Article

Systematic Finite-Temperature Reduction of Crystal Energy Landscapes

Nicholas F. Francia, Louise S. Price, Jonas Nyman, Sarah L. Price, and Matteo Salvalaglio*

Cite This: https://dx.doi.org/10.1021/acs.cgd.0c00918

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Crystal structure prediction methods are prone to overestimate the number of potential polymorphs of organic molecules. In this work, we aim to reduce the overprediction by systematically applying molecular dynamics simulations and biased sampling methods to cluster subsets of structures that can easily interconvert at finite temperature and pressure. Following this approach, we rationally reduce the number of predicted putative polymorphs in crystal structure prediction (CSP)-generated crystal energy landscapes. This uses an unsupervised clustering approach to analyze independent finite-temperature molecular dynamics trajectories and hence identify a representative structure of each cluster of distinct lattice energy minima that are effectively equivalent at finite temperature and pressure. Biased simulations are used to reduce the impact of limited sampling time and to estimate the work associated with polymorphic transformations. We demonstrate the proposed systematic approach by studying the polymorphs of urea and succinic acid, reducing an initial set of over 100 energetically plausible CSP structures to 12 and 27 respectively, including the experimentally known polymorphs. The simulations also indicate the types of disorder and stacking errors that may occur in real structures.
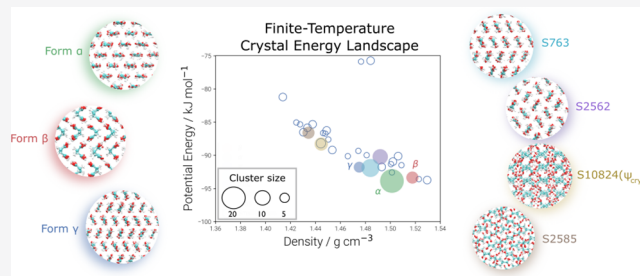
INTRODUCTION

In this paper, we introduce and test a protocol to tackle the issue of overprediction,[1,2] which plagues all currently used crystal structure prediction (CSP) methods, by using molecular dynamics (MD) simulations and enhanced sampling methods. We reduce, or coarsen, the landscapes of crystal structures generated by CSP into a smaller set of crystal structures that are persistent, distinct at finite temperatures, and more likely to correspond to real polymorphs.

CSP has evolved over the past decade to the point where the identification of local minima in the rugged potential energy surface associated with the packing of irregularly shaped organic molecules has been established as the standard approach in the field.[3] Search algorithms are now able to efficiently generate putative packings, and while the existence of suitably accurate methods for the description of the potential energy cannot be assumed, *ab initio* and semiempirical methods[4−8] have evolved to provide increasingly accurate lattice energy estimates.[2] Applications of CSP methods based on the estimate of (temperature agnostic) relative lattice energies, here identified by the abbreviation CSP_0,[9] are blossoming within both industry and academia, finding applications in pharmaceutical manufacturing and functional materials design.[10−12] Nevertheless, CSP_0 techniques only partially capture the physics underlying polymorphism. Thermal effects on thermodynamic stability, as well as kinetics of nucleation and growth, play a determining role in selecting which putative structures can be experimentally observed and under which specific conditions.[1,13,14]

The computational protocol presented in this work aims to reduce the number of crystal structures predicted by CSP_0 methods—hence achieving a coarsening of CSP_0 crystal energy landscapes. This idea finds its roots in the work of Mooij & van Eijck[15] and Gavezzotti,[16] and builds on the recent contributions by Tuckerman et al., who have used molecular dynamics[17,18] and introduced enhanced sampling methods[19,20] to complement CSP_0 in several applications, including the sixth blind test of organic CSP.[2] Our approach is based on identifying:

- Labile structures—i.e., crystal structures that, while representing valid potential energy minima, are unstable when simulated at finite temperature and can thus be discarded.[2,15,16,21]
- Effectively equivalent structures—i.e., structures representing distinct potential energy minima, which can easily interconvert and merge into the same dynamic ensemble at finite temperature.[22]
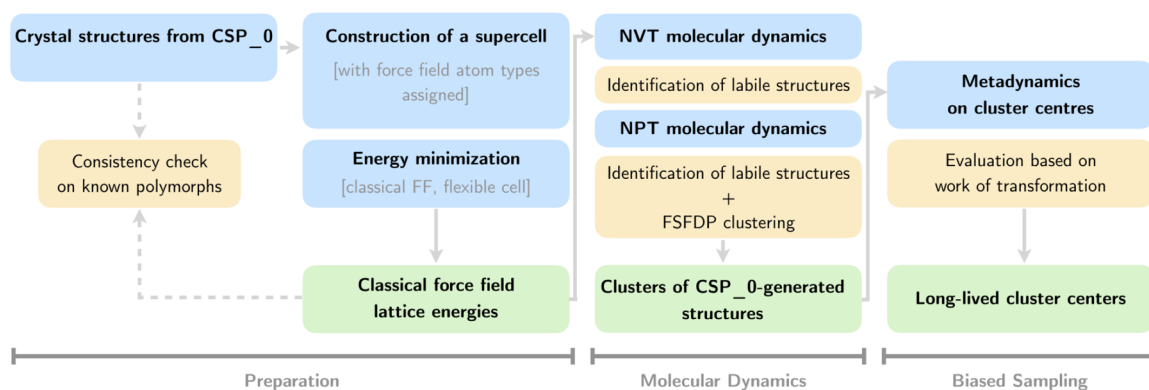
Figure 1. Workflow for the finite-temperature coarsening of crystal energy landscapes, showing in blue the simulations performed, in orange the analysis, and in green the results obtained. The preparation stage is devoted to the construction of supercells from CSP_0 and experimental data. The molecular dynamics section allows removal of structures that are unstable at the temperature of interest (T = 300 K in this work) and identification of dominant structural clusters. The biased sampling section involves computing the work necessary to observe transitions in the structures that emerge as stable from the MD screening.

Coarsening crystal energy landscapes by eliminating labile structures and merging ensembles of essentially equivalent structures enables the identification of significantly fewer target structures, thus reducing the computational cost of sophisticated calculations such as those necessary to implement highly accurate methods for the assessment of the relative thermodynamic stability (CSP_thd),[23] or even for the explicit simulation of out-of-equilibrium nucleation and growth processes in solution.

This paper is organized as follows. First, we describe the coarsening workflow, detailing the approach to clustering finite-temperature crystal structures, and in estimating the work associated with crystal structure transitions. We then illustrate the reduced landscapes obtained for two representative systems: urea and succinic acid. The former is a common example of a nearly rigid organic molecule able to generate a multitude of hydrogen-bonding patterns in the solid state, while the latter is an example of a small but conformationally flexible molecule, subject to conformational polymorphism. For both cases, we significantly reduce the number of predicted polymorphs and identify the most relevant packing motifs among those generated by CSP_0.

In essence, we present methods for reducing the number of low energy crystal structures from temperature-free CSP, using molecular dynamics and enhanced sampling techniques, which result in a small number of thermally stable putative polymorphs.

■ METHODS

In order to coarsen CSP_0-generated crystal energy landscapes, we aim to computationally assess the persistence of plausible crystal packings at constant temperature and pressure. To this aim, we perform MD simulations at finite temperature and pressure on supercells obtained from CSP_0 structures. Statistically meaningful MD simulations of supercells containing hundreds of molecules are impractically computationally demanding for *ab initio* MD methods, thus requiring traditional force fields. Here we describe the computational workflow adopted in this work, highlighting the clustering approach that allows us to identify recurrent packing motifs independently from the symmetry of the parent crystal structure, the size of the supercell, the orientation of the supercell, Z', or even the presence of point defects. The workflow described in the paper is implemented in an *ad hoc* Python library, available upon request.

Coarsening Workflow. The CSP_0 crystal energy landscape is the set of lattice energy minima generated in a search through a defined range of degrees of freedom (space groups, Z') which are sufficiently low in energy to be thermodynamically plausible as polymorphs, allowing for the expected range of lattice energies between polymorphs and the likely errors. This often contains hundreds of structures, with the number often increasing with the flexibility and size of the molecule. In order to screen such a large number of crystal structures at finite T and P, it is necessary to efficiently set up and perform molecular simulations in various ensembles of relatively large supercells containing hundreds of molecules. Here we follow the workflow detailed in Figure 1, which is composed of a *preparation*, a *molecular dynamics*, and a *biased sampling* section. In the following, we discuss each section in detail.

Structure Preparation and Atom Typing. In order to perform classical simulations of supercells starting from crystallographic files generated from either CSP calculations or experiments, it is necessary to replicate the asymmetric unit according to the space group symmetry, and to deduce any bonding between atoms, to obtain a simulation cell containing a finite number of whole molecules. The preparation of input files is a tedious process that can be carried out by hand only for a limited number of crystal structures; however, when dealing with sets containing hundreds of structures, a systematic approach is required. These tasks were carried out by interfacing to existing utilities with an *ad hoc* Python library.

Force field parameters need to be assigned in order to carry out classical MD simulations. In this work, we used the General Amber Force Field (GAFF), and force field atom types were assigned by interfacing with the AmberTools suite.[24] The framework that we have developed is however flexible enough to handle other force fields as well. For instance, by interfacing with Tinker,[25] we can use the AMOEBA polarizable force field.[26,27]

Finally, the asymmetric cell is replicated to obtain a supercell with a shape as close as possible to cubic, complying with the minimum image convention dictated by direct space cutoff for nonbonded interactions (see the following paragraphs). This led to the inclusion of at least 256 and 400 molecules for succinic acid and urea, respectively.

Energy Minimization. The first stage of the coarsening process is a lattice energy minimization in which we relax the crystal supercell. In this work, the optimization of the atomic coordinates is carried out in GROMACS,[28] using the steepest descent algorithm. To carry out this first step, the group cutoff scheme was used to generate neighbor lists every 10 steps, and long-range electrostatic interactions were treated with the smooth particle-mesh Ewald (PME) method.[29] Van der Waals and electrostatic interactions in real space were calculated with cutoffs of 0.8 and 1.0 nm for urea and succinic acid, respectively. In order to relax cell parameters under an applied pressure of 1 atm, we used the LAMMPS Molecular Dynamics package,[30] which implements this option. The input conversion between the two packages is made possible by interfacing with the InterMol software tool.[31] After cell relaxation in LAMMPS, a final coordinates optimization is performed in

GROMACS, to avoid small differences in the crystal cell energies derived from implementation differences between the two codes.[31]

After minimization, we estimate the lattice energy E^{latt} of each crystal structure with the same force field that will be used in subsequent stages of the coarsening process, using the following definition:

$$E^{latt} = \frac{E^{crystal}}{n_{mol}} - E^{vacuum} \tag{1}$$

where $E^{crystal}$ is the potential energy of the supercell, n_{mol} is the number of molecules in it, and E^{vacuum} is the energy of an isolated molecule in its lowest-energy conformation.

The use of a different approximation for the calculation of energies than the one used for CSP_0 ranking induces differences in the ranking of crystal structures across the set of CSP_0-generated crystal structures. It should be noted that similar differences are also observed when comparing CSP_0 energies with dispersion-corrected density functional (DFT+D) estimates of the same quantities. In this work, we have compared the CSP_0 lattice energies with those computed with GAFF for both urea and succinic acid. In the case of succinic acid, we have also compared lattice energies with those obtained with a tailor-made force field[32,33] and periodic DFT+D. The two force fields performed similarly in the calculation of lattice energies; however, GAFF emerged as a better choice to model conformational transitions in MD simulations. Therefore, we have chosen to use GAFF for our study. A comparison between the lattice energies and structures obtained from supercell optimization with GAFF and those used as input from the CSP_0 stage is used to check that they are sufficiently consistent and realistic for the force field to be used in the following stages. A comparison of the lattice energies for the MD force field (GAFF) and those from the CSP_0 studies is given in SI section S1.

MD Simulations. In the context of this study, we define as labile structures the CSP_0-generated structures that spontaneously melt at room temperature and pressure, losing long-range order during MD simulations. In order to identify such labile crystal structures, we perform MD simulations in three steps. At first, we simulate the supercells for 5 ns in the canonical ensemble. In such simulations, performed with GROMACS, PME[29] is used to treat long-range electrostatics, while a real space cutoff for the calculation of van der Waals and electrostatic interactions is set to 0.8 and 1.0 nm for urea and succinic acid, respectively. In all simulations, we adopted a time step of 1 fs, and controlled the temperature with the Bussi–Donadio–Parrinello thermostat. The canonical simulation is followed by an isothermal–isobaric simulation, in which pressure is initially equilibrated for 1 ns with a Berendsen anisotropic barostat. The pressure equilibration step is followed by a production simulation performed using the Parrinello–Rahman barostat. The typical simulation length for this step is 15 ns. We identify labile structures and effectively equivalent structures by analyzing the trajectory generated in both the canonical and isothermal–isobaric trajectories using probabilistic structural fingerprints discussed in detail in the next paragraph.

Defining a Dissimilarity Measure. In order to calculate the similarity between crystal packings, we needed to be able to consistently compare supercells with different numbers of molecules, different shapes, and different unit cells. More importantly, an effective similarity metric needs to handle atomic configurations generated at finite temperature, in which atoms are fluctuating around their equilibrium position, and the perfect symmetry of static structures generated in CSP_0 is absent. To this aim, we base our approach on the definition of hierarchical, probabilistic fingerprints, able to capture key characteristics of the fluctuating local molecular environment of crystal structures. Inspired by the descriptors proposed in a recent paper on CO$_2$ packing polymorphism,[35] and by results published by other groups,[36,37] we define fingerprints encoding information on crystal structures in the form of a set of probability densities. Following the extended pair distribution function approach at the heart of the definition of order parameters developed for molecular crystals,[38–40] we include in the fingerprints information on molecular position, relative orientation, and molecular conformation. The fingerprint associated with each structure i is the set $F_i = \{p_i(r_{COM}), p_i(\vec{\theta}), p_i(\vec{\phi})\}$.

The first level component of the fingerprint set is $p_i(r_{COM})$, the probability density, in the form of a kernel density estimate, of distances between molecular centers of mass, r_{COM}. The second level of the fingerprint is $p_i(\vec{\theta})$ a distribution of intermolecular angles $\vec{\theta}$, describing the relative orientation of neighboring molecules. Finally, the third level fingerprint is the distribution of intramolecular torsional angles $p_i(\vec{\phi})$, capturing the conformational state of a molecule. For molecules that are rigid and very isotropic in shape, the first component of the fingerprint is sufficient; for anisotropic, rigid molecules the first two components are necessary, while for anisotropic flexible molecules all three components contribute in capturing the salient features of a local molecular environment.

By quantitatively comparing the structural fingerprint of two different structures generated from the analysis of finite-temperature MD simulations, we can assess the similarity of simulation snapshots of different structures. Let us begin by considering structures i and j, characterized by fingerprints F_i and F_j. We define the dissimilarity between each of the components of the fingerprint, say p_i and p_j by computing the Hellinger distance H_{ij}:

$$H_{ij} = \sqrt{1 - BC(p_i, p_j)} \tag{2}$$

where $BC(p_i, p_j)$ is the Bhattacharyya coefficient, defined as

$$BC(p_i, p_j) = \int \sqrt{p_i(\vec{\xi}) p_j(\vec{\xi})} \, d\vec{\xi} \tag{3}$$

where for the sake of generality, $\vec{\xi}$ is used to indicate a vector variable. We note that in this work $\vec{\xi}$ reduces to a scalar in the case of r_{COM}, while it will usually have dimensionality larger than one in the case of $\vec{\theta}$ and $\vec{\phi}$. While $\vec{\theta}$ can be a 3D distribution at most, the dimensionality of $\vec{\phi}$ depends on the relevant torsional angles of the molecule. In contrast to other measures of similarity between probability distributions, such as the KL divergence used by Gobbo et al.[36] or the generalized entropy introduced by Piaggi et al.,[37] the Hellinger distance is inherently symmetric, does not diverge when p_i and p_j are sparse, and therefore does not require regularization. By comparing all the components of F_i and F_j, we obtain a vector of Hellinger distances $\mathbf{H}_{ij} = [H_{ij}^{r_{COM}}, H_{ij}^{\vec{\theta}}, H_{ij}^{\vec{\phi}}]$. The dissimilarity metric between structures i and j, Δ_{ij}, is thus defined as the norm of vector $\mathbf{H}_{ij}$: $\Delta_{ij} = \|\mathbf{H}_{ij}\|$, appropriately normalized as discussed in the Supporting Information (SI). We can apply this dissimilarity metric to each pair of a set of N crystal structures to finally construct a $N \times N$ distance matrix, $\mathbf{\Delta}$.

Identifying Labile Structures. Having defined a measure of the dissimilarity between crystal packings, we can automate the identification of labile structures, i.e., structures that despite being local minima of the potential energy are unstable at finite temperature and melt during NVT or NPT simulations. To this aim, we focus on $p_i(\vec{\theta})$, the intermolecular orientation component of F_i. In order to identify the loss of order associated with melting in the structure, we compute the dissimilarity between $p_i(\vec{\theta})$ and a uniform distribution in $\vec{\theta}$, corresponding to a completely disordered packing,[35] here labeled with l.

In this paper, we consider one intermolecular torsional angle, θ, as a descriptor of the relative orientation of molecules since it is extremely sensitive to melting, and limiting the analysis to this component reduces the overall computational cost of the procedure. This choice is effective in both the urea and succinic acid cases and can be generalized to properly describe the orientation of any molecule by using three angles.[38] However, a single angle θ may come up short for molecules that form plastic crystals in which they freely rotate. In case this would emerge as an issue for specific systems, it would be straightforward to extend the criterion for the identification of a labile structure to the full fingerprint.

The uniform distribution of a single intermolecular torsional angle, representative of a disordered packing, is $p_U(\theta) = 1/2\pi$, hence:

$$H_{il} = \sqrt{1 - \int \sqrt{\frac{1}{2\pi} p_i(\theta)} \, d\theta} \tag{4}$$
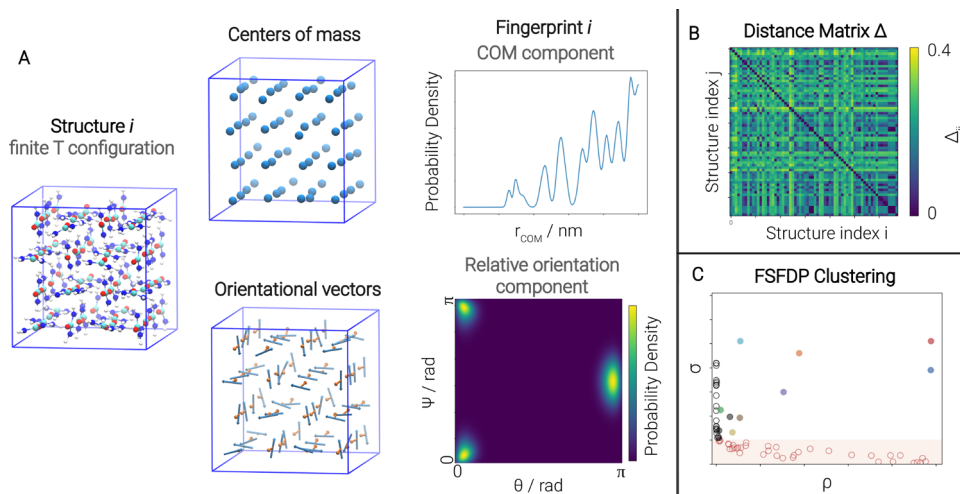
Figure 2. Steps in clustering of finite-temperature supercell configurations obtained from MD simulations. (A) Starting from a generic structure i, we calculate the centers of mass and determine a set of vectors able to identify the relative orientation of molecules. From these, we define a structural fingerprint as a set of probability densities, as described in the Defining a Dissimilarity Measure section, obtained considering each pair of molecules and averaging over the different time frames. To simplify, a small supercell with 64 molecules is shown in this image, but the simulation boxes for urea contain from 400 to 800 molecules. (B) By computing the norm of the dimensionless Hellinger distances, Δ_{ij}, between distributions of each pair of structures, we obtain a distance matrix suitable for clustering analysis. (C) The decision graph of urea related to the FSFDP clustering algorithm,[34] introduced in the Clustering finite-temperature snapshots of crystal structures section, using a distance cutoff of $\sigma = 0.05$ that divides the cluster centers from the other structures (in the red shaded area). These other structures are assigned to the different clusters based on the vicinity to the structures of higher densities, i.e., σ. Groups of more than two structures are shown with different colors, while unique geometries are identified with a black circle.

For a labile structure, $H_{il} \rightarrow 0$, which means that by monitoring this quantity over the simulation time it is possible to detect structures that melt without visual inspection, which would be impractically time-consuming for large sets of structures. The labile structures are removed from the analysis.

Clustering Finite-Temperature Snapshots of Crystal Structures. The distance matrix Δ can be exploited to identify clusters of structures which are likely to be indistinguishable at finite temperature. With it at our disposal, we can apply distance-based clustering methods, such as the Fast Search and Find of Density Peaks algorithm (FSFDP)[34] to identify the number of clusters, their centers, and structures identified as members of the same cluster. For details on this clustering approach, we refer the interested reader to ref 34. In Figure 2B, we report the distance matrix Δ computed for the urea data set after the NPT stage and in Figure 2C the ρ, σ decision graph identifying the CSP_0 structures that correspond to cluster centers according to the FSFDP algorithm. Cluster center structures are used to compute further properties. The choice of clustering parameters is discussed in the Supporting Information.

Metadynamics. After the MD stage and the clustering of all the *surviving* structures (see Figure 1), we have reduced the initial set of crystal packings included in the CSP_0 crystal energy landscape to a smaller set of cluster centers. Any practical length of the MD step is inherently insufficient to ensure that all the cluster centers will not readily transform, and some cluster centers may be separated by such a small barrier that only the most stable would crystallize. Therefore, in order to efficiently assess the persistence of the structures corresponding to cluster centers, and to estimate the work associated with their transition, we introduce an additional step based on the application of an adaptive bias potential through metadynamics. We perform Well Tempered metadynamics (WTmetaD) on the crystal supercells corresponding to the cluster centers identified after the NPT stage. The bias is added as a function of two collective variables (CV) that correspond to the coordinates typically used to represent structures in a crystal energy landscape, namely, potential energy and density. These variables are chosen for their generality, and low computational cost, and are expected to enhance transitions between similar structures. We note that the choice of these CVs would be suboptimal for accurately computing free energy differences between polymorphs. Typically, CVs that can be used for free energy calculations need to be

able to discriminate between specific crystal structures and to reversibly enhance transitions between them. Several examples of CVs used for this scope can be found in the enhanced sampling literature.[17,19,20,37,38,41,42] In typical WTmetaD performed in this work, the bias potential is updated every 1 ps with Gaussians characterized by an initial height of 2 kJ mol^{-1} and a width of 20 kg m^{-3} for the density and 2 kJ mol^{-1} for the lattice energy. These simulations are performed using GROMACS patched with PLUMED 2.[43] Typical PLUMED input files required to reproduce the results reported in this paper are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the PLUMED consortium,[44] as plumID:20.019.

To perform a statistical analysis on the transition pathways that emerge from biased simulations and to converge estimates of the work necessary to induce a transition, we performed 10 independent simulations for every cluster center structure.

In order to detect transitions without visually inspecting the trajectories obtained from WTmetaD we analyzed, on the fly, the reweighting factor $C(t)$, defined as

$$C(t) = \frac{1}{\beta} \log \frac{\int ds\ e^{-\beta G(s)}}{\int ds\ e^{-\beta(G(s)+V(s,t))}} \tag{5}$$

where $\beta = 1/k_B T$, $G(s)$ is the Gibbs free energy, and $V(s, t)$ the bias potential. Following the argument introduced by Tiwary and Parrinello,[45] transitions to a new basin produce a marked change in the slope of $C(t)$. Furthermore, $C(t)$ represents the work performed on the system through the introduction of a bias potential. By averaging the reweighting factor of the different trajectories obtained at the time when a transition occurs, we estimate the average work associated with a transition as $W = \langle C(t) \rangle$. The behavior of $C(t)$ associated with the transformation of a form I crystal supercell of urea during metadynamics is reported in Figure 4A.

Together with the work associated with a transition, we are also interested in tracking to which form each structure converts. To this aim, we compute structural fingerprints and compare them to those of the cluster centers obtained at the end of the NPT stage. If the fingerprint of the structure obtained by evolving a supercell with WTmetaD does not match any of the existing ones, a new reference crystal structure is created. We also check if the system visits any intermediate structure by comparing configurations at the beginning
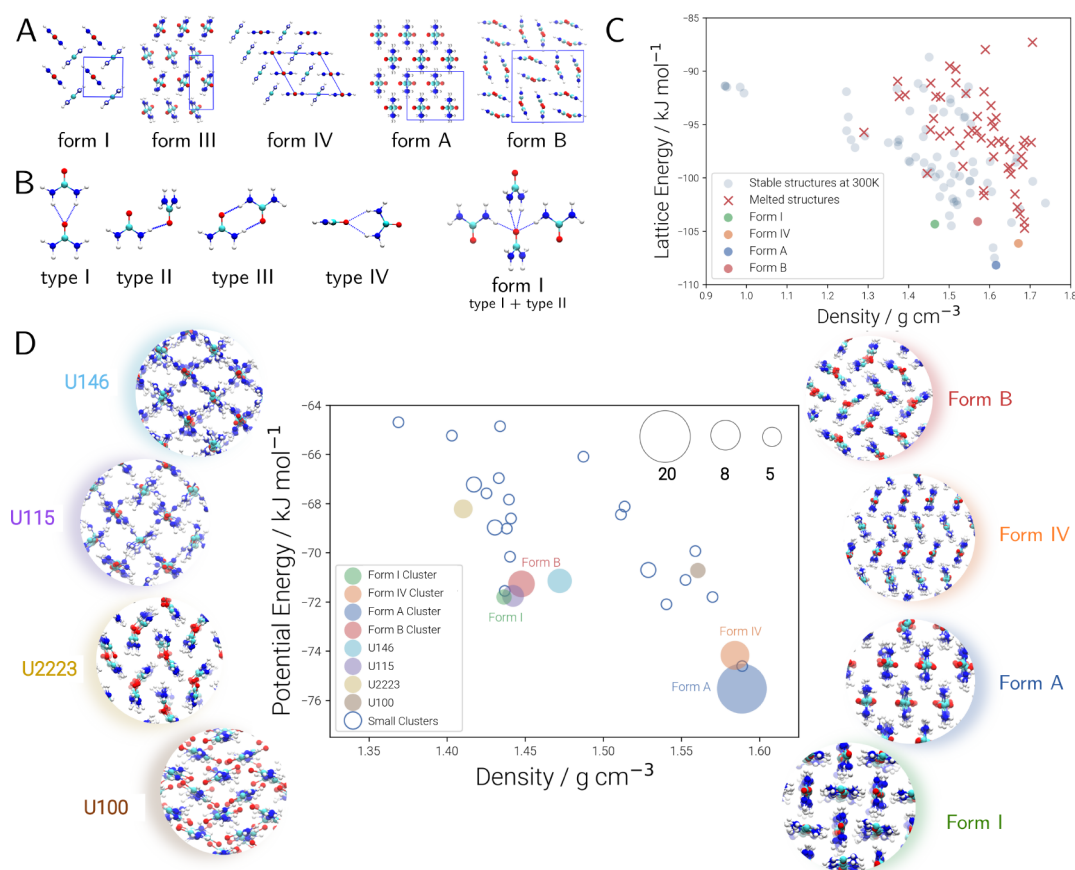
Figure 3. Analysis of the finite-temperature structures of urea. (A) The known and proposed putative polymorphs of urea.[37] From left to right form I, form III, and form IV are the experimental polymorphs, while form A and form B are structures previously identified as stable with the GAFF force field. (B) H-bonding types present in most of the CSP_0-generated structures of urea. Following the nomenclature in ref [46], these are labeled as type I, type II, type III, and type IV, according to H-bond distances. The characteristic H-bond motif found in form I, which combines type I and type II, is shown. (C) Lattice energy landscape of the CSP_0-generated structures of urea optimized with the GAFF force field. The plot shows those structures that melt during the equilibration at 300 K and 1 bar as red crosses. (D) Finite-temperature crystal energy landscape. The size of each circle corresponds to the number of CSP_0 structures that convert to the same geometry according to the legend in the upper right side.

and at the end of the portion of the trajectory in which $C(t)$ plateaus. We check whether the system goes back to the initial state when $C(t)$ begins increasing again, i.e., if the transition is temporary. This is done by comparing the structures before and after the time-span in which $C(t)$ is constant.

RESULTS

Urea: An Example of Packing Polymorphism. Urea is a small and relatively rigid organic molecule whose crystal structures have been heavily studied by both the experimental and modeling communities. Many X-ray diffraction studies have been performed on form I, which is the stable phase of urea at ambient pressure and temperature.[47] Other high-pressures phases of urea were identified in the first half of the twentieth century, but their structures were only recently determined.[47−52] Urea form I transforms to form III at 0.48 GPa, and to form IV at 2.8 GPa.[51] A form I ↔ form II transition was reported in 1916 by Bridgman at 373 K and 0.60 GPa, but this has not been reproduced,[47,51] and since a direct transition from form I to form IV was observed at 0.70 GPa and 396 K, it seems likely that form IV coincides with form II.[52] In 2002, a new polymorph was discovered at 7.2 GPa, but no structure was obtained.[50] However, the structures of the high pressure forms III and IV show variations in the molecular structure consistent with changes in its electronic structure, so it is unclear how well any

atomistic simulations could model those high pressure phases. Increasing the pressure leads to polymorphs with a greater tilt of the molecules and a more compressed structure.[53]

Together with the experimental work, urea has been the subject of several theoretical studies on both nucleation and crystal growth.[37,39,46,54−60] In particular, from force field based MD simulations, a new form similar in energy to form I has been found.[37,39,46,58,59] This form exhibits a H-bonding motif similar to the experimental forms I, III, and IV but has a higher density. Following the nomenclature reported in ref [37], this form will be referred to as form A. In the same paper, a new stable structure called form B was proposed. Unlike form A, no similarities between the experimental structures and form B have been observed. It was not clear whether the simulated favorability of structures A and B is an artifact of the MD force field, or whether they could be undiscovered polymorphs, so they are tracked during the workflow. Representative images of the key urea structures and the main hydrogen bonding types are shown in Figure 3, panels A and B, respectively.

CSP_0 Study and Initial Structure Set. The crystal structure of urea form I has a low-frequency high-amplitude normal mode which combines out-of-plane deformations of the amino group with molecular librations, such that the pyramidalization at the N atom allows the hydrogen bonding to be maintained.[61] The experimental atomic displacement parameters that show the

Table 1. Stable and Metastable Structures of Urea with Their H-Bonding Motif Types, Ordered by Their Potential Energy[a]

Initial Structure	Potential Energy (kJ mol^{-1})	H-bonding motifs	Resulting Structure	H-bonding motifs	W (kJ mol^{-1})
Form A	−75.52	I + III	–	–	$>k_BT$
U1955	−74.61	I + III	Form I	I + II	2.34
Form IV	−74.17	I + II	Form A	I + III	0.61
U785	−72.10	I + II	Form I	I + II	0.72
U499	−71.80	I + II	U146	II + III	1.37
Form I	−71.80	I + II	Form A*	I + III	0.87
U115	−71.76	I + II	Form A*	I + III	1.29
Form B	−71.27	III	–	–	$>k_BT$
U146	−71.14	II + III	–	–	$>k_BT$
U638	−71.10	–	–	–	$>k_BT$
U100	−70.72	I + III	–	–	$>k_BT$
U1831	−70.16	I + II	Form IV	I + II	0.90
U81	−69.93	I + III	–	–	$>k_BT$
U1540	−69.02	III	–	–	$>k_BT$
U2015	−68.98	III	–	–	$>k_BT$
U302	−68.61	I + III	–	–	$>k_BT$
U721	−68.45	III	melted	–	1.88
U2223	−68.22	III	Form A	I + III	0.70
U49	−68.13	I + II	Form A	I + III	1.87
U321	−67.84	–	–	–	$>k_BT$
U771	−67.59	III	melted	–	2.35
U802	−67.23	III	–	–	$>k_BT$
U2449	−66.97	I + III	Form A	I + III	1.00
U562	−66.10	III	melted	–	2.48
U1960	−65.24	III	–	–	$>k_BT$
U792	−64.86	III	melted	–	1.91
U1179	−64.70	–	melted	–	1.03

[a]The 12 structures in white result as putative polymorphs in this study. Excluded structures are shown in red if their work of transformation W is less than k_BT, and in blue if they exhibit a partial melting of the simulation box or defects. If a structure undergoes a phase transition, the resulting geometry and its H-bonding motif types are shown. The labile structures U2403, U1140, and U628 are not listed as they melt in the first steps of the simulation, and so no estimation of W is possible.

motion of the NH_2 group in the crystal and gas-phase spectra show that there is a very low barrier to pyramidalization. The crystal structure is an average over vibrational motions, with both the planar NH_2 group and exact tetragonal symmetry being transition states on many potential energy surfaces, leading to considerable challenges in modeling urea by lattice energy minimization.[62,63] Hence, to avoid these complications, the urea molecule was kept rigid throughout this study, using a PBE0/aug-cc-pVTZ optimized conformation with urea constrained to be planar using Gaussian 09.[64] A rigid search was carried out using CrystalPredictor[65] in 58 space groups with one molecule in the asymmetric unit cell, using PBE0/aug-cc-pVTZ atomic point charges and the empirical FIT repulsion-dispersion potential.[5,66] After clustering to remove duplicates, the structures were labeled by their rank order. The structures were then lattice energy minimized using DMACRYS[7] with the electrostatic model improved by using distributed multipoles obtained from the PBE0/aug-cc-pVTZ charge density with GDMA 2.2,[67] and reducing the symmetry of all structures that were transition states within the space-group constraints. Further clustering was performed to remove duplicates, as defined by RMSD$_{15}$, the optimum root-mean-square deviation of the non-hydrogen atoms in an overlay of 15 molecule clusters, calculated with COMPACK.[68] Structure U472 was found to correspond to experimental form I with an RMSD$_{15}$ of 0.017 Å, while structure U411 corresponds to form IV with an RMSD$_{15}$ of 0.011 Å. Theoretical forms A and B correspond to structures U2363 and U849, respectively.

Following the nomenclature in ref 46, most of the structures present some of the three main hydrogen-bonding motifs shown in Figure 3B. All experimental forms and form A exhibit a type I H-bond network, while form B is formed of type III chains. However, it is important to notice that experimental forms are actually combinations of type I and type II, while form A presents also type III H-bonds.

Energy Minimization and Preliminary Ranking. The geometry of the different structures can now be used in GROMACS. The GAFF force field was used, constraining the improper dihedrals so that the molecules are planar as in the CSP_0 study. A preliminary ranking can be done based on the lattice energies, as reported in the Supporting Information. The previously known and proposed geometries are all among the most stable structures with form A being the most stable. In the CSP_0 ranking, experimental forms I and IV are placed in 60th and 41st positions with an energy difference with respect to the global minimum of +9.0 kJ mol^{-1} and +7.8 kJ mol^{-1}, respectively. This difference is smaller when the GAFF force field is used, with both the experimental structures appearing among the 10 most stable crystalline configurations.

Finite-Temperature Simulations and Clustering. In order to assess the lifetime of CSP_0-generated crystal structures, we apply the methodology introduced in the previous section. After equilibration, an additional 10 ns simulation in the NPT ensemble is then performed to check the stability of the crystal structures that did not melt during the initial NVT/NPT equilibration of 5 ns each and to collect statistics on their the properties.
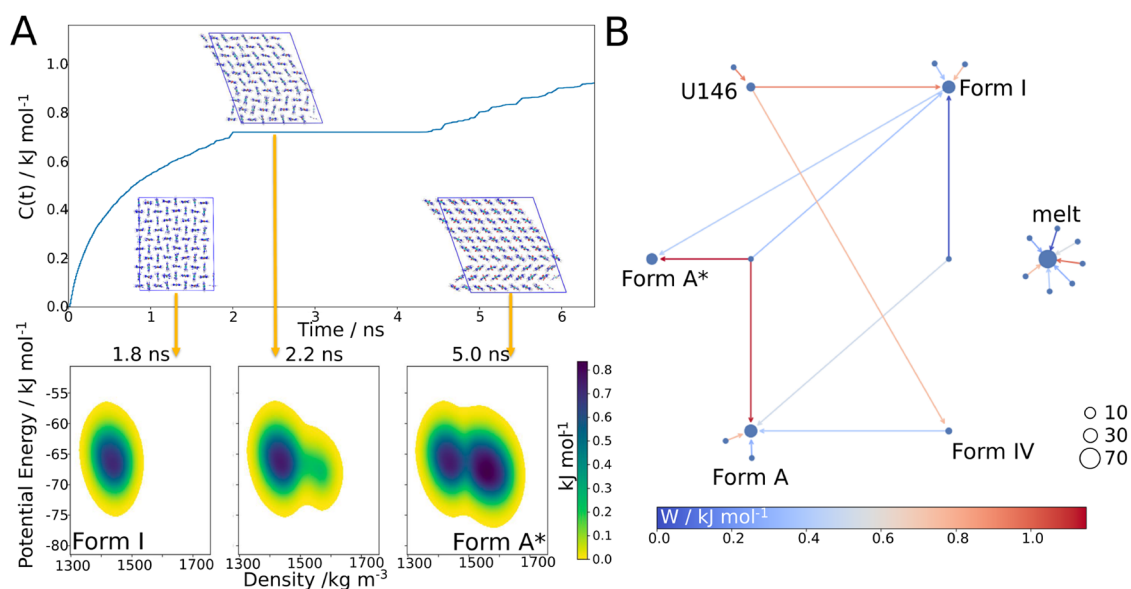
Figure 4. (A) Variation of the reweighting factor, $C(t)$, for the transition from urea form I to form A* in a WTmetaD simulation. The asterisk in A* indicates the presence of stacking defects. On the bottom of the image, the bias potential in the CVs space before (1.8 ns), during (2.2 ns), and after (5.0 ns) the solid−solid phase transition. (B) Conversion network showing the first phase transition of the different cluster centers in WTmetaD. Replicas of the same structure can transform to different geometries. The circle size refers to how many replicas converted to that specific structure, while the color of the arrows refers to the value of the free energy difference between the initial and transition states.

In Figure 3C, we report the CSP_0 crystal energy landscape, showing the structures that spontaneously melt during MD equilibration at 300 K. We note that around 40% of the structures that were obtained from CSP_0 melt at 300 K and 1 bar.

After removing the labile structures, we analyzed the trajectories obtained from MD with structural fingerprints defined following the procedure outlined in the Methods section. To characterize urea crystal structures, we considered the contribution of relative position and orientation of urea molecules in the definition of a structural fingerprint. The orientation component corresponds to the 2D probability density function of the angles ψ and θ between vectors parallel to the C−O bond and to the N−N direction, respectively, shown in Figure 2A.

In Figure 3D, we report a coarsened crystal energy landscape, in which the configurations corresponding to the cluster centers identified by the FSFDP algorithm are reported as a function of the density and the potential energy. All the structures considered in the clustering are putative polymorphs, which did not melt after a 15 ns simulation in the NPT ensemble.

All the previously reported structures of urea, both experimentally isolated or obtained from nucleation simulations, spontaneously emerge as cluster centers in Figure 3D. Form A emerges as the most populated cluster. Forms I and IV, the low and high pressure forms of urea, also emerge as centers of highly populated clusters at low and high density, respectively. Another notable highly populated cluster is centered around form B configurations—the entropically favored structure identified using the same force field by Piaggi et al.[37] from nucleation simulations, reportedly stable at elevated temperature.

Interestingly, a coarsening of the crystal energy landscape leads to clusters of structures that correspond to disordered variants of known polymorphs. For instance, structures in clusters U146 and U115 present the characteristic packing of form I, with a stacking disorder along the axis a. However, while structure U146 is dominated by type-IV H-bonding along the a

direction (see Figure 3B), structure U115 presents a combination of type-I and type-IV H-bonds resulting into one misaligned layer of urea molecules every five.

We note that by compressing the unit cell of cluster U146, we can obtain the characteristic packing of cluster U100. Cluster U2223 exhibits a type III H-bonding motif, and it appears to be a variation of form B. How individual structures were clustered after the unbiased MD stage is reported in the Supporting Information, section S2.

Assessing the Persistence of Long-Lived Crystal Structures. For every cluster center, the work necessary to obtain a phase transformation W is obtained as an average of over 10 independent WTmetaD simulations. The WTmetaD analysis reveals 12 cluster centers do not undergo any transition after at least 30 ns of biased evolution, sampling an energy scale of 2.5 kJ mol^{-1} per molecule. Forms A and B are among these supercell configurations, while forms I and IV can transform as a result of fluctuations in density and energy, enhanced through WTmetaD. A summary of the qualitative behavior of the cluster centers under the effect of biased dynamics is reported in Table 1.

In Figure 4B, we can see the network deriving from the transformations observed between structures during WTmetaD. All the cluster centers considered either convert during WTmetaD to three previously known or proposed geometries (form I, form IV, form A), to the cluster center U146, melt or do not undergo any phase transformation. In addition to perfect supercell structures, we also sampled transitions to supercells compatible with form A, but displaying packing defects. Such structures are labeled in the following as form A*, to indicate their similarity to form A. Snapshots from a typical trajectory in which form I transforms to form A with packing defects are reported as inset in Figure 4A. We note that the possibility that in experiments crystals are produced with packing defects is physically consistent and often found in MD simulations of molecular crystals.[20]
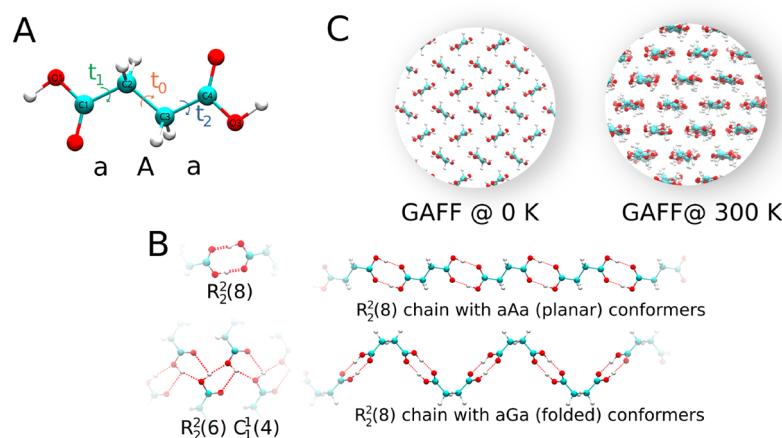
Figure 5. (A) Main torsional angles used to identify the different conformers of succinic acid with the planar (aAa) conformation of forms α and β. (B) The $R_2^2(8)$ H-bonding type and relative motifs for the planar and folded conformers (on the right). On the lower left, the alternative H-bonding motif $R_2^2(6)\,C_1^1(4)$ present in structures S1107, S18857, and S15665. (C) The α structure at 0 K (left) and its transformation at 300 K and 1 bar (right) using the GAFF force field.

In the coarsened landscape, we identify two regions at high and low density, where forms A and I are the most common structures exhibiting a type I H-bonding network. We note that other type I cluster centers during WTmetaD preferentially transform to form I when in low-density regions of the crystal energy landscape, and to form A in high-density regions. In Table 1, we report the structures with their H-bonding motif. Those characterized by type III networks are particularly stable. In most of the cases, even if the putative polymorph undergoes a phase transition, the type III motif is preserved. On the other hand, structures characterized by a type II network, such as the two experimental forms I and IV, easily transform to type III.

In order to use the information gathered at this stage to screen putative polymorph structures in a CSP workflow, we identify crystal structures that are persistent or thermodynamically stable. Here we ascribe to this category all cluster centers that do not exhibit partial melting or defects, that are found within 5 kJ mol^{-1} from the global potential energy minimum, or are associated with a $W > k_B T$. Implementing these criteria gives us a set of 12 structures that are stable and persistent, out of an initial set of 124. The list of remaining structures, ranked by their potential energy, is reported in Table 1.

Succinic Acid: An Example of Conformational Polymorphism. Succinic acid is a small molecule that can adopt a range of conformations in all phases. Three polymorphs of succinic acid are currently known, namely, the α, β, and γ forms. The recently reported γ conformational polymorph has only been obtained once in a failed cocrystallization experiment.[69,70] The high-temperature α phase of succinic acid is stable above 137 °C and can be crystallized from the melt or by sublimation[71] and only slowly transforms to the more stable β form at ambient temperature, depending on sample and conditions.

In this study, crystal structures were classified based on the conformer adopted by their building blocks according to the nomenclature in ref 69, where three dihedral angles t_0, t_1, and t_2, shown in Figure 5A are used to identify the different families of conformers in the gas or liquid phase. Conformers are labeled with three letters associated with the value of t_1, t_0, and t_2. Each one of these are named with the letter *a* or *g* if they are in the *anti* or *gauche* position, respectively. Capital letters are used to indicate the configuration adopted by t_0. In ref 69, a conformational cluster analysis was performed using dSNAP[72]

on all crystal structures containing succinic acid available in the Cambridge Structural Database (CSD).[73] Even though free molecules in the gas phase or in solution can adopt many conformations,[69] this search identified two dominant conformers, namely, the planar (**aAa**, shown in Figure 5A) and folded (**aGa**) conformations. Isolated succinic acid molecules in the gas phase or in solution favor the folded conformation.[69] The two most common experimentally observed polymorphs of succinic acid, β and α, are comprised of planar conformers, whereas the folded conformer of succinic acid is in the recently discovered γ form.

The metastable α form has been found at room temperature, as a contaminant after grinding the β polymorph[74] and in liquid segmented flow crystallization[75] or by spray-drying from water.[76] This suggests that the α polymorph is observed at ambient conditions only when its solvent-mediated transformation to the more stable β is arrested. Although the crystallography of succinic acid dates back to the work of Kathleen Lonsdale in 1924, in the Cambridge Structural Database[73] there is a single structure determination of the atomic coordinates of polymorph α (ID:SUCACB07). The structure has two half molecules in the asymmetric unit, and an improbable location of one of the hydrogen atoms bonded to carbon. This suggests that the α structure may involve a dynamical disorder. Hence, while we would expect the coarsened landscape to include at least the α, β, and γ forms as crystals have been observed at ambient temperature, the ability of the two metastable forms to convert to the stable form β may provide insights into the elusive nature of the metastable polymorphs in this system.

CSP_0 Structure Set. Comparison of CSP_0 studies performed with different potential energy surfaces should produce structures that are quite similar. The two potential energy surfaces may have a slightly different location of their approximation to a well-defined sharp minimum in a structure where the molecules are tightly interlocked, and there would be a substantial barrier to any transformation. Alternatively, the potential energy surface can be very broad and contain a variety of very shallow minima, and this differs between different model potential energy surfaces, so the actual minima found also depend on the optimization route and clustering used in the CSP study. In the former case, using the GAFF force field as a
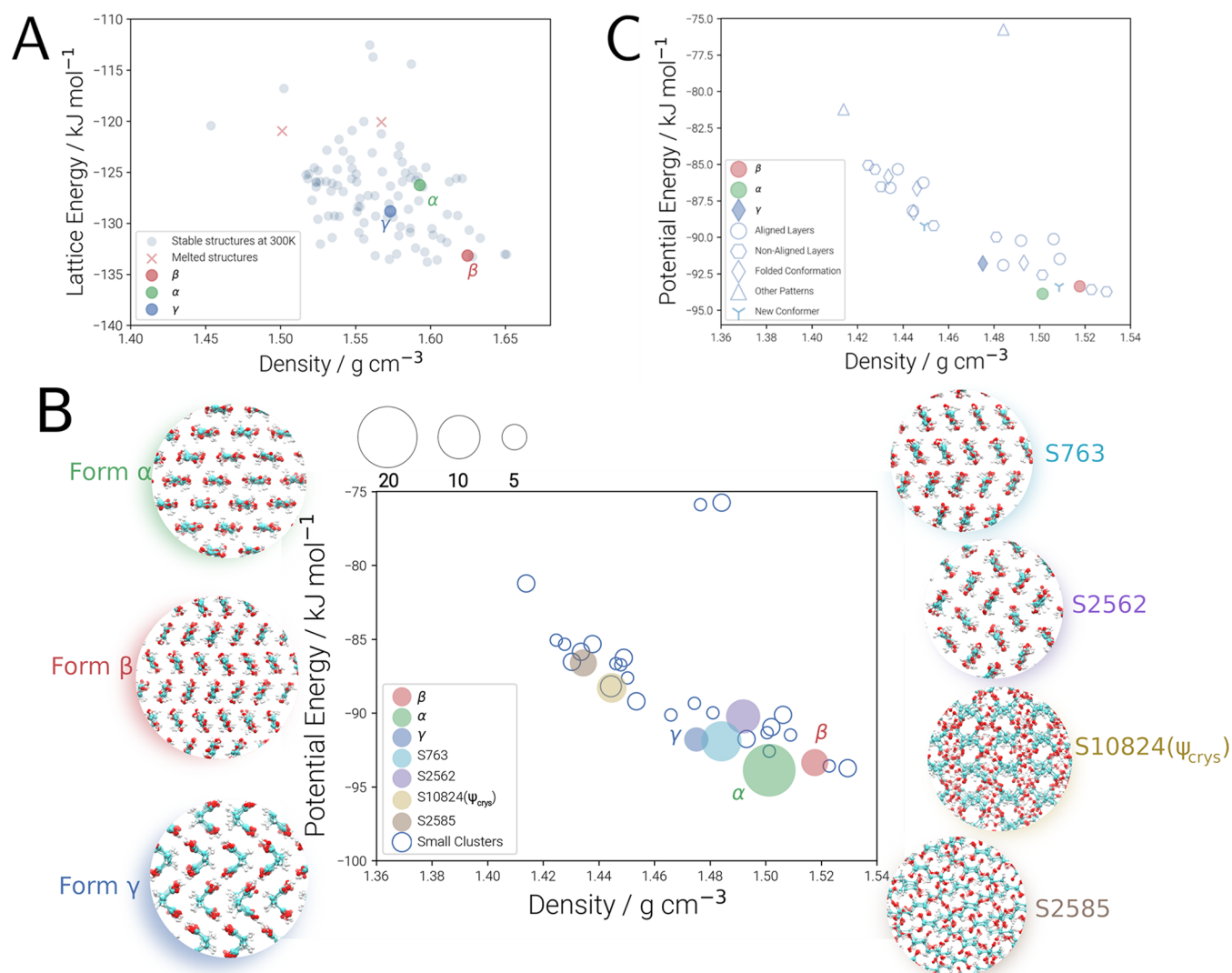
Figure 6. Analysis of the finite-temperature structures of succinic acid. (A) Lattice energy landscape of the CSP_0-generated structures of succinic acid optimized with the GAFF force field at 0 K showing those structures that melt at 300 K as red crosses. (B) Finite-temperature crystal energy landscape of the cluster centers. The size of each point refers to the number of structures that converted to the same geometry. Known structures and most populated clusters are shown on either side of the plot. The different clusters are labeled based on their cluster center. (C) Final finite-temperature crystal energy landscape classified by the motifs observed and including the new structures found during WTmetaD simulations according to the color bar on the lower left side.

third potential surface from either starting structure would result in the same lattice energy minimum and dynamical structure. In the latter case, where the well was broad and weakly undulating in reality, this would differ between approximate potential energy surfaces producing different lattice energy minima, and the dynamic crystal would have large amplitude motions covering the broad well. To test the extent to which the outcome of our workflow is independent of the finer details of the CSP_0 method, we have analyzed a joint set of CSP_0 succinic acid crystal structures obtained with two different methods. The first, here referred to as Ψ_{mol}, is the result of the CSP study published in ref 69. A COMPACK overlay of 30 molecules was performed between the Ψ_{mol} set and the experimental structures. Structures S23778 and S19217 correspond to forms β and γ, respectively, while structure S9393 is a good approximation of form α. Form β is the global potential energy minimum, with the high-temperature form α at +8.9 kJ mol^{-1} and γ at +4.7 kJ mol^{-1}.

A second set of putative crystal structures for the system, here labeled Ψ_{crys}, is obtained through an optimization and reranking of the Ψ_{mol}-generated candidate structures, using a periodic DFT+D method.[77,78] In the Ψ_{crys} set, all structures were geometry-optimized with VASP version 5.4.1.[79] The PBE functional was used,[80] supplemented with the D3BJ dispersion correction.[77,78] The plane wave cutoff energy was 700 eV, and default PAW pseudopotentials were used. For the **k**-points, a Γ-centered Monkhorst−Pack grid with approximately $(2\pi \times)$ 0.06 Å^{-1} spacing was used. Both atomic positions and lattice parameters were optimized together until the forces on all atoms were smaller than 0.01 eV Å^{-1}. Space group symmetry was recovered with PLATON.[81] In the Ψ_{crys} set, experimental forms β, α, and γ are included. For the Ψ_{crys} set, the global energy minimum is the labile polymorph γ, with α at +0.3 kJ mol^{-1} and β at +1.6 kJ mol^{-1}.

Considering the joint Ψ_{mol} and Ψ_{crys} sets of structures, we analyzed a total of 101 CSP-generated crystal structures. As each Ψ_{crys} structure was derived by reoptimizing a Ψ_{mol} structure, in
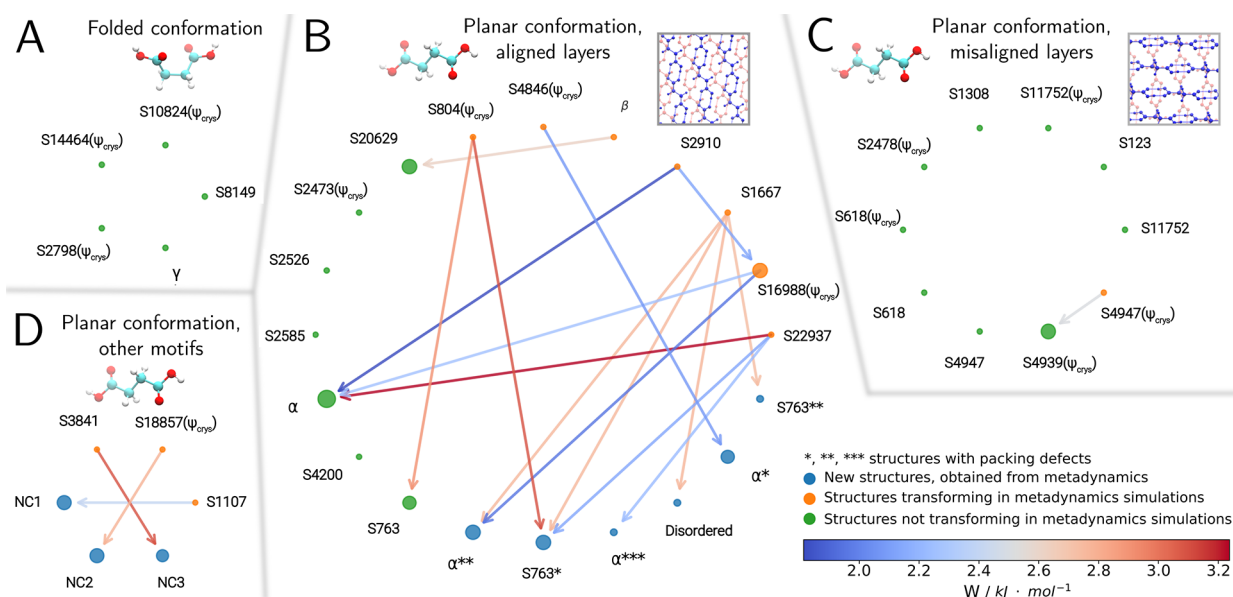
Figure 7. Conversion Network for succinic acid crystal structures. The different structures are named by their cluster centers, and they are divided based on the conformation of the molecules and the orientation of the layers. The circle size refers to how many replicas converted to that specific structure, while their color is orange if a phase transition is observed, green if the structure is kinetically stable, and blue for new structures. The color of the arrows refers to the value of the free energy difference between the initial and transition states. The presence of an asterisk in the representation of the conversion network denotes structures presenting defects within the supercell.

most cases both lead to the same dynamical structure, which is denoted by one label. When the finite-temperature structures were different, they are distinguished by adding (Ψ_{crys}) to the label.

The majority of the structures identified, including experimental ones, present an H-bonding motif of type $R_2^2(8)$[82] that gives rise to H-bonded chains of two types, depending on the conformational state of the succinic acid molecules, as shown in Figure 5B. Only three structures exhibit the more complicated H-bonding network shown on the lower left of Figure 5B. The crystal structures mainly differ in the conformation of the molecule and the different orientations of chains and layers. This determined the choice of fingerprints implemented for the clustering of finite-temperature structures.

The force field in the finite-temperature simulations used to carry out a systematic coarsening is the General Amber Force Field (GAFF),[83] with the AM1-BCC scheme in which the Mulliken charges obtained from semiempirical AM1 calculations are modified by a correction term that depends on the bond type.[84]

Energy Minimization and Finite-Temperature Equilibration. Geometry optimization and energy minimization calculations are performed for all the structures. The resulting lattice energy is different from that obtained from CSP_0. In the Ψ_{mol} ranking, form β is the global minimum, while using GAFF it is in the sixth position with a difference of 0.6 kJ mol^{-1} from the most stable structure. Nevertheless, β remains the most stable among the experimental forms. Focusing on the experimentally observed polymorphs, the GAFF force field is in good agreement with the Ψ_{mol} CSP_0 study results showing the β polymorph as the most stable, followed by γ at +4.34 kJ mol^{-1} and α at 6.91 kJ mol^{-1}. In order to eliminate labile structures and to begin the coarsening process, all the structures are equilibrated at 300 K and 1 bar. In Figure 6A, we see that only a few structures are labile compared to the case of urea illustrated earlier.

Looking at the experimental structures, the α form is distorted after equilibration. In Figure 5C, we see the α structure optimized with GAFF at 0 K and relaxed with MD at 300 K and 1 bar. The experimental structure of α polymorph, with its closest approximation identified by CSP_0, is reported for comparison in the Supporting Information, Tab. S1. The crystal structure obtained by finite-temperature relaxation of the α form presents the same hydrogen bonded chains of the experimental structure, but it is organized in layers of perfectly parallel molecules. This transformation is probably due to the inability of the GAFF force field to reproduce the α form at ambient temperature, and other force fields may be considered in the future. The γ form exhibits the same relaxation of the cell parameters seen in ref 69, with the β angle changing from nearly 90° to 99°.

Cluster Analysis. In order to reduce the relevant structures contained in the initial set of CSP_0 structures, we define a structural fingerprint for succinic acid molecules. As discussed in the Methods section, the structural fingerprint includes the distribution of distances between centers of mass, $p_i(r_{COM})$. The molecular orientation is defined through two vectors: one connecting the two central carbon atoms, and the other connecting the terminal ones, resulting in the intermolecular angles θ and ψ, respectively. θ captures how the central carbon atoms of a molecule are located with respect to neighboring molecules. The angle ψ instead captures the relative orientation of hydrogen bonded layers. Succinic acid H-bonded chains in different layers, shown in Figure 5B, are not necessarily parallel to each other, with the angle ψ able to vary as shown on the top-right of Figure 7B−C. Moreover, different packing motifs of H-bonded chains are observed, including alternate (ABABAB...) or doubly alternate (AABBAABB...) aligned layers. The varying alignment of H-bonded chains is illustrated in the Supporting Information, Figure S6. Since succinic acid is a flexible molecule, we include in the fingerprint a descriptor of the conformational state of individual molecules. To this aim, we select the distribution of the t_0 torsional angles (see Figure 5A) indicated

Table 2. Stable and Metastable Succinic Acid Crystal Structures with the Conformation of the Succinic Acid Molecule and Ordered According to Their Potential Energy[a]

Initial Structure	Potential Energy (kJ mol^{-1})	Conformation	Resulting Structure	W (kJ mol^{-1})
Form α	−93.87	**aAa**		> 3.5
S123	−93.73	**aAa**		> 3.5
S11752(Ψ_{crys})	−93.58	**aAa**		> 3.5
Form β	−93.35	**aAa**	S20629	2.61
S618(Ψ_{crys})	−92.58	**aAa**		> 3.5
S763	−91.91	**aAa**		> 3.5
Form γ	−91.79	**aGa**		> 3.5
S8149	−91.75	**aGa**		> 3.5
S804(Ψ_{crys})	−91.48	**aAa**	S763	2.90
S11752	−91.33	**aAa**		> 3.5
S1308	−90.95	**aAa**		> 3.5
S2526	−90.23	**aAa**		> 3.5
S16988(Ψ_{crys})	−90.13	**aAa**	Form α**	1.96
S20629	−90.12	**aAa**		> 3.5
S618	−89.98	**aAa**		> 3.5
S2910	−89.34	**aAa**	Form α	1.81
S2478(Ψ_{crys})	−89.20	**aAa**		> 3.5
S10824(Ψ_{crys})	−88.30	**aGa**		> 3.5
S4200	−88.19	**aAa**		> 3.5
S4846(Ψ_{crys})	−87.62	**aAa**	Form α*	2.08
S22937	−86.74	**aAa**	S763*	2.14
S14464(Ψ_{crys})	−86.65	**aGa**		> 3.5
S2585	−86.61	**aAa**		> 3.5
S4939(Ψ_{crys})	−86.53	**aAa**		> 3.5
S2473(Ψ_{crys})	−86.26	**aAa**		> 3.5
S2798(Ψ_{crys})	−85.85	**aGa**		> 3.5
S4947	−85.34	**aAa**		> 3.5
S1667	−85.33	**aAa**	S763**	2.73
S4947(Ψ_{crys})	−85.06	**aAa**	S4939(Ψ_{crys})	2.51
S3841	−81.23	**aAa**	NC3	3.09
S1107	−75.90	**aAa**	NC1	2.41
S18857(Ψ_{crys})	−75.75	**aAa**	NC2	2.78
NC1	−93.35	**gA-g**		−
NC2	−86.90	**gA-g**, **aAa**		−
NC3	−89.28	**gA-g**		−

[a]The structures that are excluded from the final ranking are shown in red if their work of transformation W was less than than $k_{B}T$, and in blue if they exhibited a partial melting or defects. Structures NC1, NC2, and NC3 emerged from transitions sampled during the metadynamics step and were not included in the initial CSP_0 set. These structures include conformers that were not considered in the initial search. We note that NC2 and NC3 also have $Z' = 2$ and $Z' = 4$, respectively.

with $p_i(\phi)$. The resulting fingerprint for succinic acid is thus expressed as

$$F_i = \{p_i(r_{COM}), p_i(\theta, \psi), p_i(\phi)\} \qquad (6)$$

The reduced energy landscape is reported in Figure 6B with a representation of the most populated structures. We observe that the α-like form is the most populated cluster, while the β and γ forms are also among the largest ones. Moreover, we notice that the geometry typical of the cluster centered on structure S763 is very similar to that of the β polymorph with a change in orientation occurring every two layers. Cluster S10824(Ψ_{crys}) is among the few with the succinic acid in the folded conformation. We observe that this cluster displays the same structure as form γ, but with stacking disorder along one axis. On the other hand, clusters S2562 and S2585 exhibit crystal structures significantly different from the experimental ones. They present the H-bonding motif shown on the top right of Figure 5B with chains in different layers oriented at an angle to one another, rather than in a parallel arrangement as in polymorph β. It should be noted that most of the structures in

the Ψ_{crys} data set coalesced to the same cluster as the Ψ_{mol} structure from which it was derived, as would be expected from a structure that had been reminimized on a different potential energy surface. However, there are exceptions (SI Tab. S4), which are mainly structures that are based on layers of hydrogen bonded molecules, suggesting that the potential energy surfaces differ significantly in the interlayer stacking energies. We obtain a single cluster for every experimentally known crystal structure α, β, and γ. In the unbiased relaxation of crystal supercells performed with NPT simulations, we do not observe any spontaneous conversion between planar and folded conformers.

Metadynamics. In this case, we investigated transitions by introducing bias up to 3.5 kJ mol^{-1} per molecule. As before, 10 independent replicas per cluster center were simulated, leading to the conversion network reported in Figure 7. In this representation, we divided the structures into four groups based on the molecules' conformation and the relative orientation of hydrogen bonded layers. Structures with folded molecules (Figure 7A) or misaligned layers of planar molecules (Figure 7C) emerge as extremely stable and do not transform

into other crystal structures under the effect of biased dynamics. The only exception to these groups is structure S4947(Ψ_{crys}) which undergoes a transition to the cluster centered around structure S4939(Ψ_{crys}). On the other hand, structures characterized by hydrogen bonded chains of planar molecules in aligned layers can easily interconvert (Figure 7B). Most of them transform to α or adopt a structure typical of the cluster centered around structure S763. In biased simulations, we also observe the transition to three putative conformational polymorphs that were not included in the original set of structures generated from the CSP_0 search and classification. Such transitions are represented in the conversion network reported in Figure 7D from the cluster centers S1107, S18857(Ψ_{crys}), and S3841 to structures NC1, NC2 and NC3. In particular, we note that structures with the H-bonding network motif $R_2^2(6)C_1^1(4)$ reported in Figure 5B convert to crystal structures in which the building block is the **gA-g** conformer.[35] We observe that the H-bonding network resulting from such transitions presents only the chain type $C_1^1(4)$ reported in the Supporting Information.

To conclude, from the initial CSP_0 set containing 101 structures, we identify 27 low energy, persistent crystal structures. This includes the conformational polymorphs resulting from the biased sampling of unit cells presenting the $R_2^2(6)C_1^1(4)$ H-bonding motif. These structures, with their cluster center, potential energy, and conformer are reported in Table 2.

■ DISCUSSION

The workflow presented here relies on, and confirms, the idea that the number of structures in a CSP_0 process can be significantly reduced by a Molecular Dynamics shakeup[15,16,22,85] and provides static perfect crystal structures (cf. crystallographic structures) to represent the most likely putative polymorphs generated by CSP_0. The reduction in the number of structures from CSP_0 represents a coarsening of the underlying potential energy surface, and the removal of many shallow minima with barriers that are lower than the thermal energy. The number of local lattice energy minima that form one finite-temperature ensemble strongly depends on the type of crystal packing and the intermolecular forces involved. For example, a strongly hydrogen bonded layer of molecules can have many shallow minima in the dispersion dominated interactions between the layers, whereas a structure with strong directional hydrogen bonds defining all three crystallographic directions will have only one minimum. The coarsening refines the CSP_0 output to allow focus on the dominant structural types and hence the structures most likely to be long-lived polymorphs with distinct properties. The approach is also crude or coarse in the sense that the practicalities of processing a large number of structures require an approximate force field and short simulation times in relatively small simulation boxes.

In an ideal world, if we could use an accurate force field, and perform MD simulations on each structure without time scale limitations, the MD stage of the coarsening process would identify the free energy minima at the simulation temperature and pressure. It would provide representative crystallographic structures for each free energy minimum and could also give their relative free energies. The metadynamics step then identifies which structures are sufficiently closely related for there to be a small barrier for a solid-state transformation and estimates the barrier height. This would identify relationships between the structures, and an easy transformation could

suggest that the structures would not remain distinct throughout the crystallization process.

A key observation is that the metadynamics step has found structures such as A* of urea, which do not have a small unit cell, but contain an interface boundary between different regions of the simulation cell. The observation of such structures is common when polymorphic transformations are investigated using enhanced sampling methods.[19,20,35,39] The interface, or stacking defect, involves sufficient molecules that the fingerprints of the simulation cells differ. Such structures are realistic and may correspond to the boundaries of polymorphic domains (cf. aspirin[86]), twin boundaries,[87] or stacking defects which are very common for organic crystal structures built up of layers of molecules with weak dispersion forces between the layers.[88] Thus, such simulations are entirely realistic in revealing the possibility of disorder in structures that may hinder the growth of single crystals.

Within the current approach (using a force field that is inaccurate with short, small MD simulations), the metadynamics step is very much part of the coarsening process, showing which structures would have probably melted if longer simulations or a slightly higher temperature had been used in the first stage. It may also reveal cases where the barrier and distinction between structures is so small to be an artifact of the force field used, such as holding urea rigid. The workflow should not be used in a completely blind, black box manner, because some crystal structures show significant dynamic disorder and are thermally stabilized, to the extent that a single lattice energy minimum CSP_0 structure does not give a good representation. Examples would include plastic crystals, structures with fairly freely rotating groups such as fluoromethyl groups, and probably the high temperature α form of succinic acid, as well as structures with stacking faults and grain boundaries. However, the clustering step is able to highlight that these structures need manual visualization of the trajectory, and careful consideration as to whether they represent the behavior of real crystals.

Urea and succinic acid are two relatively small molecules, characterized by limited conformational flexibility. Looking forward to applications of this protocol to industrially relevant cases where size, conformational complexity and Z' are likely to increase, we will need to address the scalability of our approach. Besides the increase in computational cost associated with larger system sizes both in MD and in CSP_0 studies, we foresee that the complexity in the molecular structure will impact the dimensionality, and thus the computational cost, of the fingerprint components used to generate a dissimilarity matrix necessary to perform clustering. In fact, while the dimensionality of the positional ($p(r_{com})$) and orientational ($p(\theta)$) components are fairly independent from size and complexity, the dimensionality of the conformational component ($p(\phi)$) is bound to increase with the size and conformational flexibility of the system. To tackle scaling issues of $p(\phi)$ for a system with an arbitrary number of torsion angles, we are currently exploring using a set of univariate probability distributions as an approximation to the fully correlated multidimensional probability density. We are currently testing this working hypothesis on larger molecules and larger CSP_0 data sets. We also note that, since the purpose of enhanced sampling in the workflow is not that of obtaining a thorough exploration of the configurational ensemble of crystal structures, but the facilitation of the conversion of crystal structures that are already similar, the CVs used (energy and density) are fairly independent of conformational complexity and size. Therefore,

L

we do not anticipate the cost of this step to be severely affected by an increase in size, flexibility, or Z' of the systems studied.

Finally, we note that our workflow currently necessitates the use of standard molecular dynamics force fields, which are not sufficiently accurate for crystal structure prediction studies. The differences between the GAFF force field and that used in the CSP_0 studies, or derived by reoptimizing the succinic acid structures using periodic electronic structure methods (see the SI section S1) are sufficient to raise questions as to whether the stability of some structures is an artifact of the force field. As demonstrated by periodic high Z' structures (NC2 and NC3) being generated in the biased simulations of succinic acid, answering such questions requires further consistency checks through calculations with the more realistic potential energy surfaces. Indeed, revaluating the energies of the new (NC) structures generated by the metadynamics step with the Ψ_{mol} and Ψ_{crys} methods showed that they were too high in energy to be considered as putative polymorphs. In the future, when we can develop a sufficiently accurate potential energy surface for a specific pharmaceutical,[89] that can be used for both the initial CSP_0 structure generation and this workflow, then the resulting, much smaller set of CSP_thd structures could be used with more confidence to guide experimental polymorph screening efforts.[90]

CONCLUSIONS

In this study, we have demonstrated the application of a systematic coarsening approach of CSP_0-generated crystal energy landscapes, based on the application of both unbiased and biased MD simulation methods, coupled with the clustering of finite-temperature structures based on probabilistic structural fingerprints.

The method was tested on two small organic molecules exhibiting polymorphism, namely, urea and succinic acid. In both cases, our analysis allowed us to reduce the number of candidates to be further evaluated as potential polymorphs by one order of magnitude. In the case of urea, we reduced the interesting candidates from 124 CSP_0-generated crystal structures to 12 putative polymorphs. In the case of succinic acid, we went from 101 CSP_0 minima to two dozen putative polymorphs.

It is worth noting that for both urea and succinic acid all previously known structures both experimentally observed or computationally predicted in other studies were obtained in the final reduced set.

Finally, we note that finite-temperature sampling of the density/energy collective variable space starting from crystal structures generated from CSP_0 yielded several examples of crystal supercells that have a structure consistent with the most populated clusters, but incorporating defects. The fingerprint representation and clustering method used in this work is able to identify and distinguish such structures from their perfectly ordered counterparts. This is because, while probabilistic structural fingerprints are robust with respect to isolated point defects, they are sensitive to packing defects that introduce a statistically significant number of molecules arranged differently along the defect. We believe this is an important consequence of this analysis, as it allows the identification of structures that are potentially prone to exhibit packing defects in reality.

In order to perform a systematic coarsening of crystal energy landscapes based on finite-temperature MD simulations, we have used force fields that differ from those employed in the CSP_0 ranking. While among the surviving crystal structures,

we can find all the known experimental and previously proposed structures, the (in)consistency of the energy approximation is an issue throughout the entire CSP workflow. For instance, similar deviations to those found with GAFF are observed comparing Ψ_{crys} and Ψ_{mol} lattice energies across the entire pool of succinic acid structures. A comparison of lattice energies computed with different methods is reported in the Supporting Information.

We are convinced that the ideas described in this work on a systematic screening of structures generated by CSP_0 aimed at identifying long-lived, stable polymorphs are general and will provide a useful, physics-based strategy for the systematic reduction of crystal energy landscapes. This enables a systematic identification of promising candidates to be further analyzed in computationally expensive accurate thermodynamic calculations, and in explicit nucleation and growth simulations for the assessment of crystallization kinetics.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.cgd.0c00918.

> Additional details on the CSP_0 data sets, clustering analysis, force field comparison, and input files for trajectory analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Matteo Salvalaglio — *Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, U.K.;* ⓞ orcid.org/0000-0003-3371-2090; Email: m.salvalaglio@ucl.ac.uk

Authors

Nicholas F. Francia — *Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, U.K.;* ⓞ orcid.org/0000-0003-0936-2342

Louise S. Price — *Department of Chemistry, University College London, London WC1H 0AJ, U.K.;* ⓞ orcid.org/0000-0002-7633-1987

Jonas Nyman — *Small Molecule Design & Development, Eli Lilly and Company, Indianapolis, Indiana 46285, United States; Holly Chemistry, S-69135 Karlskoga, Sweden;* ⓞ orcid.org/0000-0001-9011-3962

Sarah L. Price — *Department of Chemistry, University College London, London WC1H 0AJ, U.K.;* ⓞ orcid.org/0000-0002-1230-7427

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.cgd.0c00918

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Price, S. L. Why Don't We Find More Polymorphs? *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2013**, *69*, 313−328.

(2) 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.

(3) Price, S. L.; Reutzel-Edens, S. M. The Potential of Computed Crystal Energy Landscapes to Aid Solid-Form Development. *Drug Discovery Today* **2016**, *21*, 912−923.

(4) Mortazavi, M.; Hoja, J.; Aerts, L.; Quéré, L.; van de Streek, J.; Neumann, M. A.; Tkatchenko, A. Computational Polymorph Screening Reveals Late-Appearing and Poorly-Soluble Form of Rotigotine. *Comm. Chem.* **2019**, *2*, 1−7.

(5) Nyman, J.; Pundyke, O. S.; Day, G. M. Accurate Force Fields and Methods for Modelling Organic Molecular Crystals at Finite Temperatures. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15828−15837.

(6) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *J. Chem. Theory Comput.* **2011**, *7*, 1998−2016.

(7) Price, S. L.; Leslie, M.; Welch, G. W.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478−8490.

(8) Neumann, M. A.; Leusen, F. J.; Kendrick, J. A Major Advance in Crystal Structure Prediction. *Angew. Chem., Int. Ed.* **2008**, *47*, 2427−2430.

(9) Price, S. L. Is Zeroth Order Crystal Structure Prediction (CSP_0) Coming to Maturity? What Should We Aim for in an Ideal Crystal Structure Prediction Code? *Faraday Discuss.* **2018**, *211*, 9−30.

(10) Abramov, Y. A. Current Computational Approaches to Support Pharmaceutical Solid Form Selection. *Org. Process Res. Dev.* **2013**, *17*, 472−485.

(11) Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M. Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids? *Chem. Commun.* **2016**, *52*, 7065−7077.

(12) Nyman, J.; Reutzel-Edens, S. M. Crystal Structure Prediction is Changing From Basic Science to Applied Technology. *Faraday Discuss.* **2018**, *211*, 459−476.

(13) Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences Between Polymorphs. *CrystEngComm* **2015**, *17*, 5154−5165.

(14) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and Fictions About Polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619−8635.

(15) Mooij, W. T.; van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. Crystal Structure Predictions for Acetic Acid. *J. Comput. Chem.* **1998**, *19*, 459−474.

(16) Gavezzotti, A. A Molecular Dynamics Test of the Different Stability of Crystal Polymorphs Under Thermal Strain. *J. Am. Chem. Soc.* **2000**, *122*, 10724−10725.

(17) Shtukenberg, A. G.; Zhu, Q.; Carter, D. J.; Vogt, L.; Hoja, J.; Schneider, E.; Song, H.; Pokroy, B.; Polishchuk, I.; Tkatchenko, A.; Oganov, A. R.; Rohl, A. L.; Tuckerman, M. E.; Kahr, B. Powder diffraction and crystal structure prediction identify four new coumarin polymorphs. *Chem. Sci.* **2017**, *8*, 4926−4940.

(18) 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.

(19) 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.

(20) 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**, *72*, 542−550.

(21) Dunitz, J. D.; Filippini, G.; Gavezzotti, A. Molecular Shape and Crystal Packing: a Study of C12H12 Isomers, Real and Imaginary. *Helv. Chim. Acta* **2000**, *83*, 2317−2335.

(22) Dybeck, E. C.; McMahon, D. P.; Day, G. M.; Shirts, M. R. Exploring the Multi-Minima Behavior of Small Molecule Crystal Polymorphs at Finite Temperature. *Cryst. Growth Des.* **2019**, *19*, 5568−5580.

(23) Červinka, C.; Beran, G. J. Ab Initio Prediction of the Polymorph Phase Diagram for Crystalline Methanol. *Chem. Sci.* **2018**, *9*, 4622−4629.

(24) Case, D. et al. *AMBER 2018*; University of California: San Francisco, 2018.

(25) Rackers, J. A.; Laury, M. L.; Lu, C.; Wang, Z.; Lagardère, L.; Piquemal, J. P.; Ren, P.; Ponder, J. W.; Schnieders, M. TINKER 8: A Modular Software Package for Molecular Design and Simulation. *J. Chem. Theory Comput.* **2018**, *14*, 5273.

(26) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(27) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.

(28) Abraham, M.; van der Spoel, D.; Lindahl, E.; Hess, B.; the GROMACS development team, *GROMACS User Manual version 2018*, http://www.gromacs.org; 2018.

(29) 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.

(30) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19 http://lammps.sandia.gov..

(31) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 147−161.

(32) Neumann, M. A. Tailor-Made Force Fields for Crystal-Structure Prediction. *J. Phys. Chem. B* **2008**, *112*, 9810−9829.

(33) This feature is available in the GRACE software, but no article was found about the use of such tailor-made force fields in MD packages.

(34) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science* **2014**, *344*, 1492−1496.

(35) Gimondi, I.; Salvalaglio, M. CO2 Packing Polymorphism Under Confinement in Cylindrical Nanopores. *Mol. Syst. Des. Eng.* **2018**, *3*, 243−252.

(36) 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.

(37) Piaggi, P. M.; Parrinello, M. Predicting Polymorphism in Molecular Crystals Using Orientational. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10251−10256.

(38) Santiso, E. E.; Trout, B. L. A General Set of Order Parameters for Molecular Crystals. *J. Chem. Phys.* **2011**, *134*, 064109.

(39) Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Insight Into the Nucleation of Urea Crystals from the Melt. *Chem. Eng. Sci.* **2015**, *121*, 51−59.

(40) Tribello, G. A.; Giberti, F.; Sosso, G. C.; Salvalaglio, M.; Parrinello, M. Analyzing and Driving Cluster Formation in Atomistic Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 1317−1327.

(41) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics studies of crystal nucleation. *IUCrJ* **2015**, *2*, 256−266.

(42) Gimondi, I.; Salvalaglio, M. CO2 packing polymorphism under pressure: Mechanism and thermodynamics of the I-III polymorphic transition. *J. Chem. Phys.* **2017**, *147*, 114502.

(43) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(44) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670−673.

(45) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602.

(46) Shang, C.; Zhang, X.-J.; Liu, Z.-P. Crystal Phase Transition of Urea: What Governs the Reaction Kinetics in Molecular Crystal Phase Transitions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32125−32131.

(47) Olejniczak, A.; Ostrowska, K.; Katrusiak, A. H-bond Breaking in High-Pressure Urea. *J. Phys. Chem. C* **2009**, *113*, 15761−15767.

(48) Swaminathan, S.; Craven, B. M.; McMullan, R. K. The Crystal Structure and Molecular Thermal Motion of Urea at 12, 60 and 123 K from Neutron Diffraction. *Acta Crystallogr., Sect. B: Struct. Sci.* **1984**, *40*, 300−306.

(49) 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.

(50) Marshall, W. G.; Francis, D. J. Attainment of Near-Hydrostatic Compression Conditions Using the Paris−Edinburgh Cell. *J. Appl. Crystallogr.* **2002**, *35*, 122−125.

(51) Roszak, K.; Katrusiak, A. Giant Anomalous Strain Between High-Pressure Phases and the Mesomers of Urea. *J. Phys. Chem. C* **2017**, *121*, 778−784.

(52) 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.

(53) Weber, H. P.; Marshall, W. G.; Dmitriev, V. High-Pressure Polymorphism in Deuterated Urea. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2002**, *58*, c174.

(54) Piana, S.; Gale, J. D. Understanding the Barriers to Crystal Growth: Dynamical Simulation of the Dissolution and Growth of Urea from Aqueous Solution. *J. Am. Chem. Soc.* **2005**, *127*, 1975−1982.

(55) Piana, S.; Reyhani, M.; Gale, J. D. Simulating Micrometre-Scale Crystal Growth from Solution. *Nature* **2005**, *438*, 70−73.

(56) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *J. Am. Chem. Soc.* **2012**, *134*, 17221−17233.

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

(58) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-Dynamics Simulations of Urea Nucleation from Aqueous Solution. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E6−E14.

(59) Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Urea Homogeneous Nucleation Mechanism is Solvent Dependent. *Faraday Discuss.* **2015**, *179*, 291−307.

(60) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular Dynamics Simulations of Solutions at Constant Chemical Potential. *J. Chem. Phys.* **2015**, *142*, 144113.

(61) Bürgi, H. B.; Capelli, S. C. Dynamics of Molecules in Crystals from Multi-Temperature Anisotropic Displacement parameters. I. Theory. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2000**, *56*, 403−412.

(62) Day, G. M.; Price, S. L.; Leslie, M. Elastic Constant Calculations for Molecular Organic Crystals. *Cryst. Growth Des.* **2001**, *1*, 13−27.

(63) Civalleri, B.; Doll, K.; Zicovich-Wilson, C. Ab Initio Investigation of Structure and Cohesive Energy of Crystalline Urea. *J. Phys. Chem. B* **2007**, *111*, 26−33.

(64) Frisch, M. J. et al. *GAUSSIAN 09*. Inc.: Wallingford, CT, 2009; *121*, 150−166.

(65) Sugden, I.; Adjiman, C. S.; Pantelides, C. C. Accurate and Efficient Representation of Intramolecular Energy in Ab Initio Generation of Crystal Structures. I. Adaptive Local Approximate Models. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 864−874.

(66) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996**, *100*, 7352−7360.

(67) Stone, A. J. *GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System*; University of Cambridge, 2010.

(68) Chisholm, J. A.; Motherwell, S. COMPACK: A Program for Identifying Crystal Structure Similarity Using Distances. *J. Appl. Crystallogr.* **2005**, *38*, 228−231.

(69) Lucaioli, P.; Nauha, E.; Gimondi, I.; Price, L. S.; Guo, R.; Iuzzolino, L.; Singh, I.; Salvalaglio, M.; Price, S. L.; Blagden, N. Serendipitous Isolation of a Disappearing Conformational Polymorph of Succinic Acid Challenges Computational Polymorph Prediction. *CrystEngComm* **2018**, *20*, 3971−3977.

(70) Bučar, D.-K.; Lancaster, R. W.; Bernstein, J. Disappearing Polymorphs Revisited. *Angew. Chem., Int. Ed.* **2015**, *54*, 6972−6993.

(71) Yu, Q.; Dang, L.; Black, S.; Wei, H. Crystallization of the Polymorphs of Succinic Acid Via Sublimation at Different Temperatures in the Presence or Absence of Water and Isopropanol Vapor. *J. Cryst. Growth* **2012**, *340*, 209−215.

(72) Barr, G.; Dong, W.; Gilmore, C. J.; Parkin, A.; Wilson, C. C. dSNAP: A Computer Program to Cluster and Classify Cambridge Structural Database Searches. *J. Appl. Crystallogr.* **2005**, *38*, 833−841.

(73) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171−179.

(74) Chikhalia, V.; Forbes, R.; Storey, R.; Ticehurst, M. The Effect of Crystal Morphology and Mill Type on Milling Induced Crystal Disorder. *Eur. J. Pharm. Sci.* **2006**, *27*, 19−26.

(75) Robertson, K.; Flandrin, P.-B.; Klapwijk, A. R.; Wilson, C. C. Design and Evaluation of a Mesoscale Segmented Flow Reactor (KRAIC). *Cryst. Growth Des.* **2016**, *16*, 4759−4764.

(76) Carver, K. M.; Snyder, R. C. Unexpected Polymorphism and Unique Particle Morphologies from Monodisperse Droplet Evaporation. *Ind. Eng. Chem. Res.* **2012**, *51*, 15720−15728.

(77) 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.

(78) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(79) Kresse, G.; Furthmüller, J. VASP the Guide (Universität Wien, Wien, Austria, 2007); G. Kresse and J. Hafne. *Phys. Rev. B* **1993**, *47*, 558.

(80) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(81) Spek, A. Single-Crystal Structure Validation with the Program PLATON. *J. Appl. Crystallogr.* **2003**, *36*, 7−13.

(82) Etter, M. C.; MacDonald, J. C.; Bernstein, J. Graph-Set Analysis of Hydrogen-Bond Patterns in Organic Crystals. *Acta Crystallogr., Sect. B: Struct. Sci.* **1990**, *46*, 256−262.

(83) 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.

(84) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(85) Gray, A.; Day, G.; Leslie, M.; Price, S. Dynamics in Crystals of Rigid Organic Molecules: Contrasting the Phonon Frequencies Calculated by Molecular Dynamics with Harmonic Lattice Dynamics for Imidazole and 5-Azauracil. *Mol. Phys.* **2004**, *102*, 1067−1083.

(86) Bond, A. D.; Boese, R.; Desiraju, G. R. On the Polymorphism of Aspirin: Crystalline Aspirin as Intergrowths of Two "Polymorphic" Domains. *Angew. Chem., Int. Ed.* **2007**, *46*, 618−622.

(87) Coquerel, G. The 'Structural Purity' of Molecular Solids—An Elusive Concept? *Chem. Eng. Process.* **2006**, *45*, 857−862.

O

(88) Price, L. S.; McMahon, J. A.; Lingireddy, S. R.; Lau, S.-F.; Diseroad, B. A.; Price, S. L.; Reutzel-Edens, S. M. A Molecular Picture of the Problems in Ensuring Structural Purity of Tazofelone. *J. Mol. Struct.* **2014**, *1078*, 26−42.

(89) Uzoh, O. G.; Galek, P. T.; Price, S. L. Analysis of the Conformational Profiles of Fenamates Shows Route Towards Novel, Higher Accuracy, Force-Fields for Pharmaceuticals. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7936−7948.

(90) Bhardwaj, R. M.; McMahon, J. A.; Nyman, J.; Price, L. S.; Konar, S.; Oswald, I. D. H.; Pulham, C. R.; Price, S. L.; Reutzel-Edens, S. M. A Prolific Solvate Former, Galunisertib, Under the Pressure of Crystal Structure Prediction, Produces ten Diverse Polymorphs. *J. Am. Chem. Soc.* **2019**, *141*, 13887−13897.