

Solubility Prediction of Organic Molecules with Molecular Dynamics Simulations

Zoran Bjelobrk, Dan Mendels, Tarak Karmakar, Michele Parrinello,* and Marco Mazzotti*

Cite This: *Cryst. Growth Des.* 2021, 21, 5198–5205

Read Online

ACCESS |



Metrics & More

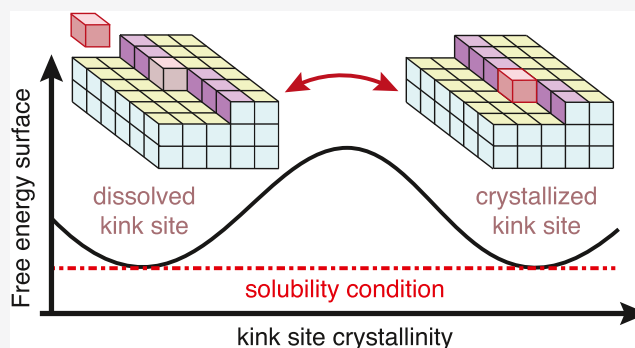


Article Recommendations



Supporting Information

ABSTRACT: We present a molecular dynamics simulation method for the computation of the solubility of organic crystals in solution. The solubility is calculated based on the equilibrium free energy difference between the solvated solute and its crystallized state at the crystal surface kink site. To efficiently sample the growth and dissolution process, we have carried out well-tempered metadynamics simulations with a collective variable that captures the slow degrees of freedom, namely, the solute diffusion to and adsorption at the kink site together with the desolvation of the kink site. Simulations were performed at different solution concentrations using constant chemical potential molecular dynamics, and the solubility was identified at the concentration at which the free energy values between the grown and dissolved kink states were equal. The effectiveness of this method is demonstrated by its success in reproducing the experimental trends of solubility of urea and naphthalene in a variety of solvents.



INTRODUCTION

Molecular solubility is a crucial indicator of functionality in pharmaceutical drugs. Not only does it dictate the driving force for crystallization of drugs in their purification process, but it also dictates their bio-availability in the body upon their intake. To date, determining the solubility of candidate drug molecules has been primarily done using experiments. However, experiments often are time-consuming and for some systems expensive and tedious. Given these deficiencies, molecular dynamics (MD) simulations have been deemed to be a potentially useful tool for predicting the solubilities of candidate drug molecules and hence providing guidance in screening processes.

In approaches involving MD simulations, there are two major ways to extract the solubility for a two-phase crystal-liquid system: either indirectly through thermodynamic cycles, where the solubility is obtained by the combination of individual simulations of the crystallized and fluid phases,^{1–6} or through direct coexistence simulations, where the crystal surface is exposed to the solution to sample growth and dissolution events.^{7–10} Although direct coexistence simulations are conceptually a more straightforward approach compared to thermodynamic cycles, they suffer from a time scale limitation. Namely, the time scale required to obtain sufficient statistics for the estimation of the solubility of a given system is considerably longer than the time scale which can be simulated with present-day computational capabilities.^{7–9} To overcome this limitation in the context of crystallization, enhanced sampling was introduced with considerable success to study

layered growth and homogeneous nucleation of small organic molecules in solution.^{11–15} Here, we shall follow this approach with the aim of calculating the solubilities for given solute–solvent systems.

We focus on the growth and dissolution process of solute molecules at kink sites, which are sites that are located at the ends of unfinished molecule rows of crystal surface edges.^{14–16} The study of kink site growth and dissolution allows us to extract the free energy difference between a solute molecule in the state of the crystal and its dissolved state in solution, while excluding surface free energy differences. It is also a natural choice, because at solution concentrations around solubility, growth and dissolution at these sites is the rate-limiting crystallization step;^{17,18} by modeling and simulating it, we are therefore able to extract the solubility. Figure 1a illustrates the crystal surface exposing a kink site to the solution, showing on the left the dissolved kink site (state A) and on the right the crystallized kink site (state B). The cubes indicate growth units (solute molecules), whereby the growth unit undergoing the phase transition is colored in red. As the kink site grows by one growth unit, the free energy of the system changes by the

Received: May 10, 2021

Revised: July 20, 2021

Published: July 30, 2021



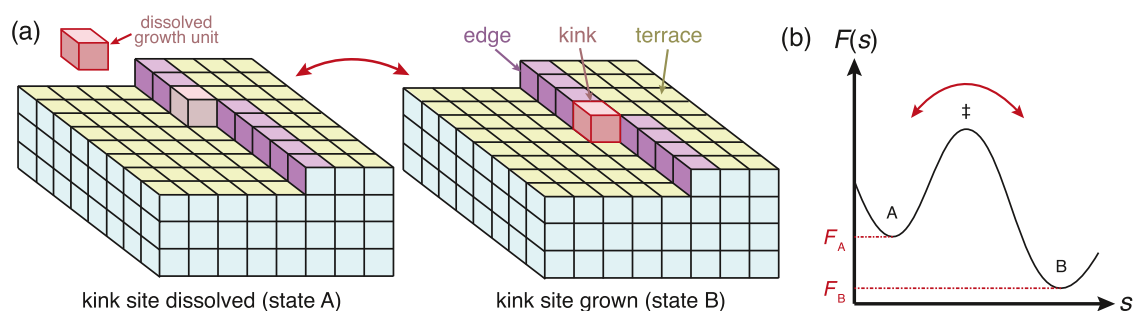


Figure 1. (a) Scheme of the crystal surface with an unfinished layer comprising a kink site; the growth units are shown as cubes. The cube with red contours indicates a growth unit which is either dissolved in solution (state A) or incorporated into the kink site (state B). Growth units at the surface are colored in green for terraces, violet for edges, and red for kink sites. (b) Corresponding scheme of the free energy profile along the reaction coordinate, s , which describes appropriately the two states A and B separated by transition state \ddagger . The energy difference between the grown and dissolved states, $\Delta F = F_B - F_A$, defines whether the solution is undersaturated, $\Delta F > 0$, at solubility, $\Delta F = 0$, or supersaturated, $\Delta F < 0$.

energy difference between that corresponding to having the growth unit incorporated in the crystal lattice and that corresponding to the growth unit dissolved in solution. The surface free energy remains constant in this process, since the kink growth regenerates a new kink site and therefore preserves the number of crystalline growth units at edges, kinks, or in terraces as illustrated in Figure 1a.

The corresponding scheme of the free energy surface, $F(s)$, of the kink growth process is shown in Figure 1b, where s corresponds to a reaction coordinate that captures the dissolved (A) or crystalline (B) kink site states. The symbol \ddagger labels the transition state, which being a local maximum in energy indicates that growth and dissolution are activated processes. The energy difference between the two states, $\Delta F = F_B - F_A$, for a given solute mole fraction defines whether the solution is undersaturated, $\Delta F > 0$, at solubility, $\Delta F = 0$, or supersaturated, $\Delta F < 0$.^{18–21}

The growth and dissolution of molecules at the kink site is enhanced through well-tempered metadynamics (WTMetaD).²² For this work, we have developed a collective variable (CV), which captures the slow degrees of freedom for the kink site growth. The biased simulations allow us to obtain sufficient number of crossings between the states of grown and dissolved kink site to be able to resolve the difference in free energies for the different solution concentrations, which further allows us to identify the solubility with remarkable accuracy.

As the kink site grows, the solution is depleted, which is especially significant for low solution concentrations. To prevent solution depletion, the solute concentration in the region adjacent to the crystal surface is kept constant using the constant chemical potential molecular dynamics (C μ MD) method developed by Perego et al.²³

The methodology presented here is applied to two organic species, namely, urea and naphthalene. Particularly, we aim at predicting the solubility of these species in different solvents to capture the trends observed in experiments.

COLLECTIVE VARIABLES (CVS)

CV for Enhanced Kink Growth Simulations. To overcome the time scale limitation of kink site growth, we use enhanced sampling with the WTMetaD method. An integral part of a CV-based enhanced sampling method like WTMetaD is the appropriate choice of CVs; they are functions of atomic coordinates and embed the system's slow degrees of freedom.^{24,25} In WTMetaD, a time-dependent bias potential is

constructed as a function of these CVs, with the aim to discourage frequently visited states and encourage the system to overcome free energy barriers.

For a kink site to grow, a solute molecule needs to diffuse to and adsorb at the kink site. During the adsorption event, the solute and kink site need to undergo partial desolvation. Both diffusion and adsorption are often rate-limiting.²⁰ In the following, we systematically define a set of functions that are used to describe each of these phenomena.

To describe the solute diffusion and its adsorption and desorption at the kink site, we define the following function

$$s_s = \sum_i \exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_k|^2}{2\sigma_s^2}\right) \quad (1)$$

which is composed of sums of Gaussian-like bell curves summed over each solute molecule i in solution or at the biased kink site. Vectors \mathbf{r}_i and \mathbf{r}_k correspond to the positions of solute molecule i and of kink site, respectively; σ_s is the variance of the Gaussian function. Figure 2 shows the contour lines of s_s (in red) at the biased kink site of the unfinished layer of urea projected along all three spatial directions. By an appropriate choice of σ_s , the nonzero function values of s_s extend into the liquid phase. This allows us, by applying bias, to push solute molecules from the liquid toward the kink site and vice versa, enhancing diffusion as well as adsorption and desorption.

To describe the solvent's desorption and adsorption at the kink site, we introduce a function with the same functional form as s_s

$$s_1 = \sum_j \exp\left(-\frac{|\mathbf{r}_j - \mathbf{r}_k|^2}{2\sigma_1^2}\right) \quad (2)$$

where we sum over all solvent molecules j . Vector \mathbf{r}_j corresponds to the position of solvent molecule j . The bell curve width σ_1 is chosen such that s_1 possesses nonzero values only at the solvent's adsorption site at the kink. Joswiak et al. reported a similar function to s_1 used in their enhanced sampling simulations for water desorption and adsorption at kink sites of a rock salt.^{26,27} In contrast to the referenced work, we employ this functional form for both the solute and the solvent.

The CV, s_b , used for the WTMetaD simulations in this work is a function of s_s and s_1

$$s_b = w_s(s_s^{\chi_{s1}} + s_s^{\chi_{s2}}) + w_1 s_1^{\chi_1} \quad (3)$$

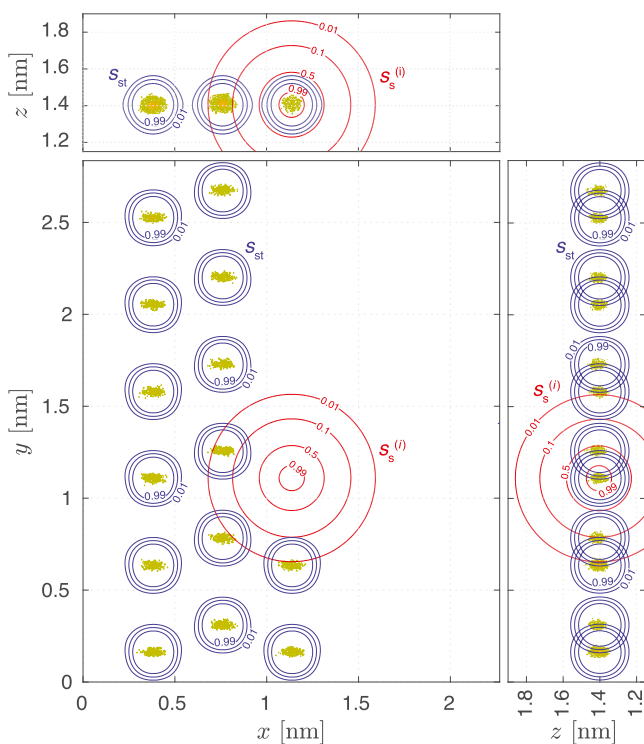


Figure 2. Histogram of urea carbon atom positions (yellow dots) of the unfinished layer sampled from 100 000 time frames (100 ns of simulation) projected along all three spatial directions. The red contour lines correspond to the Gaussian-like function, s_s , which is a part of the biased CV, s_b . s_s is introduced to accelerate the solute diffusion and adsorption/desorption in the biased simulations. The blue contour lines correspond to the surface structure CV, s_{st} , through which a harmonic potential is introduced to prevent the dissolution of the unfinished layer.

where w_s and w_l are scalar weights that are obtained by harmonic linear discriminant analysis.^{28–31} To improve the WTMetaD sampling performance, s_s is potentiated through positive scalar exponents $\chi_{s1} < 1$ and $\chi_{s2} > 1$. This variable transformation helps us to map the corresponding free energy surface $F: s_s \rightarrow s_s^{\chi_{s1}} + s_s^{\chi_{s2}}$ into a function with wider local minima, which otherwise would be too narrow for efficient bias deposition performance. Also s_l is transformed similarly, so as the mapping of the free energy surface $F: s_l \rightarrow s_l^{\chi_l}$ leads to a wider well of the local energy minimum at $s_l = 0$. These transformations follow the approach utilized earlier.^{32,33}

It is worth noting that s_b does not take into account the reorientation of the solute at the kink site, which might be a further rate-limiting step for organic molecules exhibiting a more complex structure.³⁴

Surface Structure CV. While growing the kink sites that are present in an unfinished layer, it is likely that some of the crystalline molecules of the unfinished layer dissolve. This might lead to undesired difficulties in the sampling of growth and dissolution events at the kink site.

To prevent such dissolution, a harmonic potential is introduced. This potential is defined as a function of a surface structure CV, s_{st} , whose value for each molecule is 1 if the molecule in the unfinished layer is in a lattice position or 0 otherwise [see the Supporting Information (SI) for further details]. In this way, s_{st} counts the total number of molecules in the unfinished layer. Figure 2 shows the contour lines of s_{st} projected along all three spatial directions together with the

histogram of the molecule centers of the unfinished layer for the case of urea. Note that we should not alter the natural lattice vibrations of the crystalline molecules by applying the harmonic potential. This is achieved by setting the parameters of s_{st} such that its potential only counteracts the molecule's motion out of its lattice position. The amplitudes of the natural lattice vibrations are known from unconstrained simulations.

Crystallinity CV. The biased CV, s_b , discussed in the CV for Enhanced Kink Growth Simulations section does not suffice to define the kink site crystallinity since it lacks information about the solute molecule's orientational ordering. To account for this, we have introduced a pair of CVs, s_{c1} and s_{c2} , defined using the logistic function as follows

$$s_{c_j} = \sum_i \left(1 - \frac{1}{1 + \exp(-\sigma_c(|\mathbf{r}_{c_j,i} - \bar{\mathbf{r}}_{c_j}| - d_c))} \right) \quad (j = 1, 2) \quad (4)$$

These CVs discriminate whether the atom positions $\mathbf{r}_{c1,i}$ and $\mathbf{r}_{c2,i}$ of solute molecule i at the kink site are at their concordant crystal lattice positions, namely, $\bar{\mathbf{r}}_{c1}$ and $\bar{\mathbf{r}}_{c2}$. Using two appropriate atom positions allows taking into account the molecule's orientation at the kink site as well, as requested. Moreover, d_c is the position of the logistic function's step and is set roughly at the distance between the crystalline lattice position and the kink growth region of transition, i.e., the region where the biased kink site is neither fully dissolved nor fully crystalline. σ_c is the steepness of the step and is chosen so as the values of the switching function, which are larger than 0 but smaller than 1, correspond roughly to the molecule's positions that are contained within the region of transition. For s_{c1} and s_{c2} , values around 0 correspond to a fully dissolved biased kink site while values around 1 correspond to a crystalline biased kink site (see the SI for further details).

COMPUTATIONAL DETAILS

Force Fields. The general AMBER force field (GAFF)^{35,36} with full atomistic description was used for all molecular species considered in this work. The force field parameters of naphthalene are reported in a previous paper.³² For naphthalene, the electrostatic potential was calculated using Gaussian 09³⁷ at the B3LYP/6-31G(d,p) level and the atom partial charges were fitted with the restrained electrostatic potential method.^{38,39} The force field parameters of all other molecules studied in this work were taken from the literature.⁴⁰

Simulation Runs. The simulations were performed with Gromacs 2016.5^{41–45} patched with a private version of Plumed 2.5.0.⁴⁶ The temperature was kept constant with the velocity rescaling algorithm.⁴⁷ The nonbonded electrostatic interactions cutoff was set to 1 nm. For the long-range electrostatics, the Ewald particle mesh algorithm⁴⁸ was used. To run simulations at an integration time step of 0.002 ps, the LINCS algorithm was used to constrain the hydrogen bonds.^{44,49}

Kink growth simulations were performed for urea polymorph I⁵⁰ grown from acetonitrile, ethanol, and methanol solutions, as well as for naphthalene polymorph I⁵¹ grown from ethanol and toluene solutions. Urea simulations were run at 300 K, while naphthalene simulations were run at 280 K, in accordance with our previous work.³² For both solutes, the slowest growing face was exposed to the solution. Both solutes exhibit the same slowest growing face in all solvents

considered, namely, face $\{110\}$ in the case of urea¹¹ and face $\{001\}$ in the case of naphthalene.^{52,53} The unfinished layer was cut along the slowest growing edge, i.e., for urea along the $[001]$ direction and for naphthalene along the $[010]$ direction. For each simulation setup, one edge of the surface comprises an unfinished row and the biased kink site. The simulation box equilibration procedure⁵⁴ to obtain the initial configurations is reported in the SI. The visualizations⁵⁵ of the simulation setups are shown in Figure 3 for the cases of naphthalene and urea, each grown from an ethanol solution.

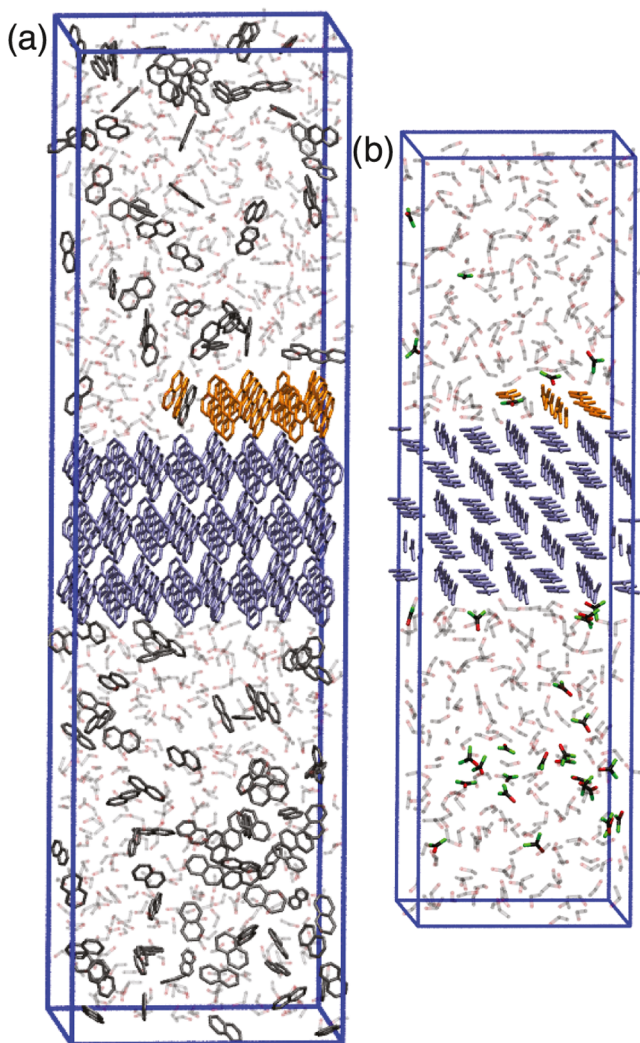


Figure 3. Visualizations of the kink growth simulation setups for (a) naphthalene grown in ethanol and (b) urea grown in ethanol. The biased kink site is located in the center of the upper surface layer. The bulk crystal molecules are colored in blue, and the unfinished layer molecules are colored in orange. The atoms of the molecules in solution are colored in black for carbon, red for oxygen, and green for nitrogen. Hydrogens are omitted for clarity. Ethanol molecules are shown in faded colors.

Small displacements of the crystal can occur during the simulations, which can affect the localization protocol of the biased kink site. To avoid such displacements, the movement of the bulk crystal's center of mass was constrained in all three spatial directions by means of a harmonic potential.

All simulations were run for at least 750 ns. The solution concentration was kept constant with the $C\mu$ MD algo-

rithm.^{23,32,56–58} The CV s_b was biased through WTMetaD. The WTMetaD Gaussian bias deposition was switched off after 300 ns to increase the convergence performance. All parameter values and further details can be found in the SI.

RESULTS

To compute the solubilities of the different systems, we employed WTMetaD to accelerate the growth and dissolution of the solute molecules from kink sites. Using WTMetaD, we sample the free energy difference between the grown and dissolved states, ΔF , thereby allowing us to obtain the solubility values corresponding to the solute mole fraction of the system when $\Delta F = 0$.

A typical time evolution of the biased CV, s_b , is shown in Figure 4 for the case of urea face $\{110\}$ exposed to an acetonitrile solution at a solute mole fraction of $x = 0.0018$. One can clearly see that through the WTMetaD bias, many growth and dissolution events at the kink site are obtained for a simulation time of 1 μ s. This is in sharp contrast to an unbiased simulation, where only a few growth or dissolution events would be observed within the same simulation time interval. Figure 4b shows representative visualizations of the dissolved and grown kink site states.

Although s_b is a potent CV for enhancing the growth and dissolution of kink sites, it does not capture the crystallinity of the kink site, which is necessary to compute ΔF . Therefore, we reweigh the results using the approach of Tiwary et al.⁵⁹ with the aforementioned crystallinity CVs, s_{c1} and s_{c2} . For urea, we take the carbon and oxygen atoms as reference for s_{c1} and s_{c2} . For naphthalene, we take the centers of the two outermost carbon atom pairs along the naphthalene's long axis as reference (see the SI for details). The first 300 ns of the simulation was used to construct the WTMetaD bias potential. After the first 300 ns, the Gaussian bias deposition was stopped to continue the simulation with a static bias. Only the parts of the simulation with a static bias were used for the reweighing. The corresponding reweighed free energy surface at $t = 1 \mu$ s is shown in Figure 5a, from which the energy difference between the crystallized and dissolved kink site states, $\Delta F = F_B - F_A$, is obtained. Figure 5b shows the time evolution of the reweighed ΔF for five different mole fractions. ΔF converges in all cases at $t \approx 700$ ns.

From the sampling of ΔF for different mole fractions, one can obtain the predicted solubility, x_{sim}^* , through interpolation. The graphs in Figure 6 show the results for all sampled ΔF series of urea grown in acetonitrile, ethanol, methanol, and naphthalene grown in ethanol and toluene (circles). The corresponding experimental solubilities reported in the literature^{60–63} are also shown for comparison (asterisks). We approximate the correlation between ΔF and x as linear within the sampled mole fraction ranges. The graphs show the linear regression for each solvent's ΔF series together with the lower and upper bound determined by the corresponding standard deviation. We locate x_{sim}^* where the linear regression intersects the horizontal axis, i.e., $\Delta F = 0$. The numerical values of the predicted solubility values are listed in Table 1 together with the experimental values.

The comparison between simulations and experiments clearly shows that in all cases, the simulations predict the order of magnitude of the values of solubility correctly. Moreover, for both urea and naphthalene, the solubility trends in the different solvents are predicted correctly as well.

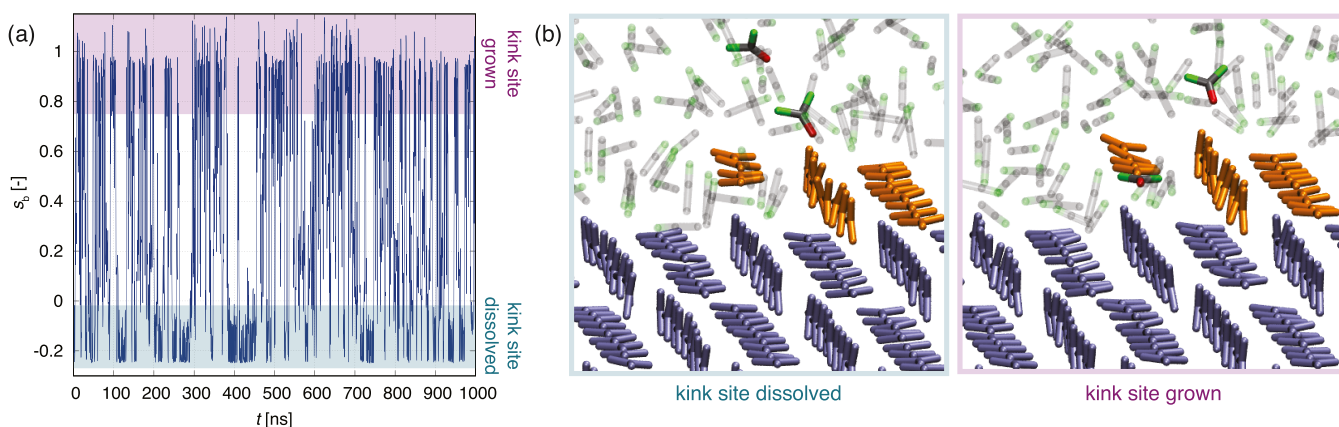


Figure 4. (a) Biased CV, s_b , in dependence of time, t , for urea grown from acetonitrile at a mole fraction of $x = 0.0018$. The regions of the dissolved and grown kink site are shaded in blue and violet, respectively. Since we are using negative numbers for weights of the solvent adsorption/desorption, s_b exhibits negative values for the dissolved kink site states. (b) Representative visualizations of the dissolved (left) and grown (right) kink site. The biased kink site is located in the center of each frame. The same color code was used as in Figure 3. Hydrogens are omitted for clarity. Acetonitrile molecules are shown in faded colors.

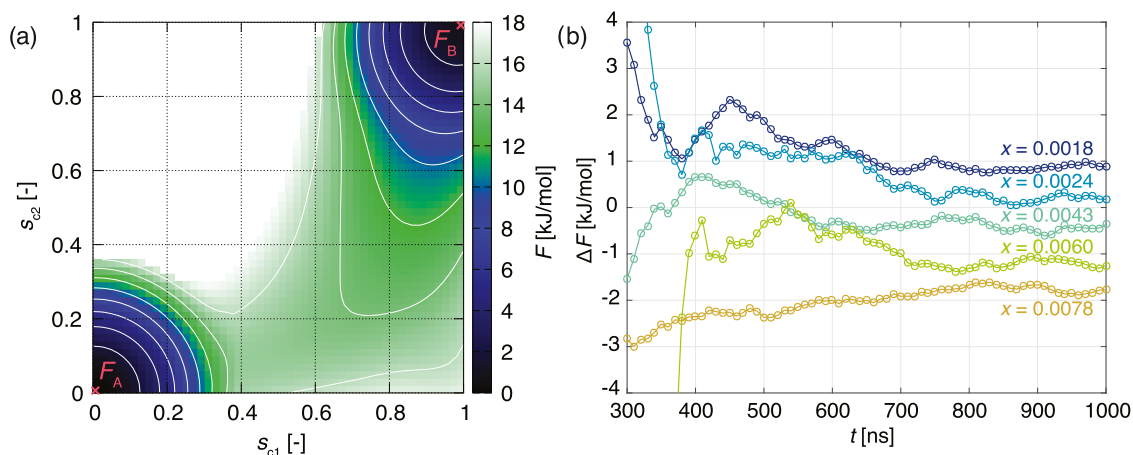


Figure 5. (a) Free energy surface in dependence of the crystallinity CVs, s_{c1} and s_{c2} , for the case of urea grown from acetonitrile solution at a mole fraction of $x = 0.0018$. The difference in energy of the grown kink site, F_B , and dissolved kink site, F_A , provides ΔF . (b) Time evolution of ΔF for five different mole fractions of urea in acetonitrile.

More specifically for urea, the solubility in ethanol and methanol is underestimated while that in acetonitrile is predicted exactly. We know from previous studies that the calculated force field melting point, $T_{\text{sim}}^m \approx 420$ K, is above its experimental counterpart, $T_{\text{exp}}^m = 406$ K.^{12,64} This indicates that the cohesive energy between the urea molecules described by the empirical potential is larger than the actual one and it further implies that the release of urea molecules from the crystal surface is less easy than in reality. The solubility of urea in the alcohols is therefore underestimated. Nevertheless, in the case of the urea solubility in acetonitrile, this effect is compensated by the force fields' overemphasis of the solute–solvent interactions through the amine (urea) and nitrile (acetonitrile) functional groups.

Contrarily, in the case of naphthalene, the force field's calculated melting point, $T_{\text{sim}}^m \approx 328$ K, is below the experimental one, $T_{\text{exp}}^m = 353$ K. Thus, for naphthalene, the lower cohesive force results in facile dissolution of crystalline naphthalene molecules in the solvents and thereby in an overestimated solubility. The large deviation of solubility of naphthalene in ethanol is due to the force field's underemphasis of solute–solute interactions. While in the case of

toluene, the smaller discrepancy in solubility between simulation and experiment can be anticipated based on the underestimation of the solute–solvent interactions mediated by the underemphasis of the π – π stacking interactions between their aromatic moieties.

For urea and naphthalene, it suffices to sample only one kink site to obtain the solubility, since the chemical environment upon the growth of one molecule along the chosen edge does not change in terms of crystal surface energy. For molecules with more complex growth units such as dimers, one has to sample the growth and dissolution of each subunit at the particular kink site to obtain the solubility by averaging the energy difference between dissolved and crystalline states over all of these sites.^{34,65}

It is important to underline that the simulation method used in this work allowed us to obtain converged values of ΔF for mole fractions as low as $x = 0.0018$, i.e., corresponding to a mass fraction of $w = 0.006$, as demonstrated for the case of urea in acetonitrile. As a consequence, this simulation setup can extract the solubility in the relevant concentration range for small real-life drug intermediates and APIs. For combinations of solute and solvent, where the solubility is

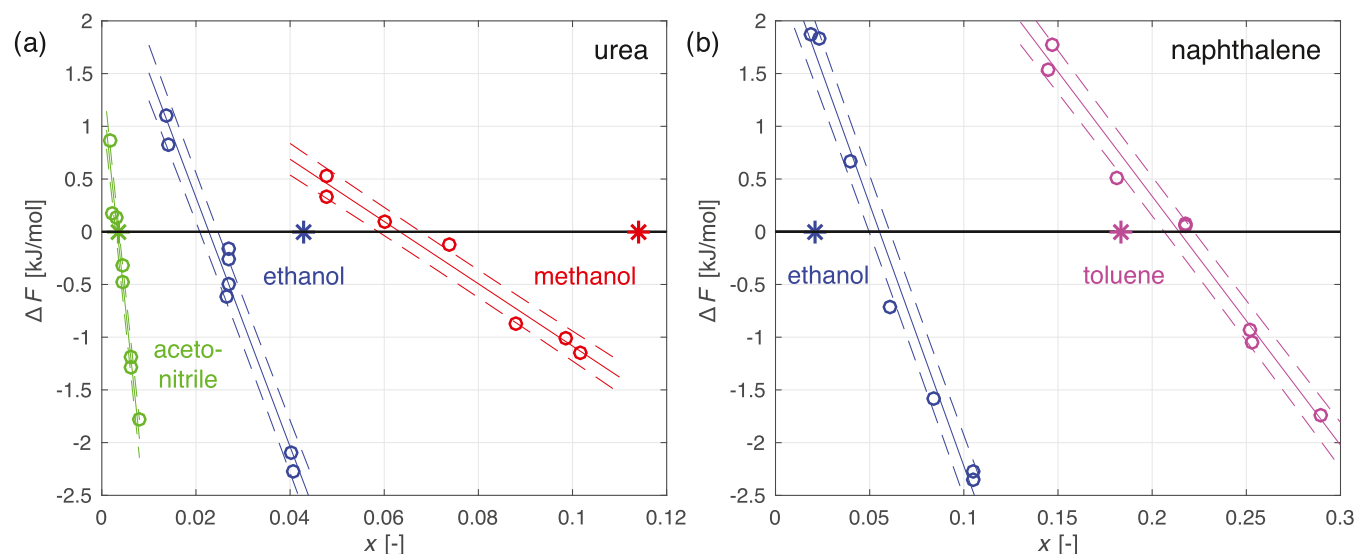


Figure 6. Sampled ΔF in dependence of mole fraction x (circles) in comparison to experimental solubilities (asterisks): (a) results for urea grown from acetonitrile (green), ethanol (blue), and methanol (red); (b) naphthalene grown from ethanol (blue) and toluene (violet). The straight lines correspond to linear fits of the sampled ΔF , and the dashed lines correspond to standard deviations. The simulated solubilities are obtained from interpolation of the linear fit at $\Delta F = 0$.

Table 1. Predicted Solubilities, x_{sim}^* , Compared to Experimental Solubilities, x_{exp}^* ^a

| | urea | | | naphthalene | |
|-------------------------------------|-----------|------------|------------|-------------|------------|
| | MeCN | EtOH | MeOH | EtOH | MePh |
| x_{sim}^* | 0.0033(4) | 0.0228(20) | 0.0633(45) | 0.0552(55) | 0.2143(79) |
| x_{exp}^* ^{60–63} | 0.0037 | 0.0429 | 0.1140 | 0.0213 | 0.183 |

^aThe simulations were performed at 1 bar and 300 K for urea and at 280 K for naphthalene.

lower than a mass fraction of roughly $w = 0.01$, such a setup cannot be used to estimate solubility. It is also worth noting that in these cases the species is considered to be insoluble as far as the design of a crystallization process in the pharmaceutical industry is concerned.

CONCLUSIONS

We have introduced a new approach to predict the solubility of organic molecular crystals in a variety of solvents using molecular dynamics. The approach samples energy differences between grown and dissolved kink site states at controlled solution concentrations.²³ To achieve sufficient sampling performance of the growth and dissolution events, well-tempered metadynamics was used through a collective variable that has been developed as a part of this work. The collective variable addresses all relevant slow degrees of freedom of the kink growth process, namely, the solute diffusion to and adsorption/desorption at the kink site as well as the solvent desorption/adsorption at the kink site.

We have shown that using General Amber Force Fields for urea crystals grown in acetonitrile, ethanol, and methanol solutions and for naphthalene crystals grown in ethanol and toluene solutions, the computed solubility values exhibit the correct order of magnitude and the correct trends in the different solvents compared with experimental measurements.

The deviation of the melting point of the force field from its experimental counterpart allows us to predict whether in the simulations the solubility will be rather underestimated, as for urea, whose force field's melting temperature is too high (solute–solute interactions are overemphasized), or over-

estimated, as for naphthalene, whose force field's melting temperature is too low in comparison to experiments (solute–solute interactions are underemphasized). It is worth noting that also the solute–solvent interactions play a similarly important role and can either reinforce or compensate the effect of the solute–solute interactions described above. In the case of urea in acetonitrile, both the solute–solvent interactions and the solute–solute interactions are overestimated, thus leading to a compensating effect and an estimated solubility value close to the experimental one. In the case of naphthalene in toluene, the solute–solvent interactions and solute–solute interactions are both underestimated, hence leading again to a solubility prediction close to the experimental value. Such compensating effect is not present in the case of the other combinations of solute and solvent studied in this work.

Naturally, here, we obtain the solubility of the force field. However, with the rapid advance of machine learning-based force fields, which can represent the atomistic properties to *ab initio* accuracy,⁶⁶ it is reasonable to expect that in the near future, the presented method will be able to compute solubility significantly closer to its experimental counterpart.

We anticipate that an extension of the setup used here will allow also the calculation of the activation energies of kink growth, which can be important for the estimation of reliable parameters of spiral growth models.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.cgd.1c00546>.

Simulation setup equilibration; constant chemical potential method; collective variables (CVs); biased CV; surface structure CV; adsorption site CVs; crystallinity CVs; and sampling of solubility with chemically distinct kink sites (PDF)

(PDF)

AUTHOR INFORMATION

Corresponding Authors

Michele Parrinello – Istituto Italiano di Tecnologia (IIT),
16163 Genova GE, Italy; Email: michele.parrinello@iit.it

Marco Mazzotti – Institute of Energy and Process Engineering,
ETH Zürich, CH-8092 Zürich, Switzerland; orcid.org/0000-0002-4948-6705; Email: marco.mazzotti@ipe.mavt.ethz.ch

Authors

Zoran Bjelobrk – Institute of Energy and Process Engineering,
ETH Zürich, CH-8092 Zürich, Switzerland; orcid.org/0000-0002-2494-8133

Dan Mendels – Pritzker School of Molecular Engineering,
University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0002-2408-3133

Tarak Karmakar – Istituto Italiano di Tecnologia (IIT),
16163 Genova GE, Italy; orcid.org/0000-0002-8721-6247

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.cgd.1c00546>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Z.B. and M.M. thank Novartis Pharma AG for the partial financial support to this project. Z.B. thanks Pablo Piaggi, Ashwin Rajagopalan, Michele Invernizzi, Thilo Weber, Philipp Müller, and Marco Holzer for valuable discussions. The computational resources were provided by ETH Zürich and the Swiss Center for Scientific Computing at the Euler Cluster.

REFERENCES

- (1) Ferrario, M.; Ciccotti, G.; Spohr, E.; Cartailler, T.; Turq, P. Solubility of KF in water by molecular dynamics using the Kirkwood integration method. *J. Chem. Phys.* **2002**, *117*, 4947–4953.
- (2) Paluch, A. S.; Maginn, E. J. Predicting the Solubility of Solid Phenanthrene: A Combined Molecular Simulation and Group Contribution Approach. *AIChE J.* **2013**, *59*, 2647–2661.
- (3) Mester, Z.; Panagiotopoulos, A. Z. Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations. *J. Chem. Phys.* **2015**, *143*, No. 044505.
- (4) Benavides, A. L.; Aragonés, A. L.; Vega, C. Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route. *J. Chem. Phys.* **2016**, *144*, No. 124504.
- (5) Li, L.; Totton, T.; Frenkel, D. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *J. Chem. Phys.* **2017**, *146*, No. 214110.
- (6) Khanna, V.; Monroe, J. I.; Doherty, M. F.; Peters, B. Performing solvation free energy calculations in LAMMPS using the decoupling approach. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 641–646.
- (7) Manzanilla-Granados, H. M.; Saint-Martín, H.; Fuentes-Azcatl, R.; Alejandre, J. Direct Coexistence Methods to Determine the Solubility of Salts in Water from Numerical Simulations. Test Case NaCl. *J. Phys. Chem. B* **2015**, *119*, 8389–8396.

(8) Kolafa, J. 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.* **2016**, *145*, No. 204509.

(9) Espinosa, J. R.; Young, J. M.; Jiang, H.; Gupta, D.; Vega, C.; Sanz, E.; Debenedetti, P. G.; Panagiotopoulos, A. Z. On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures. *J. Chem. Phys.* **2016**, *145*, No. 154111.

(10) Asadi, E.; Asle Zaeem, M.; Nouranian, S.; Baskes, M. I. Two-phase solid–liquid coexistence of Ni, Cu, and Al by molecular dynamics simulations using the modified embedded-atom method. *Acta Mater.* **2015**, *86*, 169–181.

(11) Salvalaglio, M.; Vetter, T.; Mazzotti, M.; Parrinello, M. Controlling and Predicting Crystal Shapes: The Case of Urea. *Angew. Chem., Int. Ed.* **2013**, *52*, 13369–13372.

(12) Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Insight into the nucleation of urea crystals from the melt. *Chem. Eng. Sci.* **2015**, *121*, 51–59.

(13) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E6–E14.

(14) Kossel, W. Zur Theorie des Kristallwachstums. *Nachr. Ges. Wiss. Göttingen* **1927**, 135–143.

(15) Stranski, I. N. Zur Theorie des Kristallwachstums. *Z. Phys. Chem.* **1928**, *136*, 259–278.

(16) Burton, W. K.; Cabrera, N.; Frank, F. C.; Mott, N. F. The growth of crystals and the equilibrium structure of their surfaces. *Philos. Trans. R. Soc., A* **1951**, *243*, 299–358.

(17) Snyder, R. C.; Doherty, M. F. Faceted crystal shape evolution during dissolution or growth. *AIChE J.* **2007**, *53*, 1337–1348.

(18) Li, J.; Tilbury, C. J.; Joswiak, M. N.; Peters, B.; Doherty, M. F. Rate Expressions for Kink Attachment and Detachment During Crystal Growth. *Cryst. Growth Des.* **2016**, *16*, 3313–3322.

(19) Snyder, R. C.; Doherty, M. F. Predicting Crystal Growth by Spiral Motion. *Proc. R. Soc. A* **2009**, *465*, 1145–1171.

(20) Li, J.; Tilbury, C. J.; Kim, S. H.; Doherty, M. F. A design aid for crystal growth engineering. *Prog. Mater. Sci.* **2016**, *82*, 1–38.

(21) Tilbury, C. J.; Green, D. A.; Marshall, W. J.; Doherty, M. F. Predicting the Effect of Solvent on the Crystal Habit of Small Organic Molecules. *Cryst. Growth Des.* **2016**, *16*, 2590–2604.

(22) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.

(23) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *J. Chem. Phys.* **2015**, *142*, No. 144113.

(24) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(25) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(26) Joswiak, M. N.; Doherty, M. F.; Peters, B. Ion dissolution mechanism and kinetics at kink sites on NaCl surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 656–661.

(27) Joswiak, M. N.; Peters, B.; Doherty, M. F. In Silico Crystal Growth Rate Prediction for NaCl from Aqueous Solution. *Cryst. Growth Des.* **2018**, *18*, 6302–6306.

(28) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *J. Phys. Chem. Lett.* **2018**, *9*, 2776–2781.

(29) Piccini, G.; Mendels, D.; Parrinello, M. Metadynamics with Discriminants: A Tool for Understanding Chemistry. *J. Chem. Theory Comput.* **2018**, *14*, 5040–5044.

(30) Mendels, D.; Piccini, G.; Brotzakis, F. Z.; Yang, Y. I.; Parrinello, M. Folding a small protein using harmonic linear discriminant analysis. *J. Chem. Phys.* **2018**, *149*, No. 194113.

(31) Brotzakis, F. Z.; Mendels, D.; Parrinello, M. Augmented Harmonic Linear Discriminant Analysis. 2019, arXiv:1902.08854. arXiv.org e-Print archive. <http://arxiv.org/abs/1902.08854>.

- (32) Bjelobrk, Z.; Piaggi, P. M.; Weber, T.; Karmakar, T.; Mazzotti, M.; Parrinello, M. Naphthalene crystal shape prediction from molecular dynamics simulations. *CrystEngComm* **2019**, *21*, 3280–3288.
- (33) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The role of water in host-guest interaction. *Nat. Commun.* **2021**, *12*, No. 93.
- (34) Chernov, A. A. Crystal Growth and Crystallography. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1998**, *54*, 859–872.
- (35) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (36) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (37) Frisch, M. J.; et al. *Gaussian 09*, revision B.01; Gaussian Inc.: Wallingford, CT, 2009.
- (38) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem. A* **1993**, *97*, 10269–10280.
- (39) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE-AnteChamber Python Parser interface. *BMC Res. Notes* **2012**, *5*, No. 367.
- (40) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **2012**, *28*, 752–753.
- (41) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (42) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
- (43) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (44) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (45) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (46) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (47) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130*, No. 074101.
- (48) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (49) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (50) Sklar, N.; Senko, M. E.; Post, B. Thermal effects in urea: the crystal structure at -140°C and at room temperature. *Acta Crystallogr.* **1961**, *14*, 716–720.
- (51) Cruickshank, D. W. J. A detailed refinement of the crystal and molecular structure of naphthalene. *Acta Crystallogr.* **1957**, *10*, 504–508.
- (52) Lovette, M. A.; Doherty, M. F. Multisite models to determine the distribution of kink sites adjacent to low-energy edges. *Phys. Rev. E* **2012**, *85*, No. 021604.
- (53) Elwenspoek, M.; van der Eerden, J. P. Kinetic roughening and step free energy in the solid-on-solid model and on naphthalene crystals. *J. Phys. A: Math. Gen.* **1987**, *20*, 669–678.
- (54) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (55) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (56) Karmakar, T.; Piaggi, P. M.; Perego, C.; Parrinello, M. A Cannibalistic Approach to Grand Canonical Crystal Growth. *J. Chem. Theory Comput.* **2018**, *14*, 2678–2683.
- (57) Karmakar, T.; Piaggi, P. M.; Parrinello, M. Molecular Dynamics Simulations of Crystal Nucleation from Solution at Constant Chemical Potential. *J. Chem. Theory Comput.* **2019**, *15*, 6923–6930.
- (58) Han, D.; Karmakar, T.; Bjelobrk, Z.; Gong, J.; Parrinello, M. Solvent-mediated morphology selection of the active pharmaceutical ingredient isoniazid: Experimental and simulation studies. *Chem. Eng. Sci.* **2019**, *204*, 320–328.
- (59) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736–742.
- (60) Loeser, E.; Delacruz, M.; Madappalli, V. Solubility of Urea in Acetonitrile–Water Mixtures and Liquid–Liquid Phase Separation of Urea-Saturated Acetonitrile–Water Mixtures. *J. Chem. Eng. Data* **2011**, *56*, 2909–2913.
- (61) Lee, F. M.; Lahti, L. E. Solubility of urea in water-alcohol mixtures. *J. Chem. Eng. Data* **1972**, *17*, 304–306.
- (62) Speyers, C. L. Solubilities of some carbon compounds and densities of their solutions. *Am. J. Sci.* **1902**, *14*, 293–302.
- (63) Ward, H. L. The Solubility Relations of Naphthalene. *J. Phys. Chem. B* **1926**, *30*, 1316–1333.
- (64) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *J. Am. Chem. Soc.* **2012**, *134*, 17221–17233.
- (65) Kuvadia, Z. B.; Doherty, M. F. Spiral Growth Model for Faceted Crystals of Non-Centrosymmetric Organic Molecules Grown from Solution. *Cryst. Growth Des.* **2011**, *11*, 2780–2802.
- (66) Bonati, L.; Parrinello, M. Silicon Liquid Structure and Crystal Nucleation from Ab Initio Deep Metadynamics. *Phys. Rev. Lett.* **2018**, *121*, No. 265701.