PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2017, **19**, 5285

Solubility curves and nucleation rates from molecular dynamics for polymorph prediction – moving beyond lattice energy minimization†

Conor Parks,[a] Andy Koswara,[a] Frank DeVilbiss,[a] Hsien-Hsin Tung,[b] Nandkishor K. Nere,[b] Shailendra Bordawekar,[b] Zoltan K. Nagy[a] and Doraiswami Ramkrishna*[a]

Received 20th October 2016,
Accepted 23rd January 2017

DOI: 10.1039/c6cp07181c

rsc.li/pccp

Current polymorph prediction methods, known as lattice energy minimization, seek to determine the crystal lattice with the lowest potential energy, rendering it unable to predict solvent dependent metastable form crystallization. Facilitated by embarrassingly parallel, multiple replica, large-scale molecular dynamics simulations, we report on a new method concerned with predicting crystal structures using the kinetics and solubility of the low energy polymorphs predicted by lattice energy minimization. The proposed molecular dynamics simulation methodology provides several new predictions to the field of crystallization. (1) The methodology is shown to correctly predict the kinetic preference for β-glycine nucleation in water relative to α- and γ-glycine. (2) Analysis of nanocrystal melting temperatures show γ- nanocrystals have melting temperatures up to 20 K lower than either α- or β-glycine. This provides a striking explanation of how an energetically unstable classical nucleation theory (CNT) transition state complex leads to kinetic inaccessibility of γ-glycine in water, despite being the thermodynamically preferred polymorph predicted by lattice energy minimization. (3) The methodology also predicts polymorph-specific solubility curves, where the α-glycine solubility curve is reproduced to within 19% error, over a 45 K temperature range, using nothing but atomistic-level information provided from nucleation simulations. (4) Finally, the methodology produces the correct solubility ranking of β- > α-glycine. In this work, we demonstrate how the methodology supplements lattice energy minimization with molecular dynamics nucleation simulations to give the correct polymorph prediction, at different length scales, when lattice energy minimization alone would incorrectly predict the formation of γ-glycine in water from the ranking of lattice energies. Thus, lattice energy minimization optimization algorithms are supplemented with the necessary solvent/solute dependent solubility and nucleation kinetics of polymorphs to predict which structure will come out of solution, and not merely which structure has the most stable lattice energy.

Introduction

Twenty-eight years after John Maddox provocatively referred to the inability to predict crystal structures as one of the continuing scandals in the physical sciences,[1] the problem remains unresolved. In the absence of a robust prediction methodology, costly solvent form screenings must be employed, which never truly answer the question of whether all solid-state forms have been observed, and leaves unanswered whether another crystal structure with more desirable solid-state properties remains undiscovered. Advances in lattice energy minimization algorithms[2] have improved the answer to the question of whether crystal structures are predictable to a conditional yes.[3] This is reflected in the crystal structure prediction blind tests performed by the Cambridge Crystallographic Data Centre which have shown that attempts to predict crystal structures through lattice energy minimization are becoming increasingly reliable.[4] However, challenges remain as lattice energy minimization predicts polymorphs based solely off which form has the most stable lattice free energy, often approximated as the potential energy, and hence neglects key crystallization physics. In reality, the form to crystallize from solution will be determined by the solubility, and the kinetics of nucleation and growth, which is determined by the complex atomic interactions between not only solute molecules in the solid-state, but also (1) solute and solvent molecules for homogeneous nucleation mechanisms,

[a] *School of Chemical Engineering, Purdue University, 480 West Stadium Mall, West Lafayette, Indiana, 47907, USA. E-mail: ramkrish@ecn.purdue.edu*
[b] *Process Research & Development, AbbVie, Inc., 1 North Waukegan Road, North Chicago, Illinois 60064, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp07181c

and (2) solute, solvent, and crystallization-container molecules for heterogeneous nucleation mechanisms. This current shortcoming renders lattice energy minimization unable to address the kinetic accessibility of the predicted low energy structures. This is well illustrated by the amino acid glycine, where although lattice energy minimization correctly predicts the experimentally observed structures of glycine, α-, β-, and γ-, to be among the possible low energy structures, the lattice energies are separated by only a few kT,[5–8] preventing a definitive conclusion as to what polymorph will appear in what solvent. The polymorphic outcome of glycine crystallization has been to shown to depend on a multitude of factors ranging from not only on the lattice energies,[5–8] but also on whether the crystallization was performed under confinement,[9] on self-assembled monolayers,[10] in the presence of an electromagnetic field,[11,12] or simply in the presence of methanol.[13] To further advance polymorph prediction algorithms, a computational procedure for predicting the polymorph specific kinetics of nucleation, and the solvent dependent thermodynamic stability as function of length scale, for the low energy structures predicted by lattice energy minimization is proposed. The combination of lattice energy minimization predicted low energy structures with polymorph specific nucleation rates and solubility is shown to provide a more robust polymorph prediction for glycine. Furthermore, this would provide the necessary information for supersaturation feedback control of polymorphic forms,[14–16] and is a step in the direction of fully atomistic crystallization unit design.

Herein, we introduce a methodology to determine polymorph specific nucleation kinetics, bulk (crystal size >1000 nm) solubility curves, and nanocrystal stability rankings, from embarrassingly parallel multiple replica[17] molecular dynamics (MD) seeded cluster nucleation simulations.[18–24] The utility of this method is the ability to screen the low energy structures predicted by lattice energy minimization using the kinetics (nucleation rate) and thermodynamics (solubility) of crystallization. Historically, MD has been used only sparingly to study the nucleation of organic molecules due to the large computational loads. Notable successes in this area include polyethylene,[25] urea,[26–28] benzene,[29,30] *n*-hexane,[31] and methane hydrates.[23] Successful works in this area have struggled to answer what the rate of a particular polymorphic form is, or what the solubility ranking is, either at the nanoscale or bulk scale. Although, significant achievements have been made for the systems monomolecular systems of water[22] and colloids.[32,33] In the case of water, ice Ic and ice Ih polymorphs were simulated using MD, and were shown to be kinetically competitive during nucleation using a computational method similar to the method in this work.[22] To date, no work has used MD based nucleation simulations to quantitatively predict the rankings of polymorphs, at both the nano and bulk scale, of an organic molecule. In the absence of this information at both length scales, the picture of the crystallization pathway remains incomplete, as this work shows that different polymorphic forms can be favored depending on the length scale under question in agreement with Ostwald's rule of stages. The main takeaway from this work is that the low energy structures predicted by lattice energy minimization can be supplemented with nucleation kinetics and solubility predictions to yield the correct polymorph prediction at both the nanoscale and bulk scale for a given solvent in the case of glycine. This allows for the prediction of not only which polymorph have the most stable lattice energy, but also which forms are predicted to come out of solution.

In this work, we study α-, β-, and γ-glycine as a model system, using embarrassingly parallel, Intel® Xeon Phi™ coprocessor accelerated,[17] MD. These three polymorphs correspond to the experimentally observed polymorphs under ambient conditions, as well as three of the low energy structures as predicted by lattice energy minimization.[5–8] Batch crystallization in water is known to exclusively form α-glycine.[34] However, glycine has been shown to exist in the β- form at the nanoscale, when crystallized from water under confinement,[9] neither of which can be explained by lattice energy minimization alone, which predicts γ-glycine formation from the ranking of lattice energies.[5–8] Using the proposed methodology, we achieve the following: (1) a successful prediction of the size dependent polymorphism of α- and β-glycine in water, (2) a reproduction of the α-glycine solubility curve over 45 K to within 19% error, (3) a prediction of the experimentally unmeasurable β-glycine solubility curve, (4) a correct prediction of the strong nucleation kinetics preference for β-glycine formation, and (5) the prediction of the kinetic inaccessibility and thermodynamic instability of γ-glycine critical nuclei clusters, rationalizing why it does not appear upon crystallization in water, despite being the thermodynamically preferred solid-state structure.

Classical nucleation theory

CNT views the nucleation process as a competition between the energetically favorable formation of the new crystal phase, and the energetically unfavorable formation of a liquid solid interface. The bulk limit CNT free energy expression is given by eqn (1).[26]

$$\Delta G = -n\Delta\mu + \sigma a n^{2/3} \quad (1)$$

where n is the size of the cluster, $\Delta\mu$ is the change in chemical potential between the metastable liquid and bulk solid, σ is the interfacial free energy, and a, 8.88×10^{-19} m^2 per molecules$^{2/3}$ for glycine, is a shape factor that converts $an^{2/3}$ into a surface area.

Assuming ideality of solution, the change in chemical potential can be directly related to the solubility by eqn (2).

$$\Delta\mu = kT \ln\left(\frac{c_0}{c_{\mathrm{sat}}(T)}\right) \quad (2)$$

where k is Boltzmann's constant, T is the temperature, c_0 is the concentration of the solute in solvent in the absence of any crystal like nuclei,[26] and $c_{\mathrm{sat}}(T)$ is the solubility of the polymorph in the solvent at temperature T. At the critical size, the free energy is at a maximum, and the height of this free energy barrier is given by eqn (3).

$$\Delta G_{\mathrm{crit}} = n_{\mathrm{crit}}\frac{\Delta\mu}{2} \quad (3)$$

Assuming spherical nuclei, n_{crit} is given by eqn (4).

$$n_{crit} = \left(\frac{2\sigma a}{3\Delta\mu}\right)^3 \quad (4)$$

From eqn (2) and (4), it is evident that for fixed c_0 and solvent molecule, n_{crit} is determined by the temperature of the system. The number density of clusters at the free energy maximum, multiplied by a kinetic prefactor, yields the CNT rate expression, given in eqn (5):

$$J = \sqrt{\frac{\sigma a}{9\pi k T {n_{crit}}^{4/3}}} C_0 D \exp\left[\frac{-\Delta G_{crit}}{kT}\right] \quad (5)$$

where D is the diffusivity in cluster size space, and J is the nucleation rate. To determine the polymorph specific nucleation rate and solubility, the above set of equations require the estimation of the parameters σ, D, $\Delta\mu$, and n_{crit} for a given polymorph. To this end, a stochastic model is derived, which relates σ, D, and $\Delta\mu$ to particle size mean squared displacements (MSD), in the MSD model derivation section. In the temperature screening section, a computational methodology is discussed for determining n_{crit} as well as calculating σ, D, and $\Delta\mu$ through parameter estimation.

MSD model derivation

When determining diffusivity coefficients of critical nuclei clusters, the classical Einstein relationship will overestimate the diffusivity coefficient at long time scales, due to the presence of curvature in the free energy surface, which is not accounted for in the Einstein relationship.[35] The classical Einstein diffusivity for critical nuclei clusters is shown in eqn (6), where the MSD is predicted to increase linearly in time

$$\langle (n(t) - n_{crit})^2 \rangle = 2tD \quad (6)$$

where $n(t)$ is the cluster size at time t. To modify eqn (6) to account for the free energy curvature induced drift, we build upon the work of Knott, Duff, Doherty, and Peters[35] to derive a MSD equation that accounts for the curvature of the free energy surface, as well as deviations in the initial condition critical size estimate, n_{crit}, from the true critical size. The latter was deemed important, as the estimated critical temperature from the seeded technique will only be approximate, and accounting for this uncertainty in the critical size and temperature estimate will yield better γ, D, and $\Delta\mu$ estimates, and hence more accurate solubility and nucleation rate predictions.

Following the spirit of Knott, Duff, Doherty, and Peters,[35] we begin by Taylor series expanding the free energy expression as a function of cluster size around our n_{crit} estimate, as shown in eqn (7).

$$\beta G(n) = \beta G(n_{crit}) + \varepsilon(n - n_{crit}) - \tfrac{1}{2}\omega^2(n - n_{crit})^2 \quad (7)$$

where β is $1/kT$, ε is the first derivative of βG evaluated at n_{crit}, and ω^2 is the second derivate of βG evaluated at n_{crit}. Again, we stress that this n_{crit} is our critical size estimate from the temperature screening, and need not be the true critical size. Hence, we include the first derivative in the free energy expansion.

Taking the second derivative of the CNT expression, ω^2 is directly related to the interfacial free energy *via* eqn (8).

$$\omega^2 = \frac{2}{9}\beta a \sigma {n_{crit}}^{4/3} \quad (8)$$

The first derivative is directly related to the chemical potential and interfacial free energy as follows in eqn (9).

$$\varepsilon = \beta\left[-\Delta\mu + \frac{2}{3}\sigma a {n_{crit}}^{-1/3}\right] \quad (9)$$

The diffusion along the free energy surface can be modeled with a Smoluchowski equation, which tracks the time evolution of the probability density function in cluster size space, given that system originated at n_{crit} at time zero. Treating n as a continuous variable, the Smoluchowski equation is given in eqn (10):

$$\frac{\partial p(n,t)}{\partial t} = \frac{\partial}{\partial n}\left[D\frac{\partial \beta G(n)}{\partial n}p(n,t) + D\frac{\partial p(n,t)}{\partial n}\right] \quad (10)$$

where the above equation is subject to the initial condition $p(n,0) = \delta(n - n_{crit})$, and where delta is the Dirac delta function.

The Smulochowski equation can be nondimensionalized using dimensionless variables $z = \omega(n - n_{crit})$ and $\tau = \omega^2 Dt$.[35] Using the Taylor series expansion for $\beta G(n)$, along with the nondimensionless variables, yields the partial differential equation given in eqn (11).

$$\frac{\partial p(z,\tau)}{\partial \tau} = -p + (\varepsilon - z)\frac{\partial p}{\partial z} + \frac{\partial^2 p}{\partial z^2} \quad (11)$$

A moment generating function solution can then be used to generate an ordinary differential equation for the moments. The solution for the second moment is given in eqn (12).

$$\langle z^2 \rangle = e^{2\tau} - 1 + \varepsilon^2(e^{\tau} - 1)^2 \quad (12)$$

Transforming back to (n,t) yields eqn (13) form for the second moment.

$$\left\langle (n - n_{crit})^2 \right\rangle = \frac{e^{2\omega^2 Dt} - 1 + \varepsilon^2\left(e^{\omega^2 Dt} - 1\right)^2}{\omega^2} \quad (13)$$

Therefore, we see that the MSD of critical nuclei clusters is intimately linked with σ through ω^2 and eqn (8), D, and $\Delta\mu$. Specifically, a parameter fit of the MSD data to the above model allows a direct calculation of D, σ, and $\Delta\mu$. To generate MSD data emanating from critical nuclei clusters, a computational MD procedure is employed, and is discussed in the seeded cluster simulation strategy section.

Seeded cluster simulation strategy

In order to calculate the relevant parameters for a polymorph specific nucleation rate and solubility, we employ an MD seeded cluster nucleation simulation strategy, in combination with the derived stochastic model for nucleation discussed in the previous section. For all MD computational details employed, see the force field selection and MD computational

details section in the ESI.† The seeded cluster simulation strategy requires determining the critical, *i.e.* melting, temperature of an embedded spherical, polymorph specific, seed cluster through temperature screening. The critical temperature is the temperature at which the embedded nanocrystal has a 50 percent chance of growing or dissolving by definition. At the critical temperature, the crystal is in unstable equilibrium at the top of the CNT free energy barrier, and corresponds to n_{crit} in eqn (4). To determine this temperature, many trajectories are launched from a polymorph specific seed crystal, at a fixed temperature, to determine if the crystal grows or dissolves on average. The temperature is systematically varied to find the temperature at which the crystal has an equal probability of growing or dissolving. Through temperature screenings, the nanoscale stability rankings of polymorphs are determined directly by ascertaining which polymorph has the highest melting temperature for a given crystal size. Once the critical temperature is determined for each polymorph, embarrassingly parallel MD trajectories are launched from clusters of size n_{crit}, for MSD data generation, at the top of CNT barrier. The stochastic model links the polymorph specific parameters in the CNT free energy and rate expression directly to the mean squared displacement (MSD) data generated by MD. These CNT parameters are determined through parameter estimation. This makes performing costly thermodynamic integration calculations to determine $\Delta\mu$ potentially unnecessary.[21,22] The polymorph specific chemical potential, interfacial free energy, and critical are then used to determine both the bulk solubility and nucleation kinetics of a specific polymorph. Fig. 1 provides a pictorial flowchart outlining the method.

A similar seeded cluster simulation methodology has been recently compared against forward flux sampling,[35–39] umbrella sampling,[40–46] and brute force MD simulations for four different systems:[18] mW water, Tosi-Fumi NaCl, Lennard-Jones, and hard spheres. The results showed that the seeded cluster simulation methodology produced nucleation rate and interfacial energy predictions in good agreement with more rigorous biased sampling methods. This success of prior seeded cluster simulations motivated this work, as the seeded cluster strategy enables nucleation simulations over a broad range of supersaturations, allowing the kinetic screening of polymorphs to be performed over a wide temperature window. However, the proposed methodology is axiomatic in that it assumes the free energy surface to be correctly modeled by the CNT free energy expression, which could prove troublesome for systems that follow a two-step nucleation pathway.[47–51] However if a two-step free energy expression was found to be correct, a new polymorph specific MSD relation could be derived using the method outlined in this work and by Knott, Duff, Doherty, and Peters.[35] The material variables D, σ, and $\Delta\mu$, could then be calculated through parameter estimation using the two-step nucleation free energy MSD model. As the system sizes required to perform these nucleation simulations are well outside the range of accessibility for quantum mechanical density functional theory (DFT) calculations, classical molecular force fields must be used. Therefore, even if a system follows a CNT pathway, a molecular force field

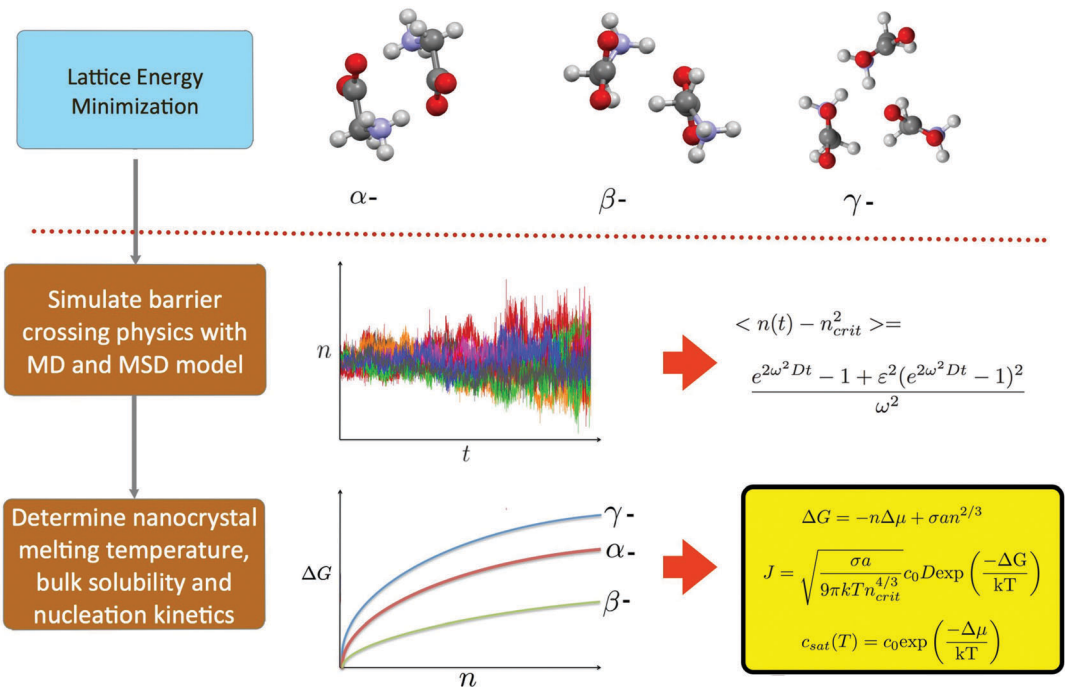

Fig. 1 Methodology overview for polymorph prediction. Step 1 is the traditional lattice energy minimization procedure to generate low energy polymorph candidates. Step 2 is simulating the nucleation physics using MD, for each polymorph under analysis, using the seeded cluster simulation strategy, in combination with the derived MSD model. Step 3 is estimating the polymorph specific nucleation kinetics and solubility. This allows the determination of which polymorph nucleates, and which polymorph is the equilibrium form in the specified solvent at bulk length scales.

that accurately reproduces the solubility, as well as solid–liquid interfacial energy, for each polymorph is required to produce a correct prediction. These polymorph specific variables are not known *a priori* for a given force field. Thus, the scientist runs the risk of performing costly simulations to only determine that the force field is incapable of simulating crystallization,[52] or potentially predicting incorrect forms due to force field inaccuracies. Next generation force fields that use AI trained on DFT data to generate system specific force field parameters may alleviate this risk in the future.[53–55] In total, the proposed methodology should be most applicable for systems that (1) follow a CNT pathway, and (2) have a force field that produces the correct polymorph ranking of solubility and interfacial tension.

Results

The derived MSD model and seeded cluster simulation strategy provide a framework for the prediction of polymorphs based off crystallization unit operation variables, such as temperature, crystal size, and solvent. It is shown in the following paragraphs and methods section that with a suitable force field, the proposed methodology is capable of capturing a wealth of physics unreferenced in traditional lattice energy minimization for glycine. Specifically, it will be shown that the β-glycine nanocrystals simulated in this work possess elevated melting temperatures and nucleation rates relative to α- and γ-glycine, reflecting the thermodynamic and kinetic preference for β-glycine nanocrystals. Furthermore, γ-glycine critical clusters are shown to be uncompetitive in the nucleation process, as they possess melting temperatures up to 20 K lower than either α- or β-glycine. Contrary to the nanoscale polymorph prediction, it will be shown correctly that β-glycine bulk crystals are more soluble than α-glycine in water, thus yielding the prediction of α-glycine at bulk length scales. This size dependent polymorphism riddle will be resolved through determination of σ, D, and $\Delta\mu$ for both polymorphs.

Temperature screening and melting temperature determination

To perform temperature screening of polymorphs, spherical nanocrystals of the α-, β-, and γ-glycine polymorphic forms were prepared by cutting them out of their respective bulk structures, obtained from the Cambridge Structural Database.[56] Spherical crystals with radii of 1.5 nm, 1.85 nm, 2.0 nm, 2.15 nm, and 2.70 nm for α- and β-glycine, and radii of 1.85 nm, 2.0 nm, 2.15 nm for γ-glycine, were studied. The computational details of the temperature screening procedure are discussed in the temperature screening computational methodology section of the ESI.† The average nucleus size drift data at the critical temperature is shown for all polymorphic forms in Fig. 2a–c. The final critical sizes are plotted as a function of the critical temperature in Fig. 2d.

For a fixed particle size, analysis of the nanocrystal melting temperatures leads to the CNT critical cluster stability ranking of β- > α- > γ-. Whereas both β- and α- appear to be competitive in terms of stability, γ-glycine clusters show an approximately 15–20 K lower melting point temperature relative to α- and β-. Furthermore, γ-glycine melting temperatures are shown to vary little with size, in comparison to the other polymorphs. This makes γ-glycine increasingly inaccessible, from a thermodynamic and kinetic perspective, relative to α- and β-, as the critical size, or equivalently temperature at fixed concentration, increases. These results predict that the glycine homogenous nucleation mechanism does not proceed through γ-glycine, as this would require the formation of highly unstable transition state complex (critical nuclei cluster). This would suggest that the fact that γ-glycine does not appear at bulk length scales upon batch crystallization in water emanates from the kinetic inaccessibility of the γ-glycine nucleation pathway. Finally, the prediction of β- > α- > γ- is in direct agreement with nanoconfinement crystallization experiments, which observed that β-glycine was the thermodynamically preferred form below pore sizes of 24 nm,[9] a size much larger than the particle sizes simulated here, showing that the proposed methodology is capable of predicting the correct polymorphic ranking of nanocrystals in solvent.

Contrary to what has been suggested in the literature,[57] we find no correlation between particle stability and the size of the diffusive interface. Although γ- is the least stable nanocrystal, and develops the smallest diffusive interface, β- develops a smaller diffusive interface than α-, and yet is found to be more stable. To the contrary, the smaller diffusive interface of β- results in smaller critical nuclei clusters for a given temperature, which is shown to magnify the nucleation rate of β- relative to α- in the polymorph specific nucleation rate estimation section.

MSD analysis and CNT parameter determination

For MSD analysis, where the polymorph specific D, ω^2, and $\Delta\mu$ parameters are determined through parameter fitting to the MSD model, the α-glycine clusters with 415 molecules at 295 K, and the β-glycine clusters with 369 molecules at 297.5 K, were selected for MSD data generation. These clusters were chosen as these critical temperatures corresponded approximately to the midpoint of the calculated critical temperatures. Furthermore, lower temperatures are preferred due to the decreased solubility, which minimizes possible spurious finite size effects due to solute depletion.[19] MSD analysis was not performed for γ-glycine, given the large computational load associated with performing MSD analysis, and that the polymorph had already been deemed thermodynamically uncompetitive for nucleation through the temperature screening simulations. The methodology employed for parameter estimation and MSD curve generation is discussed in the MSD analysis and parameter estimation methods section of the ESI.† The final MSD curves and model fits for both polymorphs are shown in Fig. 3a and b. In the case of α-glycine, the final parameter estimates are 1.77×10^{-4} fs^{-1} for D, 4.87×10^{-4} for ω^2, and 0.36 kcal mol^{-1} for $\Delta\mu$, with a model fit of $r^2 = 0.99$. In the case of β-glycine, the final parameter estimates are 2.38×10^{-4} for D, 3.27×10^{-4} for ω^2, and 0.21 kcal mol^{-1} for $\Delta\mu$ with a model fit of $r^2 = 0.98$. For both polymorphic forms, excellent r^2 values are obtained, showing the ability of the model

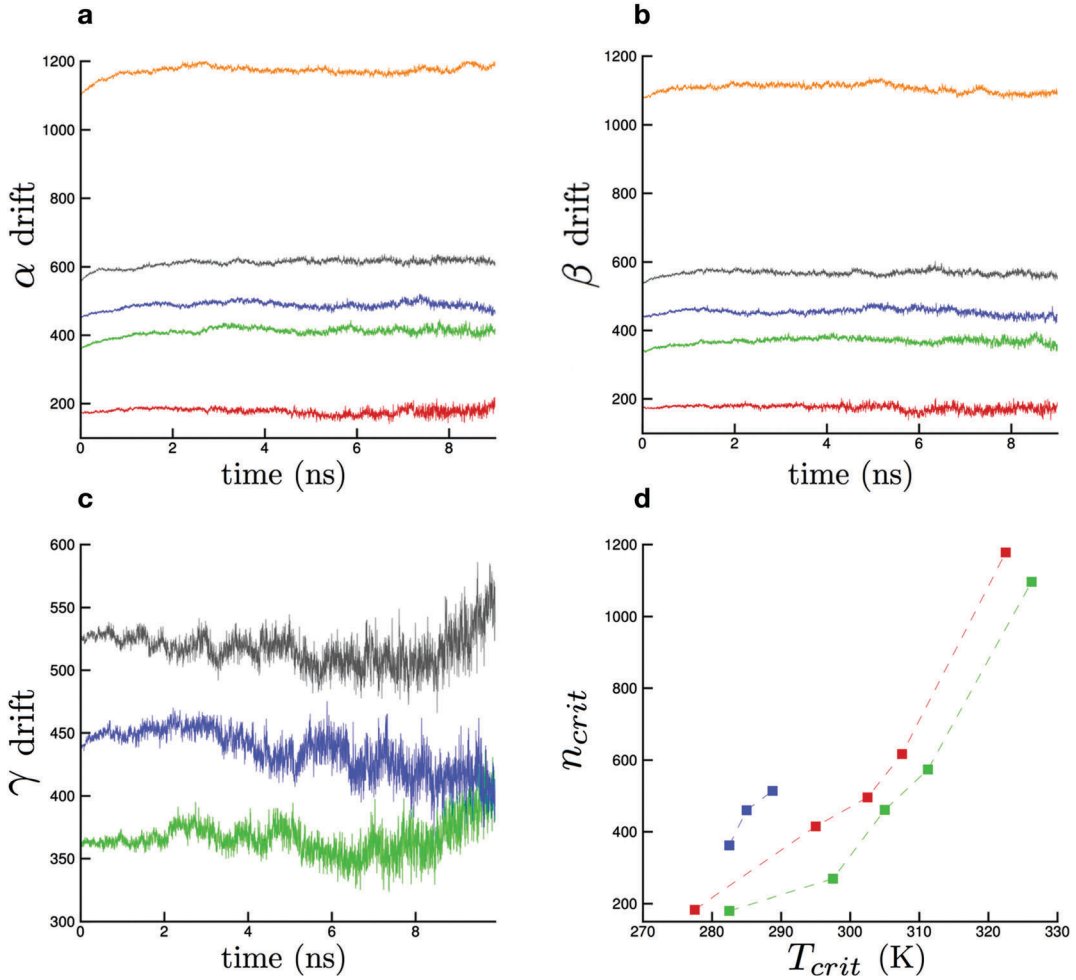

Fig. 2 Nucleus size drift data for α- (a), β- (b), and γ-glycine (c) at the determined critical temperatures demonstrating the zero drift. Plot (d) contains the final critical size estimate and temperature for every initial seed particle and polymorph. Estimated critical sizes are determined *via* linear fits to the drift data at the critical temperature after interface equilibration. For plots (a–c), red corresponds to the initially 1.5 nm particle, green to the initially 1.85 nm particle, blue to the initially 2.0 nm particle, grey to the initially 2.15 nm particle, and orange to the initially 2.70 nm particle. For plot (d), blue points correspond to melting temperatures and critical sizes for γ-, red for α-, and green for β-glycine. Dashed lines are only meant to guide the eyes of the reader.

to fit the simulated physics. In this work, and in CNT, it is assumed that ω^2 and $\Delta\mu$ are size and temperature independent variables. Thus, it is not necessary to simulate the number of trajectories required for full MSD analysis for every critical cluster studied in this work. The diffusivities of the remaining particles are calculated using the classical Einstein relation, which requires significantly shorter lengths of simulation and number of trajectories, and are contained in Table 1. The MSD data for Einstein diffusivity coefficient determination is plotted in Fig. S3 and S4 in the ESI.† The details of the Einstein relation diffusivity calculation are discussed in the MSD analysis and parameter estimation section in the ESI.† As remarked by Knott, Duff, Doherty, and Peters,[35] the Einstein diffusivity coefficient predictions are consistently above the values predicted by the MSD model.

Using eqn (8), the interfacial free energy, σ is calculated to be 31.0 mJ m^{-2}, for α-glycine, and 17.97 mJ m^{-2}, for β-glycine. As the stability of nanocrystals is dictated increasingly by the interfacial free energy, due to a large surface area to volume ratio, these interfacial free energies rationalize the nanocrystal stability ratio rankings of polymorph β- > α-, obtained from the temperature screening analysis, and hence provide physical insight into what drives the stability of the β-glycine polymorph, at the nanoscale, in water. Contrary to interfacial free energy ranking, the estimated values of $\Delta\mu$ predict that α-glycine is thermodynamically preferred over β-glycine in the absence of interfacial free energy contributions at bulk length scales. This result resolves the riddle of why β-glycine nanocrystals are thermodynamically preferred over α-glycine nanocrystals, and yet α-glycine becomes the exclusive polymorph formed from batch crystallization. These results provide the prediction of the size dependent polymorphism of glycine.

Polymorph specific solubility estimation

The bulk solubility of a polymorph, $c_{\mathrm{sat}}(T)$, which is related to $\Delta\mu$ *via* eqn (2), is the correct solvent dependent thermodynamic variable, not merely lattice energy, for determining which polymorph is most stable in a given solvent for micron size crystals (negligible surface area to volume ratio). As seen in the

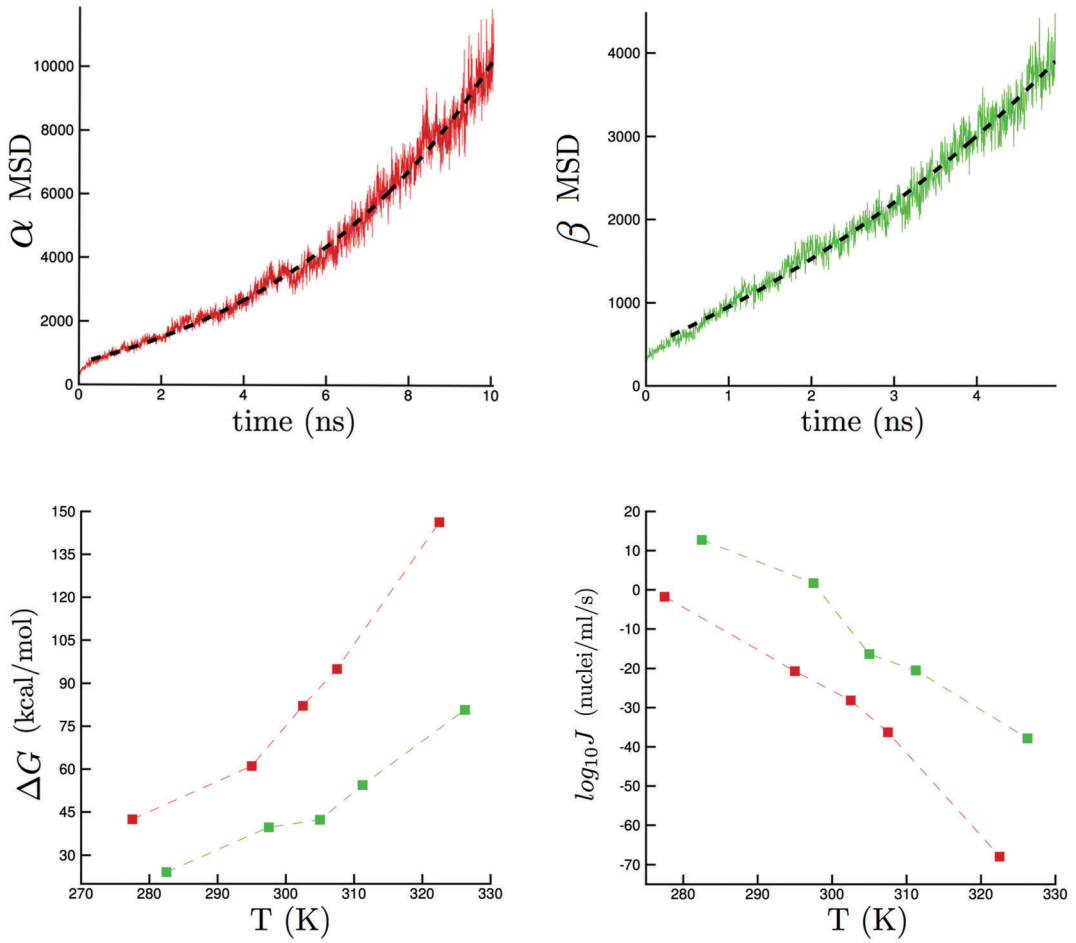

Fig. 3 (a) α-Glycine MSD (molecules^2) *vs.* simulation length (ns) curve (red) along with model fit (dashed black). (b) β-Glycine MSD (molecules^2) *vs.* simulation length (ns) curve (green) along with model fit (dashed black). (c) ΔG_{crit} (kcal mol^{-1}) of nucleation for α-(red) and β-(green) *vs.* temperature (K) (d) $\log_{10} J$ *vs.* temperature for α-glycine (red) and β-glycine (green). Dashed lines in plots (c) and (d) are only meant to guide the eyes of the reader.

Table 1 Cluster diffusivities obtained from Einstein relationship for particles not studied *via* MSD model analysis at each critical temperature. Diffusivities were calculated from linear fits over a 600 ps window

Polymorph	Critical temp. (K)	Diffusivity × 10^{11} (s^{-1})
α-	277.5	6.32
α-	302.5	5.75
α-	307.5	7.42
α-	322.5	8.29
β-	282.5	9.37
β-	305	5.75
β-	311.25	7.42
β-	326.25	8.29

previous section, the polymorph specific $\Delta\mu$ is parameter fitted directly from the MSD analysis for one of the critical temperatures. For the remaining particles, for which only temperature screening was performed, the exact critical size approximation is used to estimate $\Delta\mu$. At the critical size, CNT states that the chemical potential is related to the interfacial free energy through eqn (14).

$$\Delta\mu = \frac{2\gamma a}{3} n_{crit}^{-1/3} \quad (14)$$

Using the interfacial free energy estimated from the MSD analysis, and the n_{crit} estimates from the temperature screening, $\Delta\mu$ can be calculated at each temperature. From the MD data directly, the concentration of glycine in water, c_0, can be calculated. The solubility can then be directly calculated through eqn (2). In Table 2, we report the predicted and measured solubility values, along with the percent error in the prediction, for all the conditions tested in this study. All measured solubility values reported were taken from the work of Wang, Yang and Ching.[58] Because the experimental values were not available at the exact temperatures simulated, a linear fit to the experimental data was used to estimate the α-glycine solubility at the desired temperature for comparison.

The predicted α-glycine solubility agrees with the experimental values to within 23% error, with 4 of the 5 predicted solubility values being less than 10% error. Among the predicted solubility values, the 295 K α-glycine solubility predictions agree best with the experimental solubility data. We hypothesize that this the quality agreement for the 295 K simulation is due to the fact that this simulation yielded the interfacial free energy and chemical potential estimate. We attribute the increased percent error of 23% at 277.5 K in the solubility prediction to not

Table 2 MD simulation critical and experimental temperature (T), simulation concentration (c_0), simulation predicted chemical potential ($\Delta\mu$), simulation supersaturation (c_0/c_{sat}), simulation predicted solubility (c_{sat}), Experimentally-measured solubility (c_{sat}) (Adapted with permission from Yang *et al.*[58] Copyright (2008) American Chemical Society), and % error in solubility prediction, defined as |100 × (predicted − measured)/measured|. Cells marked ~ correspond are due to no existing experimental data

Polymorph	T (K)	c_0 (g ml^{-1})	$\Delta\mu$ (kcal mol^{-1})	c/c_{sat}	Predicted c_{sat} (g ml^{-1})	Measured c_{sat} (g ml^{-1})	% error
α-	277.5	0.399	0.462	2.31	0.172	0.136	23
α-	295	0.414	0.360	2.04	0.224	0.210	0.7
α-	302.5	0.434	0.331	1.85	0.250	0.246	4.8
α-	307.5	0.432	0.308	1.73	0.261	0.271	9.1
α-	322.5	0.484	0.248	1.47	0.328	0.353	8.9
β-	282.5	0.399	0.269	1.61	0.247	—	—
β-	297.5	0.413	0.210	1.42	0.290	—	—
β-	305	0.421	0.196	1.38	0.304	—	—
β-	311.25	0.431	0.183	1.34	0.320	—	—
β-	326.25	0.480	0.147	1.25	0.383	—	—

accounting for the size dependence of the interfacial free energy and neglecting the interfacial free energy dependence on temperature. A larger value of the interfacial energy for the 180 molecule α-glycine cluster at 277.5 K would result in a solubility prediction with reduced error. However, the small percent errors in the solubility estimates suggest that variation in the interfacial free energy with size and temperature is small over the simulations performed. If it were necessary to reduce the error further, performing MSD analysis for each critical temperature and cluster could be performed if the computational resources are available.

To determine if the simulations performed were sufficient to extrapolate the predicted solubility values for each polymorph over the temperature range of 278–323 K performed experimentally, a linear fit analysis was performed. The line fit for the predicted α-glycine solubility has an R^2 value above 0.99, and is given by $c_{sat}(T) = 3.4 \times 10^{-3} \times T - 0.7787$. The line fit for the predicted β-glycine has a R^2 value of 0.99 and is given by $c_{sat}(T) = 3.0 \times 10^{-3} \times T - 0.6131$. The percent error in the linear fit extrapolation α-glycine solubility prediction is given in Fig. 4a. The linear fit extrapolation solubility values for α- and β-glycine, along with the experimentally measured α-glycine solubility values, are given in Fig. 4b. Over the 45 K temperature range, the maximum percent error in the α-glycine prediction is only 19%, and has a minimum percent error of 0.8%. The solubility of β-glycine is not measurable in water, due to rapid recrystallization kinetics to α-glycine.[59] However, MD allows for β-solubility prediction. In agreement with the experimentally observed recrystallization kinetics,[59] it is shown that the predicted solubility of β-glycine is higher than the predicted α-glycine solubility over the entire temperature range. From a polymorph prediction perspective, we would predict correctly that the stability ranking of the bulk glycine structures in water is α-> β- from the solubility prediction over the 45 K temperature range.

From Fig. 4, it is shown that the MD data underestimates the experimental solubility *vs.* temperature slope for α-glycine. The vant Hoff equation, given in eqn (15), provides the differential

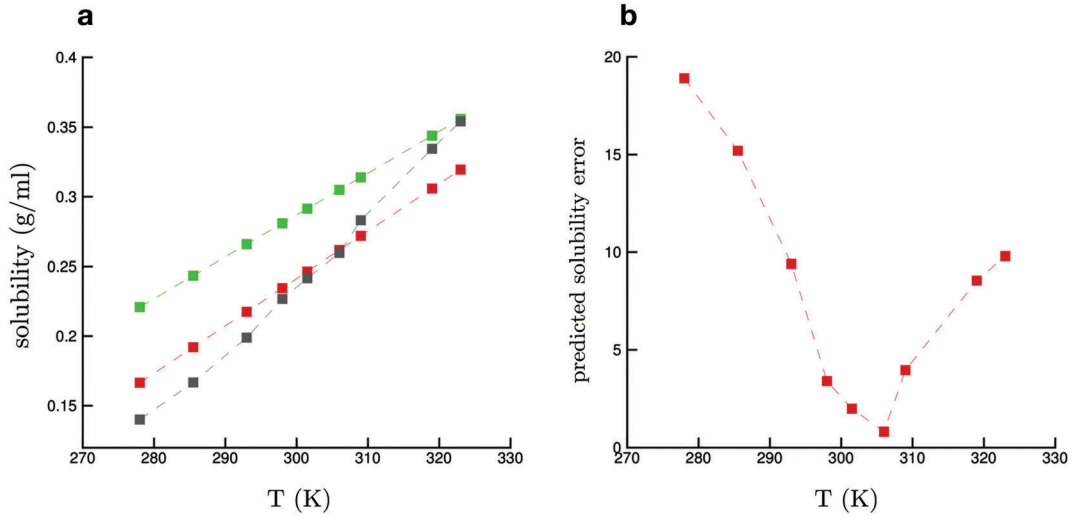

Fig. 4 (a) Percent error in linearly extrapolated α-glycine solubility predictions *vs.* temperature (K) (b) predicted α-glycine solubility (red), measured alpha glycine solubility (grey) (adapted with permission from Yang *et al.*[25] Copyright (2008) American Chemical Society), and predicted beta glycine solubility (green) curve *vs.* temperature (K). Predicted points are estimated through a linear fit extrapolation to solubility predictions from the MD simulations. Linear fit solubility prediction successfully predicts β-glycine (green) to be more soluble than the predicted α-glycine (red) over the entire temperature window. Dashed lines are only meant to guide the eyes of the reader.

Table 3 Free energy of critical nuclei clusters, ΔG_{crit} (kcal mol^{-1}), and $\log_{10}$ of nucleation rates, J (nuclei per ml per s)

Polymorph	Critical temperature (K)	ΔG_{crit} (kcal mol^{-1})	$\log_{10} J$ (nuclei per ml per s)
α-	277.5	42.25	−1.77
α-	295	69.38	−20.68
α-	302.5	82.13	−28.15
α-	307.5	95.00	−36.29
α-	322.5	146.20	−67.94
β-	282.5	24.19	12.74
β-	297.5	39.64	1.68
β-	305	45.29	−1.29
β-	311.25	52.42	−5.54
β-	326.25	80.72	−22.79

equation relating solubility, temperature, and the enthalpy of solvation:

$$\frac{d \ln S}{dT} = \frac{\Delta H_{solv}}{RT^2} \quad (15)$$

with R being the ideal gas constant, T the temperature, ΔH_{solv} the enthalpy of solvation, and S the solubility. In the integrated form of the vant Hoff equation, given in eqn (16), the enthalpy of solvation can be directly calculated from the solubility values at two distinct temperatures.

$$\Delta H_{solv} = \frac{T_1 T_2}{T_1 - T_2} R \ln\left(\frac{S_2}{S_1}\right) \quad (16)$$

Using the 277.5 K and 322.5 K α-glycine solubility predictions, the predicted enthalpy of solvation is 10.64 kJ mol^{-1}. Using the experimentally measured values at the same temperatures, for direct comparison, yields an enthalpy of solvation value of 17.3 kJ mol^{-1}. This misrepresentation of the enthalpy of solvation is likely a result of employing a classical molecular force field, and rationalizes the disagreement between the slopes of the predicted and measured α-glycine curves.

Polymorph specific nucleation rate estimation

The remaining unreported parameter necessary for the polymorph specific nucleation rate (J) estimate is the nucleation free energy barrier. Using the determined parameters (n_{crit}, $\Delta\mu$, γ, D), the free energy barrier is calculated using eqn (3), and finally the rate is calculated using the CNT rate expression, eqn (5). The results are given in Table 3, and are shown graphically in Fig. 3c and d.

The calculated free energy barrier to nucleation for β-glycine is less than α-glycine by 20 to 60 kcal mol^{-1}, depending on the temperature. This serves as further direct numerical confirmation of the stability of the β-glycine clusters, relative to α-glycine. Comparing the respective nucleation rates of β- and α-glycine over the whole temperature range, the nucleation rate of β- is between 14 orders of magnitude faster at the lowest temperatures, to 60 orders of magnitude faster at the highest temperature, than α-glycine. This result of β-glycine nucleation at these length scales is again in agreement with the nanoconfinement crystallization experiments performed by Hamilton, Hillmyer, and Ward.[9] In their work, it was found that β-glycine was the form to appear, suggesting that the nucleation pathway proceeds through β-glycine formation. From the analysis performed in this work, a strong kinetic preference for β-glycine formation, relative to α-glycine, is predicted.

Conclusions

Current lattice energy minimization algorithms for polymorph prediction fail to capture any crystallization process parameters, either than possibly temperature and pressure, but remain an essential first step towards determining which polymorphic forms are energetically possible. To have a more robust prediction procedure, MD is required to further determine nucleation kinetics and solubility for the low energy structures at length scales ranging from the metastable liquid to the final bulk crystal as a function of the crystallization conditions employed, such as solvent and anti-solvent. In this work, MD based seeded cluster nucleation simulations were combined with a stochastic model to calculate polymorph specific nucleation kinetics, nanocrystal stability rankings, and solubility to augment current polymorph prediction algorithms. This method resulted in the successful prediction of β- at the nanoscale, α- at the bulk scale, and showed the kinetic inaccesability of γ-glycine critical clusters. Lattice energy minimization alone incorrectly predicts γ-glycine from the ranking of lattice energies. By predicting the correct polymorphic form both at the nanoscale and bulk scale, the results herein predict correctly the size dependent polymorphism of glycine.

Analysis of the nanocrystal melting temperatures revealed that γ-glycine nanocrystals possess melting temperatures 15–20 K lower than α- and β- nanocrystals of equivalent size, rendering it the polymorph with the least stable critical nuclei cluster. This suggests that the reason γ-glycine is not observed upon batch crystallization in water stems from the instability of the γ-glycine transition state complex during nucleation, rendering it uncompetitive kinetically in comparison to α- and β-glycine. The melting temperatures of the three polymorphs show β-glycine nanocrystals to be the preferred form at the nanoscale in aqueous solvent. This observation was subsequently rationalized through the results provided through the MSD model where β-glycine was predicted to have a lower interfacial energy (17.97 mJ m^{-2}) than α-glycine (31.0 mJ m^{-2}). The enhanced stability of the β-glycine clusters is shown to yield a strong kinetic preference for the formation of β-glycine, where free energy barriers to nucleation are between 20 to 60 kcal mol^{-1} lower for β-glycine than α-glycine The lower nucleation free energy barrier of β-glycine translates to the nucleation kinetics of β-glycine being 14 to 60 orders of magnitude faster than α-glycine over the temperature range simulated. These results are in agreement with nanoconfinement crystallization experiments, where β-glycine was observed exclusively below pore sizes of 24 nm upon evaporative crystallization in water.[9]

The bulk solubility of α- and β-glycine was predicted over a 45 K temperature range. A comparison with experimental

values show that the trend-line estimated α-glycine solubility values are within 19% error of the measured values. Furthermore, β-glycine is predicted to be more soluble in water at bulk length scales than α-glycine, over the entire temperature range. This provides the correct prediction that α-glycine is more stable than β-glycine at bulk length scales in aqueous medium.[59] Experimentally, the solubility of β-glycine is not measurable due to rapid recrystallization kinetics to α-glycine. MD however, provides for a prediction of the β-glycine solubility curve over a 45 K temperature range.

The nanocrystal melting temperatures, nucleation rates, and solubility results illustrate the complexity of the crystallization pathway, and the necessity of including crystallization unit operation variables, such as crystal size, temperature, and solvent into future polymorph prediction frameworks to answer the question "which polymorph will appear" and not merely "which polymorph has the most stable lattice energy." Furthermore, the phenomenon of size dependent polymorphism necessitates rigorous clarification of the length scale under study when we say "polymorph prediction." The results of this work demonstrate this necessity as we correctly predict (1) a bulk stability ranking of α- > β- from the solubility calculation, and (2) a nanocrystal stability ranking of β- > α- > γ- from the melting temperatures, and nucleation rate calculations.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

This study was sponsored by AbbVie Inc. through grant number 8000053025 and 8000069224. Hsien-Hsin Tung was an AbbVie employee at the time that he collaborated on this project. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. This work used the XSEDE Extended Collaborative Support Service (ECSS) program. The authors would like to acknowledge XSEDE and Purdue RCAC, as all simulations were performed on Stampede supercomputer at TACC and Conte supercomputer at Purdue. The authors were given access to Stampede supercomputer through XSEDE grant CHE150072. The authors would like to acknowledge the tremendous help provided by Lei Huang and Yang Wang at TACC, and Xiao Zhu at Purdue, for their assistance in parallelizing the MD platform on Stampede and Conte.

References

1 J. Maddox, *Nature*, 1988, **345**, 201.
2 S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.
3 J. D. Dunitz, *Chem. Commun.*, 2003, 545–548.
4 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. Van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. Van De Streek, T. S. Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2011, **67**, 535–551.
5 J. A. Chisholm, S. Motherwell, P. R. Tulip, S. Parsons and S. J. Clark, *Cryst. Growth Des.*, 2005, **5**, 1437–1442.
6 C. M. Freeman, J. W. Andzelm, C. S. Ewig, J. Hill and B. Delley, *Chem. Commun.*, 1998, 2455–2456.
7 N. Marom, R. A. Distasio, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz and A. Tkatchenko, *Angew. Chem., Int. Ed.*, 2013, **52**, 6629–6632.
8 A. M. Lund, G. I. Pagola, A. M. Orendt, M. B. Ferraro and J. C. Facelli, *Chem. Phys. Lett.*, 2015, **626**, 20–24.
9 B. D. Hamilton, M. A. Hillmyer and M. D. Ward, *Cryst. Growth Des.*, 2008, **8**, 3368–3375.
10 X. Yang and A. S. Myerson, *CrystEngComm*, 2015, **17**, 723–728.
11 J. E. Aber, S. Arnold, B. A. Garetz and A. S. Myerson, *Phys. Rev. Lett.*, 2005, **94**, 1–4.
12 C. Parks, A. Koswara, H.-H. Tung, N. Nandkishor, B. Shailendra, Z. K. Nagy and D. Ramkrishna, *Phys. Chem. Chem. Phys.*, 2017, submitted.
13 I. Weissbuch, V. Y. Torbeev, L. Leiserowitz and M. Lahav, *Angew. Chem. Int., Ed.*, 2005, **44**, 3226–3229.
14 E. Simone, A. N. Saleemi, N. Tonnon and Z. K. Nagy, *Cryst. Growth Des.*, 2014, **14**, 1839–1850.
15 M. R. A. Bakar, Z. K. Nagy and C. D. Rielly, *Org. Process Res. Dev.*, 2009, **13**, 1343–1356.
16 M. R. Abu Bakar, Z. K. Nagy, C. D. Rielly and S. E. Dann, *Int. J.* Pharm., 2011, **414**, 86–103.
17 C. Parks, L. Huang, Y. Wang and D. Ramkrishna, *Mol. Simul.*, 2016, submitted.
18 J. R. Espinosa, C. Vega, C. Valeriani and E. Sanz, *J. Chem. Phys.*, 2016, **144**, 34501.
19 N. E. R. Zimmermann, B. Vorselaars, D. Quigley and B. Peters, *J. Am. Chem. Soc.*, 2015, **137**, 13352–13361.
20 E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal and C. Valeriani, *J. Am. Chem. Soc.*, 2013, **135**, 15008–15017.
21 J. R. Espinosa, E. Sanz, C. Valeriani and C. Vega, *J. Chem. Phys.*, 2014, **141**, 1–21.
22 A. Zaragoza, M. M. Conde, J. R. Espinosa, C. Valeriani, C. Vega and E. Sanz, *J. Chem. Phys.*, 2015, **143**, 134504.
23 M. Lauricella, S. Meloni, N. J. English, B. Peters and G. Ciccotti, *J. Phys. Chem. C*, 2014, **40**, 22847–22857.
24 X.-M. Bai and M. Li, *J. Chem. Phys.*, 2006, **124**, 124707.
25 C. R. Locker, G. C. Rutledge and C. Link, *Macromolecules*, 2013, **11**, 4723–4733.
26 M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E6–14.
27 M. Salvalaglio, T. Vetter, F. Giberti, M. Mazzotti and M. Parrinello, *J. Am. Chem. Soc.*, 2012, **134**, 17221–17223.

28 F. Giberti, M. Salvalaglio, M. Mazzotti and M. Parrinello, *Chem. Eng. Sci.*, 2015, **121**, 51–59.
29 M. Shah, E. E. Santiso and B. L. Trout, *J. Phys. Chem. B*, 2011, **115**, 10400–10412.
30 E. E. Santiso and B. L. Trout, *J. Chem. Phys.*, 2015, **143**, 174109.
31 A. Gavezzotti, *CrystEngComm*, 2011, **13**, 3573.
32 B. Peters, *J. Chem. Phys.*, 2009, **131**, 244103.
33 A. R. Browning, M. F. Doherty and G. H. Fredrickson, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2008, **77**, 1–6.
34 C. S. Towler, R. J. Davey, R. W. Lancaster and C. J. Price, *J. Am. Chem. Soc.*, 2004, **126**, 13347–13353.
35 B. C. Knott, N. Duff, M. F. Doherty and B. Peters, *J. Chem. Phys.*, 2009, **131**, 224112.
36 R. J. Allen, C. Valeriani and P. Rein Ten Wolde, *J. Phys.: Condens. Matter*, 2009, **21**, 463102.
37 E. Sanz, C. Valeriani, T. Vissers, A. Fortini, M. E. Leunissen, A. van Blaaderen, D. Frenkel and M. Dijkstra, *J. Phys.: Condens. Matter*, 2009, **20**, 494247.
38 L. Filion, M. Hermes, R. Ni and M. Dijkstra, *J. Chem. Phys.*, 2010, **133**, 244115.
39 T. Li, D. Donadio, G. Russo and G. Galli, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19807.
40 J. Kästner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 932–942.
41 S. Auer and D. Frenkel, *Annu. Rev. Phys. Chem.*, 2004, **55**, 333–361.
42 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
43 R. Radhakrishnan and B. L. Trout, *J. Am. Chem. Soc.*, 2003, **2139**, 7743–7747.
44 P. Rein ten Wolde, M. J. Ruiz-Montero and D. Frenkel, *J. Chem. Phys.*, 1996, **104**, 9932.
45 S. Auer and D. Frenkel, *Nature*, 2001, **409**, 1020–1023.
46 C. Valeriani, E. Sanz and D. Frenkel, *J. Chem. Phys.*, 2005, **122**, 194501.
47 D. Erdemir, A. Y. Lee and A. S. Myerson, *Acc. Chem. Res.*, 2008, **6**, 3–4.
48 C. M. Völkle, D. Gebauer and H. Cölfen, *Faraday Discuss.*, 2015, **179**, 59–77.
49 A. Jawor-Baczynska, B. D. Moore and J. Sefcik, *Faraday Discuss.*, 2015, **179**, 141–154.
50 M. Vorontsova, D. Maes and P. G. Vekilov, *Faraday Discuss.*, 2015, **4**, 1166–1169.
51 A. Sauter, F. Roosen-runge, F. Zhang and G. Lotze, *Faraday Discuss.*, 2015, **179**, 41–58.
52 S. Banerjee and H. Briesen, *J. Chem. Phys.*, 2009, **131**, 184705.
53 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 1–4.
54 R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler and M. Parrinello, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 18–21.
55 J. Behler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930.
56 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.
57 N. Duff and B. Peters, *J. Chem. Phys.*, 2011, **135**, 134101.
58 X. Yang, X. Wang and C. B. Ching, *J. Chem. Eng. Data*, 2008, **53**, 1133–1137.
59 A. Bouchard, G. W. Hofland and G. J. Witkamp, *J. Chem. Eng. Data*, 2007, **52**, 1626–1629.