Predicting polymorphism in molecular crystals using orientational entropy

Pablo M. Piaggi[a,b,c] **and Michele Parrinello**[b,c,d,1]

[a]Theory and Simulation of Materials, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; [b]Facoltà di Informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland; [c]National Center for Computational Design and Discovery of Novel Materials (MARVEL), Università della Svizzera Italiana, CH-6900 Lugano, Switzerland; and [d]Department of Chemistry and Applied Biosciences, ETH Zurich, CH-6900 Lugano, Switzerland

Contributed by Michele Parrinello, August 14, 2018 (sent for review June 27, 2018; reviewed by Julian Gale and Valeria Molinero)

We introduce a computational method to discover polymorphs in molecular crystals at finite temperature. The method is based on reproducing the crystallization process starting from the liquid and letting the system discover the relevant polymorphs. This idea, however, conflicts with the fact that crystallization has a timescale much longer than that of molecular simulations. To bring the process within affordable simulation time, we enhance the fluctuations of a collective variable by constructing a bias potential with well-tempered metadynamics. We use as a collective variable an entropy surrogate based on an extended pair correlation function that includes the correlation between the orientations of pairs of molecules. We also propose a similarity metric between configurations based on the extended pair correlation function and a generalized Kullback–Leibler divergence. In this way, we automatically classify the configurations as belonging to a given polymorph, using our metric and a hierarchical clustering algorithm. We apply our method to urea and naphthalene. We find different polymorphs for both substances, and one of them is stabilized at finite temperature by entropic effects.

crystal structure prediction | polymorphism | enhanced sampling | molecular simulation | urea

Polymorphism is the ability that substances have to crystallize into different structures. A paradigmatic example is carbon that in its two main polymorphs, graphite and diamond, exhibits amazingly different properties. Polymorphism is also important from a practical point of view since controlling which crystal structure forms is of the utmost importance in many manufacturing processes. The pharmaceutical industry suffers in particular the consequences of polymorphism (1, 2). Active pharmaceutical ingredients are usually small, organic molecules that frequently exist in a plethora of crystalline forms. Different polymorphs can be patented separately and usually lead to different drug performances. Therefore a comprehensive screening of polymorphs is crucial to prevent a rival company from releasing to the market the same molecule in a different polymorph (3) and to anticipate the transformation of one polymorph into another during the manufacturing process or the shelf life (4).

The screening of polymorphs was traditionally performed experimentally despite the large costs involved (2). In the last 15 years, the increase in computer power and the development of algorithms able to screen a large number of polymorphs have led to very significant successes in polymorph prediction (5–9). Such methods are based on the search of local minima on the potential energy surface. The minima are ordered by energy and typically corrected for thermal effects, using the harmonic approximation. Typically these methods find hundreds of structures, most of which are thermodynamically unstable. In addition, entropic effects beyond the harmonic approximation can be significant. They can not only alter the delicate energetic balance between the different polymorphs but also even stabilize structures that are not local minima of the potential energy surface, for instance in the case of superionic conductors such as AgI (10). Another issue that is often overlooked is the kinetic side of crystallization; for instance a given polymorph can be favored relative to energetically lower ones by the fact that is kinetically more accessible. For all these reasons we take here a different approach and we try to reproduce on the computer the crystallization process starting from the liquid state and letting the system discover different polymorphs.

The above ambition conflicts with the fact that crystallization is a process that occurs on a timescale that is much longer than that of computer simulations. This requires the use of enhanced sampling methods that bring the timescale of crystallization within affordable simulation time (11). Some enhanced sampling methods require the definition of order parameters or collective variables (CVs). These methods channel and enhance the fluctuations to favor the reversible observation of multiple freezing and melting events. Thus far, such order parameters have been based on some structural geometrical information on the phase the system is going to crystallize into. If one is interested in discovering polymorphs, this approach defeats the purpose. Recently, however, we showed that in simple systems this can be circumvented by using as CVs surrogates of enthalpy and entropy (12). The idea is to mimic what happens in a real system in which there is a tradeoff between entropy and enthalpy. We dealt with simple one-component (12) or two-component (10) atomic systems. In this case the following expression is used as a surrogate for entropy,

$$S_2 = -2\pi\rho k_B \int_0^\infty \left[g(r)\ln g(r) - g(r) + 1\right] r^2\, dr, \qquad [1]$$

Significance

Small organic molecules often exhibit an amazing polymorphism. Since most drugs are based on organic molecules, this has important practical consequences. Not only does the deliverability of the drugs depend on the crystal structure but also different polymorphs can be separately patented. We address this problem by appropriately designed enhanced sampling simulations that start from the liquid and let the system crystallize spontaneously at the freezing temperature. In such a way entropy effects are automatically included. We successfully apply the method to the cases of urea and naphthalene and discover that entropy plays an important role.

Author contributions: P.M.P and M.P. designed research, performed research, analyzed data, and wrote the paper.

Reviewers: J.G., Curtin University; and V.M., The University of Utah.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: parrinello@phys.chem.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811056115/-/DCSupplemental.

Published online September 20, 2018.

where r is a distance, $g(r)$ is the radial distribution function, k_B is the Boltzmann constant, and ρ is the density of the system. Together with enthalpy, S_2 proved successful in predicting the lattice into which the system was going to crystallize. For a discussion of S_2 we refer the reader to ref. 12. This is a simple proof of principle to show that crystal structures, even the ones that are stabilized by entropy (10), can be predicted.

Molecular systems that are of interest to the pharmaceutical industry present a complexity much larger than that of the relatively simple systems so far handled in which most of the time only one polymorph was stable. Here we enlarge considerably the scope of these calculations and move to study molecular systems that, as we shall see, present a large number of polymorphs.

We consider a system of molecules and, for the purpose of developing a CV, we represent each molecule by the position of its center of mass and a vector that characterizes its orientation in space. We can define a correlation function $g(r,\theta)$ akin to $g(r)$ but including the relative orientation between two molecules. θ is defined as $\theta = \arccos\left(\frac{\mathbf{v}_i\cdot\mathbf{v}_j}{|\mathbf{v}_i||\mathbf{v}_j|}\right)$, where $\mathbf{v}_i$ and $\mathbf{v}_j$ are the vectors characterizing the orientation of molecules i and j. Statistical mechanics provide us with an expression for the entropy of such a system equivalent to the one in Eq. **1**; this is (13)

$$S_\theta = -\pi\rho k_B \int_0^\infty \int_0^\pi \left[g(r,\theta)\ln g(r,\theta) - g(r,\theta) + 1\right] r^2 \sin\theta \, dr \, d\theta. \quad [2]$$

We use S_θ as the CV to drive simulations. A similar CV was introduced in ref. 14 although in that case the probability as a function of the angle of the molecules with respect to a fixed reference frame was used to define the entropy.

An important part in the definition of our CV is the choice of angles to characterize the relative orientation between neighboring molecules. In principle, three angles are needed to specify completely the relative orientation between two rigid molecules, for instance the three Euler angles ϕ,θ,ψ. This would imply the construction of a function $g(r,\phi,\theta,\psi)$ whose calculation would be cumbersome. Here we take a different approach and we use several CVs each involving one angle.

We choose two systems to test the ability of S_θ to explore polymorphism, namely urea and naphthalene. We have chosen two CVs and therefore two angles for each system. In the case of urea we use the angles θ_1 and θ_2 to define the CVs s_{θ_1} and s_{θ_2}. The first one is defined using the direction of the dipole moment and the second one with the direction of the vector joining the two nitrogens. In the case of naphthalene we use the direction of the longest axis of the molecule and the direction perpendicular to the aromatic rings to define the CVs s_{θ_2} and s_{θ_1}. Other choices are possible, for instance the directions of the eigenvectors of the moment of inertia tensor. At variance with our previous work (12), here we do not bias enthalpy. This would add a third CV, making the calculation more cumbersome. Also, the scope here is different since our objective is to discover structures, and observing multiple and reversible transitions is less of an issue.

Since the use of $g(r,\theta)$ is not so widespread, we thought it useful to help the reader get a feeling of its behavior by plotting $g(r,\theta_1)$ for the liquid and polymorph I of urea at 450 K (Fig. 1). The liquid $g(r,\theta_1)$ exhibits some structure at very short distances and almost no correlations at distances larger than 0.8 nm. On the other hand the $g(r,\theta_1)$ of polymorph I shows a well-defined structure that persists at long distances as expected from a solid phase. As can be observed in Fig. 1, one of the main characteristics of polymorph I is that molecules have parallel or antiparallel dipole moments. Thus, $g(r,\theta)$ contains important orientational information that can help to distinguish between phases.

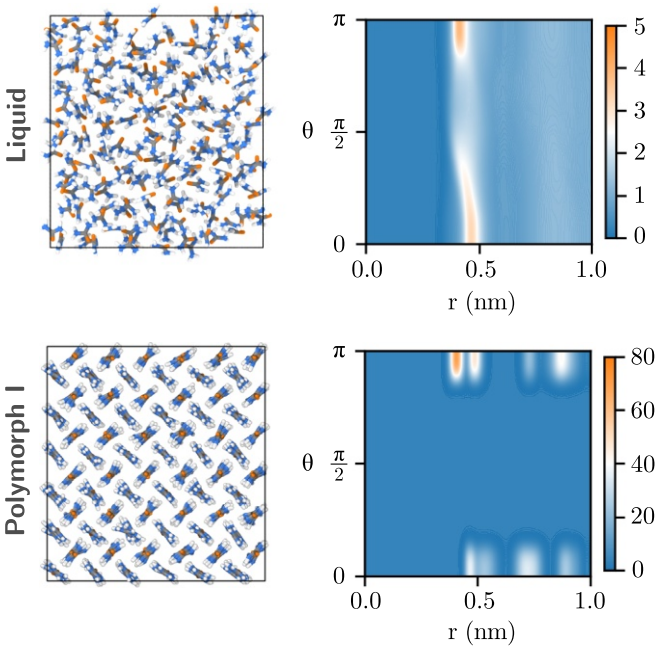

Fig. 1. $g(r,\theta)$ for the liquid and polymorph I of urea at 450 K. Snapshots of the system in each of the phases are shown. Polymorph I is viewed down the *c* axis. C, O, N, and H atoms are shown in orange, gray, blue, and white, respectively.

We briefly describe the polymorphs found experimentally so far for each system. Urea shows a rich polymorphism and up to five polymorphs have been reported (15–17). The most stable form at ambient conditions is form I and it has been extensively studied. Two other forms exist at higher pressures, namely forms III and IV. Another high-pressure polymorph, form V, has been found although to our knowledge the details of the structure have not been reported. There has also been theoretical work that found other polymorphs (18, 19). In particular, for urea as described by the Amber force field, the so-called form A (18, 19) is highly relevant, having an energy very close to that of the ground state. At variance with urea, naphthalene has only one solid form and despite several investigations at high pressure (20, 21) no more forms have yet been found.

We used well-tempered metadynamics (WTMetaD) (22) to enhance the fluctuations of s_{θ_1} and s_{θ_2}. In WTMetaD a time-dependent potential is constructed as a sum of kernels, typically chosen to be Gaussians. The potential discourages frequently visited configurations, thus boosting the exploration of configuration space. We simulated urea at 450 K and naphthalene at 300 K. Both temperatures were chosen close to the melting point of each substance. Further details can be found in *Materials and Methods*. In the 200-ns biased simulations both urea and naphthalene explore thoroughly the space spanned by the CVs, although understanding the nature of the configurations explored requires further analysis. A visual inspection of the trajectories shows many transitions to different crystal forms. The crystalline configurations have different orientations in space and some of them contain small crystalline defects. The wealth of information that these simulations contain, however, cannot be analyzed with the naked eye. It would therefore be useful to have an automatic method to identify and classify the polymorphs that crystallize in the course of the simulation. In the following paragraphs we propose one such automatic method.

A key ingredient for an automatic method to identify and classify polymorphs is a metric for the similarity between two given configurations. Several structural similarity metrics exist in the literature (23) but in this work we propose another one. In the present context, it is natural to use for this purpose the very function $g(r,\theta)$ that defines the CVs to characterize the configuration of the system. However, we still need a measure of distance between two $g(r,\theta)$s. We can define a distance by taking inspiration from the pair entropy expression. We first note that Eq. **2** is a measure of the distance between the $g(r,\theta)$ of the present configuration and the $g(r,\theta)$ of the ideal gas; i.e., $g(r,\theta)=1\,\forall\, r,\theta$. Inspired by this observation we introduce a divergence of $g_1(r,\theta)$ with respect to $g_2(r,\theta)$,

$$D(g_1||g_2)=\int_0^\infty\int_0^\pi\left[g_1(r,\theta)\ln\frac{g_1(r,\theta)}{g_2(r,\theta)}-g_1(r,\theta)+g_2(r,\theta)\right]\times r^2\sin\theta\,dr\,d\theta. \quad [3]$$

This is a generalization of the Kullback–Leibler divergence for nonnormalized functions. This divergence is a special case of Bregman divergence and has some interesting properties such as that of being convex and having a minimum at $g_1=g_2$ (24). Strictly speaking, $D(g_1||g_2)$ is not a distance since it is not symmetric. For applications in which a well-defined distance is needed we use a symmetrized version of Eq. **3**, namely,

$$d(g_1,g_2)=\frac{D(g_1||g_2)+D(g_2||g_1)}{2}. \quad [4]$$

Equipped with this metric, we can compare configurations and analyze the rich and complex trajectories resulting from the biased simulations. We exemplify our approach by analyzing the trajectory of urea. The configurations in the trajectory were clustered using a hierarchical clustering approach (25, 26) based on the distance defined in Eq. **4**. We used the average distance between points in two clusters as a linkage criterion. As a result of the clustering, we obtain a tree diagram (Fig. 2) that shows the similarity between different configurations in the trajectory. We can now choose a threshold distance d_c and join together all configurations that belong to a branch with maximum distance d_c between configurations. The choice of d_c allows us to focus on the dominant structures that appear in the simulation. In Fig. 2 d_c is shown with a dashed gray line and the resulting clusters of structures are shown with different colors.

We still have to determine the phases that each cluster represents. A possible way to do so is by choosing the minimum energy configuration within each cluster. This configuration will be the one with the least number of defects and less affected by the thermal motion of molecules. In some cases this approach is not appropriate, for instance when structures are stabilized by large entropic effects. In these cases one can choose the configuration that has an energy close to the average energy of the cluster. We have chosen with this criterion the configurations that are used to determine the nature of each cluster. Some of these configurations are shown in Fig. 2.

We now describe the phases that were found. The tree diagram has two main branches. The right branch contains liquid-like configurations (violet cluster in Fig. 2) and interesting partially ordered configurations (brown cluster in Fig. 2) in which the dipole moments are oriented in the same direction but do not exhibit long-range translational order. The left branch contains solid-like configurations and it can be further subdivided into five relevant clusters. One of these clusters contains an unstable structure and we disregard it (gray cluster in Fig. 2). The other four clusters correspond to form I, to a polymorph that we name form B, to form IV, and to form A. Form B is reported in this paper. The other structures were expected based on previous studies (16, 18, 19). All polymorphs are metastable at 450 K and they do not transform during a 1-ns unbiased simulation. We include the configurations of all relevant structures in Datasets S1–S3. We also performed a similar analysis for naphthalene. The clustering identifies the experimentally known form I, the liquid, and a structure that we name form A. The results can be found in *SI Appendix*.

We have estimated the free energy difference between the polymorphs and the liquid using

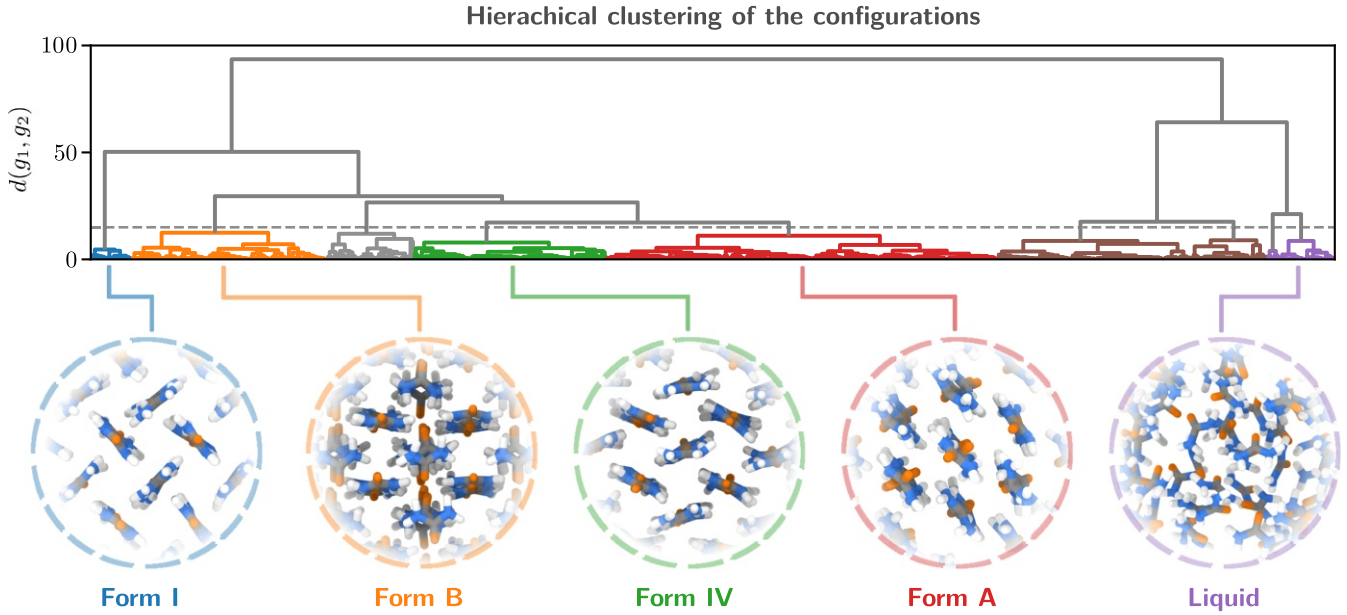

Fig. 2. Tree diagram resulting from the clustering according to the distance in Eq. **4** of the trajectory of urea at 450 K. The threshold distance used to join clusters is shown with a gray dashed line. Configurations at 450 K for selected clusters are shown.

$$\Delta G = -\frac{1}{\beta}\log\left(\frac{p_i}{p_l}\right), \quad [5]$$

where p_i and p_l are the probabilities to observe polymorph i and the liquid, respectively. The states that we consider in these calculations are either purely solid or purely liquid. In an unbiased molecular dynamics (MD) simulation one could calculate the probabilities p_i and p_l directly from the simulation. However, since we have introduced the WTMetaD potential that alters the probability of observing a given configuration, the p_is must be calculated with the reweighting procedure described in ref. 27. We used the clustering described above to identify the phase of each configuration. The resulting free energy differences are shown in Fig. 3. The error bars in Fig. 3 are relatively large since free energy differences calculated in this way are not easy to converge and the simulation contains transitions between many different structures. Fig. 3 shows also the enthalpy ΔH of the polymorphs with respect to the liquid phase. Using ΔG and ΔH the entropy ΔS can be calculated from the definition of free energy $\Delta G = \Delta H - T\,\Delta S$. The results show that form I of urea is close to equilibrium with the liquid at 450 K, in line with refs. 18 and 28 where the melting temperature was found to be around 420 K. Similarly, form I of naphthalene is close to equilibrium with the liquid at 300 K, as expected from the estimated melting temperature (330 K).

We now consider in detail the newly discovered polymorphs. We first discuss form B of urea that has a $P4_2/mbc$ space group and is shown in Fig. 4. This polymorph is particularly interesting because it has a relatively high enthalpy, roughly k_BT above form I (Fig. 3). Based only on energy arguments one would conclude that this structure cannot compete with form I. However, strong entropic effects stabilize it. The entropies shown in Fig. 3 indeed show a greater contribution to the stability in form B than in form I. We suggest that an important factor that contributes to the entropy is the fast rotation about the C–O axis. We calculated the characteristic rotation time using the time autocorrelation function of the N-N unit vector and fitting an exponential function to it. We show the results in *SI Appendix*, Fig. S2 and we compare them with those of form I. The characteristic time of rotation in form I is ∼800 ps while in form B it is ∼7 ps. We also computed the probability $p(\theta)$ as a function of the rotation angle θ about the C–O axis. From $p(\theta)$ the free energy can be calculated as $G(\theta) = -k_B\,T\log p(\theta)\sin\theta$. We show the results in *SI Appendix*, Fig. S2. Both in form I and in form B $G(\theta)$ exhibits a barrier separating two molecular configurations in which the Ns are exchanged. The barrier height is ∼18 kJ/mol in form B while it is ∼34 kJ/mol in form I. The entropy contribution from this rotation can be calculated from $k_B\,T\int p(\theta)\,\log p(\theta)\sin\theta\,d\theta$. The difference in entropy between form I and form B accounts for about 1.5 kJ/mol (0.4 k_BT). As the temperature is lowered, the structure undergoes a phase transition at around 200 K. Therefore, methods that search structures at zero temperature would find only the low-temperature form instead of the high-temperature one. The change in structure cannot be accounted for using harmonic corrections.

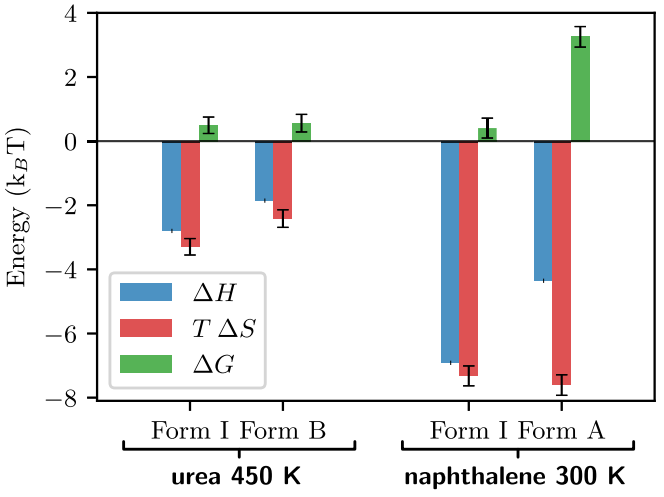

Fig. 3. Enthalpy, entropy, and free energy for selected polymorphs of urea at 450 K and naphthalene at 300 K. All quantities have the liquid as reference state.

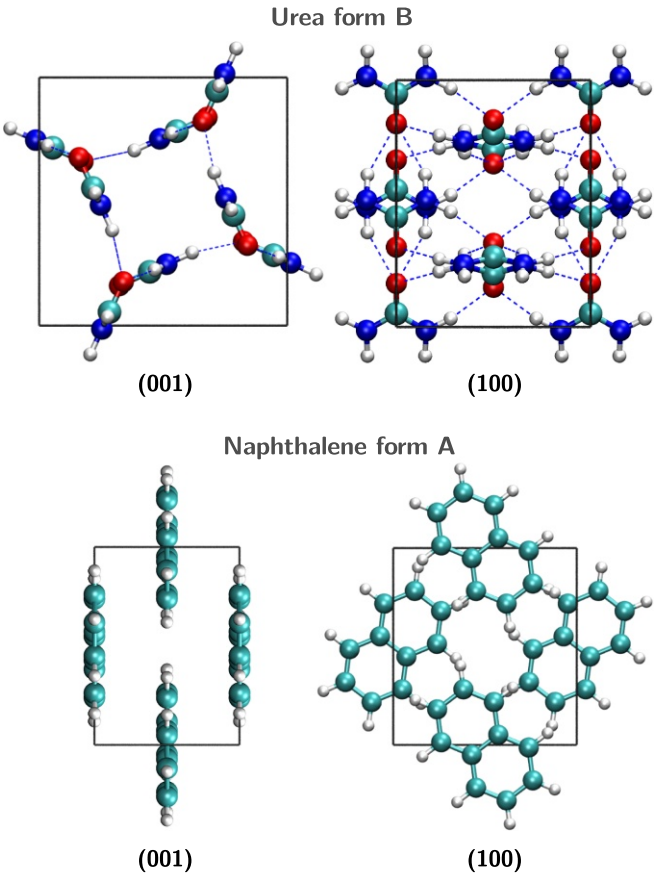

Fig. 4. Crystal structures of form B of urea and form A of naphthalene. C, O, N, and H atoms are shown in cyan, red, blue, and white, respectively. Images were obtained with Visual Molecular Dynamics (VMD) (29).

We now turn to discuss polymorph A of naphthalene. Form A has a layered structure and its space group is Pnnm (30). The structure is shown in Fig. 4. During an unbiased simulation at 300 K, form A decays to the liquid. This is consistent with the calculated free energy (Fig. 3) that shows that form A has a free energy around 3 $k_B\,T$ higher than form I and the liquid. Despite the relatively high free energy, it is possible that this polymorph could be kinetically trapped.

We have presented a method to explore polymorphism in molecular crystals in finite-temperature MD simulations. An important feature of our method is that not only does it discover polymorphs but also it pinpoints which are the relevant ones at finite temperature. In fact, form B of urea could not have been predicted from a zero-temperature search with harmonic corrections. A key ingredient of our approach is the structure similarity metric defined using $g(r,\theta)$ and the distance in Eq. **3**. This metric allows us to automatically assign configurations to a given polymorph, thus reducing the burden of the

Table 1. Parameters used in the definition of CVs S_{θ_1} and S_{θ_2} for urea and naphthalene

CV	r_m, nm	σ_r, nm	σ_θ	Mirror symmetry
Urea				
S_{θ_1}	0.6	0.05	0.25	no
S_{θ_2}	0.6	0.05	0.125	yes
Naphthalene				
S_{θ_1}	0.7	0.05	0.125	yes
S_{θ_2}	0.7	0.05	0.125	yes

See text for details.

analysis of the simulations. We are also able to calculate free energies and entropies from the simulation using a reweighting procedure (27).

Materials and Methods

Urea and naphthalene were described using the generalized Amber force field (GAFF) (31). For naphthalene, the electrostatic potential was calculated at the B3LYP/6-31+G (d,p) level using Gaussian 09 (32) and the partial charges of the atoms were fitted using the restrained electrostatic potential (RESP) method (33). The partial charges of urea were those provided with the Amber 03 database (34). Biased MD simulations were performed using Gromacs 5.1.4 (35) patched with a development version of PLUMED 2 (36). Van der Waals interactions and the electrostatic interaction in real space were calculated with cutoffs of 0.9 nm and 0.75 nm for urea and naphthalene, respectively. The electrostatic interaction in reciprocal space was calculated using the particle mesh Ewald (PME) method (37). The atomic bonds involving hydrogen were constrained using the LINCS algorithm (38) and the equations of motion were integrated with a 2-fs time step. The temperature was controlled using the stochastic velocity rescaling thermostat (39) with a relaxation time of 0.1 ps. The target temperatures of the thermostat were 450 K and 300 K for urea and naphthalene, respectively. We maintained the pressure at its atmospheric value using the isotropic version of the Parrinello–Rahman (40) barostat with a 10-ps relaxation time. We used systems of 108 and 36 molecules for urea and naphthalene, respectively.

It must be added here that we do not expect that the observed nucleation process is realistic. In fact, small size effects and periodic boundary conditions will artificially promote crystallization. In the present context this is a desirable feature. If we were to study the nucleation process, much larger systems should be considered.

We now provide the parameters used for the WTMetaD simulations (22). The Gaussians had a width of 0.1 k_B and 0.2 k_B for urea and naphthalene, respectively. In all cases the Gaussians had an initial height of 5 k_BT and were deposited every 1 ps. The bias factor was 200 for all simulations. The maximum free energy explored by WTMetaD is roughly the bias factor in k_BT. In our case we use a relatively large bias factor that ensures the exploration of high free energy regions.

We now discuss some practical aspects of the use of S_θ as a CV. To calculate the forces arising from the WTMetaD bias, S_θ should be continuous and differentiable. This can be achieved by constructing the function $g(r, \theta)$ using Gaussian kernels of width σ_r and σ_θ, as done in previous work. Furthermore, the integration in Eq. **2** cannot have an infinite upper limit, and in practice a finite cutoff r_m is taken. The integration is performed numerically using the trapezoid rule with steps of size σ_r and σ_θ in the r and $\cos\theta$ dimensions, respectively. We report in Table 1 the chosen parameters. σ_θ is reported in units of $\cos\theta$. A subtlety in the calculation of S_θ is the periodicity of $g(r, \theta)$ in its θ argument. For a general molecule $g(r, \theta)$ is periodic in θ with period π. However, for a molecule that has a mirror symmetry with respect to the plane perpendicular to the vector v defining the orientation of the molecule, $g(r, \theta)$ has a period $\pi/2$. We report in Table 1 whether a given CV is defined based on a direction of the molecule with mirror symmetry.

ACKNOWLEDGMENTS. We are grateful to Zoran Bjelobrk for providing the force field for naphthalene. We also thank Haiyang Niu for his valuable assistance in the analysis using the structure factor. This research was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation. The authors also acknowledge funding from European Union Grant ERC-2014-AdG-670227/VARMET. The computational time for this work was provided by the Swiss National Supercomputing Center (CSCS) under Project ID mr3. Calculations were performed in CSCS cluster Piz Daint.

1. Bernstein J (2002) *Polymorphism in Molecular Crystals* (Oxford Univ Press, New York), Vol 14.
2. Hilfiker R (2006) *Polymorphism: In the Pharmaceutical Industry* (Wiley, Weinheim, Germany).
3. Cabri W, Ghetti P, Pozzi G, Alpegiani M (2007) Polymorphisms and patent, market, and legal battles: Cefdinir case study. *Org Process Res Dev* 11:64–72.
4. Bauer J, et al. (2001) Ritonavir: An extraordinary example of conformational polymorphism. *Pharm Res* 18:859–866.
5. Bazterra VE, Ferraro MB, Facelli JC (2002) Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *J Chem Phys* 116: 5984–5991.
6. Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J Chem Phys* 124:244704.
7. Price SL (2008) From crystal structure prediction to polymorph prediction: Interpreting the crystal energy landscape. *Phys Chem Chem Phys* 10:1996–2009.
8. Pickard CJ, Needs R (2011) Ab initio random structure searching. *J Phys Condens Matter* 23:053201.
9. Yu TQ, Tuckerman ME (2011) Temperature-accelerated method for exploring polymorphism in molecular crystals based on free energy. *Phys Rev Lett* 107:015701.
10. Mendels D, McCarty J, Piaggi PM, Parrinello M (2018) Searching for entropically stabilized phases: The case of silver iodide. *J Phys Chem C* 122:1786–1790.
11. Valsson O, Tiwary P, Parrinello M (2016) Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annu Rev Phys Chem* 67:159–184.
12. Piaggi PM, Valsson O, Parrinello M (2017) Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. *Phys Rev Lett* 119:015701.
13. Prestipino S, Giaquinta PV (2004) The entropy multiparticle-correlation expansion for a mixture of spherical and elongated particles. *J Stat Mech Theory Exp* 2004: P09008.
14. Gobbo G, Bellucci MA, Tribello GA, Ciccotti G, Trout BL (2018) Nucleation of molecular crystals driven by relative information entropy. *J Chem Theory Comput* 14: 959–972.
15. Lamelas F, Dreger Z, Gupta Y (2005) Raman and x-ray scattering studies of high-pressure phases of urea. *J Phys Chem B* 109:8206–8215.
16. Olejniczak A, Ostrowska K, Katrusiak A (2009) H-bond breaking in high-pressure urea. *J Phys Chem C* 113:15761–15767.
17. Dziubek K, Citroni M, Fanetti S, Cairns AB, Bini R (2017) High-pressure high-temperature structural properties of urea. *J Phys Chem C* 121:2380–2387.
18. Giberti F, Salvalaglio M, Mazzotti M, Parrinello M (2015) Insight into the nucleation of urea crystals from the melt. *Chem Eng Sci* 121:51–59.
19. Shang C, Zhang XJ, Liu ZP (2017) Crystal phase transition of urea: What governs the reaction kinetics in molecular crystal phase transitions. *Phys Chem Chem Phys* 19:32125–32131.
20. Fabbiani FP, Allan DR, Parsons S, Pulham CR (2006) Exploration of the high-pressure behaviour of polycyclic aromatic hydrocarbons: Naphthalene, phenanthrene and pyrene. *Acta Crystallogr B* 62:826–842.
21. Likhacheva AY, Rashchenko SV, Litasov KD (2014) High-pressure structural properties of naphthalene up to 6 GPa. *J Appl Crystallogr* 47:984–991.
22. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100:020603.
23. De S, Bartók AP, Csányi G, Ceriotti M (2016) Comparing molecules and solids across structural and alchemical space. *Phys Chem Chem Phys* 18:13754–13769.
24. Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge Univ Press, New York).
25. Müllner D, et al. (2013) Fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *J Stat Softw* 53:1–18.
26. Jones E, Oliphant T, Peterson P (2014) Scipy: Open source scientific tools for Python. Available at https://www.scipy.org/. Accessed September 6, 2018.
27. Tiwary P, Parrinello M (2014) A time-independent free energy estimator for metadynamics. *J Phys Chem B* 119:736–742.
28. Salvalaglio M, Vetter T, Giberti F, Mazzotti M, Parrinello M (2012) Uncovering molecular details of urea crystal growth in the presence of additives. *J Am Chem Soc* 134:17221–17233.
29. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graphics* 14:33–38.
30. Stokes H, Hatch D, Campbell B (2018) Isotropy software suite. Available at iso.byu.edu. Accessed September 6, 2018.
31. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174.
32. Frisch M, et al. (2016) Gaussian 09, Revision a. 02 (Gaussian Inc., Wallingford, CT).
33. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J Phys Chem* 97:10269–10280.
34. Case DA, et al. (2005) The amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
35. Abraham MJ, et al. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1: 19–25.
36. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: New feathers for an old bird. *Comp Phys Commun* 185:604–613.

37. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
38. Hess B (2008) P-lincs: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4:116–122.
39. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.
40. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182–7190.