The Journal of Chemical Physics

RESEARCH ARTICLE | JULY 28 2023

Prediction of aqueous solubility of a strongly soluble solute from molecular simulation

James Carruthers ; Mauro Ferrario ; Jamshed Anwar

J. Chem. Phys. 159, 044114 (2023)
https://doi.org/10.1063/5.0159402

AIP Publishing

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Prediction of aqueous solubility of a strongly soluble solute from molecular simulation

Cite as: J. Chem. Phys. 159, 044114 (2023); doi: 10.1063/5.0159402
Submitted: 23 May 2023 • Accepted: 3 July 2023 •
Published Online: 28 July 2023

James Carruthers,[1] Mauro Ferrario,[2] and Jamshed Anwar[1,a)]

AFFILIATIONS

[1] Department of Chemistry, Lancaster University, Lancaster LA1 4YB, United Kingdom
[2] Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/A, 41125 Modena, Italy

[a)] **Author to whom correspondence should be addressed:** j.anwar@lancaster.ac.uk

ABSTRACT

The prediction of solubilities of compounds by means of molecular simulation has been receiving increasing attention due to the key role played by solubility in countless applications. We have predicted the aqueous solubility of urea at 300 K from chemical potential calculations for two urea model combinations: Özpinar/TIP3P and Hölzl/(TIP4P/2005). The methodology assumes that the intramolecular contribution of the urea molecule to the chemical potentials is identical in the crystal and in solution and, hence, cancels out. In parallel to the chemical potential calculations, we also performed direct coexistence simulations of a urea crystal slab in contact with urea-water solutions with the aim to identify upper and lower bounds to the solubility value using an independent route. The chemical potential approach yielded similar solubilities for both urea models, despite the actual chemical potential values showing a significant dependence on the force field. The predicted solubilities for the two models were 0.013–0.018 (Özpınar) and 0.008–0.012 (Hölzl) mole fraction, which are an order of magnitude lower than the experimental solubility that lies in a range of 0.125–0.216 mole fraction. The direct coexistence solubility bounds were relatively wide and did not encompass the chemical potential based solubilities, although the latter were close to the lower bound values.

I. INTRODUCTION

The solubility of a substance is a fundamental thermodynamic quantity that has an impact across a wide spectrum of physicochemical phenomena. It is an important determinant of the phase behavior of biomolecules and the development of pathologies such as atherosclerosis plaques, the formation of gall and kidney stones, and the formation of amyloid plaques in diseases such as Alzheimer's and diabetes. In pharmaceuticals, the aqueous solubility can dictate the bioavailability (the rate and extent of absorption) of a therapeutic agent.[1,2] In the chemical industry, differences in solubility between the product and impurities form the basis for the purification of compounds. The solubility is also a key determinant of the formation and dynamics of ecological environments such as soil and the fate of pollutants.[3]

The ability to accurately predict the solubility would be immensely useful, for instance, in *in silico* screening[4] and at conditions inaccessible by experiment, e.g., in supercritical fluids or the earth's mantle. There are three distinct approaches to solubility prediction, each characterized by varying success.[5] The long-established (and perhaps the most commonly employed) approach is that based on quantitative structure–property relationships (QSPRs), in which molecule descriptors are correlated with experimentally determined solubility values. Although simple to develop, the QSPR approach is restricted to the chemical space (solute structure and choice of solvent) and conditions of temperature and pressure similar to those in the training set. Extrapolation to similar but unseen molecules can yield significant errors.[6] The second approach is quantum mechanical, involving the immersion of the solute in a uniform, isotropic continuum with the characteristics of the solvent being defined by a specified dielectric constant. This method therefore neglects solvent degrees of freedom. It also involves parametrization to fit the free energies of fusion of the solid phase. The third approach involves molecular simulation to access thermodynamic quantities via statistical mechanics. In this case, the input parameters are the intermolecular forces. It is pertinent to note that whilst the intermolecular interaction parameters (the force field) are optimized, for the established so-called "universal"

forcefields, they are optimized not just to reproduce a single particular property but to reproduce most of the accessible properties of the system simultaneously, albeit within a selected domain. The molecular simulation approach is potentially the most powerful given that it is based on first principles rather than correlation, opening up the possibility of accessing the full chemical and thermodynamic parameter space (a range of temperatures and pressures) without recourse to any experimental data. In practice, its success will be limited by the accuracy of the intermolecular interactions.

The use of molecular simulation to predict solubility is still very much in its infancy, although the seminal study on the aqueous solubility of KF was carried out by Ferrario *et al.* back in 2002.[7] Since then, there has been a flurry of papers employing a variety of molecular simulation approaches that include chemical potential calculations,[7–9] expanded ensembles,[10] density of states,[11,12] osmotic ensemble,[13–15] constant chemical potential simulations involving metadynamics,[16] and direct coexistence simulations.[17–19] A notable feature of some of these studies is the focus on reproducing the solubility of NaCl, which has become a *de facto* test system for new methods.[8]

Whilst some of the solubility predictions from molecular simulation have been in good agreement with experimental data, there are cases of reported predictions where the departure from experimental data is gross. Notable examples include the aqueous solubility of caffeine, for which the predicted solubility was 2 orders lower than experiment;[20] aspirin in water, which was predicted to be practically insoluble compared with its experimental value of 0.038 mol %;[21] and urea in water, where the predicted solubility was some 45 fold lower than experiment, and in methanol, where the predicted solubility was roughly 5 fold lower than experiment.[12] These results suggest an issue with sparingly soluble and soluble molecules possibly associated with the failure of current force fields.

Here we present the prediction of the aqueous solubility of the strongly soluble solute, urea in polymorphic form I (the stable form at ambient conditions), for two distinct urea models. Urea solutions have fascinated scientists for some time attracting considerable interest, especially in regard to the use of urea as a protein denaturant. Urea is known to enhance the solubility of hydrophobic molecules, and there is controversy about its aggregation behavior in water.[22,23] Furthermore, the urea molecule is well characterized by molecular simulation, notably molecular dynamics simulation studies of crystal growth and dissolution.[24]

With respect to the methodology, we have opted for the route of chemical potential calculations using standard alchemical free energy methods, including thermodynamic integration (TI) and perturbation. To give us confidence in the methodology and associated statistical mechanics, we have also estimated the solubility by means of direct coexistence simulations of the dissolution of the urea crystal as well as its crystallization from a super-saturated solution, with the estimated solubilities from the coexistence simulations serving as upper and lower bounds. The consistency of these values (or otherwise) with the predictions from the chemical potential calculations, therefore, serves to validate the latter.

II. THEORY

A. Solubility

At the solubility limit, the solid phase coexists with its solution, and the chemical potential of the solute in the solid is equal to the chemical potential of the solute in solution $\mu_{cry}(T,p) = \mu_{sol}(T,p)$, where T and p are the temperature and pressure, respectively. The determination of the solubility, in general, therefore, requires the knowledge of the chemical potential of the solid phase and the chemical potential of the solute in solution for a series of solute concentrations. The solubility is the solute concentration given by the intersection of the chemical potential of the solid phase and that of the solution as a function of solute concentration.

B. Chemical potential of the crystal phase

The chemical potential of the solute in the solid phase, being a pure phase, is given by

$$\mu_{cry} = \frac{G_{cry}}{N_{cry}} = \frac{A_{cry} + pV_{cry}}{N_{cry}}, \tag{1}$$

where G_{cry}, A_{cry}, N_{cry}, and V_{cry} are the Gibbs and Helmholtz free energies, the number of molecules, and the volume of the crystal phase, respectively, and p is the external pressure.

Since it is not feasible, in general, to directly calculate the absolute free energy for such a system, one usually starts with a simple reference state for which the free energy can be analytically calculated along with defining a reversible thermodynamic pathway between the reference and the real crystal. We determine the Helmholtz free energy of the solid using the Einstein crystal[25] as a reference state. We employ the molecule adaptation,[20,26–32] in which the molecules in the crystal are independent (non-interacting with each other; intramolecular interactions remain) but tethered to their lattice sites using a harmonic potential to fix the molecule position in space using a reference atom, with further harmonic potentials added later on to at least two more atoms to restrain the molecular orientations. The chemical potential of the solid phase is then determined using either thermodynamic integration or the free energy perturbation route.

Following Bellucci *et al.*,[31] we start by separating the intramolecular potential contribution in the canonical partition function $Q(V,T)$ for a single molecule, described collectively by the $3n$-dimensional vector $\mathbf{R} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ representing the atomic coordinates with n being the number of atoms comprising the molecule. For a flexible molecule where the intramolecular potential depends only on pair distances, one can separate the coordinate of a chosen atom, labeled here by $i = 1$, and perform the integral over its coordinates $\mathbf{r}_1$ to obtain the following expression for the single molecule canonical partition function

$$Q(V,T) = \frac{\int d\mathbf{R} e^{-U(\mathbf{R})/k_BT}}{\prod_{i=1}^{n} \Lambda_i^3(T)} = \frac{V}{\Lambda_1^3(T)} \cdot \frac{q(V,T)}{\prod_{i=2}^{n} \Lambda_i^3(T)}, \tag{2}$$

$$q(V,T) = \int d\mathbf{R}' \, e^{-\tilde{U}(\mathbf{R}')/k_BT}, \tag{3}$$

where the $3(n-1)$-dimensional vector $\mathbf{R}'$ describes the solute molecule in the reference frame with atom-1 in the origin, $d\mathbf{R}' = \prod_{i=2}^{n} d(\mathbf{r}_i - \mathbf{r}_1)$ and $\Lambda_i(T) = \sqrt{h^2/2\pi m_i k_B T}$. Then, for a fully flexible molecule, $\tilde{U}(\mathbf{R}')$ is the translationally invariant intramolecular potential energy, and $q(V,T)$ is the configurational partition function of a solute molecule with atom-1 in a given position. For

volumes V large in comparison to the characteristic molecular volume, q can be assumed to be only a function of the temperature T, and the same term $q(V,T) \cong q(T)$ can be factorized out in both the expressions of the Gibbs free energy for the crystal and for the solution phase, making it possible to cancel out such contributions when comparing the chemical potential of the two phases.

As an example, let us see how factorizing the $q(T)$ term results in a split of the expression for the chemical potential $\mu_s^{(id)}$ of an ideal gas of N_s solute molecules in a given volume V at the temperature T

$$\mu_s^{(id)}(N_s,V,T) = -k_BT\ln\frac{Q(N_s,V,T)}{Q(N_s-1,V,T)} = k_BT\ln\left(\rho_s\Lambda_1^3\right) + k_BT\ln\left(\frac{\prod_{i=2}^{n}\Lambda_i^3}{q}\right) \quad (4)$$

into two terms, the first of which is akin to the chemical potential of an ideal gas of atom-1 particles with $\rho_s = N_s/V$. The second term in the right-hand side of Eq. (4) ultimately depends only on the temperature T and contains the intramolecular contribution $q(T)$ together with the product of the other $(n-1)$ thermal wavelengths $\Lambda_i(T)$, which is also needed to make the argument of the logarithm dimensionless. Moreover, as we shall see, such a term will appear in both the expressions of the crystal chemical potential and in the solute chemical potential, so that the calculation of $q(T)$ can actually be avoided.

For the Einstein crystal, the appropriate choice is to determine the Helmholtz free energy (i.e with reference to the NVT ensemble) rather than the Gibbs free energy $G(N,p,T)$. In fact, in the NpT ensemble, volume fluctuations would require a pressure-induced scaling of the Einstein lattice positions with an additional complication in the definition of the reference system. Moreover, the pV difference between the two free energies is going to be a small term for standard pressure conditions, usually of the same order or even lower than the statistical errors.

In the case of molecules, the Einstein crystal reference state is the one with N isolated molecules distributed on the lattice sites of the specific crystal. A selected atom for each molecule, with Eq. (2) in mind, is tethered with a harmonic spring with a force constant γ_E to the reference fixed lattice position in space, $\mathbf{r}_E$. The relative single molecule potential term in the Hamiltonian takes the form

$$U = \frac{\gamma_E}{2}(\mathbf{r}_1 - \mathbf{r}_E)^2, \quad (5)$$

where $\mathbf{r}_1$ is the position of the selected atom.

The expression for the free energy A_{EC} of this reference state can be written, per molecule, as

$$\frac{A_{EC}}{N} = -\frac{3}{2}k_BT\ln\left(\frac{2\pi k_BT}{\gamma_E\Lambda_1^2}\right) - k_BT\ln\left(\frac{q}{\prod_{i=2}^{n}\Lambda_i^3}\right). \quad (6)$$

The free energy of the crystal phase can then be computed by adding to the above-mentioned reference state value, the free energy difference for reversibly transforming the reference state to the real crystal. This is accomplished in a number of steps, starting first with turning on additional harmonic restraints on selected atoms of each molecule to control the molecular orientation, for which the associated free energy contribution is specified as ΔA_{o-res}. The next two steps involve switching on the real intermolecular interactions, with an associated 'cost' in free energy ΔA_{vdw} for turning on the intermolecular van der Waals forces and ΔA_{qq} for turning on the electrostatic Coulomb forces. Finally, all harmonic restraints are removed to yield the crystal phase with an associated free energy change ΔA_{rm-res}.

Summing up all terms gives the required chemical potential of the crystal phase

$$\mu_{cry} = \frac{A_{EC} + \Delta A_{o-res} + \Delta A_{vdw} + \Delta A_{qq} + \Delta A_{rm-res}}{N} + \Delta A_{symm} + \frac{pV}{N}, \quad (7)$$

where we have included the correction $\Delta A_{symm} = k_BT\ln(\sigma)$ to consider the effects of molecular symmetry, where σ is the symmetry number of the molecule. In the case of urea, it is just $\sigma = 2$ corresponding to the mirror plane in the molecule.

It is convenient to separate out the second term in Eq. (5) to define

$$\mu'_{cry} = \mu_{cry} - k_BT\ln\frac{\prod_{i=2}^{n}\Lambda_i^3}{q}, \quad (8)$$

which will enable us, later on, to explicitly cancel out the right-hand term containing the contribution from the intra-molecular configurational partition function.

C. Chemical potential of solute solutions

We calculate the chemical potential of the solute in solution, comprising N_s solute molecules and N_w solvent molecules, as

$$\mu_{sol}(N_s,N_w,p,T) = G(N_s,N_w,p,T) - G(N_s-1,N_w,p,T) = -k_BT\ln\frac{\mathcal{Q}(N_s,N_w,p,T)}{\mathcal{Q}(N_s-1,N_w,p,T)}, \quad (9)$$

where the Gibbs free energy $G(N_s,N_w,p,T)$ is given in terms of the isothermal-isobaric partition function $\mathcal{Q}(N_s,N_w,p,T)$ for the system. The chemical potential μ_{sol} can then be determined using thermodynamic integration or the perturbation approach using Bennett Acceptance Ratio (BAR) along a reversible path realizing the insertion of one solute molecule into the solution at fixed temperature, pressure, and number of solvent molecules.

Determining μ_{sol} involves the reversible insertion of the additional solute molecule into the solution, which is achieved by gradually switching on the interaction potential between the inserted molecule and the solution by means of the λ-coupling parameter. This transforms the Hamiltonian of the system according to

$$\mathcal{H}(\lambda) = \mathcal{K}(N_s-1,N_w) + \mathcal{U}_0(N_s-1,N_w) + \sum_{i=1}^{n}\frac{\mathbf{p}_i^2}{2m_i} + \lambda\mathcal{U}_1(N_s,N_w), \quad (10)$$

where $\mathcal{K}(N_s-1,N_w)$ and $\mathcal{U}_0(N_s-1,N_w)$ are the kinetic and potential energies, respectively, of the starting solution. The momentum term is the kinetic energy of the additional solute molecule comprising n atoms, and $\mathcal{U}_1(N_s,N_w)$ is the potential energy resulting from the insertion of the additional solute molecule into the solution.

The isothermal-isobaric partition function for the system described by $\mathscr{H}(\lambda)$, as a function of λ assuming water as the solvent, can then be written as

$$\begin{aligned}\mathscr{Q}(N_s-1,N_w,p,T,\lambda) &= \frac{1}{(N_s-1)!\,N_w!}\left(\prod_{i=1}^{n}\Lambda_i\right)^{-3N_s}\left(\prod_{j=1}^{3}\Lambda_j\right)^{-3N_w} \\ &\times \int dV \exp^{-\beta pV}\int d\mathbf{R}^{N_w}\, d\mathbf{R}^{N_s-1}\, e^{-\beta\mathscr{U}_0}\int d\mathbf{R}_s e^{-\lambda\beta\mathscr{U}_1}, \quad (11)\end{aligned}$$

where $\mathbf{R}_s$ are the coordinates of the n atoms of the inserted molecule.

The respective partition functions for the two systems, with the inserted solute molecule, ($\lambda = 1$), and without, ($\lambda = 0$), can then be written

$$\mathscr{Q}(\lambda=0) = \mathscr{Q}(N_s-1,N_w,p,T)\frac{\langle V\, q(V,T)\rangle_{N_s-1,N_w,p,T}}{\Lambda_1^3(T)\prod_{i=2}^{n}\Lambda_i^3(T)}, \quad (12)$$

$$\mathscr{Q}(\lambda=1) = N_s\,\mathscr{Q}(N_s,N_w,p,T), \quad (13)$$

where $\langle\ldots\rangle_{N_s-1,N_w,p,T}$ is the average over the ensemble with fixed (N_s-1,N_w,p,T), and for $\lambda = 0$, we have performed the integral over the non-interacting solute molecule as done in Eq. (2).

Then, using the assumption that the volume dependence of q can be ignored, the corresponding free energies can be written

$$\begin{aligned}G(\lambda=0) \cong G(N_s-1,N_w,p,T) &- k_BT\ln\left(\frac{\langle V\rangle_{N_s-1,N_w,p,T}}{\Lambda_1^3(T)}\right) \\ &- k_BT\ln\left(\frac{q(T)}{\prod_{i=2}^{n}\Lambda_i^3(T)}\right), \quad (14)\end{aligned}$$

$$G(\lambda=1) \cong G(N_s,N_w,p,T) - k_BT\ln(N_s). \quad (15)$$

The chemical potential of the solute is then written as the sum of the free energy difference between the two end states $G(\lambda = 1) - G(\lambda = 0) = \Delta G_s$ plus an *ideal* term plus the term containing the unknown intramolecular contribution q

$$\begin{aligned}\mu_{sol}(N_s,N_w,p,T) &= G(N_s,N_w,p,T) - G(N_s-1,N_w,p,T) \\ &\cong G(\lambda=1) - G(\lambda=0) + k_BT\ln\left(\frac{N_s}{\langle V\rangle}\Lambda_1^3\right) \\ &+ k_BT\ln\frac{\prod_{i=2}^{n}\Lambda_i^3}{q}, \quad (16)\end{aligned}$$

where ΔG_s is the solvation free energy that we determine by either thermodynamic integration or perturbation methodology using soft-core potentials.[33] The latter removes the divergence in the ensemble averages resulting from the annihilation or creation of atoms.

In terms of the solute density $\rho_s = N_s/\langle V\rangle$, then the *ideal* term in the above-mentioned equation is the same as the first one on the right-hand side of Eq. (4).

As done for the chemical potential in the solid phase [Eq. (7)], we can define

$$\mu'_{sol}(N_s,N_w,p,T) = \mu_{sol}(N_s,N_w,p,T) - k_BT\ln\frac{\prod_{i=2}^{n}\Lambda_i^3}{q}. \quad (17)$$

Having assumed the intra-molecular partition function q is independent of volume, the contribution of q is identical both in solution and in the solid state; therefore, the right-hand side term in Eq. (16) cancels out with the same term contained in the expression for μ'_{cry} in Eq. (7), yielding

$$\mu'_{sol} \cong \mu'_{cry}$$

at coexistence. This is a fair assumption for urea given that it is a near-rigid molecule due to the high rotational barrier of the N–C bond. We would expect the assumption to fail for large flexible molecules, for which the intramolecular partition function is likely to be more significantly coupled to the intramolecular partition function, being a particular issue when the structures sampled in the crystal and solution environments are markedly different.

In Eqs. (6) and (15), we have hidden a number of correction terms that are required when simulations are performed, as usual, at zero constant total momentum $\mathbf{P} = 0$ for the simulation box. Actually, their contribution is very small (of order N^{-1}, where N is the number of molecules), which we checked and found that it could be safely neglected for the system sizes used in our calculations. We refer the interested reader to Polson *et al.*[34] and the appendices in Bellucci *et al.*[31]

III. METHODS

There are two notable and distinct force field models of urea, one developed by Özpınar[35] and the other by Hölzl.[36] In view of this, we carried out two sets of solubility predictions, one for each of the models. The Özpınar model is a refinement of the GAFF force field[37] developed to reproduce the crystal structure, whilst the Hölzl model is an improvement on the Kirkwood–Buff force field (KBFF) of Weerasinghe and Smith[38] being optimized using various experimental and computational data with a focus on reproducing aqueous solution properties. The respective parameters and associated partial charges for the Özpınar and Hölzl urea are given in Table I. The cross terms were evaluated using the Lorentz–Berthelot combining rules. The TIP3P water model was used with the Özpınar model given that it is associated with the GAFF force field, and the TIP4P/2005 model with the Hölzl urea model as specified in the Hölzl study.

All simulations were carried out using the GROMACS molecular dynamics package version 2018.1.[39] The simulations were carried out in the isothermal-isobaric ensemble at a temperature of 300 K and a pressure of 1 bar, employing a stochastic integrator to enhance sampling, unless explicitly specified for a particular set of simulations. The time step was 2 fs, and all bond lengths were constrained using the LINCS algorithm. The Coulombic interaction forces were calculated using particle mesh Ewald (PME) with a Fourier spacing of 0.12 nm, a PME order of 6, and an Ewald tolerance of 1.0×10^{-6}. The van der Waals and real space components of the Coulombic interactions were truncated at 1.2 nm, with the van der Waals being moderated by a potential-switch function starting at 1.0 nm. The pressure was regulated using the Parrinello–Rahman barostat with a compressibility of 4.5×10^{-5} bar^{-1}.

The urea form I crystal structure used for the free energy and direct coexistence simulations was taken from the Crystallography Open Database (entry 1278490), which had been determined using neutron diffraction at room temperature.[40] The crystal structure was equilibrated for both urea models, Özpınar and Hölzl, at 300 K and

TABLE I. Force field parameters for the Özpınar and Hölzl urea models. The cis and trans designations on the hydrogens are in relation to the oxygen. The dihedral parameters are common to both models. All bonds were constrained to their r_0 values.

Özpınar			
Atom	σ (nm)	ε (kJ mol^{-1})	q (e_c)
H	0.106 91	0.006 57	0.388
C	0.339 97	0.359 80	0.884
N	0.325 00	0.711 30	−0.888
O	0.295 99	0.878 60	−0.660
Bond		r_0 (nm)	
H–N		0.1010	
C–N		0.1383	
C=O		0.125	
Angle	$U = k/2\,(\theta - \theta_0)^2$ k (kJ mol^{-1})	θ_0 (deg)	
H–N–H	292.9	120.0	
H–N–C	251.0	120.0	
N–C–N	585.8	118.6	
N–C–O	669.4	120.9	
Hölzl			
Atom	σ (nm)	ε (kJ mol^{-1})	q (e_c)
H(cis)	0.113 33	0.065 69	0.4026
H(trans)	0.113 33	0.065 69	0.4421
C	0.360 39	0.359 82	0.6068
N	0.344 52	0.511 14	−0.8400
O	0.313 77	0.594 32	−0.6162
Bond		r_0 (nm)	
H–N		0.1010	
C–N		0.1335	
C=O		0.1229	
Angle	$U = k/2\,(\theta - \theta_0)^2$ k (kJ mol^{-1})	θ_0 (deg)	
H–N–H	445	120.0	
H–N–C	390	120.0	
N–C–N	670	117.2	
N–C–O	670	121.4	
Common			
Proper dihedral	$U = k(1 + \cos(n\phi - \phi_0))$ k (kJ mol^{-1})	ϕ_0 (deg)	n
H–N–C–O	8.368	0	1
H–N–C–O	10.46	180	2
X-C-N-X	10.46	180	2
Improper dihedral	$U = k/2\,(\phi - \phi_0)^2$ k (kJ mol^{-1})	ϕ_0 (deg)	
C	43.932	180	
N	4.6024	180	

1 bar using molecular dynamics simulation with anisotropic pressure coupling and Parrinello–Rahman boundary conditions, where the angles and dimensions of the simulation cell are allowed to vary. The deviation in the lattice parameters was minimal, being below 5% for both models. The average cell parameters and atomic coordinates from the equilibration trajectory were minimized in an *NVT* ensemble simulation to yield the low potential energy crystal structure that was used for all subsequent simulations.

The direct coexistence studies were carried out by exposing the surfaces of a urea crystal slab to pure water or water supersaturated with urea. The coexistence simulations were attempted for the slow growing face (110), which is normally exhibited by the urea crystal, and the more faster growing face (001). For the slower growing face, little or no dissolution was observed over a trajectory comprising 640 ns, although some dissolution could be observed by removing a few of the surface molecules from the solution. This overcomes the 2-D nucleation barrier to dissolution. In view of this, further coexistence studies were restricted to the faster growing (001) face.

In the coexistence studies, the exposed faces of the urea crystal slab were aligned perpendicular to the *z*-axis of the simulation box and exposed to water. The (001)-exposing slab comprised 10 (*a*-axis) × 10 (*b*-axis) × 12 (*c*-axis) unit cells, which equated to 12 molecular layers (perpendicular to the *z*-axis) of urea molecules. The thickness of the crystal slab was 5.5 nm. Periodic boundaries were applied in all 3 dimensions. The thickness of the water layer between the crystal surfaces was ~16 nm (comprising 15 697 water molecules), which is about 13 times the non-bonded interaction cutoff, thus, minimizing the influence of either of the two exposed crystal surfaces on the processes occurring at the other. The initial simulation box dimensions were 5.604 nm in the *x* and *y* axes and 28.009 nm in the *z* axis. Pressure scaling was semi-isotropic with the scaling along the *z*-axis being decoupled from the *x*- and *y*-axes.

The pure water coexistence simulation (dissolution) of the Özpınar model crystal was simulated for 400 ns, whilst that of the Hölzl model was simulated for 1000 ns. The simulations were stopped after observing convergence with respect to the number of solute molecules in solution and, hence, the different time periods. The supersaturated water coexistence (crystal growth) simulations employed the same slab-water set up as above, exposing the same surface. The water layer in this case was super-saturated with urea molecules at a concentration of 0.2 M fraction, which translates to 2000 urea molecules in 8000 water molecules.

The free energy simulations for the urea crystal phase were carried out in a simulation box comprising 5 (*a*-axis) × 5 (*b*-axis) × 6 (*c*-axis) unit cells of the equilibrated urea crystal (300 molecules, 2400 atoms). The initial simulation box dimensions were 2.831 nm in the *x* and *y* dimensions and 3.286 nm in the *z* dimension. The component free energy changes [see Eq. (6)] were determined in the *NVT* ensemble for reasons given earlier in the Theory section. The positional harmonics restraints on the urea molecule defining the Einstein crystal were located on the central carbon atom using a force constant of $\gamma_E = 0.50 \times 10^6$ kJ mol^{-1} nm^{-2}. The thermodynamic pathway from the reference Einstein crystal to the real crystal comprised the following stages: (i) gradually turning on additional positional restraints on the nitrogen atoms to enforce correct orientation; (ii) gradually turning on van der Waals interactions (the soft-core van der Waals is unnecessary here as the required voids to accommodate the emerging atoms are unoccupied); (iii) gradually switching on partial charges; and finally (iv) removing all positional restraints. As urea is a relatively rigid molecule, one could use any three non-colinear atoms (colinear atoms would create an axis of rotation) to restrain the orientation but we decided on the symmetrical selection of the carbon and two nitrogens. The calculation of free energies from the thermodynamic integration and perturbation ensemble averages was carried out using the Alchemical Analysis program by The Mobley Lab.[41]

Whilst the determined crystal free energy is theoretically independent of the harmonic restraint force constant, there may be an optimum value of the restraint force for minimizing numerical errors. In view of this, we explored the effect of varying the force constant γ_E for the Özpınar model using the values of 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 1.00, 2.00, and 4.00 $\times 10^6$ kJ mol^{-1} nm^{-2}. All of the force constant values other than the lowest ($\gamma_E = 0.01 \times 10^6$ kJ mol^{-1} nm^{-2}) yielded crystal chemical potential values (μ'_{cry}) that were very close, differing only by about 0.1 kJ mol^{-1} (see Table IV and Fig. 3). In view of this, we proceeded to use the results for $\gamma_E = 0.50 \times 10^6$ kJ mol^{-1} nm^{-2}, which lies in the middle of the range investigated, for the solubility prediction.

The Gibbs free energy of solvation of urea was determined for a set of aqueous urea solutions of varying concentration comprising a total of 1001 molecules (urea plus water) using thermodynamic integration.[25] The composition ratios (urea:water) of the urea solutions were: 0.0005 (with 2001 molecules), 0.001, 0.011, 0.031, 0.051, 0.101, 0.151, 0.201, 0.251, and 0.301. The simulations employed the soft-core van der Waals potential to minimize divergence of the ensemble averages due to particle creation or annihilation.[33] The soft-core parameters were sc-alpha: 0.5, sc-r-power: 6, sc-power: 1, and sc-sigma: 0.3. The van der Waals parameters were gradually switched on first, followed by the partial charges. Switching on both the van der Waals and Coulombic interactions simultaneously is problematic unless one employs a damping potential for the electrostatics.[42] The number of λ-states for van der Waals and electrostatics was 26 and 11, respectively, with linear spacing (λ_{vdw} : 0.0, 0.04, 0.08, . . . , 0.92, 0.96, 1.0; λ_{qq} : 0.0, 0.1, 0.2, . . . , 0.8, 0.9, 1.0). Only the intermolecular non-bonded forces were transformed by the lambda-coupling.

IV. RESULTS AND DISCUSSION

A. Reproduction of the crystal structure of urea

The averaged equilibrated lattice parameters for the urea crystal structure resulting from the molecular dynamics simulations at 300 K and 1 bar are compared with experimental data for both the Özpınar and the Hölzl models in Table II.

For both urea models, all of the predicted lattice parameters show a relatively low percentage deviation, being ≤3.3% for the Özpınar model and ≤2.5% for the Hölzl model (which is impressive considering this model was not optimized for the crystal phase). The simulation cell angles did not drift from 90.0 indicating no breaking of the unit cell symmetry. These results give confidence that the crystal structure is well described by both urea models. The lattice energy for the Özpınar model was calculated to be 96.5 kJ mol^{-1}, which is appreciably lower than the experimental value of 99.43 kJ mol^{-1}. In contrast, the calculated lattice energy for the Hölzl model was

TABLE II. Average lattice parameters for urea polymorph I from molecular dynamics simulation at 300 K and 1 bar for the Özpınar and Hölzl models of urea, compared with experimental data.[40] Percentage deviation from the experimental data is given in brackets. U_{latt} is the lattice potential energy per molecule, $U_{molecule}$ is the potential energy of the equilibrated isolated molecule, and E_{latt} is the lattice energy determined using $E_{latt} = -(U_{latt} - U_{molecule}) - 2RT$. Urea polymorph I is tetragonal (unit cell parameters: $a = b \neq c, \alpha = \beta = \gamma = 90°$) in the space group P-4 2_1 m with 2 molecules in the unit cell (Z = 2).

Cell parameter	Exp.	Özpınar	Hölzl
a (nm)	0.5661	0.5477 (−3.3)	0.5647 (−2.5)
c (nm)	0.4712	0.4714 (0.04)	0.4713 (0.02)
U_{latt} (kJ mol^{-1})	⋯	−775.8	−777.9
$U_{molecule}$ (kJ mol^{-1})	⋯	−674.3	−662.0
E_{latt} (kJ mol^{-1})	99.43	96.50 (−2.9)	110.90 (11.5)

110.90 kJ mol^{-1}, which is significantly (~11%) higher than the experimental value.

B. Solubility bounds from direct coexistence simulation

The evolution of the molar fraction of urea in solution as a function of time for the direct coexistence simulations in which the crystal slab is exposed to pure water and to a supersaturated solution is given in Fig. 1 for both urea models. In general, the rate at which the solution concentration decays for the crystal growth simulations is faster than the rate at which the concentration increases for the crystal dissolution simulation, which is expected being given by the substantial supersaturation in the case of the former. The deposited layers had the same structure in terms of the molecular orientations as the crystal slab.

The simulations were carried out to apparent convergence not equilibrium, hence, the differences in simulation lengths. Convergence here does not necessarily imply equilibrium. A complete crystal layer with no terraces or kink sites has a significant (2-D nucleation) barrier to having molecules added to or removed from it.[43] At a sufficiently low solute chemical potential difference between the crystal surface and solution, the simulation converges to a stage when the crystal surface comprises a complete layer. As there is a barrier to further dissolution or growth, the kinetics of either of these processes become really slow, and the simulation appears to show convergence. True equilibration can require simulations on the microsecond timescale, which is a known limitation of the direct coexistence approach. The objective here was not to determine the solubility using direct coexistence but rather to identify solubility *bounds* resulting from the simulation of two inverse processes, dissolution and crystal growth.

The solubility bounds estimated from Fig. 1 were 0.03–0.06 (2–3M) mole fraction for the Özpınar model and 0.03–0.11 mole fraction (2–5M) for the Hölzl model. It is important to note that these bounds are likely to be dependent on the chosen exposed crystal surface, here (001), given that the chemical potential difference driving the dissolution (and crystal growth) is not just determined by the bulk crystal potential but also the interfacial free energy of the exposed surface. The reported experimental aqueous solubility values of urea at 298K vary between 5.87–9.07M,[44] which is

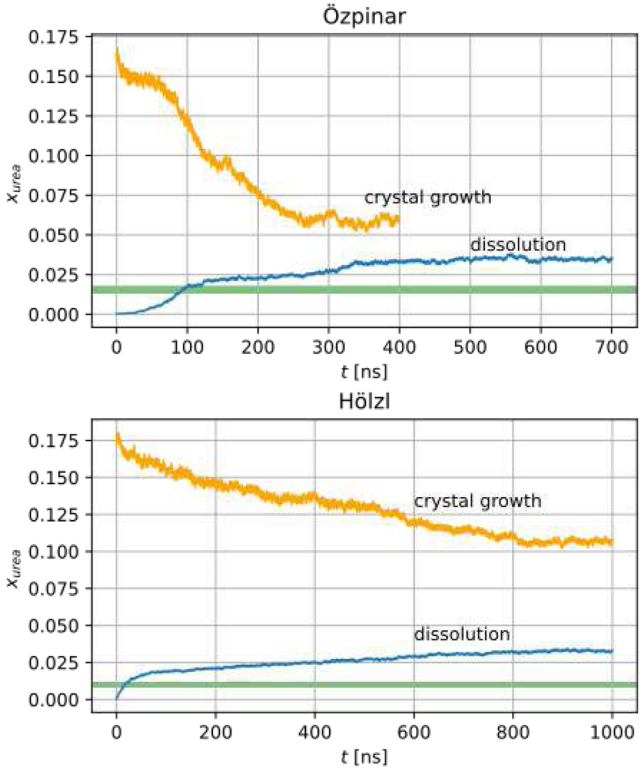

FIG. 1. Evolution of the mole fraction of urea in aqueous solution for the dissolution of a crystal slab of urea in an initially pure system of water and the growth of the crystal slab from a 0.2 mole fraction of urea in aqueous solution for both the Özpınar/TIP3P and Hölzl/(TIP4P/2005) urea model combinations. The horizontal, green band is the estimated solubility range from free energy calculations.

2–6 fold higher, suggesting an issue with the urea force field parameters, including its interaction with the employed water model. An earlier simulation study using the urea model of Duffy *et al.*[45,46] exploring the dissolution and crystal growth of urea in aqueous solution, recorded final solution concentrations in a range of 7–11.6M from a simulation trajectory of up to 70 ns.[24] The Duffy urea model, however, showed marked deviations in the crystal structure, necessitating isotropic pressure coupling to maintain the crystal structure.

C. Chemical potential of the crystal phase

The free energy change for the transformation of the Einstein crystal to a real crystal was determined using both the multistate Bennett acceptance ratio (MBAR) and thermodynamic integration (TI). The pathway involved including additional harmonic restraints to enforce the correct molecular orientation, switching on the van der Waals, followed by Coulombic interactions, and finally removal of the restraints. The key ensemble average for thermodynamic integration, $\langle \partial H/\partial \lambda \rangle$, as a function of λ is shown for each of these stages for the Özpınar model in Fig. 2. The respective curves for the Hölzl model are similar. It is clear from these curves that the integration involving van der Waals and Coulombic interactions is well behaved, whilst the onset of

the orientational restraints and the end-points of the restraint removal stage show marked variation in the $\langle \partial H/\partial\lambda \rangle$ value with the possibility of divergence. To minimize integration errors, these regions were sampled more intensively with 51 cubic spaced lambda states (λ_{o-res} : $0.00, 0.02^3, 0.04^3, \ldots, 0.96^3, 0.98^3, 1.0$; λ_{rm-res} : $0.0, 1 - 0.98^3, 1 - 0.96^3, \ldots, 1 - 0.04^3, 1 - 0.02^3, 1.0$).

The various component free energy values [corresponding to Eq. (6)] are given in Table III calculated using MBAR and TI. Both approaches essentially give the same results with similar estimated errors, which gives us confidence in the convergence of the simulations and the choice of λ states. All further numerical free energy calculations employed thermodynamic integration.

The chemical potential μ'_{cry} of the Hölzl model is on average about 13 kJ mol^{-1} lower than that for the Özpınar model, suggesting a stronger affinity of the Hölzl model for the lattice. With respect to the component free energies, the determined values differ markedly between the two urea models, including the free energy for switching-on the orientational restraints. The latter arises because the models have slightly different C–N and C=O bonds and N–C–N and N–C–O bond angles (see Table I). As expected from the $\langle \partial H/\partial\lambda \rangle$ as a function of λ plots in Fig. 2, the largest error contributions come from the switching-on of the orientational restraints and the last stage of switching-off all restraints.

We also explored the effect of varying the magnitude of the harmonic restraint force constant γ_E on μ'_{cry} and the free energy component values and associated errors for the Özpınar model. The component free energies and associated errors for the various force constant values determined using thermodynamic integration are tabulated in Table IV and plotted in Fig. 3. It is clear from Fig. 3 that the chemical potential values (μ'_{cry}) for all force constants other than the lowest (γ_E = 0.01 ×10^6 kJ mol^{-1} nm^{-2}) are consistent and very close, differing only by about 0.1 kJ mol^{-1}. Based on these considerations, we have selected the data for γ_E = 0.50 × 10^6 kJ mol^{-1} nm^{-2}, which lies in the middle of the range investigated, for the solubility calculations. The lowest restraint strength fails to agree with higher restraint strengths as it no longer stops atoms from significantly overlapping and leads to badly behaved data collection. Higher restraint strengths have greater errors as the free energy curves for applying and removing restraints become more extreme.

D. Chemical potential of aqueous solutions of urea

The thermodynamic ensemble average $\langle \partial H/\partial\lambda \rangle$ for the two component free energies, ΔG_{vdw} and ΔG_{qq}, of the solvation free energy as a function of λ is shown plotted for the Özpınar/TIP3P urea model combination for a urea molecule in pure water in Fig. 4. The respective plots for other concentrations and for the Hölzl/(TIP4P/2005) urea model combination were similar. The $\langle \partial H/\partial\lambda \rangle$ ensemble average is well behaved, and there are no integration issues. Nevertheless, the plots indicate that the estimated errors on the individual $\langle \partial H/\partial\lambda \rangle$ data points are significant. When propagated to the solution chemical potential, the errors in the latter are an order of magnitude larger than crystal chemical potential errors but are still decent at below 0.2 kJ/mol—this is because only a single molecule is being transformed, so sampling efficiency is lower than for the crystal calculation.

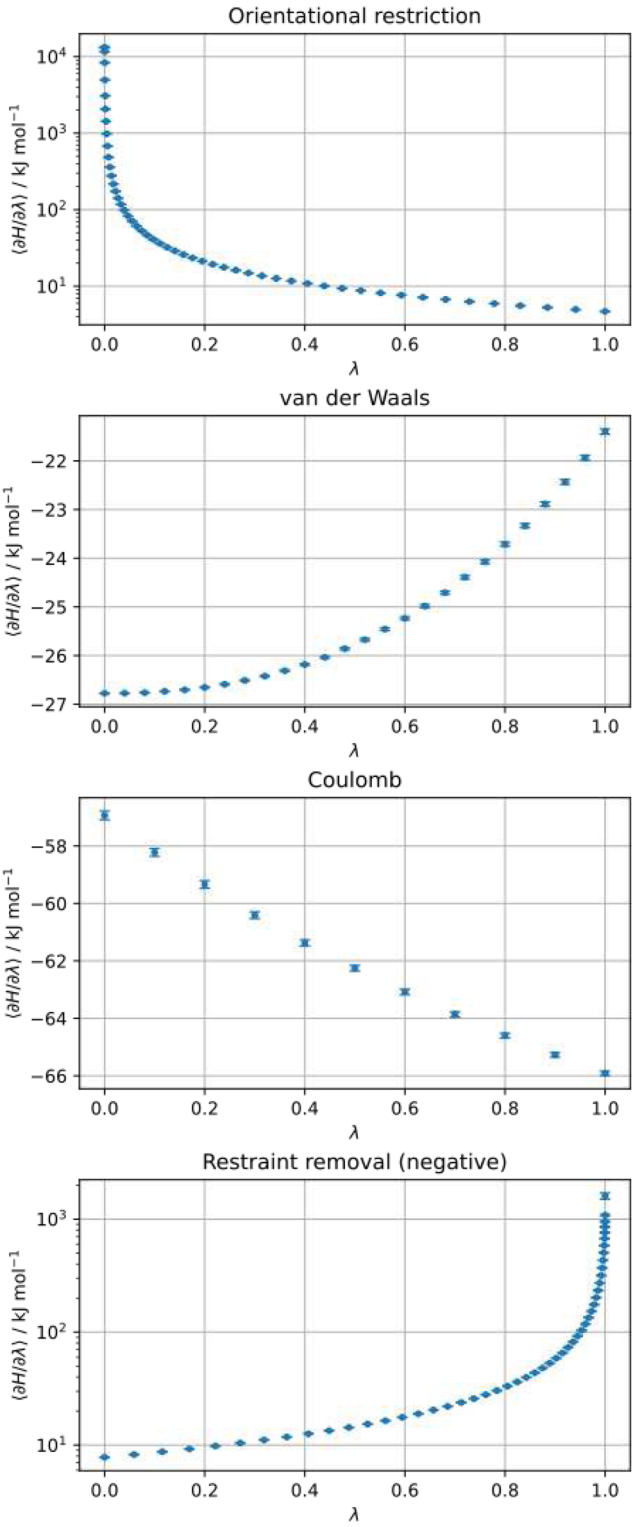

FIG. 2. The ensemble average $\langle \partial H/\partial\lambda \rangle$ as a function of λ for the various contributions to the crystal chemical potential for the Özpınar model of urea. The restraint removal component is plotted as the negative to enable the use of a logarithmic scale.

TABLE III. Chemical potential of the urea crystal μ'_{cry} as defined by Eq. (7) for the Özpınar and Hölzl urea models, showing the various component free energy contributions determined using both thermodynamic integration (TI) and multistate Bennett acceptance ratio (MBAR). μ'_{cry} excludes the intramolecular partition term since we assume it to be identical to that in solution and hence cancels out. Estimated standard errors are given in brackets.

Component ΔA	Özpınar-TI (kJ mol^{-1})	Özpınar-MBAR (kJ mol^{-1})	Hölzl-TI (kJ mol^{-1})	Hölzl-MBAR (kJ mol^{-1})
A_{EC}/N	12.398	12.398	12.398	12.398
$\Delta A_{o-res}/N$	33.773(2)	33.479(3)	35.005(2)	34.13(1)
$\Delta A_{vdw}/N$	−22.415(1)	−22.435(1)	−15.176(1)	−14.701(3)
$\Delta A_{qq}/N$	−81.667(5)	−81.677(5)	−103.345(3)	−103.28(3)
$\Delta A_{rm-res}/N$	−27.809(5)	−27.468(6)	−27.675(4)	−27.578(4)
ΔA_{sym}	−1.729	−1.729	−1.729	−1.729
pV/N	0.004	0.004	0.005	0.005
μ'_{cry}	−87.445(7)	−87.428(8)	−100.517(5)	−100.76(3)

TABLE IV. Chemical potential of the urea crystal μ'_{cry} as defined by Eq. (7) as a function of restraint force constant γ_E for the Özpınar urea model, showing the various component free energy contributions along with associated standard errors given in brackets. Free energies were calculated using thermodynamic integration.

Component ΔA (kJ mol^{-1})	γ_E (10^6 kJ mol^{-1} nm^{-2})								
	0.01	0.05	0.10	0.25	0.50	0.75	1.00	2.00	4.00
A_{EC}/N	−2.239	3.783	6.376	9.804	12.398	13.915	14.991	17.584	20.178
$\Delta A_{o-res}/N$	17.566(1)	23.748(1)	26.566(2)	30.548(2)	33.773(2)	35.723(2)	37.129(2)	40.563(2)	44.038(2)
$\Delta A_{vdw}/N$	−20.217(4)	−21.198(2)	−21.583(2)	−22.118(1)	−22.415(1)	−22.539(1)	−22.608(1)	−22.724(1)	−22.789(1)
$\Delta A_{qq}/N$	−74.108(8)	−79.254(5)	−80.394(5)	−81.264(5)	−81.667(5)	−81.824(4)	−81.918(4)	−82.068(4)	−82.150(4)
$\Delta A_{rm-res}/N$	−6.071(2)	−12.909(3)	−16.773(3)	−22.740(4)	−27.809(5)	−31.002(5)	−33.328(6)	−39.177(7)	−45.124(9)
ΔA_{sym}	−1.729	−1.729	−1.729	−1.729	−1.729	−1.729	−1.729	−1.729	−1.729
pV/N	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
μ'_{cry}	−86.798(9)	−87.559(7)	−87.537(6)	−87.500(7)	−87.449(7)	−87.456(7)	−87.463(8)	−87.551(9)	−87.58(1)

The chemical potential μ'_{sol} [as defined by Eq. (16)] of the various urea solutions determined using thermodynamic integration along with the various free energy components is tabulated in Tables V and VI as a function of concentration for the Ozpınar/TIP3P and Hölzl/(TIP4P/2005) urea model combinations.

The chemical potential μ'_{sol} values at any given concentration are markedly different for the two models, with the value for the Hölzl model being about 10–12 kJ mol^{-1} lower suggesting greater affinity for the employed water model. The chemical potential of urea in water has previously been studied by Kokuba *et al.* (2007)[47] for two urea models, OPLS[45] and KBFF,[38] with a focus on calculating activity coefficients of aqueous urea solutions. The shift that defines μ'_{sol} relative to μ_{sol} in Eq. (16) is equivalent to Kokuba *et al.* disregarding the unknown contribution from the internal degrees of freedom of the molecule and assuming $q_A = 1$ as defined in their Eq. (61). Comparison of the chemical potentials on this basis with Kokuba *et al.*'s study reveals that our μ'_{sol} values for the Ozpınar/TIP3P model practically superimpose on their values for the OPLS model. For instance, at $x = 0.251$ $\mu_{sol} = -83.8$ kJ mol^{-1}, whilst the OPLS value at $x = 0.27$ in Kokuba *et al.*'s study is −85.29 kJ mol^{-1}. The largest deviation comes in at the "infinite dilution" as a consequence of the different sizes of the system (hence different non-zero concentrations), which result in different values of the ideal term. Unfortunately, it is not possible to directly compare the activity coefficients as calculated by Kokuba *et al.* as the reference infinite dilution states are different between the two studies.

Considering the two free energy components, the component ΔG_{vdw} is very small (<4% contribution) but interestingly takes a negative value for the Ozpınar model at high urea concentrations. This we believe reflects the relatively high value of the Lennard-Jones individual homo ε parameter for both the N and the O atom in the Ozpınar model: ε(N) = 0.711 30 kJ mol^{-1} and ε(O) = 0.878 60 kJ mol^{-1} compared with ε(N) = 0.511 14 kJ mol^{-1} and ε(O) = 0.594 32 kJ mol^{-1} for the Hölzl model. The ideal component G_{ideal} contributes about 25% to μ'_{sol} and differs a little between the models due to the slightly different solution densities. The Coulombic component ΔG_{qq} is the most dominant, contributing between 70% and 80% to the chemical potential.

E. Prediction of solubility

The chemical potential of the crystal μ'_{cry} and that of the series of urea solutions μ'_{sol} is plotted in Fig. 5 for the two urea models. Both of these chemical potentials, μ'_{cry} and μ'_{sol}, by definition [see Eqs. (7) and (16)] exclude the intramolecular term, which we assume to be identical in both the crystal and in solution. This is a fair assumption, as the urea molecule has only limited flexibility. The solubility is given by the intersection of the crystal and solution curves.

Using a second-order rational function to fit the data, 1σ confidence bands gave a solubility interval of 0.013–0.018 M fraction for the Özpınar model and 0.008–0.012 for the Hölzl model. The experimental aqueous solubility of urea at 298 K is reported to

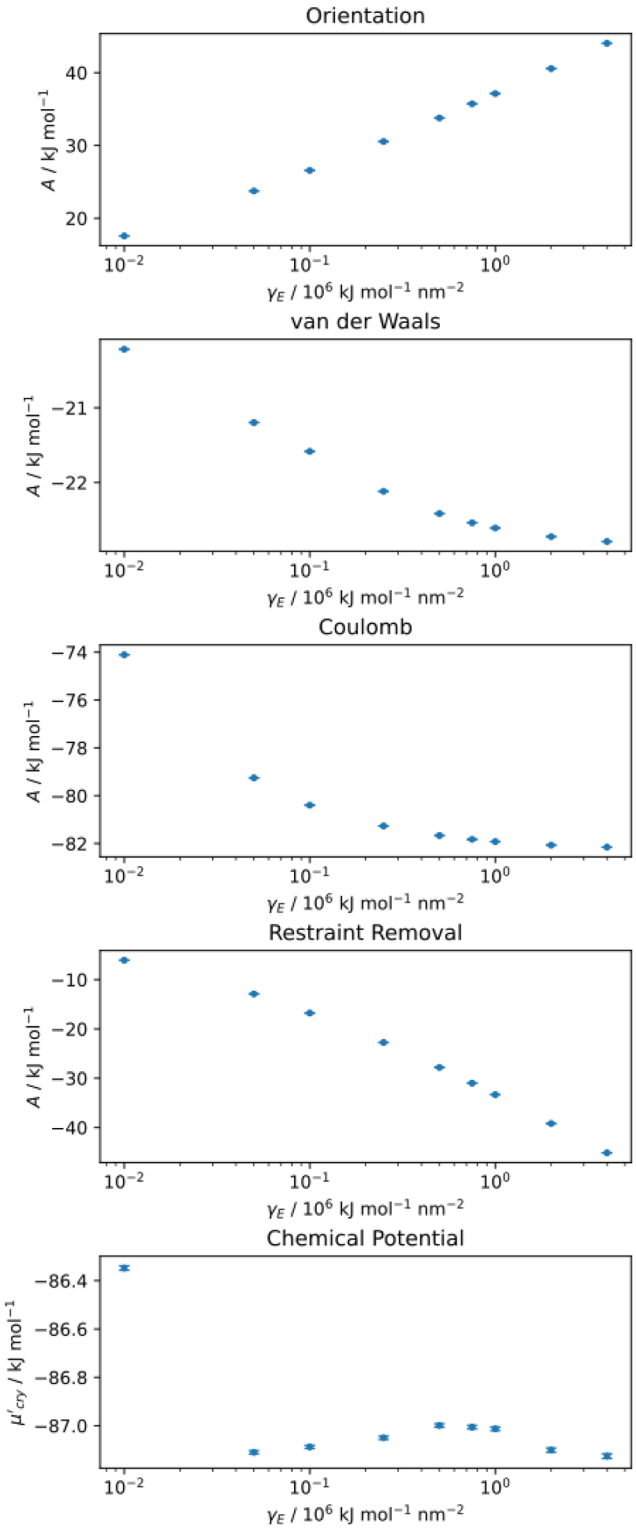

FIG. 3. Chemical potential μ'_{cry} of the urea crystal as a function of the force constant γ_E of the harmonic restraints for the Özpınar urea model, showing the various component free energy contributions along with estimated standard errors.

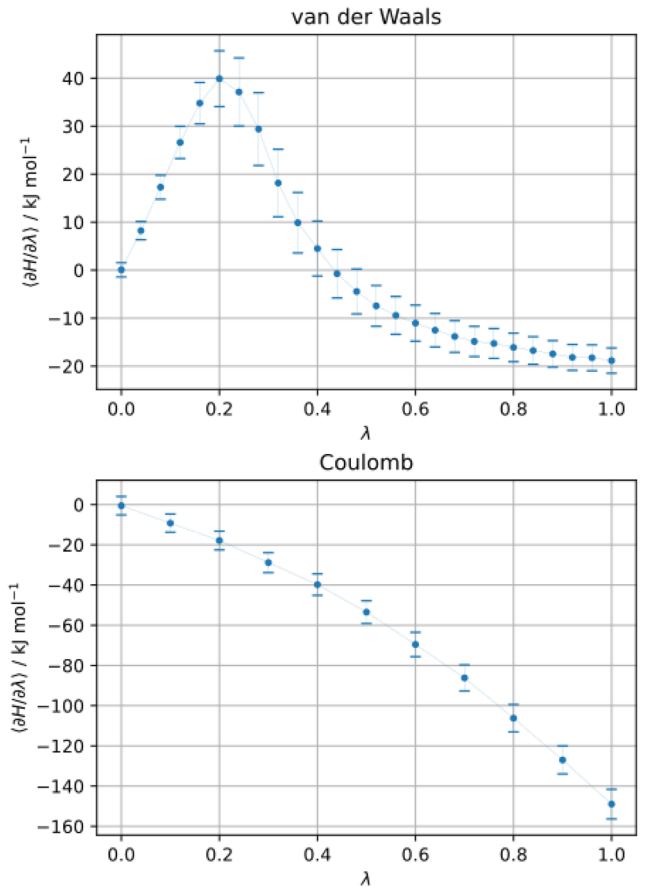

FIG. 4. The ensemble average $\langle \partial H/\partial\lambda \rangle$ for the two free energy contributions, ΔG_{vdw} and ΔG_{qq}, as a function of λ for the insertion of a urea molecule into pure water for the Özpınar urea model.

be in a range of 5.874–9.058M,[44] which equates to a 0.125–0.216 mole fraction.[48] The predicted solubilities are, therefore, an order of magnitude lower than the experimental solubility. The solubility of urea in water (and in ethanol) was previously investigated by Boothroyd and Anwar[12] but by the density of states methodology rather than alchemical free energy methods as done here. However, they employed the Amber GAFF force field with TIP3P water coupled with two distinct sets of partial charges and treated the urea molecule as rigid, making a direct comparison with our results difficult. Having said that, their predicted solubility of urea in water was in a range of 0.008–0.009 mole fraction, which is within the confidence band of the Hölzl model.

We note that whilst the chemical potential of the crystal phase of the Hölzl model (−100.517 kJ mol^{-1}) is markedly lower than that of the Özpınar model (−87.445 kJ mol^{-1}), their predicted solubilities are not too different. It appears that for the Hölzl model, the solution chemical potential as a function of urea concentration is also lower, causing the chemical potential curves to intersect at a concentration similar to that for the Özpınar model. So, for the Hölzl model, whilst the cohesive interaction in the crystal is stronger (the lattice energy

TABLE V. Chemical potential of urea in solution determined using thermodynamic integration as a function of urea concentration (mole fraction), showing the various free energy components for the Özpınar/TIP3P urea model combination. The chemical potential μ'_{sol} is the sum of the ideal component G_{ideal}, calculated with the effective solute density ρ_s, and the two component terms comprising the free energy of solvation, ΔG_{vdw} and ΔG_{qq}, and excludes the intramolecular term (which is assumed to be identical to that in the crystal and, hence, cancels out in the solubility calculation) as defined by Eq. (16). Estimated standard errors are given in brackets.

Urea concentration		Component free energies (kJ mol^{-1})			
X	ρ_s (nm^{-3})	G_{ideal}	ΔG_{vdw}	ΔG_{qq}	μ'_{sol}
0.0005	0.02	−36.714	2.43(6)	−61.37(8)	−95.7(1)
0.001	0.03	−34.986	2.26(6)	−61.30(8)	−94.0(1)
0.011	0.36	−29.040	2.26(6)	−61.42(8)	−88.2(1)
0.031	0.98	−26.508	1.72(7)	−61.34(8)	−86.1(1)
0.051	1.58	−25.324	1.33(7)	−61.43(8)	−85.4(1)
0.101	2.97	−23.759	0.53(7)	−61.35(8)	−84.6(1)
0.151	4.21	−22.888	0.12(7)	−61.44(8)	−84.2(1)
0.201	5.33	−22.302	−0.30(7)	−61.48(8)	−84.1(1)
0.251	6.33	−21.869	−0.29(7)	−61.63(8)	−83.8(1)
0.301	7.25	−21.527	−0.63(7)	−61.61(9)	−83.8(1)

TABLE VI. Chemical potential of urea in solution determined using thermodynamic integration as a function of urea concentration (mole fraction) showing the various free energy components for the Hölzl/(TIP4P/2005) urea model combination. The chemical potential μ'_{sol} is the sum of the ideal component G_{ideal}, calculated with the effective solute density ρ_s, and the two component terms comprising the free energy of solvation ΔG_{vdw} and ΔG_{qq}, and excludes the intramolecular term (which is assumed to be identical to that in the crystal and, hence, cancels out in the solubility calculation) as defined by Eq. (16). Estimated standard errors are given in brackets.

Urea concentration		Component free energies (kJ mol^{-1})			
X	ρ_s (nm^{-3})	G_{ideal}	ΔG_{vdw}	ΔG_{qq}	μ'_{sol}
0.0005	0.02	−36.677	5.15(7)	−76.6(1)	−108.1(1)
0.001	0.03	−34.929	5.17(7)	−76.5(1)	−106.2(1)
0.011	0.37	−28.980	5.09(7)	−76.6(1)	−100.5(1)
0.031	1.00	−26.473	5.04(7)	−76.7(1)	−98.1(1)
0.051	1.61	−25.289	4.95(7)	−76.9(1)	−97.2(1)
0.101	2.98	−23.750	4.85(7)	−77.1(1)	−96.0(1)
0.151	4.18	−22.903	4.66(7)	−76.88(9)	−95.1(1)
0.201	5.23	−22.343	4.95(7)	−76.85(9)	−94.2(1)
0.251	6.19	−21.925	4.77(7)	−76.95(9)	−94.1(1)
0.301	7.03	−21.607	5.00(8)	−77.17(9)	−93.8(1)

is about 11% higher than the experimental value—see Table II), its interaction with the water model is also stronger.

The chemical potential-based solubilities along with the solubility bounds predicted by direct coexistence are shown in Figs. 1 and 5. The data reveal that for both urea models, the direct coexistence solubility bounds are quite broad and (against our expectations) do not encompass the chemical potential-based solubility, although the lower bound is close to it. The breadth of the solubility bounds from coexistence simulations is substantially larger than that observed in previous studies, which are all restricted to either NaCl or the Lennard-Jones system. Unlike NaCl, urea is a molecule. Its deposition on its crystal surfaces involves not just translational but also orientational alignment, along with the decoupling of the solvent from both the attaching molecule and the crystal surface. We should, therefore, expect slow kinetics, particularly for the deposition process, and hence extremely long equilibration times and broad solubility bounds. Furthermore, when the crystal surface is perfectly flat, that is, without kinks or terraces, both surface deposition and dissolution will require a 2-D nucleation event to initiate a new crystal layer or the dissolution of the surface. (In the case of dissolution, the 2-D nucleus takes the form of a circular trough in the surface layer characterized by missing molecules.) The critical size of the 2-D nucleus is given by the classical nucleation theory[49] (in 2-D)[43] and depends on the extent of supersaturation or degree of under-saturation in the case of dissolution. This critical nucleus size approaches infinity at the solubility limit (saturation), the implication being that the dissolution or growth of a new crystal layer becomes increasingly more challenging as we approach saturation from either direction—dissolution in an under-saturated solution or crystal growth from a super-saturated solution. The issue is considered to be particularly acute for molecular crystals, where the entropic component of the nucleation free energy barrier is substantive due to the need for orientational alignment of the molecules at the crystal surface. We believe this to be the cause of the broad solubility bounds observed in the direct coexistence simulation for urea. These considerations suggest that the simulations should transiently converge at solubility values that correspond to a perfect flat surface, which is indeed what we observe.

With respect to the lower solubility bound being slightly above the chemical potential-based solubility, the disparity possibly arises from the choice of using the faster dissolving (001) crystal face with its higher interfacial free energy (surface chemical potential) for the coexistence simulations. There is a possibility that this high, surface chemical potential causes the amount of solute that dissolves from the crystal face to overshoot the solution saturation limit. Whilst this should be countered by the inverse process of solute deposition, the over-saturation may be insufficient to initiate the 2-D nucleation required to initiate the deposition of a new crystal layer once a complete surface without kinks or terraces has developed. This would result in a slightly over-saturated solution at convergence, yielding a

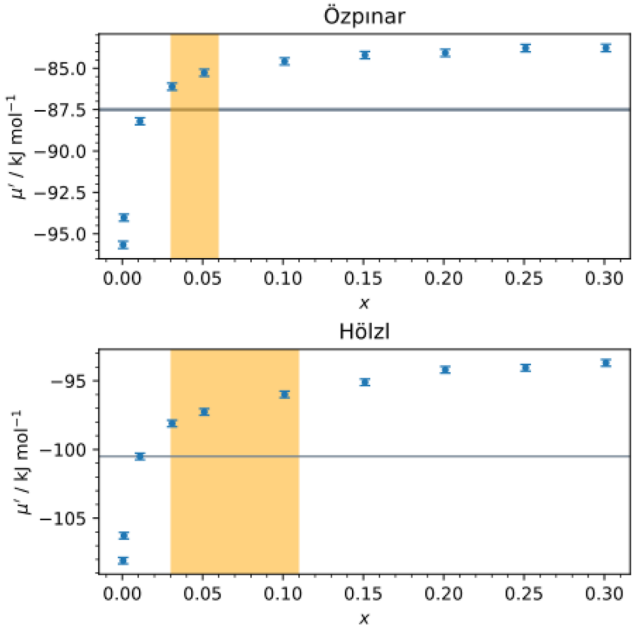

FIG. 5. Solubility prediction for urea in aqueous solution for the two urea model combinations, Özpınar/TIP3P and Hölzl/(TIP4P/2005). The horizontal, gray line is the chemical potential of the crystal phase, whilst the blue data points are the chemical potential of urea as a function of the urea mole fraction x of the solution (excluding the inserted solute molecule). The solubility is given by the intersection of the crystal chemical potential line and the solution chemical potential curve. The vertical, orange band is the region defined by the upper and lower solubility bounds determined from direct coexistence simulations. The predicted solubility from the chemical potentials for both models lies just below the lower solubility bound determined from direct coexistence. The estimated solubilities are 0.015 and 0.010 mole fractions for the Özpınar and the Hölzl models, respectively.

higher apparent solubility. Clearly, in-depth direct coexistence studies are required for molecular crystals, which we have now initiated for urea.

On the basis of confidence in our chemical potential-based solubility calculations, urea is yet another example for which the predicted solubility from molecular simulation is markedly underestimated relative to experiment, suggesting significant issues with the force fields. Whilst it is hard to generalize as we have only a few examples (caffeine,[20] aspirin,[21] and urea[12]), we note that these molecules are either sparingly soluble or soluble molecules, and the medium is water. The gross disparity in the predicted solubility should not be surprising. Phase equilibria continue to challenge our current force fields, with an illustrative example being the predicted freezing points of the commonly employed simple models of water (TIP3P, TIP4P, and SPC/E) being 40–60 K below the experimental value.[50–52] Whilst the simple water models have now been optimized to reproduce the liquid and solid phase properties, we expect the optimization of force fields to reproduce phase equilibria (such as solubility) for heterogeneous systems to be a significant challenge because of the subtle interplay of the affinity of a molecule for its solid phase and for the solvent, which may require revisiting some commonly employed assumptions like the use of the Lorentz-Berthelot rules.[53,54]

Summarizing, the aqueous solubility of urea has been estimated at 300 K using two distinct urea models, Özpınar/TIP3P and Hölzl/(TIP4P/2005) model combinations, using chemical potential calculations employing both thermodynamic integration and perturbation methodology. Although the simulated urea molecule was flexible (but with the bonds being constrained), the methodology assumes that the intramolecular contribution to the chemical potential is identical for both the crystal and for the solution and, hence, cancels out. We believe this to be a fair assumption, as the flexibility of the urea molecule is limited. The predicted solubilities for the two models are about 10 fold lower than the experimental solubility.[44,48] The chemical potential calculations have been supplemented by direct coexistence dissolution and crystal growth simulations to yield upper and lower bounds of the solubility for the two models. The predicted solubility bounds are broad, with the upper estimate being 2–4 times the lower estimate, and do not encompass the chemical potential-based solubility, although the lower bound is close to it. Whilst we have rationalized these results on the basis of significant entropic energy barriers associated with the need to align molecules at the interface for the case molecular crystals, more in-depth direct coexistence is needed. Assuming confidence in our chemical potential-based calculation, we attribute the disparity in the predicted and experimental solubility to the inadequacy of the force field models.

ACKNOWLEDGMENTS

The simulations were carried out on Lancaster University's high performance computing facility, HEC.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

James Carruthers: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Mauro Ferrario**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jamshed Anwar**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Input files for all simulations carried out for this study can be obtained on request.

REFERENCES

[1] H. D. Williams, N. L. Trevaskis, S. A. Charman, R. M. Shanker, W. N. Charman, C. W. Pouton, and C. J. H. Porter, Pharmacol. Rev. **65**, 315 (2013).
[2] C. R. Gardner, C. T. Walsh, and Ö. Almarsson, Nat. Rev. Drug Discovery **3**, 926 (2004).
[3] N. Earl, C. Cartwright, S. Horrocks, M. Worboys, S. Swift, J. Kirton, A. Askan, H. Kelleher, and D. Nancarrow, Environment Agency, Bristol, 2003.
[4] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Adv. Drug Delivery Rev. **64**, 4 (2012).
[5] R. E. Skyner, J. L. McDonagh, C. R. Groom, T. van Mourik, and J. B. O. Mitchell, Phys. Chem. Chem. Phys. **17**, 6174 (2015).
[6] L. D. Hughes, D. S. Palmer, F. Nigsch, and J. B. O. Mitchell, J. Chem. Inf. Model. **48**, 220 (2008).
[7] M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, and P. Turq, J. Chem. Phys. **117**, 4947 (2002).
[8] A. L. Benavides, J. L. Aragones, and C. Vega, J. Chem. Phys. **144**, 124504 (2016).
[9] J. L. Aragones, E. Sanz, and C. Vega, J. Chem. Phys. **136**, 244508 (2012).
[10] A. S. Paluch, D. L. Mobley, and E. J. Maginn, J. Chem. Theory Comput. **7**, 2910 (2011).
[11] S. Boothroyd, A. Kerridge, A. Broo, D. Buttar, and J. Anwar, Phys. Chem. Chem. Phys. **20**, 20981 (2018).
[12] S. Boothroyd and J. Anwar, J. Chem. Phys. **151**, 184113 (2019).
[13] M. Lísal, W. R. Smith, and J. Kolafa, J. Phys. Chem. B **109**, 12956 (2005).
[14] F. Moucka, M. Lísal, J. Škvor, J. Jirsák, I. Nezbeda, and W. R. Smith, J. Phys. Chem. B **115**, 7849 (2011).
[15] F. Moucka, I. Nezbeda, and W. R. Smith, J. Chem. Theory Comput. **11**, 1756 (2015).
[16] Z. Bjelobrk, D. Mendels, T. Karmakar, M. Parrinello, and M. Mazzotti, Cryst. Growth Des. **21**, 5198 (2021).
[17] H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, and J. Alejandre, J. Phys. Chem. B **119**, 8389 (2015).
[18] J. Kolafa, J. Chem. Phys. **145**, 204509 (2016).
[19] J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, and A. Z. Panagiotopoulos, J. Chem. Phys. **145**, 154111 (2016).
[20] L. Li, T. Totton, and D. Frenkel, J. Chem. Phys. **146**, 214110 (2017).
[21] G. Duarte Ramos Matos and D. L. Mobley, F1000Research **7**, 686 (2018).
[22] J. A. Schellman, Biophys. Chem. **37**, 121 (1990).
[23] M. C. Stumpe and H. Grubmüller, J. Phys. Chem. B **111**, 6220 (2007).
[24] S. Piana and J. D. Gale, J. Am. Chem. Soc. **127**, 1975 (2005).
[25] D. Frenkel and A. J. C. Ladd, J. Chem. Phys. **81**, 3188 (1984).
[26] C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, J. Phys.: Condens. Matter **20**, 153101 (2008).
[27] J. L. Aragones, C. Valeriani, and C. Vega, J. Chem. Phys. **137**, 146101 (2012).
[28] J. L. Aragones, E. G. Noya, C. Valeriani, and C. Vega, J. Chem. Phys. **139**, 034104 (2013).
[29] K. Ioannidou, M. Kanduč, L. Li, D. Frenkel, J. Dobnikar, and E. Del Gado, Nat. Commun. **7**, 12106 (2016).
[30] M. S. Sellers, M. Lísal, and J. K. Brennan, Phys. Chem. Chem. Phys. **18**, 7841 (2016).
[31] M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout, J. Chem. Phys. **150**, 094107 (2019).
[32] G. Gobbo, G. Ciccotti, and B. L. Trout, J. Chem. Phys. **150**, 201104 (2019).
[33] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. Van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).
[34] J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, J. Chem. Phys. **112**, 5339 (2000).
[35] G. A. Özpınar, F. R. Beierlein, W. Peukert, D. Zahn, and T. Clark, J. Mol. Model. **18**, 3455 (2012).
[36] C. Hölzl, P. Kibies, S. Imoto, J. Noetzel, M. Knierbein, P. Salmen, M. Paulus, J. Nase, C. Held, G. Sadowski, D. Marx, S. M. Kast, and D. Horinek, Biophys. Chem. **254**, 106260 (2019).
[37] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[38] S. Weerasinghe and P. E. Smith, J. Phys. Chem. B **107**, 3891 (2003).
[39] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1-2**, 19 (2015).
[40] J. E. Worsham, H. A. Levy, and S. W. Peterson, Acta Crystallogr. **10**, 319 (1957).
[41] P. V. Klimovich, M. R. Shirts, and D. L. Mobley, J. Comput.-Aided Mol. Des. **29**, 397 (2015).
[42] J. Anwar and D. M. Heyes, J. Chem. Phys. **122**, 224117 (2005).
[43] P. Ectors, J. Anwar, and D. Zahn, Cryst. Growth Des. **15**, 5118 (2015).
[44] P. Jain, Y. He, and S. Yalkowsky, *Handbook of Aqueous Solubility Data*, 2nd ed. (CRC Press, 2010), p. 1608; arXiv:1011.1669v3.
[45] E. M. Duffy, D. L. Severance, and W. L. Jorgensen, Isr. J. Chem. **33**, 323 (1993).
[46] L. J. Smith, H. J. C. Berendsen, and W. F. van Gunsteren, J. Phys. Chem. B **108**, 1065 (2004).
[47] H. Kokuba, J. Rosgen, D. W. Bolen, and B. M. Pettitt, Biophys. J. **93**, 3392–3407 (2007).
[48] D. Niether, S. Di Lecce, F. Bresme, and S. Wiegand, Phys. Chem. Chem. Phys. **20**, 1012 (2018).
[49] J. Anwar and D. Zahn, Angew. Chem., Int. Ed. **50**, 1996–2013 (2011).
[50] T. Bryk and A. D. J. Haymet, Mol. Simul. **30**, 131 (2004).
[51] G. T. Gao, X. C. Zeng, and H. Tanaka, J. Chem. Phys. **112**, 8534 (2000).
[52] C. Vega, E. Sanz, and J. L. F. Abascal, J. Chem. Phys. **122**, 114507 (2005).
[53] H. A. Lorentz, Ann. Phys. **248**, 127 (1881).
[54] D. Berthelot, Weekly Records of the Meetings of the Academy of Sciences (1898), Vol. 126, pp. 1703–1855.