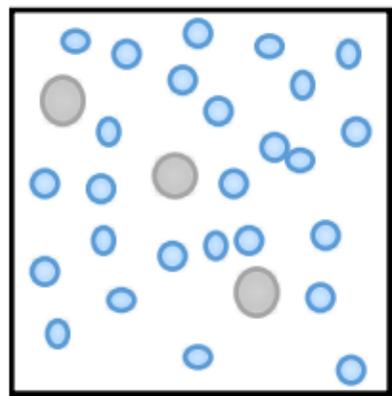
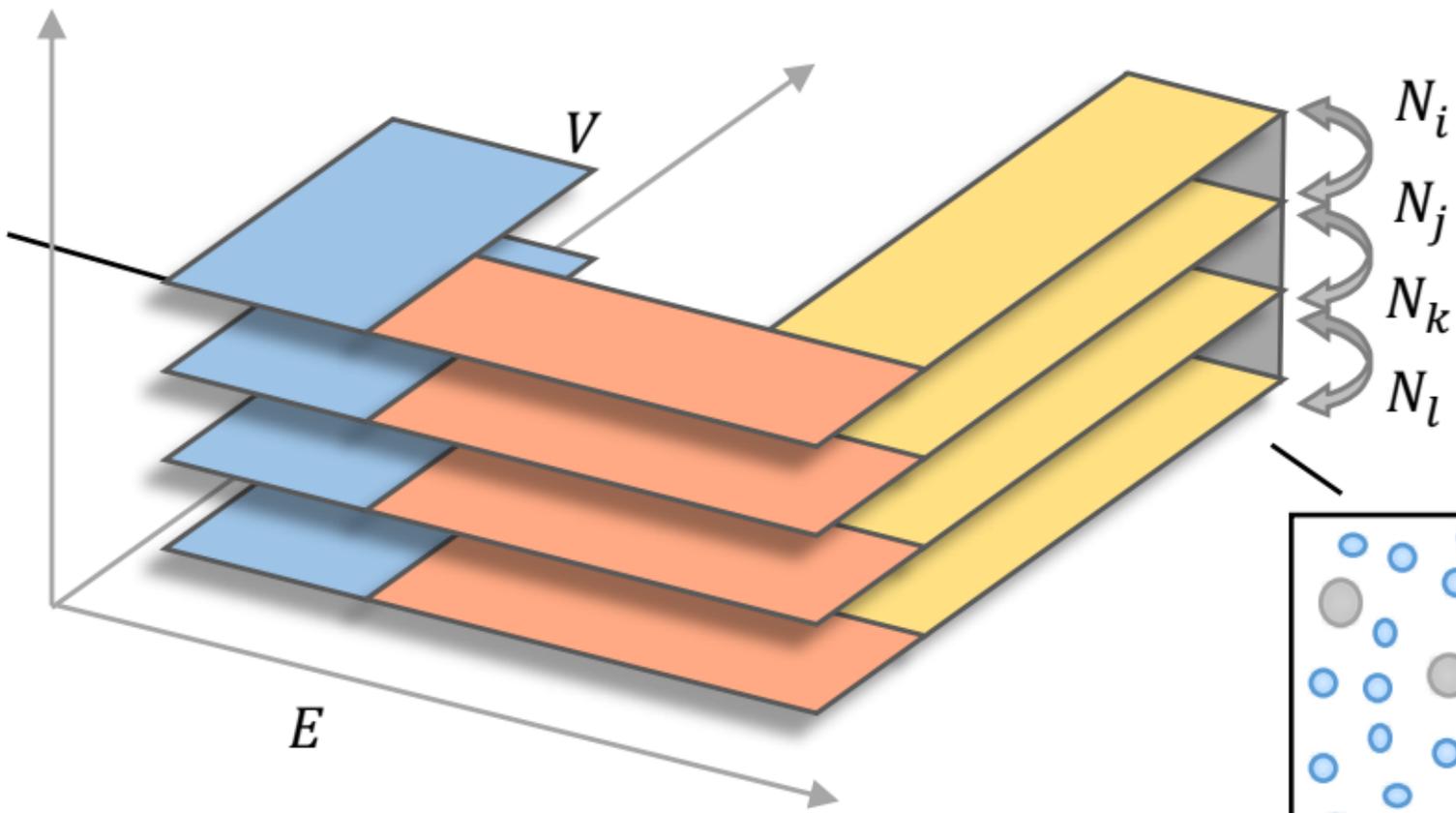
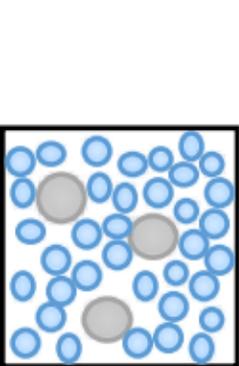
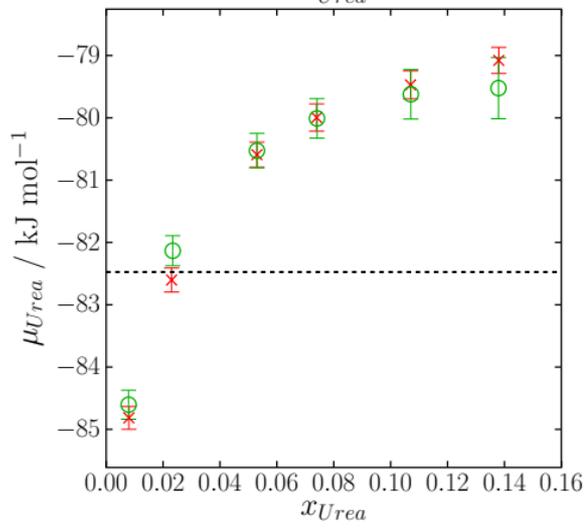
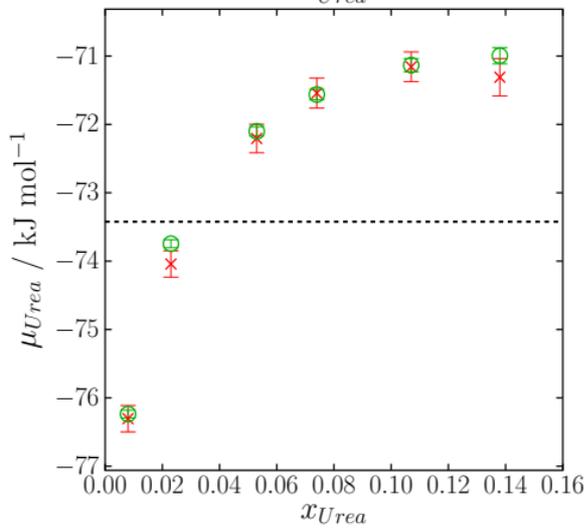
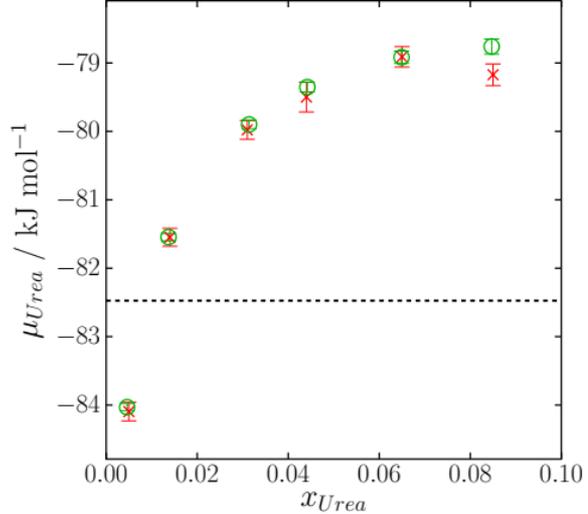
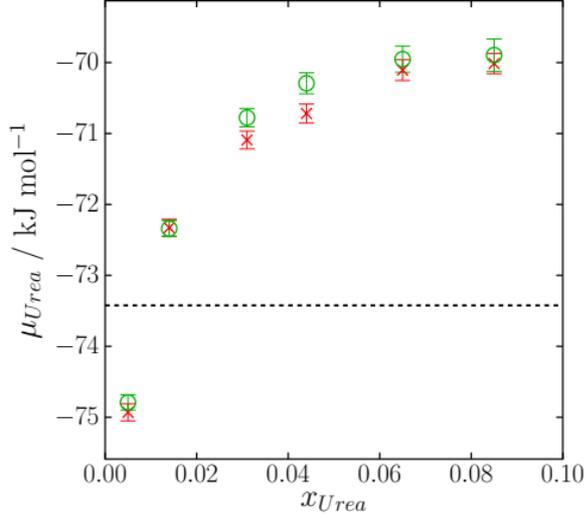
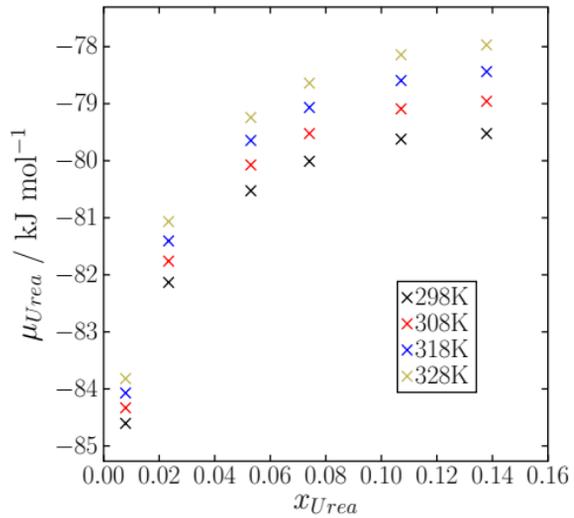
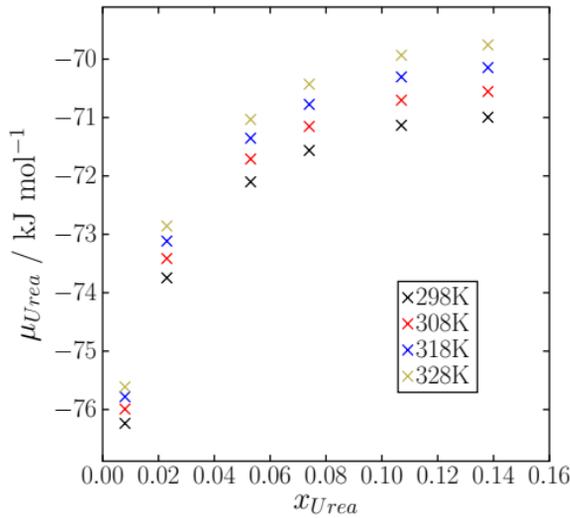
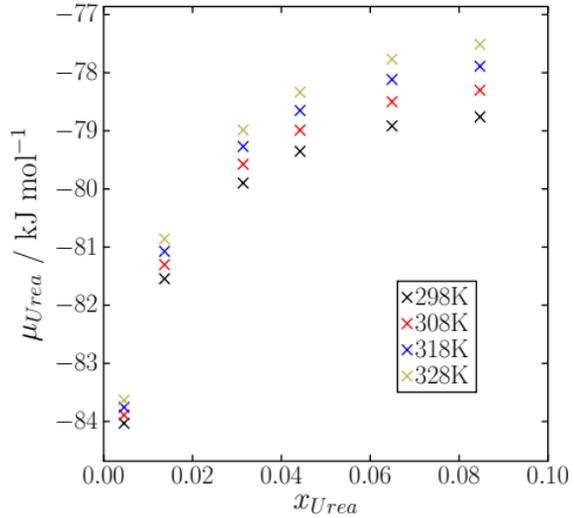
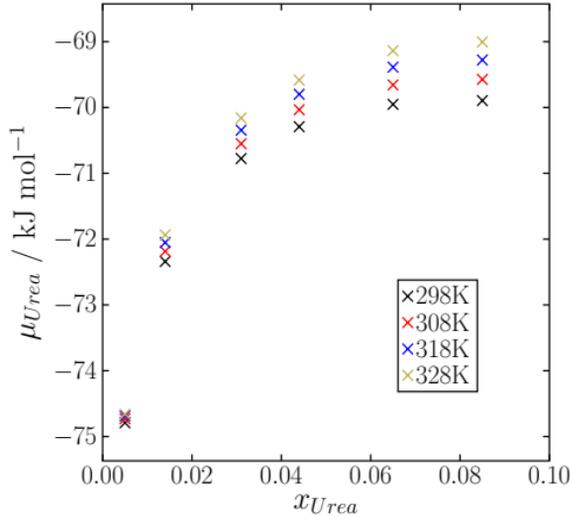
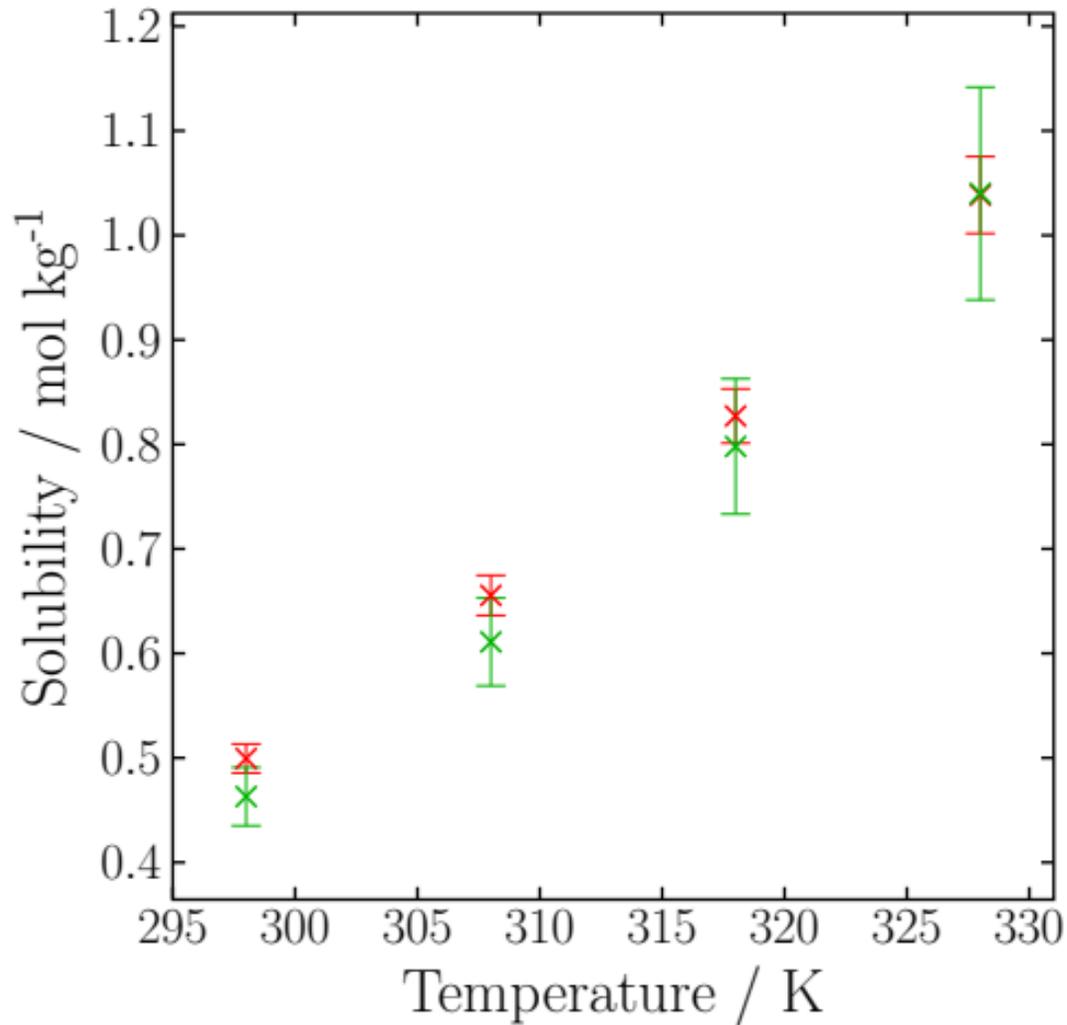
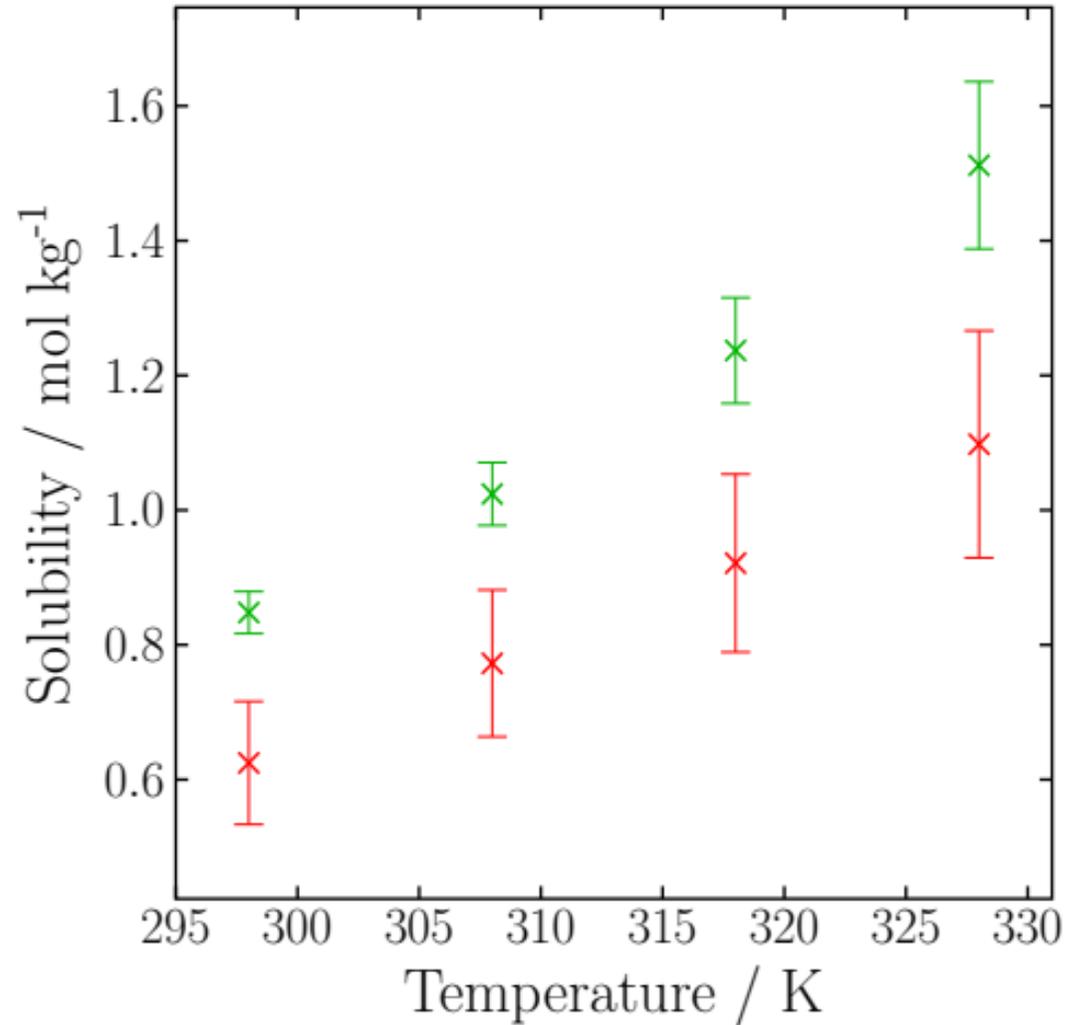


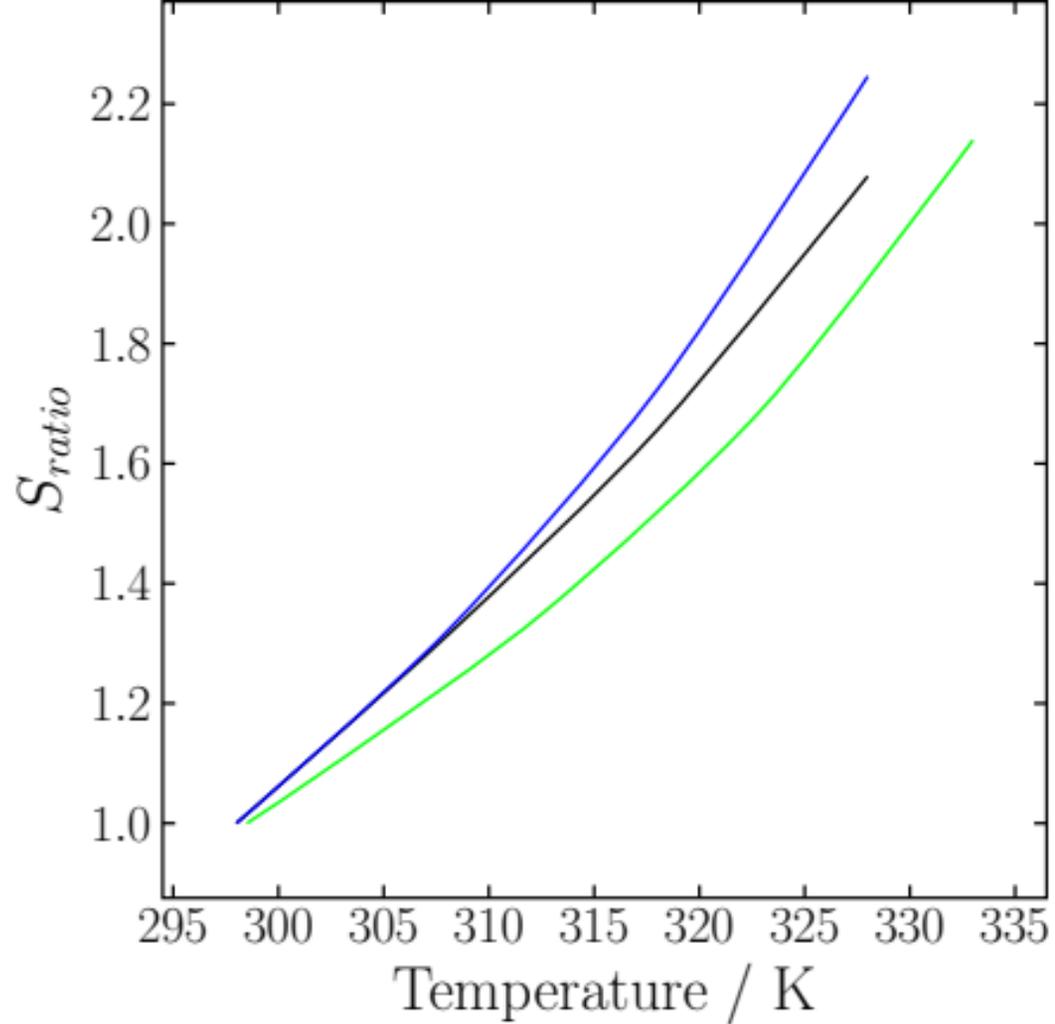
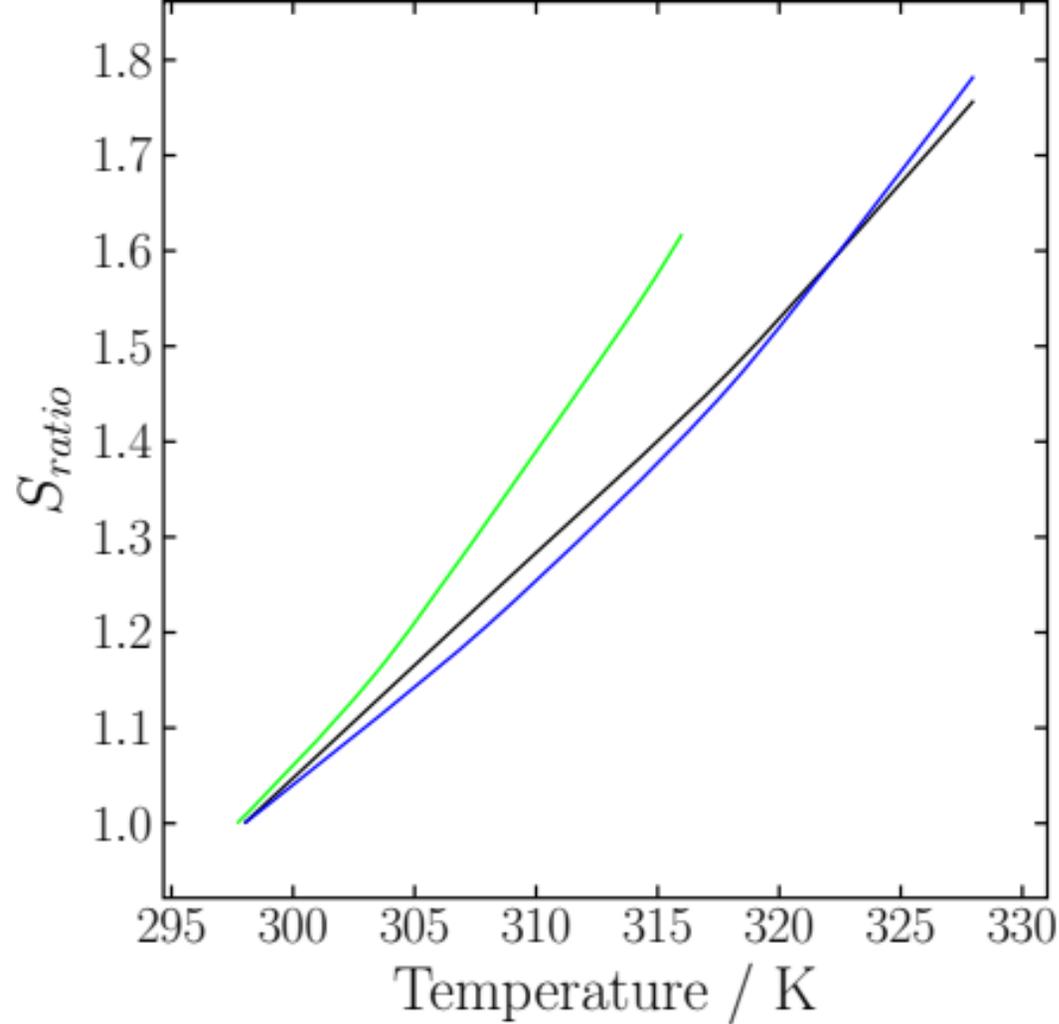
$\ln \Omega(E, V)$











Solubility prediction for a soluble organic molecule via chemical potentials from density of states

Simon Boothroyd and Jamshed Anwar *

Chemical Theory and Computation, Department of Chemistry, Lancaster University, Lancaster
LA1 4YB U.K.

*Corresponding author: j.anwar@lancaster.ac.uk

ABSTRACT

Whilst the solubility of a substance is a fundamental property of widespread significance, its prediction from first principles (starting from only the knowledge of the molecular structure of the solute and solvent) remains a challenge. Recently, we proposed a robust and efficient method to predict the solubility from the density of states (DOS) of a solute-solvent system using classical molecular simulation. The efficiency, and indeed the generality, of the method has now been enhanced by extending it to calculate solution chemical potentials (rather than probability distributions as done previously), from which solubility may be accessed. The method has been employed to predict the chemical potential of Form 1 of urea in both water and in methanol for a range of concentrations at ambient conditions, and for two charge models. The chemical potential calculations were validated by thermodynamic integration with the two sets of values being in excellent agreement. The solubility determined from the chemical potentials for urea in water ranged from 0.46-0.50 mol kg⁻¹, whilst that for urea in methanol ranged from 0.62-0.85 mol kg⁻¹, over the temperature range 298-328 K. In common with other recent studies of solubility prediction

from molecular simulation, the predicted solubilities differ markedly from experimental values, reflecting limitations of current forcefields.

1. INTRODUCTION

Solubility is a fundamental property in chemistry, arising from a complex interplay of solute–solute, solute–solvent and solvent–solvent interactions. It is of considerable importance across a spectrum of physico-chemical phenomena with application domains including the structure and organisation of biomolecules in the body as well as deposition of bone and pathologies such as amyloid formation, geology, materials development, toxicology, food processing, the oil industry¹, pharmaceutical development,² and the fate of pollutants to name a few. Consequently, for any interventions, the ability to accurately and efficiently predict the solubility would be a significant utility. Such a predictive ability would also give access to solubilities at conditions inaccessible to experiment e.g. at high temperatures and pressures, or for compounds which are difficult or dangerous to handle manually. Further, for some applications, e.g. the development of pharmaceuticals, there is a need to predict the solubility of molecules that have yet to be synthesised, though this would first require the prediction of the crystalline structure which is becoming feasible³.

A route to predicting solubility from molecular simulation can employ a direct coexistence approach, wherein we monitor either dissolution from a crystal surface or the growth of a crystal surface exposed to a supersaturated solution, to equilibrium⁴. Whilst this is promising, there are limitations, a key one being the relatively long simulation time required to reach equilibrium (on the order of microseconds)⁴ which is barely accessible. In contrast, the chemical potential route to solubility prediction is more robust and efficient. At the solubility limit, the chemical potential of the solute in solution $\mu_{solution}(T, p)$ and that in the solid phase $\mu_{solid}(T, p)$ are equal

$$\mu_{solid}(T, p) = \mu_{solution}(T, p) \quad - \text{Equation (1)}$$

where T and p are temperature and pressure respectively. While μ_{solid} is readily calculated by employing the Einstein molecule⁵ (or crystal⁶) method described below, calculating $\mu_{solution}$ is typically more involved. The objective, therefore, is to ascertain the solution concentration when the identity in Equation (1) holds. Determining the concentration at a given chemical potential can in principle be achieved by performing simulations in the grand or semi-grand canonical ensemble⁷. In the grand-canonical approach the solute molecules are inserted (or grown) or removed (or gradually annihilated) from the solvent, or the solvent molecule is transformed into a solute until equilibrium is achieved. While this method may work well for spherical solutes⁷⁻⁹ or where the solvent and solute molecules are structurally similar and can be readily transformed, large flexible solute molecules that bear little resemblance to the solvent pose a challenge. The alternative approach, determining the chemical potential as a function of concentration and looking for intersection with the chemical potential of the solid¹⁰ is more general and established. The methods here include thermodynamic integration¹¹⁻¹⁵ (TI), perturbation¹⁶⁻¹⁹ or expanded ensemble calculations^{20,21}. These calculations, however, are very demanding in terms of compute resource. For example, TI requires dozens of simulations to calculate a single chemical potential value for one particular concentration at a specified (T, p) . Such a calculation would then need to be repeated for each concentration, for the particular (T, p) of interest. Further, these methods too are challenged by larger molecules, although the use of soft-core potentials²² or the recently employed cavity method¹³ go some way in overcoming this.

Recently we proposed a novel method to predict solubility directly from a system's density of states (DOS) that, in principle, can deal with larger molecules and enable solubilities to be calculated for a range of temperatures and pressures from a single DOS calculation²³. The DOS

gives access to most properties of a system, including the probability of the system existing at different concentrations as a function chemical potential from which solubility can be determined. The approach employed a variant of the Wang—Landau algorithm^{24,25}, where solution simulations are bridged to the vapour phase for the required insertion/deletion moves, so that insertion of even large molecules may be facilitated. Further, as the density of states is independent of temperature and pressure, the DOS gives access to solubility for wide range of conditions from a single DOS calculation. The method was employed to predict the solubility of NaCl²³.

In this contribution we extend the DOS-based solubility prediction methodology, moving the focus from calculating co-existence distributions to a more efficient approach of predicting solubility from chemical potentials calculated from DOS (effectively switching the independent variables). In the original DOS solubility approach, one identifies the location of the probability distribution in the discrete solution concentration space (N , $N+1$, $N+2$, $N+3$... systems) at a particular chemical potential – the chemical potential of the solid phase. To accurately capture this distribution, the DOS must be determined for all concentrations that have a non-zero probability of existing at the given chemical potential. When the solubility limit is unknown *a priori*, it is then necessary to include a large spectrum of discrete solute concentrations within the DOS calculation as one does not know the location of the probability distribution in concentration space. Much of this information, however, is redundant, since the important concentrations are only those that contribute to the distribution peak identifying the solubility concentration. In the original DOS-based solubility study, we exploited prior knowledge of the solubility of the NaCl model and covered a selective region of the concentration space with 12 discrete concentrations.

Here we reformulate the DOS-based solubility prediction method to calculate chemical potentials as a function of concentration, rather than the other way around. The free energy of

solution of a given concentration is directly accessible (to within an additive constant) from its density of states as calculated by our approach. If several such free energies are determined, an analytical function may be fitted to these as a function of solute concentration, with the chemical potential being simply its derivative. While a certain number of free energies (equating to the number of discrete concentrations) must be determined to produce an accurate fit, the number of required determinations are in general substantially fewer than would be required for the distribution route. It is pertinent to note that while this method is presented in the context of calculating solution chemical potentials and predicting solubilities, it is in fact a general approach for calculating the chemical potential of pure fluid phases. We demonstrate and apply the method to predict the solubility of Form I of the organic molecular crystal urea in both methanol and water for a range of temperatures. Urea was chosen as it is a small, bio-relevant molecule (being the end product of metabolism, a source of nitrogen in fertilisers, and extensively used in plastics) which has a high solubility; the high solubility makes the solubility prediction quite demanding as the solution chemical potential must be determined for a large range of concentrations. Further, the chemical potentials calculated using DOS have been validated by thermodynamic integration.

2. THEORY

Chemical potential of solution from DOS

The isothermal-isobaric partition function of a system of N_{sol} solute molecules (each comprising n_{sol} solute atoms) and N_{solv} solvent molecules (each comprising n_{solv} solvent atoms) is given by

$$\begin{aligned}
 & Q(T, p, N_{sol}, N_{solv}) \\
 &= \frac{\Lambda_{sol}^{-3n_{sol}N_{sol}} \Lambda_{solv}^{-3n_{solv}N_{solv}}}{(n_{sol}N_{sol})! (n_{solv}N_{solv})!} \\
 & \times \int e^{-\beta pV} dV \int e^{-\beta U(\mathbf{r}_{sol}^{n_{sol}N_{sol}}, \mathbf{r}_{solv}^{n_{solv}N_{solv}})} d\mathbf{r}_{sol}^{n_{sol}N_{sol}} d\mathbf{r}_{solv}^{n_{solv}N_{solv}}
 \end{aligned}
 \tag{2}$$

where the configurational integral is over all solute (\mathbf{r}_{sol}) and solvent (\mathbf{r}_{solv}) atomic positions and all momentum degrees of freedom have been integrated out analytically and appear as the solute (Λ_{sol}) and solvent (Λ_{solv}) *de Broglie* wavelengths.

Given the degeneracy of the configurational microstates, Equation 2 can be rewritten as

$$Q(T, p, N_{sol}, N_{solv}) = \frac{1}{\Lambda_{sol}^{3n_{sol}N_{sol}} \Lambda_{solv}^{3n_{solv}N_{solv}}} \sum_V \sum_E \Omega_{conf}(V, E) e^{-\beta(E+pV)} \tag{3}$$

where Ω_{conf} is the configurational density of states²⁵

$$\begin{aligned}
 \Omega_{conf}(V, E) &= \frac{1}{(n_{sol}N_{sol})! (n_{solv}N_{solv})!} \\
 & \times \int \delta(E - U(\mathbf{r}_{sol}^{n_{sol}N_{sol}}, \mathbf{r}_{solv}^{n_{solv}N_{solv}})) d\mathbf{r}_{sol}^{n_{sol}N_{sol}} d\mathbf{r}_{solv}^{n_{solv}N_{solv}}
 \end{aligned}
 \tag{4}$$

and the integrals over all states have been replaced with summations over all volumes and energy levels, and δ is the Dirac delta function. In the case where the solute is treated as a rigid body (as is the case in this study), the configurational density of states becomes

$$\begin{aligned}
 \Omega_{conf} &= \frac{1}{N_{sol}! (n_{solv}N_{solv})!} \\
 & \times \int \delta(E - U(\mathbf{R}_{sol}^{N_{sol}}, \boldsymbol{\theta}_{sol}^{N_{sol}}, \mathbf{r}_{solv}^{n_{solv}N_{solv}})) d\mathbf{R}_{sol}^{N_{sol}} d\boldsymbol{\theta}_{sol}^{N_{sol}} d\mathbf{r}_{solv}^{n_{solv}N_{solv}}
 \end{aligned}
 \tag{5}$$

where the configurational integral is now over the solute molecule position (\mathbf{R}_{sol}) and orientation ($\boldsymbol{\theta}_{sol}$), and over atomic coordinates of the solvent molecules.

The free energy of such a system of solute and solvent is related to its partition function by

$$G(T, p, N_{sol}, N_{solv}) = -\frac{1}{\beta} \ln Q(T, p, N_{sol}, N_{solv}) = -\frac{1}{\beta} \ln \left[\frac{1}{\Lambda_{sol}^{3n_{sol}N_{sol}} \Lambda_{solv}^{3n_{solv}N_{solv}}} \sum_V \sum_E \Omega_{conf}(V, E) e^{-\beta(E+pV)} \right] \quad \text{- Equation (6)}$$

Given that the density of states is independent of temperature and pressure, Equation 6 can in principle be used to determine the free energy for a range of temperatures and pressures, all from a single density of states calculation. Should the free energy be determined for a series of concentrations (enforcing the condition that number of solvent particles is fixed, and only the number of solute particles is allowed to vary), the solution chemical potential is found by fitting a polynomial (or another such analytical function) as a function of N_{sol} , and analytically taking the derivative. As noted by Vega *et al*¹⁰, a more accurate fit can be achieved by splitting the free energy into an ideal (G_{id}), and an excess (G_{ex}) component $G = G_{id} + G_{ex}$ and fitting to the excess, rather than full free energy. The rationale behind this is that at low concentrations, the free energy profile is dominated by the log term of the ideal free energy, while the excess free energy varies more smoothly. For the systems studied here, we found that the excess free energy can be fitted to a good approximation to a second order polynomial, such that

$$G_{ex} = a_0 N_{sol}^2 + a_1 N_{sol} + a_2 \quad \text{- Equation (7)}$$

where a_0 , a_1 and a_2 are coefficients to be determined by least squares fitting. While we believe that this will generally be a good choice, care must be taken that the fitted function does indeed capture the behaviour of the excess free energy. The excess chemical potential is then

$$\mu_{ex} = 2a_0 N_{sol} + a_1 \quad \text{- Equation (8)}$$

and the full chemical potential is recovered by

$$\mu = \mu_{ex} + \frac{1}{\beta} \ln \frac{N_{sol}}{V} \quad \text{- Equation (9)}$$

where the righthand term is the ideal chemical potential. Here the *de Broglie* wavelength of the solute has been chosen to be unity for convenience, as it is common to both the solid and the solute in solution. Further, the *de Broglie* wavelength of the solvent is not present in Equation 9 provided that the number of solvent molecules remains constant, as is the case in this study.

The challenge then is to calculate the density of states of the solute and solvent system for a range of concentrations. This is accomplished by the 3-d DOS method developed by us previously²⁶. For completeness, we present the salient aspects of the approach here (see also Figure 1). The key feature is that the particle insertion/deletion moves that link the N , $N+1$, $N+2$ $N+..$ systems are carried out in the gas phase with a bridge to the discrete solution systems. There are 4 components to the DOS determination which are then stitched together:

- i) DOS of the solute in solution for each concentration of interest, ensuring that the selected windows cover the energies and volumes accessible to the temperatures and pressures of interest for each concentration.
- ii) DOS for each concentration of interest in windows that *extend the energies* to those accessible to the system in a supercritical state.
- iii) DOS of the solute in solution for each concentration of interest in windows limited to energies in the supercritical state that *extend the volume* of those windows into the gas phase.
- iv) A single 2-d DOS calculation in the gas phase involving multiple systems of varying solute concentration connected by grand-canonical like insertion/deletion moves between the different concentration windows (Figure 1).

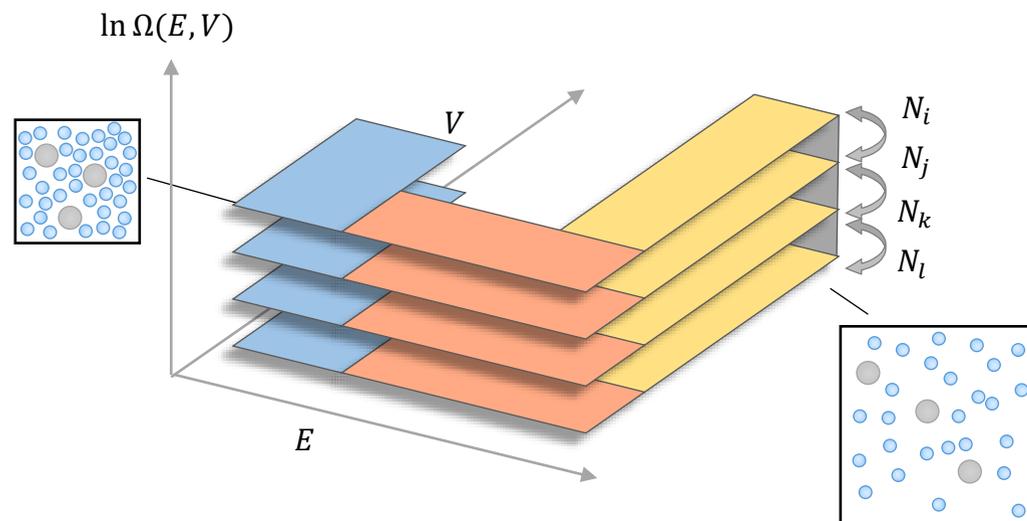


Figure 1. 3-d DOS for a solute-solvent system. The DOS of the solute in solution is calculated for each concentration of interest ($N_i, N_j, N_k, N_l, \dots$) over the range of energies and volumes that would be accessible to the system over the temperature and pressure ranges of interest (blue regions). A second set of DOS windows are calculated, which extend the energy range sampled in the original windows to energies which would be accessible at a temperature and pressure above the critical point (orange regions). A third set of DOS is calculated over windows which extend the volume range sampled in the supercritical state to also cover the volume corresponding to low-density gas phase (yellow regions). Finally, in the gas phase, a single DOS of the full set of concentration windows employing solute insertion/deletion moves to transition between the concentrations (grey regions).

The advantage of calculating the DOS in this way is two-fold: firstly, by first transitioning the system to a supercritical state we bypass a first order transition which would introduce irreversibility and associated hysteresis; secondly, as the insertion/deletion moves are performed in the gas phase, such moves become feasible for larger molecules (with scope to simply reduce the gas density as required to facilitate move acceptance).

Chemical potential of solution from TI

As a check on the DOS calculation, the solution chemical potentials of urea in water and urea in methanol were also calculated using the established TI approach. TI calculations extract free energy as the reversible work in going from one state to another, where the progress of the transition between states is controlled by a coupling parameter λ

$$\Delta G(A \rightarrow B) = \int \frac{\delta G(\lambda)}{\delta \lambda} d\lambda = \int \langle \frac{\delta U(\lambda)}{\delta \lambda} \rangle d\lambda \quad - \text{Equation (10)}$$

where the angular brackets represent an ensemble average and we assume that the kinetic component of the partition function is unperturbed. If U varies linearly with respect to lambda i.e. $U(\lambda) = U_A(1 - \lambda) + U_B(\lambda)$, Equation 10 becomes

$$\Delta G(A \rightarrow B) = \int \langle U_B(\lambda) - U_A(\lambda) \rangle d\lambda \quad - \text{Equation (11)}$$

For calculating solution chemical potentials, the solute molecule starts in an ideal state, and is gradually grown within the solution using the λ parameter to switch on the interactions between the emerging solute molecule and the rest of the system²⁷.

Chemical potential of solid

The chemical potential of solid urea was calculated using the Einstein molecule method. The reference state in this calculation is an ideal Einstein lattice of (rigid) urea molecules, which are restrained to their lattice sites by harmonic potentials of the form

$$U_{pos} = \sum_i^{N_{solid}} k(\mathbf{r}_{i,C} - \mathbf{r}_{i,C,0})^2 \quad - \text{Equation (12)}$$

where N_{solid} is the number of molecules in the solid, k is the spring stiffness, and $\mathbf{r}_{i,C}$ and $\mathbf{r}_{i,C,0}$ are the instantaneous and mean lattice positions of the carbon atom of urea molecule i , respectively. Here the restraints are attached to the central carbon atom as a good approximation

of the molecule's centre of mass. To prevent diffusion of the centre of mass of the system, the position of one of the carbon atoms in the system is kept fixed. The chemical potential of this Einstein molecule reference state is given by²⁸

$$\beta\mu_0 = \frac{3}{2} \left(1 - \frac{1}{N_{solid}}\right) \ln \frac{\beta k}{\pi} + \frac{1}{N_{solid}} \ln \frac{N_{solid}}{V_{solid}} \quad - \text{Equation (13)}$$

where V_{solid} is the volume of the solid. The *de Broglie* wavelength terms were again chosen here to be unity as they are common to both the solid and the solute in solution.

The above reference state is transformed into the full, unrestrained crystal by three successive steps, such that the total chemical potential of the crystal is given by

$$\mu_{solid} = \mu_0 + \Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 + \mu_{sym} \quad - \text{Equation (14)}$$

The first stage is the chemical potential difference for restraining the orientation of the urea molecules in the Einstein molecule lattice using two additional tethers per urea molecule

$$U_{or} = \sum_i^{N_{or}} k(\mathbf{r}_i - \mathbf{r}_{i,0})^2 \quad - \text{Equation (15)}$$

where N_{or} is the number of orientational restraints introduced. In the case of urea, these restraints are attached to each of the nitrogen atoms. The change in the chemical potential associated with this step is then calculated by thermodynamic integration (Equation 10) as outlined earlier. The restraint spring constant was scaled directly yielding

$$\Delta\mu_1 = \frac{1}{N_{solid}} \int_0^k \langle \sum_i^{N_{or}} k(\mathbf{r}_i - \mathbf{r}_{i,0})^2 \rangle_k dk \quad - \text{Equation (16)}$$

We introduced an extra constant c and changed the limits of the integral to improve the accuracy as proposed by others^{6,29}

$$\Delta\mu_1 = \frac{1}{N_{solid}} \int_{\ln(c)}^{\ln(k+c)} \langle \sum_i^{N_{or}} (\mathbf{r}_i - \mathbf{r}_{i,0})^2 \rangle_k (k+c) d \ln(k+c) \quad - \text{Equation (17)}$$

In the second stage we introduce the intermolecular interactions. The difference between the chemical potential of the ideal, non-interacting crystal and the fully interacting one ($\Delta\mu_2$) was calculated by free energy perturbation

$$\beta\Delta\mu_2 = \frac{1}{N_{solid}} (\beta U_{lattice} - \ln\langle e^{-\beta(U_{solid}-U_{lattice})} \rangle) \quad - \text{Equation (18)}$$

where the average is evaluated over configurations sampled employing the ideal potential energy function (i.e. one that only evaluates the tethered and intramolecular interactions), U_{solid} is the instantaneous energy of the solid evaluated using the full system potential energy function and $U_{lattice}$ is the energy of the perfect lattice. The final step involved removing all restraints from the system, whilst calculating the corresponding change in chemical potential $\Delta\mu_3$ by thermodynamic integration

$$\Delta\mu_3 = -\frac{1}{N_{solid}} \int_{\ln(c)}^{\ln(k+c)} \langle \sum_i^{N_{tethers}} (\mathbf{r}_i - \mathbf{r}_{i,0})^2 \rangle_k (k+c) d \ln(k+c) \quad - \text{Equation (19)}$$

where $N_{tethers}$ is the total number of restrained atoms. The last term in Equation 14, μ_{sym} , accounts for the orientation field not having the same symmetry as the molecule of interest⁵. As urea has a point group of C_{2v} , $\beta \Delta\mu_{sym} = -\ln 2$.

3. TECHNICAL DETAILS

The solubility of urea in methanol, and urea in water was calculated as a function of temperature using the above methodology. The methanol solubility calculations employed 125 methanol and between 1 and 20 urea molecules, spanning a concentration range of ~ 0.25 - 5.00 mol kg^{-1} . For the aqueous solubility, we employed 216 water molecules and between 1 and 20 urea molecules, spanning a concentration range of ~ 0.26 - 5.14 mol kg^{-1} . The urea, methanol and water molecules were modelled using the Amber GAFF force field³⁰ and the TIP3P water model³¹. Two sets of partial charges were employed for the urea molecule – the Caldwell set originally shipped with

Amber³⁰ (referred to as set **A**), and a reportedly improved set³² fitted against properties of solid urea (referred to as set **B**). Both charge sets are given in Table 1.

Table 1. The Caldwell (set **A**) and improved charges (set **B**) employed for urea.

GAFF Atom Type	Set A / <i>e</i>	Set B / <i>e</i>
o	-0.612	-0.660
n	-0.924	-0.888
c	0.880	0.884
hn	0.395	0.388

The urea molecule was treated as a rigid body, an approximation that enables more efficient sampling of phase space without markedly compromising the accuracy of the calculations. This is a fair assumption given that the urea molecule is a relatively rigid molecule. The geometry of the rigid urea molecules using charge sets **A** and **B** was constructed using the bonded parameters provided by GAFF³⁰ and those provided in ref 29 respectively. All van der Waals interactions were truncated after 0.85 nm and the standard long-range correction applied, while Ewald summations with a direct space cut off of 0.85 nm and a precision of $1.0e^{-5}$ were employed for the electrostatic interactions. For the density of states calculations, an energy bin size of 10 kJ mol^{-1} was used, while a logged volume bin size of 0.008 and 0.011 was used for the methanol and aqueous systems respectively. The Wang-Landau modification factor was initially set to 1.0 and was allowed to decrease to 1.907×10^{-6} at which point the results were well converged. For the thermodynamic integration calculations, a general scheme utilising soft core potentials²² was followed²⁷. A 16-point gaussian quadrature was employed to evaluate both the van der Waal and Coulomb integrals. All solution-phase calculations were performed using an in-house Monte Carlo simulation code.

The chemical potential of the solid phase of urea was calculated at 298, 308, 318, 328 K using a simulation cell comprising $4 \times 4 \times 4$ unit cells with periodic boundaries. The structure of crystalline

urea Form I was taken from the Cambridge Structural Database³³ (reference code UREAXX29). Equilibration of the crystal at each temperature of interest was performed using the DLPOLY 4.07³⁴ molecular dynamics (MD) simulation code in the constant stress ensemble ($N\sigma T$) in which the cell vectors and angles are allowed to vary. These simulations were run for 100,000 steps with a timestep of 5 fs using the Nose-Hoover thermostat (coupling constant $\tau_T = 0.1\text{ps}$) and barostat (coupling constant $\tau_p = 1.0\text{ps}$). In each case, the simulation box angles, while permitted to change, remained orthogonal to a good degree. The Einstein-crystal thermodynamic integration simulations were performed using our in-house Monte Carlo code. The starting configuration for the Einstein crystal calculations was a perfect lattice constructed using the equilibrated cell dimensions for each respective temperature. Spring constants of 8000 kT \AA^{-2} and 10000 kT \AA^{-2} were used for calculations employing charge models **A** and **B** respectively. The integrals in Equations 17 and 19 were evaluated using a 32- and 16- point Gauss-Legendre quadrature respectively.

4. RESULTS AND DISCUSSION

The density of states of urea in water, and of urea in methanol were calculated for a range of concentrations. These were then employed to calculate the solute free energy (Equation 6), and hence the solution chemical potentials, by employing the fitting procedure outlined in Section 2. The coefficients produced by the fitting procedure are presented in Tables 2 and 3.

Table 2. The coefficients calculated by fitting the excess free energies (calculated by the DOS approach) of urea in water to Equation 8 for charge sets **A** and **B**.

<i>T</i>	$a_0 / \text{kJ mol}^{-1}$		$a_1 / \text{kJ mol}^{-1}$		$a_2 / \text{kJ mol}^{-1}$	
	A	B	A	B	A	B
298	-0.055 (5)	-0.045 (2)	-52.89 (11)	-62.15 (5)	-328.82 (45)	-147.37 (22)

308	-0.054 (5)	-0.043 (2)	-52.08 (12)	-61.26 (5)	-392.45 (49)	-213.21 (22)
318	-0.054 (6)	-0.041 (3)	-51.28 (13)	-60.37 (5)	-455.67 (54)	-278.37 (22)
328	-0.053 (7)	-0.041 (3)	-50.49 (14)	-59.49 (6)	-518.31 (59)	-343.12 (24)

Table 3. The coefficients calculated by fitting the excess free energies (calculated by the DOS approach) of urea in methanol to Equation 8 for charge sets **A** and **B**.

<i>T</i>	$a_0 / \text{kJ mol}^{-1}$		$a_1 / \text{kJ mol}^{-1}$		$a_2 / \text{kJ mol}^{-1}$	
	A	B	A	B	A	B
298	-0.049 (3)	-0.053 (11)	-53.80 (6)	-62.16 (23)	-305.57 (23)	-136.03 (96)
308	-0.050 (3)	-0.052 (10)	-52.77 (7)	-61.11 (22)	-372.22 (28)	-204.76 (90)
318	-0.051 (4)	-0.051 (10)	-51.77 (8)	-60.06 (21)	-438.34 (35)	-272.43 (87)
328	-0.052 (4)	-0.052 (10)	-50.81 (10)	-59.02 (21)	-503.90 (40)	-339.48 (87)

The chemical potentials of urea in methanol and in water as a function of urea concentration at 298 K are tabulated in Tables 4 and 5 and shown graphically in Figure 2. In addition, the chemical potentials of both systems were determined by thermodynamic integration and are also shown in Figure 2.

Table 4. The ideal, excess and total solution chemical potentials of urea in water as a function of increasing concentration using charge sets **A** and **B** calculated at 298K. The slight difference in μ_{id} arises from models **A** and **B** having slightly different volumes.

N_{solute}	$\mu_{id} / \text{kJ mol}^{-1}$		$\mu_{ex} / \text{kJ mol}^{-1}$		$\mu / \text{kJ mol}^{-1}$	
	A	B	A	B	A	B
1	-21.79	-21.79	-53.00 (11)	-62.24 (5)	-74.79 (11)	-84.03 (5)
3	-19.12	-19.13	-53.22 (11)	-62.42 (5)	-72.34 (11)	-81.54 (5)
7	-17.12	-17.12	-53.66 (13)	-62.78 (6)	-70.78 (13)	-79.90 (6)
10	-16.31	-16.31	-53.98 (15)	-63.04 (7)	-70.29 (15)	-79.35 (7)
15	-15.42	-15.42	-54.53 (19)	-63.49 (9)	-69.95 (19)	-78.92 (9)
20	-14.82	-14.82	-55.08 (23)	-63.94 (11)	-69.90 (23)	-78.76 (11)

Table 5. The ideal, excess and total solution chemical potentials of urea in methanol as a function of increasing concentration using charge sets **A** and **B** calculated at 298K. The slight difference in μ_{id} arises from models **A** and **B** having slightly different volumes.

N_{solute}	$\mu_{id} / \text{kJ mol}^{-1}$		$\mu_{ex} / \text{kJ mol}^{-1}$		$\mu / \text{kJ mol}^{-1}$	
	A	B	A	B	A	B
1	-22.34	-22.34	-53.90 (6)	-62.27 (23)	-76.24 (6)	-84.60 (23)
3	-19.65	-19.66	-54.09 (6)	-62.48 (24)	-73.75 (6)	-82.13 (24)
7	-17.61	-17.62	-54.49 (7)	-62.9 (28)	-72.10 (7)	-80.53 (28)
10	-16.78	-16.79	-54.78 (8)	-63.22 (32)	-71.56 (8)	-80.01 (32)
15	-15.86	-15.87	-55.27 (10)	-63.75 (40)	-71.13 (10)	-79.62 (40)
20	-15.23	-15.24	-55.76 (12)	-64.28 (49)	-71.00 (12)	-79.52 (49)

As can be seen in Figure 2, there is an excellent agreement between the DOS-based calculation and the TI methodology for both solvent systems and both charge sets, within the uncertainties of each method. This good agreement gives confidence that the DOS approach is able to accurately calculate the chemical potential of molecular systems, given that it was first tested on the simple ionic system NaCl^{23} . Although urea is a relatively small molecule, the vapourisation pathway employed in the DOS approach appears to transfer well to molecules, without any modification. As previously anticipated, our current experience suggests that the DOS approach should scale further to more complex and flexible organic molecules without any significant technical difficulties.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.5117281

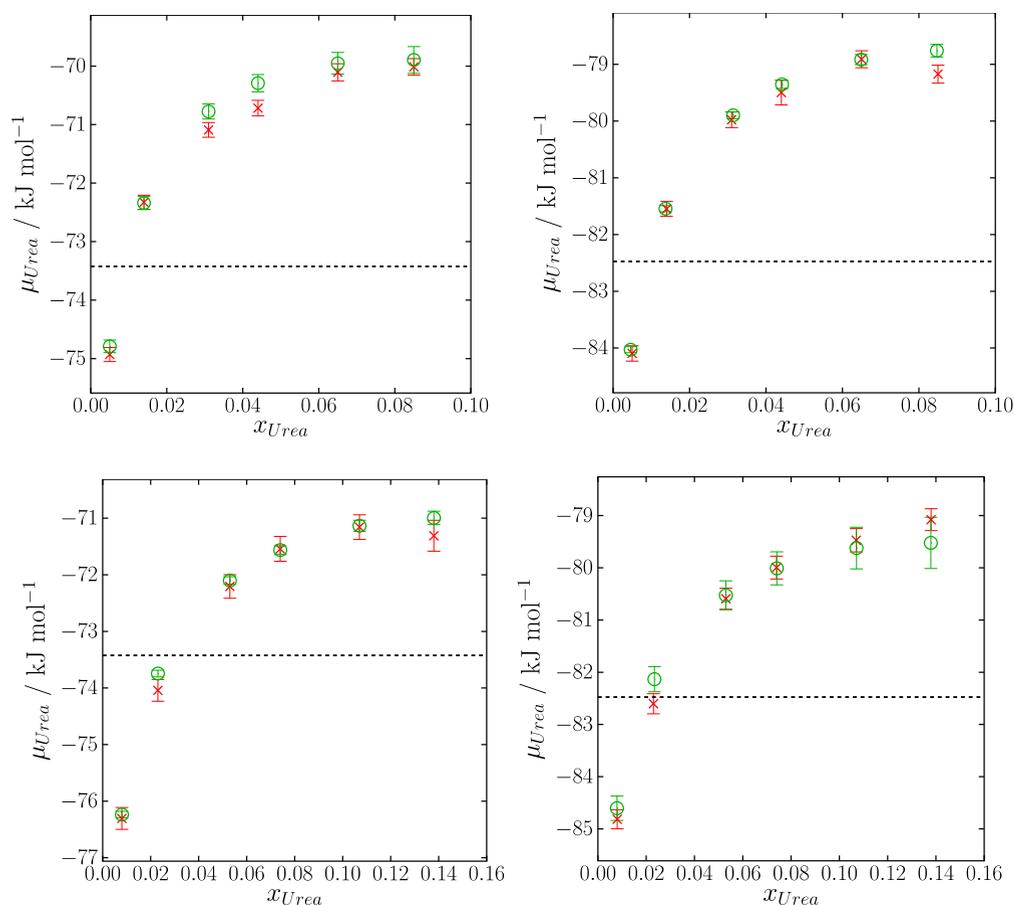


Figure 2. Chemical potential (total) of urea in water using charge sets **A** (top left) and **B** (top right), and of urea in methanol using charge sets **A** (bottom left) and **B** (bottom right) as a function of urea mole fraction (x_{Urea}). DOS results are plotted as green data points, whilst thermodynamic integration results are plotted as red crosses. The dashed horizontal line represents the chemical potential of the solid phase at 298 K, calculated using the Einstein molecule method.

For both solvent systems, the solution chemical potentials as a function of urea concentration were further calculated at 308 K, 318 K and 328 K by reweighting the DOS used in the 298 K calculations (Figure 3). As would be expected, the chemical potential is seen to increase smoothly as a function of temperature for all systems considered.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.5117281

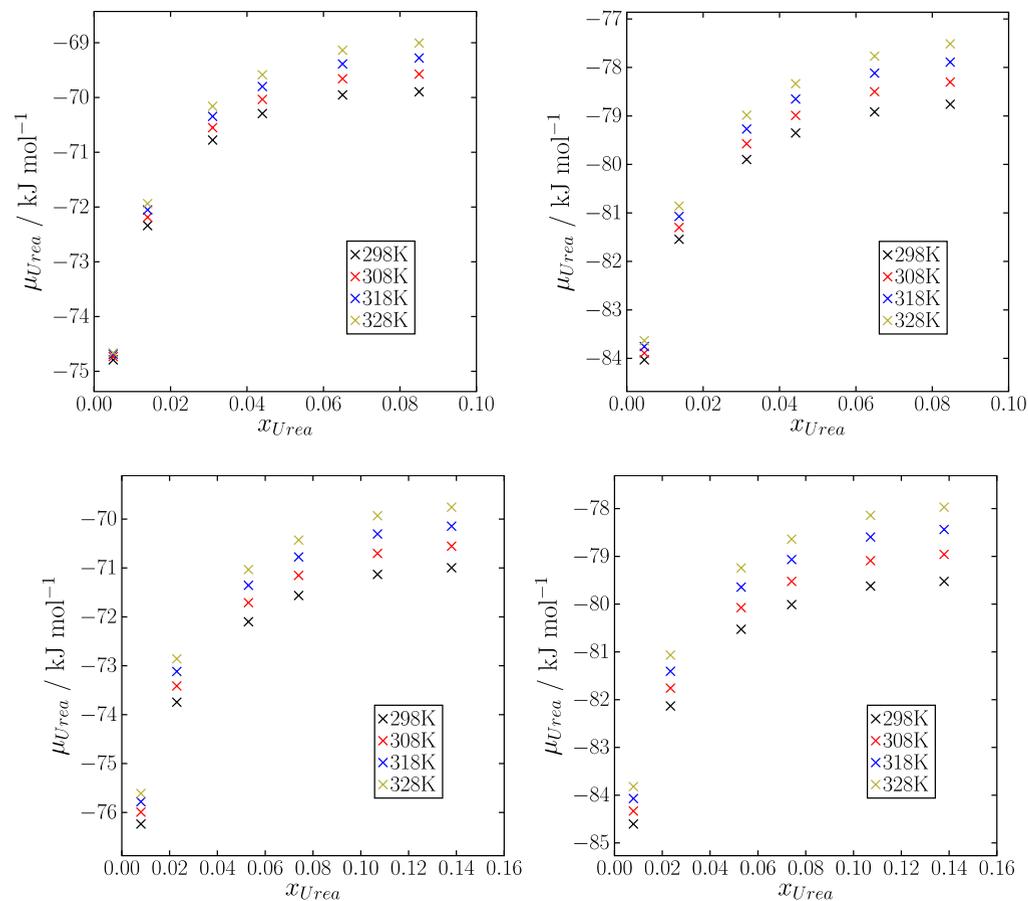


Figure 3. The chemical potentials of urea in water using charge sets **A** (top left) and **B** (top right) and of urea in methanol using charge sets **A** (bottom left) and **B** (bottom right) as a function of urea mole fraction (x_{Urea}).

The chemical potential of the solid phase calculated using the Einstein molecule method as a function of temperature is presented in Tables 6 (charge set **A**) and 7 (charge set **B**).

Table 6. The individual components of the total solid phase chemical potential as a function temperature calculated by the Einstein molecule method for charge set **A**.

T	$\beta\mu_{sym}$	$\beta\mu_0$	$\beta\Delta\mu_1$	$\beta\Delta\mu_2$	$\beta\Delta\mu_3$	$\beta\mu_{solid}$
298	-1.717	28.836	45.289	-103.255	-42.587 (4)	-73.435 (4)

308	-1.775	29.803	46.805	-103.204	-44.332 (5)	-72.703 (5)
318	-1.833	30.771	48.326	-103.137	-46.101 (5)	-71.974 (5)
328	-1.890	31.738	49.841	-103.069	-47.871 (5)	-71.251 (5)

Table 7. The individual components of the total solid phase free energies as a function of temperature calculated by the Einstein molecule method for charge set **B**.

T	$\beta\mu_{sym}$	$\beta\mu_0$	$\beta\Delta\mu_1$	$\beta\Delta\mu_2$	$\beta\Delta\mu_3$	$\beta\mu_{solid}$
298	-1.717	29.658	46.313	-113.245	-43.425 (5)	-82.416 (5)
308	-1.775	30.653	47.866	-113.222	-45.161 (4)	-81.638 (4)
318	-1.833	31.649	49.417	-113.170	-46.940 (4)	-80.877 (4)
328	-1.890	32.644	50.966	-113.124	-48.713 (5)	-80.118 (5)

The solubility at each temperature was then determined by finding the point of intersection between the solid and solution curves. Rather than plotting the solid and solution chemical potential curves and interpolating to where they intersect by eye, we have instead fitted polynomial equations to these terms. Given that the excess solution chemical potential is already expressed as a polynomial (Equation 8) and the solid chemical potential is constant, only the ideal part (Equation 9) must be fit. Further, as it is only the volume in Equation 9 that is unknown as a function of solute concentration, and the volume varies much more slowly than the ideal chemical potential, only the volume was fit to a polynomial of the form

$$V_{sol} = b_0 N_{sol}^2 + b_1 N_{sol} + b_2 - \text{Equation (20)}$$

where b_0 , b_1 and b_2 are coefficients determined from least squares fitting (Tables 8 and 9).

Table 8. The coefficients calculated by fitting the solution phase volumes (calculated from the DOS) of urea in water, to Equation 20 for charge sets **A** and **B**.

T	$b_0 / \text{\AA}^3$		$b_1 / \text{\AA}^3$		$b_2 / \text{\AA}^3$	
	A	B	A	B	A	B
298	0	0	68.98 (20)	69.08 (50)	6534 (2)	6541 (6)
308	0	0	69.47 (11)	69.56 (26)	6593 (1)	6595 (3)

318	0	0	69.88 (12)	70.06 (15)	6658 (1)	6659 (2)
328	0	0	70.22 (17)	70.49 (11)	6729 (2)	6727 (1)

Table 9. The coefficients calculated by fitting the solution phase volumes (calculated from the DOS) of urea in methanol to Equation 20 for charge sets **A** and **B**.

<i>T</i>	$b_0 / \text{\AA}^3$		$b_1 / \text{\AA}^3$		$b_2 / \text{\AA}^3$	
	A	B	A	B	A	B
298	0.30 (4)	0.14 (8)	53 (1)	58 (2)	8187 (4)	8181 (7)
308	0.31 (7)	0.22 (2)	53 (1)	56	8295 (6)	8287 (2)
318	0.27 (7)	0.29 (6)	54 (1)	55 (1)	8404 (6)	8403 (5)
328	0.20 (6)	0.25 (7)	55 (1)	55 (1)	8518 (5)	8517 (6)

Combining Equations 1, 9 and 20 yields

$$2a_0N_{sol} + a_1 + \frac{1}{\beta} \ln \frac{N_{sol}}{b_0N_{sol}^2 + b_1N_{sol} + b_2} = \mu_{solid} \quad - \text{Equation (21)}$$

which can be readily solved for the solubility concentration (N_{sol}) by applying the Newton—Raphson algorithm. The solubilities calculated by this approach are tabulated in Table 10 and shown graphically in Figure 4.

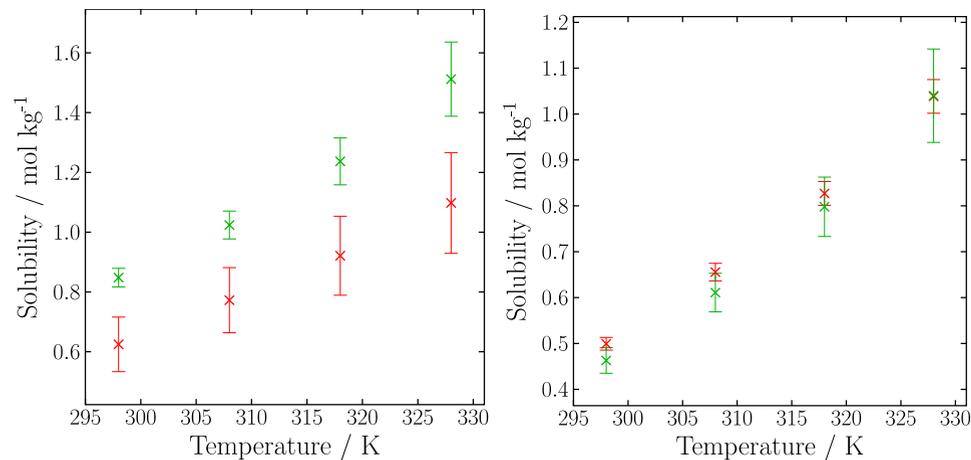


Figure 4. The solubility of urea in methanol (left) and in water (right) as a function of temperature using charge sets **A** (green points) and **B** (red points).

Table 10. The solubility of urea in water and urea in methanol for the two employed charge sets as a function of temperature calculated by the DOS approach.

<i>T</i>	Water / mol kg ⁻¹		Methanol / mol kg ⁻¹	
	A	B	A	B
298	0.46 (3)	0.50 (1)	0.85 (3)	0.62 (9)
308	0.61 (4)	0.66 (2)	1.02 (5)	0.77 (11)
318	0.80 (6)	0.83 (3)	1.24 (8)	0.92 (13)
328	1.04 (10)	1.04 (4)	1.51 (12)	1.10 (17)

For both solvent systems and for both charge sets, the calculated solubility of the model is markedly lower than experiment for all temperatures. The predicted solubility of urea in methanol at 298K is 0.85 mol kg⁻¹ and 0.62 mol kg⁻¹ for charge sets **A** and **B** respectively, which is roughly 5 fold lower than the experimental solubility of 4.01 mol kg⁻¹³⁵. Similarly, the solubility of urea in water at 298K was calculated to be 0.46 mol kg⁻¹ and 0.50 mol kg⁻¹ for charge sets **A** and **B** respectively, which is close to 45 fold lower than the experimental solubility of 20.15 mol kg⁻¹³⁵. The solubility of urea in water was little affected on changing the partial atomic charges of urea. In contrast, the solubility in methanol was roughly 1.5X larger when using charge set **A** compared to set **B**. It is unclear why the change in charge sets improves the solubility in methanol relative to experiment, and yet does nothing in water. Further, it is notable that the charge set that was explicitly derived to better reproduce the structure of solid urea yielded the least accurate set of solubilities.

The predicted solubility of urea as a function of temperature for both solvent systems increases smoothly as a function of temperature (Figure 4), as is in line with experiment³⁵. We investigated whether relative trends in solubility as a function of temperature (rather than absolute values) might be a decent match to the relative trends observed experimentally. The relative solubility was calculated as $S_{ratio}(T) = S(T)/S_{298}$ where S_{298} is the solubility at 298 K, for both the solubilities

calculated here and for an experimental data set³⁵ (Figure 5). If the trend in relative solubilities is accurately captured by the model, plots of the calculated and experimental ratios are expected to coincide. It can be seen that the curves Figure 5 are indeed somewhat similar, although not quite coincidental. The curves deviate by between 0.2-0.4% (which is significantly better than the 45-fold difference in the absolute aqueous solubilities), thus suggesting that the relative solubilities offer a better agreement than the absolute values. This opens the possibility that while the currently available models may not be able to predict solubilities in absolute terms, they may have utility at least in calculating trends in solubilities.

What might be the source of error for the gross disparity between predicted and experimental solubility values? The solution phase chemical potentials were calculated by both the DOS approach and the TI methodology, both of which were found to be in excellent agreement (to within the statistical uncertainties of each method). The solid chemical potentials were calculated using the established Einstein molecule method that we also applied to methanol and sodium chloride crystals, and reproduced results that others have reported⁵. It thus appears that the model interaction parameters may not be accurate enough. A number of other recent solubility prediction studies have published predicted solubilities differing markedly from experimental values (the most frequently reported being NaCl, whose calculated solubility is of the order of 2-10 fold less than experiment³⁶) and attribute the disparity to limitations in the existing forcefield parameters and possibly functional form^{14,37}. It appears that whilst current parameters may be able to reasonably reproduce other aspects of phase behaviour such as melting points^{13,36-38}, solid-state phase transitions³⁹, and interfacial properties^{40,41}, solubility prediction is clearly challenging current parameters. In general, the observed trend in the solubility predictions is that solubility is under-predicted suggesting something systematic may be missing from the forcefields. Perhaps

this challenging behaviour should be expected as the solubility predictions probe the second derivatives of free energy, the first derivative being the chemical potential itself (variation in free energy as a function of number molecules) and the second derivative being the variation in chemical potential as function of concentration.

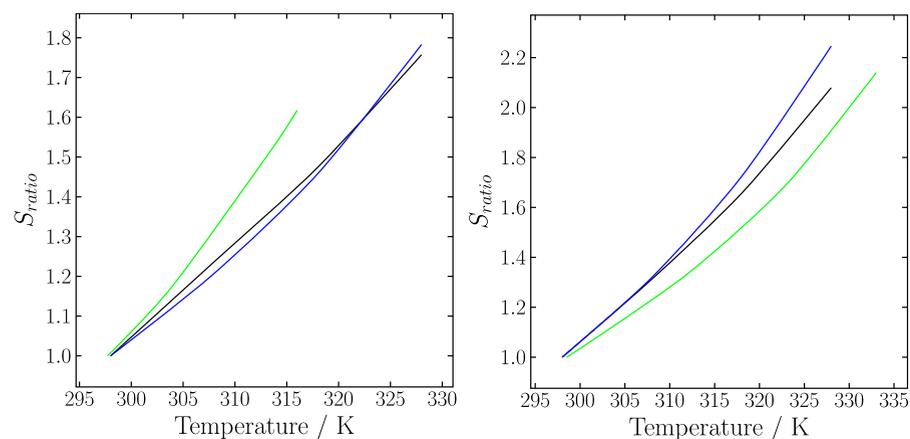


Figure 5. Predicted (blue curve: charge set A; black curve: charge set B) and experimental³⁵ (green curve) reduced solubilities (relative to the respective solubility at 298K) as a function of temperature for urea in methanol (left) and in water (right).

These results highlight the critical need to have a robust and efficient method for predicting solubility from molecular simulation, as such a method will be invaluable in not only testing, but also helping to optimise existing force fields. A higher quality of force field is clearly required if solubility predictions from classical molecular simulations are to be routinely performed (or more importantly, trusted). This must be one of the major focuses of future work. The DOS method proposed here is robust and is expected to scale to larger and more flexible molecules without much technical challenge. Should there be any difficulties it could be combined with configurational bias Monte Carlo simulation moves⁴², which would enhance sampling of the internal degrees of freedom of flexible molecules.

A key feature of the proposed method is its marked efficiency relative to both the (our) previously reported solubility prediction using density of states approach via co-existence distributions and well-established methods such as exponential averaging and thermodynamic integration. With regards to the comparison with established methods, the density of state methods have the generic advantage of being able to access chemical potentials over a range of temperatures and pressures from a single DOS calculation, whereas established methods would need the simulations to be repeated for each new condition of interest. This is of particular importance to those application domains whereby the solubility profile of a solute must be known for multiple temperatures and humidity's. A prime example is in drug development whereby pharmaceutical molecules are typically developed at ambient lab conditions (25 °C) but are exposed to elevated temperatures when stored in warmer climates, and when ingested into the body (37 °C) – the solubility over the entirety of such a range would be readily accessible from a single DOS calculation, whereas traditional techniques could require at least twice as many calculations.

With the DOS-based co-existence distribution method (our previous study), it is necessary to approximately locate the co-existence distribution in solute concentration (N_{sol})-space and then sample the DOS at all concentrations that have a non-zero contribution to the co-existence distribution. With the current (DOS-based chemical potential) approach, the number of free energy determinations need not be extensive as we fit an analytical function to determine the chemical potential as a function of concentration. For example, in the co-existence distribution study on NaCl, we carried out simulations at 11 different solute concentrations, with prior knowledge of the NaCl predicted solubility using the selected potentials. For the current urea solubility study, we were able to cover the appropriate solubility range (with decent accuracy) with simulations at only 6 solute concentrations.

In conclusion, we have demonstrated that the chemical potentials route from DOS is significantly more efficient for predicting solubility than via DOS-based co-existence distributions. This is a further enhancement in efficiency for the DOS-based solubility prediction approach. Furthermore, we were able to predict the solubilities of urea in methanol and of urea in water as a function of temperature from knowledge of a single density of states surface, the latter being estimated from a relatively modest number of simulations compared to what would be required for more traditional approaches. Both accuracy and precision have been shown to be comparable with more established methods such as thermodynamic integration. In common with other recent studies, the predicted solubilities differ markedly from experimental values, reflecting limitations of current forcefields.

AUTHOR INFORMATION

Corresponding Author

Corresponding author, E-mail: j.anwar@lancaster.ac.uk

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

S. B. would like to thank the EPSRC, Lancaster University, and AstraZeneca for his PhD CASE Studentship (EP/L504804/1). We acknowledge the use of Lancaster University's high-performance computing (HPC) facility (HEC) and the N8 (consortium of the U.K.'s northern universities) HPC facility (Polaris).

REFERENCES

- (1) Park, S. J.; Ali Mansoori, G. Aggregation and Deposition of Heavy Organics in Petroleum Crudes. *Energy Sources* **1988**, *10* (2), 109–125.
- (2) Di, L.; Fish, P. V.; Mano, T. Bridging Solubility between Drug Discovery and Development. *Drug Discov. Today* **2012**, *17* (9–10), 486–495.
- (3) Nyman, J.; Reutzler-Edens, S. Crystal Structure Prediction Is Changing from Basic Science to Applied Technology. *Faraday Discuss.* **2018**.
- (4) Espinosa, J. R.; Young, J. M.; Jiang, H.; Gupta, D.; Vega, C.; Sanz, E.; Debenedetti, P. G.; Panagiotopoulos, A. Z. On the Calculation of Solubilities via Direct Coexistence Simulations: Investigation of NaCl Aqueous Solutions and Lennard-Jones Binary Mixtures. *J. Chem. Phys.* **2016**, *145* (15), 154111.
- (5) Aragonés, J. L.; Noya, E. G.; Valeriani, C.; Vega, C. Free Energy Calculations for Molecular Solids Using GROMACS. *J. Chem. Phys.* **2013**, *139* (3), 034104.
- (6) Frenkel, D.; Ladd, A. J. C. New Monte Carlo Method to Compute the Free Energy of Arbitrary Solids. Application to the Fcc and Hcp Phases of Hard Spheres. *J. Chem. Phys.* **1984**, *81* (7), 3188–3193.
- (7) Lísal, M.; Smith, W. R.; Kolafa, J. J. Molecular Simulations of Aqueous Electrolyte Solubility: 1. The Expanded-Ensemble Osmotic Molecular Dynamics Method for the Solution Phase. *J. Phys. Chem. B* **2005**, *109* (26), 12956–12965.
- (8) Moučka, F.; Lísal, M.; Škvor, J.; Jirsák, J.; Nezbeda, I.; Smith, W. R. Molecular Simulation of Aqueous Electrolyte Solubility. 2. Osmotic Ensemble Monte Carlo Methodology for Free Energy and Solubility Calculations and Application to NaCl. *J. Phys. Chem. B* **2011**, *115* (24), 7849–7861.

- (9) Moučka, F.; Lísal, M.; Smith, W. R. Molecular Simulation of Aqueous Electrolyte Solubility. 3. Alkali-Halide Salts and Their Mixtures in Water and in Hydrochloric Acid. *J. Phys. Chem. B* **2012**, *116* (18), 5468–5478.
- (10) Aragonés, J. L.; Sanz, E.; Vega, C. Solubility of NaCl in Water by Molecular Simulation Revisited. *J. Chem. Phys.* **2012**, *136* (136), 244508.
- (11) Straatsma, T. P.; Berendsen, H. J. C. Free Energy of Ionic Hydration: Analysis of a Thermodynamic Integration Technique to Evaluate Free Energy Differences by Molecular Dynamics Simulations. *J. Chem. Phys.* **1988**, *89* (9), 5876–5886.
- (12) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43* (1), 407–435.
- (13) Li, L.; Totton, T.; Frenkel, D. Computational Methodology for Solubility Prediction: Application to the Sparingly Soluble Solutes. *J. Chem. Phys.* **2017**, *146* (21), 214110.
- (14) Bellucci, M. A.; Gobbo, G.; Wijethunga, T. K.; Ciccotti, G.; Trout, B. L. Solubility of Paracetamol in Ethanol by Molecular Dynamics Using the Extended Einstein Crystal Method and Experiments. *J. Chem. Phys.* **2019**, *150* (9), 094107.
- (15) Wand, C. R.; Fayaz-Torshizi, M.; Jiménez-Serratos, G.; Müller, E. A.; Frenkel, D. Solubilities of Pyrene in Organic Solvents: Comparison between Chemical Potential Calculations Using a Cavity-Based Method and Direct Coexistence Simulations. *J. Chem. Thermodyn.* **2019**, *131*, 620–629.
- (16) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300–313.
- (17) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426.

- (18) Chipot, C.; Pohorille, A. Calculating Free Energy Differences Using Perturbation Theory; Springer, Berlin, Heidelberg, 2007; pp 33–75.
- (19) Cabeza de Vaca, I.; Zarzuela, R.; Tirado-Rives, J.; Jorgensen, W. L. Robust FEP Protocols for Creating Molecules in Solution. *J. Chem. Theory Comput.* **2019**, acs.jctc.9b00213.
- (20) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96* (3), 1776–1783.
- (21) Lyubartsev, A. P.; Laaksonen, A.; Vorontsov-Velyaminov, P. N. Free Energy Calculations for Lennard-Jones Systems and Water Using the Expanded Ensemble Method A Monte Carlo and Molecular Dynamics Simulation Study. *Mol. Phys.* **1994**, *82* (3), 455–471.
- (22) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222* (6), 529–539.
- (23) Boothroyd, S.; Kerridge, A.; Broo, A.; Buttar, D.; Anwar, J. Solubility Prediction from First Principles: A Density of States Approach. *Phys. Chem. Chem. Phys.* **2018**, *20* (32), 20981–20987.
- (24) Wang, F.; Landau, D. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86* (10), 2050–2053.
- (25) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Generalization of the Wang-Landau Method for off-Lattice Simulations. *Phys. Rev. E* **2002**, *66* (5 Pt 2), 056703–056709.
- (26) Boothroyd, S.; Kerridge, A.; Broo, A.; Buttar, D.; Anwar, J. Why Do Some Molecules Form Hydrates or Solvates? *Cryst. Growth Des.* **2018**, *18* (3), 1903–1908.

- (27) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.
- (28) Aragones, J. L.; Valeriani, C.; Vega, C. Note: Free Energy Calculations for Atomic Solids through the Einstein Crystal/Molecule Methodology Using GROMACS and LAMMPS. *J. Chem. Phys.* **2012**, *137* (14), 146101.
- (29) D. Frenkel; B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: London, 2002.
- (30) 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* (9), 1157–1174.
- (31) 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* (2), 926–935.
- (32) Ozpinar, G. A.; Peukert, W.; Clark, T. An Improved Generalized AMBER Force Field (GAFF) for Urea. *J. Mol. Model.* **2010**, *16* (9), 1427–1440.
- (33) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C.; IUCr. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72* (2), 171–179.
- (34) Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T. DL_POLY_3: New Dimensions in Molecular Dynamics Simulations via Massive Parallelism. *J. Mater. Chem.* **2006**, *16* (20), 1911–1918.
- (35) Lee, F.-M.; Lahti, L. E. Solubility of Urea in Water-Alcohol Mixtures. *J. Chem. Eng. Data* **1972**, *17* (3), 304–306.
- (36) Benavides, A. L.; Aragones, J. L.; Vega, C. Consensus on the Solubility of NaCl in Water from Computer Simulations Using the Chemical Potential Route. *J. Chem. Phys.* **2016**,

- 144 (12), 124504.
- (37) Li, L.; Totton, T.; Frenkel, D. Computational Methodology for Solubility Prediction: Application to Sparingly Soluble Organic/Inorganic Materials. *J. Chem. Phys.* **2018**, *149* (5), 054102.
- (38) Anwar, J.; Frenkel, D.; Noro, M. G. Calculation of the Melting Point of NaCl by Molecular Simulation. *J. Chem. Phys.* **2003**, *118* (2), 728.
- (39) Anwar, J.; Zahn, D. Polymorphic Phase Transitions: Macroscopic Theory and Molecular Simulation. *Adv. Drug Deliv. Rev.* **2017**, *117*, 47–70.
- (40) Handel, R.; Davidchack, R. L.; Anwar, J.; Brukhno, A. Direct Calculation of Solid-Liquid Interfacial Free Energy for Molecular Systems: TIP4P Ice-Water Interface. *Phys. Rev. Lett.* **2008**, *100* (3), 036104.
- (41) Davidchack, R. L.; Handel, R.; Anwar, J.; Brukhno, A. V. Ice Ih-Water Interfacial Free Energy of Simple Water Models with Full Electrostatic Interactions. *J. Chem. Theory Comput.* **2012**, *8* (7), 2383–2390.
- (42) Mooij, G. C. A. M.; Frenkel, D.; Smit, B. Direct Simulation of Phase Equilibria of Chain Molecules. *J. Phys. Condens. Matter* **1992**, *4* (16), L255–L259.