

Generating Cocrystal Polymorphs with Information Entropy Driven by Molecular Dynamics-Based Enhanced Sampling

Hongxing Song, Leslie Vogt-Maranto,* Ren Wiscons, Adam J. Matzger, and Mark E. Tuckerman*

Cite This: *J. Phys. Chem. Lett.* 2020, 11, 9751–9758

Read Online

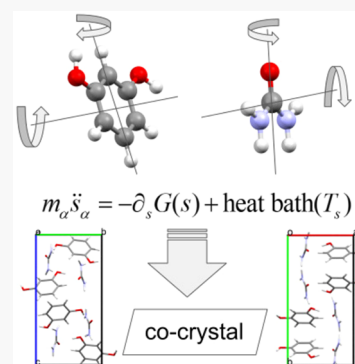
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Predicting structures of organic molecular cocrystals is a challenging task when considering the immense number of possible intermolecular orientations. Use of the Shannon information entropy, constructed from an intermolecular orientational spatial distribution function, to drive a search for crystal structures via enhanced molecular dynamics can be an efficient way to map out a landscape of putative polymorphs. Here, the Shannon entropy is used to generate a set of collective variables for differentiating polymorphs of a 1:1 cocrystal of resorcinol and urea. We show that driven adiabatic free energy dynamics, a particular enhanced-sampling approach, combined with these entropy variables, can transform the stable phase into alternate polymorphs. Density functional theory calculations confirm that a structure obtained from the enhanced molecular dynamics is stable at pressures above 1 GPa. We thus show that enhanced sampling should be considered an integral component of crystal structure searching protocols for systems with multiple independent molecules.



The determination of polymorphs of organic molecular crystals is an important component in the study of pharmaceuticals, energetic materials, organic semiconductors, and agrochemicals, as the performance of these materials can be affected by the existence of multiple kinetically accessible structures. In the area of organic molecular crystalline materials, leveraging cocrystal formulations, in which multiple components are crystallized together in a stoichiometric ratio, can be a useful avenue for tuning certain desired properties required by a particular target application. This is especially true in the pharmaceutical industry, where cocrystallizing an active pharmaceutical ingredient (API) with other molecules can significantly affect physicochemical properties such as the aqueous solubility, dissolution rate, and stability.^{1,2} Although cocrystallization can allow properties to be accessed that are not available in single-component crystals, understanding the interactions that favor cocrystallization over assembly of the individual components into pure crystals remains an area of active interest.³ While generating a set of theoretical polymorphs via crystal structure prediction (CSP) is becoming routine for organic molecules, determining experimentally relevant structures with more than one molecule in the asymmetric unit cell remains challenging.⁴ By definition, cocrystals have at least two independent molecules, making it critical to find new methods to predict the crystalline forms of these systems.

In this Letter, we demonstrate that the use of molecular dynamics (MD)-based enhanced sampling is a promising approach for cocrystal polymorph prediction. In this class of MD simulations, a trajectory is guided by modifying the original potential energy surface and corresponding forces, by

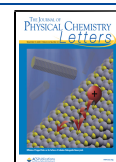
varying the temperature of the system, or both, to facilitate sampling of otherwise rare transitions between local minima. Indeed, it is becoming increasingly recognized that such molecular simulation approaches are an important component of a CSP protocol,^{4–6} as they subject putative structures to experimental conditions, can check their stability under these conditions, and produce relevant thermal averages at experimentally relevant temperatures and pressures. The need for enhanced-sampling approaches arises from the fact that polymorphs may be separated by high-energy barriers even if the free energy differences between them are small.

One approach to enhancing sampling is the popular metadynamics approach, in which a history-dependent bias potential adds to the underlying potential of visited structures, promoting exploration of new regions of the structural space.⁷ The biased potentials are built from one or more predefined collective variables (CVs), which distinguish different local minima and describe the transition path between them. Using the elements of the simulation cell matrix as CVs, for example, it has been possible to determine the solid forms of diamond and silicon⁸ and polymorphs of benzene.⁹ An alternate approach to the promotion of barrier crossings is to assign a high temperature to the heat bath of the CVs, or indirectly to the heat bath of auxiliary variables, known as coarse-grained

Received: August 29, 2020

Accepted: October 26, 2020

Published: November 3, 2020



variables (CGVs), coupled harmonically to the CVs. These algorithms are known as temperature-accelerated or adiabatic methods and include adiabatic free energy dynamics (AFED),¹⁰ driven adiabatic free energy dynamics (d-AFED),¹¹ and temperature-accelerated molecule dynamics (TAMD).¹² Their advantage lies in the fact that they generate quick sweeps over the configurational landscape, allowing new structures to be discovered with greater efficiency than is usually afforded by metadynamics. These methods, in their constant-pressure forms,^{13,14} were successfully used to generate structural transitions in atomic crystals such as xenon¹⁴ and molybdenum,¹⁵ to elucidate mechanisms of melting,¹⁶ and to search and rank polymorphs and predict disordered motifs in simple molecular crystals such as benzene^{5,13} and naphthalene⁵ under high pressure. This approach was also employed to confirm the thermodynamic stability of polymorphs of benzamide¹⁷ and one of the targets in the sixth CCDC Blind Test of Organic Crystal Structure Methods.⁴

The methods described above require a choice for the CVs, and for crystals, no universally applicable choice exists, although important strides have been made.¹⁸ Our previous studies of molecular crystals^{4,5,13,17} employed the three supercell vectors as CVs, which can drive orientational changes indirectly through changes in the supercell shape. However, for cocrystals, which is our focus here, use of the supercell vectors, alone, as CVs may be insufficient, as they do not differentiate between intermolecular orientations. Recently, the Shannon information entropy and relative entropy [also known as Kullback–Leibler divergence (KLD)] were introduced as CVs in metadynamics studies of the nucleation and recrystallization in atomic materials¹⁹ and small-molecule organic crystals.²⁰ It was also shown that the information entropy built from the intermolecular spatial distribution function, which includes the relative orientation between neighboring molecules, could be used to predict new metastable crystal polymorphs for rigid molecules such as urea and naphthalene.²¹

In this study, we demonstrate that the d-AFED scheme,¹⁴ when used with two information entropy-based CVs, can be used to predict new polymorphs of a 1:1 cocrystal of resorcinol and urea. Resorcinol, in particular, is a major commodity chemical that is used in the manufacturing of polymers, agrochemicals, and pharmaceuticals. Its unusual crystal growth patterns have remained a mystery for more than 100 years.^{22,23} In individual monocrystals of resorcinol and of urea, resorcinol has several different polymorphs corresponding to different relative orientations of the hydroxyl groups (see refs 24–27), and for urea, a number of polymorphs have also been reported (see refs 28–30 for urea); however, only one structure is reported for the cocrystal,^{31,32} and indeed, only one structure was observed for the 1:1 cocrystal in our own experimental crystallization attempts. Here, we show that a d-AFED trajectory generates a new structure in <25 ps that is confirmed as a low-energy polymorph in a standard (independent) random-search CSP approach. Although the predicted structure is metastable under ambient conditions, dispersion-corrected density functional theory (DFT-D) calculations predict that this new structure is more stable at pressures in the range of 1–2 GPa.

In order to explore the cocrystal free energy landscape, we employ CVs first introduced by Piaggi et al.,²¹ which measure translational and orientational order based on a construction of the Shannon information entropy. In this work, the construction is performed using the following procedure: Let

i and j index two molecules in a system of N_{mol} total molecules, and let $\mathbf{v}_i^{(\alpha)}$ ($\alpha = 1, \dots, n$) denote one of n internal vectors in molecule i . We then construct an approximate pair distribution or spatial distribution function (SDF) $g_\alpha(R, \cos \theta)$ via a sum of Gaussians as

$$g_\alpha(R, \cos \theta) \approx \frac{2}{4\pi R^2 \rho N_{\text{mol}}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{2\pi \sigma_R \sigma_{\cos \theta}} \exp \left\{ -\frac{(R - R_{ij})^2}{2\sigma_R^2} - \frac{[\cos \theta - \cos \theta_{ij}^{(\alpha)}]^2}{2\sigma_{\cos \theta}^2} \right\} \quad (1)$$

where $\rho = N_{\text{mol}}/V$ is the molecular number density, σ_R and $\sigma_{\cos \theta}$ are the Gaussian width parameters, $R_{ij} = |\mathbf{R}_i - \mathbf{R}_j|$ is the distance between centers of mass $\mathbf{R}_i$ and $\mathbf{R}_j$ of molecules i and j , respectively, and $\cos \theta_{ij}^{(\alpha)} = \mathbf{v}_i^{(\alpha)} \cdot \mathbf{v}_j^{(\alpha)} / \|\mathbf{v}_i^{(\alpha)}\| \|\mathbf{v}_j^{(\alpha)}\|$ is the relative orientation between $\mathbf{v}_i^{(\alpha)}$ and $\mathbf{v}_j^{(\alpha)}$. These n distribution functions are then used to construct n CVs given by

$$q_\alpha(\mathbf{r}) = -2\pi\rho \int_0^\infty dR R^2 \int_{-1}^1 d(\cos \theta) \{g_\alpha(R, \cos \theta) \ln g_\alpha(R, \cos \theta) - [g_\alpha(R, \cos \theta) - 1]\} \quad (2)$$

where $\mathbf{r} \equiv \mathbf{r}_1, \dots, \mathbf{r}_N$ denotes the full set of N atomic coordinates in the system. In practice, the integrals in eq 2 are evaluated using a trapezoidal rule approximation and a distance cutoff of R_c on a uniform two-dimensional grid with all molecule pairs whose center-of-mass distances are within $R_c + 5\sigma_R$ contributing as Gaussian centers in eq 1.

Once the n CVs are defined, they are fed into the d-AFED enhanced-sampling approach, whose equations of motion are

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i - \sum_{\alpha=1}^n \kappa_\alpha [q_\alpha(\mathbf{r}) - s_\alpha] \frac{\partial q_\alpha}{\partial \mathbf{r}_i} + \text{Bath}(T) \\ \mu_\alpha \ddot{s}_\alpha = \kappa_\alpha [q_\alpha(\mathbf{r}) - s_\alpha] + \text{Bath}(T_s) \quad (3)$$

where $\mathbf{F}_i$ is the force on atom i arising from the interatomic interaction potential $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$, m_i is the mass of atom i , s_α is a coarse-grained variable (CGV), μ_α is a fictitious mass associated with s_α that determines the time scale on which it moves, κ_α determines the strength of a harmonic coupling introduced between the CVs and the CGVs, and $T_s \gg T$ is the temperature used to accelerate sampling of the CGVs. Here, “Bath(τ)” generically refers to a thermostat coupling, such as a Langevin bath or Nosé–Hoover chain,³³ used to maintain the associated variables at temperature τ . These equations provide a sampling of the marginal distribution and free energy landscape as a function of the CGVs, which is formally exact in the limit that $\kappa_\alpha \rightarrow \infty$; typical values for practical applications to molecular crystal systems are described in the Computational Methods section.

Figure 1 shows chemical diagrams of the two monomers, resorcinol (a) and urea (b). Resorcinol is shown in the conformation it adopts in the cocrystals with urea that are studied here. All heavy atoms (hydrogen atoms excluded) are used to calculate the center of mass of each molecule, but only two or three atoms are required to define the internal vectors $\mathbf{v}_i^{(\alpha)}$ needed to specify the relative orientation between neighboring molecules. Here we chose two such internal vectors ($n = 2$): an out-of-plane vector and an in-plane vector. The first is the normal (out-of-plane) vector of each molecule,

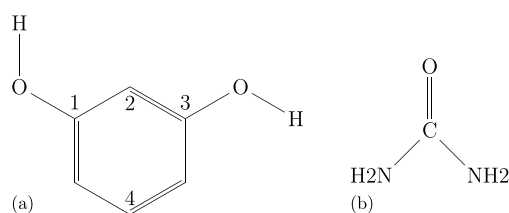


Figure 1. Schematic representation of the two monomers: (a) resorcinol and (b) urea in their 1:1 cocrystal conformations. The selected heavy atoms (C atoms labeled as 1–4 for resorcinol, C, O, and both N atoms for urea) are used to calculate the orientation of each molecule.

i.e., $\mathbf{v}_i^{(1)} = (\mathbf{r}_{C1} - \mathbf{r}_{C2}) \times (\mathbf{r}_{C3} - \mathbf{r}_{C2})$ for the i th resorcinol molecule, and $\mathbf{v}_i^{(1)} = (\mathbf{r}_{N1} - \mathbf{r}_C) \times (\mathbf{r}_{N2} - \mathbf{r}_C)$ for the i th urea molecule. The second vector is the symmetry axis (in-plane) of the molecule (without hydrogen atoms considered), i.e., $\mathbf{v}_i^{(2)} = \mathbf{r}_{C2} - \mathbf{r}_{C4}$ for the i th resorcinol molecule, and $\mathbf{v}_i^{(2)} = \mathbf{r}_O - \mathbf{r}_C$ for the i th urea molecule. As mentioned above, the dot product between the unit vectors corresponding to the vectors $\mathbf{v}_i^{(\alpha=1,2)}$

and $\mathbf{v}_j^{(\alpha=1,2)}$ gives the relative orientation $\cos \theta_{ij}^{(\alpha=1,2)}$ between two neighboring molecules i and j , which then can be used to build the spatial distribution functions $g_{\alpha=1,2}(R, \cos \theta)$ and their corresponding CVs $q_{\alpha=1,2}(\mathbf{r})$ in the enhanced-sampling MD simulations.

The simulations (see **Computational Methods**) start from the reported orthorhombic structure (space group $P2_12_12_1$) shown in **Figure 3a**, which was previously determined experimentally,^{31,32} observed in our own crystallization experiments (see **Experimental Methods** and the **Supporting Information**), and confirmed by us as the lowest-energy polymorph using standard zero-temperature crystal structure prediction protocols. (See **Computational Methods**. Cocrystals between urea and each of the two remaining conformers of resorcinol were also considered, as discussed in the **Supporting Information**, but they are not considered further here, as no experimental evidence exists, as yet, for such structures.) **Figure 2** shows the evolution of the instantaneous potential energy during a 50 ps long d-AFED run at 100 K and 1 bar. The Gaussian broadening parameters are $\sigma_R = 0.5 \text{ \AA}$ and $\sigma_{\cos \theta}$

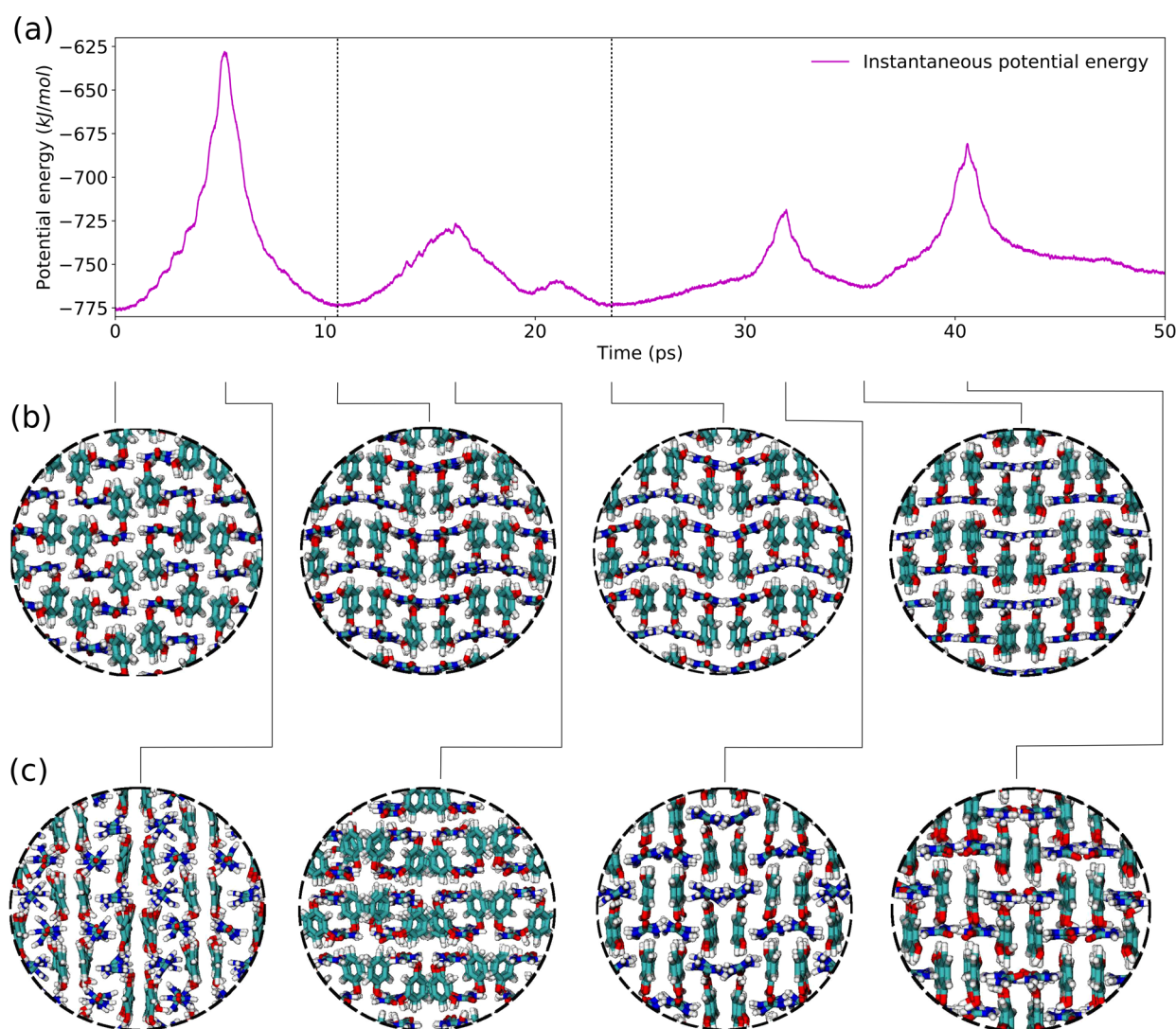


Figure 2. Evolution of (a) the instantaneous potential energy [per formula unit (f.u.)] along a $T = 100 \text{ K}$ and $P = 1 \text{ bar}$ d-AFED simulation for a 1:1 cocrystal of resorcinol and urea. The initial structure ($t = 0 \text{ ps}$) is the experimental structure in space group $P2_12_12_1$. The vertical dashed lines indicate the predicted $P2_1/c$ structure visited at both $t = 10.59 \text{ ps}$ and $t = 23.64 \text{ ps}$. Snapshots for selected local (b) minima and (c) maxima are also shown. The local minimum at $t = 19.86 \text{ ps}$ is also $P2_1/c$ -like, which is not shown here.

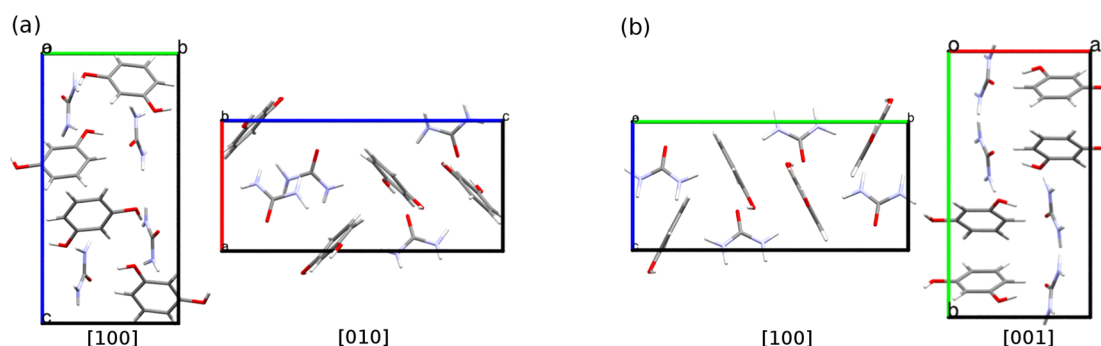


Figure 3. Unit cell of (a) experimental $P2_12_12_1$ and (b) predicted $P2_1/c$ structures for the 1:1 cocrystal of resorcinol and urea.

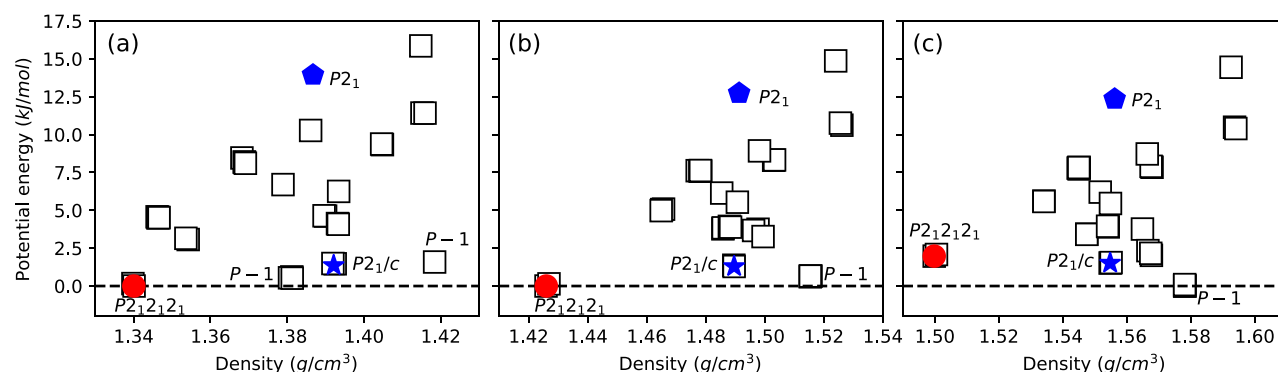


Figure 4. Potential energy vs density plots for the top 30 structures (empty black squares) from standard CSP. The potential energy (given per f.u.) and the density are both averaged from flexible cell NPT MD at 300 K but at different pressures: (a) 1 bar, (b) 1 GPa, and (c) 2 GPa. The experimental structure (filled red circle) and the metastable polymorphs predicted in the d-AFED trajectories (filled blue star and pentagon) are also shown, including their space groups.

= 0.25, respectively. We also provide snapshots visited at the selected local minima and maxima in potential energy along the trajectory (see panels b and c of Figure 2; full animations and more snapshots can also be found in the Supporting Information).

Figure 2a shows two local minima in potential energy ($t = 10.59$ ps, and $t = 23.64$ ps), which are somewhat similar to the initial experimental structure. Therefore, we equilibrated the corresponding structures using a standard MD simulation to verify their stabilities and to characterize new features in their intermolecular packing. The equilibrated structures at both $t = 10.59$ ps and $t = 23.64$ ps result in the same new polymorph with space group $P2_1/c$ and $Z' = 2$. Similar transformations to the $P2_1/c$ polymorph were also observed at 200 and 300 K. Thus, although the d-AFED trajectory passes through structures with potential energies of >150 kJ/mol per formula unit (f.u.) above the starting structure, it ultimately identifies the $P2_1/c$ form within 25 ps, which is within 2 kJ/mol of the experimental orthorhombic structure (see Figure 2a). The free energy difference at 1 bar, as calculated using the Frenkel–Ladd method³⁴ with the classical force field, is only ~ 1 kJ/mol in the temperature range of 50–300 K (see the Supporting Information). Despite the relatively small energy difference, no experimental work under atmospheric conditions to date has identified this $P2_1/c$ polymorph. In Figure 3, we provide the unit cell of both the experimental structure and the new polymorph predicted in the d-AFED trajectory. Their packing differences are provided in the Supporting Information. Compared to the experimental structure, the predicted monoclinic polymorph is slightly more dense. The MD-averaged lattice parameters of the $P2_1/c$ polymorph at room

temperature and ambient pressure ($T = 300$ K, and $P = 1$ bar) are as follows: $a = 7.9678$ Å, $b = 14.8209$ Å, $c = 6.9249$ Å, and $\beta = 96.66^\circ$.

We note that running the d-AFED simulations using the Gaussian kernel width $\sigma_{\cos \theta} = 0.3$ at 200 K reveals another new polymorph in the $P2_1$ space group with $Z' = 4$, but this lower-symmetry structure is much higher in energy than the experimental structure (~ 14 kJ/mol). This structure is discussed in more detail in the Supporting Information.

We also used a standard CSP-based random structure search³⁵ to validate the new polymorphs predicted by MD-based enhanced sampling. Figure 4 shows the MD-averaged energy–density plots for the top 30 structures. Both the potential energy and the density for each structure are averaged from standard isothermal–isobaric flexible cell (NPT) MD at 300 K and different pressures ($P = 1$ bar, 1 GPa, and 2 GPa). As expected, the experimental polymorph in space group $P2_12_12_1$ is ranked as the lowest-energy structure in a $Z' = 2$ search (two independent monomers) under ambient pressure. The density calculated by MD is ~ 1.34 g/cm³, which is in good agreement with the experimental density (1.31–1.32 g/cm³), and the predicted and experimental structures match with a root-mean-square displacement of 20 molecules (RMSD_{20}) of 0.34 Å against the previously reported structure³² and the better resolution structure determined in this work. The predicted $P2_1/c$ polymorph is also confirmed to be one of the low-energy forms found in the same search, as shown in Figure 4a. The d-AFED-predicted $P2_1$ polymorph was not observed in a separate $Z' = 4$ search, which generated an additional 3000 random structures in the $P2_1$ space group. This highlights the fact that an MD-based enhanced sampling

approach can identify structures not readily observed in a random CSP search.

Although the experimental structure is confirmed to be the lowest-energy structure in the random search at ambient pressure, there are several other polymorphs that are energetically comparable. For example, the d-AFED-predicted $P2_1/c$ polymorph and two other $P\bar{1}$ random structures are within 2 kJ/mol of the experimental structure, as shown in Figure 4a. More importantly, the experimental structure has the lowest density among the top 30 structures, which means it might not be preferable at higher pressures. Therefore, we also minimized the top structures at higher pressures ($P = 1$ and 2 GPa) and then compared their MD-averaged energy and density in panels b and c of Figure 4. We note that at 1 GPa, the two low-energy $P\bar{1}$ random structures generated at 1 bar merge to the same $P\bar{1}$ structure (see the Supporting Information). With increasing pressure, the PV contribution begins to dominate the relative enthalpy (far larger than entropic contributions to the Gibbs free energy), making the crystal density an important factor in ranking possible polymorphs. Enthalpy–density plots can be found in the Supporting Information. While the experimentally reported structure still has the lowest energy at 1 GPa, other predicted structures have lower enthalpies. At 2 GPa, the $P2_12_12_1$ structure is higher in both energy and enthalpy than the two other candidates, $P2_1/c$ and $P\bar{1}$.

The $P2_1/c$ structure was found in a d-AFED trajectory starting from the experimentally known form of the cocrystal. These simulations can also be run using other structures, e.g., from a standard $Z' = 2$ CSP search as the starting point. Using this approach, the d-AFED simulations find the low-energy $P\bar{1}$ polymorph with the highest density as well as two other variants for this packing (see the Supporting Information for more details and the structure files). One form is ~ 4.7 kJ/mol higher in potential energy in the $P\bar{1}$ space group. The other is only 2 kJ/mol higher in energy with double the number of molecules in the asymmetric unit cell ($Z' = 4$) in the $P\bar{1}$ space group. The $P\bar{1}$ variant with $Z' = 2$ is found in the standard CSP search, while the $P\bar{1}$ variant with $Z' = 4$ was not found in an additional CSP search of 3000 structures having the same Z' value in the $P\bar{1}$ space group.

As all of the MD simulations and the standard CSP are based on a classical force field, we turn to DFT-D calculations to confirm the stability of the predicted low-energy polymorph at various pressures. In the DFT-D calculations, we also include the representative $P\bar{1}$ form identified in the random CSP search (see the Supporting Information for a more detailed description of the structure). The relative energy and enthalpy differences are shown in Figure 5. The $P2_1/c$ form is predicted to be the most stable at pressures between 1.02 and 2.07 GPa. The relative stabilities of the three selected structures are primarily due to the interplay of total energy and the PV contribution (see the Supporting Information for more details). To consider finite temperatures, the vibrational energy and the entropic contributions to the Gibbs free energy could be evaluated using the harmonic approximation, which is computationally expensive and beyond the scope of this work. In any case, the PV contribution would eventually dominate any entropic differences with an increase in pressure. Although the experimental structure is predicted to be most stable at low pressure, the denser structures are predicted to become enthalpically favored at high pressures. The d-AFED-predicted $P2_1/c$ structure is enthalpically preferred between 1.02 and

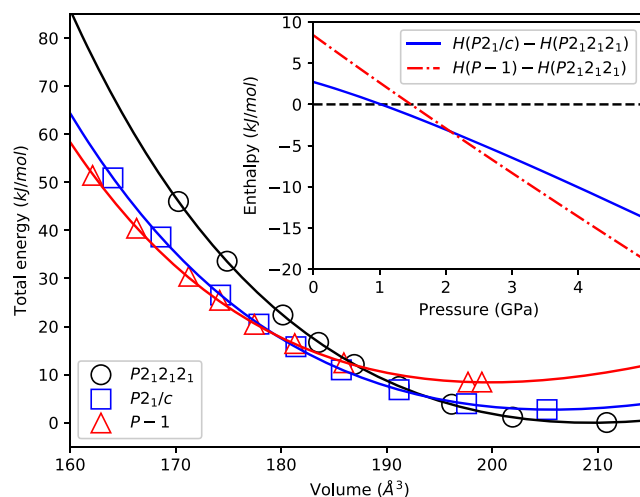


Figure 5. Relative total energy vs cell volume from DFT-D calculations (see Computational Methods) for three low-energy structures. The E – V curves are fitted to the Murnaghan equation of state model, which is then used to calculate the enthalpy (per f.u.) as a function of pressure (see the inset). The $P2_1/c$ structure is predicted to become more stable than the experimentally observed $P2_12_12_1$ polymorph (zero reference of energy and enthalpy) at pressures between 1.02 and 2.07 GPa.

2.07 GPa. This behavior suggests that the predicted monoclinic structure, which is metastable at ambient pressure, could be more thermodynamically accessible if future experiments are conducted at moderate pressures.³⁶

From a computational perspective, cocrystals are inherently more complex than monocystals owing to the richness of the effective $Z' = 2$ crystal structure landscape. In this study, we demonstrated the value of enhanced-sampling techniques in rapidly identifying polymorphs not easily found in a standard CSP search and generating transitions between different structures. The ability to generate structures readily with higher Z' values makes this method complementary to a standard CSP search with the minimal number of molecules in the asymmetric unit cell. A further advantage of MD-based enhanced-sampling methods like d-AFED is that they return structures that are equilibrated at experimentally relevant temperatures and pressures, as opposed to returning zero-temperature structures, as is the case in standard CSP approaches. This is particularly relevant for cocrystals with multiple molecular species in the asymmetric unit cell, where entropic effects could play a role in stabilizing certain polymorphs. Although this particular study employed entropy-based CVs within the enhanced-sampling protocol, it would be interesting to explore the use of recently introduced environment variables³⁷ within the d-AFED scheme as an alternative approach to the exploration of the cocrystal free energy landscape, and this will be the subject of future work.

COMPUTATIONAL METHODS

For the classical force field molecular dynamics simulations, the intramolecular (bond, bend, and torsion) interaction was taken from the generalized AMBER force field (GAFF) parametrization implemented in AMBER,^{38,39} while the intermolecular interaction employed the standard Lennard-Jones potential, parametrized in the optimized potentials for liquid simulations (OPLS) force field.⁴⁰ Atomic charges for resorcinol were assigned using the AmberTools package,⁴¹

with a Boltzmann-weighted RESP fit⁴² to HF/6-31G* calculations at the three local MP2/6-31G* minima, as calculated in Gaussian09.⁴³ Charges for urea were taken from the OPLS-AA force field.⁴⁰

All standard MD and crystal d-AFED simulations were performed in the fully flexible cell isothermal–isobaric (*NPT*) ensemble, if not specified explicitly, as implemented in the PINY_MD package.⁴⁴ There were 128 resorcinol and the same number of urea molecules (1:1 cocrystal) in the simulation box with periodic boundary conditions. The basic time step for integration was $\Delta t = 0.25$ fs, and the total time for each crystal d-AFED simulation was 50 ps at different temperatures (100, 200, and 300 K). In evaluating the long-range Coulomb potential, we used an Ewald screening factor of $\alpha = 0.35 \text{ \AA}^{-1}$ and a real space cutoff $r_c = 10 \text{ \AA}$ for both resorcinol and urea. The smooth particle mesh Ewald (SPME) summation⁴⁵ was also used with the interpolation order of 10. The massive Nosé–Hoover chain (NHC) thermostat^{33,46,47} was used for all system variables; each NHC has a length of 4 with a characteristic time scale of $\tau = 20$ fs. The factorization scheme of the NHC integrator was the Suzuki–Yoshida^{48,49} scheme up to the sixth order or $n_{sy} = 7$ with a multiple time step factor $n_c = 4$.⁵⁰ In the integration of the Martyna–Tobias–Klein (MTK) equations in the flexible cell *NPT* ensemble,^{50,51} a characteristic time scale $\tau = 1000$ fs was used for both the barostat and its NHC thermostat. For the thermostat of the CGVs in crystal d-AFED, we used the generalized Gaussian moment thermostat (GGMT)^{11,52} up to the fourth moment ($M = 2$), with the fourth-order Suzuki–Yoshida factorization, or $n_{sy} = 3$, a multiple time step factor $n_c = 4$, and a characteristic time scale of $\tau = 50$ fs.

In the crystal d-AFED simulations, the mass (μ_a) and temperature (T_s) coupled to the CGVs were set to $3.4 \times 10^{12} \text{ K fs}^2$ and $1 \times 10^6 \text{ K}$, respectively. The harmonic coupling constant between CGVs and CVs was taken to be $\kappa_a = 2.8 \times 10^7 \text{ K}$. The cutoff distance was set to $R_c = 7.0 \text{ \AA}$ in the numerical integration to calculate the entropy CVs. The Gaussian width was $\sigma_R = 0.5 \text{ \AA}$ and $\sigma_{\cos \theta} = 0.25\text{--}0.3$ for the distance and angle, respectively. The local minima visited in the crystal d-AFED simulations were re-equilibrated in a 25 ps standard *NPT* MD simulation. The supercell after the equilibration was averaged and collapsed into a unit cell, with symmetry information and lattice parameters determined by PLATON.⁵³

The randomly packed structures were generated by the UPACK package³⁵ for 13 space groups, which include the most common space groups for molecular crystals: *C2*, *C2/c*, *Cc*, *P2₁2₁2₁*, *Pbca*, *Pbcn*, *P2₁*, *P2₁/c*, *Pca2₁*, *Pc*, *Pna2₁*, *P1*, and *P1*. For each space group, more than 3000 random structures are generated with $n_{gr} = 1$ or $Z' = 2$ (number of the two coformer molecules in the asymmetric primitive cell). We also ran another search for space group *P2₁* with $n_{gr} = 2$ or $Z' = 4$ to check the higher-energy metastable polymorph predicted by d-AFED, but no match was found. The rigid molecule packing step was followed by a minimization of the potential energy with the OPLS force field in the same package. These random structures were then ranked according to their energy; the top 30 structures were then expanded to a supercell, reminimized, and equilibrated (one *NVT* MD run followed by two other flexible cell *NPT* MD runs) to the target pressure and temperature.

The DFT-D total energy calculations were performed using the QUANTUM ESPRESSO package^{54,55} within the framework

of the generalized gradient approximation (GGA) parametrized by the Perdew–Burke–Ernzerhof (PBE) functional.⁵⁶ The dispersion-correction energy term was described by the Grimme DFT-D3 method⁵⁷ with Becke–Jonson (BJ) damping.⁵⁸ The projector-augmented wave (PAW) pseudopotential⁵⁹ was employed with a plane wave basis set cutoff of 80 Ry and the Monkhorst–Pack k-point mesh.⁶⁰ The convergence of the kinetic energy cutoff and the k-point grid were carefully tested to ensure that the change in the total energy was $< 5 \times 10^{-5}$ Ry per atom. All force components were $< 1 \times 10^{-3}$ Ry/bohr in the optimized structure.

EXPERIMENTAL METHODS

Single crystals of the urea/resorcinol cocrystal were grown by liquid-assisted grinding using acetonitrile with a 1:1 molar ratio of urea and resorcinol. The compounds were ground together at room temperature for 15 min using only enough acetonitrile to wet the solids. Single-crystal X-ray diffraction data were collected using a Rigaku XtaLAB Synergy-S X-ray diffractometer configured in a κ goniometer geometry. The diffractometer is equipped with a low-temperature device and a PhotonJet-S microfocus Cu source ($\lambda = 1.54187 \text{ \AA}$) and operated at 50 kV and 1 mA. X-ray intensities were measured under ambient conditions with the HyPix-6000HE detector placed 32.01 mm from the sample. The data were processed with CrysAlisPro version 38.46 (Rigaku Oxford Diffraction) and corrected for absorption. The structures were determined in OLEX2⁶¹ using SHELXT⁶² and refined using SHELXL.⁶³ All non-hydrogen atoms were refined anisotropically with hydrogen atoms placed at idealized positions. Single crystals were mounted on a 150 μm MiTeGen MicroMount using mineral oil.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.0c02647>.

Selected snapshots of the structural transitions observed in the crystal d-AFED simulation, MD-averaged enthalpy–density plot for the predicted polymorphs and the top 30 random structures, simulated powder X-ray diffraction (PXRD) of the *P2₁2₁2₁* and *P2₁/c* phases under ambient conditions, DFT-D-calculated pressure dependence of the lattice parameters for the *P2₁2₁2₁*, *P2₁/c*, and *P1* polymorphs, a table of experimental crystallographic parameters, and an ORTEP figure (PDF)

Crystallographic information of the predicted *P2₁/c* structure and another *P2₁* structure (much higher in energy), the low-energy *P1* polymorph, and its two variants (*P1* and *P1*) in the crystal d-AFED simulations (labeled “dAFED”), crystallographic information of the experimental *P2₁2₁2₁* structure, the metastable *P2₁/c* form, and the *P1* family from the standard structure prediction (labeled “CSP”), minimized and at various temperatures and pressures (among the selected CSP structures are those that were also discovered in crystal d-AFED simulations), and crystallographic information of the *P2₁2₁2₁*, *P2₁/c*, and *P1* structures at selected pressures determined by DFT-D optimization (labeled “DFT-D”) (CIF)

Movie showing the transition from the $P2_12_12_1$ to $P2_1/c$ polymorph over the d-AFED simulation viewed along the z axis (MPG)

Movie showing the transition from the $P2_12_12_1$ to $P2_1/c$ polymorph over the d-AFED simulation viewed along the x axis (MPG)

Movie showing the transition from the high-energy $P2_1$ to the low-energy $P\bar{1}$ polymorph over the d-AFED simulation viewed along the x axis (MPG)

Movie showing the transition from the high-energy $P2_1$ to the low-energy $P\bar{1}$ polymorph over the d-AFED simulation viewed along the z axis (MPG)

AUTHOR INFORMATION

Corresponding Authors

Leslie Vogt-Maranto – Department of Chemistry, New York University, New York, New York 10003, United States; orcid.org/0000-0002-7006-4582; Email: leslie.vogt@nyu.edu

Mark E. Tuckerman – Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003, United States; NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China; orcid.org/0000-0003-2194-9955; Email: mark.tuckerman@nyu.edu

Authors

Hongxing Song – Department of Chemistry, New York University, New York, New York 10003, United States

Ren Wiscons – Department of Chemistry and Macromolecular Science and Engineering Program, University of Michigan, Ann Arbor, Michigan 48109–1055, United States; orcid.org/0000-0002-8281-0343

Adam J. Matzger – Department of Chemistry and Macromolecular Science and Engineering Program, University of Michigan, Ann Arbor, Michigan 48109–1055, United States; orcid.org/0000-0002-4926-2752

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpclett.0c02647>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Nicolas Vigilante for preliminary work generating cocrystal structures. R.W. and A.J.M. acknowledge Sabrina Archer for assistance with crystal growth. L.V.-M. acknowledges support from the Army Research Office (ARO), Grant W911NF-19-1-0095 P0002 and W911NF-13-0387. M.E.T. acknowledges support from the National Science Foundation (NSF), Grants CHE-1565980 and CHE-1955381.

REFERENCES

- (1) Blagden, N.; de Matas, M.; Gavan, P.; York, P. Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Adv. Drug Delivery Rev.* **2007**, *59*, 617–630.
- (2) Karki, S.; Friščić, T.; Fábán, L.; Laity, P. R.; Day, G. M.; Jones, W. Improving Mechanical Properties of Crystalline Solids by Cocrystal Formation: New Compressible Forms of Paracetamol. *Adv. Mater.* **2009**, *21*, 3905–3909.
- (3) Sandhu, B.; Mclean, A.; Sinha, S. A.; Desper, J.; Sarjeant, A. A.; vyas, S.; Reutzel-Edens, S. M.; Aakeroy, C. B. Evaluating competing intermolecular interactions through molecular electrostatic potentials

and hydrogen-bond propensities. *Cryst. Growth Des.* **2018**, *18*, 466–478.

(4) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 439–459.

(5) Schneider, E.; Vogt, L.; Tuckerman, M. E. Exploring Polymorphism of Benzene and Naphthalene with Free Energy Based Enhanced Molecular Dynamics. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *B72*, 542–550.

(6) Chan, E. J.; Neumann, M. A. Evaluation of General and Tailor Made Force Fields via X-ray Thermal Diffuse Scattering Using Molecular Dynamics and Monte Carlo Simulations of Crystalline Aspirin. *J. Chem. Theory Comput.* **2018**, *14*, 2165.

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

(8) Martoňák, R.; Laio, A.; Parrinello, M. Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Phys. Rev. Lett.* **2003**, *90*, 075503.

(9) Raiteri, P.; Martoňák, R.; Parrinello, M. Exploring Polymorphism: The Case of Benzene. *Angew. Chem., Int. Ed.* **2005**, *44*, 3769–3773.

(10) Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. E. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.* **2002**, *116*, 4389.

(11) Abrams, J. B.; Tuckerman, M. E. Efficient and Direct Generation of Multidimensional Free Energy Surfaces via Adiabatic Dynamics without Coordinate Transformations. *J. Phys. Chem. B* **2008**, *112*, 15742–15757.

(12) Maragliano, L.; Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **2006**, *426*, 168–175.

(13) Yu, T.-Q.; Tuckerman, M. E. Temperature-Accelerated Method for Exploring Polymorphism in Molecular Crystals Based on Free Energy. *Phys. Rev. Lett.* **2011**, *107*, 015701.

(14) Yu, T.-Q.; Chen, P.-Y.; Chen, M.; Samanta, A.; Vanden-Eijnden, E.; Tuckerman, M. Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions. *J. Chem. Phys.* **2014**, *140*, 214109.

(15) Rogal, J.; Schneider, E.; Tuckerman, M. E. Neural-network based path collective variables for enhanced sampling of phase transformations. *Phys. Rev. Lett.* **2019**, *123*, 245701.

(16) Samanta, A.; Tuckerman, M. E.; Yu, T.-Q.; E, W. Microscopic mechanisms of equilibrium melting of a solid. *Science* **2014**, *346*, 729.

(17) Fellah, N.; Shtukenberg, A. G.; Chan, E. J.; Vogt-Maranto, L.; Xu, W.; Li, C.; Tuckerman, M. E.; Kahr, B.; Ward, M. D. Disorderly conduct of benzamide IV: Crystallographic and computational analysis of high entropy polymorphs of small molecules. *Cryst. Growth Des.* **2020**, *20*, 2670–2682.

(18) Santiso, E. E.; Trout, B. L. A general set of order parameters for molecular crystals. *J. Chem. Phys.* **2011**, *134*, 064109.

(19) Piaggi, P. M.; Valsson, O.; Parrinello, M. Enhancing Entropy and Enthalpy Fluctuations to Drive Crystallization in Atomistic Simulations. *Phys. Rev. Lett.* **2017**, *119*, 015701.

(20) Gobbo, G.; Bellucci, M. A.; Tribello, G. A.; Ciccotti, G.; Trout, B. L. Nucleation of Molecular Crystals Driven by Relative Information Entropy. *J. Chem. Theory Comput.* **2018**, *14*, 959–972.

(21) Piaggi, P. M.; Parrinello, M. Predicting polymorphism in molecular crystals using orientational entropy. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10251–10256.

(22) Wallerant, F. C. On the helicooidal windings in crystallised bodies. *R. Acad. Sci.* **1906**, *143*, 555.

(23) Wallerant, F. C. The origin helicooidal rolling body in crystallised bodies. *R. Acad. Sci.* **1906**, *143*, 1169.

(24) Robertson, J. M. The structure of resorcinol a quantitative X-ray investigation. *Proc. R. Soc. London, Ser. A* **1936**, *157*, 79–99.

- (25) Robertson, J. M.; Ubbelohde, A. R.; Bragg, W. H. A new form of resorcinol I. Structure determination by X-rays. *Proc. R. Soc. London, Ser. A* **1938**, *167*, 122.
- (26) Rao, R.; Sakuntala, T.; Godwal, B. K. Evidence for high-pressure polymorphism in resorcinol. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *65*, 054108.
- (27) Zhu, Q.; Shtukenberg, A. G.; Carter, D. J.; Yu, T.-Q.; Yang, J.; Chen, M.; Raiteri, P.; Oganov, A. R.; Pokroy, B.; Polishchuk, I.; et al. Resorcinol Crystallization from the Melt: A New Ambient Phase and New "Riddles". *J. Am. Chem. Soc.* **2016**, *138*, 4881–4889.
- (28) Lamelas, F. J.; Dreger, Z. A.; Gupta, Y. M. Raman and X-Ray Scattering Studies of High-Pressure Phases of Urea. *J. Phys. Chem. B* **2005**, *109*, 8206–8215.
- (29) Olejniczak, A.; Ostrowska, K.; Katrusiak, A. H-Bond Breaking in High-Pressure Urea. *J. Phys. Chem. C* **2009**, *113*, 15761–15767.
- (30) Dziubek, K.; Citroni, M.; Fanetti, S.; Cairns, A. B.; Bini, R. High-Pressure High-Temperature Structural Properties of Urea. *J. Phys. Chem. C* **2017**, *121*, 2380–2387.
- (31) Pickering, M.; Small, R. W. H. Structure of the 1:1 Complex of Resorcinol and Urea. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1982**, *B38*, 3161–3163.
- (32) Krishnakumar, V.; Nagalakshmi, R.; Manohar, S.; Piasecki, M.; Kityk, I.; Bragiel, P. Parametrical optical effects in the 1:1 complex of resorcinol and urea—a non linear optical crystal. *Phys. B* **2010**, *405*, 839–842.
- (33) Martyna, G. J.; Klein, M. L.; Tuckerman, M. E. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (34) Frenkel, D.; Ladd, A. J. C. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.* **1984**, *81*, 3188.
- (35) van Eijck, B. P.; Kroon, J. UPACK Program Package for Crystal Structure Prediction: Force Fields and Crystal Structure Generation for Small Carbohydrate Molecules. *J. Comput. Chem.* **1999**, *20*, 799–812.
- (36) Neumann, M. A.; van de Streek, J.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat. Commun.* **2015**, *6*, 7793.
- (37) Piaggi, P. M.; Parrinello, M. Calculation of phase diagrams in the multithermal-multibaric ensemble. *J. Chem. Phys.* **2019**, *150*, 244119.
- (38) 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.
- (39) 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.
- (40) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (41) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J. Chem. Inf. Comput. Sci.* **2001**, *222*, U403.
- (42) 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.* **1993**, *97*, 10269–10280.
- (43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, rev. E.01; Gaussian Inc.: Wallingford, CT, 2009.
- (44) Tuckerman, M. E.; Yarne, D.; Samuelson, S. O.; Hughes, A. L.; Martyna, G. J. Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Comput. Phys. Commun.* **2000**, *128*, 333–376.
- (45) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (46) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (47) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695–1697.
- (48) Suzuki, M. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.* **1991**, *32*, 400–407.
- (49) Yoshida, H. Construction of higher order symplectic integrators. *Phys. Lett. A* **1990**, *150*, 262–268.
- (50) Tuckerman, M.; Alejandre, J.; López-Rendón, R.; Jochim, A.; Martyna, G. A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal-isobaric ensemble. *J. Phys. A: Math. Gen.* **2006**, *39*, S629–S651.
- (51) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (52) Liu, Y.; Tuckerman, M. E. Generalized Gaussian moment thermostating: A new continuous dynamical approach to the canonical ensemble. *J. Chem. Phys.* **2000**, *112*, 1685–1700.
- (53) Spek, A. L. Single-crystal structure validation with the program PLATON. *J. Appl. Crystallogr.* **2003**, *36*, 7–13.
- (54) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.
- (55) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.
- (56) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (57) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (58) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (59) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758.
- (60) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **1976**, *13*, 5188.
- (61) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. K.; Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.* **2009**, *42*, 339–341.
- (62) Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr., Sect. C: Struct. Chem.* **2015**, *71*, 3–8.
- (63) Sheldrick, G. M. SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr., Sect. A: Found. Adv.* **2015**, *A71*, 3–8.