Computational methodology for solubility prediction: Application to the sparingly soluble solutes

Lunna Li, , Tim Totton, and , and Daan Frenkel

Articles you may be interested in

 Simulations of simple linoleic acid-containing lipid membranes and models for the soybean plasma membranes
The Journal of Chemical Physics **146**, 215103 (2017); 10.1063/1.4983655

The entropy of a complex molecule
The Journal of Chemical Physics **146**, 224106 (2017); 10.1063/1.4984965

 Communication: Density functional theory embedding with the orthogonality constrained basis set expansion procedure
The Journal of Chemical Physics **146**, 211101 (2017); 10.1063/1.4984777

 Osmotic and diffusio-osmotic flow generation at high solute concentration. II. Molecular dynamics simulations
The Journal of Chemical Physics **146**, 194702 (2017); 10.1063/1.4981794

Polymorphic transitions in single crystals: A new molecular dynamics method
Journal of Applied Physics **52**, 7182 (1998); 10.1063/1.328693

 Perspective: Theory and simulation of hybrid halide perovskites
The Journal of Chemical Physics **146**, 220901 (2017); 10.1063/1.4984964

Computational methodology for solubility prediction: Application to the sparingly soluble solutes

Lunna Li,[1,a)] Tim Totton,[2] and Daan Frenkel[1,b)]

[1]*Department of Chemistry, University of Cambridge, Cambridgeshire CB2 1EW, United Kingdom*
[2]*BP Exploration Operating Co. Ltd., Sunbury-on-Thames TW16 7LN, United Kingdom*

The solubility of a crystalline substance in the solution can be estimated from its absolute solid free energy and excess solvation free energy. Here, we present a numerical method, which enables convenient solubility estimation of general molecular crystals at arbitrary thermodynamic conditions where solid and solution can coexist. The methodology is based on standard alchemical free energy methods, such as thermodynamic integration and free energy perturbation, and consists of two parts: (1) systematic extension of the Einstein crystal method to calculate the absolute solid free energies of molecular crystals at arbitrary temperatures and pressures and (2) a flexible cavity method that can yield accurate estimates of the excess solvation free energies. As an illustration, via classical Molecular Dynamic simulations, we show that our approach can predict the solubility of OPLS-AA-based (Optimized Potentials for Liquid Simulations All Atomic) naphthalene in SPC (Simple Point Charge) water in good agreement with experimental data at various temperatures and pressures. Because the procedure is simple and general and only makes use of readily available open-source software, the methodology should provide a powerful tool for universal solubility prediction. © *2017 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).* [http://dx.doi.org/10.1063/1.4983754]

I. INTRODUCTION

Solubility is an important concept in many areas of science. For the development of pharmaceuticals,[1–3] it is crucial to know the solubility of drugs, as this influences bioavailability. In the oil industry, organic/inorganic compounds with low aqueous solubility can form unwanted deposits (scaling/fouling) to cause blockages.[4–7] These are but two examples to illustrate that it is extremely important to understand the precipitation of common materials under normal conditions and predict their solubility under conditions for which experimental information is difficult to obtain.

Numerical simulations provide a powerful tool for solubility prediction. For many users, it would be helpful to have a generic tool to predict solubility, which uses standard, open-source software. There is no lack of solubility calculations in the literature.[8–29] However, on the whole, simulation-based predictions of solubility are not widely used. We argue that this is due to the fact that generic, open-source tools are lacking. Yet, as we argue below, the prediction of solubility should be a standard test of new force fields. This becomes particularly clear when we consider the cases where existing force-fields have been used to compute solubility: very often, force-fields that seem to perform well for other properties fail the solubility-prediction test. Conceptually, the simplest way to compute solubilities from simulation is to

carry out "brute-force" direct coexistence simulations.[8,9,15–17] This method has the advantage of simplicity, but it generally requires long simulations (in some cases up to microseconds) to achieve solubility equilibrium, even for highly soluble compounds.

One "thermodynamic" approach determines solubility by computing the chemical potential in the solid and liquid phases.[14–20] The alternative "thermodynamic" approach determines the saturation solute composition by imposing the chemical potential of the solid phase to the solution in a grand canonical style. Such a route has been followed in the osmotic ensemble method/OEMC/OEMD.[10,12,13,21] The osmotic ensemble method has typically been applied to relatively small (ionic) solutes and focuses on the simulation of the liquid phase, assuming the chemical potential of the crystal is known from other sources. Also, it exhibits an advantage for multi-electrolyte solutions. However, we find that, in particular for large solutes, there is a need for a robust generic approach that integrates the calculation of the properties of both the crystal and of the solution, and is easily applicable in standard commercially available codes.

In the present work, we describe a generic protocol to predict solubility via computing the absolute chemical potentials of the solid and solution phases in solubility equilibrium. The method can be readily incorporated in standard codes for all-atom molecular dynamics (MD) simulations such as LAMMPS,[30] using classical all-atomic force fields available in the literature. We note that extension to other software packages should be straightforward. All the numerical techniques used in this work are standard alchemical free energy methods[31–35] such as free energy perturbation (FEP) and

---

a)This research was performed while Lunna Li was at Department of Chemistry, University of Cambridge, Cambridgeshire CB2 1EW, United Kingdom.
b)Author to whom correspondence should be addressed: df246@cam.ac.uk

thermodynamic integration (TI) that are commonly used in molecular dynamics (MD) or Monte Carlo (MC) simulations for computing solvation free energies[36,37] as well as absolute solid free energies.[38–44] The modeling challenge is focused on developing atomistic modeling tools that will make it possible to predict the solubility of simple organic crystals (we use naphthalene as an example) for potential generalization to a broader range of compounds.

II. METHODOLOGY

A. Theoretical background

As the basic approach to compute solubilities is well known, we only sketch the bare essentials here and refer the reader to Appendices A–C for more details. We start from the standard thermodynamic expressions.[45] G, μ, A, and V are the Gibbs free energy, chemical potential, Helmholtz free energy, and volume and number of solute or solvent molecules. p and T are the pressure and temperature. At constant T and p, the condition for equilibrium of a solute between the crystal phase and the solution is the equality of chemical potential in the two phases,

$$\mu_{\text{solution}}^{\text{solute}}(p, T) = \mu_{\text{solid}}^{\text{solute}}(p, T). \tag{1}$$

Hence, we need tools to compute the chemical potential in the crystal phase and in solution.

The chemical potential of the solute in the solution is

$$\mu_{\text{solution}} = k_B T \ln\left(\rho_{\text{solute}}^0 \Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1}\right) + k_B T \ln\left(\rho_{\text{solute}}/\rho_{\text{solute}}^0\right)$$
$$- k_B T \ln\left\langle \exp\left[-\beta U_{\text{solute - solution}}^{\rho_{\text{solute}}}(\mathbf{R}_{\text{solute}})\right]\right\rangle_0. \tag{2}$$

In Eq. (2), the only part that requires computation is the last term on the right hand side of the equation (the excess chemical potential $\Delta\mu_{\text{excess}}$). ρ_{solute} denotes the number density of the solute, with the common ideal gas reference state $\rho_{\text{solute}}^0 = 1$ (1 molecule in 1 Å^3 = 0.001 nm^3). Λ_{solute} is the thermal de Broglie wavelength of the solute, and q_{solute} is the intramolecular (vibrational and electronic) partition function of the solute; $\mathbf{R}_{\text{solute}}$ is the fixed position of the extra single solute in the solution, and $U_{\text{solute - solution}}(\mathbf{R}_{\text{solute}})$ is the interaction energy of the extra solute with the rest of solution. As explained in Appendix A, we can ignore the constant term $\ln\left(\rho_{\text{solute}}^0 \Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1}\right)$ for both coexisting phases.

The chemical potential of the solute in the solid is

$$\mu_{\text{solid}}^{\text{solute}} = G_{\text{solid}}/N_{\text{solid}} = (A_{\text{solid}} + pV_{\text{solid}})/N_{\text{solid}}. \tag{3}$$

For the solid phase, we derive the chemical potential from a simulation in which we compute the absolute free energy of the crystal. We compute the free energy of the crystal using an (straightforward) extension of the Einstein crystal method (see Sec. II B).

We describe details of the calculation of the chemical potentials A_{solid} and $\Delta\mu_{\text{excess}}$ in Secs. II B and II C and in Appendices A–C. Equation (1) becomes

$$k_B T \ln(\rho_{\text{solute}}) - k_B T \ln \left\langle \exp\left[-\beta U_{\text{solute - solution}}(\mathbf{R}_{\text{solute}})\right]\right\rangle_0$$
$$= (A_{\text{solid}} + pV_{\text{solid}})/N_{\text{solid}}. \tag{4}$$

In the case of a sparingly soluble solute (the example studied in this paper), we may set the second term on the left side to

the hydration free energy of a single solute molecule in the solvent, and then solve analytically for ρ_{solute}. In the case of higher solubilities, the hydration free energy becomes a function of composition, and we would need to calculate this dependence and determine the composition that solves the above equation.

B. Extended Einstein crystal method

Figure 1 is a pictorial description of the extended Einstein crystal method to compute the absolute solid free energy A_{solid} of a molecular crystal. The method is based upon the original Einstein crystal method[34,38,39] and its recent adaptations.[40–44] In the present work, it has been adapted to be used in MD simulations in LAMMPS without extra codes.

As the extended Einstein crystal method is similar in spirit to the original Einstein method, we describe the practical details of the approach including the various steps and contributions as illustrated in Fig. 1 as well as finite-size corrections[46,47] in Appendix B. Here we just mention the salient features: in the original Einstein crystal method for atomic systems, the free energy is computed by TI from the fully interaction crystal to an atomic Einstein crystal for which the free energy can be computed analytically. In the extended Einstein crystal method, we transform the molecular crystal into an Einstein crystal of independent molecules that are tethered by harmonic springs to reference positions. The number of springs needed should be sufficient to fix the position and orientation of molecule, but otherwise there is considerable freedom of choice. A second aspect is that we perform TI in stages, as illustrated in Fig. 1. Finally, we should consider $\Delta A_{\text{symmetry}}$, the analytic contribution to A_{solid} due to the symmetry of the constituting molecule in the crystal. For this term, we need to consider the molecular point group. Such an explicit consideration of $\Delta A_{\text{symmetry}}$ would not be necessary if a molecule can occupy all the symmetrically equivalent orientations during a crystal simulation. However, in practice in MD, a molecule in a non-rotator phase only samples one of the symmetrically equivalent orientations, because of kinetic reasons. In contrast, a solute in the solution can explore all orientations. Therefore $\Delta A_{\text{symmetry}}$ has to be explicitly added for a consistent consideration of the same reference state for the two phases. In the present work, for the compounds that we study, we have $\beta\Delta A_{\text{symmetry}}^{\text{naphthalene}}/N_{\text{solid}} = -\ln 4$ per molecule for naphthalene, because the point group is D_{2h}.

C. Cavity method for excess solvation free energy

Figure 2 gives a schematic description of the cavity method to compute the excess solvation free energies $\Delta\mu_{\text{excess}}$ of general solutes in isothermal-isobaric NpT ensembles.

In Fig. 2, part label (1) represents a pure solvent with the total potential energy $U_{(1)} = U_{\text{solvent-solvent}}$, with $U_{\text{solvent-solvent}} = \sum_i^{N_{\text{solvent}}} u_{\text{solvent-solvent}}$ and N_{solvent} as the total number of solvent molecules. Part label (4) represents the final solution with a fully solvated solute, and the total potential energy is $U_{(4)} = U_{\text{solvent-solvent}} + U_{\text{solute-solvent}}$ with $U_{\text{solute-solvent}} = \sum_i^{N_{\text{solvent}}} u_{\text{solute-solvent}}$. The free energy difference between these two states is hence the excess chemical potential that we wish to compute,

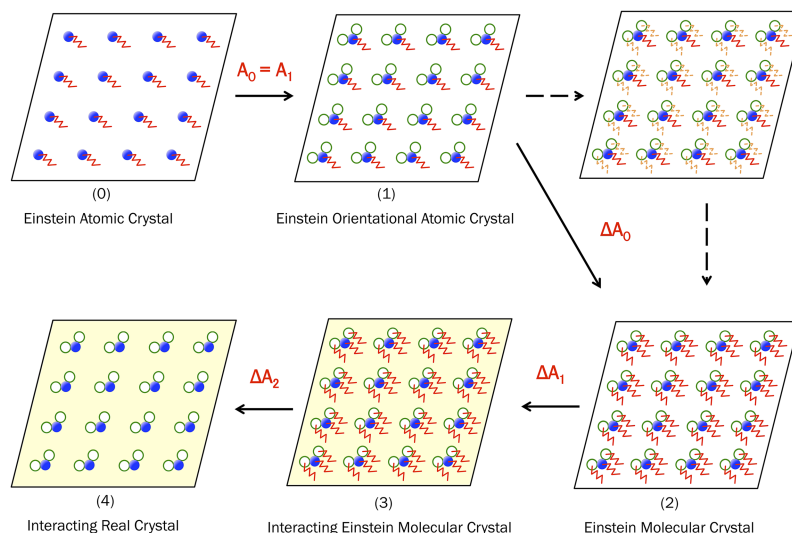$$\Delta\mu_{\text{excess}} = G_{(4)} - G_{(1)}. \tag{5}$$

FIG. 1. Pictorial representation of the extended Einstein crystal method to compute the absolute solid free energies A_{solid} of single-component molecular crystal. (0) The Einstein atomic crystal for which we know the analytic Helmholtz free energy A_0; (1) the Einstein orientational atomic crystal where each molecule is freely rotatable, with $A_0 = A_1$ because the solute is also considered to be freely rotatable in the solution phase (see Appendix B); (4) the real molecular crystal whose A_{solid} we wish to compute. The various ΔA represent different thermodynamically reversible paths and their accompanying free energy changes. Within each molecule, there is one "Central Atom" represented by the blue solid particle. All the "Central Atoms" are attached to their initial positions with the strongest Einstein springs K_{max} represented by the red solid springs throughout (1)-(3). There are also "Orientational Atoms" represented by the green empty particles. During (1) and (2), the "Orientational Atoms" are attached with varying K from $K_{min} = 0$ to K_{max}, represented by the orange dashed springs, for calculating the orientational contribution ΔA_0. During (2) and (3), both the "Central Atoms" and "Orientational Atoms" are attached with the strongest Einstein springs K_{max}, now all represented by the red solid springs, to sample the intermolecular energy contribution ΔA_1. Finally, during (3) and (4), all the red solid springs on "Central Atoms" and "Orientational Atoms" are reduced from K_{max} to $K_{min} = 0$ to recover the real crystal. Such a central-orientational arrangement is called the "Einstein Arrangement" or "EA" in our work. The yellow background filling in (3) and (4) means that the intermolecular interaction is active, but it is absent in (0)-(2).

In principle, we can compute $\Delta\mu_{excess}$ by switching on the interaction between the solute and the solvent directly. However, growing a solute in solution directly suffers from inaccuracies (mainly due to the well-known, "end-point" singularity associated with Lennard-Jones (LJ) solute-solvent interactions[48,49]). The cavity method eliminates the "end-point" singularity by introducing a softly repulsive cavity in
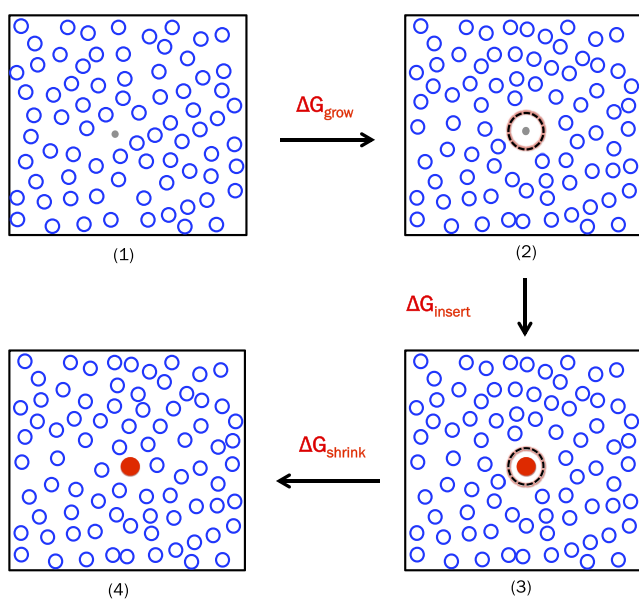

FIG. 2. Schematic representation of the cavity method to compute excess solvation free energies. The blue empty particles are solvent molecules, the red solid particle is the solute, and the black empty dashed particle around the solute is the cavity that we grow and shrink.

the solution, inside which the solute will be grown, and the cavity can be removed after the solute is fully grown. This procedure is described by (2) and (3) in Fig. 2. The cavity is attached to the center of the cubic box.

In (1) and (2) in Fig. 2, we slowly switch on the softly repulsive cavity to a predetermined size, and ΔG_{grow} is the Gibbs free energy change of creating the cavity. The potential energy expressions for states (1) and (2) are $U_{(1)} = U_{solvent-solvent}$ and $U_{(2)}(\lambda) = U_{solvent-solvent} + U_{cavity}(\lambda)$.

There is no fixed functional form for $U_{cavity}(\lambda)$ but $U_{cavity}(\lambda)$ should not suffer the same divergence problem as the LJ potential. For example, Postma *et al.*[50] have studied the thermodynamics of cavity formation in water, in the form of $U_{cavity}(\lambda) = \lambda(B/r)^{12}$, where λ is a linear scaling parameter, B a repulsive parameter, and r is the distance from the water oxygen to the cavity center. However, we cannot use this potential form for the cavity method because it is essentially the repulsive part of the Lennard-Jones potential causing the end-point singularity. In the present work, the cavity potential form we use is $U_{cavity}(\lambda) = \sum_i^{N_{solvent}} A \exp\left(-r_{solute-i}/B + \lambda\right)$, where $r_{solute-i}$ is the distance between the center of the cavity and a fixed point of the solvent molecule. The fixed point can be the center of mass of the solvent molecule, but it is more convenient to choose a specific atom, such as the O atom of water. The size of the cavity is indicated by the variable λ. In our simulations, we kept the parameters A and B fixed at $A = 1255$ kJ/mol and $B = 0.1$ nm for the typical case of a single solute molecule in $N_{water} = 864$ water molecules, and the largest range of λ we vary is between -10 and 5. At $\lambda = -10$, the interaction between the cavity potential and solvent is negligible, so it is practically equivalent to having no cavity;

at $\lambda = 5$, the volume of the cavity is roughly comparable to that of the original cubic box of $N_{water} = 864$.

In summary, the excess solvation free energy is calculated via

$$\Delta\mu_{excess} = \Delta G_{grow} + \Delta G_{insert} + \Delta G_{shrink}. \tag{6}$$

The computed excess solvation free energies should also include any tail corrections of long-range dispersion energy, which we obtain from our NpT simulations.

III. SIMULATION DETAILS

All simulations were run using the MD software package LAMMPS.[30] The van der Waals parameters and intramolecular interactions were based upon the classical all-atomic force OPLS-AA (optimized potentials for liquid simulations all atomic)[51] and SPC (simple point charge) water.[52] We used a time step dt between 0.5 and 1.0 fs. The Lorentz-Berthelot combining rules[53–55] were used for interatomic mixings. We have chosen the cutoff distances $r_{cut} = 1.2$ nm for all solution simulations and $r_{cut} = 1.0$ nm for all solid simulations. We used the default simulation settings in LAMMPS. In LAMMPS, the equations of motion are based on the work of Shinoda *et al.*,[56] Martyna *et al.*,[57] and Parrinello and Rahman.[58] For time integration schemes, we have consistently chosen the velocity-Verlet integrator.[59] For simulations in isobaric-isothermal NpT ensembles, a Nosé-Hoover[60,61] thermostat and a barostat were used to equilibrate all solution and solid systems. For simulations in canonical NVT ensembles, both Nosé-Hoover and Langevin thermostats[62] were used. We have used $N_{water} = 512$ for cyclohexane solution simulations. We have used $N_{water} = 894$ for naphthalene solution simulations between $p = 0.1$-200 MPa for $T = 298$ K and $N_{water} = 2048$ for naphthalene solution at $T = 320$ K and $T = 340$ K for $p = 0.1$ MPa, as well as $N_{naphthalene} = 240$ for naphthalene crystal simulations. We used real atomic masses. LAMMPS' default PPPM (Particle–Particle-Particle–Mesh) summation[63,64] with tin-foil boundary conditions[65] was used for all long-range electrostatics. Gauss-Legendre quadrature method[66] and the trapezoid rule[67] were used for TI.

Because the Einstein crystal method requires strong harmonic springs, it is important to choose the time step dt carefully for MD simulations of ΔA_2, ΔA_0, and ΔA_1. We have consistently chosen dt that is below $1/20$ of the period $T_p = 2\pi\sqrt{m/(2K)}$ of a harmonic spring $u = K(\mathbf{r}_i - \mathbf{r}_{0i})^2$. For example, for the biggest $K_{max} = 2\,092\,000$ (kJ/mol)/nm^2 and carbon atomic mass of 12 g/mol, $T_p = 2\pi\sqrt{m/(2K)} = 10.64$ fs so $T_p/20 = 0.53$ fs, and we used $dt = 0.5$ fs for all Einstein crystal simulations. We stress that some individual aspects of the simulations could have been carried out in a more efficient way, but this would have made it difficult to use standard software.

All the site-site interactions in this work are described by the 12-6 Lennard-Jones (LJ) potential,[68] $u_{LJ}(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^6\right]$, with ε as depth of the potential and σ is the finite intermolecular distance where $u_{LJ}(r) = 0$.

In both the extended Einstein crystal method and the cavity method, statistical errors were calculated from block averaging,[34,69] and errors of calculated values are obtained from error propagations of the individual block averaging errors.

IV. RESULTS AND DISCUSSIONS

A. Cyclohexane: Excess solvation free energy

We first illustrate the implementation of our cavity method by computing the excess solvation free energies of an OPLS-AA-based cyclohexane solute in 512 SPC water at $T = 298$ K and $p = 0.1$ MPa.

Figure 3 shows our computed excess solvation free energies of cyclohexane in water at different cavity sizes λ. We have an average of $\Delta\mu_{excess} = 7.47(35)$ kJ/mol, compared with the computed value of 6.86 kJ/mol in the OPLS-AA + SPC simulation in Ref. 37 and the experimental value[29,37,70] of 5.15 kJ/mol. Both simulations overestimate the excess solvation free energy of cyclohexane slightly but they agree within each other within errors.

Figure 3 shows a plot of the variation of our integrand contributions to ΔG_{insert}^{vdwl}, namely, $\left\langle \partial U_{solute-solvent}^{vdwl} \middle/ \partial\lambda_{vdwl} \right\rangle_{\lambda_{vdwl}}$ $= \left\langle U_{solute-solvent}^{vdwl} \middle/ \lambda_{vdwl} \right\rangle_{\lambda_{vdwl}}$ with the solute-solvent van der Waal's coupling parameter λ_{vdwl} for three different cavity sizes λ, during the cyclohexane insertion. It is evident that, as the cavity size we used for cyclohexane insertion increases (indicated by the larger λ), the weak singularity of $\left\langle \partial U_{solute-solvent}^{vdwl} \middle/ \partial\lambda_{vdwl} \right\rangle_{\lambda_{vdwl}}$ as $\lambda_{vdwl} \to 0$ becomes less pronounced. At $\lambda = 0$, using the trapezoid rule, we estimate $\Delta G_{insert}^{vdwl} = -18.20(4)$ kJ/mol, while a single-step FEP gives $\Delta G_{insert}^{vdwl} = -17.88(4)$ kJ/mol.

Figure 4 shows the computed excess solvation free energies for cyclohexane in water, computed (i.e., the cyclohexane solute inserted) at different cavity sizes. Figure 4 illustrates that all values for $\Delta\mu_{excess}$ agree within error, and the cavity at $\lambda = -0.5$ seems already large enough to suppress the weak singularity at the beginning of the switching on of the solute-solvent interaction (see Fig. 3).

B. Naphthalene: Excess solvation free energy

Figures 5(a)–5(c) show the computed excess solvation free energies of OPLS-AA-naphthalene in SPC water with increasing cavity size at $p = 0.1$ MPa, 100 MPa, and 200 MPa, respectively, at $T = 298$K. For all pressures, the values of $\Delta\mu_{excess}(\lambda)$ become independent of λ, to within statistical error for $\lambda \gtrsim 0$, indicating that the cavities beyond $\lambda = 0$
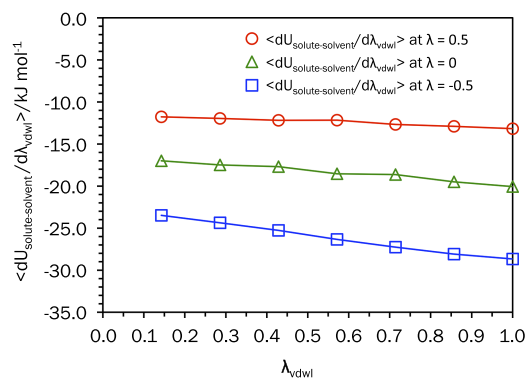

FIG. 3. $\left\langle \partial U_{solute-solvent}^{vdwl} \middle/ \partial\lambda_{vdwl} \right\rangle_{\lambda_{vdwl}}$ during the cyclohexane insertion.
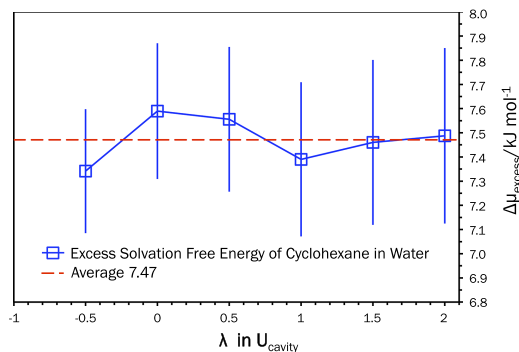
FIG. 4. Computed excess solvation free energy of cyclohexane in water at T = 298 K and p = 0.1 MPa. The red dashed line is the average value of 7.47 kJ/mol.

are sufficiently large to suppress the "end-point" singularity as $\lambda_{\mathrm{vdwl}} \to 0$.
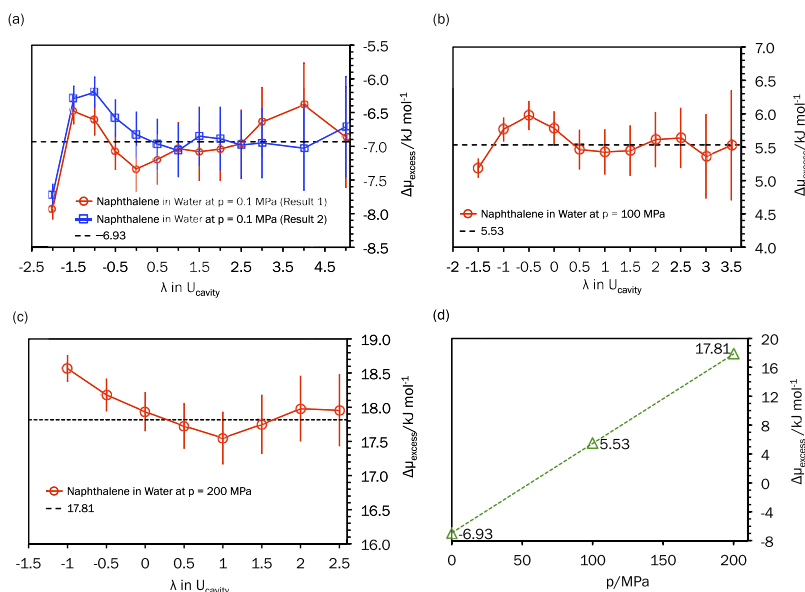
In Fig. 5(a), at p = 0.1 MPa, we obtain $\Delta\mu_{\mathrm{excess}}$ = −6.93 (42) kJ/mol, in good agreement with the value of −6.82 kJ/mol reported in the OPLS-AA + SPC simulation in Ref. 37. It has been mentioned that we have used slightly longer cutoff distance r_{cut} = 1.2 nm while the simulation reference has used r_{cut} = 1.0 nm, but $\Delta\mu_{\mathrm{excess}}$ should be nearly independent of r_{cut} for reasonable r_{cut} values if tail corrections are included.[37]

Both numerical results are in fair agreement with the experimental values[29,37,70–72] of −9.58 kJ/mol or −10.00 kJ/mol, with the difference most likely due to deficiencies of the force field.

Similarly, in Figs. 5(b) and 5(c), we plot the computed excess solvation free energies of naphthalene in water with increasing cavity size at p = 100 MPa and 200 MPa and T = 298 K.

Figure 5(d) shows the variation of the computed excess solvation free energies with pressure from p = 0.1 MPa to p = 200 MPa at T = 298 K. The pressure derivative of $\Delta\mu_{\mathrm{excess}}$ at fixed T at infinite dilution is the excess partial molar volume at infinite dilution,[73,74] $v_{\mathrm{excess}}^{\mathrm{solute}}$,

$$(\partial\Delta\mu_{\mathrm{excess}}/\partial p)_T = v_{\mathrm{excess}}^{\mathrm{solute}}. \qquad (7)$$

At T = 298 K, we estimate $v_{\mathrm{excess}}^{\mathrm{solute}}$ = 123.9(3.0) cm^3/mol, comparable with the experimental[75] value of 122.0 cm^3/mol.

Again, as in the case of cyclohexane, we verify to what extent the cavity method reduces the "end-point" singularity during the naphthalene insertion. Figure 6 shows a plot of $\left\langle \partial U_{\mathrm{solute\text{-}solvent}}^{\mathrm{vdwl}} / \partial\lambda_{\mathrm{vdwl}} \right\rangle_{\lambda_{\mathrm{vdwl}}}$ for four cavity sizes λ = −0.5, −1, 0, and 5. In Fig. 6(a), λ = −1 is apparently still too small for a complete avoidance of the "end-point" singularity because $\left\langle \partial U_{\mathrm{solute\text{-}solvent}}^{\mathrm{vdwl}} / \partial\lambda_{\mathrm{vdwl}} \right\rangle_{\lambda_{\mathrm{vdwl}}}$ still shows divergent behavior as $\lambda_{\mathrm{vdwl}} \to 0$. In Fig. 6(b), we have a larger cavity at λ = −0.5: here, the divergence of $\left\langle \partial U_{\mathrm{solute\text{-}solvent}}^{\mathrm{vdwl}} / \partial\lambda_{\mathrm{vdwl}} \right\rangle_{\lambda_{\mathrm{vdwl}}}$ is already less pronounced. In Fig. 6(c), the "end-point" divergence at λ_{vdwl} = 0.005 is still visible but in Fig. 6(d) (λ = 5), the divergence, although in principle still present, has become unobservably small. We note that whilst divergences in $\left\langle \partial U_{\mathrm{solute\text{-}solvent}}^{\mathrm{vdwl}} / \partial\lambda_{\mathrm{vdwl}} \right\rangle_{\lambda_{\mathrm{vdwl}}}$ may affect the results of TI, they have no influence on FEP. In practice, single-step FEP tends to be less accurate at small λ, but the agreement between single-step-FEP and multi-steps FEP/TI gets better for larger λ. On the other hand, it is not favorable to grow an "oversized" cavity for small solutes, because this implies larger accumulated statistical errors.

Finally, we have also used the cavity method to compute the excess solvation free energies of naphthalene in water at higher temperatures of T = 320 K and 340 K at p = 0.1 MPa. All the results are plotted in Fig. 7 in comparison with experimental data.[29,37,70–72,76,77] All computed excess solvation free energies are summarized in Table I.

From the slope of Fig. 7(d), we can estimate the partial molar entropy of solvation of naphthalene in water via

$$(\partial\Delta\mu_{\mathrm{excess}}/\partial T)_p = -\Delta s_{\mathrm{excess}}^{\mathrm{solute}}. \qquad (8)$$

The experimental line in Fig. 7(d) has a slope of 0.12 (kJ/mol)/K, giving $\Delta s_{\mathrm{excess}}^{\mathrm{solute}}$ = −0.12 (kJ/mol)/K according to Eq. (7); similarly, our computed line in Fig. 7(d) gives $\Delta s_{\mathrm{solute}}$ = −0.092(14) (kJ/mol)/K. Using

$$\Delta h_{\mathrm{excess}}^{\mathrm{solute}} = \Delta\mu_{\mathrm{excess}} + T\Delta s_{\mathrm{excess}}^{\mathrm{solute}}, \qquad (9)$$

FIG. 5. Computed excess solvation free energies of naphthalene in water at T = 298 K for p at (a) 0.1 MPa, (b) 100 MPa, and (c) 200 MPa. The dashed lines in (a)–(c) represent the average values for $\lambda \gtrsim 0$ for the three pressures. These three average values are also plotted in (d).
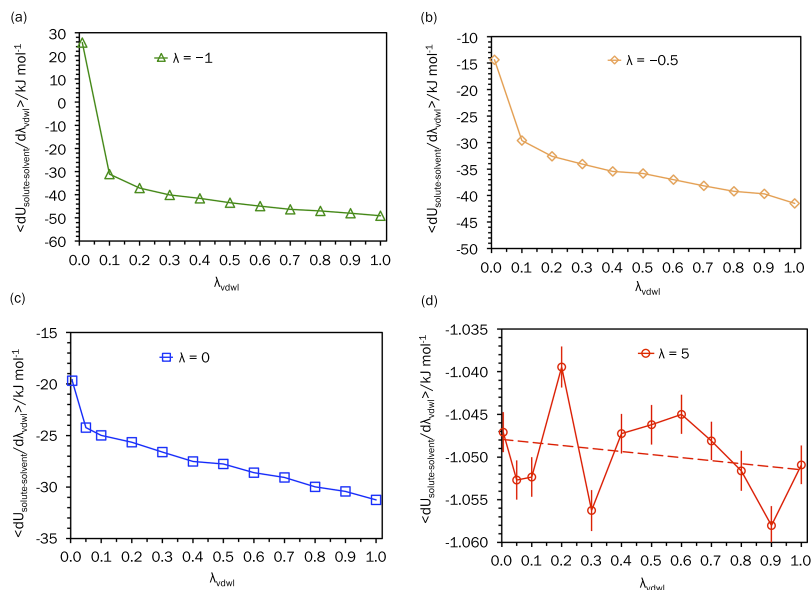
FIG. 6. $\left\langle \partial U_{\text{solute-solvent}}^{\text{vdwl}} \middle/ \partial \lambda_{\text{vdwl}} \right\rangle_{\lambda_{\text{vdwl}}}$ as a function of λ, for the naphthalene insertion at $T = 298$ K and $p = 0.1$ MPa at cavity size (a) $\lambda = -1$; (b) $\lambda = -0.5$; (c) $\lambda = 0$; and (d) $\lambda = 5$.

we obtain an estimate for the excess partial molar enthalpy of the solute: $\Delta h_{\text{excess}}^{\text{solute}} = -34.35$ kJ/mol comparing with the experimental $\Delta h_{\text{excess}}^{\text{solute}} = -46.86$ kJ/mol in Ref. 29.

C. Naphthalene: Absolute solid free energy and solubility estimation

Next, we consider the implementation of the extended Einstein crystal method to compute the free energy of the naphthalene crystals. A 20-point Gauss-Legendre quadrature series was used to evaluate ΔA_2 and ΔA_0. Obviously, we should use the same naphthalene force field for the solid and the liquid. Unfortunately, not all force fields describe the solid and liquid phases equally well. However, we found that the force field that we have used for the naphthalene solute can also reproduce the naphthalene monoclinic crystal lattice parameters reasonably well, with around 3% discrepancy between measured and simulated densities (see Fig. 8). Of course, this discrepancy could be fixed by improving the force field, but this was not our aim:

we wished to investigate how well "off-the-shelf" force fields perform.

Polycyclic aromatic hydrocarbons often occur in several polymorphs.[78] At the ambient pressure and temperature, naphthalene has a monoclinic unit cell with two molecules per unit cell. Fortunately, according to X-ray data, naphthalene remains as a single polymorph at room temperature for pressures up to at least 0.5 GPa.[79,80] Raman studies also indicate that naphthalene undergoes no structural phase transition[81] at least up to 3.5 GPa and 450 K. We used the ambient naphthalene crystal structure[82] to initiate all crystal equilibration simulations between $p = 0.1$-200 MPa at $T = 298$ K and $T = 298$-340 K at $p = 0.1$ MPa.

The final configuration of our simulated naphthalene unit cell at $T = 298$ K and $p = 0.1$ MPa has $a = 0.8320$ nm, $b = 0.5946$ nm, $c = 0.8454$ nm, $\beta = 118.07°$, and $\alpha = 89.8°$ and $\gamma = 90.24°$ with a unit cell volume of $0.373\,00$ nm^3, compared with the experimental lattice parameters[83] at $T = 296$ K: $a = 0.8261(7)$ nm, $b = 0.5987(2)$ nm, $c = 0.8682(8)$ nm,
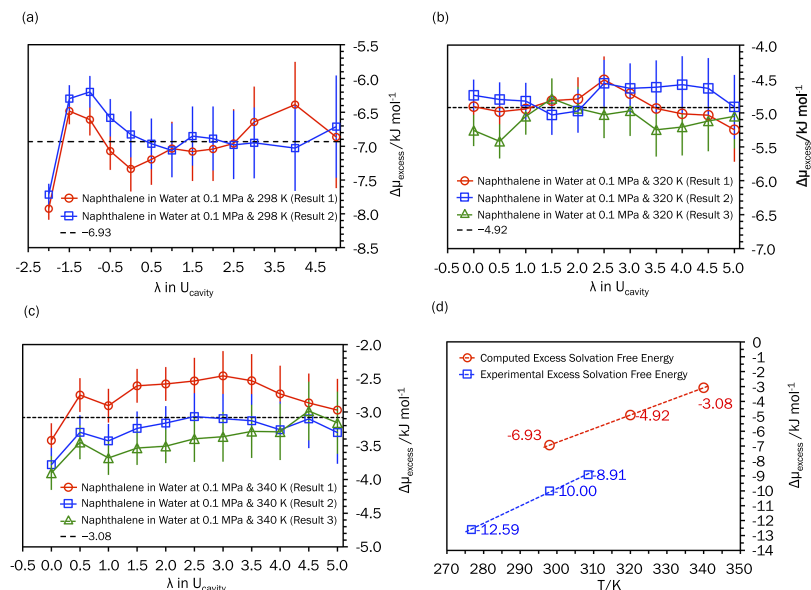

FIG. 7. Computed excess solvation free energies of naphthalene in water at $p = 0.1$ MPa and T of (a) 298 K; (b) 320 K; and (c) 340 K. The black dashed lines in (a)–(c) are the average from specific λ for the three temperatures ($\lambda \gtrsim 0$ for 298 K and 320 K, and $\lambda \gtrsim 1.0$ for 340 K), and they are also plotted (d). In (d), the empty blue squares are experimental data.

TABLE I. Summary of the computed excess solvation free energies of naphthalene and the infinitely dilute solution density. The statistical errors are in the range of 0.4-0.8 kJ/mol for medium to large cavities, and we consistently choose a typical 0.42 kJ/mol ($\approx$ 0.1 kcal/mol) as the common error.

p/MPa	T/K	$\Delta\mu_{excess}$/kJ mol^{-1}	$\rho_{solution}$/kg m^{-3}
0.1	298	−6.93 (42)	978.0(2)
100	298	5.53 (42)	1024.7(2)
200	298	17.81 (42)	1062.4(2)
0.1	320	−4.92 (42)	960.5(2)
0.1	340	−3.08 (42)	942.9(2)

$\beta = 122.67(8)°$, and $\alpha = \gamma = 90°$ with a unit cell volume of $0.361\,44$ nm^3.

In the extended Einstein crystal method, we tether a number of atoms (either real or virtual) in each molecule, to restrain the molecular position and orientation. As described in Appendix B, within each molecule, these particles can be categorized into two groups, (1) a single "Central Atom" and (2) a small number of "Orientational Atoms." The "Central Atom" is the atom attaching the molecule to its molecular site. The "Orientational Atoms" are the particles that, when tethered, fix the molecular orientation. For non-linear molecules, the number of "Orientational Atoms" per molecule should be at least two, but in practice, more "Orientational Atoms" are often necessary to reduce the effect of molecular flexibility in the extended Einstein crystal method. On the other hand, using too many "Orientational Atoms" increases the total statistical error because more mean-squared displacements have to be sampled for ΔA_0 and ΔA_2. The specific spring arrangement in the extended Einstein crystal method is called the "Einstein Arrangement" or "EA." In the example of naphthalene, we have used two Einstein arrangements. In Fig. 9, (a) is "EA 1", the first Einstein arrangement that we used for naphthalene; (b) is "EA 2", the second Einstein arrangement. For both Einstein arrangements, the "Central Atom" is the virtual central particle between the two aromatic carbon atoms. For
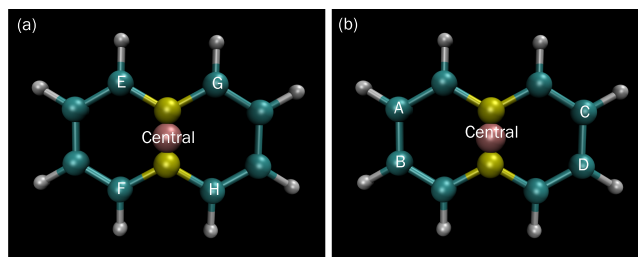

FIG. 9. (a) Einstein arrangement (EA) 1; (b) Einstein arrangement (EA) 2.

"EA 1," E, F, G, and H are the "Orientational Atoms," and for EA 2, A, B, C, and D are the "Orientational Atoms." Ideally, the results of the simulations do not depend on the choice of the EAs.

Table II shows the computed absolute solid free energies of naphthalene for $p = 0.1$-200 MPa and $T = 298$-340 K (see Tables S1-S5 in the supplementary material for details). pV_{solid} corresponds to the densities and target pressures of the corresponding reference structures.

As a test of the internal consistency of our results, we can compare the internal energy U as obtained from $\partial((A_{solid}/N_{solid})/T)/\partial(1/T) = U_{solid}/N_{solid}$, with the value obtained directly from the simulations. Figure 10 shows a plot of $(A_{solid}/N_{solid})/T$ versus $1/T$ with data from Table II, and the least squares fit slope is −65.7(9) kJ/mol. For comparison, we have run independent NVT simulations using the reference structures at these three temperatures, with U_{solid}/N_{solid} = −67.38(1) kJ/mol, −66.34(1) kJ/mol, and −65.78(1) kJ/mol at $T = 298$ K, 320 K, and 340 K, respectively. These values are all close to the slope of −65.7(9) kJ/mol. It should be noted that U_{solid}/N_{solid} consistently includes only the intermolecular interactions.

Similarly, we can compare the molar volume (more precisely, the "volume per molecule") obtained from $\partial(G_{solid}/N_{solid})/\partial p = v_{solid}$, with the value obtained directly from NpT simulations. Figure 11 shows the plot of G_{solid}/N_{solid}
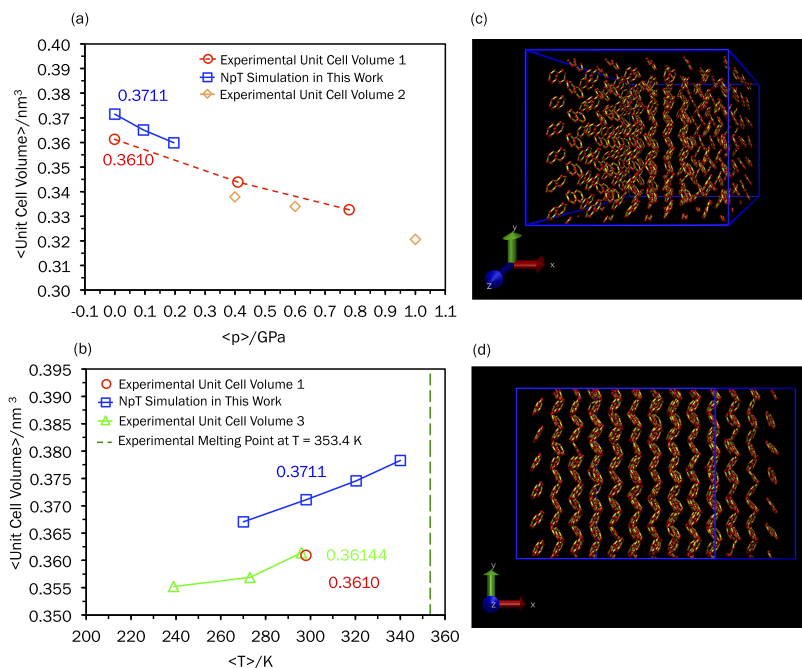

FIG. 8. (a) Simulated and experimental unit cell volumes of naphthalene with pressure. The empty blue squares are the averaged simulated results in this work, the empty red circles ("experimental unit cell volume 1" in the figure) are from the experiments of Ref. 81, and finally the empty orange rhombuses ("experimental unit cell volume 2" in the figure) are from the experiments of Ref. 78; (b) variation of simulated and experimental unit cell volumes with temperature. The empty blue squares are the averaged simulated results in this work, and the empty light green triangles ("experimental unit cell volume 3" in the figure) are from the experiments of Ref. 83. We cite the experimental melting point of 353.4 K from Ref. 84; (c) and (d) represent snapshots of the simulated naphthalene at $p = 0.1$ MPa and $T = 298$ K (hydrogen atoms are not shown and all bonds/atoms are colored equivalently).

TABLE II. Summary of the absolute solid free energies of naphthalene and densities of the reference structures and equilibrated crystals in NpT simulations. Energies are in kJ/mol.

p/MPa	T/K	A_{solid}/N_{solid}	pV_{solid}/N_{solid}	G_{solid}/N_{solid}	$\rho_{solid}/$ kg m^{-3}	$\langle\rho_{solid}\rangle/$ kg m^{-3}
0.1	298	−45.80	0.00	−45.80(8)	1141.2	1146.9(7)
100	298	−45.83	10.99	−34.84(8)	1164.9	1167.9(7)
200	298	−45.63	21.62	−24.01(8)	1184.0	1183.8(6)
0.1	320	−44.83	0.00	−44.83(8)	1132.1	1135.7(3)
0.1	340	−42.97	0.00	−42.97(8)	1129.5	1125.2(3)

FIG. 11. G_{solid}/N_{solid} against p. Energies are in kJ/mol and T is in K.

against p with data from Table II. After unit conversions, the gradient gives us $v_{solid} = 0.181(1)$ nm^3/molecule, compared with $v_{solid} = 0.1865$ nm^3, $v_{solid} = 0.1827$ nm^3, and $v_{solid} = 0.1798$ nm^3 of the reference structures at $p = 0.1$ MPa, 100 MPa, and 200 MPa, respectively. Also, $v_{solid} = 0.181(1)$ nm^3 corresponds to a partial molar volume of $v_{solid} = 109.0(6)$ cm^3/mol in the crystal phase, compared with the experimental value[75] of $v_{solid} = 108.2(3)$ cm^3/mol.

We are now in a position to estimate the aqueous solubility of naphthalene at these pressures and temperatures according to Eq. (4). In Fig. 12, we compare the computed solubilities with the experimental data.[75,85] In Fig. 12, our simulated result at $p = 0.1$ MPa and $T = 298$ K predicts a (mole fraction) solubility limit of $x = 4.74 \times 10^{-6}$ compared with the experimental value of $x = 4.40 \times 10^{-6}$. Also, as shown in Fig. 12(a), the negative pressure dependence of naphthalene solubility is reproduced fairly well, and in Fig. 12(b), the positive temperature dependence of naphthalene solubility is again well captured. Considering the simplicity of the model, such an agreement is almost certainly fortuitous.

From the slopes of the least squares fits in Fig. 12(a), we can estimate the experimental and computed volume change accompanying the dissolution,[75]

$$\Delta v = -RT\left[\partial \ln x/\partial p\right]_T, \tag{10}$$

where Δv is defined as the difference between the partial molar volume at infinite dilution in the solution v_{excess}^{solute} and the partial molar volume in the crystal v_{solid},

$$\Delta v = v_{excess}^{solute} - v_{solid}. \tag{11}$$

The linear least squares fit of the experimental data in Fig. 12(a) gives $\Delta v = 14.2(5)$ cm^3/mol, while the direct calculation of Ref. 75 gives 13.8(4) cm^3/mol. Our computed

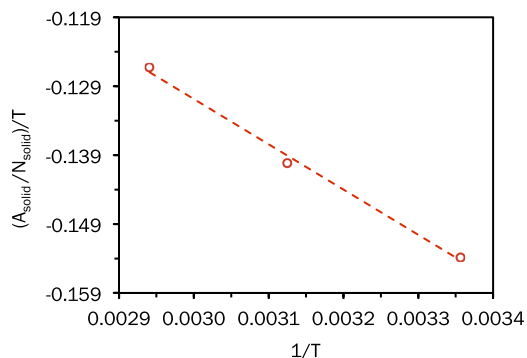line in Fig. 12(a) gives $\Delta v = 15.8(3.8)$ cm^3/mol. In our analysis of the excess solvation free energy of naphthalene, we had obtained $v_{excess}^{solute} = 123.9(3.0)$ cm^3/mol. Therefore we can deduce $v_{solid} = 15.8 - 123.9 = 108.1(6.8)$ cm^3/mol, and this is comparable with our direct calculation $v_{solid} = 109.0(6)$ cm^3/mol estimated from the slope of Fig. 12(a).

We have used a "flexible" model for naphthalene, and hence we have to check whether the effect of flexibility is different in the solid and the liquid phase. To this end, we tracked the intramolecular van der Waals plus electrostatic energy per naphthalene E_{intra}/N_{solid} at $p = 0.1$ MPa and $T = 298$ K (see the supplementary material). We find that the computed intramolecular energies in solution and in the solid agree within the statistical errors. We have also tracked the intramolecular bond-angle-torsion energy per molecule E_{mol}/N_{solid} throughout the complete route of the extended Einstein crystal method at $p = 0.1$ MPa and $T = 298$ K. We find that E_{mol}/N_{solid} can vary by 2 kJ/mol, depending on the spring constants K during the computation of ΔA_2 and ΔA_0. Also, using different Einstein arrangements seems to have a small but noticeable effect on the value of A_{solid}. For example, the largest difference within the range of computed A_{solid} is around

FIG. 12. Computed and experimental mole fraction solubility ln x of naphthalene in water (a) with varying pressure at $T = 298$ K; (b) with varying temperature at $p = 0.1$ MPa. The procedure by which we estimated the error bars in this figure is described in the supplementary material.

FIG. 10. $(A_{solid}/N_{solid})/T$ against $1/T$. Energies are in kJ/mol and T is in K.

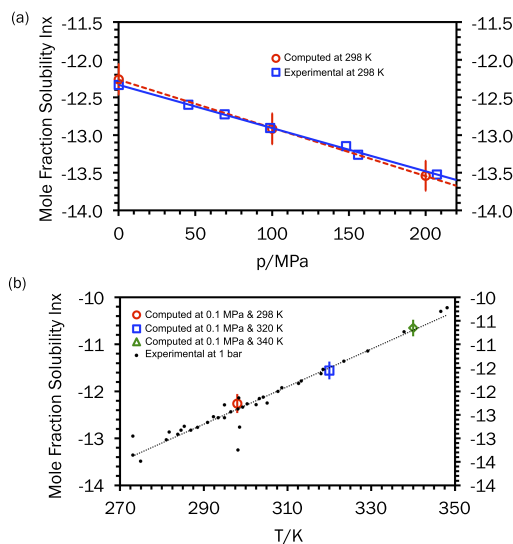0.8 kJ/mol, approximately 10 times of the statistical error 0.08 kJ/mol. Since in the current scheme, we compute ΔA_2 and ΔA_0 via sampling atomic mean-squared-displacements from their initial tethering atomic coordinates, it is not surprising that molecular flexibility in the extended Einstein crystal method has an effect. However, it is not straightforward to take these effects into account quantitatively. Fortunately, they have limited impact on the solubility prediction of naphthalene. However, to obtain higher accuracy, we would have to include the intramolecular parts in the Einstein crystal Hamiltonian, so as to compute explicitly the intramolecular contributions to the "Einstein free energy." An alternative is to use rigid molecular models, but for bending and torsion potentials, this would not be realistic.

As explained in Appendix B, we have used final configurations from long equilibrations as reference states for the extended Einstein crystal method. This should have no effect on the free energy estimates. We verified this by computing the absolute solid free energy of another naphthalene reference structure at $T = 298$ K and $p = 0.1$ MPa with $K_{max} = 1\,046\,000$ (kJ/mol)/nm^2. The values are $A_{solid}/N_{solid} = -46.19(8)$ kJ/mol and $-45.69(8)$ kJ/mol for Einstein Arrangement 2 and Einstein Arrangement 1, respectively, which are within the range of values (see the supplementary material).

V. CONCLUSION

In this work, we have developed a robust computational methodology to enable the solubility prediction of general crystalline solutes. The method is based on the equality of the chemical potentials of the solute in coexisting solution and solid phases. Thus, by separately computing the chemical potentials of the solute in the two phases via the cavity method and the extended Einstein crystal method, we obtain a numerical estimate of its solubility limit. The methodology is formulated by integrating existing alchemical numerical techniques such as free energy perturbation and thermodynamic integration with tailor-made protocols in the molecular dynamics software package LAMMPS.[30] Extensions to other software packages or homemade Monte Carlo codes should be straightforward. To demonstrate the methodology, we have estimated the solubility of naphthalene in water for a range of temperatures and pressures using simple all-atomic force fields. These predictions are all in reasonable agreement with experiments within a factor of 2.

Regarding the solution part of the methodology, the computation of excess solvation free energies of general solutes is already a well-established procedure in molecular simulations. The cavity method is computationally cheap when the solubilities of a range of compounds in the same solvent must be computed (the cavity only needs to be created once). For the solid part of the methodology, we have applied simple extensions of the original Einstein crystal method so that the method can be easily used in LAMMPS[30] without extra code. Potentially, our approach can predict the solubility of other polymorphs, not just the one that is most stable. Our current results show that a better treatment of molecular flexibility may be needed to improve the accuracy of the results.

The development of computing numerical solubilities has been hampered by two factors: (1) the limited availability of accurate force fields for molecules in crystals and in solution and (2) the fact that computing absolute solid free energies of molecular solids is not very popular, because the method, although straightforward, is cumbersome. In the literature, popular classical force fields, such as OPLS-AA[51] used in the present work, are mostly targeted at liquid phases, and in general, there are fewer accurate parameterizations for crystalline compounds. These problems may not always be relevant for the compound of interest, but they have indeed heightened the barrier for applying the Einstein crystal method and its adaptions to solubility estimation. Another obvious complication is that we have to know the experimental structure of the crystal (polymorph) that we study. In the present work, we have conducted all our simulations in the MD software package LAMMPS,[30] which at present does not allow for isobaric-isothermal NpT simulations of crystals of rigid molecules with a non-cubic simulation cell. We stress that many of the problems that we face are in no way exclusive to the topic of solubility prediction, but are generic to all numerical studies requiring high accuracy. Our key message is that there is no fundamental obstacle in developing a generic computational methodology for solubility prediction. Therefore, with the ever-increasing computational power, improving classical force fields (and hopefully, more experimental validation data appearing in the literature), and more solubility studies in the literature, it is likely that accurate yet general solubility predictions will soon be commonplace.

SUPPLEMENTARY MATERIAL

See supplementary material for details of the LAMMPS protocol to calculate the absolute chemical potentials of the solute in the aqueous solution and solid phases A_{solid} and $\Delta\mu_{excess}$ at $T = 298$ K and $p = 0.1$ MPa, and the discussion of the effect of phase on the intramolecular energy of the solute.

ACKNOWLEDGMENTS

APPENDIX A: CHEMICAL POTENTIAL OF THE SOLUTE

We adopt the definition of the chemical potential of a solute on the work of Widom[86,87] and the subsequent work of Ben-Naim.[73,88]

In the thermodynamic limit,

$$\mu_{solution} = G_{solution}(T, p, N_{solute} + 1, N_{solvent}) - G_{solution}(T, p, N_{solute}, N_{solvent}).$$

(A1)

μ_{solution} is therefore the change in Gibbs free energy associated with introducing an extra solute into the solution system $(T, p, N_{\text{solute}}, N_{\text{solvent}})$.

Denoting $G_{\text{solution}}(T, p, N_{\text{solute}} + 1, N_{\text{solvent}}) = G_1$ and $G_{\text{solution}}(T, p, N_{\text{solute}}, N_{\text{solvent}}) = G_0$, the Gibbs free energies of the two states are

$$G_1 = -kT \ln \Delta_1 = \ln \left\{ \begin{array}{l} \int \exp(-\beta pV) \, dV \dfrac{q_{\text{solute}}^{(N_{\text{solute}}+1)} q_{\text{solvent}}^{N_{\text{solvent}}}}{\Lambda_{\text{solvent}}^{3N_{\text{solvent}}} \Lambda_{\text{solute}}^{3(N_{\text{solute}}+1)} N_{\text{solvent}}! \, (N_{\text{solute}}+1)!} \\ \times \int \exp\left[-\beta U(N_{\text{solvent}}, N_{\text{solute}}+1)\right] d\mathbf{R}(N_{\text{solvent}}, N_{\text{solute}}+1) \end{array} \right\}, \tag{A2}$$

$$G_0 = -kT \ln \Delta_0 = \ln \left\{ \begin{array}{l} \int \exp(-\beta pV) \, dV \dfrac{q_{\text{solute}}^{N_{\text{solute}}} q_{\text{solvent}}^{N_{\text{solvent}}}}{\Lambda_{\text{solute}}^{3N_{\text{solute}}} \Lambda_{\text{solvent}}^{N_{\text{solvent}}} N_{\text{solute}}! \, N_{\text{solvent}}!} \\ \times \int \exp\left[-\beta U(N_{\text{solvent}}, N_{\text{solute}})\right] d\mathbf{R}(N_{\text{solvent}}, N_{\text{solute}}) \end{array} \right\}, \tag{A3}$$

with

$$\begin{aligned} U_1(N_{\text{solute}} + 1, N_{\text{solvent}}) &= U_0(N_{\text{solvent}}, N_{\text{solute}}) \\ &\quad + U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}}). \end{aligned} \tag{A4}$$

Δ is the isobaric-isothermal NpT partition function, $\mathbf{R}_{\text{solute}}$ is the fixed position of the extra single solute in the

solution, and $U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}})$ is the interaction energy of the extra solute with the rest of solution. Λ_{solute} and Λ_{solvent} are the de Broglie wavelength for the solute and solvent, and q_{solute} and q_{solvent} are the intramolecular (vibrational and electronic) partition function of the solute in the solution phase.

First, for the FEP procedure, we can write Eq. (4) as

$$\begin{aligned} \mu_{\text{solution}} &= G_1 - G_0 = -k_B T \ln(\Delta_1/\Delta_0) = -k_B T \ln\left\{ q_{\text{solute}} \middle/ \left[\Lambda_{\text{solute}}^3 (N_{\text{solute}} + 1) \right] \right\} \\ &\quad - k_B T \ln \left\{ \frac{\int \exp(-\beta pV) \, dV \int \exp[-\beta U(N_{\text{solvent}}, N_{\text{solute}} + 1)] \, d\mathbf{R}(N_{\text{solvent}}, N_{\text{solute}} + 1)}{\int \exp(-\beta pV) \, dV \int \exp[-\beta U(N_{\text{solvent}}, N_{\text{solute}})] \, d\mathbf{R}(N_{\text{solvent}}, N_{\text{solute}})} \right\} \\ &= k_B T \ln\left[\Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1} (N_{\text{solute}} + 1) \right] - k_B T \ln\langle V \exp[-\beta U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}})] \rangle_0. \end{aligned} \tag{A5}$$

The second term in Eq. (A5) is the FEP expression to compute the free energy difference between the solution with the extra solute and the solution without the extra solute, by sampling the quantity in the bracket at each equilibrium configuration of the solution with the extra solute, i.e., $(T, p, N_{\text{solute}}, N_{\text{solvent}})$, and taking the ensemble average. In Eq. (A5), if the systems are far away from critical points, it is safe to assume that the correlation between insertion energy and volume is weak,[89,90] so

$$\begin{aligned} \mu_{\text{solution}} &= k_B T \ln\left(\Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1} \right) + k_B T \ln(\rho_{\text{solute}}) \\ &\quad - k_B T \ln \langle \exp[-\beta U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}})] \rangle_0. \end{aligned} \tag{A6}$$

In Eq. (A6), we have number density concentration $\rho_{\text{solute}} = (N_{\text{solute}} + 1)/V \approx N_{\text{solute}}/V$. Substituting an ideal gas reference state of $\rho_{\text{solute}}^0 = 1$ molecule per unit volume into Eq. (A6) gives Eq. (2).

The excess solvation free energy/chemical potential of the solute is hence defined as

$$\Delta \mu_{\text{excess}}^{\rho_{\text{solute}}} = -k_B T \ln \left\langle \exp\left[-\beta U_{\text{solute-solution}}^{\rho_{\text{solute}}}(\mathbf{R}_{\text{solute}}) \right] \right\rangle_0. \tag{A7}$$

Although Eq. (A7) is general, we focus on the case of infinite dilution,

$$\begin{aligned} \Delta \mu_{\text{excess}}^{\rho_{\text{solute}} \to 0} &= k_B T \ln\left(\rho_{\text{solute}}^{\rho_{\text{solute}} \to 0} \right) \\ &\quad - k_B T \ln \left\langle \exp\left[-\beta U_{\text{solute-solution}}^{\rho_{\text{solute}} \to 0}(\mathbf{r}_{\text{solute}}) \right] \right\rangle_0. \end{aligned} \tag{A8}$$

So the activity coefficient γ_{solute} is defined as

$$\ln \gamma_{\text{solute}} \equiv \Delta \mu_{\text{excess}}^{\rho_{\text{solute}}} - \Delta \mu_{\text{excess}}^{\rho_{\text{solute}} \to 0}. \tag{A9}$$

This work considers only the case $\gamma_{\text{solute}} \to 1$, i.e., infinite dilution when $\Delta \mu_{\text{excess}}^{\rho_{\text{solute}}} = \Delta \mu_{\text{excess}}^{\rho_{\text{solute}} \to 0}$. However, we stress that the method that we describe will also in principle work for concentrated solutions. For materials with medium to high solubility, we will need to compute the coupling work required to introduce a single solute into the solution at a finite concentration ρ_{solute}. In the case of electrolytes, this procedure will be for each ion.

In Eq. (2) or (A6), the first term $\mu^0 = k_B T \ln\left(\Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1} \right)$ includes all the intramolecular (i.e., molecular internal degrees of freedom: vibrational and electronic) and molecular translational (i.e., de Broglie wavelength) contributions to the absolute chemical potential of the solute.

For the solid phase, with Q_{solid} as the canonical partition function, we can write the absolute Helmholtz free energy as

$$A_{\text{solid}} = -k_B T \ln Q_{\text{solid}}(N, V, T) = -k_B T \ln Z_{\text{solid}} - k_B T \ln P_{\text{solid}}$$

$$= -k_\text{B} T \ln \left\{ \int d\mathbf{r}^{N_{\text{solid}}} \exp \left[-\beta U_{\text{inter}} \left(\mathbf{r}^{N_{\text{solid}}} \right) \right] \right\}$$

$$- k_\text{B} T \ln \left(q_{\text{solid}}^{N_{\text{solid}}} \Big/ \Lambda_{\text{solid}}^{3N_{\text{solid}}} \right). \tag{A10}$$

In Eq. (A10), Z_{solid} is the configurational contribution and P_{solid} the momenta and intramolecular (vibrational and electronic) contribution to Q_{solid}. $U_{\text{inter}} \left(\mathbf{r}^{N_{\text{solid}}} \right)$ denotes the intermolecular potential energy. Λ_{solid} is the de Broglie wavelength for the solute in the solid phase, and q_{solid} is the intramolecular (vibrational and electronic) partition function of the solute in the solid phase.

Finally, substituting Eqs. (A10) and (A6) into Eq. (1) gives

$$\mu_{\text{solution}} = k_B T \ln \left(\Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1} \right) + k_B T \ln(\rho_{\text{solute}}) - k_B T \ln \langle \exp \left[-\beta U_{\text{solute-solution}} (\mathbf{R}_{\text{solute}}) \right] \rangle_0$$

$$= \mu_{\text{solid}} = \frac{-k_\text{B} T \ln \left\{ \int d\mathbf{r}^{N_{\text{solid}}} \exp \left[-\beta U_{\text{inter}} \left(\mathbf{r}^{N_{\text{solid}}} \right) \right] \right\} - k_\text{B} T \ln \left(q_{\text{solid}}^{N_{\text{solid}}} \Big/ \Lambda_{\text{solid}}^{3N_{\text{solid}}} \right)}{N_{\text{solid}}}. \tag{A11}$$

If we adopt $\Lambda_{\text{solid}} = \Lambda_{\text{solute}}$ and $q_{\text{solid}} = q_{\text{solute}}$ (the assumption that the intramolecular contributions in the solid/solution phases cancel each other out), the expression for solubility estimation in the unit of $k_B T$ per solute molecule is

$$k_B T \ln(\rho_{\text{solute}}) - k_B T \ln \langle \exp \left[-\beta U_{\text{solute-solution}} (\mathbf{R}_{\text{solute}}) \right] \rangle_0 = \frac{-k_\text{B} T \ln \left\{ \int d\mathbf{r}^{N_{\text{solid}}} \exp \left[-\beta U_{\text{inter}} \left(\mathbf{r}^{N_{\text{solid}}} \right) \right] \right\} + p V_{\text{solid}}}{N_{\text{solid}}}$$

$$\times k_B T \ln(\rho_{\text{solute}}) + \Delta \mu_{\text{excess}} = (A_{\text{solid}} + p V_{\text{solid}})/N_{\text{solid}}. \tag{A12}$$

Going from Eq. (A11) to Eq. (A12), the part of the intramolecular partition function associated with the molecular internal rotational degree of freedom is different for the solution and crystalline phase, because in the solution phase the solute is equally likely to have any orientation, while this is not the case in the crystalline phase. If we take an asymmetric top as an example (a similar argument would work for other types), in principle, we can define the molecular rotation with reference to any reference point because configurational properties do not depend on the choice of reference point. The proper orientational partition function for a solute molecule in the solution is $q_{\text{or}}^{\text{solution}} = \int \exp\left(-\beta u_{\text{or}}\right) \sin \theta d\theta d\phi d\gamma = 8\pi^2$, where θ, φ, and γ are the Euler angles, and u_{or} is the single molecular orientation Hamiltonian,[40] and we have $u_{\text{or}} = 0$ for a solute molecule that can freely orient in the liquid phase. Then, the important message is that, at one intermediate stage of the extended Einstein crystal method (see Appendix A), each freely rotating solute molecule has the same rotational partition function to the solute in the liquid phase, so the whole crystal has the total rotational partition function as $Q_{\text{or}}^{\text{solid}} = \left[\int \exp(-\beta u_{\text{or}}) \sin \theta d\theta d\phi d\gamma \right]^{N_{\text{solid}}} = \left(8\pi^2 \right)^{N_{\text{solid}}}$ with $\ln q_{\text{or}}^{\text{solid}} = \left[N_{\text{solid}} \ln \left(8\pi^2 \right) \right] \Big/ N_{\text{solid}} = \ln q_{\text{or}}^{\text{solution}}$. Here the rotational partition function is effectively cancelled out and hence Eq. (A12) is valid.

APPENDIX B: EXTENDED EINSTEIN CRYSTAL METHOD

Prior to the implementation of the extended Einstein crystal method, we need to have a suitable crystalline reference structure. Although it is common to take the minimum energy crystal structure as a reference, this is not necessary, as the computed free energy of the interacting crystal does not depend on this choice. In order to facilitate the use of the Einstein crystal method with standard MD software, we take a final configuration of the crystal after a long isobaric-isothermal NpT equilibration run as a reference.

There are several ways to constrain the position and orientation of molecules in a crystal using harmonic springs. The most elegant approach would be to tether the center of mass of the molecule to a lattice site using a harmonic spring, and to fix the orientation of the molecule using a "harmonic" constraint on the quaternion specifying the molecular orientation. However, as many existing codes use atomic coordinates, it is in practice more convenient (though certainly less elegant) to restrain the orientation of the molecules by tethering three or more atoms harmonically to the corresponding "atomic lattice sites" (see Fig. 1). Again, this choice should not introduce systematic errors in the estimate of the solid free energy.

In Fig. 1, (0) is an ideal Einstein atomic crystal with a "zero total momentum." When computing the partition function, we can, as before, ignore that part due to the momenta, so that the absolute Helmholtz free energy with no constraint on the total momentum or the center of mass is

$$\beta A_0 = 3/2 N_{\text{solid}} \ln \left(\beta K_{\max} / \pi \right). \tag{B1}$$

The potential energy for state (0) is $U_{(0)} = \sum_i^{N_{\text{solid}}} K_{\max} (\mathbf{r}_i - \mathbf{r}_{0i})^2$. N_{solid} is the total number of "atomic molecules" in (0), which is the same as the total number of molecules.

The constraint of a "zero total momentum" is similar to the requirement of fixing the center of mass in the original Einstein crystal method. We know that this constraint is important in MC simulations in the limit of vanishing spring constants, because the center of mass of the system could diffuse.[34,38]

However, precisely in this limit, it is sufficient to carry out MD simulations at zero total momentum to constrain the center of mass. The reason of choosing this condition is that in MD simulations it is straightforward to initiate the total momentum at zero. The particles are initially at their tethering positions, so the net force on the center of mass is zero. If the system is then initialized with zero total momentum, the center of mass will not move. This is strictly true if all atoms are involved in the Einstein crystal simulations and are of equal masses. There might be small fluctuations of the center of mass of the system if: (1) Einstein springs of the same strength K are attached to particles with different masses by using real atomic masses, this is because only when K_i/m_i is constant that we can maintain a fixed total momentum; (2) for molecules with non-linear molecular structures, during the thermodynamic stage of (1) and (2), "Orientational Atoms" are attached with different spring strength K from "Central Atoms," so K_i/m_i is not constant even when equal atomic masses are used. For practical applicability, we choose to remove the constraint of fixing the center of mass and total momentum from our partition function and simulations. In doing so, the only difference introduced to our computed absolute solid free energies is the analytic finite size correction associated with fixing the center of mass and total momentum. Along with the finite-size correction from harmonic crystal approximation,[43,44] they comprise all the finite-size corrections that are usually important to be included in theoretical studies of, e.g., Lennard-Jones (LJ) or hard spheres (HSs) crystals, where uncertainties can be as small as 0.001 $k_B T$ per molecule.[34,38] However, for solubility predictions of more complex molecular crystals, these finite-size corrections are negligible compared to other sources of error. This is because the accuracy of solubility estimation is dominated by the statistical errors of the computed excess solvation free energies and solid free energies, as well as the quality of the force field used. For the harmonic approximation, the dominant finite-size correction per molecule is proportional to $+\ln N_{solid}/N_{solid}$. For typical system sizes in the present work, this amounts to around $+\ln N_{solid}/N_{solid} = +0.02 \, k_B T/$molecule for $N_{solid} = 240$. As we will show, this is comparable with typical statistical errors in our computed solid free energies. For this reason, we simply ignore this term. The finite-size correction due to the constraint on the center of mass and total momentum is again of the order of $1/N$ and small compared with typical statistical errors. To estimate these magnitudes, we can compare Eq. (A1) $\beta A_0 = 3/2N\ln(\beta K_{max}/\pi)$, the analytic expression for an Einstein atomic crystal at state (0) without the constraints on the center of mass and the total momentum, and $\beta A_0' = 3/2(N-1)\ln(\beta K_{max}/\pi) + \ln(N/V) - 3/2\ln(N)$, the analytic expression for an Einstein atomic crystal with these constraints plus the general free energy difference between any two solids with and without these constraints.[41] The value of $(\beta A_0 - \beta A_0')/N$ is around 0.1 $k_B T/$molecule for $N_{solid} = 240$ and $K_{max} = 1\,046\,000$ (kJ/mol)/nm^2, comparable to or smaller than errors of computed excess solvation free energies.

For steps (0) and (1), we have $A_1 = A_0$. In (0), each atom is attached to its initial atomic lattice site through its center of mass. We can treat these blue solid atoms as the "Central Atoms" in (1). In (1), each molecule is attached to its

initial molecular site through its blue solid "Central Atom," around which the molecule is freely rotatable. Using atomic coordinates is a less elegant way to specify the molecular orientation, but we gain much practical convenience. For (1), the proper orientational partition function for a lattice containing N_{solid} molecules is $Q_{or} = [\int \exp(-\beta u_{or}) \sin\theta d\theta d\phi d\gamma]^{N_{solid}}$, where θ, φ, and γ are the Euler angles, and u_{or} is the single molecule orientation Hamiltonian.[40] For solubility estimation, the important information is that the same orientational partition function applies to a single solute in solution as well. Therefore $A_1 = A_0$ is valid as long as this same reference of an ideal gas is used for both phases.

In steps (1) and (2), the springs on the "orientational atoms" gradually grow from $K_{min} = 0$ to K_{max} to recover the molecular orientation of the real crystal. In (1), the potential energy is $U_{(1)} = \sum_j^{N_{solid}} \sum_i^{N_{springs}^{central}} K_{max}(\mathbf{r}_{ij} - \mathbf{r}_{0ij})^2$, with the summation over all the "Central Springs" and the total number of molecules in the solid; in (2), the potential energy is $U_{(2)} = \sum_j^{N_{solid}} \sum_i^{N_{springs}^{all}} K_{max}(\mathbf{r}_{ij} - \mathbf{r}_{0ij})^2$, with the summation over all the springs.

The accompanying free energy change ΔA_0 can be computed via TI,

$$\Delta A_0 = \int_{\ln c_1}^{\ln(K+c_1)} \left\langle \sum_j^{N_{solid}} \sum_i^{N_{springs}^{orientational}} (\mathbf{r}_{ij} - \mathbf{r}_{0ij})^2 \right\rangle_K$$
$$\times (K + c_1) \, d\ln(K + c_1). \quad (B2)$$

In Eq. (A4), the constant c_1 is added to make a change of the original variable K to allow for more accurate integration.[34,38]

To compute the free energy change ΔA_1 in steps (2) and (3), we implement a TI procedure, first switching on the electrostatic interactions, going from the non-interacting Einstein molecular crystal with the potential energy $U_{(2)} = \sum_j^{N_{solid}} \sum_i^{N_{springs}^{all}} K_{max}(\mathbf{r}_{ij} - \mathbf{r}_{0ij})^2$ to the electrostatically interacting Einstein molecular crystal with the potential energy $U_{(2.5)} = U_{(2)} + U_{inter}^{electrostatic}(\alpha = 1)$, with $0 \le \alpha \le 1$ as the atomic charge ratio in the crystal,

$$\Delta A_1^{electrostatic} = \int_0^1 \left\langle \partial U_{inter}^{electrostatic}(\alpha) / \partial\alpha \right\rangle_{\alpha, \lambda_{vdwl}} d\alpha. \quad (B3)$$

In the classical all-atomic MD, Columbic electrostatics is based on pairwise-additive point (partial) charges. Therefore, for a system of (partial) charges αq_i with i indicating the atom types, we can write the total electrostatic potential energy for the system as $U_{inter}^{electrostatic}(\alpha) = \sum_{i \ne j}^{N_{solid}} (\alpha q_i)(\alpha q_j)\phi(\mathbf{r}_{ij})$, where $\phi(\mathbf{r}_{ij})$ is the position-dependent electrostatic potential, so $\left\langle \partial U_{inter}^{electrostatic}(\alpha)/\partial\alpha \right\rangle_\alpha = \left\langle 2\alpha \sum_{i \ne j}(q_i)(q_j)\phi(\mathbf{r}_{ij}) \right\rangle = \left\langle 2U_{inter}^{electrostatic}(\alpha)/\alpha \right\rangle_\alpha$.

After computing $\Delta A_1^{electrostatic}$, we turn on the van der Waals interactions, going from $U_{(2.5)} =, U_{(2)} + U_{inter}^{electrostatic}(\alpha = 1)$ to $U_{(3)} = U_{(2)} + U_{inter}^{vdwl}(\lambda_{vdwl} = 1) + U_{inter}^{electrostatic}(\alpha = 1)$ via the Kirkwood linear coupling[91] with $0 \le \lambda_{vdwl} \le 1$,

$$\Delta A_1^{vdwl} = \int_0^1 \left\langle U_{inter}^{vdwl}(\lambda_{vdwl}) / \lambda_{vdwl} \right\rangle_{\lambda_{vdwl}} d\lambda_{vdwl}, \quad (B4)$$

$$\Delta A_1 = \Delta A_1^{electrostatic} + \Delta A_1^{vdwl}. \quad (B5)$$

The free energy change in (3) and (4) is denoted by ΔA_2. In (3), all the Einstein springs on the "Central Atoms" and "Orientational Atoms" are now equivalent, and thereafter during (3) and (4), all of them are reduced from at K_{max} to $K_{min} = 0$ together. For (4), the potential energy finally corresponds to the real crystal, i.e., $U_{(4)} = U_{inter}^{vdwl} + U_{inter}^{electrostatic}$.

The TI expression for ΔA_2 is

$$\Delta A_2 = -\int_{\ln c_2}^{\ln(K_{max}+c_2)} \left\langle \sum_j^{N_{solid}} \sum_i^{N_{springs}^{all}} \left(\mathbf{r}_{ij} - \mathbf{r}_{0ij} \right)^2 \right\rangle_K$$
$$\times (K + c_2)\, d\ln(K + c_2). \quad (B6)$$

Again, the constant c_2 is added to make a change of the original variable K to allow for more accurate integration.[34,38] In summary, the final expression for the absolute Helmholtz free energy A_{solid} of a molecular crystal is

$$A_{solid} = A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{symmetry}. \quad (B7)$$

APPENDIX C: CAVITY METHOD

In Fig. 1, the TI expression for ΔG_{grow} is

$$\Delta G_{grow} = G_{(2)} - G_{(1)} = \int \left\langle \partial U_{cavity}(\lambda) \big/ \partial\lambda \right\rangle_\lambda d\lambda$$
$$= \int \left\langle U_{cavity}(\lambda) \right\rangle_\lambda d\lambda. \quad (C1)$$

Having created the cavity, a solute can now be inserted into the cavity.

First, we switch on the solute-solvent van der Waals interactions and compute the van der Waals contribution via TI,

$$\Delta G_{insert}^{vdwl} = \int_0^1 \left\langle \partial U_{(2)\to(2.5)}(\lambda_{vdwl}) \big/ \partial\lambda_{vdwl} \right\rangle_{\lambda_{vdwl}} d\lambda_{vdwl}. \quad (C2)$$

Of course, ΔG_{insert}^{vdwl} can also be computed via FEP.

Then we switch on the atomic (partial) charges of the solute and compute the electrostatic contribution via FEP,

$$\Delta G_{insert}^{electrostatic} = -k_B T \ln\left\langle \exp\left[-\beta\left(U_{(3)} - U_{(2.5)}\right)\right]\right\rangle_{(2.5)}. \quad (C3)$$

The total insertion free energy is therefore

$$\Delta G_{insert} = \Delta G_{insert}^{vdwl} + \Delta G_{insert}^{electrostatic}. \quad (C4)$$

ΔG_{insert}, ΔG_{insert}^{vdwl}, and $\Delta G_{insert}^{electrostatic}$ are all computed in the same cavity λ created by ΔG_{grow}.

For small neutral solutes, or if the size of the cavity is large enough, ΔG_{grow} can be easily obtained by a single-step FEP. In principle, we can always create a cavity that is sufficiently large for the single-step FEP insertion of any large solutes, by using a larger λ. However, our purpose is to reduce the "end-point" singularity as much as possible, and as long as this is satisfied, there is no need to create a larger cavity, in which case we can use two-steps/multi-steps FEP/TI.

Finally, in (3) and (4), we gradually switch off the cavity potential to recover the final state of a fully solvated solute in the solution,

$$\Delta G_{shrink} = G_{(4)} - G_{(3)} = -\int \left\langle \partial U_{cavity}(\lambda) \big/ \partial\lambda \right\rangle_\lambda d\lambda$$
$$= -\int \left\langle U_{cavity}(\lambda) \right\rangle_\lambda d\lambda. \quad (C5)$$

Because the removal of "end-point" singularity is a gradual process, there is no strict criterion to determine when a cavity becomes "sufficiently large," but we use as a rule of thumb that the solvent should never experience the repulsive part of the solute-solvent interaction. Of course, the singularity problem can be avoided by using a small FEP calculation to replace the first step in the TI.

[1] G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, "A theoretical basis for a biopharmaceutic drug classification: The correlation of *in vitro* drug product dissolution and *in vivo* bioavailability," Pharm. Res. **12**, 413 (1995).

[2] W. L. Jorgensen and E. M. Duffy, "Prediction of drug solubility from structure," Adv. Drug Delivery Rev. **54**, 355–366 (2002).

[3] W. L. Jorgensen and E. M. Duffy, "Prediction of drug solubility from Monte Carlo simulations," Bioorg. Med. Chem. Lett. **10**, 1155 (2000).

[4] S. J. Park and G. A. Mansoori, "Aggregation and deposition of heavy organics in petroleum crudes," Energy Sources **10**, 109 (1988).

[5] K. Raju and G. Atkinson, "Thermodynamics of 'scale' mineral solubilities. 1. BaSO$_4$(s) in H$_2$O and aqueous NaCl," J. Chem. Eng. Data **33**, 490 (1988).

[6] K. U. G. Raju and G. Atkinson, "Thermodynamics of 'scale' mineral solubilities. 3. Calcium sulfate in aqueous NaCl," J. Chem. Eng. Data **35**, 361 (1990).

[7] T. Chen, A. Neville, and M. Yuan, "Calcium carbonate scale formation—Assessing the initial stages of precipitation and deposition," J. Pet. Sci. Eng. **46**, 185–194 (2005).

[8] 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).

[9] J. Kolafa, "Solubility of NaCl in water and its melting point by molecular dynamics in the slab geometry and a new BK3-compatible force field," J. Chem. Phys. **145**, 204509 (2016).

[10] 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 (2005).

[11] F. Moučka, M. Lísal, and W. R. Smith, "Molecular simulation of aqueous electrolyte solubility. 3. Alkali–halide salts and their mixtures in water and in hydrochloric acid," J. Phys. Chem. B **116**, 5468 (2012).

[12] F. Moučka, I. Nezbeda, and W. R. Smith, "Chemical potentials, activity coefficients, and solubility in aqueous NaCl solutions: Prediction by polarizable force fields," J. Chem. Theory Comput. **11**, 1756 (2015).

[13] 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 (2016).

[14] M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, and P. Turq, "Solubility of KF in water by molecular dynamics using the Kirkwood integration method," J. Chem. Phys. **117**, 4947 (2002).

[15] H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, and J. Alejandre, "Direct coexistence methods to determine the solubility of salts in water from numerical simulations. Test case NaCl," J. Phys. Chem. B **119**, 8389 (2015).

[16] J. L. Aragones, E. Sanz, and C. Vega, "Solubility of NaCl in water by molecular simulation revisited," J. Chem. Phys. **136**, 244508 (2012).

[17] 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).

[18] 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).

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

[20] Z. Mester and A. Z. Panagiotopoulos, "Temperature–dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations," J. Chem. Phys. **143**, 044505 (2015).

[21] F. Moucka, M. Lisal, J. Skvor, J. Jirsak, 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 (2011).

[22]M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren, "The structure, thermodynamics, and solubility of organic crystals from simulation with a polarizable force field," J. Chem. Theory Comput. **8**, 1721 (2012).

[23]D. S. Palmer, A. Llinàs, I. Morao, G. M. Day, J. M. Goodman, R. C. Glen, and J. B. O. Mitchell, "Predicting intrinsic aqueous solubility by a thermodynamic cycle," Mol. Pharm. **5**, 266 (2008).

[24]D. S. Palmer, J. L. McDonagh, J. B. O. Mitchell, T. Van Mourik, and M. V. Fedorov, "First-principles calculation of the intrinsic aqueous solubility of crystalline druglike molecules," J. Chem. Theory Comput. **8**, 3322 (2012).

[25]J. C. Dearden, "In silico prediction of aqueous solubility," Expert Opin. Drug Discovery **1**, 31 (2006).

[26]A. S. Paluch and E. J. Maginn, "Predicting the solubility of solid phenanthrene: A combined molecular simulation and group contribution approach," AIChE J. **59**, 2647 (2013).

[27]M. J. Lazzaroni, D. Bush, C. A. Eckert, T. C. Frank, S. Gupta, and J. D. Olson, "Revision of MOSCED parameters and extension to solid solubility calculations," Ind. Eng. Chem. Res. **44**, 4075 (2005).

[28]R. T. Ley, G. B. Fuerst, B. N. Redeker, and A. S. Paluch, "Developing a predictive form of MOSCED for nonelectrolyte solids using molecular simulation: Application to acetanilide, acetaminophen, and phenacetin," Ind. Eng. Chem. Res. **55**, 5415 (2016).

[29]S. Cabani, G. Mollica, and V. Lepori, "Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution," J. Solution Chem. **10**, 563 (1981).

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

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

[32]R. W. Zwanzig, "High-temperature equation of state by a perturbation method. II. Polar gases," J. Chem. Phys. **23**, 1915 (1955).

[33]D. A. Kofke and P. T. Cummings, "Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation," Mol. Phys. **92**, 973 (1997).

[34]D. Frenkel and B. Smith, *Understanding Molecular Simulations*, 2nd ed. (Academic, London, 2002).

[35]P. Kollman, "Free energy calculations: Applications to chemical and biochemical phenomena," Chem. Rev. **93**, 2395 (1993).

[36]W. L. Jorgensen, J. F. Blake, and J. K. Buckner, "Free energy of TIP4P water and the free energies of hydration of CH_4 and Cl^- from statistical perturbation theory," Chem. Phys. **129**, 193 (1989).

[37]D. Shivakumar, J. Williams, Y. J. Wu, W. Damm, J. Shelley, and W. Sherman, "Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field," J. Chem. Theory Comput. **6**, 1509 (2010).

[38]D. Frenkel and A. J. 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 (1984).

[39]E. J. Meijer, D. Frenkel, R. A. LeSar, and A. J. C. Ladd, "Location of melting point at 300 K of nitrogen by Monte Carlo simulation," J. Chem. Phys. **92**, 7570 (1990).

[40]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).

[41]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).

[42]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).

[43]K. Ioannidou, M. Kanduč, L. Li, D. Frenkel, J. Dobnikar, and E. D. Gado, "The crucial effect of early-stage gelation on the mechanical properties of cement hydrates," Nat. Commun. **7**, 12106 (2016).

[44]M. S. Sellers, M. Lísal, and J. K. Brennan, "Free–energy calculations using classical molecular simulation: Application to the determination of the melting point and chemical potential of a flexible RDX model," Phys. Chem. Chem. Phys. **18**, 7841 (2016).

[45]D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).

[46]W. G. Hoover, "Entropy for small classical crystals," J. Chem. Phys. **49**, 1981 (1968).

[47]J. Polson, E. Trizac, S. Pronk, and D. Frenkel, "Finite-size corrections to the free energies of crystalline solids," J. Chem. Phys. **112**, 5339 (2000).

[48]T. Steinbrecher, D. L. Mobley, and D. A. Case, "Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculation," J. Chem. Phys. **127**, 214108 (2007).

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

[50]J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, "Thermodynamics of cavity formation in water. A molecular dynamics study," Faraday Symp. Chem. Soc. **17**, 55 (1982).

[51]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," J. Am. Chem. Soc. **118**, 11225 (1996).

[52]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.

[53]D. Berthelot, Compt. Rend. Acad. Sci. Paris **126**, 1703 (1889).

[54]H. A. Lorentz, Ann. Phys. **248**, 127 (1881).

[55]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1987).

[56]W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," Phys. Rev. B **69**, 134103 (2004).

[57]G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," J. Chem. Phys. **101**, 4177 (1994).

[58]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182 (1981).

[59]L. Verlet, "Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," Phys. Rev. **159**, 98 (1967).

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

[61]W. G. Hoover, "Constant-pressure equations of motion," Phys. Rev. A **34**, 2499 (1986).

[62]T. Schneider and E. Stoll, "Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions," Phys. Rev. B **17**, 1302 (1978).

[63]P. P. Ewald, Ann. Phys. **369**, 253 (1921).

[64]R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1989).

[65]S. W. de Leeuw, J. W. Perram, and E. R. Smith, "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants," Proc. R. Soc. A **373**, 27 (1980).

[66]M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[67]K. E. Atkinson, *An Introduction to Numerical Analysis*, 2nd ed. (John Wiley & Sons, New York, 1989).

[68]J. E. Lennard-Jones, "On the determination of molecular fields," Proc. R. Soc. A **106**, 463 (1924).

[69]H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," J. Chem. Phys. **91**, 461 (1989).

[70]J. Hine and P. K. Mookerjee, "Structural effects on rates and equilibriums. XIX. Intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions," J. Org. Chem. **40**, 292 (1975).

[71]R. D. Wauchope and R. Haque, "Aqueous solubility of polynuclear aromatic hydrocarbons," Can. J. Chem. **50**, 133 (1972).

[72]W. Shiu and K. Ma, "Temperature dependence of physical–chemical properties of selected chemicals of environmental interest. I. Mononuclear and polynuclear aromatic hydrocarbons," J. Phys. Chem. Ref. Data **29**, 41 (2000).

[73]A. Ben-Naim and Y. Marcus, "Solvation thermodynamics of nonionic solutes," J. Chem. Phys. **81**, 2016 (1984).

[74]J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen, "Determination of partial molar volumes from free energy perturbation theory," Phys. Chem. Chem. Phys. **17**, 8407 (2015).

[75]S. Sawamura, M. Tsuchiya, T. Ishigami, Y. Taniguchi, and K. Suzuki, "Effect of pressure on the solubility of naphthalene in water at 25 °C," J. Solution Chem. **22**, 727 (1993).

[76]A. C. Chamberlin, C. J. Cramer, and D. G. Truhlar, "Predicting aqueous free energies of solvation as functions of temperature," J. Phys. Chem. B **110**, 5665 (2006).

[77]M. Alaee, R. M. Whittal, and W. M. Strachan, "The effect of water temperature and composition on Henry's law constant for various PAH's," Chemosphere **32**, 1153 (1996).

[78]F. P. A. Fabbiani, D. R. Allan, S. Parsons, and C. R. Pulham, "Exploration of the high-pressure behaviour of polycyclic aromatic hydrocarbons: Naphthalene, phenanthrene and pyrene," Acta Crystallogr., Sect. B: Struct. Sci. **62**, 826 (2006).

[79] H. C. Alt and J. Kalus, "X-ray powder diffraction investigation of naphthalene up to 0.5 GPa," Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **38**, 2595 (1982).

[80] A. Y. Likhacheva, S. V. Rashchenko, A. D. Chanyshev, T. M. Inerbaev, K. D. Litasov, and D. S. Kilin, "Thermal equation of state of solid naphthalene to 13 GPa and 773 K: *In situ* x-ray diffraction study and first principles calculations," J. Chem. Phys. **140**, 164508 (2014).

[81] M. Nicol, M. Vernon, and J. T. Woo, "Raman spectra and defect fluorescence of anthracene and naphthalene crystals at high pressures and low temperatures," J. Chem. Phys. **63**, 1992 (1975).

[82] S. C. Capelli, A. Mason, S. A. Willis, and B. T. M. Willis, "Molecular motion in crystalline naphthalene: Analysis of multi-temperature x-ray and neutron diffraction data," J. Phys. Chem. A **110**, 11695 (2006).

[83] C. P. Brock and J. D. Dunitz, "Temperature dependence of thermal motion in crystalline naphthalene," Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **38**, 2218 (1982).

[84] M. J. O'Neil, *The Merck Index—An Encyclopedia of Chemicals, Drugs, and Biologicals* (Royal Society of Chemistry, Cambridge, UK, 2013).

[85] D. G. Shaw, A. Maczynski, M. Goral, B. Wisniewska-Goclowska, A. Skrzecz, I. Owczarek, K. Blazej, M.-C. Haulait-Pirson, G. T. Hefter, P. L. Huyskens, F. Kapuku, Z. Maczynska, and A. Szafranski, "IUPAC-NIST solubility data series. 81. Hydrocarbons with water and seawater—Revised and updated. Part 9. C_{10} hydrocarbons with water," J. Phys. Chem. Ref. Data **35**, 93 (2006).

[86] B. Widom, "Some topics in the theory of fluids," J. Chem. Phys. **39**, 2808 (1963).

[87] B. Widom, "Potential-distribution theory and the statistical mechanics of fluids," J. Phys. Chem. **86**, 869 (1982).

[88] A. Ben-Naim, "Standard thermodynamics of transfer. Uses and misuses," J. Phys. Chem. **82**, 792 (1978).

[89] D. M. Heyes, "Molecular dynamics at constant pressure and temperature," Chem. Phys. **82**, 285 (1983).

[90] K. S. Shing and S. T. Chung, "Computer simulation methods for the calculation of solubility in supercritical extraction systems," J. Phys. Chem. **91**, 1674 (1987).

[91] J. G. Kirkwood, "Statistical mechanics of fluid mixtures," J. Chem. Phys. **3**, 300 (1035).