



# MIT Open Access Articles

*Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>As Published</b>	10.1063/1.5086706
<b>Publisher</b>	AIP Publishing
<b>Version</b>	Final published version
<b>Citable link</b>	<a href="https://hdl.handle.net/1721.1/135116">https://hdl.handle.net/1721.1/135116</a>
<b>Terms of Use</b>	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

# Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments

Cite as: J. Chem. Phys. **150**, 094107 (2019); <https://doi.org/10.1063/1.5086706>

Submitted: 22 December 2018 . Accepted: 19 February 2019 . Published Online: 05 March 2019

Michael A. Bellucci , Gianpaolo Gobbo , Tharanga K. Wijethunga , Giovanni Ciccotti , and  
Bernhardt L. Trout 



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[A coarse-grained polymer model for studying the glass transition](#)

The Journal of Chemical Physics **150**, 091101 (2019); <https://doi.org/10.1063/1.5089417>

[Lack of an equation of state for the nonequilibrium chemical potential of gases of active particles in contact](#)

The Journal of Chemical Physics **150**, 094108 (2019); <https://doi.org/10.1063/1.5085740>

[A preconditioning scheme for minimum energy path finding methods](#)

The Journal of Chemical Physics **150**, 094109 (2019); <https://doi.org/10.1063/1.5064465>

# Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments

Cite as: J. Chem. Phys. 150, 094107 (2019); doi: 10.1063/1.5086706

Submitted: 22 December 2018 • Accepted: 19 February 2019 •

Published Online: 5 March 2019



Michael A. Bellucci,<sup>1,a)</sup> Gianpaolo Gobbo,<sup>1,b)</sup> Tharanga K. Wijethunga,<sup>1</sup> Giovanni Ciccotti,<sup>2,3</sup>   
and Bernhardt L. Trout<sup>1,c)</sup>

## AFFILIATIONS

<sup>1</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

<sup>2</sup>Università di Roma La Sapienza, Ple. A. Moro 5, 00185 Roma, Italy

<sup>3</sup>School of Physics, University College of Dublin, Belfield, Dublin 4, Ireland

<sup>a)</sup>Current address: XtalPi, Inc., One Broadway, Ninth floor, Cambridge, Massachusetts 02139, USA.

<sup>b)</sup>Contributions: Michael A. Bellucci and Gianpaolo Gobbo equally contributed to this work.

<sup>c)</sup>Electronic mail: trout@mit.edu

## ABSTRACT

Li and co-workers [Li *et al.*, J. Chem. Phys. **146**, 214110 (2017)] have recently proposed a methodology to compute the solubility of molecular compounds from first principles, using molecular dynamics simulations. We revise and further explore their methodology that was originally applied to naphthalene in water at low concentration. In particular, we compute the solubility of paracetamol in an ethanol solution at ambient conditions. For the simulations, we used a force field that we previously reparameterized to reproduce certain thermodynamic properties of paracetamol but not explicitly its solubility in ethanol. In addition, we have determined the experimental solubility by performing turbidity measurements using a Crystal16 over a range of temperatures. Our work serves a dual purpose: (i) methodologically, we clarify how to compute, with a relatively straightforward procedure, the solubility of molecular compounds and (ii) applying this procedure, we show that the solubility predicted by our force field ( $0.085 \pm 0.014$  in mole ratio) is in good agreement with the experimental value obtained from our experiments and those reported in the literature (average  $0.0585 \pm 0.004$ ), considering typical deviations for predictions from first principle methods. The good agreement between the experimental and the calculated solubility also suggests that the method used to reparameterize the force field can be used as a general strategy to optimize force fields for simulations in solution.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5086706>

## I. INTRODUCTION

The determination of solubility is important across different fields, including those with practical applications. For example, the solubility of organic molecules is especially important in the pharmaceutical field because it is one of the factors determining the bioavailability of a drug substance.<sup>1–3</sup> In addition, the solubility of a drug substance in various solvents is central to the design of crystallization processes in drug manufacturing.<sup>4,5</sup> While these examples are already sufficient to illustrate the importance of solubility, there are countless other physical and industrial processes where solubility is important, in addition to its scientific importance in chemical physics.

Historically, many of the methods developed to compute solubilities use empirical or semiempirical correlations,<sup>6–11</sup> group contribution methods,<sup>12–16</sup> or quantitative structure-property relationships (QSPRs) combined with experimental data to predict solubility.<sup>17–21</sup> These methods have the advantage that they are computationally inexpensive, and therefore they are often used as an efficient screening tool for solubility. However, many of these methods provide a wide range of predictions, and QSPR based predictions, in particular, are known to be unreliable for molecules dissimilar to those in the training set. A more rigorous approach is to predict solubility from first-principles using molecular simulations. Molecular simulations provide a powerful tool for solubility prediction, but until recently, few such approaches have been developed, and in

general, simulation-based predictions of solubility are still not widely used. This is particularly true for the prediction of the solubility of molecular solids since most of the methods previously developed have been applied to ionic solids, such as salts,<sup>22–38</sup> and have not been extended to the more complex case of molecular solids. However, there are exceptions<sup>39–43</sup> and of particular note are the direct coexistence method<sup>36–38</sup> and the procedure, based on the extended Einstein crystal method (EECM), proposed by Li *et al.*<sup>43</sup>

The direct coexistence method is based on the equilibration of a solid with the solution. However, while this method is straightforward to use, it generally requires long simulation times to reach a solubility equilibrium, and this is especially true for drug molecules, whose dynamics is often characterized by slow relaxation time scales. Another well known method for computing the solubility from first principles is the osmotic ensemble method.<sup>28–31</sup> This method uses a grand canonical Monte Carlo procedure to dynamically insert/delete solute atoms from the solution, which makes the extension of the method to molecular systems non-trivial. By contrast, the method of Li *et al.* computes the solubility by equating the chemical potentials of the solute in the solution and solid phase. As a consequence, determining the solubility amounts to computing the free energy of the solute in the solution and solid phase, the latter of which is computed using the EECM. Avoiding the direct simulation of the coexistence of two phases, this method is an ideal candidate for calculating the solubility of pharmaceutically relevant molecules. Li *et al.* have applied their method to calculate the solubility of a sparingly soluble nonpolar molecule (naphthalene in water), but to the best of the authors' knowledge, this method has not been applied yet to compute the solubility of more challenging molecules that have higher solubility and are strongly polar.

Solubility calculations are important not only for experimental prediction, but also, in certain cases, they are necessary for molecular simulations themselves. In particular, for computational studies of nucleation<sup>44</sup> from solution, it is important to know the solubility predicted by the force field model in the solvent of interest. Since supersaturation is the driving force for nucleation, solubility is central to any crystallization study regardless of whether it is experimental or computational. Consequently, the solubility characterizing a force field is important not only as a measure of the quality of the force field but also because it is necessary to correctly determine the supersaturation of the system. However, force fields are not typically parameterized using solubility related quantities and, in general, cannot be expected to reproduce accurately experimental solubilities. In fact, in the cases where existing force fields have been used to compute the solubility, it has been shown that, even for simple ionic crystals, the value of the solubility predicted by the different force fields can show deviations up to a factor of 5–8, from the experimental value and between the different force fields.<sup>31</sup>

Recently we have developed a simple method to reoptimize the CHARMM general force field (CGenFF) to reproduce certain thermodynamic properties that are relevant for nucleation simulations (see Sec. II B), and we have applied it to paracetamol.<sup>44,45</sup> One of the objectives of this work is to reconstruct, clarify, and apply the EECM of Li *et al.* to a more challenging system characterized by a relatively high solubility and strong polar interactions.

At the same time, we wanted to test our force field reparameterization procedure to assess how well the resulting force field can predict solubility. To this end, we compute the solubility of paracetamol in ethanol and compare it with the value obtained from experiments performed by us and with values reported in the literature. Paracetamol is a good model compound due to its relevance in the pharmaceutical industry and due to the availability of existing experimental data. Moreover, the results generated here could also inform computational studies of nucleation and epitaxy.<sup>45–51</sup>

The rest of the paper is organized as follows. In Sec. II A, we reconstruct in detail and clarify the theoretical framework for the calculation of solubilities of molecular compounds introduced in Ref. 43, i.e., the so-called EECM. In Sec. II B, we discuss details of the free energy calculations and the simulation setup, while in Sec. II C we discuss the experimental measures of solubility of paracetamol in ethanol. Results are discussed in Sec. III.

## II. METHODOLOGY

### A. Theoretical background

By definition, the solubility of a substance in a solution is determined by the concentration of solute for which the chemical potential of the solute in solution is equivalent to the chemical potential of the solute substance in the crystal phase,

$$\mu_{\text{sol}} = \mu_{\text{crys}}. \quad (1)$$

If we neglect finite size effects, the chemical potential in the crystal phase is nothing but the Gibbs free energy,  $G_{\text{crys}}$ , per unit molecule. However, since the calculations we need to perform involve restraints defined by absolute positions in space, it is more convenient to perform simulations at constant volume and compute Helmholtz free energies instead of Gibbs free energies directly. For the chemical potential, the following relation holds:

$$\mu_{\text{crys}} = \frac{G_{\text{crys}}}{N_{\text{crys}}} = \frac{A_{\text{crys}} + pV}{N_{\text{crys}}}, \quad (2)$$

where  $A_{\text{crys}}$  is the Helmholtz free energy of the crystal,  $N_{\text{crys}}$  is the total number of molecules, and  $p$  and  $V$  are, respectively, the pressure and the volume of the system. The free energy for a non-harmonic crystal cannot be easily computed directly. However, it is possible to compute it by considering an ideal system as a reference, the absolute free energy of which can be computed analytically. It is then possible to compute the free energy difference between this system and the actual crystal through a series of alchemical transformations. We choose as reference an Einstein crystal in which the molecules are not interacting with each other and only their centers are tethered to the crystal lattice sites by harmonic springs. We call this a freely rotatable Einstein crystal. The free energy of the freely rotatable Einstein crystal is

$$\begin{aligned} & -\beta^{-1} N_{\text{crys}} \ln \left[ \frac{q}{\Lambda^{3n}} \right] + \beta^{-1} \frac{3}{2} N_{\text{crys}} \ln(\beta K_E / 2\pi) \\ & = -\beta^{-1} N_{\text{crys}} \ln \left[ \frac{q}{\Lambda^{3n}} \right] + A_0. \end{aligned} \quad (3)$$

In Eq. (3),  $\beta^{-1} = k_b T$ ,  $K_E$  is the constant determining the strength of the harmonic restraints, and we have defined  $A_0$  as the free energy contribution associated with the harmonic restraints. Notice that for simplicity we are considering for our derivation the case of fully flexible molecules. This guarantees that the kinetic and intramolecular (the  $q$  term) parts of the partition function are independent from each other and from the part originating from the external field. However, the overall derivation can be generalized to the case of constrained systems without altering the final result in Eq. (15) (the demonstration of this fact is not difficult but a bit convoluted and will be the subject of future work). We will make use of this fact in Sec. II B, where the simulations we describe are performed with constraints. In the first term of the summation in Eq. (3),  $\Lambda^{3n}$  represents the contribution to the partition function associated with the momenta of one molecule and is defined by

$$\Lambda^{3n} \equiv \prod_{i=1}^n \lambda_i^3, \quad (4)$$

where  $\lambda_i = \sqrt{\beta \hbar^2 / (2\pi m_i)}$  is the thermal de Broglie wavelength of the  $i$ th atom of the molecule and  $n$  is the total number of atoms in one molecule. The  $q$  term in Eq. (3) is related to the free energy in the vacuum,  $A_v$ , of a single molecule without constraints via

$$\begin{aligned} A_v &= -\beta^{-1} \ln \left[ \frac{1}{\Lambda^{3n}} \int e^{-\beta \tilde{U}(X)} dX^n \right] \\ &= -\beta^{-1} \ln \left[ \frac{V}{\Lambda^{3n}} \int e^{-\beta \tilde{U}(Y)} dY^{n-1} \right] = -\beta^{-1} \ln \left[ \frac{V}{\Lambda^{3n} q} \right]. \end{aligned} \quad (5)$$

In Eq. (5),  $\tilde{U}$  is the intramolecular interaction energy, which is invariant for translations. To go from the first to the second line of Eq. (5), we perform the change of variable,  $X \rightarrow Y$ , to the reference frame where the origin coincides with the position of one selected atom of the molecule. Integration over the dummy coordinates yields the  $V$  factor. In practice,  $q$  represents the configurational partition function of a solute molecule in a given position in space.

The free energy of the molecular crystal can be computed through a series of subsequent alchemical calculations. First, one calculates the free energy cost,  $\Delta A_0$ , of turning on additional harmonic restraints on selected atoms of each molecule so as to restrain also the orientation of the molecules and obtain a proper Einstein molecular crystal. After this, one computes the free energy cost,  $\Delta A_1$ , of turning on van der Waals (VDW) and electrostatic interactions. Finally, the free energy cost,  $\Delta A_2$ , of turning off all harmonic restraints is computed. The free energy of the crystal is determined by

$$A_{crys} = -\beta^{-1} N_{crys} \ln \left[ \frac{q}{\Lambda^{3n}} \right] + A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{symm}. \quad (6)$$

The last term in Eq. (6) accounts for the symmetry equivalent configurations that a molecule can assume, and it depends on the rotational symmetry number  $\sigma_{rot}$  of the molecular point group of the molecule,

$$\Delta A_{symm} = -\beta^{-1} N_{crys} \ln(\sigma_{rot}). \quad (7)$$

This contribution arises from the fact that a molecule in solution explores all the symmetrically equivalent configurations defined

by its molecular point group, while in the solid phase only one of these symmetrically equivalent configurations is accessible. To take this correctly into account, the correction term  $\Delta A_{symm}$  can be either subtracted from the free energy of the liquid phase or added to the crystal free energy, the latter of which we do here.

The chemical potential of the solute in solution is defined as the change in Gibbs free energy upon addition of a solute molecule (which we refer to below as the additional solute molecule),

$$\mu_{sol} = G(N+1, M, p, T) - G(N, M, p, T), \quad (8)$$

where  $M$  is the number of solvent molecules,  $N$  is the number of solute molecules,  $p$  is the pressure, and  $T$  is the temperature.

The correct expression for the Gibbs free energy of a solution of  $N$  solute molecules, with coordinates  $X_s$ , in  $M$  solvent molecules, with coordinates  $X_s$ , is

$$\begin{aligned} G(N, M) &= -\beta^{-1} \ln \left[ \frac{1}{\Lambda^{3N+n} \Lambda_s^{3M+m} N! M!} \int dV e^{-\beta p V} \right. \\ &\quad \times \left. \int dX_s^{N+n} dX_s^{M+m} e^{-\beta U(X_s, X_s, N, M)} \right] \\ &= -\beta^{-1} \ln[Q(N, M)], \end{aligned} \quad (9)$$

where  $m$  is the number of atoms in a solvent molecule and  $\Lambda_s^3$  is the equivalent, for a solvent molecule, of  $\Lambda^3$  defined in Eq. (4). The integral in Eq. (9) has the dimensions of a volume. It can be properly normalized by dividing by a chosen reference volume<sup>52</sup> or, equivalently, by fixing arbitrarily the units of measure. It must be noted that, in general, the potential energy of the whole system  $U(X_s, X_s, N, M)$  cannot be decoupled into two separated intramolecular and intermolecular contributions to be integrated over independently.

The chemical potential of the solute in solution can then be written as

$$\mu_{sol} = -\beta^{-1} \ln \left[ \frac{Q(N+1, M)}{Q(N, M)} \right] = -\beta^{-1} \ln \left[ \frac{Q(N+1, M)}{q Q(N, M)} \right] - \beta^{-1} \ln[q]. \quad (10)$$

In order to match the dimensionality of the numerator and the denominator in the first term of the right hand side, we can set the origin to be the coordinates of one of the atoms of the additional solute molecule and perform, in the integral in the numerator, the same change of variables used in Eq. (5),  $X_s, X_s \rightarrow Y_s, Y_s$ . We can then rewrite  $Q(N+1, M)$  as

$$\begin{aligned} Q(N+1, M) &= \frac{1}{\Lambda^{3(N+1)+n} \Lambda_s^{3M+m} (N+1)! M!} \int dV e^{-\beta p V} V \\ &\quad \times \int dY_s^{(N+1)+n-1} dY_s^{M+m} e^{-\beta U(Y_s, Y_s, N+1, M)}, \end{aligned} \quad (11)$$

allowing factorization of the volume  $V$ . The final expression for the chemical potential becomes

$$\begin{aligned}
\mu_{sol} &= -\beta^{-1} \ln \left[ \frac{1}{\Lambda^{3n}(N+1)} \frac{\int dV e^{-\beta p V} \int dY_S^{[(N+1)*n-1]} dY_S^{M*} e^{-\beta U(Y_S, Y_S, N+1, M)}}{q \int dV e^{-\beta p V} \int dX_S^{N*} dX_S^{M*} e^{-\beta U(X_S, X_S, N, M)}} \right] - \beta^{-1} \ln[q] \\
&= -\beta^{-1} \ln \left[ \frac{1}{\Lambda^{3n}(N+1)} \frac{\int dV e^{-\beta p V} \int dY_S^{[(N+1)*n-1]} dY_S^{M*} e^{-\beta U(Y_S, Y_S, N+1, M)}}{\int e^{-\beta \tilde{U}} dY^{n-1} \int dV e^{-\beta p V} \int dX_S^{N*} dX_S^{M*} e^{-\beta U(X_S, X_S, N, M)}} \right] - \beta^{-1} \ln[q] \\
&= -\beta^{-1} \ln \left[ \frac{1}{\Lambda^{3n}(N+1)} \frac{\int dV e^{-\beta p V} \int dY_S^{[(N+1)*n-1]} dY_S^{M*} e^{-\beta U(Y_S, Y_S, N+1, M)}}{\int dV e^{-\beta p V} \int dY_S^{[(N+1)*n-1]} dY_S^{M*} e^{-\beta [U(Y_S, Y_S, N, M) + \tilde{U}]}} \right] - \beta^{-1} \ln[q]. \quad (12)
\end{aligned}$$

To go from the second to the third line of Eq. (12), we made the reasonable assumption that the dependence of the free energy of a single molecule in the vacuum on the volume of the system is negligible. If  $\tilde{U}_E$  is the energy of interaction between the additional solute molecule and the rest of the solution, we have  $U(X_S, X_S, N+1, M) = U(X_S, X_S, N, M) + \tilde{U}_E + \tilde{U}$ . Assuming as well that the volume  $V$  of the system is negligibly correlated with  $\tilde{U}_E$ , Eq. (12) can be rewritten as

$$\begin{aligned}
\mu_{sol} &= -\beta^{-1} \ln \left[ \frac{\langle V \rangle}{\Lambda^{3n}(N+1)} \langle e^{-\beta \tilde{U}_E} \rangle_0 \right] - \beta^{-1} \ln[q] \\
&= \beta^{-1} \ln \rho_S - \beta^{-1} \ln \left[ \langle e^{-\beta \tilde{U}_E} \rangle_0 \right] - \beta^{-1} \ln \left[ \frac{q}{\Lambda^{3n}} \right] \\
&= \beta^{-1} \ln \rho_S + \Delta G_{solv} - \beta^{-1} \ln \left[ \frac{q}{\Lambda^{3n}} \right], \quad (13)
\end{aligned}$$

where  $\rho_S = N/V \approx (N+1)/V$  is the number density of the solute in the solution. The average  $\langle e^{-\beta \tilde{U}_E} \rangle_0$  is computed in the ensemble where the additional solute molecule does not interact with the other molecules, and it is at a given position in space. This represents the free energy cost of turning on the interaction between the additional solute molecule and the solution; i.e., it is the free energy of solvation of the solute molecule,  $\Delta G_{solv}$ . The calculation of this term can be performed using standard alchemical free energy techniques, as we discuss in Sec. II B.

If we equate the expressions for the two chemical potentials, we obtain

$$\begin{aligned}
\beta^{-1} \ln \rho_S + \Delta G_{solv} - \beta^{-1} \ln \left[ \frac{q}{\Lambda^{3n}} \right] \\
= -\beta^{-1} \ln \left[ \frac{q}{\Lambda^{3n}} \right] + \frac{A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{symm} + pV}{N_{crys}}. \quad (14)
\end{aligned}$$

It must be noted that the argument of the logarithm in the term  $-\beta^{-1} \ln[q/\Lambda^{3n}]$  is not dimensionless. Hence, simplifying it on both sides comes at the cost of being aware that the units of measure of the number density,  $\rho_S$ , and those of the spring constant of the harmonic restraints,  $K_E$ , involved in the definition of  $A_0$ , must be consistent. After performing this simplification, we finally obtain

$$\begin{aligned}
\beta^{-1} \ln \rho_S + \Delta G_{solv} &= \frac{A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{symm} + pV}{N_{crys}} \\
&= \frac{A_{crys}^{ex} + pV}{N_{crys}} = \frac{G_{crys}^{ex}}{N_{crys}}, \quad (15)
\end{aligned}$$

where  $A_{crys}^{ex}$  and  $G_{crys}^{ex}$  are, respectively, the Helmholtz and Gibbs excess free energy of the crystal. In the case of a sparingly soluble solute, the estimate for the solubility can be obtained by calculating the free energy of solvation of one solute molecule in pure solvent and by inverting Eq. (15) and solving for the number density  $\rho_S$ . When the solute is not sparingly soluble, as is the case with paracetamol in ethanol, one needs to carry out multiple free energy of solvation calculations at different concentrations and then determine through an interpolation the value of  $\rho_S$  for which the identity with the right hand side of Eq. (15) is satisfied.

## B. Simulation details

For all calculations, the CHARMM36 CGenFF<sup>53</sup> force field was used for ethanol, while the reoptimized CGenFF introduced in Ref. 45 was used for paracetamol, and the Lorentz-Berthelot<sup>54,55</sup> combining rules were used for interatomic mixing of Lennard-Jones parameters. The CGenFF of paracetamol was optimized using a Monte Carlo procedure that made small random changes to the partial charges of the paracetamol molecule until the force field reproduced the correct lattice energy difference of the form I and form II polymorphs of paracetamol; the experimental values of the heat of sublimation, heat of vaporization, heat of fusion; and the solvation free energy of a gas-phase molecule going into pure ethanol. The optimized partial charges of paracetamol and the computed values of the thermodynamic properties can be found in the supporting information of Ref. 45. All MD simulations were run using the Gromacs 5.1.4 software package,<sup>56</sup> and restraints were enforced using the PLUMED 2 plugin.<sup>57</sup> Short-range interactions were truncated at 15 Å, and long-range electrostatic interactions were computed using the particle mesh Ewald summation of order 4 with a Fourier grid spacing of 1.2 Å and a tolerance of  $10^{-5}$ . In addition, we utilized long-range dispersion corrections for the energy and pressure. The equations of motion were integrated with the velocity Verlet algorithm with a time step of 0.5 fs, and rigid bond constraints were enforced with the LINCS<sup>58</sup> algorithm. For simulations in the isobaric-isothermal ensemble, conditions were 298 K and 1 bar, and a Nosé-Hoover<sup>59</sup> thermostat and a Parrinello-Rahman<sup>60</sup> barostat were used to control the temperature and pressure. For simulations in the canonical ensemble, all simulations were carried out at 298 K, and a Nosé-Hoover thermostat was used to control the temperature.

The most stable crystal form of paracetamol at ambient conditions, and the form present in pharmaceutical formulations, is form I (this is also the form that homogeneously nucleates in



supersaturated ethanol solutions).<sup>61</sup> To compute the Helmholtz free energy,  $A_{\text{cryst}}$ , of the solid, a form I crystal of 240 molecules was initially equilibrated in the isothermal-isobaric ensemble for 5 ns. The average lattice vectors were then computed over 20 ns of trajectory, and the computed average volume  $V_{\text{cryst}}$  was used in all subsequent NVT simulations. The free energy differences,  $\Delta A_0$  and  $\Delta A_2$ , related to turning on and off the lattice restraints were computed with thermodynamic integration (TI). The free energy cost of turning on the restraints can be written as

$$\Delta A_0 = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda = \int_0^1 d\lambda \left\langle \sum_i \frac{1}{2} K_E (r_i - r_{0,i})^2 \right\rangle_\lambda. \quad (16)$$

In Eq. (16),  $\partial U/\partial \lambda$  is the derivative of the energy with respect to the TI parameter,  $\lambda$ . Since we used a simple linear coupling scheme, the derivative coincides with the potential energy of the interaction that is being gradually turned on, i.e., that of the restraints. The index  $i$  in the sum runs over all atoms involved in the definition of the orientational restraints, and  $r_{0,i}$  is the position defining the restraint acting on the  $i$ th atom. To perform the calculations, the interval between 0 and 1 was discretized in 61 windows. To improve the precision, 51 windows spaced by a distance of 0.02 between 0 and 1 were complemented with other additional 10 placed at 0.00005, 0.0001, 0.0002, 0.0004, 0.0008, 0.0016, 0.0032, 0.0064, 0.01, 0.015. For each window, the system was first minimized using conjugate gradients followed by a 1 ns equilibration, and the averages used in the TI were calculated over a 6 ns production run. The central and orientational atoms used to fix the molecular lattice positions and orientations are shown in Fig. 1. The value of the harmonic force constant used to restrain the central and orientational atoms was  $K_E = 10^5$  (kJ/mol)/nm<sup>2</sup>. Finally, the value of the thermodynamic integral is computed by summing the contribution from every window using the Simpson integration rule.

The  $\Delta A_1$  term that accounts for the cost of turning on intermolecular interactions was computed using the Free Energy Perturbation (FEP) implementation available in GROMACS. In FEP calculations, the free energy difference between a non-interacting state, indicated by  $\lambda = 0$ , and an interacting one, characterized by  $\lambda = 1$ , is computed by introducing a series of intermediate thermodynamic states interpolating between the two. Each thermodynamic state is characterized by a particular value of the parameter  $\lambda$  which governs the interpolation. Hence the free energy difference can be

expressed as

$$\Delta A_1 \propto \ln \left[ \frac{Q|_{\lambda=1}}{Q|_{\lambda=0}} \right] = \sum_{\lambda=0}^{1-\Delta\lambda} \ln \left[ \frac{Q|_{\lambda+\Delta\lambda}}{Q|_{\lambda}} \right] = \sum_{\lambda=0}^{1-\Delta\lambda} \ln \left[ \langle e^{-\beta \Delta U} \rangle_\lambda \right], \quad (17)$$

where  $Q|_\lambda$  is the partition function for the thermodynamic state characterized by a specific value of  $\lambda$  and  $\Delta U$  is the difference between the energy of a given configuration of the system evaluated at  $\lambda + \Delta\lambda$  and at  $\lambda$ . Provided that the spacing between intermediate states is small enough and that the interpolation is performed using mollified soft core potentials to avoid end-point singularities, the average values in the last term of Eq. (17) can be computed without numerical issues.

For each FEP calculation, each of the interaction energies was split into an electrostatic and VDW contribution, and coupling parameters,  $\lambda_e, \lambda_{\text{VDW}} \in [0, 1]$ , were used to scale each of the electrostatic and VDW interactions during the alchemical transformations. The coupling parameter was subdivided into 51 evenly spaced windows for each interaction energy. When coupling the intermolecular interactions, the calculations were carried out such that alchemical transformations involving the VDW potential were performed before the alchemical transformations involving the electrostatics. Performing the alchemical transformations in this way ensures that the interaction energy is always bounded from below and prevents the system from becoming unstable, which could lead to inaccurate sampling and therein inaccurate free energies. For each simulation for a given window, the system was first minimized using conjugate gradients, followed by a 1 ns equilibration, and the averages used in the FEP were calculated over a 6 ns production run. The free energy differences were calculated by processing the results of the simulations with the Bennett acceptance ratio (BAR) protocol.<sup>62</sup>

The solvation free energy,  $\Delta G_{\text{solv}}$ , was computed over a range of solution compositions  $X = 0.0008, 0.0292, 0.0645, 0.1182$ , where  $X$  is the mole ratio (ratio between moles of solute and moles of solvent). The number of solute and solvent molecules was varied such that the volume of the system remained approximately constant, and in particular, 1, 25, 50, and 80 paracetamol molecules were dissolved into 1286, 857, 775, and 677 molecules of ethanol, respectively. Each individual free energy calculation was performed using FEP, and all simulations were performed with the same procedure that was used in the calculation of  $\Delta A_1$  described above. For each FEP calculation, the intermolecular interaction energy of the additional paracetamol molecule was coupled to the rest of the solution, and the central atom of this molecule (see Fig. 1) was fixed at the center of the simulation cell. After determining the solvation free energy for each composition considered, a linear fit on the values of the chemical potential in solution was performed, and the solubility was determined from the fit as the composition of the system that satisfied Eq. (15). The validity of the linear approximation as model for the solvation free energy data is discussed below.

### C. Experimental details

Paracetamol was purchased from Sigma-Aldrich (St. Louis, MO) and used as received. Ethanol (absolute, 200 proof) was purchased from VWR International (Edison, NJ) and filtered through a polytetrafluoroethylene (PTFE) membrane with 0.2  $\mu\text{m}$  pores prior to use in experiments. 2 ml clear, screw thread vials were purchased

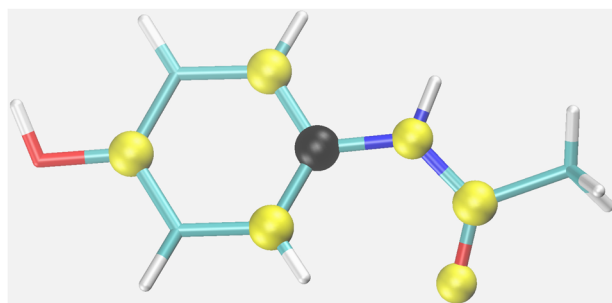
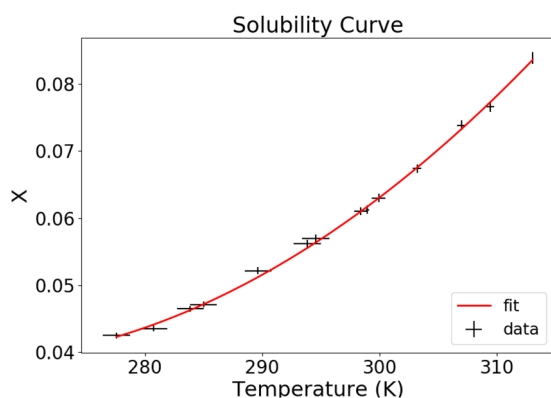


FIG. 1. The central atom (black) and orientational atoms (yellow) used in the Einstein crystal to fix the molecular lattice positions and orientations.

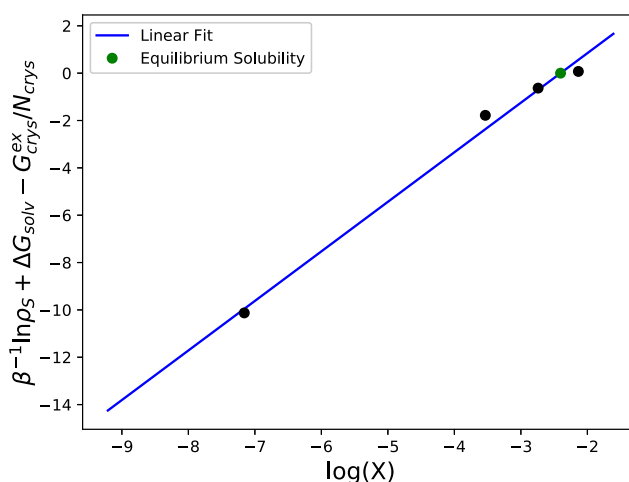


**FIG. 2.** Experimental solubility curve for paracetamol in ethanol determined from our measurements. Solubility is expressed in mole ratio. Data are shown as black crosses with experimental uncertainties. The curve fit is shown as a red solid line.

from Fisher Scientific (Pittsburgh, PA). The solubility of paracetamol in ethanol was determined by turbidity measurements using a Crystal16 (Avantium Technologies, The Netherlands) in the temperature range 275–315 K. For this, 16 vials with 100–220 mg of paracetamol in 1 ml of ethanol were prepared. The vials were subjected to three cycles of heating and cooling with a stirring speed of 700 rpm. The data were processed with CrystalClear (Avantium Technologies, The Netherlands) software, and the average solubility temperature for each concentration was obtained.

### III. RESULTS AND DISCUSSION

The experimental solubility curve of paracetamol in ethanol, as a function of temperature, determined from this study is shown in Fig. 2. The solubility in mole ratio ( $X$ ) at 298 K from this study and other studies in the literature is reported in Table I. The variability of values reported in the literature may be ascribed to the purity of the substances used and the specific experimental technique adopted for each experiment.<sup>61</sup> Our experimental results are in good agreement with the other solubility measurements taken from the literature. In the last column of Table I, the average experimental solubility,  $\langle X \rangle_{\text{exp}}$ , is reported. The uncertainty on this quantity has been estimated from the variance of the experimental values.



**FIG. 3.** Linear fit of the difference between the left and the right hand sides of Eq. (15). We note that in Eq. (15),  $\beta^{-1} \ln \rho_S + \Delta G_{\text{solv}}$  is a function of the number density of solute molecules in the solution. However, the latter can be considered as a deterministic function of the mole ratio and we do the fit and solve directly for this quantity. Data points are shown in black, while the blue line represents the linear fit. The green dot corresponds to the determined equilibrium solubility.

In Fig. 3, the difference between the quantities  $\beta^{-1} \ln \rho_S + \Delta G_{\text{solv}}$  and  $G_{\text{crys}}^{\text{ex}}/N_{\text{crys}}$  from Eq. (15) is plotted as a function of the logarithm of the mole ratio concentration,  $X$ , and the linear fit to the data is also shown. The green point on the plot in Fig. 3 represents the equilibrium solubility determined through Eq. (15), and the precise value is shown in Table II along with the solid form free energy calculated from the EECM.

The value of the solubility calculated from simulations ( $X_{\text{calc}} = 0.085 \pm 0.014$ ) and the mean experimental solubility ( $\langle X \rangle_{\text{exp}} = 0.0585 \pm 0.004$ ) differ roughly by a factor 1.5. This is a more than reasonable result considering the uncertainties and typical agreement reported in the literature,<sup>66</sup> where deviations around a factor 3 are not unusual, as well as the simplicity of the force field model and the fact that the solubility value itself was not directly used in the force field reoptimization procedure.

It is known in the literature<sup>22,67</sup> that a weak numerical divergence may arise in the calculations involved in estimating the free energy of a crystal with the Einstein crystal method; however,

**TABLE I.** Experimental mole ratio solubilities for paracetamol in ethanol at 298 K. The last column reports the averaged value.

	exp <sup>a</sup>	exp <sup>b</sup>	exp <sup>c</sup>	exp <sup>d</sup>	exp <sup>e</sup>	$\langle X \rangle_{\text{exp}}$
Solubility( $X$ )	0.0542	0.0546	0.0595	0.0605	0.0639	$0.0585 \pm 0.004$

<sup>a</sup>Reference 63.

<sup>b</sup>Reference 64.

<sup>c</sup>Reference 65.

<sup>d</sup>This study.

<sup>e</sup>Reference 61.



**TABLE II.** The upper part of the table shows the crystal Helmholtz and Gibbs free energies, the value of the free energy of solvation corresponding to the green dot in Fig. 3, and the estimated value of the solubility. Free energies are reported in kJ/mol, while the solubility is reported as mole ratio. The lower part of the table compares the mole ratio and the number density at the concentrations described in the text, as well as the corresponding solvation free energies. The last row shows the numeric value of the quantity fitted in Fig. 3,  $\beta^{-1} \log(\rho) + \Delta G_{\text{solv}}(\rho) - G_{\text{crys}}^{\text{ex}}/N_{\text{crys}}$ .

$A_{\text{crys}}^{\text{ex}}/N_{\text{crys}}$	$pV_{\text{crys}}/N_{\text{crys}}$	$G_{\text{crys}}^{\text{ex}}/N_{\text{crys}}$	$\Delta G_{\text{solv}}$	$X_{\text{calc}}$
$-75.68 \pm 0.03$	0.01	$-75.67 \pm 0.03$	$-57.98$	$0.085 \pm 0.014$
<b>Mole ratio</b>	0.0008	0.0292	0.0645	0.1182
$\beta^{-1} \log(\rho)$	$-29.03$	$-20.2$	$-18.4$	$-17.13$
$\Delta G_{\text{solv}}(\rho)$	$-56.77 \pm 0.06$	$-57.26 \pm 0.11$	$-57.91 \pm 0.13$	$-58.47 \pm 0.15$
$\beta^{-1} \log(\rho) + \Delta G_{\text{solv}}(\rho) - G_{\text{crys}}^{\text{ex}}/N_{\text{crys}}$	$-10.13 \pm 0.07$	$-1.78 \pm 0.12$	$-0.63 \pm 0.13$	$0.07 \pm 0.16$

it is argued that, in the case of molecular crystals, the error induced should be negligible with respect to the expected uncertainties of the calculations due to the precision associated with the force field parameterization.<sup>43</sup> In principle, this issue can be prevented using a reference state for the alchemical transformations, that is, an Einstein crystal with a fixed center of mass (COM) and performing all the alchemical transformations with a fixed COM. The free energy of the unconstrained physical crystal can then be obtained by adding a correction that can be computed analytically. In the Appendix, we derive the expressions necessary to compute the free energy of the crystal using this procedure. We have repeated the calculation of the free energy of the unconstrained crystal according to this procedure and, repeating the fitting procedure performed above, we obtain for the solubility,  $X_{\text{calc}} = 0.094 \pm 0.018$ . This result is compatible, within the error, with the value obtained from the simulations without constraints on the COM. This validates the assumption that, in the case of a molecular crystal, the error caused by neglecting this potential issue is expected to be small.

#### IV. CONCLUSIONS

Solubility is a property that is important across a variety of different fields, particularly in the pharmaceutical field. Despite this, there are not many methods available to compute the solubility of pharmaceutically relevant molecules from first principles simulations. Beyond computational methods, useful applications also require one to have reliable molecular models. Indeed, many force fields typically do a poor job at reproducing experimental solubilities. In this work, we successfully address both of these issues. (i) We have further explored and clarified the recently proposed EECM in order to use it to calculate the solubility of paracetamol in ethanol at room temperature; indeed, we have found that the calculated solubility ( $X_{\text{calc}} = 0.085 \pm 0.014$ ) is in reasonable agreement with the experimental measurements performed by us and to those found in the literature ( $\langle X \rangle_{\text{exp}} = 0.0585 \pm 0.004$ ). (ii) We have shown that the force field reparameterization procedure presented in Ref. 45 can potentially be used as a general and relatively simple method to reoptimize existing force fields for use in studies of solubility and nucleation from solution. The approach presented in this work may lay the foundation for computing reliable solubilities for molecular simulations of pharmaceutical interest.

#### ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the MIT-Novartis Center for Continuous Manufacturing.

#### APPENDIX: FREE ENERGY OF THE CRYSTAL FROM A FIXED CENTER OF MASS REFERENCE STATE

We detail here the procedure to compute the free energy of the crystal when performing all necessary simulations with a fixed COM.<sup>67</sup> In this case, the starting point for the alchemical transformations is an Einstein crystal with a fixed COM. We start by deriving the analytical expression for the free energy of such a system. We consider the case of an atomic Einstein crystal, but the calculation directly generalizes to the molecular case. Moreover, we only need to consider the contribution to the free energy coming from configuration space, neglecting momenta since their contribution to the free energy is computed analytically in Eq. (3).

In simulations, the crystal system is typically set in an initial condition where the position of every atom,  $\tilde{r}_i$ , coincides with the position of the corresponding ideal center,  $r_{0,i}$ . If this condition is satisfied, the COM of the system coincides with the COM of the ideal centers, and if the COM is then held fixed, this condition is conserved thereafter. Hence, in the configurational part of the partition function, the constraint on the COM can be expressed as a  $\delta$  function depending on the position of the ideal centers. The partition function takes the form

$$Z_{\text{EI}}^{\text{CM}} = \int d^d N \tilde{r} \prod_{i=1}^N \exp(-(\beta K_E (\tilde{r}_i - r_{0,i})^2 / 2)) \delta\left(\sum_{i=1}^N v_i (\tilde{r}_i - r_{0,i})\right), \quad (\text{A1})$$

where  $v_i = m_i / \sum_{j=1}^N m_j$  is the relative mass of the  $i$ th particle. We note that here we treat explicitly the dimensionality  $d$  of the physical space as a parameter, while in the calculations detailed in Sec. II A the dimensionality of an atomic position was implicitly assumed to be equal to 3. If we apply the change of variables  $r_i = v_i (\tilde{r}_i - r_{0,i})$ , we obtain

$$\prod_{i=1}^N \frac{1}{v_i^d} \int d^d N r \prod_{i=1}^N \exp(-(\beta K_E (r_i)^2 / 2 v_i^2)) \delta\left(\sum_{i=1}^N r_i\right), \quad (\text{A2})$$

where the prefactor in front of the integral comes from the determinant of the transformation. Integrating over  $r_1$ , we get

$$\prod_{i=1}^N \frac{1}{v_i^d} \int d^{d(N-1)} r \exp \left[ - \left( \beta K_E \left( \sum_{j=2}^N r_j \right)^2 / 2v_1^2 \right) \right] \times \prod_{i=2}^N \exp \left( - (\beta K_E (r_i)^2 / 2v_i^2) \right). \quad (\text{A3})$$

The integral in Eq. (A3) can be factorized into  $d$  independent and identical gaussian integrals. The inverse of the  $(N-1)$ -dimensional covariance matrix for each of these gaussians has the form

$$\Gamma_{ij}^{-1} = \frac{\beta K_E}{\prod_{k=1}^N v_k^2} \times \begin{cases} (v_1^2 + v_{i+1}^2) \prod_{k=2, k \neq i+1}^N v_k^2 & \text{if } i = j \\ \prod_{k=2}^N v_k^2 & \text{if } i \neq j, \end{cases} \quad (\text{A4})$$

and it can be verified inductively that

$$\det \Gamma^{-1} = \frac{(\beta K_E)^{N-1} \sum_{i=1}^N v_i^2}{\prod_{i=1}^N v_i^2}. \quad (\text{A5})$$

The configurational partition function for the Einstein crystal with fixed COM then becomes

$$Z_{EI}^{CM} = \left( \prod_{i=1}^N \frac{1}{v_i} \right)^d \left( \frac{\beta K_E}{2\pi \sum_{i=1}^N v_i^2} \right)^{d/2} \prod_{i=1}^N \left( \frac{2\pi v_i^2}{\beta K_E} \right)^{d/2} = \left( \frac{\beta K_E}{2\pi \sum_{i=1}^N v_i^2} \right)^{d/2} \prod_{i=1}^N \left( \frac{2\pi}{\beta K_E} \right)^{d/2}, \quad (\text{A6})$$

which differs from the positional partition function of the unconstrained Einstein crystal,  $Z_{EI} = \prod_{i=1}^N (2\pi/\beta K_E)^{d/2}$ , only for the  $(\beta K_E/2\pi \sum_{i=1}^N v_i^2)^{d/2}$  prefactor [see Eq. (3)], and the following relation holds:

$$Z_{EI}^{CM} = \left( \frac{\beta K_E}{2\pi \sum_{i=1}^N v_i^2} \right)^{d/2} Z_{EI}. \quad (\text{A7})$$

The free energy of the physical crystal with fixed COM is then calculated as

$$A_{crys}^{CM} = A_{EI}^{CM} + \Delta A_0^{CM} + \Delta A_1^{CM} + \Delta A_2^{CM}, \quad (\text{A8})$$

where  $A_{EI}^{CM} = -\beta^{-1} \ln(Z_{EI}^{CM})$ , and the various  $\Delta A_i^{CM}$  terms corresponds to the alchemical transformations discussed in the text and are computed in the same way, except for the presence of the constraint on the COM (we also note that, as discussed in the main text, in the case of a molecular crystal, a  $\Delta A_{Symm}$  term properly accounting for the symmetries of the molecule must be added). If we denote  $\Delta A_{crys}$  as the free energy difference between the unconstrained physical crystal and the crystal with fixed COM, we get

$$A_{crys} = A_{crys}^{CM} + \Delta A_{crys}^{CM} = A_{EI}^{CM} + \Delta A_0^{CM} + \Delta A_1^{CM} + \Delta A_2^{CM} + \Delta A_{crys}. \quad (\text{A9})$$

This term can be computed analytically following the argument presented in Ref. 67. The ratio between the configurational partition function,  $Z_{crys}^{CM}$ , of a physical crystal with a constrained COM and that of a free crystal,  $Z_{crys}$ , represents the probability for the COM of the free crystal to be in a specific point in space. For a crystal without any constraint and subject to periodic boundary conditions, the probability distribution of the position of the COM is uniformly distributed in a volume equal to that of a Wigner-Seitz cell of the crystal and centered in the middle of the simulation box.

Hence, the probability for the COM to be in a specific point in space is the inverse of the volume of the Wigner-Seitz cell. If the number of molecules per Wigner-Seitz cell is one, the volume of the cell coincides with the ratio between the total volume available to the system and the total number of molecules,  $V/N$ . Hence, we have

$$A_{crys} = -\beta^{-1} \ln(Z_{crys}) = -\beta^{-1} \ln(Z_{crys}^{CM}) + \beta^{-1} \ln(N/V) \quad (\text{A10})$$

and  $\Delta A_{crys} = \beta^{-1} \ln(N/V)$ .

Combining Eqs. (A8) and (A10), we get

$$\frac{A_{crys}}{N} = \frac{A_{EI}^{CM} + \Delta A_0^{CM} + \Delta A_1^{CM} + \Delta A_2^{CM} + \Delta A_{crys}}{N} = -\beta^{-1} \frac{d}{2} \ln \left( \frac{2\pi}{\beta K_E} \right) - \beta^{-1} \frac{d}{2N} \ln \left( \frac{\beta K_E}{2\pi \sum_{i=1}^N v_i^2} \right) + \frac{\Delta A_0^{CM} + \Delta A_1^{CM} + \Delta A_2^{CM}}{N} + \beta^{-1} \frac{\ln(N/V)}{N}. \quad (\text{A11})$$

In the case of a single atomic or molecular species, as is the case for the paracetamol crystal we consider in this paper,  $\sum_{i=1}^N v_i^2 = 1/N$ , and the second term in the second line of the summation above reduces to

$$-\beta^{-1} \frac{d}{2N} \ln \left( \frac{\beta K_E}{2\pi} \right) - \beta^{-1} \frac{d}{2N} \ln(N). \quad (\text{A12})$$

If we compute the free energy of the paracetamol crystal following the procedure just detailed, we get a value of the free energy per molecule of  $-75.49$  kJ/mol which is about 0.3% higher than the estimate of  $-75.68$  kJ/mol obtained from the simulations without the COM constraint, confirming that the overall effect is small.

## REFERENCES

- G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, *Pharm. Res.* **12**, 413 (1995).
- K. T. Savjani, A. K. Gajjar, and J. K. Savjani, *ISRN Pharm.* **2012**, 195727.
- H. D. Williams, N. L. Trevaskis, S. A. Charman, R. M. Shanker, W. N. Charman, C. W. Pouton, and C. J. Porter, *Pharmacol. Rev.* **65**, 315 (2013).
- J. Chen, B. Sarma, J. M. Evans, and A. S. Myerson, *Cryst. Growth Des.* **11**, 887 (2011).
- A. Myerson, *Handbook of Industrial Crystallization* (Butterworth-Heinemann, 2002).
- N. Jain and S. H. Yalkowsky, *J. Pharm. Sci.* **90**, 234 (2001).
- S. Ruppert, S. Sandler, and A. Lenhoff, *Biotechnol. Prog.* **17**, 182 (2001).
- C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *Adv. Drug Delivery Rev.* **46**, 3 (2001).
- A. Jouyban, H.-K. Chan, and N. R. Foster, *J. Supercrit. Fluids* **24**, 19 (2002).
- A. Jouyban, *J. Pharm. Pharm. Sci.* **11**, 32 (2008).
- F. L. Nordström and Å. C. Rasmuson, *Eur. J. Pharm. Sci.* **36**, 330 (2009).
- S. Cabani, P. Gianni, V. Mollica, and L. Lepori, *J. Solution Chem.* **10**, 563 (1981).
- W. L. Jorgensen and E. M. Duffy, *Adv. Drug Delivery Rev.* **54**, 355 (2002).
- M. J. Lazzaroni, D. Bush, C. A. Eckert, T. C. Frank, S. Gupta, and J. D. Olson, *Ind. Eng. Chem. Res.* **44**, 4075 (2005).
- A. S. Paluch and E. J. Maginn, *AIChE J.* **59**, 2647 (2013).
- R. T. Ley, G. B. Fuerst, B. N. Redeker, and A. S. Paluch, *Ind. Eng. Chem. Res.* **55**, 5415 (2016).

- <sup>17</sup>J. Bajorath, *Nat. Rev. Drug Discovery* **1**, 882 (2002).
- <sup>18</sup>A. R. Katritzky, A. A. Oliferenko, P. V. Oliferenko, R. Petrukhin, D. B. Tatham, U. Maran, A. Lomaka, and W. E. Acree, *J. Chem. Inf. Comput. Sci.* **43**, 1794 (2003).
- <sup>19</sup>J. C. Dearden, *Expert Opin. Drug Discovery* **1**, 31 (2006).
- <sup>20</sup>K. V. Balakin, N. P. Savchuk, and I. V. Tetko, *Curr. Med. Chem.* **13**, 223 (2006).
- <sup>21</sup>A. J. Hopfinger, E. X. Esposito, A. Llinàs, R. C. Glen, and J. M. Goodman, *J. Chem. Inf. Model.* **49**, 1 (2009).
- <sup>22</sup>D. Frenkel and A. J. Ladd, *J. Chem. Phys.* **81**, 3188 (1984).
- <sup>23</sup>M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailier, and P. Turq, *J. Chem. Phys.* **117**, 4947 (2002).
- <sup>24</sup>K. Raju and G. Atkinson, *J. Chem. Eng. Data* **33**, 490 (1988).
- <sup>25</sup>K. U. Raju and G. Atkinson, *J. Chem. Eng. Data* **35**, 361 (1990).
- <sup>26</sup>T. Chen, A. Neville, and M. Yuan, *J. Pet. Sci. Eng.* **46**, 185 (2005).
- <sup>27</sup>J. Kolafa, *J. Chem. Phys.* **145**, 204509 (2016).
- <sup>28</sup>M. Lísál, W. R. Smith, and J. Kolafa, *J. Phys. Chem. B* **109**, 12956 (2005).
- <sup>29</sup>F. Moučka, M. Lísál, and W. R. Smith, *J. Phys. Chem. B* **116**, 5468 (2012).
- <sup>30</sup>F. Moučka, I. Nezbeda, and W. R. Smith, *J. Chem. Theory Comput.* **11**, 1756 (2015).
- <sup>31</sup>I. Nezbeda, F. Moučka, and W. R. Smith, *Mol. Phys.* **114**, 1665 (2016).
- <sup>32</sup>A. Benavides, J. Aragones, and C. Vega, *J. Chem. Phys.* **144**, 124504 (2016).
- <sup>33</sup>A. S. Paluch, S. Jayaraman, J. K. Shah, and E. J. Maginn, *J. Chem. Phys.* **133**, 124504 (2010).
- <sup>34</sup>Z. Mester and A. Z. Panagiotopoulos, *J. Chem. Phys.* **142**, 044507 (2015).
- <sup>35</sup>Z. Mester and A. Z. Panagiotopoulos, *J. Chem. Phys.* **143**, 044505 (2015).
- <sup>36</sup>J. Espinosa, J. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. Debenedetti, and A. Panagiotopoulos, *J. Chem. Phys.* **145**, 154111 (2016).
- <sup>37</sup>H. M. Manzanilla-Granados, H. Saint-Martin, R. Fuentes-Azcatl, and J. Alejandro, *J. Phys. Chem. B* **119**, 8389 (2015).
- <sup>38</sup>J. Aragones, E. Sanz, and C. Vega, *J. Chem. Phys.* **136**, 244508 (2012).
- <sup>39</sup>M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren, *J. Chem. Theory Comput.* **8**, 1721 (2012).
- <sup>40</sup>D. S. Palmer, A. Llinàs, I. Morao, G. M. Day, J. M. Goodman, R. C. Glen, and J. B. Mitchell, *Mol. Pharmaceutics* **5**, 266 (2008).
- <sup>41</sup>D. S. Palmer, J. L. McDonagh, J. B. Mitchell, T. van Mourik, and M. V. Fedorov, *J. Chem. Theory Comput.* **8**, 3322 (2012).
- <sup>42</sup>C. Vega, E. Sanz, J. Abascal, and E. Noya, *J. Phys.: Condens. Matter* **20**, 153101 (2008).
- <sup>43</sup>L. Li, T. Totton, and D. Frenkel, *J. Chem. Phys.* **146**, 214110 (2017).
- <sup>44</sup>G. Gobbo, M. A. Bellucci, G. A. Tribello, G. Ciccotti, and B. L. Trout, *J. Chem. Theory Comput.* **14**, 959 (2018).
- <sup>45</sup>J. Stojaković, F. Baftizadeh, M. A. Bellucci, A. S. Myerson, and B. L. Trout, *Cryst. Growth Des.* **17**, 2955 (2017).
- <sup>46</sup>Y. Diao, A. S. Myerson, T. A. Hatton, and B. L. Trout, *Langmuir* **27**, 5324 (2011).
- <sup>47</sup>A. G. Shtukenberg, S. S. Lee, B. Kahr, and M. D. Ward, *Annu. Rev. Chem. Biomol. Eng.* **5**, 77 (2014).
- <sup>48</sup>L. Tan, R. M. Davis, A. S. Myerson, and B. L. Trout, *Cryst. Growth Des.* **15**, 2176 (2015).
- <sup>49</sup>M. D. Ward, *ACS Nano* **10**, 6424 (2016).
- <sup>50</sup>D. S. Frank and A. J. Matzger, *Cryst. Growth Des.* **17**, 4056 (2017).
- <sup>51</sup>T. K. Wijethunga, F. Baftizadeh, J. Stojaković, A. S. Myerson, and B. L. Trout, *Cryst. Growth Des.* **17**, 3783 (2017).
- <sup>52</sup>M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, 2010).
- <sup>53</sup>K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov *et al.*, *J. Comput. Chem.* **31**, 671 (2010).
- <sup>54</sup>H. Lorentz, *Ann. Phys.* **248**, 127 (1881).
- <sup>55</sup>D. Berthelot, C. R. Hebd. Seances Acad. Sci. **126**, 1703 (1898).
- <sup>56</sup>M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *SoftwareX* **1**, 19 (2015).
- <sup>57</sup>G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Comput. Phys. Commun.* **185**, 604 (2014).
- <sup>58</sup>B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, *J. Comput. Chem.* **18**, 1463 (1997).
- <sup>59</sup>S. Nosé, *J. Chem. Phys.* **81**, 511 (1984).
- <sup>60</sup>M. Parrinello and A. Rahman, *J. Appl. Phys.* **52**, 7182 (1981).
- <sup>61</sup>R. A. Granberg and Å. C. Rasmuson, *J. Chem. Eng. Data* **44**, 1391 (1999).
- <sup>62</sup>C. H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).
- <sup>63</sup>G. L. Perlovich, T. V. Volkova, and A. Bauer-Brandl, *J. Pharm. Sci.* **95**, 2158 (2006).
- <sup>64</sup>J. A. Jiménez and F. Martínez, *J. Solution Chem.* **35**, 335 (2006).
- <sup>65</sup>N. A. Mitchell and P. J. Frawley, *J. Cryst. Growth* **312**, 2740 (2010).
- <sup>66</sup>L. Li, T. Totton, and D. Frenkel, *J. Chem. Phys.* **149**, 054102 (2018).
- <sup>67</sup>J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, *J. Chem. Phys.* **112**, 5339 (2000).