Supplementary Information for

Properties of Aqueous Trehalose Mixtures: Glass Transition and Hydrogen Bonding

Gil I. Olgenblum, Liel Sapir, Daniel Harries

S1. Methods

Table S1. Simulation details[a]

#	Water molecules	Trehalose molecules	Molal	Trehalose wt%
1	0	500	-	100
2	481	450	51.97	94.7
3	830	415	27.77	90.5
4	2140	450	11.68	80.0
5	2565	270	5.84	66.7
6	1902	100	2.92	50.0
7	3386	122	2.00	40.6
8	3392	110	1.80	38.1
9	3400	98	1.60	35.4
10	3409	86	1.40	32.4
11	3414	80	1.30	30.8
12	3515	76	1.20	29.1
13	3632	72	1.10	27.4
14	3660	66	1.00	25.5
15	3700	60	0.90	23.6
16	3816	55	0.80	21.5
17	3960	50	0.70	19.3
18	3976	43	0.60	17.0
19	3996	36	0.50	14.6
20	4160	30	0.40	12.1
21	4255	23	0.30	9.3
22	4430	16	0.20	6.4
23	5550	10	0.10	3.3
24	7100	0	0	0

[a] Mixtures vary from pure water ($0\ wt\%$) to anhydrous trehalose ($100\ wt\%$) in order to examine and compare mixture properties at different states. After generation, each box was energetically minimized for $\sim 20000-30000$ steps. Equilibration followed with short MD runs of $100\,ps$ for diluted and $200\,ps$ for concentrated mixtures with a time step of 1 to $2\,fs$. The concentrations of $94.7\ wt\%$ was specifically chosen to test the model against available experimental osmotic pressure data that is given as relative hydration by Simperler et al[1] (achieved by equilibration against saturated salt solution of $MgCl_2$). Furthermore, $100\ wt\%$ and $90.5\ wt\%$ enable us to probe both trehalose anhydrous and dihydrate states, respectively.

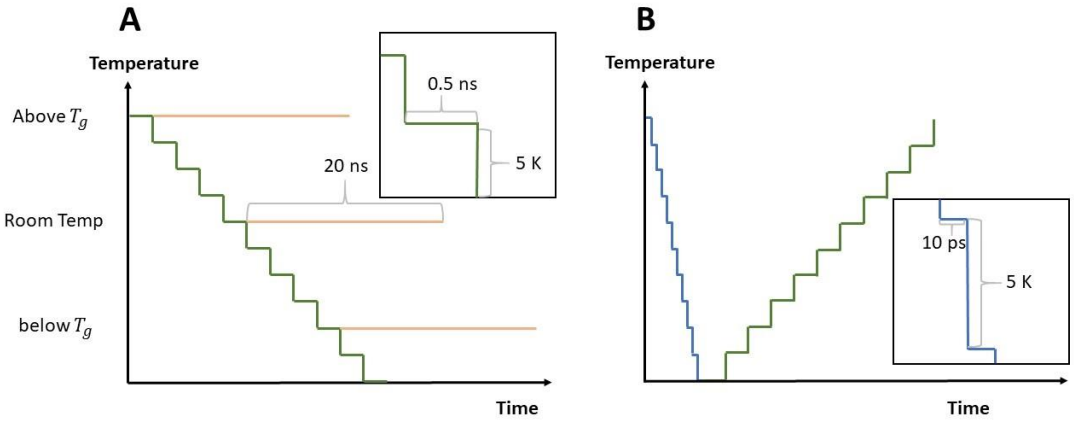

Figure S1- Schematic simulation of cooling and heating protocols. **(A)** Cooling protocol: each step lasts $0.5ns$, while the change in temperature is $5K$ ($10K/ns$) as shown in green. At specific temperatures i.e. above T_g, below T_g and room temperature, as specified in Table S2, the final trajectory was extended for longer isothermic simulations of $20ns$ as shown in orange for hydrogen bonding analysis. By repeating the cooling cycle and extension of specific trajectories for longer isothermic simulations 5 times, the calculated thermodynamical quantities (radial distribution functions, osmotic pressure and hydrogen bonding free energies) were averaged. **(B)** Heating protocol: the system was first cooled as described in (A) only with shorter steps of $10ps$ to reach $100K$. Then, a heating protocol with $0.5ns$ segments and $5K$ temperature jumps was performed to $600K$.

Table S2. Simulation temperatures for long production runs of $20ns$ [a]

Temperature \ Trehalose %wt	100	94.7	90.5	80	66.7	50	30.8	0
Below T_g	271	221	181	141	91	91	91	131
Above T_g	489	439	399	359	307	307	307	366

[a] Table S2 Summarizes of the temperature ranges at which longer simulations were conducted for low water content mixtures. The longer runs yield better statistics required for probing structural and thermodynamical properties. Runs of $20ns$ at room temperature were used to calculate the presented radial distribution functions and, by extension, the Kirkwood-Buff integrals (see Figures 1 and 3 in the main text). All the mixtures below 40.6 wt% at 298 K were simulated for 50 ns, of which the last 30 were used for analysis. Additionally, at 100, 90.5, 30.8 and 0 $wt\%$ the entire temperature range was used to calculate the hydrogen bonding interaction and its temperature dependency.

S2. Structure factor and spatial distribution function

The structure factor is calculated from MD simulations using the relation:

$$S(Q) = 1 + \frac{4\pi}{Q}\int_0^{\infty}\left[\rho^e(r) - \rho_0^e\right] r \sin(Qr)\, dr$$

where $\rho^e(r)$ is the electron density at a distance r, ρ_0^e is the bulk electron density and Q is the magnitude of the scattering vector.

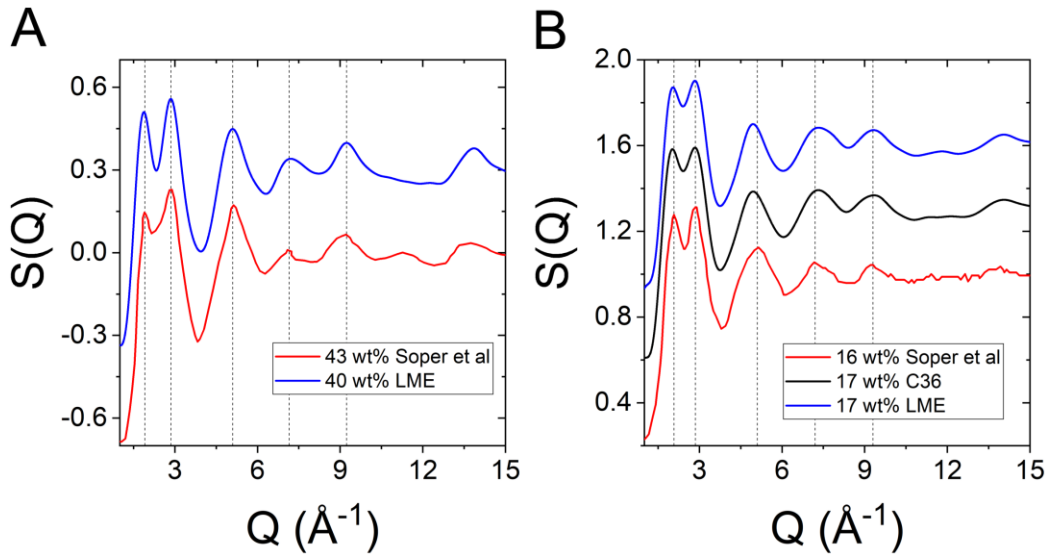

Figure S2 – Structure factor $S(Q)$ calculated from simulation and compared with those experimentally derived by Soper et al[2]. **(A)** $S(Q)$ for concentrated liquid mixture of $43\ wt\%$ corresponding to $2.22\ molal$, and a simulated mixture of $40\ wt\%$ corresponding to $2\ molal$. **(B)** $S(Q)$ for both experimental and simulated mixtures in the diluted regime with the experimental $16\ wt\%$ corresponding to $0.55\ molal$ and a simulated mixture $17\ wt\%$ corresponding to $0.6\ molal$. Data sets are shifted for clarity.

Panel A shows that the $S(Q)$ peak positions and relative heights are in close agreement between the simulation and experiment for concentrated mixtures. In panel B, the first peak in the experimental $S(Q)$ at $\sim 2.1 Å^{-1}$ is slightly lower than the second peak at $\sim 2.9 Å^{-1}$. While C36's first two peaks are of the same height, LME succeeds to reproduce the height difference. Additionally for LME, the third peak at $\sim 5.1 Å^{-1}$ is slightly higher than the fourth and the fifth peaks, which is in better agreement with the experimental $S(Q)$, compared to C36 where the third peak is lower than the other two peaks.

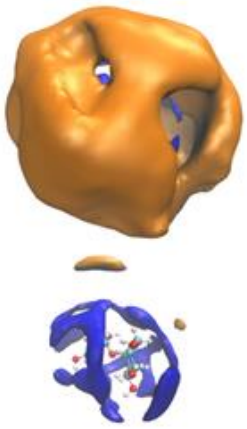

Figure S3 – Same as Fig. 2B and 2F with the same trehalose isovalue of $0.7nm^{-3}$ for both C36 (top) an LME (bottom) force fields. The sugar aggregative nature of C36 is observed clearly in the trehalose probability density surface that completely coats the reference sugar while it is hardly seen for LME.

S3. Kirkwood-Buff Integrals

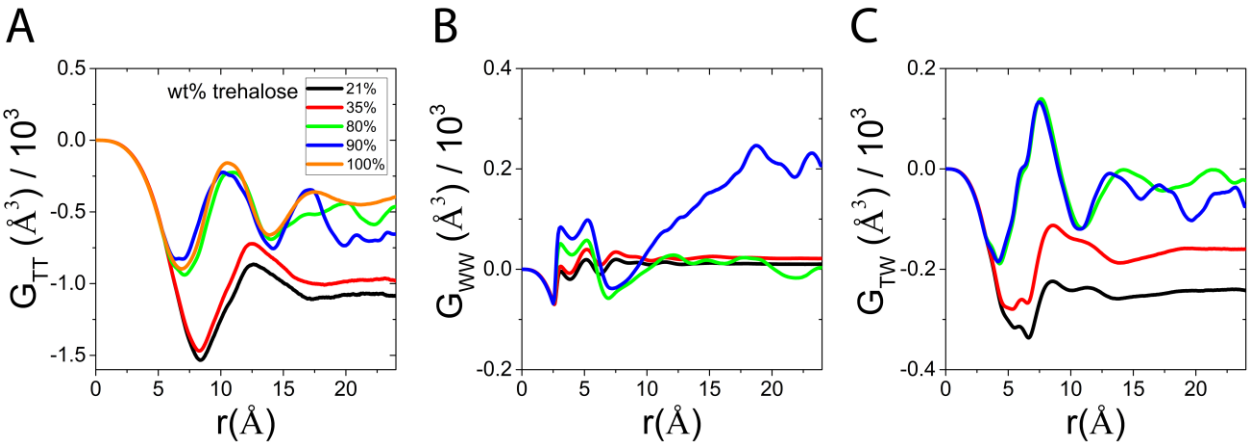

Figure S4. Example for the convergence of KBIs in several systems, simulated with the LME force field, as a function distance. The value at the bulk was taken in each case as the relevant average value of KBI over the range described in table S3. To improve the KBI's convergence, the $g_{ij}(r)$ were averaged between 5 simulations each lasting $20ns$, thus ensuring that $g_{ij}(r_{bulk}) \to 1$ for the correct bulk density.

Table S3. Bulk range of the KBI values[a]

Trehalose wt%	Bulk range (Å)
21	20-25
35	20-25
80	23-26
90	21-24
100	26-28

[a] The range of the bulk taken in the KBI calculation for the concentrations in figure S4.

The experimental KBI were derived through the Kirkwood-Buff inversion relations,[3,4]

$$G_{TW} = \frac{\bar{v}_T}{x_W\left(x_W\bar{v}_W + x_T\bar{v}_T\right)\dfrac{\partial\left(\Pi/RT\right)}{\partial x_W}} \tag{S1}$$

$$G_{WW} = G_{TW} - \frac{1}{x_W}\left(\frac{\bar{v}_T}{x_W\bar{v}_W\dfrac{\partial\left(\Pi/RT\right)}{\partial x_W}} + x_W\bar{v}_W + x_T\bar{v}_T\right) \tag{S2}$$

$$G_{TT} = G_{TW} - \frac{1}{x_T}\left(\frac{1}{x_W\dfrac{\partial\left(\Pi/RT\right)}{\partial x_W}} + x_W\bar{v}_W + x_T\bar{v}_T\right) \tag{S3}$$

where x_W and x_T are the mole fractions of water and trehalose, respectively, and $\bar{v}_W$ and $\bar{v}_T$ are their molar volumes. For the osmotic pressure derivative we used previously published osmotic pressure measurements.[5,6] The partial molar volumes were derived from the experimental densities[7] (see Fig. S5).

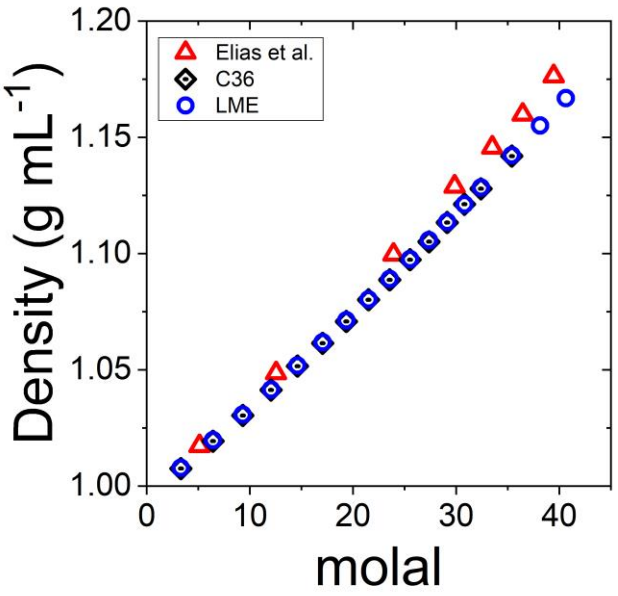

Figure S5. Comparison of experimental density measurements by Elias et al[7] with the values derive from the two examined force fields.

S4. Heat capacity

Table S4. Average temperature and heat capacity change for the glass transition[a]

	LME			C36		
Trehalose %wt	$\Delta T_{reg}\,(K)$	$\Delta C_p\left(kJ\ kg^{-1}K^{-1}\right)$	$T_{g,C_p}\,(K)$	$\Delta T_{reg}\,(K)$	$\Delta C_p\left(KJKg^{-1}K^{-1}\right)$	$T_{g,C_p}\,(K)$
100	159	0.58	390.4	188	0.68	-
94.7	203	0.74	350.2	186	0.76	384.9
90.5	167	0.83	340.5	171	0.79	-
80.0	145	1.00	296.9	160	0.99	298.8
66.7	106	1.17	264.9	121	1.06	254.7
30.8	71	1.73	214.8	43	1.51	212.8
0	17.2	1.77	211	-	-	-

[a] Comparison of the glass transition region of the two force fields. Three variables are compared: change in heat capacity along the transition, ΔC_p, the spread of the glass transition region, ΔT_{reg} and the glass transition temperature derived from heat capacity (section 3.3 in the main text). $0\ wt\%$ is located in LME column although it is the same in both force-fields as the WW interaction are the same.

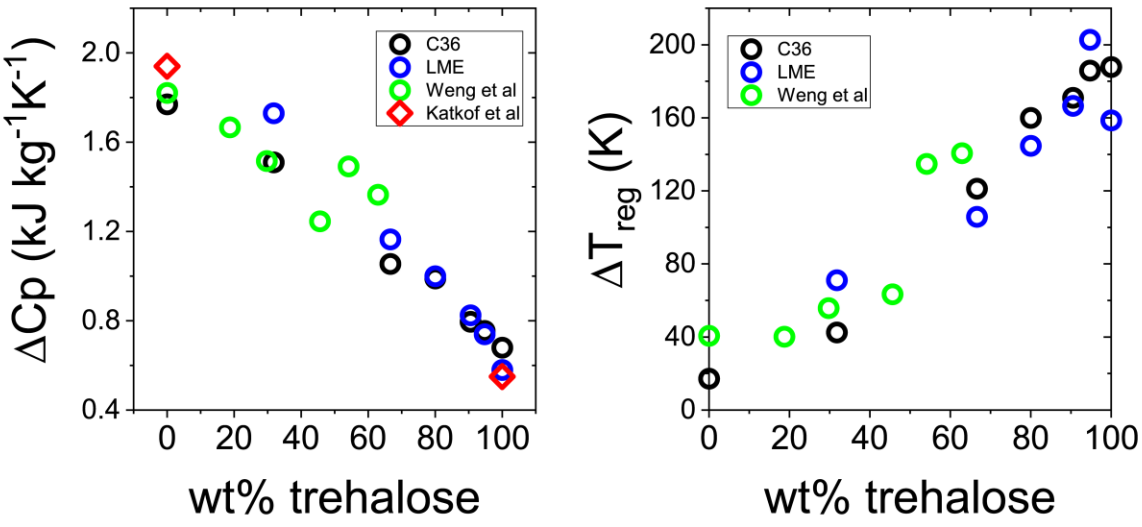

Figure S6. Same data as in Table S4, with added simulation and experimental derived data from Weng et al[8] and Katkov et al [9] respectively, shown for comparison. Both $0\ wt\%$ and $100\ wt\%$ are in good agreement with experimental (Red).

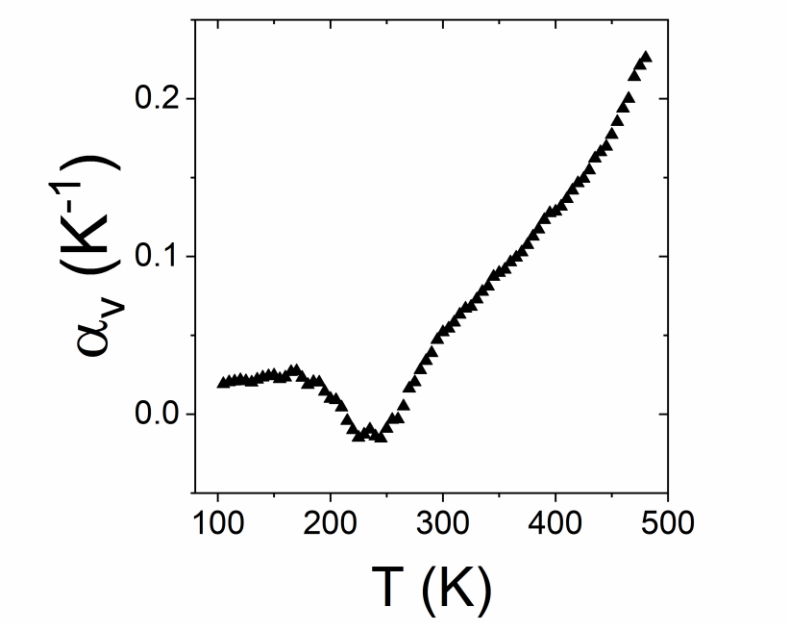

Figure S7. Thermal expansion coefficient (α_v) of $30.8\ wt\%$ from the cooling cycle. A kink appears at $\sim 170K$ to $\sim 280K$, which prevents the determination of T_g by this α_v in diluted mixtures.

S5. Hydrogen Bonding

For the calculation of Hbond strength we followed our recent proposed methodology,[10] which is based on an information-theoretic approach.[11] In short, we consider the probability distribution, $P(r,\theta)$, of donor-acceptor configurations with respect to two quantities: the donor-acceptor distance, r, and the hydrogen-donor-acceptor angle, θ. The probability distribution can be transformed to a potential of mean force for Hbond formation:

$$PMF_{\mathrm{HB}}(r,\theta) = -RT \ln \frac{P(r,\theta)}{P_{\mathrm{rand}}(r,\theta)} \tag{S4}$$

where $P_{\mathrm{rand}}(r,\theta)$ is the random distributions, which can be calculated (analytically or numerically) and depend on the type of donor-acceptor pair. Notice that in Eq. (S4) we assume the distribution functions reach bulk values (i.e. r is large enough). If the (r,θ) distributions are truncated at shorter distances, one can simply use the radial distribution function, $g(r)$, and multiply $P(r)$ by the average of $g(r)P_{\mathrm{rand}}(r)/P(r)$, where $P_{\mathrm{rand}}(r)$ is calculated in the full range. Figure S9 shows a typical contour maps of $P(r,\theta)$, $PMF_{HB}(r,\theta)$, and $\mathrm{P}(r,\theta)\cdot PMF_{HB}(r,\theta)$.

The Hbond strength can be defined through the free energy quantified by the integral:

$$\Delta G = -RT \int_{r_{\min}}^{r_{\max}} \int_{0}^{\pi} \zeta P(r,\theta) \ln \frac{\zeta P(r,\theta)}{\zeta_{\mathrm{rand}} P_{\mathrm{rand}}(r,\theta)} d\theta dr \tag{S5}$$

where we define the normalization constants $1/\zeta = \int_{r_{\min}}^{r_{\max}} \int_{0}^{\pi} P(r,\theta) d\theta dr$ and similarly $1/\zeta_{\mathrm{rand}} = \int_{r_{\min}}^{r_{\max}} \int_{0}^{\pi} P_{\mathrm{rand}}(r,\theta) d\theta dr$. The integration boundaries are chosen to cover the relevant range for hydrogen bonding (table S5).

In this study we use a modified definition for the free energy, ΔG:

$$\Delta G = -RT \int_{r_{\min}}^{r_{\max}} \int_{0}^{\pi} \zeta_{\mathrm{rand}} P_{\mathrm{rand}}(r,\theta) \ln \frac{\zeta_{\mathrm{rand}} P_{\mathrm{rand}}(r,\theta)}{\zeta P(r,\theta)} d\theta dr \tag{S6}$$

which describes the free energy change upon transition from the random distribution to the hydrogen bonded distribution. With this definition, the free energy can be well understood through the dissection to entropic and energetic contributions. Consider the probability distribution of a system:

$$P(n) = \frac{g_n e^{-\beta E_n}}{Q} \tag{S7}$$

where g_n and E_n are the degeneracy and energy of state n, respectively, Q is the partition function, and $\beta = (kT)^{-1}$, where T is the temperature and k is the Boltzmann constant. A similar expression can be written for any configuration, say the random distribution: $P_{\mathrm{rand}}(n) = Q_{\mathrm{rand}}^{-1} g_{n,\mathrm{rand}} e^{-\beta E_{n,\mathrm{rand}}}$. Then, assuming $g_n = g_{n,\mathrm{rand}}$ and that $E_{n,\mathrm{rand}}$ is constant for each n, it follows from Eq. (S6) that the energetic

and entropic contributions to ΔG are:

$$\Delta E = \langle E \rangle - \sum_n \frac{g_n e^{-\beta E_{n,\text{rand}}}}{Q_{\text{rand}}} E_n \tag{S8}$$

$$\Delta S = \left(S - S_{\text{rand}} \right) \tag{S9}$$

where S and S_{rand} are the entropy of the equilibrium and random states, respectively. From, Eq. (S8) the energetic component can be understood as the energetic change associated with redistributing the ensemble from the random distribution to the equilibrium distribution, over the same (equilibrium) energy landscape.

The motivation to use Eq. (S6) rather than Eq. (S5) in describing the free energy change for the configuration ensemble can be rationalized using a simple example. Consider a two-level system with energies $\varepsilon_1, \varepsilon_2$ and corresponding degeneracies g_1, g_2 coupled to a heat reservoir at temperature T. The exact (analytic) free energy difference calculated from the partition functions is:

$$\Delta G = -kT \ln \frac{g_1 e^{-\beta\varepsilon_1} + g_2 e^{-\beta\varepsilon_2}}{g_1 + g_2} \tag{S10}$$

whereas Eq. (S5) gives:

$$\Delta G = -kT \ln\left(\frac{g_1 + g_2}{g_1 e^{-\beta\varepsilon_1} + g_2 e^{-\beta\varepsilon_2}} \right) + \frac{g_1 e^{-\beta\varepsilon_1}}{g_1 e^{-\beta\varepsilon_1} + g_2 e^{-\beta\varepsilon_2}} \varepsilon_1 + \frac{g_2 e^{-\beta\varepsilon_2}}{g_1 e^{-\beta\varepsilon_1} + g_2 e^{-\beta\varepsilon_2}} \varepsilon_2 \tag{S11}$$

and Eq. (S6) yields:

$$\Delta G = -kT \ln\left(\frac{g_1 e^{-\beta\varepsilon_1} + g_2 e^{-\beta\varepsilon_2}}{g_1 + g_2} \right) - \frac{g_1}{g_1 + g_2} \varepsilon_1 - \frac{g_2}{g_1 + g_2} \varepsilon_2 \tag{S12}$$

The corresponding entropic and energetic components are shown in plot S7 for the specific case of $g_1 = 1, g_2 = 2$ and $\varepsilon_1 = 0, \varepsilon_2 = \varepsilon$. It is clear that Eq. (S6) is identical to the exact result, apart for a constant shift in the energy scale, which is related to the state taken as a reference. Specifically, the reference of the energy in Eq. (S6) is taken as the value at infinite temperature, while the reference of the energy derived from the partition function is for T approaching zero.

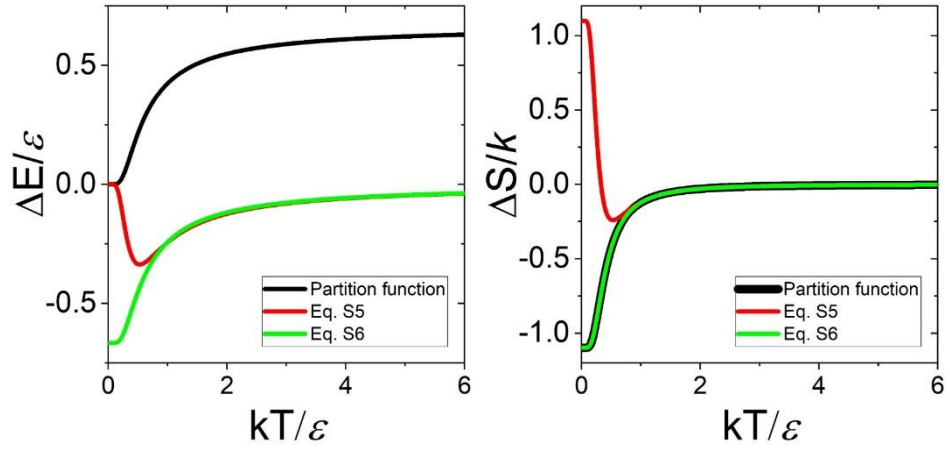

Figure S8. Comparing the different equations for the thermodynamic properties of a two-level system. The energy (left) and entropy (right) are shown as a function of temperature. The approximation of Eq. (S6) is identical to the exact result derived from the partition function, apart for a constant shift in the reference energy value.

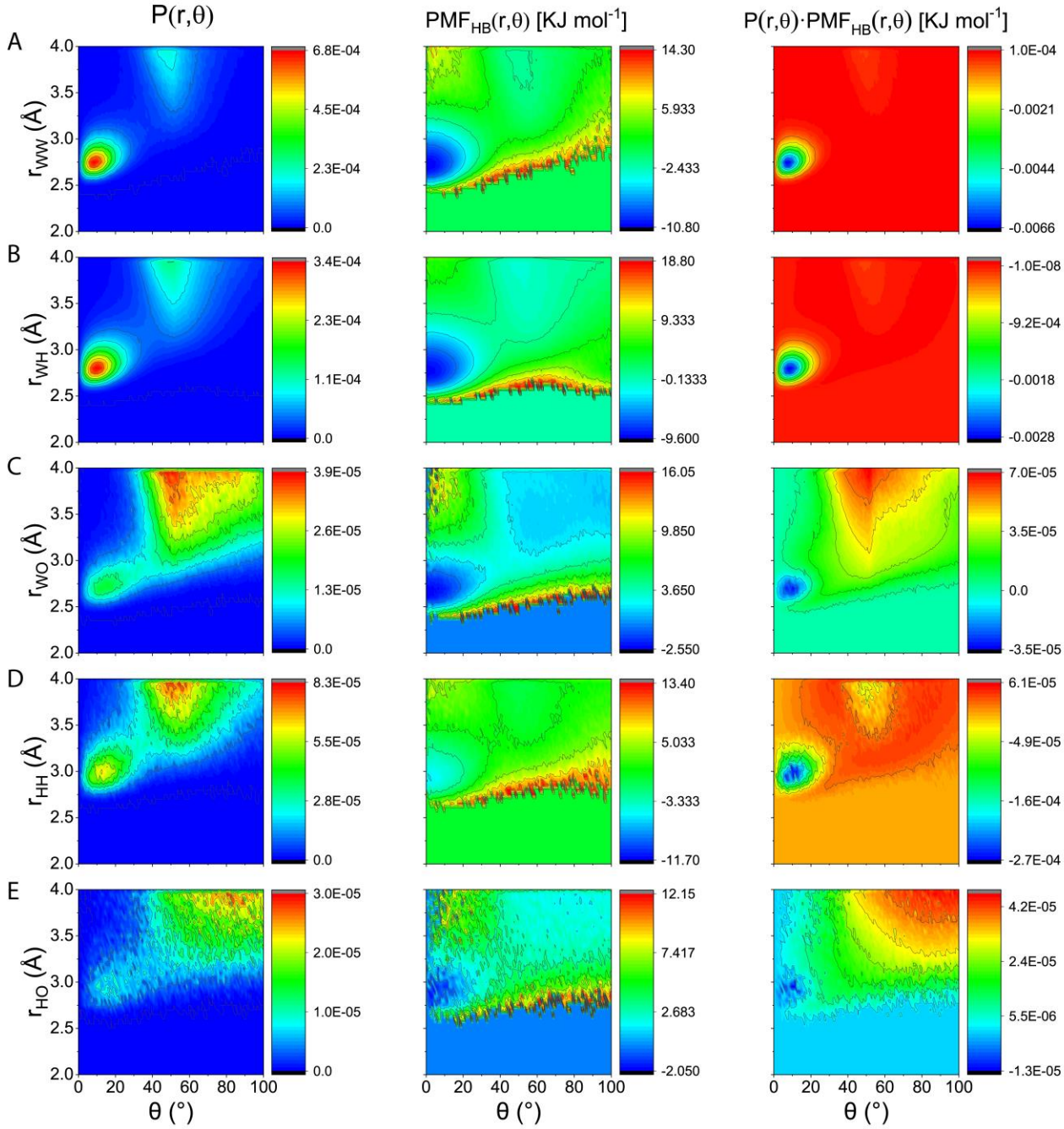

Figure S9. Hydrogen bond analysis contour maps of $30.8\ wt\%$ ($1.3\ molal$) mixture of trehalose-water at $298K$ for five different types of bonds by rows: **(A)** water-water (WW) **(B)** water-sugar hydroxyl (WH) **(C)** water-sugar etheric oxygen **(D)** sugar hydroxyl-sugar hydroxyl (HH) **(E)** sugar hydroxyl- sugar etheric oxygen (HO). The plots are ordered from left to right in columns: **(left)** Probability density function **(center)** potential of mean force (PMF_{HB}) for hydrogen bonding formation relative to the random distribution **(right)** weighted potential of mean force.

Table S5. Integration range for each type of Hbond[a]

Hbond type	Integration range (Å)
WW	2.45 –3.15
WH	2.5 –3.2
WO	2.45 –3.2
HH	2.65 –3.35
HO	2.65 –3.2

[a] The integration range corresponds to the first hydration shell. The lower bound, $r_{\min}$, is taken as the minimum pair distance where $PMF_{HB}(r,\theta \to 0) = 0$. The upper bound, $r_{\max}$, was taken as the saddle point in $PMF_{\mathrm{HB}}(r,\theta)$.

In addition to the integration range, we restricted $r_{\min}$ in the integration of Eq.(6) by different cutoffs on $g(r_{\min})$ to exclude simulation artifacts. Plotted in Figure S10 are two examples for the Hbond free energy for different values of this cutoff. The maximal cutoff was 0.4 and the minimal was 0.001. In panel A, the minimum in the free energy of pure water visibly shifts from $\sim 270K$ to $\sim 197K$ as the cutoff is decreased. The same trend can be seen for water-water interactions in diluted solution of $30.8\ wt\%$ in panel B, with a minimum shift from $\sim 250K$ to $\sim 206K$. For further analysis, the free energy in each temperature was extrapolated from $g(r)_{cutoff} = 0.04$ to $g(r)_{cutoff} \to 0$ using a polynomial fit (either linear, quadratic or cubic), panel C. The extrapolation result is seen in black circles in both panels.

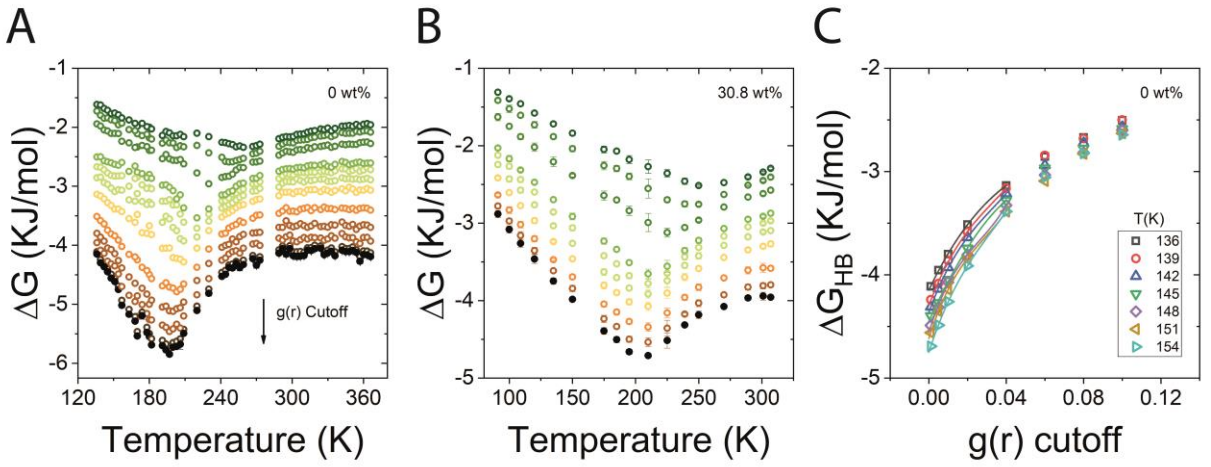

Figure S10. Free energy extrapolation to $g(r) = 0$. **(A)** $0\ wt\%$, water–water interaction (WW). **(B)** $30.8\ wt\%$ water–water interaction. The $g(r)$ cutoff decreases when shifting from green to brown (empty symbols) as follows: $0.001,\ 0.005,\ 0.01,\ 0.02,\ 0.04,\ 0.06,\ 0.08,\ 0.1,\ 0.2,\ 0.3,\ 0.4$. The extrapolated values are in black full symbols. **(C)** Extrapolation of specific temperatures in $0\ wt\%$ using a cubic polynomial fit, temperatures are indicated in the legend.

All the free energies, PMFs and information entropies reported in this work are extrapolated in the same manner and were fitted using a Padé approximant[12] of the order $[3/3]$,

$$\Delta G_{Padé} = \frac{a_0 + a_1 T + a_2 T^2 + a_3 T^3}{1 + b_1 T + b_2 T^2 + b_3 T^3} \tag{S13}$$

To compare the calculated hydrogen bonding free energy to other related thermodynamical properties, we also calculated the information entropy, defined as

$$\Delta S_{\text{Info}} = -R\left(\sum_n P \ln \frac{P}{g_n} - \sum_n P_{\text{rand}} \ln \frac{P_{\text{rand}}}{g_n} \right) \tag{S14}$$

where g_n is the degeneracy of the n-th state, calculated from P_{rand}.

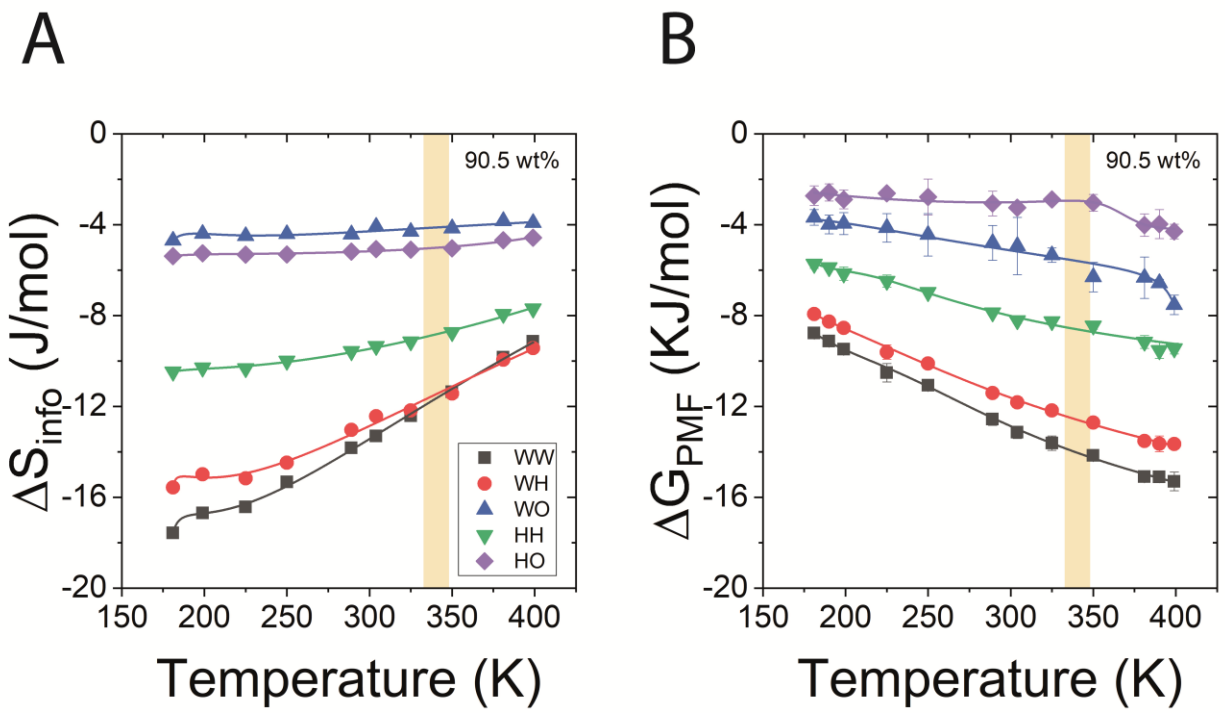

Figure S11. **(A)** Information entropy vs temperature calculated based on eq. (S14). **(B)** Potential of mean force (PMF) for all interaction pairs, eq. (S4). Both (A) & (B) are for concentrations of 90.5 wt%. A Padé approximant was fitted to both ΔS_{Info} and ΔG_{PMF}. The shaded bar represents the range of T_{g,C_p}, see main text.

The Hbond entropy is simply calculated from the derivative of the Hbond free energy:

$$\Delta S_{HB} = -\frac{\partial \Delta G_{HB}}{\partial T} \tag{S15}$$

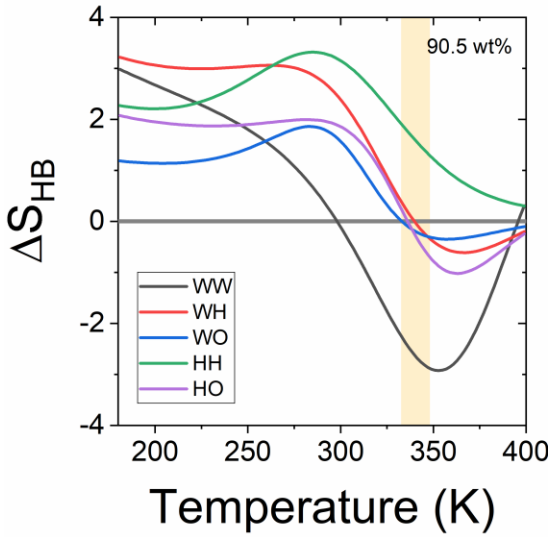

Figure S12. The entropy calculated from the Hbond interaction, as function of temperature, for 90.5 wt% trehalose. As the temperature increases the entropy decreses untill it is very close to zero, or becomes negative. The change in sign of ΔS_{HB} is highly correlated with T_g which is represented with a shaded bar.

Table S6. Hbond thermodynamic potentials at 298K[a]

Thermodynamic Potential	Hbond type / Trehalose wt%	HO	HH	WO	WH	WW
$\Delta G_{HB}\left(KJ\ mol^{-1}\right)$	0	-	-	-	-	-4.12
	30.8	-1.1	-2.19	-1.92	-4.29	-3.96
	90.5	-2.02	-3.53	-1.47	-4.26	-4.17
	100	-2.25	-4.16	-	-	-
$T\Delta S_{HB}\left(KJ\ mol^{-1}\right)$	0	-	-	-	-	-0.34
	30.8	0.86	1.65	2.7	0.68	-0.26
	90.5	1.89	3.19	1.63	2.49	0.01
	100	1.55	2.6	-	-	-
$\Delta H_{HB}\left(KJ\ mol^{-1}\right)$	0	-	-	-	-	-4.46
	30.8	-0.23	-0.54	0.77	-3.60	-4.21
	90.5	-0.13	-0.34	0.17	-1.78	-4.18
	100	-0.70	-1.52	-	-	-

[a] Comparison of the different thermodynamic quantities calculated from the fitted free energy, for different concentrations at $298K$. In liquid solution ($30.8\ wt\%$), the addition of more trehalose weakens the Hbond interaction from -4.12 in pure water to $-3.96KJ\ mol^{-1}$. This weakening is mainly enthalpy driven as ΔH_{HB} increases from $-4.46KJ\ mol^{-1}$ to $-4.21KJ\ mol^{-1}$ with sugar addition. In the glassy matrix, WW interaction are strengthened, even compared to pure water ($-4.17\,KJ\ mol^{-1}$). For WH enthalpy-entropy compensation keeps the Hbond free energy in liquid and glass quite similar, while WO interaction is weaker in the glass. Both HH and HO strengthen with trehalose content, in liquid mixture and glass, as seen in Figure 6 of the main text.

References

1. Simperler, A. *et al.* The glass transition temperatures of amorphous trehalose-water mixtures and the mobility of water: an experimental and in silico study. *Carbohydr. Res.* **342**, 1470–1479 (2007).

2. Soper, A. K., Ricci, M. A., Bruni, F., Rhys, N. H. & McLain, S. E. Trehalose in Water Revisited. *J. Phys. Chem. B* **122**, 7365–7374 (2018).

3. Ben-Naim, A. Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system. *J. Chem. Phys.* **67**, 4884–4890 (1977).

4. Matteoli, E. A study on Kirkwood-Buff integrals and preferential solvation in mixtures with small deviations from ideality and/or with size mismatch of components. Importance of a proper reference system. *J. Phys. Chem. B* **101**, 9800–9810 (1997).

5. Sapir, L. & Harries, D. Linking Trehalose Self-Association with Binary Aqueous Solution Equation of State. 624–634 (2011) doi:10.1021/jp109780n.

6. Poplinger, M., Shumilin, I. & Harries, D. Impact of trehalose on the activity of sodium and potassium chloride in aqueous solutions: Why trehalose is worth its salt. *Food Chem.* **237**, 1209–1215 (2017).

7. Elias, M. E. & Elias, A. M. Trehalose plus water fragile system: properties and glass transition. *J. Mol. Liq.* **83**, 303–310 (1999).

8. Weng, L. & Elliott, G. D. Dynamic and thermodynamic characteristics associated with the glass transition of amorphous trehalose-water mixtures. *Phys. Chem. Chem. Phys.* **16**, 11555–65 (2014).

9. Katkov, I. I. & Levine, F. Prediction of the glass transition temperature of water solutions: Comparison of different models. *Cryobiology* **49**, 62–82 (2004).

10. Sapir, L. & Harries, D. Revisiting Hydrogen Bond Thermodynamics in Molecular Simulations. *J. Chem. Theory Comput.* **13**, 2851–2857 (2017).

11. Procaccia, I. & Levine, R. D. Potential work: A statistical-mechanical approach for systems in disequilibrium. *J. Chem. Phys.* **65**, 3357–3364 (1976).

12. William H.Press, Saul A.Teukolsky, William T.Vetterling, B. P. F. Pade_Approx. in *Numerical Recipes in C: The Art of Scientific Computing (2rd ed.)* 200–203 (cambridge university press).