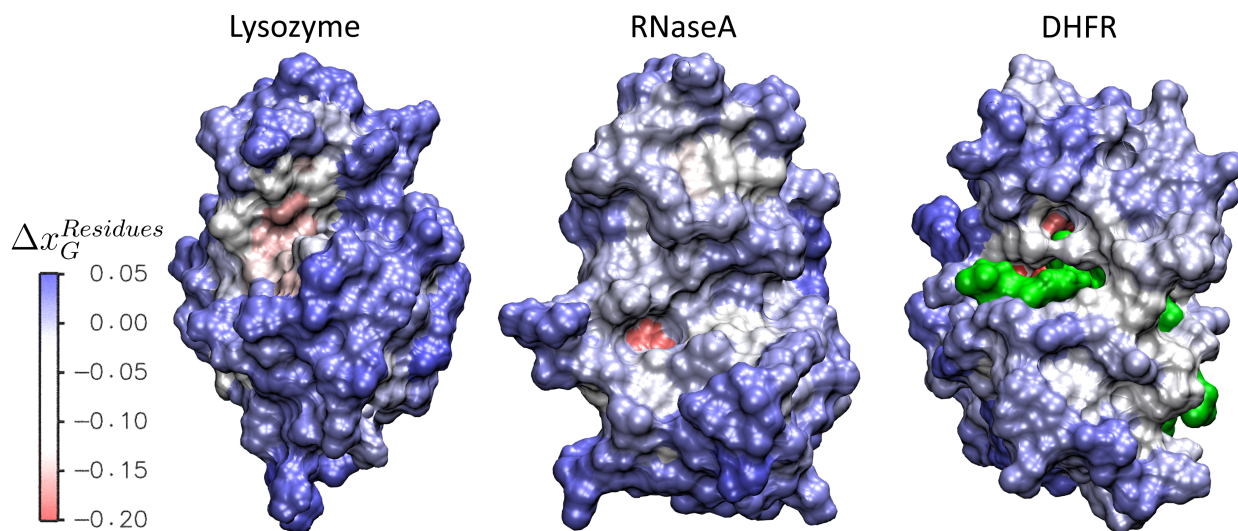


Figure 10: Color map of site-resolved  $\Delta x_G^{\text{shell}}$  within  $6\text{\AA}$  and in a  $x_G = 0.20$  mixture throughout the solvent exposed surfaces of lysozyme (left), RNase A (middle) and DHFR (right, with the dihydrofolate substrate and NADPH cofactor in green).

A more quantitative measure of the local protein topography is provided by the solvent accessible surface area (SASA). Table 2 reports the correlations between each protein exposed residue's SASA and the local  $\Delta x_G^{\text{shell}}$  (resp.  $\Gamma_{GP}$ ). Our results show that  $\Gamma_{GP}$  becomes more negative when the SASA decreases, consistent with the accessibility argument (see Figure S5 and Figure S6 in SI). However, this correlation remains very moderate, since

$\Gamma_{GP}$  is also sensitive to the solvation shell size. The local shell composition  $\Delta x_G^{\text{shell}}$  is more strongly correlated to the local SASA: glycerol is more depleted in less accessible sites. This correlation is enhanced when the solvent mixture is enriched in glycerol, probably due to a progressive saturation of accessible glycerol binding sites at the protein surface. The importance of these steric considerations for preferential solvation was already discussed<sup>78</sup> in the context of very large cosolvents including e.g. polyethylene glycol, where steric repulsion was suggested to be the main source of protein preferential hydration. We stress that for the opposite situation of small cosolvent molecules, one would expect this accessibility argument to lead to a reduced depletion, or even a preferential solvation by the cosolvent; this has been experimentally observed e.g. for mixtures of water and acetonitrile.<sup>79</sup>

**Table 2: Pearson correlation coefficients (R) between solvent accessible surface area (1.4 Å probe radius) and preferential interaction coefficient  $\Gamma_{GP}$  and difference of molar fractions  $\Delta x_G^{\text{shell}}$ , for three selected proteins and in a series of solvent mixtures.**

(vs SASA)	Protein	Lysozyme							RNaseA	DHFR
	$x_G$	0.01	0.05	0.20	0.40	0.60	0.75	0.88	0.20	0.20
$\Delta x_G^{\text{shell}}$	R	0.28	0.45	0.65	0.69	0.71	0.69	0.61	0.58	0.51
$\Gamma_{GP}$	R	0.11	0.23	0.42	0.49	0.45	0.50	0.43	0.41	0.30

The correlation between local solvent composition and protein surface topography suggests that preferential solvation has an important entropic contribution. We therefore extended our study to a series of temperatures, and considered lysozyme at two glycerol molar fractions,  $x_G = 0.05$  and  $0.20$ , at 275 K, 300 K, 325 K and 350 K. This is to our knowledge the first temperature-dependent simulation study of protein preferential solvation. Figure 11 shows that glycerol depletion within the protein solvation shell is enhanced with increasing temperature.

In order to determine the enthalpic and entropic contributions to  $\Gamma_{GP}$ , we show in the SI that Equation 6 can be expressed as

$$\Gamma_{GP} = \langle n_G^{\text{shell}} + n_W^{\text{shell}} \rangle (1 - K), \quad (8)$$

where  $K$  is the ratio between the shell–bulk partition coefficients for a water molecule and for any solvent particle. When both water and glycerol have the same partition ratio,  $K=1$  and  $\Gamma=0$  (prior studies<sup>30</sup> suggested a connection between  $\Gamma_{GP}$  and a different equilibrium constant which did not satisfy this condition). In each solvation shell site, the average free energy difference  $\Delta G_{site}$  that governs the site preferential hydration temperature dependence is thus

$$\Delta G_{site} = -k_B T \ln \left( 1 - \frac{\Gamma_{GP}}{\langle n_G^{\text{shell}} + n_W^{\text{shell}} \rangle} \right) \quad (9)$$

and for moderate preferential hydration coefficients ( $\Gamma_{GP}/\langle n_G^{\text{shell}} + n_W^{\text{shell}} \rangle \ll 1$ ), the total free energy that determines  $\Gamma_{GP}$  is approximately

$$\Delta G_{\text{tot}} \simeq k_B T \Gamma_{GP}. \quad (10)$$

The plots shown in the SI provide approximate determinations of the entropic and entropic contributions to  $\Gamma_{GP}$ . At  $x_G = 0.05$ , our simulations yield  $\Delta H_{\text{tot}} = 6.5$  kcal/mol,  $T\Delta S_{\text{tot}}(300K) = 6.9$  kcal/mol and  $T\Delta S_{\text{tot}}(350K) = 7.8$  kcal/mol; in the more concentrated glycerol solution  $x_G = 0.20$ , these values change to  $\Delta H_{\text{tot}} = 21.7$  kcal/mol,  $T\Delta S_{\text{tot}}(300K) = 27.1$  kcal/mol and  $T\Delta S_{\text{tot}}(350K) = 30.3$  kcal/mol. These results therefore first show that the enthalpic contribution – due to the different protein-water and protein-glycerol interaction energies – favors the presence of glycerol within the solvation shell. The second important point is that the entropic accessibility term leads to glycerol depletion next to the protein.

The contributions of the enthalpic and entropic terms strongly change with solvent composition. At low glycerol fraction, preferential solvation is governed by both enthalpic and entropic contributions; in contrast, at larger glycerol fractions, the relative importance of entropy becomes more pronounced. This is consistent with the increasing correlation between the entropic SASA factor and  $\Delta x_G^{\text{shell}}$  when the overall glycerol molar fraction increases. This picture also explains why at higher temperature, when the importance of the entropic term is enhanced with respect to that of the enthalpic one, correlations between  $\Delta x_G^{\text{shell}}$  (and

$\Gamma_{GP}$ ) and the entropic SASA factor are larger (for  $\Delta x_G^{shell}$ , R is respectively 0.61 and 0.65 at  $x_G = 0.05$  and 0.20; for  $\Gamma_{GP}$ , R is respectively 0.37 and 0.49). Finally, we point out that this analysis implies that preferential protein glyceration can be expected at low molar fractions and at low temperature, which is potentially important for cryopreservation. We also note that a transition from cosolvent depletion to attraction when the concentration is changed was already found experimentally.<sup>80</sup>

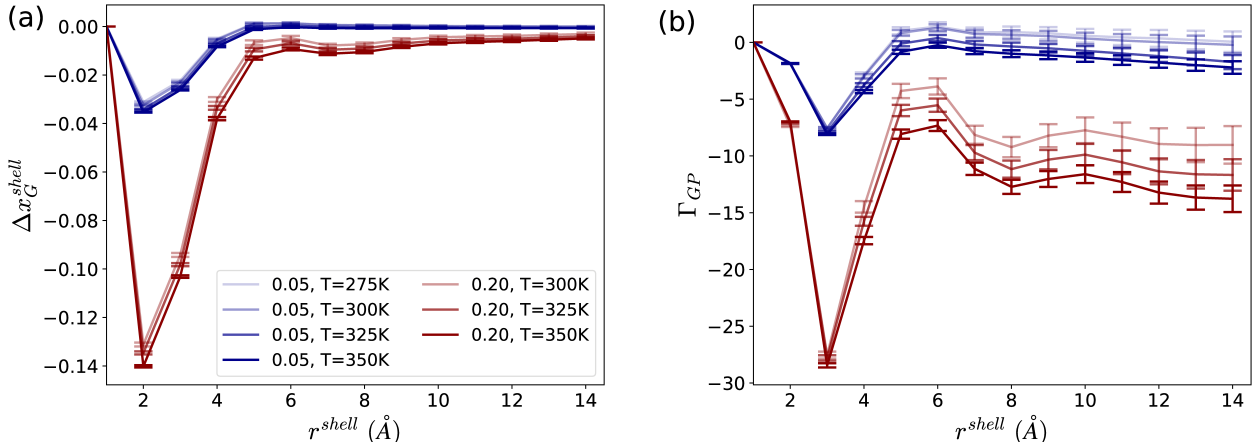


Figure 11: Temperature influence on (a) the difference of molar fractions and (b) the preferential interaction coefficient for lysozyme at two glycerol molar fractions ( $x_G = 0.05$  and 0.20).

## Conclusion

In this work, we have studied the preferential solvation of proteins in solution mixtures of water and glycerol using theoretical models and molecular dynamics simulations.

We have first assessed the usual approximations employed in such simulation studies, examined the validity of typical biochemical force fields, established an improved simulation procedure to obtain converged results, and validated our approach with direct comparisons to experiments. We have further shown that the widely used solvent-cosolvent shell exchange picture is too simplified for cases where solvent and cosolvent molecular volumes are very different, and we have suggested that the constant shell volume model should be preferred.



Our simulations show that the composition of the solvation layer next to a protein can significantly differ from that of the bulk solvent, as well established in the literature. However, our results reveal that these differences between the shell and bulk compositions exhibit dramatic changes with solvent composition, temperature and protein nature. Our simulations show that for aqueous solutions with less than 50% glycerol in volume, protein solvation shells have approximately the same composition as the bulk solvent. This implies that in contrast to the widely used and simplified image representing the first protein solvation layer as totally depleted in glycerol (or other cosolvent), there is a large number of glycerol molecules in direct contact with the protein. Such concentrations are typically those used in cryopreservation applications;<sup>4</sup> our results thus show that the way in which these mixtures protect proteins is not by encasing them within a neat water layer. In addition, our results show that the protein interface’s impact on the solvent composition goes much further than the first solvation layer and typically extends up to 10 Å from the protein surface.

Further, the molecular picture provided by our simulations shows that the local solvent composition depends on the local protein features, and thus reflect the great heterogeneity of the protein exposed surface. Our results show that a major factor affecting the local solvent composition in glycerol/water mixtures is entropic and arises from the solvent-exposed protein surface topography: larger glycerol molecules are preferentially excluded from cavities and buried sites which remain accessible to smaller water molecules.

Our present results suggest that for other cosolvent whose molecular volume is much larger than that of water, accessibility and entropy should play a key role in determining preferential solvation. In contrast, other cosolvents – such as urea and guanidinium – whose size is closer to that of water molecules are more likely to be found in the protein’s cavities. Moreover, while glycerol interacts with proteins via hydroxyl groups which are similar to the water OH groups, other cosolvents which are more polar or charged are expected to exhibit larger favorable enthalpic protein interactions, and thus to be more likely to reside in the protein solvation shell.

Molecular-level descriptions of protein solvation shells provided by molecular simulations will be a valuable guide for an improved understanding of protein’s functioning<sup>3,81</sup> and preservation<sup>4</sup> in water-cosolvent solutions.

## Supporting Information

We provide in Supporting Information additional details regarding the derivation of Eq. 6, the heavy-atom based definitions of  $\Gamma_{GP}$  and  $\Delta x_G^{shell}$ , a comparison between two approaches to compute  $\Gamma_{GP}$ , the convergence of simulations, the influence of individual residues, the calculation of the shell volume and the impact of temperature.

## Acknowledgments

The authors thank Vincent Vagenende for providing the glycerol force field used in previous simulations. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement No. 279977. This work was granted access to the HPC resources of TGCC and CINES under the allocation 2016-077156 made by GENCI.

## 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) Canchi, D. R.; García, A. E. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.

- (3) Loveridge, E. J.; Tey, L.-H.; Allemann, R. K. Solvent Effects on Catalysis by Escherichia Coli Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2010**, *132*, 1137–1143.
- (4) Hubálek, Z. Protectants Used in the Cryopreservation of Microorganisms. *Cryobiology* **2003**, *46*, 205–229.
- (5) Morris, G. J.; Goodrich, M.; Acton, E.; Fonseca, F. The High Viscosity Encountered during Freezing in Glycerol Solutions: Effects on Cryopreservation. *Cryobiology* **2006**, *52*, 323–334.
- (6) Marcus, Y. *Solvent Mixtures: Properties and Selective Solvation*; CRC Press, 2002.
- (7) Hayashi, Y.; Puzenko, A.; Balin, I.; Ryabov, Y. E.; Feldman, Y. Relaxation Dynamics in Glycerol-Water Mixtures. 2. Mesoscopic Feature in Water Rich Mixtures. *J. Phys. Chem. B* **2005**, *109*, 9174–9177.
- (8) McLain, S. E.; Soper, A. K.; Luzar, A. Investigations on the Structure of Dimethyl Sulfoxide and Acetone in Aqueous Solution. *J. Chem. Phys.* **2007**, *127*, 174515.
- (9) Puzenko, A.; Hayashi, Y.; Feldman, Y. Space and Time Scaling in Glycerol–Water Mixtures. *J. Non-Cryst. Solids* **2007**, *353*, 4518–4522.
- (10) Harpham, M. R.; Levinger, N. E.; Ladanyi, B. M. An Investigation of Water Dynamics in Binary Mixtures of Water and Dimethyl Sulfoxide. *J. Phys. Chem. B* **2008**, *112*, 283–293.
- (11) 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*, 7799–807.
- (12) Wong, D. B.; Sokolowsky, K. P.; El-Barghouthi, M. I.; Fenn, E. E.; Giammanco, C. H.; Sturlaugson, A. L.; Fayer, M. D. Water Dynamics in Water/DMSO Binary Mixtures. *J. Phys. Chem. B* **2012**, *116*, 5479–5490.

- (13) Bachler, J.; Fuentes-Landete, V.; Jahn, D. A.; Wong, J.; Giovambattista, N.; Loerting, T. Glass Polymorphism in Glycerol–Water Mixtures: II. Experimental Studies. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11058–11068.
- (14) 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.
- (15) van der Vegt, N. F. A.; Nayar, D. The hydrophobic effect and the role of cosolvents. *J. Phys. Chem. B* **2017**, *121*, 9986–9998.
- (16) Gekko, K.; Timasheff, S. N. Mechanism of Protein Stabilization by Glycerol: Preferential Hydration in Glycerol-Water Mixtures. *Biochemistry* **1981**, *20*, 4667–4676.
- (17) Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T. Vapor Pressure Osmometry Studies of OsmolyteProtein Interactions: Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of Osmotic Stress Experiments in Vitro. *Biochemistry* **2000**, *39*, 4455–4471.
- (18) Ravindra, R.; Winter, R. Pressure Perturbation Calorimetry: A New Technique Provides Surprising Results on the Effects of Co-Solvents on Protein Solvation and Unfolding Behaviour. *ChemPhysChem* **2004**, *5*, 566–71.
- (19) Spinozzi, F.; Ortore, M. G.; Sinibaldi, R.; Mariani, P.; Esposito, A.; Cinelli, S.; Onori, G. Microcalorimetric Study of Thermal Unfolding of Lysozyme in Water/Glycerol Mixtures: An Analysis by Solvent Exchange Model. *J. Chem. Phys.* **2008**, *129*, 035101.
- (20) Lehmann, M. S.; Zaccai, G. Neutron Small-Angle Scattering Studies of Ribonuclease in Mixed Aqueous Solutions and Determination of the Preferentially Bound Water. *Biochemistry* **1984**, *23*, 1939–42.

- (21) Sinibaldi, R.; Ortore, M. G.; Spinozzi, F.; Carsughi, F.; Frielinghaus, H.; Cinelli, S.; Onori, G.; Mariani, P. Preferential Hydration of Lysozyme in Water/Glycerol Mixtures: A Small-Angle Neutron Scattering Study. *J. Chem. Phys.* **2007**, *126*, 235101.
- (22) Betting, H.; Hackel, M.; Hinz, H. J.; Stockhausen, M. Spectroscopic Evidence for the Preferential Hydration of RNase A in Glycerol-Water Mixtures: Dielectric Relaxation Studies. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1688–1692.
- (23) Schellman, J. A. The Thermodynamics of Solvent Exchange. *Biopolymers* **1994**, *34*, 1015–26.
- (24) Schellman, J. A. Protein Stability in Mixed Solvents: A Balance of Contact Interaction and Excluded Volume. *Biophys. J.* **2003**, *85*, 108–125.
- (25) King, J. T.; Kubarych, K. J. Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2D-IR Spectroscopy. *J. Am. Chem. Soc.* **2012**, *134*, 18705–12.
- (26) Olsson, C.; Jansson, H.; Swenson, J. The Role of Trehalose for the Stabilization of Proteins. *J. Phys. Chem. B* **2016**, *120*, 4723–31.
- (27) Hirai, M.; Ajito, S.; Sugiyama, M.; Iwase, H.; Takata, S.-i.; Shimizu, N.; Igarashi, N.; Martel, A.; Porcar, L. Direct Evidence for the Effect of Glycerol on Protein Hydration and Thermal Structural Transition. *Biophys. J.* **2018**, *115*, 313–327.
- (28) Vagenende, V.; Trout, B. L. Quantitative Characterization of Local Protein Solvation to Predict Solvent Effects on Protein Structure. *Biophys. J.* **2012**, *103*, 1354–62.
- (29) Arthur, E. J.; King, J. T.; Kubarych, K. J.; Brooks, C. L. Heterogeneous Preferential Solvation of Water and Trifluoroethanol in Homologous Lysozymes. *J. Phys. Chem. B* **2014**, *118*, 8118–8127.

- (30) Baynes, B. M.; Trout, B. L. Proteins in Mixed Solvents: A Molecular-Level Perspective. *J. Phys. Chem. B* **2003**, *107*, 14058–14067.
- (31) Vagenende, V.; Yap, M. G. S.; Trout, B. L. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry* **2009**, *48*, 11084–96.
- (32) Vagenende, V.; Yap, M. G. S.; Trout, B. L. Molecular Anatomy of Preferential Interaction Coefficients by Elucidating Protein Solvation in Mixed Solvents: Methodology and Application for Lysozyme in Aqueous Glycerol. *J. Phys. Chem. B* **2009**, *113*, 11743–11753.
- (33) Casassa, E. F.; Eisenberg, H. Thermodynamic Analysis of Multicomponent Solutions. *Adv. Protein Chem.* **1964**, *19*, 287 – 395.
- (34) Record, M. T.; Anderson, C. F. Interpretation of Preferential Interaction Coefficients of Nonelectrolytes and of Electrolyte Ions in Terms of a Two-Domain Model. *Biophys. J.* **1995**, *68*, 786–794.
- (35) 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.
- (36) Timasheff, S. N. Protein-Solvent Preferential Interactions, Protein Hydration, and the Modulation of Biochemical Reactions by Solvent Components. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9721–9726.
- (37) Smith, P. E. Equilibrium Dialysis Data and the Relationships between Preferential Interaction Parameters for Biological Systems in Terms of KirkwoodBuff Integrals. *J. Phys. Chem. B* **2006**, *110*, 2862–2868.
- (38) Shulgin, I. L.; Ruckenstein, E. A Protein Molecule in an Aqueous Mixed Solvent: Fluctuation Theory Outlook. *J. Chem. Phys.* **2005**, *123*, 054909.

- (39) Schurr, J. M.; Rangel, D. P.; Aragon, S. R. A Contribution to the Theory of Preferential Interaction Coefficients. *Biophys. J.* **2005**, *89*, 2258–76.
- (40) Shulgin, I. L.; Ruckenstein, E. A Protein Molecule in a Mixed Solvent: The Preferential Binding Parameter via the Kirkwood-Buff Theory. *Biophys. J.* **2006**, *90*, 704–7.
- (41) Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. Recent Applications of KirkwoodBuff Theory to Biological Systems. *Cell Biochem. Biophys.* **2008**, *50*, 1–22.
- (42) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.* **1951**, *19*, 774–777.
- (43) Inoue, H.; Timasheff, S. N. Preferential and Absolute Interactions of Solvent Components with Proteins in Mixed Solvent Systems. *Biopolymers* **1972**, *11*, 737–743.
- (44) Smolin, N.; Winter, R. Effect of Temperature, Pressure, and Cosolvents on Structural and Dynamic Properties of the Hydration Shell of SNase: A Molecular Dynamics Computer Simulation Study. *J. Phys. Chem. B* **2008**, *112*, 997–1006.
- (45) Shukla, D.; Shinde, C.; Trout, B. L. Molecular Computations of Preferential Interaction Coefficients of Proteins. *J. Phys. Chem. B* **2009**, *113*, 12546–12554.
- (46) Saladino, G.; Pieraccini, S.; Rendine, S.; Recca, T.; Francescato, P.; Speranza, G.; Sironi, M. Metadynamics Study of a  $\beta$ -Hairpin Stability in Mixed Solvents. *J. Am. Chem. Soc.* **2011**, *133*, 2897–2903.
- (47) Mondal, J.; Stirnemann, G.; Berne, B. J. When Does Trimethylamine N-Oxide Fold a Polymer Chain and Urea Unfold It? *J. Phys. Chem. B* **2013**, *117*, 8723–8732.
- (48) Mondal, J.; Halverson, D.; Li, I. T. S.; Stirnemann, G.; Walker, G. C.; Berne, B. J. How Osmolytes Influence Hydrophobic Polymer Conformations: A Unified View from Experiment and Theory. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 9270–9275.

- (49) Martínez, L.; Shimizu, S. Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions. *J. Chem. Theory Comput.* **2017**, *13*, 6358–6372.
- (50) Petris, P. C.; Anogiannakis, S. D.; Tzounis, P.-N.; Theodorou, D. N. Thermodynamic Analysis of N-Hexane–Ethanol Binary Mixtures Using the Kirkwood–Buff Theory. *J. Phys. Chem. B* **2019**, *123*, 247–257.
- (51) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (52) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.
- (53) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. A Revised Potential-Energy Surface for Molecular Mechanics Studies of Carbohydrates. *Carbohydr. Res.* **1988**, *180*, 207–221.
- (54) Reiling, S.; Schlenkrich, M.; Brickmann, J. Force Field Parameters for Carbohydrates. *J. Comput. Chem.* **1996**, *17*, 450–468.
- (55) Holmberg, N.; Ryde, U.; Bülow, L. Redesign of the Coenzyme Specificity in L-Lactate Dehydrogenase from *Bacillus Stearothermophilus* Using Site-Directed Mutagenesis and Media Engineering. *Protein Eng.* **1999**, *12*, 851–856.
- (56) Liu, C. T.; Hanoian, P.; French, J. B.; Pringle, T. H.; Hammes-Schiffer, S.; Benkovic, S. J. Functional Significance of Evolving Protein Sequence in Dihydrofolate Reductase from Bacteria to Humans. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10159–10164.



- (57) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (59) Lide, D. R. *CRC Handbook of Chemistry and Physics, 85th Edition*; CRC Press, Boca Raton, FL, 2004.
- (60) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (61) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (63) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Springer International Publishing: Cham, 2015; Vol. 8759; pp 3–27.
- (64) 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.
- (65) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

- (66) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (67) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (68) 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.
- (69) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (70) Segur, J. B.; Oberstar, H. E. Viscosity of Glycerol and Its Aqueous Solutions. *Ind. Eng. Chem.* **1951**, *43*, 2117–2120.
- (71) Daschakraborty, S. How Do Glycerol and Dimethyl Sulphoxide Affect Local Tetrahedral Structure of Water around a Nonpolar Solute at Low Temperature? Importance of Preferential Interaction. *J. Chem. Phys.* **2018**, *148*, 134501.
- (72) Dubey, V.; Daschakraborty, S. Influence of Glycerol on the Cooling Effect of Pair Hydrophobicity in Water: Relevance to Proteins' Stabilization at Low Temperature. *Phys. Chem. Chem. Phys.* **2019**, *21*, 800–812.
- (73) Blicek, J.; Affouard, F.; Bordat, P.; Lerbret, A.; Descamps, M. Molecular Dynamics Simulations of Glycerol Glass-Forming Liquid. *Chemical Physics* **2005**, *317*, 253–257.
- (74) Akinkunmi, F. O.; Jahn, D. A.; Giovambattista, N. Effects of Temperature on the Thermodynamic and Dynamical Properties of Glycerol-Water Mixtures: A Computer Simulation Study of Three Different Force Fields. *J. Phys. Chem. B* **2015**, *119*, 6250–61.

- (75) Kang, M.; Smith, P. E. Preferential Interaction Parameters in Biological Systems by Kirkwood–Buff Theory and Computer Simulation. *Fluid Phase Equilib.* **2007**, *256*, 14–19.
- (76) Ma, L.; Pegram, L.; Record, M. T.; Cui, Q. Preferential Interactions between Small Solutes and the Protein Backbone: A Computational Analysis. *Biochemistry* **2010**, *49*, 1954–1962.
- (77) Schneider, C. P.; Trout, B. L. Investigation of Cosolute-Protein Preferential Interaction Coefficients: New Insight into the Mechanism by Which Arginine Inhibits Aggregation. *J. Phys. Chem. B* **2009**, *113*, 2050–2058.
- (78) Bhat, R.; Timasheff, S. N. Steric Exclusion Is the Principal Source of the Preferential Hydration of Proteins in the Presence of Polyethylene Glycols. *Protein Sci.* **1992**, *1*, 1133–1143.
- (79) Sirotkin, V. A.; Kuchierskaya, A. A. Preferential Solvation/Hydration of  $\alpha$ -Chymotrypsin in Water-Acetonitrile Mixtures. *J. Phys. Chem. B* **2017**, *121*, 4422–4430.
- (80) Shukla, D.; Trout, B. L. Preferential Interaction Coefficients of Proteins in Aqueous Arginine Solutions and Their Molecular Origins. *J. Phys. Chem. B* **2011**, *115*, 1243–1253.
- (81) Loveridge, E. J.; Evans, R. M.; Allemann, R. K. Solvent Effects on Environmentally Coupled Hydrogen Tunnelling during Catalysis by Dihydrofolate Reductase from *Thermotoga Maritima*. *Chem. Eur. J.* **2008**, *14*, 10782–8.

# TOC Graphic

