Supporting information for

Local Solvation Structures Govern the Mixing Thermodynamics of Glycerol-Water Solutions

Debasish Das Mahanta[1,2], Dennis Robinson Brown[3], Simone Pezzotti[1], Songi Han[3,4], Gerhard Schwaab[1], M. Scott Shell[3,*], Martina Havenith[1,2,*]

[1]Lehrstuhl für Physikalische Chemie II, Ruhr-Universität Bochum, 44780 Bochum, Germany

[2]Department of Physics, Technische Universität Dortmund, 44227 Dortmund, Germany

[3]Department of Chemical Engineering, University of California, Santa Barbara, California 93106-5080, United States

[4]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106-9510, United States

*Corresponding author

Email: shell@ucsb.edu, Martina.Havenith@ruhr-uni-bochum.de

Table of Content

Section 1: *Noise estimation in* the extinction coefficient spectra

The signal contrast is determined by the signal size relative to the signal difference with respect to water. We have estimated the noise level from the principal component analysis (PCA). PCA is a method of multivariate data analysis, provides mutually orthogonal principal components with an order according to their decreasing importance. We have broadly applied PCA to analyse FIR/THz spectra in our previous studies (see ref[1] for more details). In a nutshell, this is a statistical approach to decompose a large multivariate data set of partly correlated variables to a set of uncorrelated variables i.e., principal components through orthogonal transformations. For glycerol-water mixtures we found there exist two most important components which can retrieve almost 99% of the original spectra. The higher order principal components correspond to the statistical noise and small systematic error. We have reconstructed the THz extinction spectra from the two most important principal components (Fig. S1). The residual part of the spectra is calculated by subtracting the reconstructed spectra from the original measured spectra. This residual estimates the noise level. As shown in the right most panel of Fig. S1, the noise level is 100 times lower than the original measured spectra.

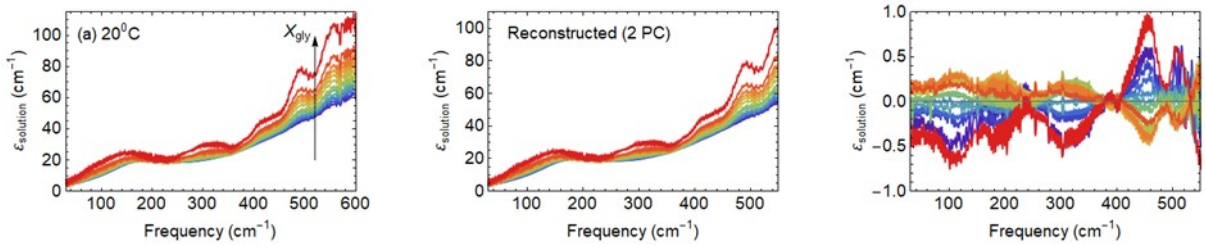

Fig. S1 The left most panel shows the original measured THz extinction spectra at 20^{0}C. We have reconstructed the spectra from the two most important principal components (in the middle panel). The residual of the spectra is calculated by subtracting the reconstructed spectra from the original measured spectra (in the right most panel).

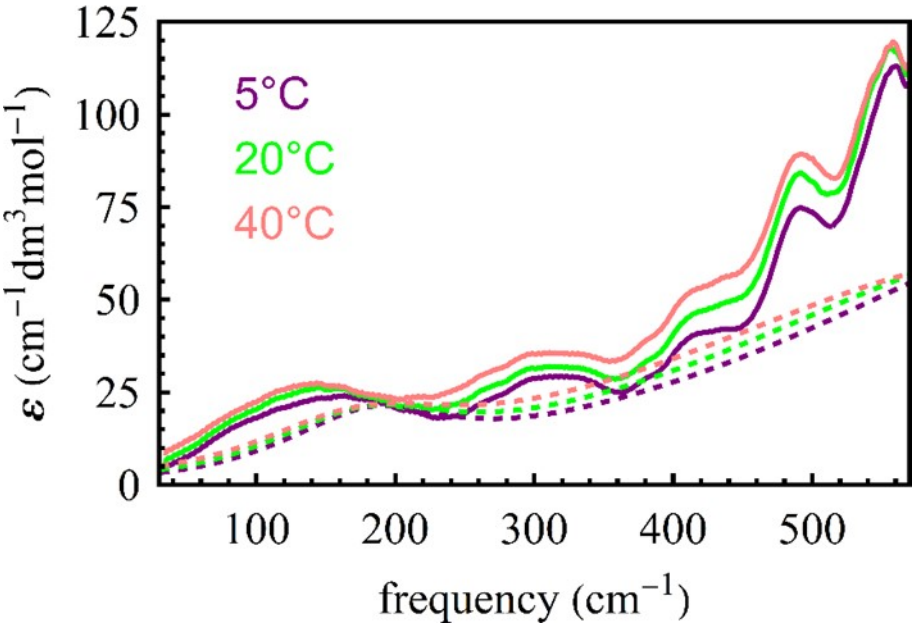

Fig. S2 Average molar extinction coefficient of pure glycerol (solid line) and pure water (dashed line) at three different temperatures.

Section 2: *Theoretical THz spectra and slope*

The THz intensity (I_{THz}) can be directly calculated from MD simulations by[2]

$$I_{THz} = \frac{2\pi\beta}{3cV}\int dt\, e^{i\omega t}\langle \dot{M}(t)\dot{M}(0)\rangle \qquad (S1)$$

where $\beta = \frac{1}{k_B T}$ (k_B is the Boltzmann constant and T is the temperature), c is speed of light and V is the volume of the system. $\langle \dot{M}(t)\dot{M}(0)\rangle$ is a correlation function, with $\dot{M}$ being the time derivative of the total dipole moment of the system. This can be further decomposed into molecular contributions by

$$I_{THz} = \frac{2\pi\beta}{3cV}\int dt\, e^{i\omega t}\sum_{i=1}^{N_{mol}}\sum_{j=1}^{N_{mol}}\langle \dot{\mu}_i(t)\dot{\mu}_j(0)\rangle \qquad (S2)$$

where μ is the dipole of each molecule in the system and the sum over i and j run over all molecules in the system. The contribution to the THz spectrum of bound water molecules can be hence separately calculated from theory by restricting the summation over i to water belonging to the bound population only (N_{bound})

$$I_{THz} = \frac{2\pi\beta}{3cV}\int dt\, e^{i\omega t}\sum_{i=1}^{N_{bound}}\sum_{j=1}^{N_{mol}}\langle \dot{\mu}_i(t)\dot{\mu}_j(0)\rangle = \sum_{i=1}^{N_{bound}} I_{bound} \qquad (S3)$$

where $I_{bound} = \frac{2\pi\beta}{3cV}\int dt\, e^{i\omega t}\sum_{j=1}^{N_{mol}}\langle\dot{\mu}_i(t)\dot{\mu}_j(0)\rangle$ is the THz spectrum of one bound water molecule. By introducing $\bar{I}_{bound}$ as the average spectrum of one bound water molecule, we hence obtain

$$I_{THz} = N_{bound}\bar{I}_{bound} \qquad (S4)$$

The contribution of bound water molecules to the slope can be hence theoretically determined by means of equation S4. To this end, we have calculated the average $\bar{I}_{bound}$ spectrum for a bound water molecule H-bonded to a –OH group by means of equation S3 from a previously performed DFT-MD simulation on a model system where one alcohol (tBuOH) molecule is solvated in water, see references[2, 3] for more details. The $\bar{I}_{bound}$ spectrum is shown in Fig. S3 (N_{bound} =1 spectrum, for which $I_{THz} = \bar{I}_{bound}$). In the same figure, we show the evolution of I_{THz} as a function of the number of bound waters (from N_{bound} = 1 to N_{bound} = 5), as calculated from equation S4. The results show a linear dependence of the slope on N_{bound} (inset of Fig. S3, as expected from equation S4). In the inset, the theoretical slope is calculated by linearly fitting the THz intensity in the 400-500 cm^{-1} range. We carefully checked that the linear relationship between the slope and the number of bound waters holds true when freely varying the extremes of the fit within the frequency range were the intensity rise (i.e., in the 400-600 cm^{-1}).

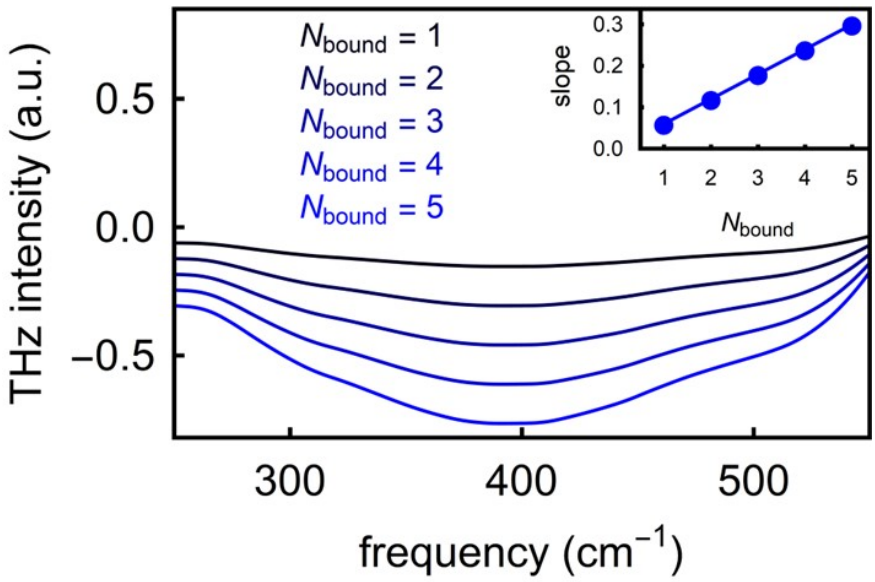

Fig. S3 Theoretical THz spectrum of bound waters as a function of the number of bound water (N_{bound}) molecules. The inset shows that the theoretically calculated slope depends linearly on N_{bound}.

Section 3: *Raman measurements in the OH-stretching frequency range*

In order to corroborate our interpretation (discussed in the main text) of the molecular arrangement of glycerol-water mixtures into three water populations, i.e., bulk, bound and wrap populations, we performed Raman measurements in the mid-IR region, 2000-3550 cm^{-1}, frequency range (Fig. S4). Pure glycerol shows three features in this frequency range: a broad OH-stretching mode around 3350 cm^{-1} and the two CH_2 stretching modes at 2886 cm^{-1} (symmetric) and 2946 cm^{-1} (asymmetric), respectively. We have analysed the spectra of the glycerol-water mixtures with the multivariate curve resolution (MCR-ALS) technique pioneered by Ben Amotz and co-workers.[4, 5]

In our measurement we observed an offset (caused by the glass slide). For further analysis we restricted our frequency range to 2000-3550 cm^{-1} and subtracted the minimum count in this frequency range which was taken as the zero baseline.

In more detail, the following steps were undertaken:

1) For further analysis we implemented an area normalization of the offset subtracted spectra. This procedure led to good agreement between pure solvent and solute measurements taken at the beginning, middle and end of the measurement series.

2) Noise removal or suppression was achieved by performing a singular value decomposition and reconstruction of the spectra based on the most relevant (three) principal components.

3) Iterative MLR-ALS analysis using three components (solvent, solute, and interaction) with the boundary conditions, that all component spectra are positive, and area normalized and that the weights of the pure solute and solvent components are 1. We want to point out that in both the SVD and the MLR-ALS three components are sufficient to describe the spectroscopic changes within experimental uncertainty.

This leads to numerically stable results that are independent of the chosen starting vector for the mixed component. The sum of the three main components as deduced by the MCR-ALS technique are plotted in Fig. S5. Based on the centre frequency and the dependency on the

glycerol mole fraction (see Fig. S5) we assign these to bulk-water, bulk-glycerol, and glycerol-water interactions, respectively.

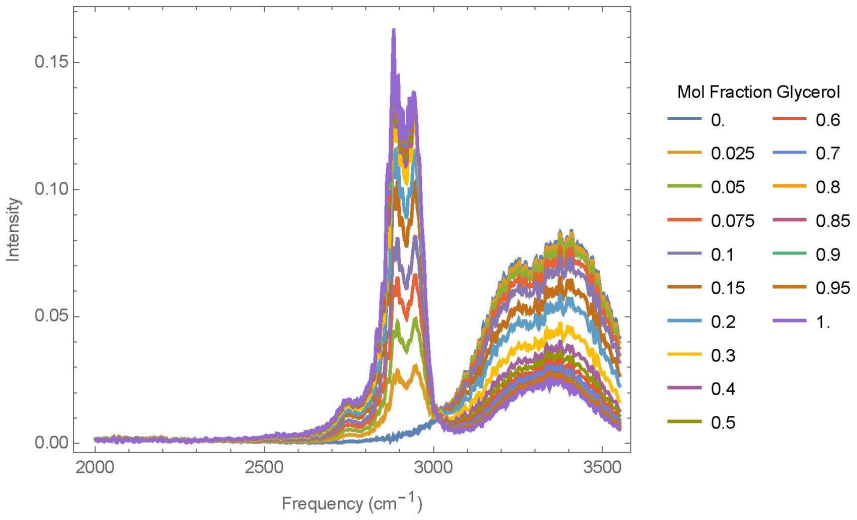

Fig. S4 Raman spectra of glycerol-water mixtures as a function of glycerol mol fraction.

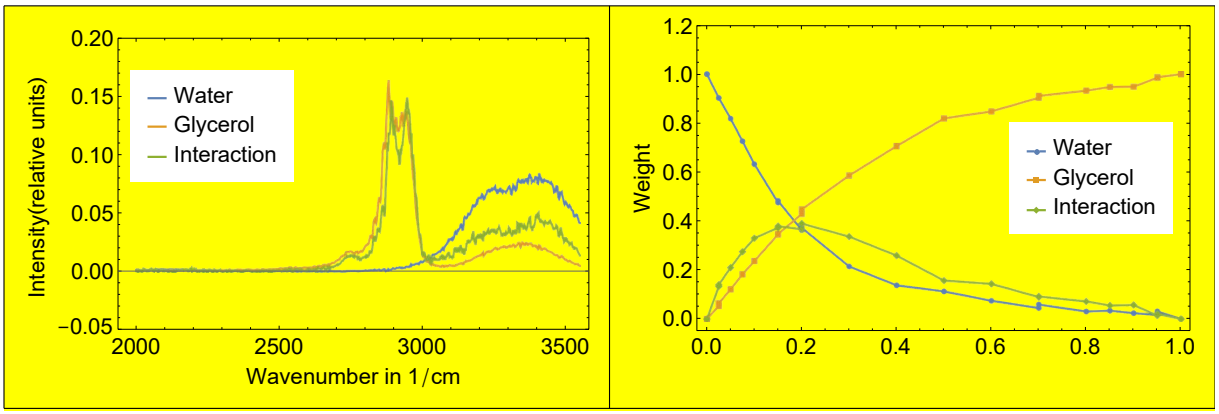

Fig. S5 (a) Three area normalized components of the Raman spectrum of glycerol-water mixtures and (b) relative contribution of each of these three components as a function of X_{gly},.as deduced from MCR-ALS.

In the mid-IR, we are able to extract one additional partial interaction component (green) in addition to the pure solvent (blue) and glycerol (orange) contributions. This interaction component peaks at $X_{gly} = 0.2$ which is in between the maxima reported for the wrap and bound populations, respectively, from MD simulations and also below the maximum at $X_{gly} = 0.4$ as described by the experimental bound water population.

The partial component attributed to bulk water (blue in Fig. S5) reaches zero around X_{gly}=0.4, in agreement with the simulations (Fig. 3c of the main text). By construction, this part is not present in the FIR/THz spectrum – we subtract bulk water. The water-glycerol interaction component (displayed in gray) has a maximum at $X_{gly} = 0.2$ which differs from the maximum

of X_{gly} = 0.4 that we find for bound water based on our THz measurements. We propose that the green "interaction" component is a spectroscopic signature of both, the "bound" water as well as the "wrap water". We found that the wrap and bound contributions as a function of X_{gly} peak at around 0.15 (from theory, Fig. 3c of the main text) and 0.4 (from both theory and THz measurements, Fig. 3), respectively. As a consequence, the maximum of the Raman scattering cross section is intermediate between the two mole fractions, i.e., at X_{gly}=0.25. In the case of glycerol, the "cavity wrap" contribution is almost bulk like and therefore may be not spectroscopically separable.

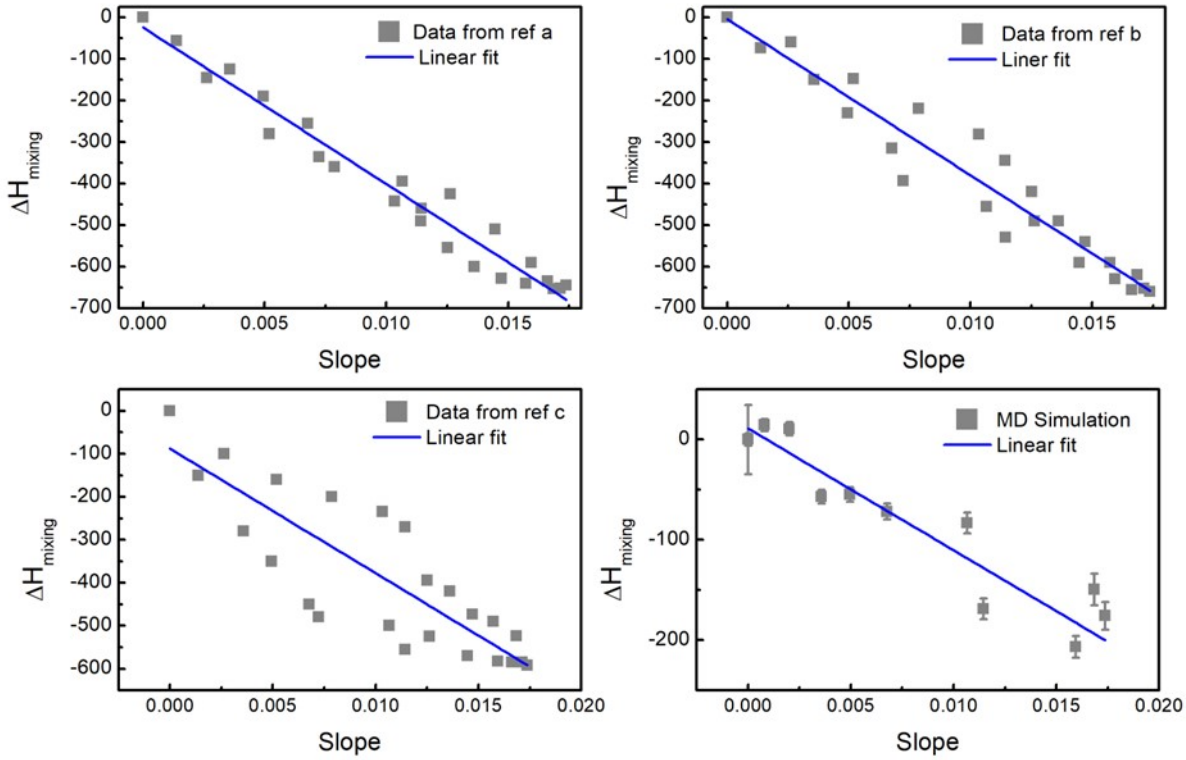

Fig. S6 Linear correlation between ΔH_{mixing} data and the experimentally obtained slope. Ref a, b and c are mentioned in the main text.

Section 4: *Defining the hydration shell*

To motivate our choice of hydration shell, we compute the average local density between glycerol heavy atoms and all water oxygens.

$$\rho_{gly-Ow}(r) = \frac{1}{N_{heavy}} \sum_{i=1}^{N_{heavy}} \sum_{j=1}^{N_{wat}} \delta(\vec{r} - \vec{r}_j) = \frac{1}{N_{heavy}} \sum_{i=1}^{N_{heavy}} \rho_{gly}{}^{i}{}_{-Ow}(r) \qquad (S5)$$

where N_{heavy}, N_{wat}, $\vec{r}_j$, $\delta(\vec{r} - \vec{r}_j)$, and $\rho_{gly}{}^{i}{}_{-Ow}(r)$ are the number of glycerol heavy atoms, the number of water oxygens, the position of the j[th] glycerol heavy atom, the Dirac delta function

centred at $\vec{r}_j$, and the local water density relative to the ith glycerol heavy atom. We obtain the radial distribution function by dividing $\rho_{gly-Ow}(r)$ by the bulk value $\rho_{gly-Ow}^{bulk} = \rho_{gly-Ow}(r \geq 14\text{Å})$. In Fig. S7, we observe a systematic increase in the amplitude of $g_{gly-Ow}(r)$ peaks as X_{gly} increases, indicating a monotonic increase in the local density of water near glycerol. However, the positions of short-range ($r \leq 5$ Å) maxima and minima remain essentially constant for the entire range of concentrations. Therefore, the underlying hydration shell structure remains qualitatively similar for all considered mixtures. To further motivate our chosen cut-off radius, we investigate the impact of the cut-off radius on the characterization of bound and wrap water populations.

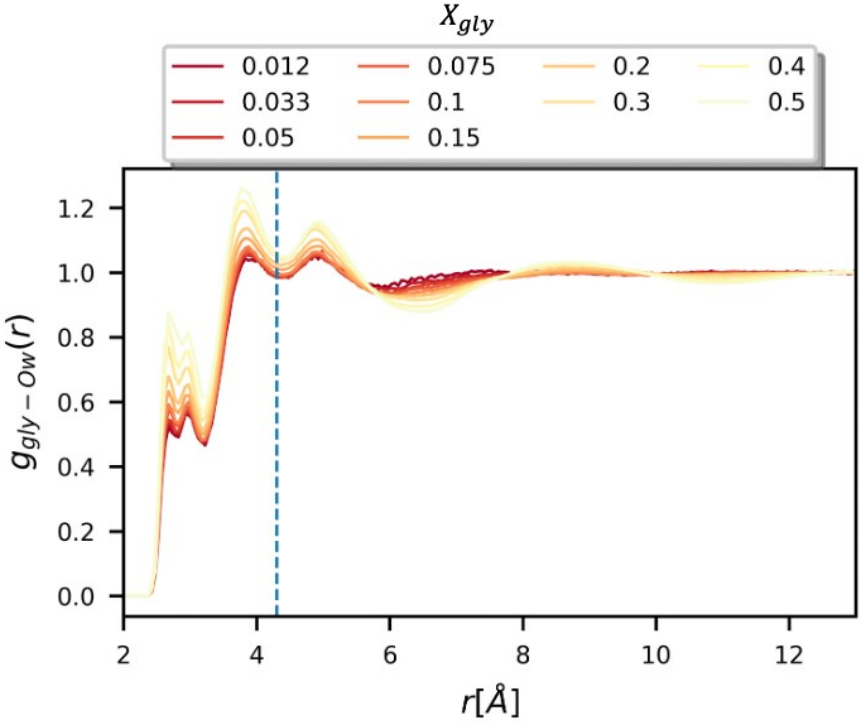

Fig. S7 As the glycerol concentration increases, the amplitude of the peaks increases systematically. For $r < 5$ Å, the location of the minima and maxima remain essentially constant for the entire concentration range from $X_{gly} =$ 0.012 to 0.5. The 4.2 Å cut-off radius mentioned in the text is indicated here with a vertical dashed line.

Given that all bound waters must lie within 3.5 Å of a glycerol hydroxyl group atom, increasing the cut-off radius beyond 3.5 Å does not alter the bound water trend shown in figure 3d. In Fig. S8, we demonstrate that increasing the cut-off radius systematically from 3.5 to 5.0 Å causes an enhancement of the wrap water peak and shifting the peak position to lower X_{gly}. The wrap water behavior is otherwise consistent for all considered cut-off radii.

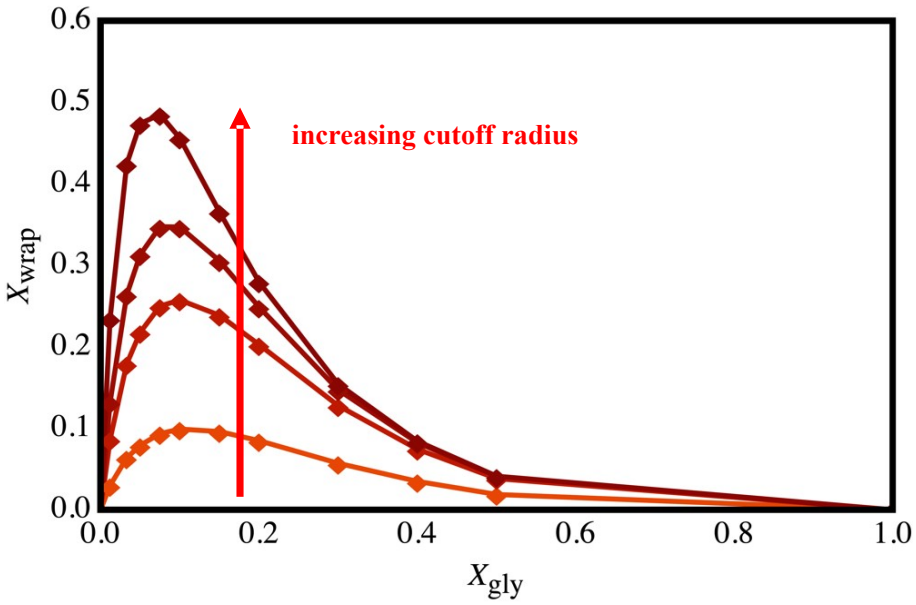

Fig. S8 We observe that a systematic increase in the cut-off radius (3.5, 3.9, 4.2, and 5.0 Å) yields an enhancement in the peak mole fraction of wrap waters while shifting the location of the peak towards lower X_{gly}.

Similarly, the choice of cut-off radius does not qualitatively impact the population of tetrahedrally-coordinated waters p_{tet}. For smaller cutoff radii, the wrap waters reside nearer to the glycerol molecules and hence form fewer three body angles (see *Results and Discussions* for further details). Thus, the wrap waters exhibit lower p_{tet} for a small cutoff radius (e.g., 3.5 Å) compared to a larger cut-off radius (e.g., 5.0 Å) (see Fig. S9).

To understand shifts in the hydration shell of glycerol with changing composition, we choose a cut-off radius that is within approximately the 2nd hydration shell of glycerol ($r \leq 4.2$ Å) while including the hydrogen bond distance criteria ($r \leq 3.5$ Å) mentioned in the *Methods* section of the text. However, choosing a smaller (e.g., 3.5 Å) or larger (e.g., 5.0 Å) does not impact the essential trend in the p_{tet} of the wrap waters.

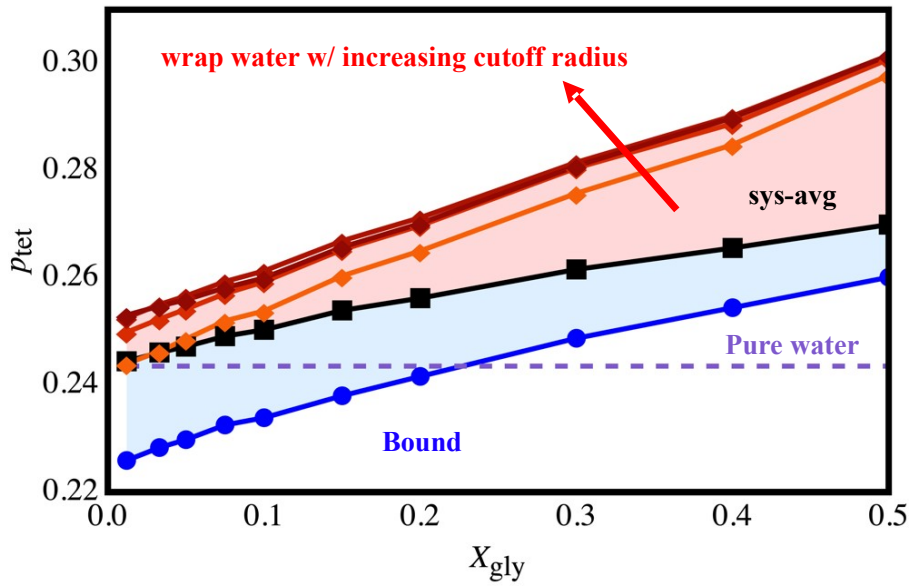

Fig. S9 We observe the same trends in p_{tet} for bound, wrap, and the system-average water as in the text (Fig. 4c). Here, we demonstrate that increasing the cut-off radius from 3.5 to 5.0 Å does not qualitatively impact the p_{tet} of the wrap waters.

Section 5: *Estimating standard error with block averaging*

We implement a block averaging scheme to estimate the standard error of the mean of quantities like the enthalpy of mixing (ΔH_{mixing}), the population of tetrahedrally coordinated waters (p_{tet}), and the mole fraction of bound/wrap waters as a function of glycerol concentration. As an example, we consider the block averaging procedure for the total energy of a box of OPC water molecules at 18°C and 1 bar: (1) split the time series of length N for an observed quantity into blocks of length $\tau < N/2$, (2) compute the mean of each block E_i (where $i = 1, 2,..,N/\tau$), (3) calculate the standard error of the block means s_E^τ, (4) repeat. For smaller block lengths, the individual samples (blocks) are more correlated, yielding lower standard error. We obtain the true s_E^τ (s_E) once the limiting behavior is attained (indicated with a horizontal dashed line in Fig. S10). After this point, s_E^τ becomes increasingly noisy due to a decreasing number of samples. To automate the calculation of s_E, we fit the data to a function

$$s_{E,fit}(\tau) = \sigma\sqrt{\frac{2}{N}}\left[aT_1\left(\frac{T_1\left(e^{-\frac{\tau}{T_1}}-1\right)}{\tau}+1\right)+(1-a)T_2\left(\frac{T_2\left(e^{-\frac{\tau}{T_2}}-1\right)}{\tau}+1\right)\right]^{1/2} \quad (S6)$$

with fit parameters a, T_1, T_2, and σ. Taking the limit as $\tau\to\infty$, we compute s_E

$$s_E = \lim_{\tau\to\infty} s_{E,fit}(\tau) = \sigma\sqrt{\frac{2}{N}}\left[aT_1 + (1-a)T_2\right]^{1/2} \quad (S7)$$

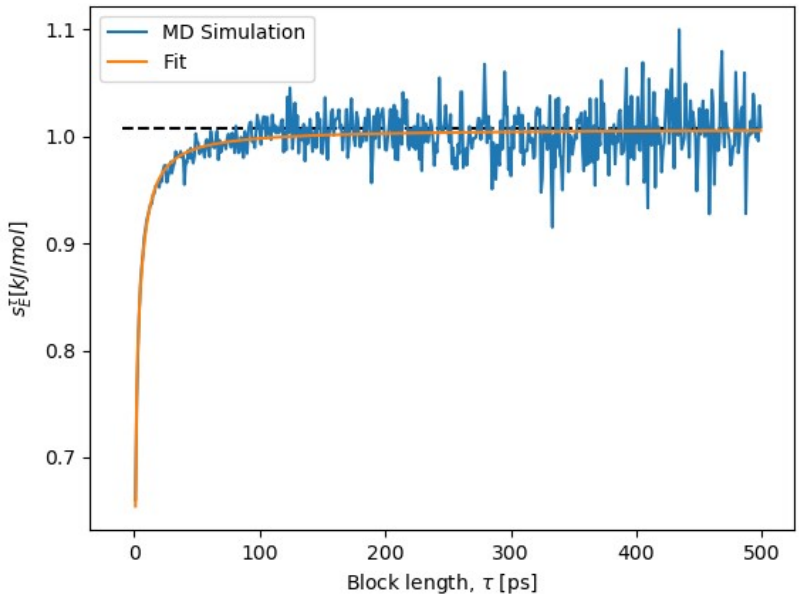

Fig. S10 Here, we provide an illustrative example of the block averaging protocol for estimating the standard error, s_E, of the total energy of the system. The raw data, fit to equation S6, and limiting behavior are represented in blue, orange, and black, respectively.

Section 6: *Hydrogen bonding in glycerol-water*

The peaks in slope at X_{gly}= 0.4 [Fig. 3c] and the mole fraction of bound waters, X_{bound}, at X_{gly}=0.3 [Fig. 3d] originate partly from an increased availability of water-glycerol HBs with increasing glycerol content. Normalizing the number of glycerol-water HBs by the total number of HBs $\frac{N_{HB,\,gly-wat}}{N_{HB,\,Tot}}$, we observe an apparent maximum at equimolar concentration (Fig. S11). However, the location of the peak in $\frac{N_{HB,\,gly-wat}}{N_{HB,\,Tot}}$ differs from the observed maximum in slope

(Fig. 3c) or X_{bound} (Fig. 3d). Instead, we observe a monotonic increase in $\frac{N_{HB,\,gly-wat}}{N_{HB,Tot}}$ and $\frac{N_{HB,\,gly-gly}}{N_{HB,Tot}}$ and simultaneous decrease in $\frac{N_{HB,\,wat-wat}}{N_{HB,Tot}}$ with higher glycerol content (Fig. S11). These trends in hydrogen bond populations yield increasing in X_{bound} for $X_{gly} \leq$ 0.30. At $X_{gly=}$ 0.3, most water molecules are characterized as bound or wrap waters ($X_{boumd} + X_{wrap} = 0.69$); hence, further increasing the concentration of glycerol leads to a simultaneous decrease in X_{bound} and X_{wrap}.

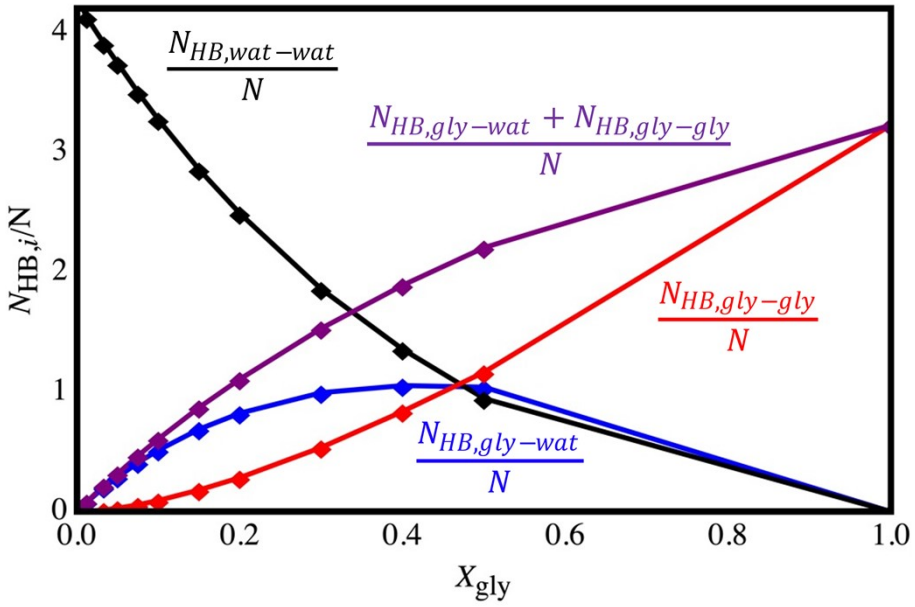

Fig. S11 The number of all three types of hydrogen bond pairs (i) water-water (black), (ii) glycerol-water (blue), and (iii) glycerol-glycerol (red) relative to the total number of molecules $N = N_{wat} + N_{gly}$. In purple, we show the fraction of glycerol-water and glycerol-glycerol hydrogen bonds $(N_{HB,gly-wat} + N_{HB,gly-gly})/N$.

References

1. G. Schwaab, F. Sebastiani and M. Havenith, Ion Hydration and Ion Pairing as Probed by THz Spectroscopy, *Angewandte Chemie International Edition*, 2019, **58**, 3000-3013.
2. V. Conti Nibali, S. Pezzotti, F. Sebastiani, D. Galimberti, G. Schwaab, M. Heyden, M.-P. Gaigeot and M. Havenith, Wrapping up hydrophobic hydration: Locality matters, *J. Phys. Chem. Lett.*, 2020, **11**, 4809-4816.
3. S. Pezzotti, F. Sebastiani, E. P. van Dam, S. Ramos, V. Conti Nibali, G. Schwaab and M. Havenith, Spectroscopic Fingerprints of Cavity Formation and Solute Insertion as a Measure of Hydration Entropic Loss and Enthalpic Gain, *Angewandte Chemie International Edition*, 2022, **61**, e202203893.

4. D. Ben-Amotz, Hydration-Shell Vibrational Spectroscopy, *J. Am. Chem. Soc.*, 2019, **141**, 10569-10580.
5. P. Perera, M. Wyche, Y. Loethen and D. Ben-Amotz, Solute-Induced Perturbations of Solvent-Shell Molecules Observed Using Multivariate Raman Curve Resolution, *Journal of the American Chemical Society*, 2008, **130**, 4576-4577.