A streamlined molecular-dynamics workflow for computing solubilities of molecular and ionic crystals

Aleks Reinhardt; Pin Yu Chew ⓘ ; Bingqing Cheng ✉ ⓘ

Check for updates

View Online   Export Citation   CrossMark

A streamlined molecular-dynamics workflow for computing solubilities of molecular and ionic crystals

View Online   Export Citation   CrossMark

Aleks Reinhardt,[1,a)] Pin Yu Chew,[1] (iD) and Bingqing Cheng[2,b)] (iD)

AFFILIATIONS

[1] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[2] Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

[a)] Electronic mail: ar732@cam.ac.uk
[b)] Author to whom correspondence should be addressed: bingqing.cheng@ist.ac.at

ABSTRACT

Computing the solubility of crystals in a solvent using atomistic simulations is notoriously challenging due to the complexities and convergence issues associated with free-energy methods, as well as the slow equilibration in direct-coexistence simulations. This paper introduces a molecular-dynamics workflow that simplifies and robustly computes the solubility of molecular or ionic crystals. This method is considerably more straightforward than the state-of-the-art, as we have streamlined and optimised each step of the process. Specifically, we calculate the chemical potential of the crystal using the gas-phase molecule as a reference state, and employ the S^0 method to determine the concentration dependence of the chemical potential of the solute. We use this workflow to predict the solubilities of sodium chloride in water, urea polymorphs in water, and paracetamol polymorphs in both water and ethanol. Our findings indicate that the predicted solubility is sensitive to the chosen potential energy surface. Furthermore, we note that the harmonic approximation often fails for both molecular crystals and gas molecules at or above room temperature, and that the assumption of an ideal solution becomes less valid for highly soluble substances.

I. INTRODUCTION

The solubility quantifies the maximum amount of a material, known as the solute, that can be dissolved in a solvent at equilibrium. A knowledge of the solubility is crucial for a wide range of applications, ranging from pharmaceutical drug formulation to crystal growth, and from chemical synthesis to the phase separation of mixtures.[1–3] For example, to ensure that a drug is absorbed at a suitable rate, it should have an appropriately high or low solubility,[4] so it is useful to estimate the solubility as part of the screening process even before the drug is synthesised experimentally. A range of approaches have been adopted in the literature to predict solubilities,[5] from informatics-based methods[6] to explicit-solvent computational models.[7–20]

However, predicting solubility computationally poses challenges for several reasons. First, the thermodynamic equilibrium between the solute and the solution is governed by both energetic and entropic factors. Addressing the entropic components often demands intricate and costly computations. Second, it is not straightforward to design the thermodynamic cycles necessary to perform these computations. Finally, it is important to model the molecular interactions accurately. Empirical force fields are often parameterised to describe either the liquid or the solid phase of a material accurately, but capturing both simultaneously is much more challenging.

The solubility of a crystal corresponds to the solution concentration c at which there is an equilibrium between the solid solute and the solution; at a known pressure P and temperature T, the equilibrium is reached when the chemical potentials of the two phases are equal, $\mu^{\text{crystal}}(P, T) = \mu^{\text{sol}}(P, T, c)$. In principle, it may be possible to determine this equilibrium using direct-coexistence simulations[21,22] by placing the crystal in the same simulation box

as a solution, and determining what the concentration of the solution is once the amount of crystal no longer changes.[23,24] However, such simulations are difficult in practice, since crystal dissolution and precipitation are complex kinetic processes which usually require equilibration times beyond typical timescales of molecular-dynamics (MD) simulations, particularly so when either solutes or solvents are large molecules and the systems of necessity entail large numbers of atoms. Moreover, many atomic and molecular systems have numerous polymorphs, i.e., several distinct possible crystal structures. Although typical solubility ratios of different crystalline polymorphs are within a factor of two, there are exceptions,[25] and each polymorph can in principle have significantly different solubility properties.[1,3] Predicting which polymorph will be the most stable or the most soluble under specific conditions requires a consideration of numerous possible crystal structures, further complicating computational efforts.

To avoid direct-coexistence simulations, one can independently compute the free energies of the crystal[26–28] and the solution.[7,29–31] For atomic crystals, the Einstein crystal method[26,32] can be used to obtain the absolute free energy by exploiting hamiltonian thermodynamic integration,[26] also commonly referred to as artificial or alchemical thermodynamic integration, to switch between the Einstein crystal potential (U_0), for which each atom is attached to its equilibrium position by a harmonic spring, and the potential of interest (U_1). For molecular crystals of flexible molecules, the extended Einstein crystal method was developed,[7,8,11,33] where the molecular crystal is transformed into an Einstein crystal of independent molecules via a multi-step workflow, including turning on additional harmonic restraints on selected atoms of each molecule to control the orientation of the molecules, subsequently turning on Van der Waals and electrostatic interactions between molecules, and finally turning off all harmonic restraints. This workflow can be somewhat convoluted, and it can be difficult to determine which atoms needs to be tethered and what spring strength is needed on each tethered atom.

The free energy of the molecule in the solution can be computed as the free-energy difference between a solvated and a gas-phase molecule. In a dilute solution, this can be determined by computing the chemical potential of a single solute molecule in a box of solvent using hamiltonian thermodynamic integration by switching on the interactions between the solute and the solvent.[7] In practice, such an approach suffers from origin singularities when the particle is first inserted.[34] This singularity can be circumvented using the cavity method: a cavity can be switched on and increased in size to accommodate the solute molecule, the solute molecule then added, and finally the cavity is gradually switched off.[7] For sufficiently dilute solutions, an ideal-dilute solution approximation can be used to estimate the chemical potential at other concentrations; in other words, if Henry's law applies to the solute, $\mu^{sol}(x) = \mu^{sol}(x \to 0) + k_B T \ln x$, where x is the mole fraction of the solute. If the solute is not very poorly soluble, it is likely that deviations from ideality will necessitate several laborious calculations of the solvation free energy as the concentration increases.[11]

The chemical potentials of the solute in solution and in the crystal must have a common origin. In the procedure outlined above, this is achieved by the stepwise calculation of the extended Einstein crystal's free energy. Since the calculation passes through

a freely rotating state, the rotational partition function cancels out, and as long as the intramolecular partition functions in the crystal and the solution are the same, a consistent thermodynamic state is obtained.[7,11] In the context of computing solid–liquid phase equilibria, it has been suggested that a simpler alternative approach can be used which avoids this effective stepwise solid–fluid transformation;[9,10] by computing the absolute free energy of the molecule in the gas phase explicitly, the liquid or solution free energy can be determined relative to the same origin as the molecular crystal.

In this work, we introduce a set of steps that simplify the calculation of solubilities. We use a combination of thermodynamic integration (TI),[26] free-energy perturbation (FEP)[35] and the S[0] method[30] to compute the chemical potentials required. In our approach, we employ a Debye crystal reference for solid phases, reducing the need for manual intervention compared to the extended Einstein-crystal method previously discussed. We show how we can compute chemical potentials with a common baseline by using a gas-phase Debye molecule reference. We showcase this workflow by calculating the solubilities of an ionic solid (sodium chloride) and molecular crystals with varied solubilities, including paracetamol in water, paracetamol in ethanol, and urea in water.

II. METHODS

The general workflow is illustrated in Fig. 1. The absolute chemical potentials of the crystalline phase, the ideal-gas molecule and the solute are each computed separately. By "absolute," we mean that the chemical potentials have a common origin, irrespective of the reference state from which they are computed. Specifically, the baseline of the chemical potential for a system comprising molecules with internal bonding, irrespective of the phase, is perhaps most naturally taken to be the chemical potential of an isolated molecule with intramolecular interactions. For instance, when determining the chemical potential of a crystalline phase relative to the isolated atoms, we should deduct the baseline chemical potential of the isolated molecule. This procedure mirrors constructing a thermodynamic cycle between the crystal and the solution via the gas-phase molecule state.

For the crystal, we use a reference Debye harmonic crystal (Sec. II A). We then determine the baseline chemical potential of an isolated molecule in the gas phase; to do this, we again use a harmonic reference state in which all pairs of atoms are connected with harmonic springs, and separately add analytical chemical potentials of translational and rotational degrees of freedom (Sec. II B). Finally, for the solute, we split the calculation into two parts. We first compute the solvation free energy at a certain low concentration using a combination of TI and FEP (Sec. II C 1), and then determine the chemical potential of the solution as a function of concentration with the S[0] method (Sec. II C 2).

A. Crystalline states

We use the Debye crystal reference[27,36,38] as the starting point of TI. This reference gives a chemical potential baseline of isolated atoms; we account for this in the next step of the calculation by shifting the baseline to that of isolated molecules. The Debye crystal has all pairs of atoms connected with harmonic springs
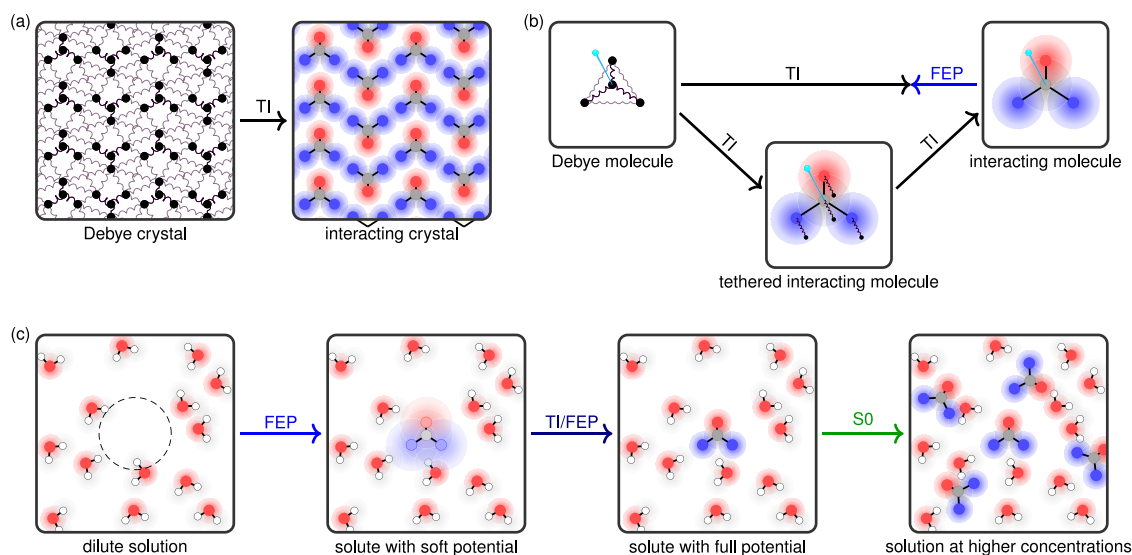
FIG. 1. A schematic illustration of the molecular-dynamics workflow for determining chemical potentials. (a) Crystalline solute. We use the normal-mode approximation to construct a reference Debye crystal with a known free energy, and then use thermodynamic integration (TI) to switch off the Debye harmonic interaction as the potential of interest is switched on. (b) Isolated molecule. We construct a normal-mode Debye-molecule landscape with a known free energy. We constrain the molecule's centre of mass and rotational motion, shown pictorially with a cyan pin, and account for these contributions analytically. We then follow one of two routes: we either (1) use hamiltonian thermodynamic integration until most of the potential is switched on and compute the final part by a free-energy perturbation (FEP) method, or (2) use thermodynamic integration to a tethered interacting molecule, and then another thermodynamic integration to switch off the tethering. (c) Solution. We start with the pure solvent or a very dilute solution. We use a free-energy perturbation to add an additional solute molecule with a soft potential, and then use either thermodynamic integration or further free-energy perturbations to switch the soft potential to the fully interacting system. Finally, we use the S^0 method to compute how the chemical potential changes from this initial low concentration as the solute is concentrated.

[as schematically shown in Fig. 1(a)], and it has the same mass-weighted hessian matrix as the physical crystal. The hessian can be computed by expanding the potential energy surface around the (local) minimum-energy state in a numerical normal-mode approximation.[39] To this end, we first find the potential-energy minimum by conjugate gradient minimisation.[40] To estimate the second derivative of the potential energy using a finite-difference approach,[38] we successively move each pair of atoms by an arbitrary small displacement relative to each other in each Cartesian direction in turn while keeping other distances unchanged, as implemented in i-PI.[41] The estimated second derivative is then scaled by the square roots of the masses of the two atoms. Once all components of the mass-weighted hessian matrix are known, we numerically find its eigenvalues. These eigenvalues, κ_i, correspond to the square roots of the normal-mode angular frequencies,[39,42] $\omega_i = \sqrt{\kappa_i}$, where i identifies each of the normal modes. The Helmholtz energy of the Debye crystal with a constrained centre of mass (CM) at T_0 is

$$A_h = k_B T_0 \sum_{i=1}^{3N-3} \ln \frac{\hbar \omega_i}{k_B T_0}, \qquad (1)$$

where the sum excludes the three normal modes with zero eigenvalues corresponding to translations of the entire system. For TI, we define a simple linear scaling of the overall potential of a system, $U(\lambda) = \lambda U_1 + (1 - \lambda) U_0$,[43] so that the potential changes from the reference harmonic potential U_0 to the potential of interest U_1 as the parameter λ changes from 0 to 1. Using hamiltonian thermodynamic integration ("λ-TI"), the Helmholtz energy of the system governed by U_1 is[26,27,32,38]

$$A_1 = A_0 + \int_0^1 \langle U_1 - U_0 \rangle_\lambda \, d\lambda, \qquad (2)$$

where in this case, $A_0 = A_h$ given by Eq. (1). Once we have the Helmholtz energy, we can obtain the Gibbs energy,[27] e.g., via $G = A + PV$, and hence the chemical potential $\mu = G/N$.

Since the harmonic approximation is just the potential energy expanded to quadratic terms, at sufficiently low temperatures, the integral is smooth and the procedure is efficient by design. As the temperature increases, however, anharmonic effects begin to play a role and the λ-TI becomes progressively more difficult. In particular, rotations about single bonds within each molecule can be activated at higher temperatures, and although in principle this should not lead to a discontinuity in $\langle U_1 - U_0 \rangle_\lambda$ at $\lambda = 1$, the potential energy difference can become large and the numerical integration may require many data points and long equilibration times. It is thus preferable to perform the λ-TI at sufficiently low temperatures when the integral is well-behaved. Once the Gibbs energy is determined at the initial low temperature T_0, we can find how it changes

with temperature by numerically integrating the Gibbs–Helmholtz equation,

$$\frac{G(T)}{T} = \frac{G(T_0)}{T_0} - \int_{T_0}^{T} \frac{H(T')}{(T')^2} \mathrm{d}T', \qquad (3)$$

where $H = U + PV$ is the enthalpy of the system.

B. Gas-phase molecule

To obtain a consistent origin for the absolute chemical potentials of the crystal and the solution, we subtract the free energy of the gas-phase molecule from the free energy of the crystal referenced to the isolated atom state. To compute the free energy of the isolated molecule, we perform a λ-TI between a harmonic reference and the physical system [Fig. 1(b)]. It is advantageous, and indeed often necessary, to constrain the centre of mass (CM) and the rotational degrees of freedom of the molecule when performing TI, and then one can add explicit free-energy contributions from the free ideal-gas particle to correct for the CM constraint (A_{cm}), as well as the rigid-body rotations of the entire molecule (A_{rot}).

The Helmholtz energy associated with the free CM in a volume V at temperature T is

$$A_{\mathrm{cm}} = -k_{\mathrm{B}}T \ln\left[V\left(\frac{mk_{\mathrm{B}}T}{2\pi\hbar^2}\right)^{3/2} \right], \qquad (4)$$

and the rotational Helmholtz energy is[44]

$$A_{\mathrm{rot}} = -k_{\mathrm{B}}T \ln\left[\sqrt{\pi}\left(\frac{2k_{\mathrm{B}}T}{\hbar^2}\right)^{3/2} \sqrt{I_1 I_2 I_3} \right], \qquad (5)$$

where I_1, I_2 and I_3 are the principal components of the moment of inertia tensor. The rotational partition function sometimes entails a division by a symmetry number to account for the number of equivalent rotations of the molecule; however, in this framework, we do not do so because the same degeneracies are also ignored when sampling the free energy of molecular crystals with restricted molecular orientations.

When performing the λ-TI from the reference system to the potential of interest, although in principle one can use any harmonic or ideal-gas reference, choosing a reference harmonic molecule that has the same frequency modes and equilibrium configuration as the real molecule is statistically efficient. Selecting such a reference system follows the same procedure as setting up the Debye crystal in Sec. II A, except that a total of 6 degrees of freedom are removed (three translational and three rotational). The classical Helmholtz energy of such a constrained Debye molecule at temperature T_0 is

$$A_{\mathrm{h}}(V, T_0) = k_{\mathrm{B}}T_0 \sum_{i=1}^{3N-6} \ln\frac{\hbar\omega_i}{k_{\mathrm{B}}T_0}, \qquad (6)$$

where ω_i is the angular frequency of normal mode i.

Finally, the Helmholtz energy of the real molecule with a fixed CM and orientation can be obtained using λ-TI [Eq. (2)]. As with the Debye crystal, in order to improve the numerical convergence of this thermodynamic integration, this step should be performed

at some low temperature T_0 at which the system is quasi-harmonic. Although for fairly rigid molecules this procedure works well, for molecules with different conformers or rotating bonds, the integrand can still adopt very large values as $\lambda \to 1$. To circumvent this, one can envisage two approaches, as illustrated in Fig. 1(b). The first approach is to perform a FEP in lieu of the final stage of the TI from $1 - \varepsilon$ to 1, i.e.,

$$A(\lambda = 1) - A(\lambda = 1 - \varepsilon) = k_{\mathrm{B}}T_0 \ln\left\langle \exp\left[-\frac{\varepsilon(U_0 - U_1)}{k_{\mathrm{B}}T_0} \right] \right\rangle_{\lambda=1}, \quad (7)$$

where ε is a small number close to zero. This procedure avoids the end-point problem at $\lambda = 1$ in TI because in FEP, for the large values of $U_0 - U_1$ involved, the exponential almost vanishes and thus has little influence on the ensemble average of Eq. (7). Note that FEP requires sufficient overlap between the two states to get statistically unbiased estimates, so a small ε is often needed.[45,46] The second approach is to perform a TI between the harmonic reference and a tethered molecule (the physical system plus harmonic springs connecting each atom to its equilibrium position), and another TI between the tethered and the untethered molecule. Finally, the temperature dependence of the free energy from T_0 to T can be computed using the Helmholtz-energy analogue of Eq. (3).

C. Solute in solution

1. Solvation free energy

The solvation free energy is the free-energy difference between a solute molecule that is fully interacting with a dilute solution and the same molecule in the ideal-gas state. To compute it, in principle we can use λ-TI [Eq. (2)] between these two states directly. Here the potential of interest U_1 corresponds to the molecule fully interacting with its surroundings at some concentration of the solution. The reference potential U_0 corresponds to the solute molecule not interacting with its surroundings, but where the intramolecular energy of the molecule, as well as the interactions between solvents, are the same as in U_1. Typically, we use a solution with one molecule dissolved in a box of the solvent for this purpose.

Many procedures for the hamiltonian switching are available.[14,47] The simplest is to use linear λ scaling and get the free energy difference using Eq. (2). However, a naïve implementation of the λ-TI results in a divergent integrand in Eq. (2) as $\lambda \to 0$: when the "ghost" molecule and the rest of the system are not interacting, atoms can overlap and result in an infinite value for $U_1 - U_0$. To circumvent this problem, one can instead perform a FEP at the end point $\lambda = 0$[48] [as highlighted with the blue arrow in Fig. 1(c)], i.e.,

$$A(\lambda = \varepsilon) - A(\lambda = 0) = -k_{\mathrm{B}}T \ln\left\langle \exp\left[-\frac{\varepsilon(U_1 - U_0)}{k_{\mathrm{B}}T} \right] \right\rangle_{\lambda=0}, \quad (8)$$

and then compute the integral in Eq. (2) from ε to 1 in λ only. Here, the large values of $U_0 - U_1$ will contribute little to the ensemble average in Eq. (8).

Another approach that avoids numerical convergence issues in TI is the use of multiple-step FEP using a λ-dependent soft-core

potential. For example, a possible soft-core functional form of the Lennard-Jones potential is[47]

$$U_{\mathrm{LJ}}^{\mathrm{soft}} = \lambda 4\varepsilon_{\mathrm{LJ}} \left\{ \left[\alpha_{\mathrm{LJ}}(1-\lambda)^2 + \left(\frac{r}{\sigma_{\mathrm{LJ}}}\right)^6 \right]^{-2} \right.$$
$$\left. - \left[\alpha_{\mathrm{LJ}}(1-\lambda)^2 + \left(\frac{r}{\sigma_{\mathrm{LJ}}}\right)^6 \right]^{-1} \right\}, \quad (9)$$

where $\varepsilon_{\mathrm{LJ}}$ and σ_{LJ} are appropriate energy and distance units, and α_{LJ} controls the switching protocol. Similarly, a soft-core Coulomb interaction is

$$U_{\mathrm{C}}^{\mathrm{soft}} = \lambda \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_{\mathrm{C}}\left[\alpha_{\mathrm{C}}(1-\lambda)^2 + r^2\right]^{1/2}}, \quad (10)$$

where q_i and q_j are the charges on the two interacting atoms, ε_0 is the electric constant, ε_{C} is the dielectric constant, and α_{C} controls the switching.

When using multiple-step FEP, one uses a number of intermediate states, for example, $U(\lambda_i)$ with $\lambda_i = i/M$ for $i \in \{0, 1, \ldots, M-1\}$. Compared to forward or backward FEP, we have found it to be statistically efficient to combine a forward and a backward FEP at half steps.[49] For example, to compute the free-energy difference between two potential-energy surfaces U_{a} and U_{b}, one runs the simulation using the potential $U_{1/2} = (U_{\mathrm{a}} + U_{\mathrm{b}})/2$, and the estimator is

$$A_{\mathrm{b}} - A_{\mathrm{a}} = -k_{\mathrm{B}}T \ln \left\langle \exp\left[-\frac{U_{\mathrm{b}} - U_{\mathrm{a}}}{2k_{\mathrm{B}}T}\right] \right\rangle_{U_{1/2}}$$
$$+ k_{\mathrm{B}}T \ln \left\langle \exp\left[-\frac{U_{\mathrm{a}} - U_{\mathrm{b}}}{2k_{\mathrm{B}}T}\right] \right\rangle_{U_{1/2}}. \quad (11)$$

As the efficiency of FEP crucially depends on the overlap between the two states involved, the half-step scheme reduces the difference in energy by half and thus enhances the efficiency.

2. Concentration-dependent chemical potential

In the previous step, we computed the solvation free energy, and hence chemical potential μ_0^{sol}, at some low concentration c_0. To find how this quantity changes as a function of solution concentration, we use the recently proposed S^0 method,[30] as shown schematically by the green arrow in Fig. 1(c). The S^0 method is based on the thermodynamic relationship between fluctuations in particle numbers and derivatives of the chemical potentials with respect to the molar concentration,[29] and only uses the static structure factors computed from equilibrium MD simulations.[30] Specifically, we perform multiple equilibrium MD simulations at different concentrations and find $\mu^{\mathrm{sol}}(c)$ by numerical integration with respect to $\ln c$,

$$\mu^{\mathrm{sol}}(c) = \mu^{\mathrm{sol}}(c_0) + k_{\mathrm{B}}T \ln (c/c_0)$$
$$+ k_{\mathrm{B}}T \int_{\ln c_0}^{\ln c} \mathrm{d} \ln (c)\left[\frac{1}{S_{\mathrm{M-M}}^0 - S_{\mathrm{M-S}}^0\sqrt{c/c_{\mathrm{S}}}} - 1\right], \quad (12)$$

where the subscripts M and S denote solute and solvent molecules, respectively. $S_{\mathrm{M-M}}^0$ is the static structure factor in the $k \to 0$ limit

between solute molecules and $S_{\mathrm{M-S}}^0$ is the equivalent between solute and solvent molecules.

III. RESULTS

A. Sodium chloride in water

To model NaCl and water, we used the JC/SPC/E[23] force field, with the Lennard-Jones interactions truncated at 1 nm with tail corrections, and long-range Coulomb interactions were treated using a particle–particle particle-mesh solver. Since NaCl is made up of individual ions, we do not need to explicitly compute the free energy of the gas-phase molecule in this case, as the reference state for the solvation free energy and the crystal free energy are both a pair of ideal-gas NaCl ions with an analytic free energy.

For the solid phase, we performed a λ-TI from the Debye crystal to the physical system at 298.15 K. The result is consistent with a λ-TI first performed at a lower T of 50 K and then an integration along an isobar [Eq. (3)]. The chemical potential for each ion pair in the solid phase 298.15 K is (-190.145 ± 0.015) kcal/mol.

To compute the solvation free energy, we used i-PI[41] to perform TI between two states: (i) a pair of Na and Cl ions that are non-interacting and a box of water (with 510 water molecules), and (ii) Na and Cl ions that are fully interacting with water and with each other and a box of water. We note that finite-size effects of such systems with long-range electrostatics may be corrected by developing methods similar to Refs. 50 and 51, but for our system we have found the finite-size effects in energy to be negligible. Crucially, a combination of a stochastic velocity rescaling thermostat[52] and a weak local Langevin thermostat was used to ensure sufficient equilibration of the ghost molecule and robust temperature control. The time step was 1 fs, and each MD run lasted for ~1.5 ns. We used FEP from $\lambda = 0$ to $\lambda = 0.0001$, and linear λ scaling in the TI over a dense grid for the rest. The FEP at $\lambda = 0$ is necessary because the TI integrand is divergent at that point due to the singularity of the LJ and the Coulomb potential related to the ions. In Fig. 2(a), we show the comparison of the excess chemical potential $\mu_0^{\mathrm{sol}}(\lambda)$ computed using FEP and the combination of TI and FEP (TIFEP),[48] as a function of the switching parameter λ. At $\lambda < 0.02$, both methods yield similar results. At higher λ, however, the direct FEP has a much higher statistical error and a systematic bias. Our estimated solvation free energy is $\mu_0^{\mathrm{sol}} = (-177.30 \pm 0.05)$ kcal/mol.

The final step is to obtain the μ^{sol} of NaCl ion pairs at different concentrations. Excess chemical potentials for NaCl ion pairs ($\mu_{\mathrm{NaCl}}^{\mathrm{ex}}$) were computed using the S^0 method [Eq. (12)] in Ref. 30. Simulations of NaCl water solutions at different molar concentrations were performed using LAMMPS[53] at 298.15 K and 1 bar using the Nosé–Hoover thermostat and barostat. As detailed in Ref. 30, a fixed number of 32 640 water molecules together with between 128 and 10 880 NaCl ion pairs were used. The ideal chemical potential is given by the first two terms on the right-hand side of Eq. (12), and is shown by the dashed black curve in Fig. 2(b). By combining the ideal and excess parts of the chemical potential, we computed μ^{sol} of NaCl ion pairs at different salt molalities m (i.e., the chemical amount of the solute per unit mass of the solvent), shown as the blue curve in Fig. 2(b). In the same figure,
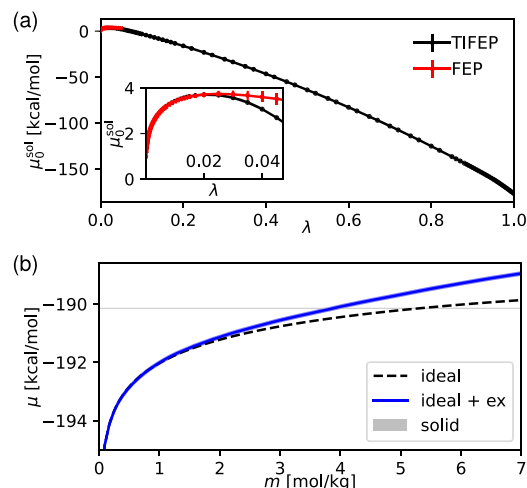
FIG. 2. (a) Chemical potential of solvation μ_0^{sol} of an NaCl ion pair computed using free-energy perturbation (FEP) and a combination of thermodynamic integration and free-energy perturbation (TIFEP). (b) The blue curve shows the chemical potentials μ^{sol} for NaCl ion pairs as a function of molality m at 298.15 K. The dashed black line is the ideal chemical potential. The grey horizontal line indicates the chemical potential of the ion pairs in the solid phase. The statistical uncertainties are indicated by the widths of the curves.

we also show the chemical potential of the ion pairs in the crystal, $\mu^{crystal}$ (grey horizontal line). The point at which the blue curve and grey line cross gives the solubility of NaCl in water; for this model, we estimate the solubility to be (3.87 ± 0.14) mol$_{NaCl}$ kg$_{water}^{-1}$. This estimate is lower than the experimental result of 6.15 mol kg^{-1} at 25 °C,[54] but is consistent with previous simulation results computed using osmotic ensemble Monte Carlo,[55] the Bennett acceptance ratio method[56] and thermodynamic integration,[18] all employing the same force field.

B. Urea in water

The properties of urea–water mixtures have been investigated using a range of empirical potentials.[57–62] For our investigation, we used the CHARMM36[63] parameterisation from CHARMM-GUI,[64,65] since unlike several other commonly used empirical potentials, we verified that the experimentally known crystalline polymorphs are at least mechanically stable at suitable thermodynamic conditions. The potential entails combinations of Lennard-Jones and Coulomb interactions, coupled with bond, angle, dihedral and improper constraints that maintain a relatively rigid molecular structure.[66] We do not use any further rigid-body constraints. We remark that since harmonic bonds are used, simulations should not be run with the Nosé–Hoover thermostat,[67,68] and alternatives such as stochastic velocity rescaling[52] or the Langevin thermostat should be used instead. For the solution phase, water molecules are modelled with a CHARMM-modified[66,69] TIP3P[70] potential so as to be compatible with the urea CHARMM force field; the water model entails a three-site potential with Coulomb and Lennard-Jones

interactions at each water atom, and harmonic bond and angle constraints.

For the bulk of our work, we considered two crystalline polymorphs of urea, namely forms I[71] and B1[72] (Fig. 3). For both polymorphs, we performed a λ-TI first at a low temperature (25 K) using a Debye crystal reference, followed by TI in T along the 1 bar isobar. The free energy computed for the crystals is thus referenced to isolated atoms. For the gas phase, the workflow for computing the free-energy reference to the isolated atoms is similar to that of crystals, except that both translations and rotations are constrained in the TI simulations. The analytic translational and rotational free-energy corrections [Eqs. (4) and (5)] are added at the end. The volume of the gas used in Eq. (4) was $1000 a_0^3$, where a_0 is the Bohr radius. The chemical potentials of the crystalline polymorphs (μ^I and μ^{B1}) are calculated by subtracting the free energy of the gas-phase molecule from the free energies of the crystal polymorphs. These chemical potentials are shown in Fig. 3(a). Consistently with previous work,[72,73] form I is the most stable form at ambient pressure. By way of comparison, the harmonic-approximation chemical potentials μ_h^I and μ_h^{B1} are also plotted in Fig. 3(a); the difference to the true chemical potential accounting for anharmonicity is rather large in both cases.

The solvation free energy is computed using multiple-step FEP with the soft-core potentials, with the switching parameters $\alpha_{LJ} = 0.5$ and $\alpha_C = 10$ Å^2 in Eqs. (9) and (10). We used a total of 20 windows, and performed the FEP in the middle step as described in Eq. (11). We confirmed that using more windows did not change the estimate. By contrast, a forward or backward FEP would require many more windows to reach convergence. Each independent run lasts for ~5 ns. We used a time step of 1 fs since the resulting μ_0^{sol} is consistent with that computed using a smaller time step of 0.1 fs. The calculations were performed using i-PI.[41] A combination of a stochastic velocity rescaling thermostat[52] and a weak local Langevin thermostat was used for ergodic sampling of the solute molecule and efficiency. The pressure was kept at 1 bar. As discussed above, the computed solvation free energy is already referenced to the gas-molecule state.

As a sanity check, we computed the solvation free energy of urea in its own melt, which is just the chemical potential of the melt μ^{melt} [the red symbols in Fig. 3(a)]. μ^{melt} computed at different temperatures agree well with the values obtained from TI of the melted phase with respect to T [Eq. (3)], as shown using the red curve in Fig. 3(a). The crossover between μ^{melt} and μ^I or μ^{B1} is the melting point of the specific crystalline phase, and we estimate $T_m^I = (425 \pm 7)$ K and $T_m^{B1} = (405 \pm 5)$ K. We then performed independent direct-coexistence simulations in elongated boxes (3840 atoms for form I with typical box dimensions of $71 \times 22 \times 22$ Å^3 and 5760 atoms for form B1 with typical box dimensions of $90 \times 27 \times 21$ Å^3). We first determined the equilibrium lattice parameters as a function of temperature at 1 bar, then melted half the box along x at a high temperature whilst keeping the remaining half the molecules frozen, and finally evolving the system at fixed P_x until one phase grew at the expense of the other. The melting point for form I determined with this method is in the range 415–430 K, and for form B1 it is in the range 400–410 K. Since both ranges are consistent with estimates from free-energy calculations, this suggests that the computed chemical potentials had a consistent choice of base-
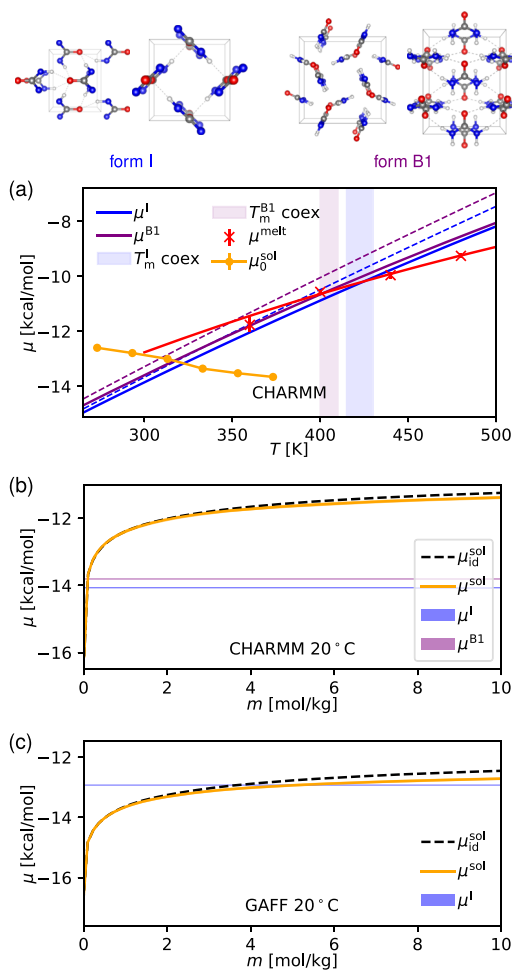
FIG. 3. (a) Chemical potentials of urea in different phases computed using the CHARMM force field. The blue and purple solid curves show the chemical potentials of crystal form I (μ^I) and crystal form B1 (μ^{B1}), respectively. The dashed curves show the harmonic free energies of the corresponding phases. The red crosses with error bars show the estimate of the chemical potential of melted urea (μ^{melt}), and the red curve shows the values from the thermodynamic integration of the liquid along the 1 bar isobar. The blue and the purple shaded areas show estimates of the melting temperatures of forms I and B1 with uncertainty, computed using the coexistence method. The orange symbols show the solvation free energy μ_0^{sol} at the concentration $c_0 = 0.49$ mol dm^{-3} with error estimates. (b) Chemical potentials of urea in solution of different molalities m (i.e., the amount of urea dissolved in a kilogram of water) computed using the CHARMM force field. (c) Chemical potentials of urea in solution of different molalities computed using the GAFF force field. In (b) and (c), the orange curve shows the chemical potentials μ^{sol} for urea as a function of the urea molality. The dashed black line is the ideal chemical potential. The blue and purple horizontal lines indicates the chemical potential of phase I and B1. The statistical uncertainties are indicated by the width of the curves.

TABLE I. Solubilities of urea in water expressed as molalities at different temperatures for the CHARMM potential.

T (°C)	T (K)	Solubility of I (mol kg^{-1})	Solubility of B1 (mol kg^{-1})
20	293.15	0.054 ± 0.002	0.083 ± 0.003
40	313.15	0.236 ± 0.009	0.36 ± 0.01
60	333.15	1.17 ± 0.04	1.82 ± 0.07
80	353.15	4.2 ± 0.1	6.9 ± 0.2
100	373.15	17.2 ± 0.7	33 ± 1

We computed the solvation free energy of urea in water (μ_0^{sol}) at 20 °C intervals between 0 and 100 °C. Once a "ghost" urea molecule was added, the simulation box contained one urea molecule and 123 water molecules, corresponding to a concentration of about 0.49 mol dm^{-3}. These results are indicated using orange symbols with error bars in Fig. 3(a). To obtain chemical potentials of the solute at other concentrations, we used the S^0 method. We simulated urea–water mixtures at different molar concentrations using LAMMPS[53] at the temperatures specified above and a pressure of 1 bar. The simulation box contained about 10 000 molecules. We used a time step of 1 fs, with runs of ~1 ns each. To compute $S(\boldsymbol{k})$, we collected a snapshot per 1000 steps of the trajectory. We used the positions of oxygen atoms as the positions of water molecules, and the positions of carbon atoms as the positions of urea molecules. The $S(\boldsymbol{k})$ values were then fitted to the Ornstein–Zernike form[30] with a maximum cutoff in the wave vector $k_{cut}^2 = 0.004/\text{Å}^2$. The S^0 values between urea–urea and urea–water can then be used to compute the concentration-dependent $\mu^{sol}(c)$ using Eq. (12). As an example, μ^{sol} at 20 °C at different molalities is shown as the orange curve in Fig. 3(b). The shaded area indicates the statistical uncertainty, which principally comes from the error in the estimation of μ_0^{sol}. The ideal chemical potential is shown as the dashed black curve. Finally, the chemical potentials of the crystalline forms I and B1 are plotted as blue and purple horizontal lines. As the concentration increases, the solution becomes less ideal. The crossover points between the orange curve and the two horizontal lines indicate the solubilities of the corresponding phases. In this case, the ideal-solution assumption turns out to be rather accurate in the solubility prediction. However, the agreement becomes worse at other temperatures considered; e.g., at 80 °C, it would lead to an underestimate in the solubility by about 15%. We summarise the solubilities at other temperatures in Table I. The solubilities increase steeply with temperature, and the solubility of form B1 is higher than that of form I at all temperatures considered.

It has been shown that using different urea models can result in considerably different thermodynamic properties in solution.[58] To check the role of the choice of force field in determining the solubility, we used the AMBER-GAFF-ESP-2018 force field (GAFF)[75,76] for both urea and water. We performed the solubility calculations for urea form I in water with GAFF, following the same workflow as with CHARMM.[77] In Fig. 3(c), we show μ^{sol} at 20 °C computed using GAFF as a function of urea molality. At this temperature, the solubility of form I of urea is (5.1 ± 0.3) mol kg^{-1}, about 100 times

line, and the correctness of the workflow. The experimental melting point of form I urea is 406 K;[74] the close agreement with simulation is perhaps somewhat surprising given the simplicity of the force field.

the solubility predicted for the CHARMM potential under the same conditions.

Experimentally, urea dissolves very readily in water; the solubility in 100 g of water ranges between ~70 g (12 mol kg^{-1}) at 0 °C to ~700 g (120 mol kg^{-1}) at 100 °C.[74,78,79] These solubilities are considerably higher than the CHARMM predictions obtained in Table I. The GAFF solubility at 20 °C is closer to, but still lower than, the experimental value. On the other hand, the crystalline phase is less well described by the GAFF potential than by CHARMM; at temperatures above 20 °C, form I is not dynamically stable under the GAFF force field, and readily transforms into an analogue of form III in MD simulations. These results thus demonstrate the sensitivity of the solubility prediction on the assumed potential energy surface, and the challenge of accurately modelling both the crystalline and the solution phases with the same potential.

C. Paracetamol in water and in ethanol

We used the CHARMM36[63] parameterisation from CHARMM-GUI[64,65] for paracetamol, water and ethanol, since all common crystalline polymorphs of paracetamol are at least metastable with this potential.

Specifically, we considered two crystalline polymorphs of paracetamol, namely forms I[80] and II.[81] For both polymorphs, we computed the free energy of the Debye crystal at 25 K and then used λ-TI to the potential of interest, followed by a TI in T along the 1 bar isobar. For both polymorphs, it is important to perform the λ-TI from the Debye crystal at $T \lesssim 40$ K, as the internal rotational modes become activated at higher T, causing numerical problems with the integrand of Eq. (2) as $\lambda \to 1$. We used a time step of 0.25 fs and, as with the case of urea above, a combination of a stochastic velocity rescaling thermostat[52] and a weak local Langevin thermostat.

For the gas phase, we computed the free-energy difference between the Debye molecule reference and the molecule with a constrained CM and rotations at $T = 100$ K. Because rotations about bonds are possible and lead to different conformers, vanilla λ-TI is thwarted by divergences in the integrand. Instead, both the tethering approach and the FEP at the end point, as outlined in Sec. II, help to circumvent the issue. We used a time step of 0.25 fs, and the same combined thermostat as for the solid phases. The analytic translational and rotational free-energy corrections are added in the end. The volume of the gas used in Eq. (4) was $1587a_0^3$. The total free energy of the gas molecule at 100 K was estimated to be (-41.84 ± 0.04) kcal/mol. The free energies at other temperatures were then determined using TI along the 1 bar isobar.

The chemical potentials of the crystalline polymorphs (μ^{I} and μ^{II}) are baseline adjusted, so the reference state is the gas-phase molecule. These chemical potentials are shown in Fig. 4(a), together with the harmonic-approximated $\mu_{\text{h}}^{\text{I}}$ and $\mu_{\text{h}}^{\text{II}}$ in dashed lines; the difference is of the order of 1 kcal/mol.

For computing solvation free energies, we used the multiple-step FEP with the soft-core potentials described in Sec. II C 1, with $\alpha_{\text{LJ}} = 0.5$ and $\alpha_{\text{C}} = 10$ Å^2. We used a total of 20 windows and performed the FEP in the middle step as described above. We confirmed that using more windows did not change the estimate. For
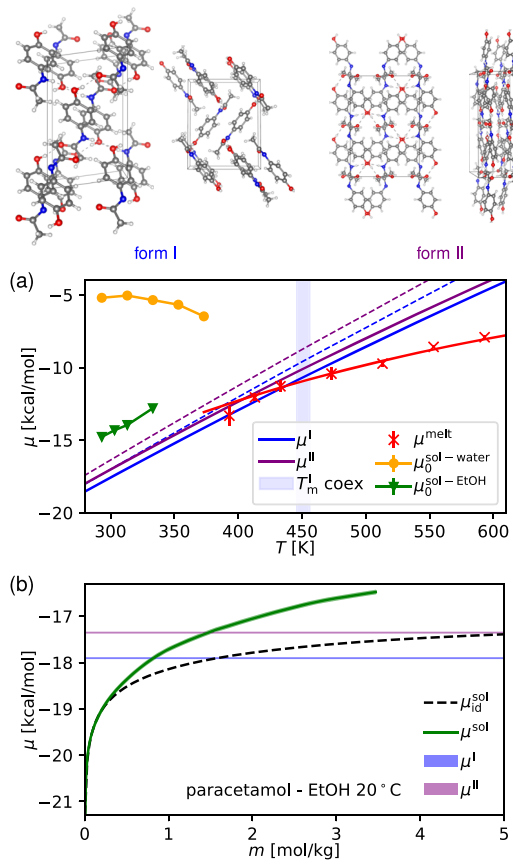

FIG. 4. (a) Chemical potentials of paracetamol in different phases computed using the CHARMM force field. The blue and purple solid curves show the chemical potentials of paracetamol in crystal forms I (μ^{I}) and II (μ^{II}), respectively. The dashed curves show the harmonic free energies of the corresponding phases. The red crosses with error bars show the estimate of the chemical potential of the paracetamol melt (μ^{melt}), and the red curve shows the values from the TI of the liquid along the 1 bar isobar. The blue shaded area shows the estimate of the melting temperature of phase I with statistical uncertainty computed using the interface-pinning method for solid–liquid coexistence. The orange symbols show the solvation chemical potentials $\mu_0^{\text{sol–water}}$ of paracetamol in water at $c_0 = 0.47$ mol dm^{-3} with error estimates, and the green symbols show the solvation chemical potentials $\mu_0^{\text{sol–EtOH}}$ in ethanol at $c_0 = 0.23$ mol dm^{-3}. (b) The green curve shows the chemical potentials μ^{sol} for paracetamol dissolved in ethanol (EtOH) as a function of molality m (i.e., amount of paracetamol per kilogram of ethanol). The dashed black line is the ideal chemical potential. The blue and purple horizontal lines indicate the chemical potentials of forms I and II, respectively. Statistical uncertainties are indicated by the widths of the curves.

the solvation free energy of paracetamol in water ($\mu_0^{\text{sol–water}}$), the simulation box contained one paracetamol molecule and 118 water molecules ($c_0 = 0.47$ mol dm^{-3}). The time step was 1 fs, and each independent run lasted about 1.5 ns. For the solvation free energy of paracetamol in ethanol ($\mu_0^{\text{sol–EtOH}}$), we used a simulation box comprising one paracetamol molecule and 73 ethanol molecules ($c_0 = 0.23$ mol dm^{-3}), and each independent FEP run lasted about 5 ns. These solvation free energies ($\mu_0^{\text{sol–water}}$ and $\mu_0^{\text{sol–EtOH}}$) are

shown in Fig. 4(a). The large gap between the two sets of solvation energies indicates that paracetamol solvation in ethanol is rather more favourable than in water when using the CHARMM force field.

In addition, we computed the solvation free energy of paracetamol in its own melt, μ^{melt}, shown using the the red symbols in Fig. 4(a). These values agree well with the TI results of the melt [Eq. (3)] [red curve in Fig. 4(a)]. We estimate the melting points of form I and II to be $T_m^I = (440 \pm 11)$ K and $T_m^{II} = (407 \pm 11)$ K. As a sanity check on our free-energy calculations, we also computed the melting point of form I of paracetamol using interface pinning.[82–84] We took an elongated box of crystal form I (roughly $107 \times 29 \times 35$ Å³ with 504 molecules, with $\alpha = \gamma = 90°$ and $\beta \approx 98°$), equilibrated it at 1 bar and a range of temperatures to determine optimal box parameters, melted half the box at a high temperature and locally equilibrated the system. We then performed interface-pinning simulations in which only the x component of the pressure was coupled to a barostat. We used a Steinhardt–Ten Wolde-style order parameter[85,86] on the carbonyl oxygen to classify molecules as being either in the crystal or the liquid phase, and then, using PLUMED,[87] applied a harmonic restraint with spring constant κ to bias the number of crystalline molecules to be close to half the total molecules in the system, $N_0 \approx N/2$. We computed the chemical-potential difference between form I and the melt via $\Delta\mu = \kappa(N_0 - \langle N_{\text{cryst}} \rangle)$,[38,82] by determining the average number of crystalline molecules as a function of temperature. Using this approach, we estimate that the melting point, at which $\Delta\mu = 0$, is (451 ± 5) K. These results are consistent with Fig. 4(a), giving us confidence that all the relevant factors have been considered in the free-energy calculations. The experimental melting point of form I paracetamol is 442 K[88] (169 °C), in close agreement with simulation despite the assumptions of the force field.

For the concentration dependence of μ^{sol}, we used the S^0 method. We simulated paracetamol–water and paracetamol–ethanol mixtures at different molar concentrations using LAMMPS,[53] although for paracetamol–water this proved not to be necessary, since at low concentrations the solution is nearly ideal. The simulation box contained about 10 000 molecules. The time step was 1 fs and each run lasted about 1 ns. Snapshots where taken every 1000 steps, and the co-ordinates of oxygen atoms in water, nitrogen atoms in paracetamol, and the carbon bonded to oxygen in ethanol were used for recording the positions of the corresponding molecules. In the estimation of S^0, the cutoff in the wave vector was $k_{\text{cut}}^2 = 0.0025/\text{Å}^2$.

TABLE II. Solubility of paracetamol in water expressed as molalities at different temperatures.

T (°C)	T (K)	Solubility of I (mol kg⁻¹)	Solubility of II (mol kg⁻¹)
20	293.15	$(3.5 \pm 0.5) \times 10^{-8}$	$(9.1 \pm 1.0) \times 10^{-8}$
40	313.15	$(5.1 \pm 0.8) \times 10^{-7}$	$(1.3 \pm 0.2) \times 10^{-6}$
60	333.15	$(1.1 \pm 0.1) \times 10^{-5}$	$(2.5 \pm 0.3) \times 10^{-5}$
80	353.15	$(1.6 \pm 0.2) \times 10^{-4}$	$(3.6 \pm 0.4) \times 10^{-4}$
100	373.15	$(3.3 \pm 0.4) \times 10^{-3}$	$(7.2 \pm 0.8) \times 10^{-3}$

TABLE III. Solubility of paracetamol in ethanol expressed as molalities at different temperatures.

T (°C)	T (K)	Solubility of I (mol kg⁻¹)	Solubility of II (mol kg⁻¹)
20	293.15	0.82 ± 0.03	1.47 ± 0.05
30	303.15	0.95 ± 0.04	1.79 ± 0.07
40	313.15	1.34 ± 0.05	2.7 ± 0.1

We show in Fig. 4(b) the solution chemical potential μ^{sol} for paracetamol in ethanol at 20 °C as a function of the paracetamol molality (green curve). The chemical potentials of the two crystalline polymorphs are plotted as blue and purple horizontal lines. At higher concentrations, the solution becomes less ideal, and indeed the ideal-solution assumption would lead to an overestimate in the solubilities by a factor of two to three. The solubilities of paracetamol in water or ethanol at all temperatures considered are provided in Tables II and III. The solubilities increase with temperature in both water and ethanol, and the solubility of form II is about double that of form I under all conditions considered.

The CHARMM22 potential with SwissParam[89] parameters has been used to identify the source of stabilisation of different polymorphs of paracetamol relative to density functional theory-level calculations;[90] this work illustrates that empirical force fields likely do not capture all the relevant physics. However, the CHARMM36 force field with optimised partial charges has been shown to be a satisfactory potential for studying the nucleation of paracetamol in acetonitrile,[91,92] and this reparameterised potential has also been used to investigate the solubility of paracetamol in ethanol[11] using the cavity method. In this work, previous experimental values[93–96] were averaged to obtain a solubility mole fraction of 0.0585 ± 0.0040 at 20 °C, or a molality of 1.35 mol kg⁻¹. The solubility of paracetamol with the CHARMM36 potential with reoptimised partial charges was (0.085 ± 0.014)[11] (or 2.0 mol kg⁻¹), a slight overestimate compared to the experimental result. By contrast, using the CHARMM-GUI parameterisation of the charges results in a slight underestimate of the solubility ($m = 0.82$ mol kg⁻¹, or $x = 0.036$). Both parameterisations appear to be reasonable, but the solubility is very sensitive to the details of the model's parameters. However, the solubility of paracetamol in water at 20 °C is $m = 0.0845$ mol kg⁻¹[93] (or $x = 0.0015$), increasing to $\sim m = 0.48$ mol kg⁻¹ ($x \approx 0.0086$) at 70 °C,[97] which is very different from the values we have obtained, suggesting that the interactions of paracetamol with water are accounted for less well.

IV. CONCLUSIONS

In summary, we present a workflow that enables simple and robust computation of solubilities of molecular or ionic crystals. As illustrated in Fig. 1, we compute the free energies of the gas, the crystal and the solution phases separately. The workflow mainly uses thermodynamic integration from reference systems with known free energies to physical systems, with free-energy perturbations incorporated at the end points of the TI to avoid numerical issues and to increase efficiency. Compared to the state-of-the-art

28 November 2023 08:28:28

methods,[7,8,10,11,20,33] we streamline many steps. For example, we compute the chemical potential of the crystal using the gas-phase molecule as a baseline, which avoids the convoluted many-step switching in the extended Einstein crystal method.[7,8] Using Debye crystals or Debye molecules as the starting points of the TI ensures high statistical efficiency, as these reference potential-energy surfaces closely resemble the potentials of interest. Constraining the CM and the rotation of the gas-phase molecule further reduces the sampling space. For the solute, the FEP at half steps [Eq. (11)] requires fewer windows than forward or backward FEP. We use the S^0 method to compute the concentration dependence of the chemical potential of the solute, which overcomes the ideal-solution assumption and avoids separate solvation free-energy calculations at different concentrations that would otherwise be necessary.[11]

We have applied the workflow to systems of a range of solubilities: sodium chloride in water, urea in water, paracetamol in water, and paracetamol in ethanol. We additionally computed the solid–liquid equilibria for urea and paracetamol using the same workflow, for validation. From these calculations, we have observed that the harmonic approximation for both the gas molecules and the crystals often breaks down at higher temperatures. Even at room temperature, for molecular crystals, the error can be of the order of a kcal mol^{-1}, as in the case of urea and paracetamol. The ideal-solution assumption can also fail: it is reasonably accurate for dilute solutions, but can result in an error of a factor of two in the solubility predictions of more soluble substances, such as paracetamol in ethanol.

We have shown that quantitative solubility predictions hinge on the accuracy of the potentials chosen. For instance, the solubility of urea in water at 20 °C predicted using GAFF is two orders of magnitude higher than the CHARMM analogue. As mentioned above, it is challenging to design empirical force fields that provide a good representation of both solution and solid phases simultaneously.[7,18,98] In part, this is because the local environment in the solution is very different from that in the solid, and the mean polarisation of the molecules is rather different. This is challenging to account for with pairwise potentials, but it may be possible to design more flexible models, for example with explicit polarisation or three-body terms, that can better represent the two phases.[98] Solubility calculations could provide a useful tool for parameterising such empirical potentials. Beyond enhancing empirical force fields, an alternative approach entails the use of machine-learning potentials (MLPs), which capture the accuracy of quantum-mechanical calculations, but at a considerably reduced computational cost.[99] Given that MLPs have the capacity to model accurately the intricate interactions at the atomic and molecular levels that are crucial for describing molecular solutions, including accounting for small differences in charge density in different phases,[100] they stand out as a promising tool for refining solubility predictions. Nonetheless, due to the higher computational demands of MLPs compared to classical force fields, integrating them with an efficient workflow, such as ours, becomes essential to ensure the tractability of calculations.

In this paper, we have focused on classical chemical potentials, but of course one could also account for the influence of nuclear quantum effects (NQEs). For instance, to obtain the chemical potentials of the molecule in solution, one can still use the workflow presented in Fig. 1(c) while using the path-integral molecular dynamics (PIMD) formalism[101] to represent the whole system. NQEs on the chemical potentials of the gas or the crystalline phases can be taken into account by integrating the quantum centroid virial kinetic energy with respect to the fictitious mass from the infinite mass to the physical masses using PIMD simulations.[102–104]

We expect that our workflow for calculating solubilities will be useful across various technologically significant systems. For example, for electrolytes, the ability to compute solubilities easily may deepen our understanding of the solvation of conducting ions in batteries. Similarly, in drug design, a clearer insight into solubility properties of drugs would enable drug delivery and absorption to be optimised. Moreover, the method can shed light on the precipitation of crystals from solutions, the behaviour of co-solvents, as well as extraction and purification methods in the chemical industry.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Aleks Reinhardt: Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Pin Yu Chew**: Investigation (equal); Writing – review & editing (equal). **Bingqing Cheng**: Conceptualization (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Simulation input files necessary to reproduce the study and Python data analysis scripts are available in the SI repository at https://github.com/BingqingCheng/solubility, and on Zenodo at http://doi.org/10.5281/zenodo.8398093.

REFERENCES

[1] A. J. Aguiar, J. Krc, Jr., A. W. Kinkel, and J. C. Samyn, "Effect of polymorphism on the absorption of chloramphenicol from chloramphenicol palmitate," J. Pharm. Sci. **56**, 847–853 (1967).

[2] G. P. Van Der Laan and A. A. C. M. Beenackers, "Kinetics and selectivity of the Fischer–Tropsch synthesis: A literature review," Catal. Rev.: Sci. Eng. **41**, 255–318 (1999).

[3] M. L. Cheney, N. Shan, E. R. Healey, M. Hanna, L. Wojtas, M. J. Zaworotko, V. Sava, S. Song, and J. R. Sanchez-Ramos, "Effects of crystal form on solubility and pharmacokinetics: A crystal engineering case study of lamotrigine," Cryst. Growth Des. **10**, 394–405 (2010).

[4]S. Byrn, R. Pfeiffer, M. Ganey, C. Hoiberg, and G. Poochikian, "Pharmaceutical solids: A strategic approach to regulatory considerations," Pharm. Res. **12**, 945–954 (1995).

[5]R. E. Skyner, J. L. McDonagh, C. R. Groom, T. van Mourik, and J. B. O. Mitchell, "A review of methods for the calculation of solution free energies and the modelling of systems in solution," Phys. Chem. Chem. Phys. **17**, 6174–6191 (2015).

[6]Z. Ye and D. Ouyang, "Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms," J. Cheminf. **13**, 98 (2021).

[7]L. Li, T. Totton, and D. Frenkel, "Computational methodology for solubility prediction: Application to the sparingly soluble solutes," J. Chem. Phys. **146**, 214110 (2017).

[8]L. Li, T. Totton, and D. Frenkel, "Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials," J. Chem. Phys. **149**, 054102 (2018).

[9]V. Khanna, M. F. Doherty, and B. Peters, "Absolute chemical potentials for complex molecules in fluid phases: A centroid reference for predicting phase equilibria," J. Chem. Phys. **153**, 214504 (2020).

[10]V. Khanna, M. F. Doherty, and B. Peters, "Predicting solubility and driving forces for crystallization using the absolute chemical potential route," Mol. Phys. **121**, e2155595 (2023).

[11]M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout, "Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments," J. Chem. Phys. **150**, 094107 (2019).

[12]E. Sanz and C. Vega, "Solubility of KF and NaCl in water by molecular simulation," J. Chem. Phys. **126**, 014507 (2007).

[13]A. S. Paluch, S. Jayaraman, J. K. Shah, and E. J. Maginn, "A method for computing the solubility limit of solids: Application to sodium chloride in water and alcohols," J. Chem. Phys. **133**, 124504 (2010).

[14]F. Moučka, M. Lísal, J. Škvor, J. Jirsák, I. Nezbeda, and W. R. Smith, "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 **115**, 7849–7861 (2011).

[15]M. Lísal, W. R. Smith, and J. Kolafa, "Molecular simulations of aqueous electrolyte solubility: 1. The expanded-ensemble osmotic molecular dynamics method for the solution phase," J. Phys. Chem. B **109**, 12956–12965 (2005).

[16]A. S. Paluch, S. Parameswaran, S. Liu, A. Kolavennu, and D. L. Mobley, "Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation," J. Chem. Phys. **142**, 044508 (2015).

[17]J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, and A. Z. Panagiotopoulos, "On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures," J. Chem. Phys. **145**, 154111 (2016).

[18]A. L. Benavides, J. L. Aragones, and C. Vega, "Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route," J. Chem. Phys. **144**, 124504 (2016).

[19]S. Boothroyd, A. Kerridge, A. Broo, D. Buttar, and J. Anwar, "Solubility prediction from first principles: A density of states approach," Phys. Chem. Chem. Phys. **20**, 20981–20987 (2018).

[20]E. C. Dybeck, A. Thiel, M. J. Schnieders, F. C. I. Pickard, G. P. Wood, J. F. Krzyzaniak, and B. C. Hancock, "A comparison of methods for computing relative anhydrous–hydrate stability with molecular simulation," Cryst. Growth Des. **23**, 142–167 (2023).

[21]A. C. L. Opitz, "Molecular dynamics investigation of a free surface of liquid argon," Phys. Lett. A **47**, 439–440 (1974).

[22]A. J. C. Ladd and L. V. Woodcock, "Triple-point coexistence properties of the Lennard-Jones system," Chem. Phys. Lett. **51**, 155–159 (1977).

[23]I. S. Joung and T. E. Cheatham III, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," J. Phys. Chem. B **112**, 9020–9041 (2008).

[24]C. Ramírez García, G. A. Méndez-Maldonado, J. Guillermo Méndez-Bermúdez, and E. Núñez-Rojas, "Free energy calculations and solubility in water of organic molecules: A numerical relation through molecular dynamics," Mol. Simul. **49**, 393–402 (2023).

[25]M. Pudipeddi and A. T. Serajuddin, "Trends in solubility of polymorphs," J. Pharm. Sci. **94**, 929–939 (2005).

[26]D. Frenkel and A. J. C. Ladd, "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. **81**, 3188–3193 (1984).

[27]B. Cheng and M. Ceriotti, "Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids," Phys. Rev. B **97**, 054102 (2018).

[28]J. L. Aragones, C. Valeriani, and C. Vega, "Note: Free energy calculations for atomic solids through the Einstein crystal/molecule methodology using GROMACS and LAMMPS," J. Chem. Phys. **137**, 146101 (2012).

[29]J. G. Kirkwood and F. P. Buff, "The statistical mechanical theory of solutions. I," J. Chem. Phys. **19**, 774–777 (1951).

[30]B. Cheng, "Computing chemical potentials of solutions from structure factors," J. Chem. Phys. **157**, 121101 (2022).

[31]N. Dawass, P. Krüger, S. K. Schnell, J.-M. Simon, and T. Vlugt, "Kirkwood-Buff integrals from molecular simulation," Fluid Phase Equilib. **486**, 21–36 (2019).

[32]C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, "Determination of phase diagrams via computer simulation: Methodology and applications to water, electrolytes and proteins," J. Phys.: Condens. Matter **20**, 153101 (2008).

[33]G. Gobbo, G. Ciccotti, and B. L. Trout, "On computing the solubility of molecular systems subject to constraints using the extended Einstein crystal method," J. Chem. Phys. **150**, 201104 (2019).

[34]T. Simonson, "Free energy of particle insertion," Mol. Phys. **80**, 441–447 (1993).

[35]R. W. Zwanzig, "High-temperature equation of state by a perturbation method. I. Nonpolar gases," J. Chem. Phys. **22**, 1420–1426 (1954).

[36]G. T. Gao, X. C. Zeng, and H. Tanaka, "The melting temperature of proton-disordered hexagonal ice: A computer simulation of 4-site transferable intermolecular potential model of water," J. Chem. Phys. **112**, 8534–8538 (2000).

[37]S. Habershon and D. E. Manolopoulos, "Free energy calculations for a flexible water model," Phys. Chem. Chem. Phys. **13**, 19714–19727 (2011).

[38]P. Y. Chew and A. Reinhardt, "Phase diagrams—Why they matter and how to predict them," J. Chem. Phys. **158**, 030902 (2023).

[39]J. F. Cornwell, *Group Theory in Physics: Volume 1*, *Techniques of Physics Vol. 7* (Academic Press, London, 1986).

[40]M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," J. Res. Natl. Bur. Stand. **49**, 409 (1952).

[41]V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. Van Speybroeck, and M. Ceriotti, "i-PI 2.0: A universal force engine for advanced molecular simulations," Comput. Phys. Commun. **236**, 214–223 (2019).

[42]N. R. Davidson, *Statistical Mechanics* (Dover, Mineola, NY, 2003).

[43]J. G. Kirkwood, "Statistical mechanics of fluid mixtures," J. Chem. Phys. **3**, 300–313 (1935).

[44]D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, 2000).

[45]M. Mezei and D. L. Beveridge, "Free energy simulations," Ann. N. Y. Acad. Sci. **482**, 1–23 (1986).

[46]C. Chipot and A. Pohorille, "Calculating free energy differences using perturbation theory," in *Free Energy Calculations: Theory and Applications in Chemistry and Biology* (Springer, Berlin, Heidelberg, 2007), pp. 33–75.

[47]T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations," Chem. Phys. Lett. **222**, 529–539 (1994).

[48]R. Schmid and B. Cheng, "Computing chemical potentials of adsorbed or confined fluids," J. Chem. Phys. **158**, 161101 (2023).

[49]C. Y. Lee and H. L. Scott, "The surface tension of water: A Monte Carlo calculation using an umbrella sampling algorithm," J. Chem. Phys. **73**, 004591–4596 (2008).

[50]M. A. Kastenholz and P. H. Hünenberger, "Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids," J. Chem. Phys. **124**, 124106 (2006).

[51]M. M. Reif and P. H. Hünenberger, "Computation of methodology-independent single-ion solvation properties from molecular simulations. IV. Optimized Lennard-Jones interaction parameter sets for the alkali and halide ions in water," J. Chem. Phys. **134**, 144104 (2011).

[52]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 014101 (2007).

[53]S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comput. Phys. **117**, 1–19 (1995).

[54]S. P. Pinho and E. A. Macedo, "Solubility of NaCl, NaBr, and KCl in water, methanol, ethanol, and their mixed solvents," J. Chem. Eng. Data **50**, 29–32 (2005).

[55]F. Moučka, I. Nezbeda, and W. R. Smith, "Molecular simulation of aqueous electrolytes: Water chemical potential results and Gibbs-Duhem equation consistency tests," J. Chem. Phys. **139**, 124505 (2013).

[56]Z. Mester and A. Z. Panagiotopoulos, "Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations," J. Chem. Phys. **142**, 044507 (2015).

[57]H. Kokubo and B. M. Pettitt, "Preferential solvation in urea solutions at different concentrations: Properties from simulation studies," J. Phys. Chem. B **111**, 5233–5242 (2007).

[58]H. Kokubo, J. Rösgen, D. W. Bolen, and B. M. Pettitt, "Molecular basis of the apparent near ideality of urea solutions," Biophys. J. **93**, 3392–3407 (2007).

[59]R. D. Mountain and D. Thirumalai, "Importance of excluded volume on the solvation of urea in water," J. Phys. Chem. B **108**, 6826–6831 (2004).

[60]C. A. Bertran, J. J. V. Cirino, and L. C. G. Freitas, "Theoretical studies of concentrated aqueous urea solutions using computational Monte Carlo simulation," J. Braz. Chem. Soc. **13**, 238–244 (2002).

[61]L. J. Smith, H. J. C. Berendsen, and W. F. van Gunsteren, "Computer simulation of urea–water mixtures: A test of force field parameters for use in biomolecular simulation," J. Phys. Chem. B **108**, 1065–1071 (2004).

[62]R. Cortes-Huerto, K. Kremer, and R. Potestio, "Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations," J. Chem. Phys. **145**, 141103 (2016).

[63]J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. J. MacKerell, and R. W. Pastor, "Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types," J. Phys. Chem. B **114**, 7830–7843 (2010).

[64]S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," J. Comput. Chem. **29**, 1859–1865 (2008).

[65]J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks III, A. D. MacKerell, Jr., J. B. Klauda, and W. Im, "CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field," J. Chem. Theory Comput. **12**, 405–413 (2016).

[66]A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. J. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," J. Phys. Chem. B **102**, 3586–3616 (1998).

[67]W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," Phys. Rev. A **31**, 1695–1697 (1985).

[68]J. L. Aragones, E. G. Noya, C. Valeriani, and C. Vega, "Free energy calculations for molecular solids using GROMACS," J. Chem. Phys. **139**, 034104 (2013).

[69]S. Boonstra, P. R. Onck, and E. van der Giessen, "CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state," J. Phys. Chem. B **120**, 3692–3698 (2016).

[70]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).

[71]H. Guth, G. Heger, S. Klein, W. Treutmann, and C. Scheringer, "Strukturverfeinerung von Harnstoff mit Neutronenbeugungsdaten bei 60, 123 und 293 K und X-N- und X-X(1s²)-Synthesen bei etwa 100 K," Z. Kristallogr. - Cryst. Mater. **153**, 237–254 (1980).

[72]P. M. Piaggi and M. Parrinello, "Predicting polymorphism in molecular crystals using orientational entropy," Proc. Natl. Acad. Sci. U. S. A. **115**, 10251–10256 (2018).

[73]A. Olejniczak, K. Ostrowska, and A. Katrusiak, "H-bond breaking in high-pressure urea," J. Phys. Chem. C **113**, 15761–15767 (2009).

[74]F. W. Miller, Jr. and H. R. Dittmar, "The solubility of urea in water. The heat of fusion of urea," J. Am. Chem. Soc. **56**, 848–849 (1934).

[75]G. A. Özpınar, W. Peukert, and T. Clark, "An improved generalized AMBER force field (GAFF) for urea," J. Mol. Model. **16**, 1427–1440 (2010).

[76]D. van der Spoel, M. M. Ghahremanpour, and J. A. Lemkul, "Small molecule thermochemistry: A tool for empirical force field development," J. Phys. Chem. A **122**, 8982–8988 (2018).

[77]For the GAFF potential, form B1 readily converts to form B2 at the conditions of interest, and so we have not considered it further.

[78]L. A. Pinck and M. A. Kelly, "The solubility of urea in water," J. Am. Chem. Soc. **47**, 2170–2172 (1925).

[79]S. Halonen, T. Kangas, M. Haataja, and U. Lassi, "Urea-water-solution properties: Density, viscosity, and surface tension in an under-saturated solution," Emiss. Control Sci. Technol. **3**, 161–170 (2016).

[80]X. Ou, X. Li, H. Rong, L. Yu, and M. Lu, "A general method for cultivating single crystals from melt microdroplets," Chem. Commun. **56**, 9950–9953 (2020).

[81]L. H. Thomas, C. Wales, L. Zhao, and C. C. Wilson, "Paracetamol form II: An elusive polymorph through facile multicomponent crystallization routes," Cryst. Growth Des. **11**, 1450–1452 (2011).

[82]U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago, "Computing Gibbs free energy differences by interface pinning," Phys. Rev. B **88**, 094101 (2013).

[83]U. R. Pedersen, "Direct calculation of the solid–liquid Gibbs free energy difference in a single equilibrium simulation," J. Chem. Phys. **139**, 104102 (2013).

[84]U. R. Pedersen, F. Hummel, and C. Dellago, "Computing the crystal growth rate by the interface pinning method," J. Chem. Phys. **142**, 044104 (2015).

[85]P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," Phys. Rev. B **28**, 784–805 (1983).

[86]P.-R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, "Simulation of homogeneous crystal nucleation close to coexistence," Faraday Discuss. **104**, 93–110 (1996).

[87]G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," Comput. Phys. Commun. **185**, 604–613 (2014).

[88]J. Ledru, C. Imrie, C. Pulham, R. Céolin, and J. Hutchinson, "High pressure differential scanning calorimetry investigations on the pressure dependence of the melting of paracetamol polymorphs I and II," J. Pharm. Sci. **96**, 2784–2794 (2007).

[89]V. Zoete, M. A. Cuendet, A. Grosdidier, and O. Michielin, "SwissParam: A fast force field generation tool for small organic molecules," J. Comput. Chem. **32**, 2359–2368 (2011).

[90]M. Rossi, P. Gasparotto, and M. Ceriotti, "Anharmonic and quantum fluctuations in molecular crystals: A first-principles study of the stability of paracetamol," Phys. Rev. Lett. **117**, 115702 (2016).

[91]J. Stojaković, F. Baftizadeh, M. A. Bellucci, A. S. Myerson, and B. L. Trout, "Angle-directed nucleation of paracetamol on biocompatible nanoimprinted polymers," Cryst. Growth Des. **17**, 2955–2963 (2017).

[92]G. Gobbo, M. A. Bellucci, G. A. Tribello, G. Ciccotti, and B. L. Trout, "Nucleation of molecular crystals driven by relative information entropy," J. Chem. Theory Comput. **14**, 959–972 (2018).

[93]R. A. Granberg and Å. C. Rasmuson, "Solubility of paracetamol in pure solvents," J. Chem. Eng. Data **44**, 1391–1395 (1999).

[94]J. A. Jiménez and F. Martínez, "Temperature dependence of the solubility of acetaminophen in propylene glycol + ethanol mixtures," J. Solution Chem. **35**, 335–352 (2006).

[95]G. L. Perlovich, T. V. Volkova, and A. Bauer-Brandl, "Towards an understanding of the molecular mechanism of solvation of drug molecules: A thermodynamic

approach by crystal lattice energy, sublimation, and solubility exemplified by paracetamol, acetanilide, and phenacetin," J. Pharm. Sci. **95**, 2158–2169 (2006).

[96] N. A. Mitchell and P. J. Frawley, "Nucleation kinetics of paracetamol–ethanol solutions from metastable zone widths," J. Cryst. Growth **312**, 2740–2746 (2010).

[97] D. Grant, M. Mehdizadeh, A.-L. Chow, and J. Fairbrother, "Non-linear van't Hoff solubility-temperature plots and their pharmaceutical interpretation," Int. J. Pharm. **18**, 25–38 (1984).

[98] I. Nezbeda, F. Moučka, and W. R. Smith, "Recent progress in molecular simulation of aqueous electrolytes: Force fields, chemical potentials and solubility," Mol. Phys. **114**, 1665–1690 (2016).

[99] V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," Adv. Mater. **31**, 1902765 (2019).

[100] T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "Accurate fourth-generation machine learning potentials by electrostatic embedding," J. Chem. Theory Comput. **19**, 3567–3579 (2023).

[101] D. Marx and M. Parrinello, "*Ab initio* path integral molecular dynamics: Basic ideas," J. Chem. Phys. **104**, 4077–4082 (1996).

[102] M. Ceriotti and T. E. Markland, "Efficient methods and practical guidelines for simulating isotope effects," J. Chem. Phys. **138**, 014112 (2013).

[103] B. Cheng, J. Behler, and M. Ceriotti, "Nuclear quantum effects in water at the triple point: Using theory as a link between experiments," J. Phys. Chem. Lett. **7**, 2210–2215 (2016).

[104] B. Cheng, A. T. Paxton, and M. Ceriotti, "Hydrogen diffusion and trapping in α-iron: The role of quantum and anharmonic fluctuations," Phys. Rev. Lett. **120**, 225901 (2018).

28 November 2023 08:28:28