Supporting Information for "Water Activity from Equilibrium Molecular Dynamics Simulations and Kirkwood-Buff Theory"

Mária Lbadaoui-Darvas and Satoshi Takahama*

ENAC/IIE Swiss Federal Institute of Technology Lausanne (EPFL), CH-1015 Lausanne, Switzerland

E-mail: satoshi.takahama@epfl.ch

Phone: +41 21 6935777

Contents

S1 System sizes

Each simulation box contained of a total number of 5000 molecules, with solute mole fractions ranging between 0.001 and the solubility limit of the given system. Exact numbers of solute molecules are summarized in Table S1 together with the time of average box volumes ($\langle V \rangle$).

S2 Convergence of the simulations

We calculate the radial distribution functions in five 20 ns long windows, which allows us to estimate the time needed to reach convergence. In order to illustrate the convergence behavior, in Figure S1 we present the 20 ns window $g(r)$s for malonic acid at $x_s = 0.10$.

Table S1: The number of solutes in the simulated systems (the total number of molecules is 5000 in every simulation) and the average volume ($\langle V \rangle$) of the simulation boxes in nm^3 for the ethanol (ETH), glyoxal (GLY), malonic acid (MAL) and sodium chloride (NaCl) systems

		$\langle V \rangle$			
x_{sol}	n_{sol}	ETH	GLY	MAL	NaCl
0.001	5	151.4	148.8	150.6	148.11
0.002	10	153.8	150.6	148.9	148.5
0.005	25	156.0	150.5	152.3	148.7
0.010	50	156.9	154.0	155.7	148.8
0.020	100	164.5	156.2	158.3	148.9
0.050	250	166.4	161.9	170.0	149.7
0.100	500	182.3	210.6	190.1	153.9
0.250	1250	233.7	332.8	251.3	160.9
0.500	2500	366.8	-	-	-
0.750	3750	378.0	-	-	-

Systematic changes diminish already after the first 20 ns and only slight statistical variations can be observed. The inset of Figure S1 shows the enlargement of the first two peaks of the solute-solvent $g(r)$s, proving the lack of both systematic and significant statistical variations. This indicates that the simulations can be considered converged already after 20 ns and thus our equilibration time (50 ns) is largely sufficient. Malonic acid is the largest of the solute molecules thus the convergence of these simulations is the slowest; indeed for the rest of the systems and all other concentrations — not shown here — we observe very similar convergence times regarding the radial distribution functions. Other faster variables such as box volume, potential energy and density converge to their equilibrium values in approximately 2–3 ns for each of the systems.

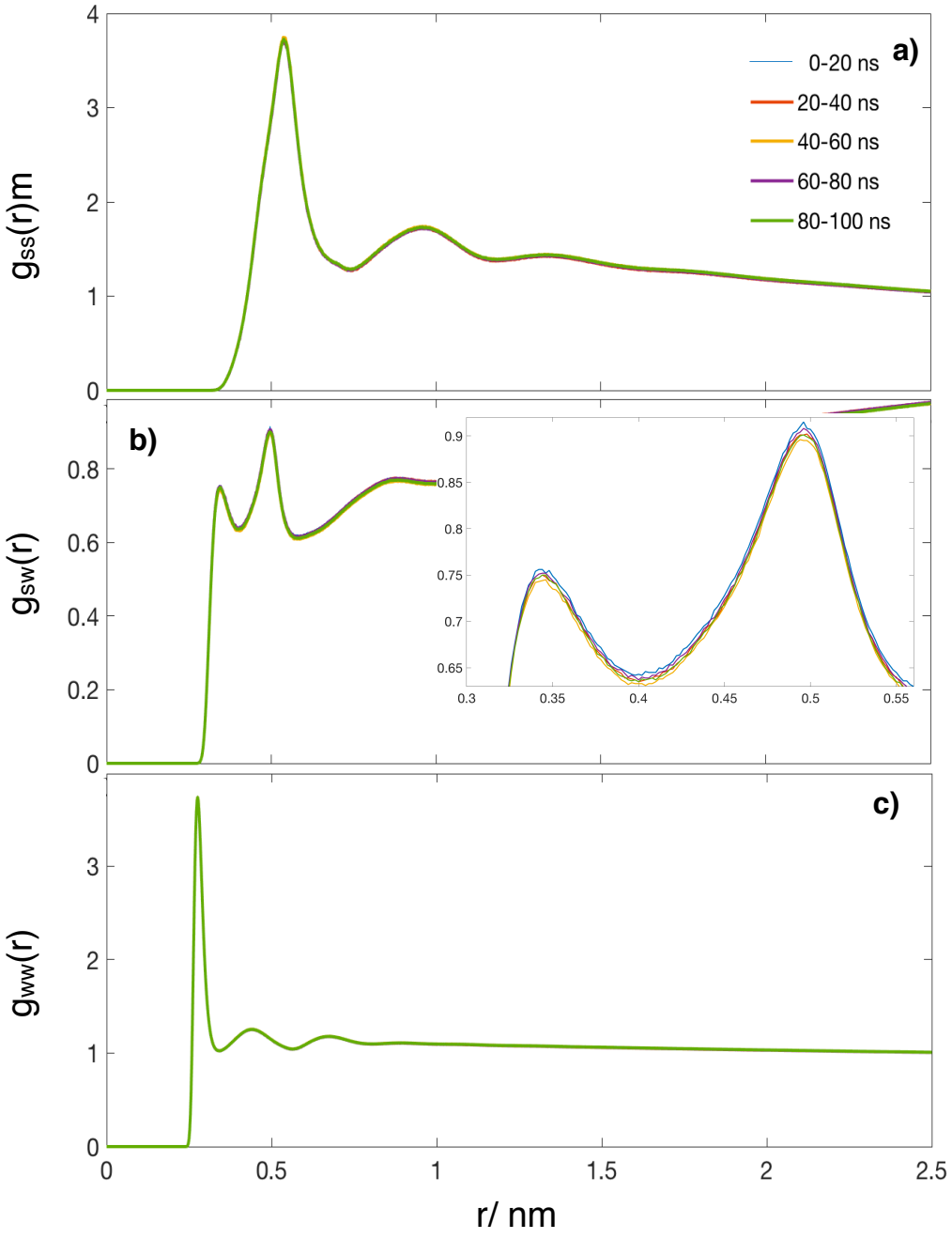

Figure S1: Radial distribution functions of the malonic acid system at $x_s = 0.10$ averaged over 20 ns-long windows. The inset is an enlargement of the first two peaks of the solute-solvent $g(r)$.

S2.1 Running Kirkwood-Buff intergrals

KB integrals as a function of integration distance are shown in figure S2 for ethanol/water (a) and malonic acid/water (b) systems, representing the best and the worst fit to water activity values. Every running KBI converges to a plateau above 2 nm, thus subvolume KBIs calculated at $R_{max} > 2$ nm are adequate for fitting and extrapolation to the thermodynamic value. The rest of the systems behave similarly.

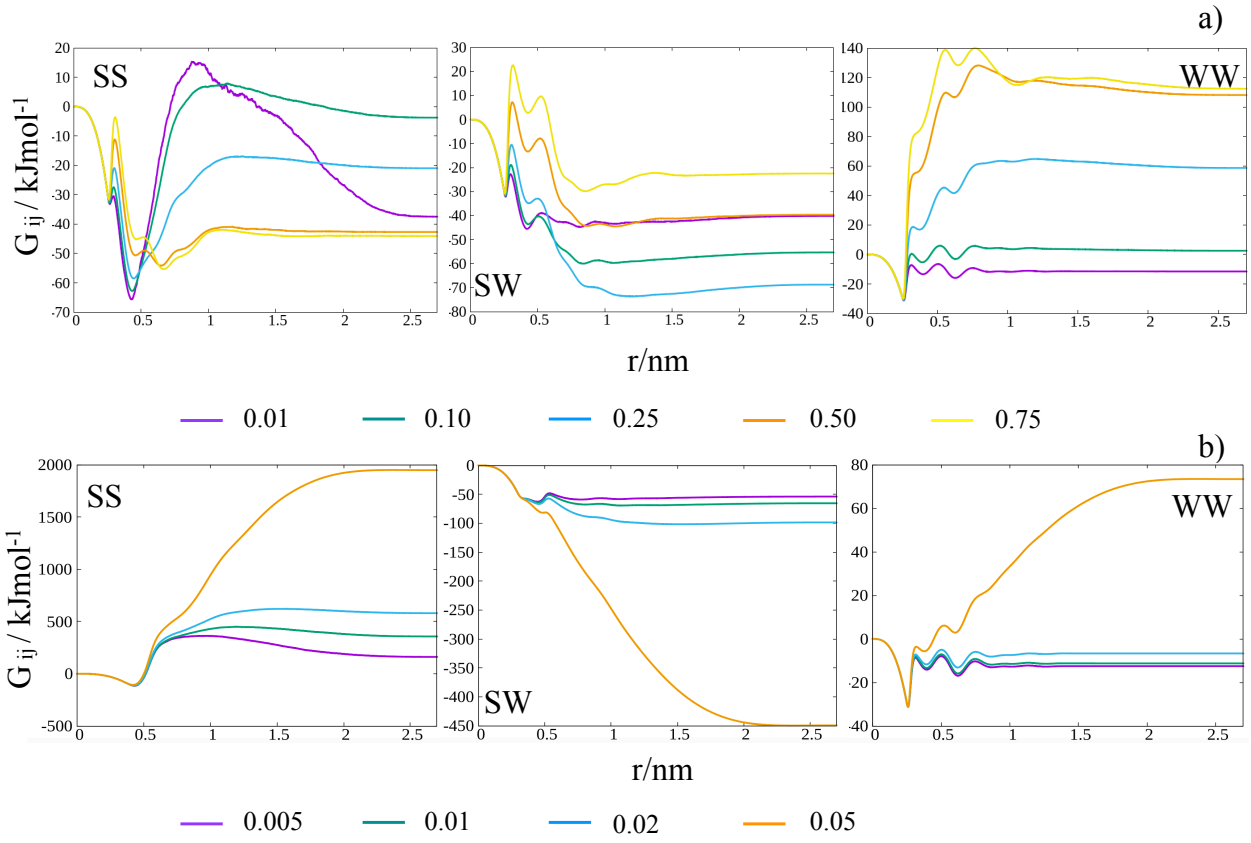

Figure S2: Solute-solute (SS), solute-water (SW) and water-water(WW) KBIs as a function of the integration distance, a) ethanol/water b) malonic acid/water systems. Colors denote concentrations (in mole fraction scale) and are resolved in the legend separately for each panel.

S3 Comparison of Force Fields

S3.1 Comparison of all-atom and united atom OPLS force field for ethanol

The transferable OPLS-UA[1] describes alkyl groups (CH, CH_2 and CH_3) as united atoms. Using united atom models for the CH_x groups results in an approximately 1.5 fold speed up to the simulations in case of ethanol at a moderate concentration ($x_s = 0.01$), which is expected to be even larger for higher concentrations. To characterize the effect of such simplifications, we repeat the simulation of the ethanol/water system at $x_s = 0.01$ with the OPLS-AA[2] all-atom potential and compared $g(r)$s, KBIs, and derivatives of the chemical potential. Figure S3 shows the radial distribution functions obtained from the all atom and the united atom simulations respectively. A significant difference is seen in the solute-solute $g(r)$s (Figure S3a). The height of first peak of the $g(r)$ obtained from the united atom simulation is three times larger than that obtained from all atom models. A similar effect is visible for the second peaks, albeit in a less pronounced manner. This is due to the difference in size between a united atom and all atom CH_2 group which influences packing. Solute-water $g(r)$s (Figure S3b) also show some variations as a function of model complexity. The united atom force field yields a $g(r)$ with two symmetric peaks, whereas the first peak is split in case of the all atom model, which might be due to attraction between the explicit hydrogen of CH_2 group and the water oxygen. Water-water curves (Figure S3c) are the same both from united and all atom force fields. Although the effect of the force field seems non-negligible on the $g(r)$s and the resulting KBIs, in reality it does not affect the value of the thermodynamic parameters needed for calculations to large extent.

Table S2 shows KBIs in their thermodynamic limit obtained from all atom and united atom simulations together with the relevant thermodynamic data obtained from them. While solute-solute and solute-water KBIs are significantly different from each other using the two models, numerical differences are small (2–3%) for the derivatives of the chemical potential;

Table S2: KBIs and thermodynamic derivatives from the OPLS-UA and OPLS-AA force fields. Square brackets indicate standard errors obtained from the linear fitting used to calculate the thermodynamic limit of the KB integrals.

	KBIs			Chemical Potentials	
force field	G_{ss}	G_{sw}	G_{ww}	$\partial\mu_w/\partial x_w$	$\partial\mu_s/\partial x_s$
OPLS UA	-87 [4]	-48.6 [0.2]	13.10 [0.01]	26.1	25.2
OPLS AA	-137 [6]	-55.9 [0.6]	12.8 [0.02]	26.9	25.7

thus the final results after integration are not significantly altered by substituting explicit alkyl groups with their corresponding united atom representations. While subtle structural details, such as the splitting of the first peak of the solute-solvent $g(r)$s are washed away by the simplified atom model, the overall solution microstructure, and thus the effective concentrations which are related to activity and chemical potential are accurately reproduced by the UA model, which is in line with the fact that these apolar groups do not play a significant role in interactions that could drastically alter effective solute concentrations, such as molecular association for instance. Ethanol/water mixtures show microsegregation whose main driving forces are hydrophobic interactions, which are largely insensitive to presence or absence explicit hydrogen atoms on the hydrocarbon backbone.

S3.2 Comparison of the KB, OPLS and NBFIX force fields for the NaCl systems

Based on prior art,[3,4] we expect that i) non-bonded fixing (NBFIX) corrections can improve nonpolarizable force fields in reproducing osmotic pressure of aqueous NaCl solution, ii) the best nonpolarizable force fields underperform in comparison to the best polarizable force fields, and iii) force fields developed on KBIs[5] (or other properties that depend on finite concentration behavior) perform better than standard ones. Without aiming for an exhaustive force field evaluation, we repeat simulations of the NaCl/water systems using the OPLS,[1,2] KBFF,[5] and NBFIX[3] force fields; the latter two matched with the SPC/E[6] and TIP3P[7]

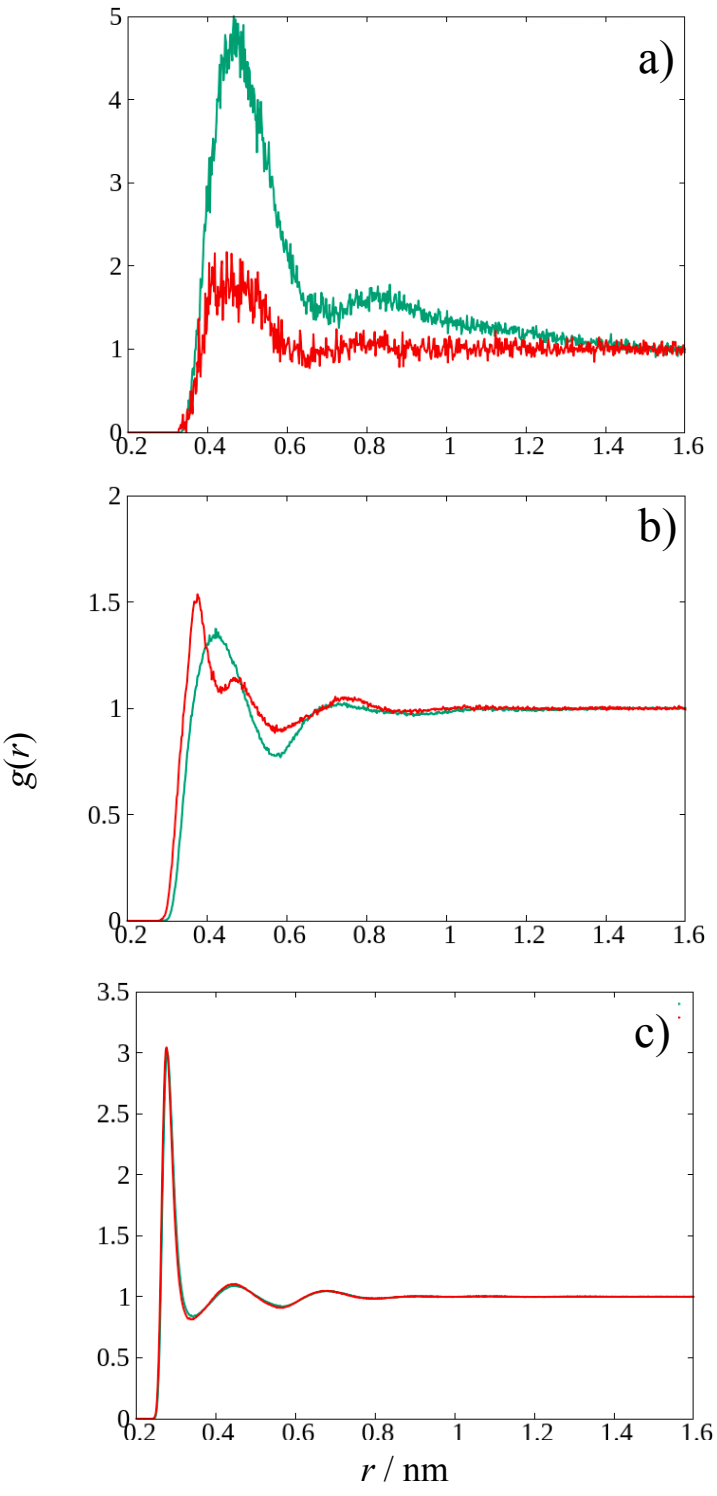

Figure S3: Radial distribution functions of an ethanol/water system ($x_s = 0.01$). Green: united atom, red: all atom, top panel: solute-solute, middle panel: solute-water, bottom panel: water-water.

water models, respectively.

Table S3: KBIs obtained from simulations with the OPLS, KBFF and NBFIX potentials. Square brackets indicate standard errors obtained from the linear fitting used to calculate the thermodynamic limit of the KB integrals.

	G_{ss}			G_{sw}			G_{ww}		
x_s	OPLS	KBFF	NBFIX	OPLS	KBFF	NBFIX	OPLS	KBFF	NBFIX
0.001	-249[73]	-284[48]	-	-0.73[0.015]	-5.35[0.04]	-	-14.2[0.016]	-14.04 [0.002]	-
0.002	-445[30]	-214[47]	-	-1.63[0.012]	-4.62[0.05]	-	-14.1[0.01]	-13.99 [0.008]	-
0.005	249[16]	-120[15]	-261	-2.40[0.011]	-5.57[0.06]	-2.92	-14.0[0.004]	-14.12 [0.006]	-13.92
0.02	139[0.5]	111[2]	-124	-6.80[0.004]	-10.59[0.02]	-3.72	-13.65[0.001]	-14.23[0.006]	-13.89
0.05	199[0.2]	94.5[2.5]	-62.1	-19.8[0.002]	-18.5[0.1]	-4.08	-13.47[0.0003]	-13.41[0.004]	-13.04
0.10	234[0.03]	160[1]	121	-47.4[0.02]	-42.9[0.2]	-31.1	-8.26[0.004]	-8.54[0.01]	-10.92
0.25	165[0.03]	571[15]	-	-107.3[0.001]	-312[7.8]	-	34.2[0.01]	140[3.8]	-

Table S3 summarizes the values and uncertainties of the KBIs obtained from these calculations. It is apparent that changing the force field has a strong effect on solute-solute and solute-water KBIs, while solvent-solvent ones are practically unaffected. Overall trends along concentration are similar for all KBIs and all force fields; pure terms show an increase with increasing concentration, while mixed terms decrease in the same direction, which considering the Δ_{ij} parameter whose positive or increasing values indicate the tendency of the system to associate. This means that, with increasing concentration, all these mixtures tend towards phase separation. The exact values however depend strongly on the force field in use. Solute-solute terms for instance are the smallest for NBFIX which is in line with the method of force field development (scaling down the attraction between the solutes), somewhat larger for KBFF and the largest for OPLS. When considering solute-solvent terms KBFF yields the largest negative values, indicating the most repulsive interactions, followed by OPLS and finally NBFIX where solute-solvent terms are the most favorable. In order to see the effect that differences in the KBIs have on thermodynamic properties we compare osmotic pressure, obtained from our simulations using the relationship $-\Pi\,\bar{v}_w = RT \ln a_w = \mu_w - \mu_w^*$,[8,9] with ideal values from the van't Hoff equation and those obtained with the VM method

of Luo and Roux as well as from experiments[10] in Figure S4. For these calculations the pure solvent molar volumes are used. OPLS and CHARMM (with the VM method) from simulations by Luo and Roux[3] yield the poorest agreement with experimental values (nearly ideal), and improves when using KBFF. The agreement with experiment is the best for the NBFIX parameters. We note that the KB method combined with the NBFIX force field and the VM simulations of Luo and Roux with the same force field are similarly accurate in reproducing experimental water activities.

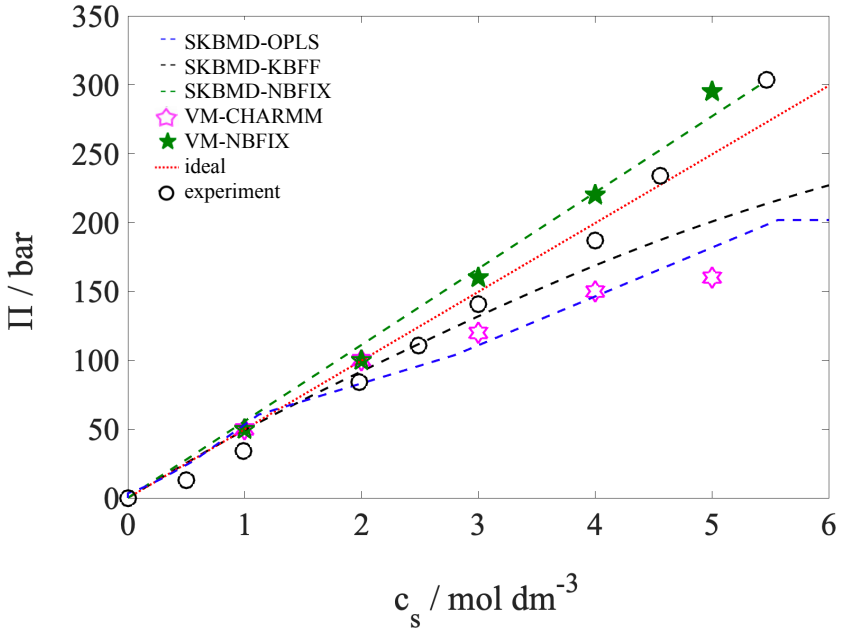

Figure S4: The osmotic pressure as a function of molar concentration. VM-CHARMM and VM-NBFIX refer to simulation data using the CHARMM force field[11] — with and without NBFIX parameters — from Luo and Roux,[3] and experimental data is taken from Hamdan et al.[10]

S4 Structure and solubility

From the MD snapshots (Figure 1a) it is evident that ethanol is solvated at any of the simulated concentrations, which is paired with small cluster sizes throughout the entire concentration range as seen in Table 1. Furthermore, solute-solute $g(r)$s do not exhibit long-range

structures (Figure S5) and fluctuations in the cluster size are small at any concentration (Figure S6). For glyoxal we observe large clusters above $x_s = 0.10$ which approximately corresponds 4.5 M in the snapshots (Figure 1b) as well as in Table 1 and this coincides with the onset concentration for large fluctuations in the time profile of cluster size (Figure S6), long-range structure of the $g(r)$s however does not appear at any of the simulated concentrations, thus large scale association of the solutes is more plausible than phase separation. However, repeating simulations with twice the number of molecules at the largest concentration reveals that the observed phase separation is an artifact of finite size effects, which is discussed in more details in the following section. The experimental solubility limit of malonic acid is underestimated by the force field in use. Onset concentration for long-range structure in the solute-solute $g(r)$s and cluster-size fluctuations is $x_s = 0.05$ (≈ 2.5 M). At the next concentration, $x_s = 0.10$ (≈ 4.3 M) all the solute molecules form a stable single cluster throughout the entire simulation, which can be considered as a separate organic phase. For NaCl, structural analysis is provided for the NBFIX force field parameterization as it is the most accurate. No sign of phase separation is observed up to $x_s = 0.10$ (≈ 5.3 M), which is close to the experimental solubility limit. An additional simulation at $x_s = 0.25$ results in a biphase system.

S4.1 Radial distribution functions

Long-range structure of the solute-solute radial distribution functions is an indicator of phase separation. Figure S5 shows solute-solute radial distribution functions for each solute at selected concentrations. Ethanol and glyoxal show almost no periodicity beyond 0.7 nm, even at the highest concentrations, while malonic acid exhibits oscillatory behavior up to 1.5 nm for any concentration higher $x_s = 0.01$ indicating extensive clustering and eventual phase separation. NaCl shows a very strongly oscillatory behavior only in the case of $x_s = 0.25$, which is clear indication of phase separation. We note that the amplitude of these oscillation is smaller at larger distances due to the pseudo-binary treatment, i.e. treating both anions

and cations as non distinguishable for the sake of $g(r)$ and KBI calculations. Individual ionic $g(r)$s exhibit somewhat more pronounced oscillations even beyond 1 nm.

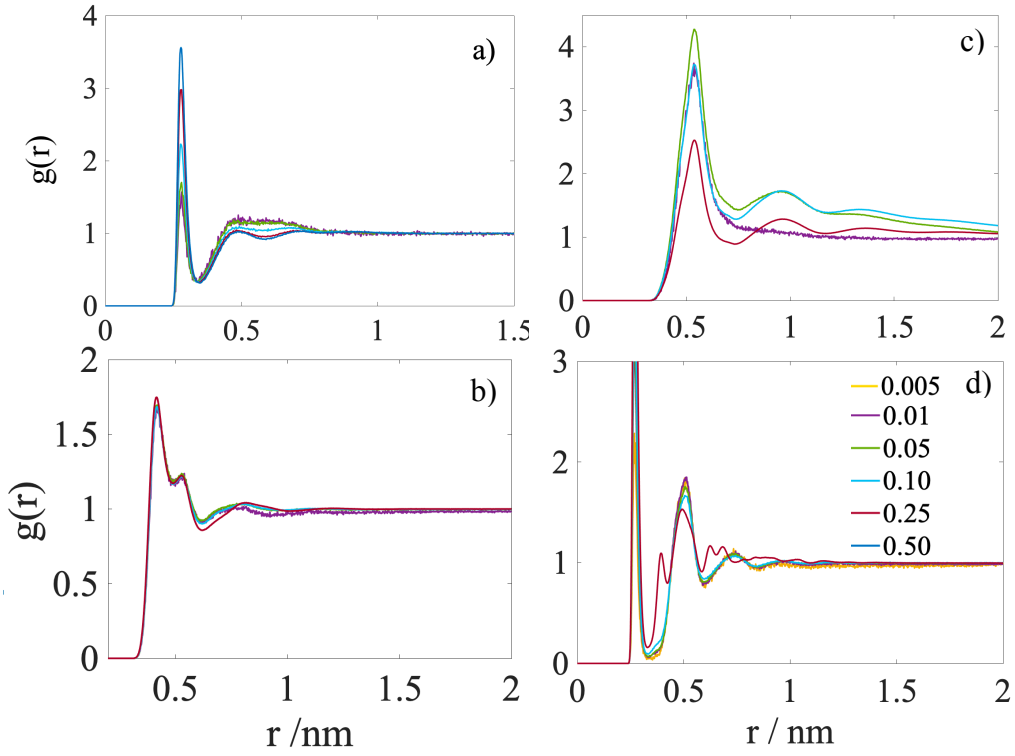

Figure S5: Solute-solute radial distribution functions of a) ethanol, b) glyoxal, c) malonic acid and d) NaCl. Colors indicate solute mole fractions.

S4.2 Cluster size evolution

We calculate the size of the largest molecular clusters in our systems using the depth first search (DFS) clustering algorithm[12] as implemented in Plumed2.1.[13] Cutoff distances for the neighbor searching are defined as the first minimum of the corresponding solute-solute radial distribution functions. Figure S6 shows the time evolution of the size of the five largest clusters in each system. The smallest extent of association is observed in the ethanol/water system (Figure S6a) with no evidence of large fluctuations. Glyoxal (Figure S6b) behaves similarly to ethanol in water at concentrations below $x_s = 0.10$, at which large fluctuations appear in the size of the largest cluster. These fluctuations are a sign of instability of

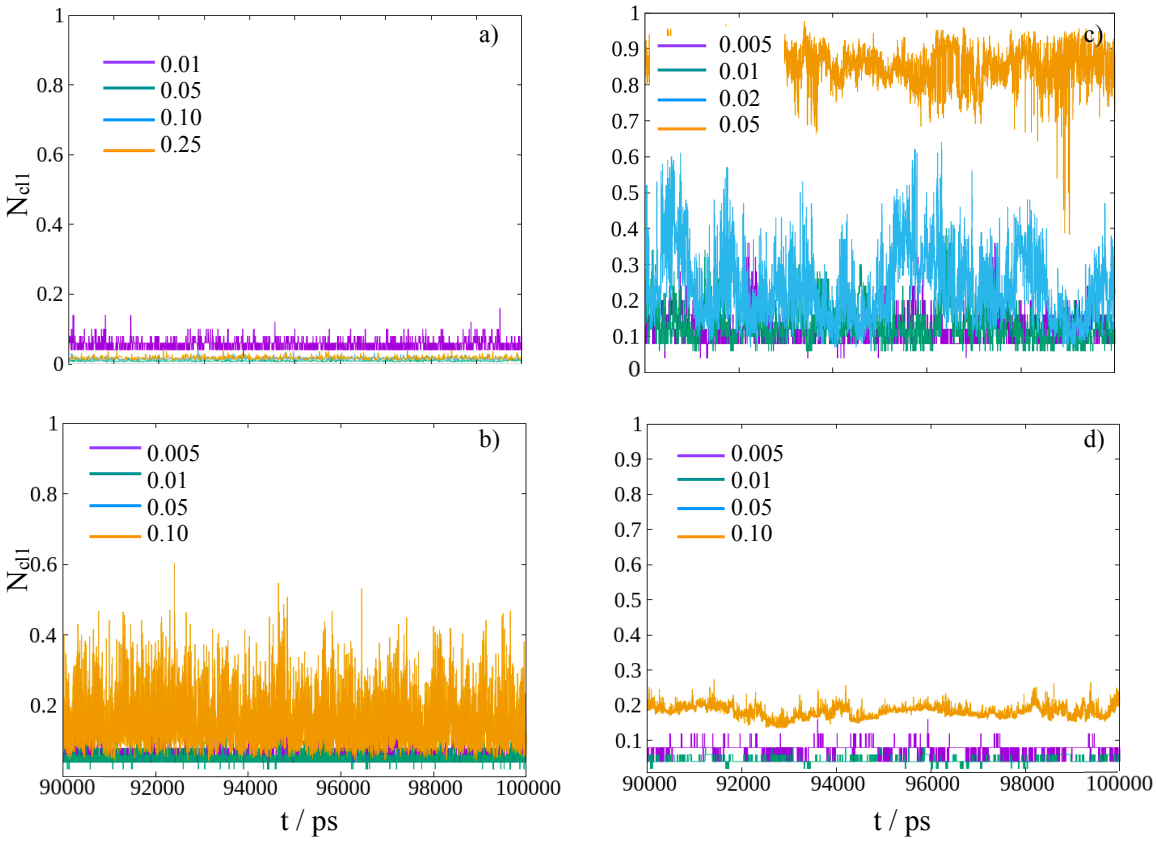

Figure S6: The size of the largest molecular cluster as a function of time in the last 10 ns window of the total simulation time in a) ethanol/water, b) glyoxal/water, c) malonic acid/water and d) NaCl/water systems. Color coding stands for different concentrations and is defined individually for each system in the legend. The data is normalized by the total number of solutes in each system, so the value of 1 means that every solute molecule belongs to the same cluster.

the system, and an indication of the proximity of the solubility limit. Malonic acid/water (Figure S6c) systems show a larger extent of clustering by H-bonds even at low concentration. Large fluctuations similar to those seen in glyoxal/water systems at $x_s = 0.10$ are visible at $x_s = 0.02$. Above this concentration, in line with what can be concluded from the KBIs, at least 80% molecules are found in the same cluster. NaCl (Figure S6) forms containing containing 2–5 pairs of ions at low concentrations, however, no sign of phase separation is visible up to $x_s = 0.10$.

S4.2.1 Finite Size Effects

The phase separation observed in the glyoxal/water systems may be related to finite size effects. To exclude or confirm this possibility we repeated our simulations with twice the number of solute and solvent molecules. Increasing the size of the system results in a more uniform equilibrium arrangement of solutes (Figure S7a), which may be seen as complete solvated system. However, cluster size analysis reveals that even in the larger systems around 98% of solutes belongs to the same cluster. This observation, together with the fact that even within the cluster glyoxal molecules tend to remain hydrated, suggests that what we observe is indeed not phase separation but hydrate formation. The same increase in the system size for the malonic acid/water system at the observed phase separation limit ($x_s = 0.10$) does not bring about a significant change in the results, the largest cluster contains 95% of the solute molecules throughout the entire simulation time. Thus, the observed aggregation and underestimated solubility of malonic acid molecules cannot be explained by finite size effects. We note that this latter simulation was initiated from a random arrangement of the molecules, which does not change conclusions about the solution structure. Figure S7 shows equilibrium snapshots of the big systems. Glyoxal (a) appears to be perfectly mixed with water, however, it is single aggregate of dihydrates, at least if cutoff values for the cluster analysis are defined as the first minimum of the corresponding solute-solute $g(r)$. Malonic acid (b) is present as a single large aggregates similar to that observed in the original smaller

simulation box, with a few oligomers around it.

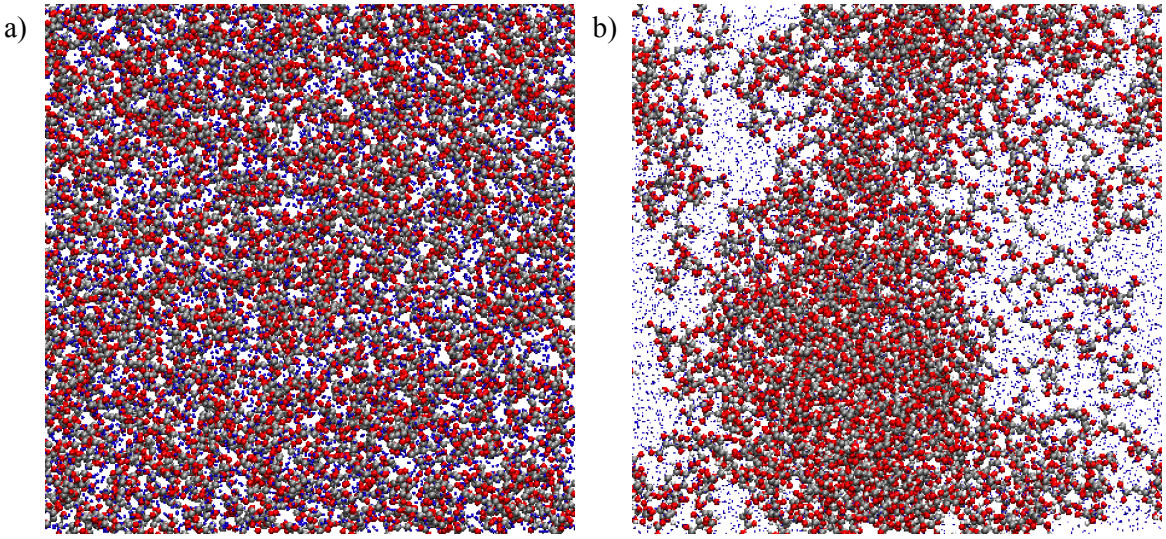

Figure S7: Equilibrium snapshots of the systems containing twice the number of molecules, a) glyoxal b) malonic acid.

S5 Partial molar volumes

Partial molar volumes calculated from KBIs as $\bar{v}_i = [1 + c_j(G_{jj} - G_{ij})]/[c_i + c_j + c_i c_j \Delta_{ij}]$ can be used to convert chemical potential or activity to osmotic pressure. Figure S8 shows a comparison between calculated and experimental values for the ethanol/water system. The partial molar volume of the solute has a very similar trend to that observed in experiments, with around 10% of maximum deviation. The low concentration structure (a minimum at $x_s \approx 0.1$) is, however, missing in the simulated curve, due probably to weaker statistics in that region. For water, simulated and experimental data agrees within 2% up to $x_s = 0.25$.

For higher concentrations relative large deviations and an inversion in the trend can be observed. For this reason, the fixed molar volume of water (of $\bar{v}_w = 18.04\ \mathrm{cm^3\,mol^{-1}}$)[14] is used instead of the calculated $\bar{v}_w$.

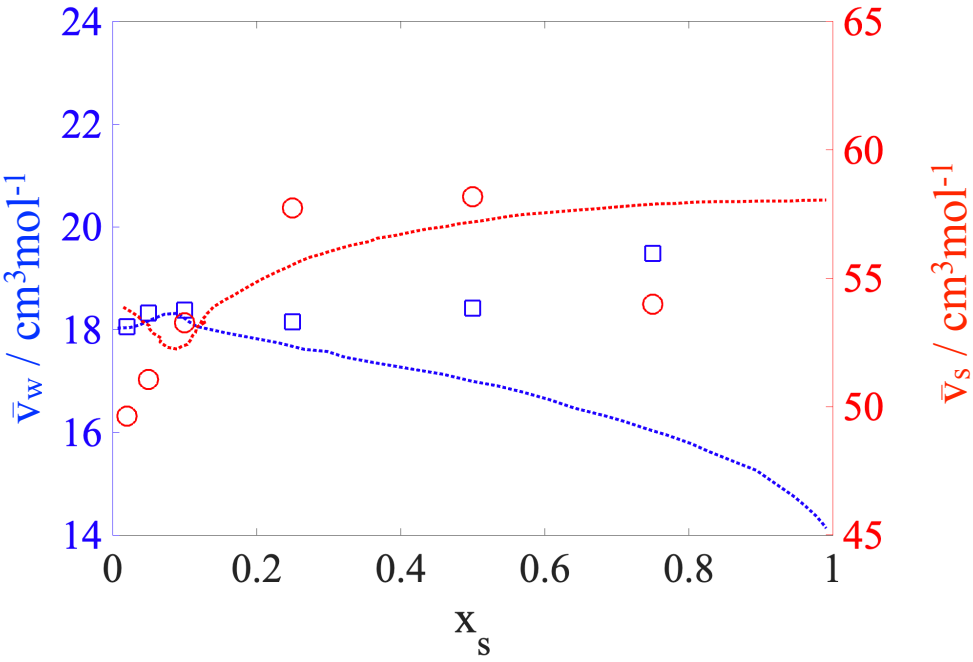

Figure S8: Partial molar volumes from the KBIs obtained from the OPLS-UA simulations (empty circles and squares) compared to experimental values (dashed lines).[15]

S6 Exploring contributions to model-measurement discrepancies for malonic acid/water systems

S6.1 Uncertainty

The uncertainty of the simulations is estimated by the block average method, which defines the variance of the data using average values of the quantity S calculated in N_b separate

non-overlapping blocks of the simulations as:

$$\sigma^2 = \frac{1}{N_b - 1} \sum_{i=1}^{N_b} (\langle S \rangle_i - \langle S \rangle)^2, \tag{1}$$

where $\langle S \rangle_i$ denotes average value of S in the i^{eth} sample and $\langle S \rangle$ is the average over the total simulation time. The magnitude of the error is obtained as $\sqrt{\sigma^2/N_b}$. Water activities were calculated from 5 consecutive 10 ns blocks of the production part of the simulation for the error calculations. Numerical errors calculated this way are always considerably smaller than the difference between simulation and the experiment. Error estimates are shown in Figure S9. Additional evaluation of numerical precision is given in Section S7.

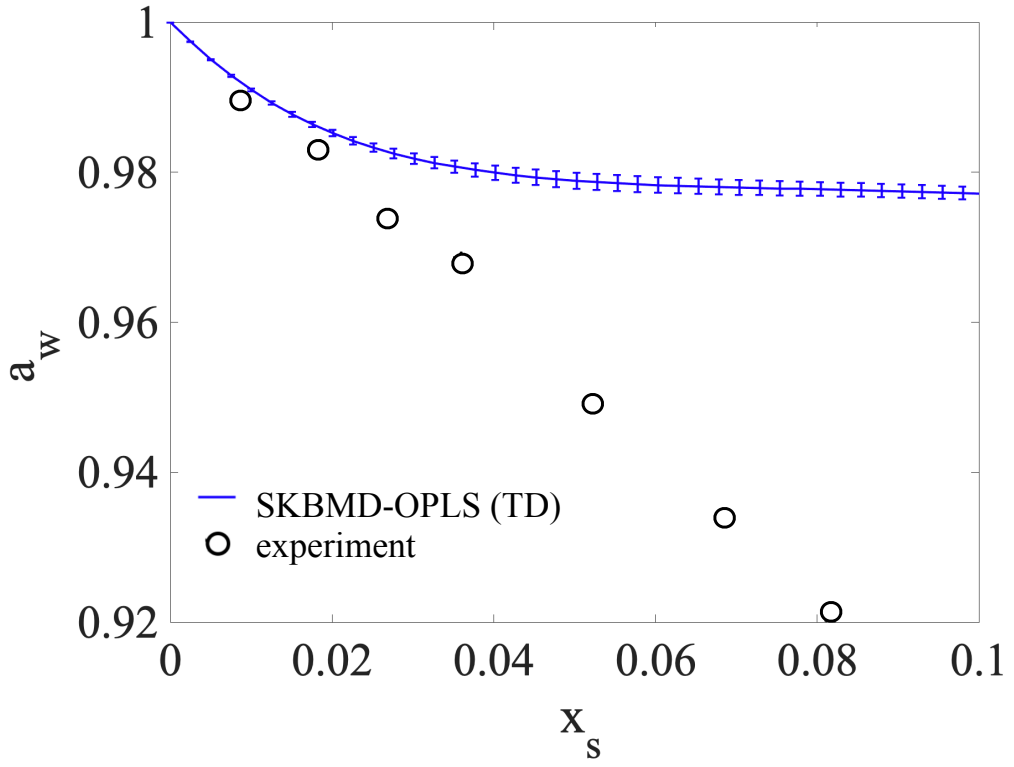

Figure S9: Water activities in the malonic acid/water systems estimated with the error bars estimated from block averaging are shown in blue. The experimental results of Peng et. al.[16] are shown as black empty circles.

S6.2 The effect of deprotonation

The largest assumption we make in this work is neglecting the acid-base equilibria of malonic acid. Malonic acid is considered relatively strong among organic acids with $\mathrm{p}K_{\mathrm{a1}} = 2.8$ for the first dissociation and $\mathrm{p}K_{\mathrm{a2}} = 5.7$ for the second one.[17] Therefore, we can expect that the percentage of non-dissociated molecules is more than 90% at any concentrations above 0.5 M ($x_s \approx 0.1$), below which it is not possible to model the fraction of dissociated molecules in a statistically meaningful way due to system size limitations. In practice it means that the lowest four of the concentrations we studied are the most affected by this systematic error. The equilibrium concentrations of non-dissociated malonic acid are shown in Figure S10, with the inset showing the same data converted to percentage of total acid concentration. The pH range studied with our deprotonated simulations is $\sim 2 - 4$.

To account for the effect of deprotonation, we repeat malonic acid simulations with the appropriate number of acid molecules replaced by malonate ($HCOOCH_2COO^-$) ions, and an equivalent number of water molecules replaced by hydronium ions (H_3O^+). The second dissociation was always neglected. As this treatment (deprotonation) results in a four-component system suffering from the issue of singularity arising from interrelated solute concentrations, a similar pseudo-binary treatment is used than for NaCl, ie. water and H_3O^+ ions and protonated and deprotonated malonic acid residues are grouped together in the phase of calculating the radial distribution functions. Water activities obtained with the two methods are shown in Figure S11.

Taking deprotonation into account during the simulation phase and using the pseudo-binary treatment from calculating the KBIs yields results whose agreement with experimental data is slightly worse than those obtained from the protonated malonic acid/water simulations. Moreover, introducing the correct ratio of deprotonated and protonated residues does not change the overall shape of the water activity vs concentration curve. Thus we conclude that neglecting deprotonation is not the main cause of deviations from experimental data.

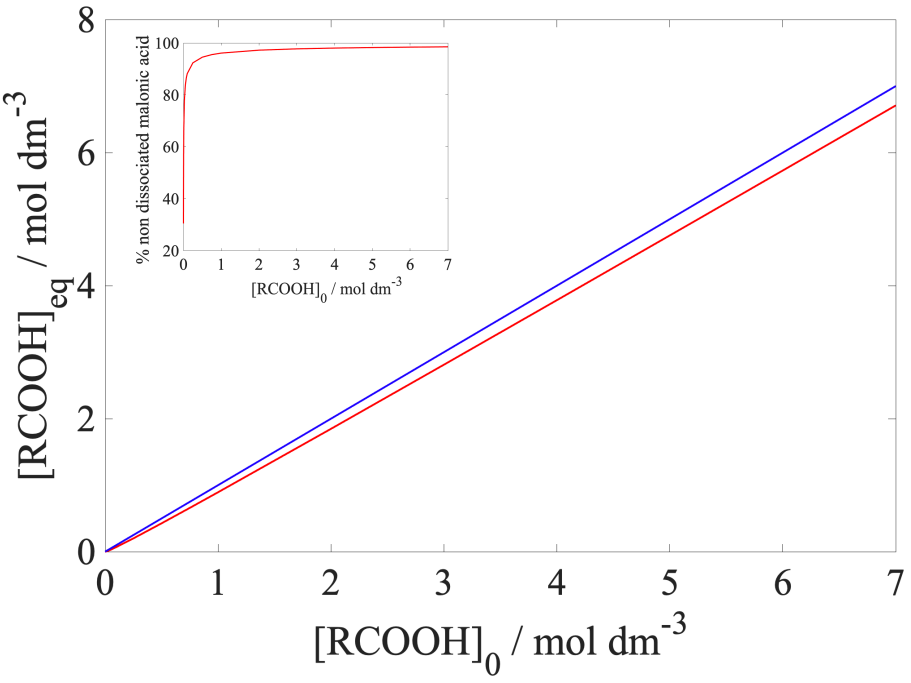

Figure S10: Concentration of the non-dissociated acid as a function of initial concentration for pK_a= 2.8 (red line). The 1:1 line (no dissociation) is shown in blue for comparison.

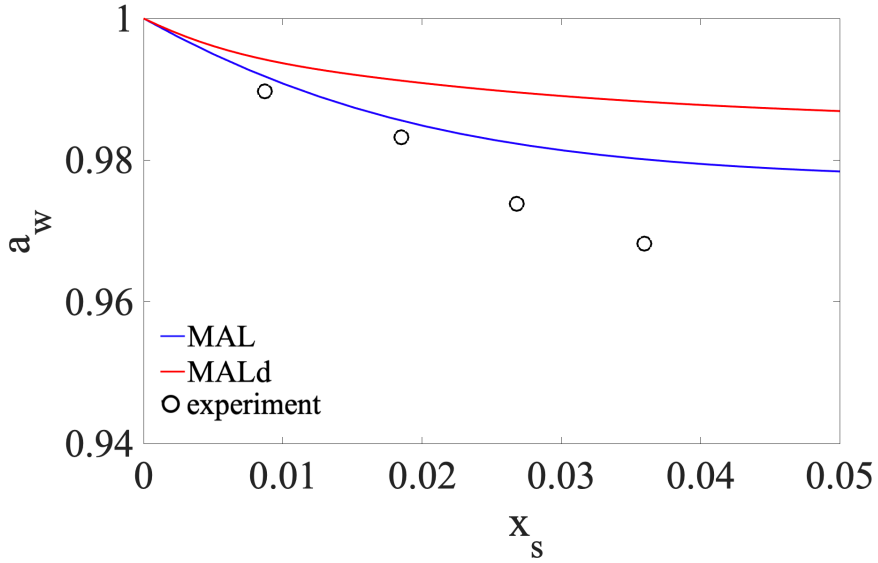

Figure S11: Water activity in MAL (blue) and MALd (red) systems compared to experimental data of Peng et. al.[16]

S6.3 Comparison of TD and CR KBIs

Correlated region (CR) KBIs are suspected to outperform thermodynamic limit (TD) KBIs for reproducing properties of strongly aggregating systems as it was demonstrated by Milzetti et al.[18] who artificially induced aggregation by modifying intermolecular interactions between the solutes in urea-water systems. They can be calculated by choosing only the correlated region (i.e., a region between the first and second peak of the corresponding $g(r)$) of the running KBIs for averaging. To highlight the differences between TD and CR approach in case of aggregating systems we calculate the integrals with both methods for glyoxal/water and malonic acid/water systems. Results (Figure S12) show that the choice of method does not make significant difference in case of weak aggregation, Figures S12b). In the case of aqueous malonic acid solutions (Figure S12c), the trends are consistent between the two methods but the peak height and position of the solute-solute term are shifted towards lower values when calculated from the CR method. The slopes of the other two KBIs also turn out to be smaller using the CR method. This confirms that the two methods differ indeed for strongly aggregating systems. However, in the lack of experimental data for malonic acid we cannot rank the accuracy of the two methods at the level of KBIs. We thus proceed to calculate water activity from both methods for malonic acid in order to better assess the differences between the TD and CR approaches (Figure S13). As it can be observed in Figure S13 the CR approach does not improve the results; differences between the two methods differences are within the uncertainty of the simulation (Figure S9). Thus we conclude that for our purposes the more easily automatable TD approach, which does not necessitate any subjective decision making is more appropriate.

S6.4 Evaluation of the NBFIX corrections for malonic acid

We evaluate the quality of the NBFIX corrections applied on the malonic acid force field by looking at the development of the agreement between simulated and experimental solute diffusion coefficients,[23] the former obtained from the Einstein relation,[24] which are presented

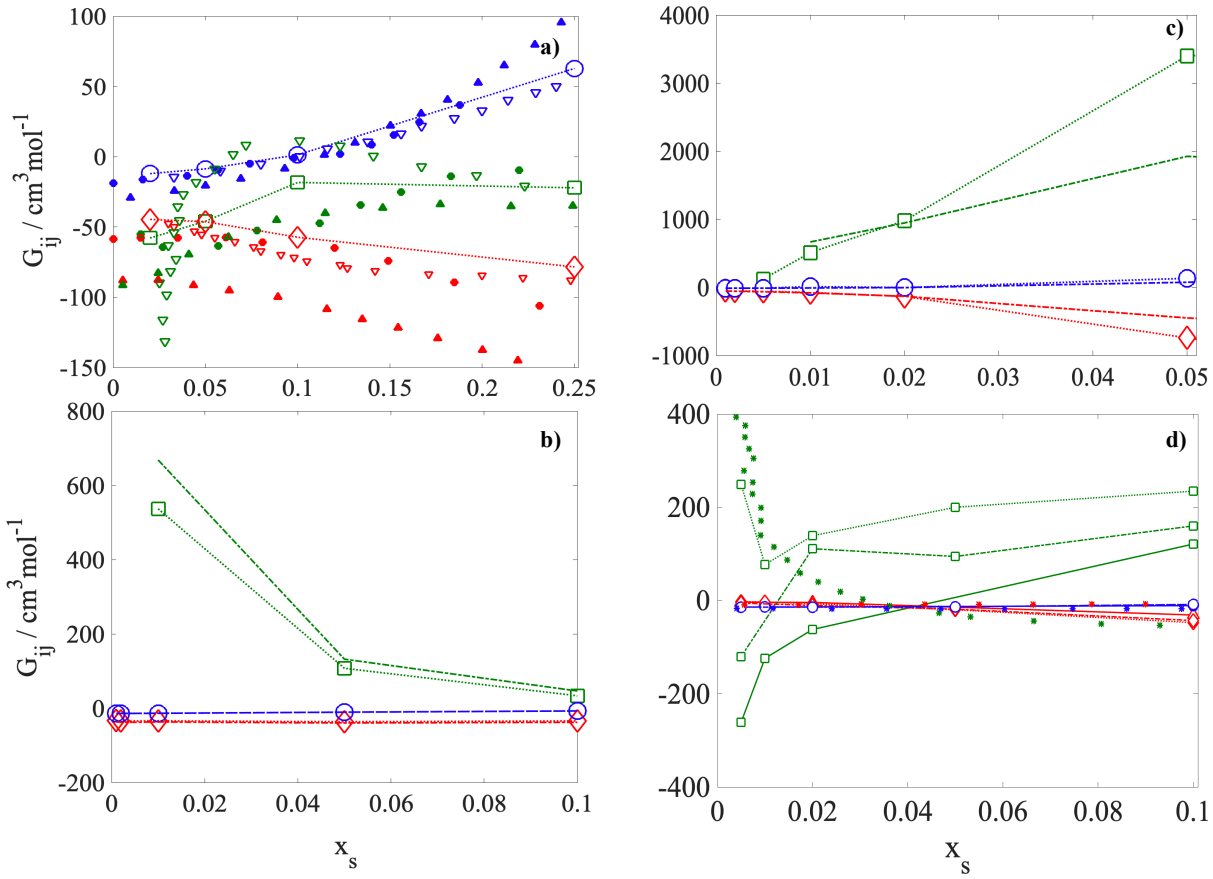

Figure S12: Kirkwood-Buff integrals as a function of solute mole fraction for a) ethanol/water, b) glyoxal/water, c) malonic acid/water and d) NaCl/water systems. Empty blue cirles and dotted lines: water-water, green squares and dotted lines: solute-solute and red diamonds and dotted lines: solute-water as obtained from the TD formulation with the OPLS forcefield, continuous lines and empty symbols in d) are obtained using the NBFIX corrections, and dash-dotted lines with empty symbols with the KBFF forcefield. Ethanol/water systems are compared with data obtain from experiments by Perera et. al.[19] (filled circles), Ben-Naim[20] (empty triangles) and Donkersloot[21] (filled triangles). NaCl/water systems are compared with experiments of Chitra et. al.[22] (asterisks). CR KBIs are plotted in dash-dotted lines for malonic acid/water and glyoxal/water sytems. Color coding: blue: water/water, red: solute/water and green: solute/solute

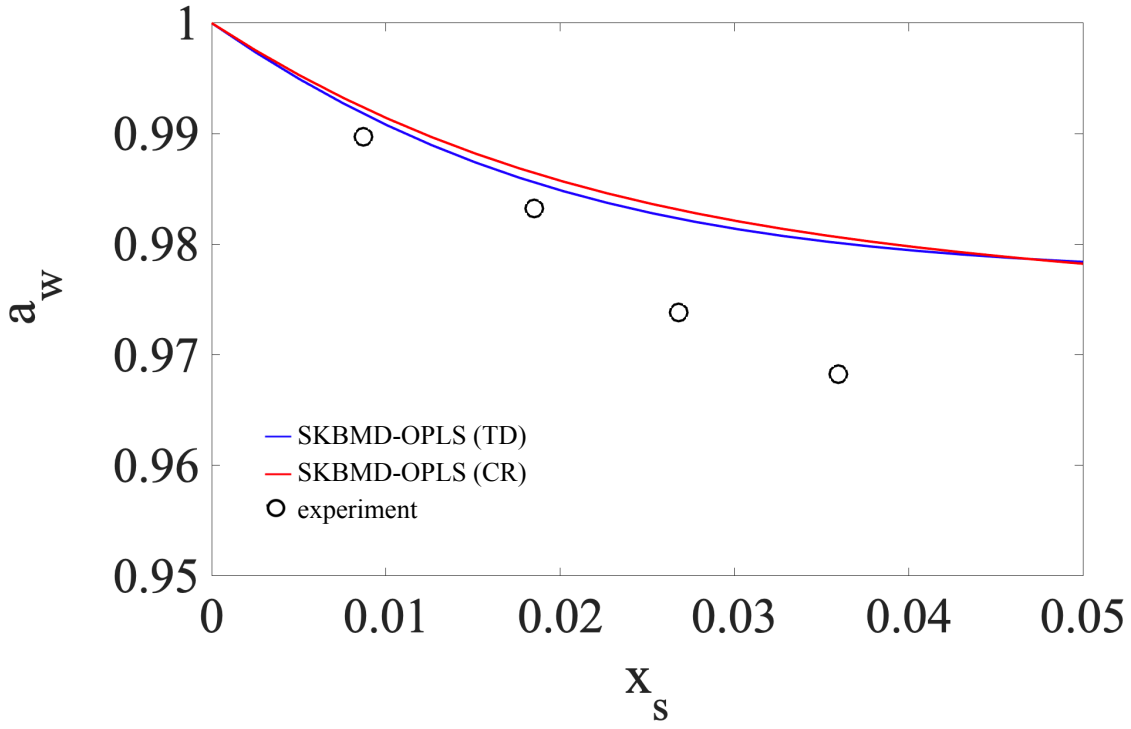

Figure S13: Water activity from the TD (blue) and CR (red) KBI approaches compared to experimental results of Peng et. al.[16].

in Figure S14. While the uncorrected OPLS potentials strongly underestimates diffusion, as molecules are bound in the large aggregate whose suppresses translational motion to a large extent, values obtained after applying the NBFIX corrections reproduce experimental diffusivity values more closely. It is apparent however, that increasing σ_{ij} between the carbonyl and hydroxyl oxygen atoms by more than 0.03 nm does not bring about any further development in neither the water activity nor the diffusivity of the system, which again is an indication of sampling insufficiency described in the main text.

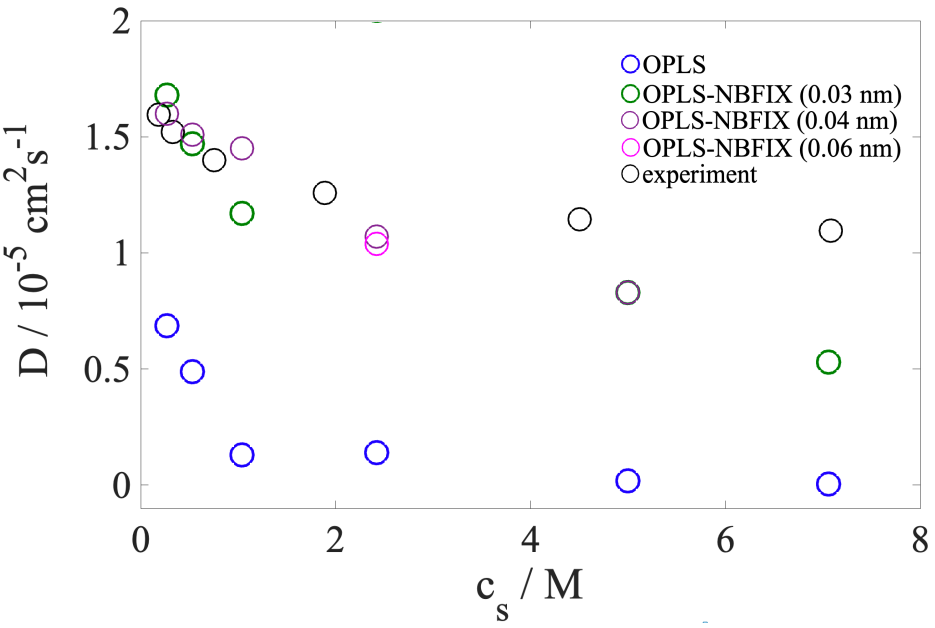

Figure S14: Diffusion coefficient of malonic acid in aqueous solution from simulations (this work) and experiments (Leaist[23]).

S7 Numerical precision

S7.1 Thermodynamic consistency through the Gibbs-Duhem relation

In addition to independent evaluations of the solute and water activity, chemical potential, and block-averaged uncertainties, a thermodynamic consistency test[25] of the Gibbs-Duhem relation (used to derive eq. 3) can be used to assess numerical precision of the KB approach. Since in a multicomponent system $\sum_i n_i d\mu_i = 0$, deviations of the left hand side from zero ("residuals") indicate this precision. Every absolute deviation is much smaller than $k_B T/1000$ (Figure S15).

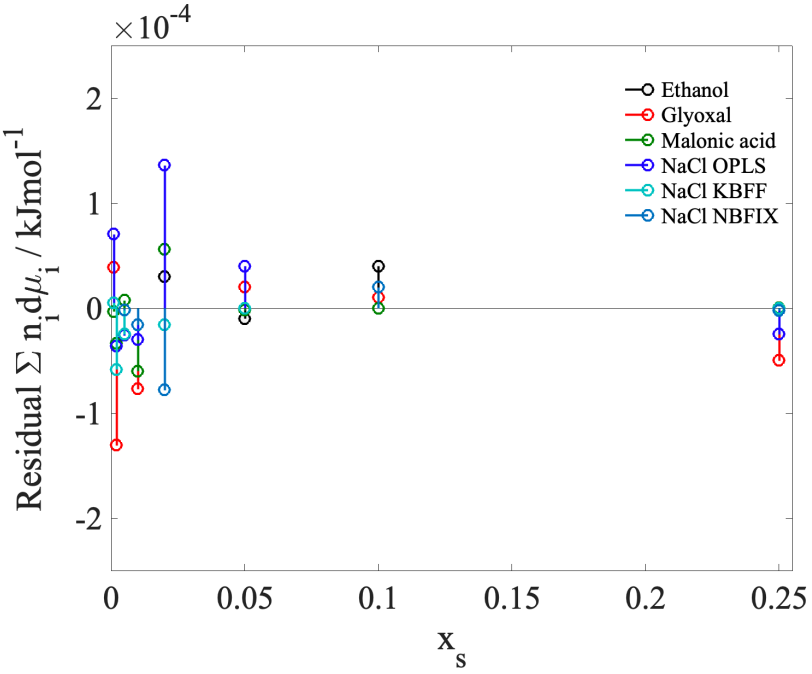

Figure S15: Residuals of the Gibbs-Duhem equation as a function of concentration.

S7.2 Interpolation and integration

In order to evaluate the challenges brought about by the necessity to interpolate and integrate a relative sparsely spaced dataset in the concentration range, we perform our analysis on a

continuous KBI dataset of ethanol/water mixtures taken from the work of Ben-Naim[20] using two sets of KBIs containing 5 and 20 datapoints respectively.

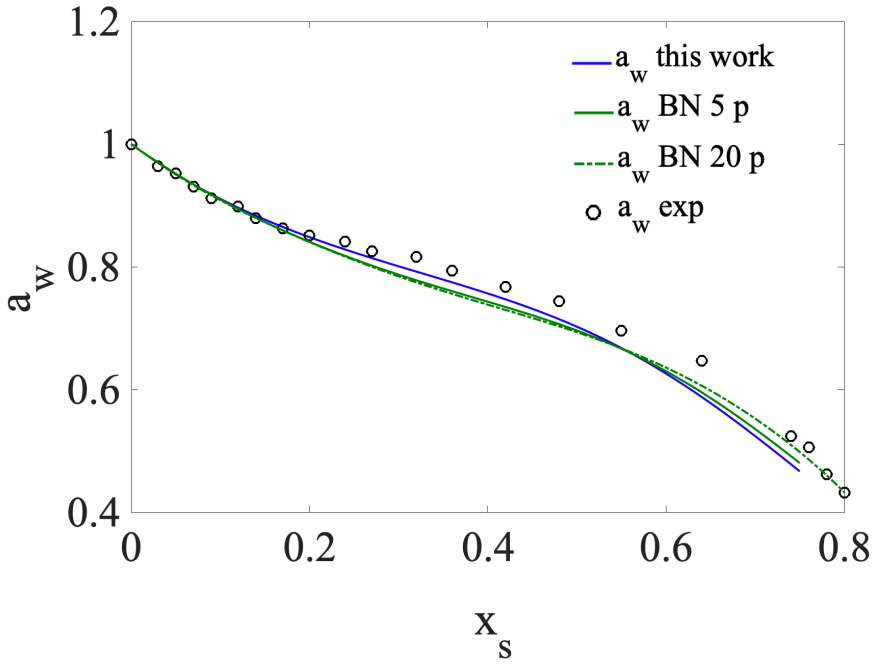

Figure S16: Experimental and KB-derived water activities using simulated points, and 5 and 20 points from the work of Ben-Naim[20].

Figure S16 shows that interpolating and integrating with a reduced number of sample points does not substantially affect the predictions. The difference between the curves obtained using 5 and 20 sample points never exceed 2%, and the main qualitative features of the curve are also reproduced. Moreover, the maximum difference between the simulated and experimentally derived curves (BN 5 p and BN 20 p) also does not exceed 3%.

References

(1) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(2) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(3) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *J. Phys. Chem. Lett.* **2010**, *1*, 183–189.

(4) Moučka, F.; Nezbeda, I.; Smith, W. R. Chemical Potentials, Activity Coefficients, and Solubility in Aqueous NaCl Solutions: Prediction by Polarizable Force Fields. *J. Chem. Theory. Comput.* **2015**, *11*, 1756–1764.

(5) Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides. *J. Chem. Theory. Comput.* **2011**, *7*, 1369–1380.

(6) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(7) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(8) Lewis, G. N. The Osmotic Pressure of Concentrated Solutions, and the Laws of the Perfect Solution. *J. Am. Chem. Soc.* **1908**, *30*, 668–683.

(9) Chinard, F. P.; Enns, T. Osmotic Pressure. *Science* **1956**, *124*, 472–474.

(10) Hamdan, M.; Sharif, A. O.; Derwish, G.; Al-Aibi, S.; Altaee, A. Draw Solutions for Forward Osmosis Process: Osmotic Pressure of Binary and Ternary Aqueous Solutions of Magnesium Chloride, Sodium Chloride, Sucrose and Maltose. *Journal of Food Engineering* **2015**, *155*, 10 – 15.

(11) Huang, J.; MacKerell, A. D. CHARMM36 All-atom Additive Protein Force Field: Validation based on comparison to NMR data. *J. Comp. Chem.* **2013**, *34*, 2135–2145.

(12) Knuth, D. E. *The Art of Computer Programming: Combinatorial Algorithms, Part 1*, 1st ed.; Addison-Wesley Professional, 2011.

(13) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Computer Physics Communications* **2014**, *185*, 604 – 613.

(14) Kell, G. S. Effects of Isotopic Composition, Temperature, Pressure, and Dissolved Gases on the Density of Liquid Water. *J. Phys. Chem. Ref. Data.* **1977**, *6*, 1109–1131.

(15) Lara, J.; Desnoyers, J. E. Isentropic Compressibilities of Alcohol-Water Mixtures at 25°C. *J. Solution. Chem.* **1981**, *10*, 465–478.

(16) Peng, C.; Chan, M. N.; Chan, C. K. The Hygroscopic Properties of Dicarboxylic and Multifunctional Acids: Measurements and UNIFAC Predictions. *Environ. Sci. Technol.* **2001**, *35*, 4495–4501.

(17) Kortüm, G.; Vogel, W.; Andrusson, K. *Dissociation Constants of Organic Acids in Aqueous Solutions*; IUPAC: Butterworths, London, 1961.

(18) Milzetti, J.; Nayar, D.; van der Vegt, N. F. A. Convergence of Kirkwood–Buff Integrals of Ideal and Nonideal Aqueous Solutions Using Molecular Dynamics Simulations. *J. Phys. Chem. B* **2018**, *122*, 5515–5526.

(19) Perera, A.; Sokolić, F.; Almásy, L.; Koga, Y. Kirkwood-Buff Integrals of Aqueous Alcohol Binary Mixtures. *J. Chem. Phys.* **2006**, *124*, 124515.

(20) Ben-Naim, A. Inversion of the Kirkwood–Buff Theory of Solutions: Application to the Water–Ethanol System. *J. Chem. Phys.* **1977**, *67*, 4884–4890.

(21) Donkersloot, M. C. A. The Structure of Binary Liquids. The Kirkwood-Buff Theory of Liquid Mixtures, Illustrated on the Basis of the Systems Water/Methanol, Water/Ethanol, and Cyclohexane/2,3-Dimethylbutane, as a Link Between Thermodynamic Data and X-Ray and Neutron Scattering results. *J. Solution. Chem.* **1979**, *8*, 293–307.

(22) Chitra, R.; Smith, P. E. Molecular Association in Solution: A Kirkwood-Buff Analysis of Sodium Chloride, Ammonium Sulfate, Guanidinium Chloride, Urea, and 2,2,2-Trifluoroethanol in Water. *J. Phys. Chem. B* **2002**, *106*, 1491–1500.

(23) Leaist, D. G. Mutual Diffusion Coefficients for Binary Aqueous Solutions of Arsenous, Arsenic, and Malonic Acids. *J. Chem. Eng. Data* **2007**, *52*, 1319–1325.

(24) Einstein, A. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik* **1905**, *322*, 132–148.

(25) Moučka, F.; Nezbeda, I.; Smith, W. R. Molecular Simulation of Aqueous Electrolytes: Water Chemical Potential Results and Gibbs-Duhem Equation Consistency Tests. *J. Chem. Phys.* **2013**, *139*, 124505.