Thermodynamic Stability is a Poor Indicator of Co-Crystallization in Models of Organic Molecules

Yulia Pimonova, John Carpenter, and Michael Gruenwald*

Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

E-mail: michael.gruenwald@utah.edu

Abstract

Co-crystallizing a given molecule with another can be useful for adjusting the physical properties of molecules in the solid state. However, most combinations of molecules do not readily co-crystallize, but either form one-component crystals or amorphous solids. Computational methods of crystal structure prediction can, in principle, identify the thermodynamically stable cocrystal and thus predict if molecules will co-crystallize or not. However, the pronounced polymorphism and tendency of many organic molecules to form disordered solids suggest that kinetic factors can play an important role in co-crystallization. The question remains: If a binary system of molecules has a thermodynamically stable cocrystal, will it indeed co-crystallize? To address this question, we simulate the crystallization of more than 2600 distinct pairs of chiral model molecules of similar size in 2D and calculate accurate crystal energy landscapes for all of them. Our analysis shows that thermodynamic criteria alone are unreliable in the prediction of co-crystallization. While the vast majority of cocrystals that form in our simulations are thermodynamically favorable, most coformer systems that have a thermodynamically stable cocrystal do not co-crystallize. We furthermore

show that co-crystallization rates increase threefold when coformers are used that do not form well-ordered single-component crystals. Our results suggest that kinetic factors of co-crystallization are decisive in many cases.

Introduction

One of the major challenges in drug development is the poor solubility of most new drug molecules.[1] Solubility, dissolution rate, and bioavailability of a solid compound can be modified via co-crystallization, the formation of crystals composed of two or more different compounds in a stoichiometric ratio.[2] In addition to adjusting pharmaceutical properties, co-crystallization can assist in chiral resolution,[3,4] increase the power of energetic materials,[5] aid in water decontamination,[6] and yield new ferroelectric organic materials.[7]

The two central questions of cocrystal formation are: Will a given combination of molecules co-crystallize? And what cocrystal will it form?[8–13] Sophisticated methods of computational crystal structure prediction (CSP) are increasingly applied to co-crystallization.[3,14–16] Based on thermodynamic principles, CSP aims to identify crystal structures with low free energy. (In practice, zero-temperature lattice energy is often used as a proxy for free energy. While entropic contributions to the free energy can be significant,[17,18] they are often ignored due to the computational cost of evaluating them.) In the case of a mixture of two kinds of molecules, cocrystals compete thermodynamically with one-component crystals. A cocrystal is often considered thermodynamically stable if its energy per molecule is lower than that of a physical mixture of one-component crystals of the two constituent molecules,

$$\Delta E = E_{\mathrm{co}} - \frac{1}{c_1 + c_2}(c_1 E_1 + c_2 E_2) < 0. \tag{1}$$

Here, E_{co} is the cocrystal energy per molecule, E_1 and E_2 are the energies per molecule of crystals of molecular components 1 and 2, respectively, and c_1 and c_2 are the stoichiometric coefficients of components 1 and 2 in the cocrystal, respectively. (Note that this thermo-

dynamic condition neglects solution effects[19,20]). If CSP energy landscapes are available for one-component structures of both molecules and for cocrystals, ΔE can be evaluated for the lowest energy structures from those landscapes: A cocrystal should form if $\Delta E < 0$ for at least one cocrystal; if several predicted cocrystals have $\Delta E < 0$, the one with the smallest value of ΔE should form.

The predictive power of CSP for co-crystallization has not been extensively tested due to the computational cost of producing accurate crystal energy landscapes, which is particularly high for cocrystals.[14,15] In 2009, Price and coworkers calculated the lattice energies of 12 cocrystals and their corresponding single-component crystals. They found the cocrystals to be thermodynamically favorable, although values of ΔE were comparable to energy differences between different predicted polymorphs and estimated computational errors.[21] In 2018, Taylor and Day analyzed a much larger set of 350 cocrystals taken from the Cambridge Structural Database and found that 95% were thermodynamically favored and many had a substantial energetic advantage.[19] While these studies strongly suggest that cocrystals usually only form when they are thermodynamically stable, they do not shed light on the thermodynamics of numerous systems that fail to form cocrystals.[22]

To reveal the role of thermodynamics in co-crystallization, one requires accurate energy landscapes of potential crystal structures of many pairs of molecules, as well as reliable data about their crystallization outcomes, including negative outcomes (i.e., formation of one-component crystals, poorly crystalline, or amorphous solids). Such comprehensive data are currently not available. We, therefore, base the analysis in this work on a large family of molecular models whose crystallization dynamics can be easily simulated and whose crystal energy landscapes can be efficiently generated.[23]

We briefly summarize our main results, which we obtained by analyzing over 2600 different model coformer systems:

1. The vast majority (99%) of systems that co-crystallized in MD simulations have a thermodynamically stable cocrystal ($\Delta E \leq 0$). However, most systems (87%) with

thermodynamically stable cocrystals do *not* actually co-crystallize, forming amorphous solids or single-component crystals instead. Even if a cocrystal forms, it is not the thermodynamically stable one in a significant fraction of cases.

2. Mixtures of coformers that are bad crystallizers (i.e., molecules that only form disordered solids from single-component solutions) are three times as likely to form a cocrystal than mixtures of molecules that are good crystallizers.

3. The likelihood of forming a cocrystal depends only weakly on the energy difference of single-component crystals of the two coformers.

4. If the coformer composition of the cocrystal with the smallest ΔE in the crystal landscape is not 1:1 or 1:2, the system is very likely to form a disordered solid instead of a cocrystal.

We present and discuss these results in detail in the following sections.

Results and Discussion

Model Molecules

We use a schematic two-dimensional model of chiral molecules to study co-crystallization.[23] We previously showed that the model reproduces the rich crystallization behavior of real racemic mixtures.[24] Molecules are rigid bodies comprising five circular beads of diameter σ, which represent different functional groups and have one of the 11 chiral shapes shown in Fig. 1a. [1] For each shape, we consider both enantiomers, indicated as L and R, respectively.

All functional groups interact through short-ranged attractive pair potentials (see Fig. 1b and Methods), but interactions may have different strengths. Specifically, each molecular

[1]Note that we only study co-crystallization of molecules of similar size (i.e., five functional groups). While cocrystals of molecules with comparable sizes are most abundant in the Cambridge Structural Database, many cocrystals of molecules with significantly different sizes have been reported (see Fig. S12). Such systems are outside the scope of this study.

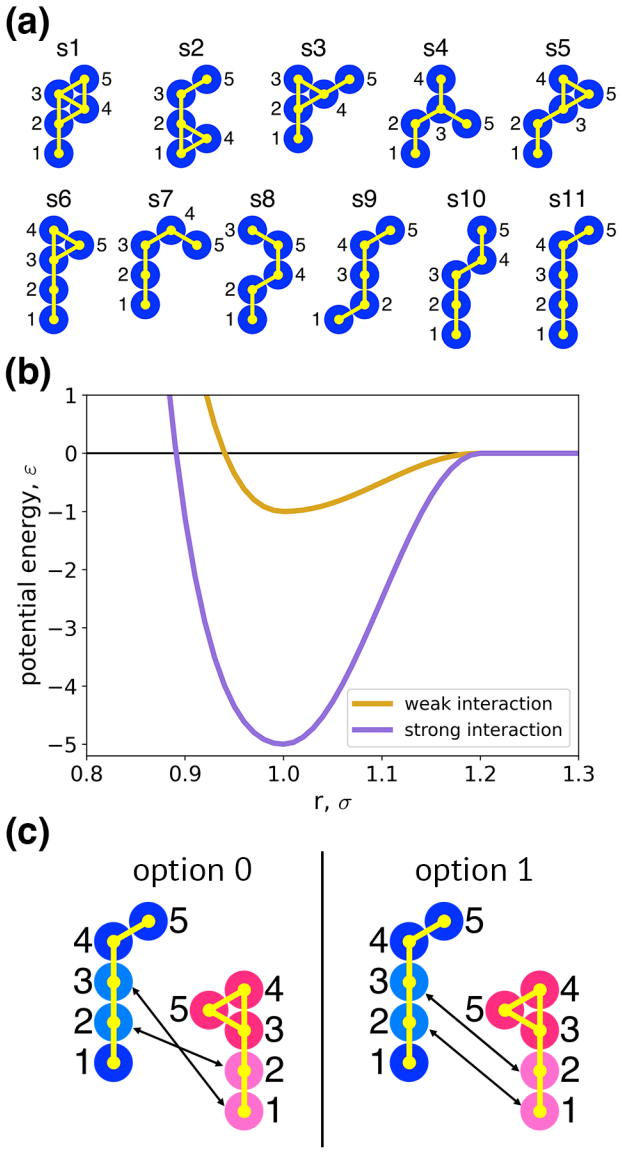

Figure 1: Molecular Model. (a) Illustration of the 11 chiral shapes used in this work. Blue beads represent functional groups. Only "left-handed" (L) enantiomers are shown for each shape, but "right-handed" (R) enantiomers are also studied. (b) Plots of pair potentials for weak and strong interactions between functional groups. (c) The two choices for strong cross interactions (indicated by arrows) between different molecules, illustrated for molecules s11L 2:3 and s6R 1:2. Blue and red colors are used for functional groups of L and R molecules, respectively. Lighter colors indicate the two SIFGs on each molecule. Note that cross interaction options 0 and 1 are *not* different interaction motifs of a particular pair of coformers but define *distinct* coformer systems.

species has a single pair of functional groups with attractive intermolecular interactions that are five times stronger than those between all other functional groups. (These strong interactions could, for instance, represent hydrogen or halogen bonds.[25,26]) A given molecular species is therefore determined by its shape (s1, s2, ..., s11), its handedness (L or R), and the indices of strongly interacting functional groups (e.g., 2:3, 1:2, 4:4, etc.). Note that our model molecules are designed to represent a variety of molecular shapes and interactions, but are not meant as accurate coarse-grained representations of specific organic molecules.

The strongly interacting functional groups (called SIFG for short) of a given molecular species also serve as centers for strong interactions between molecules of different species. Given a particular pair of coformers (*e.g.*, s11L 2:3 and s6R 1:2), we consider two different cross-interaction schemes that define two distinct coformer systems, as illustrated in Fig.1c: In option 1, strong interactions are assigned to the pair of SIFG with the lower of the two indices on both molecules (e.g., between functional group 2 of s11L 2:3 and functional group 1 of s6R 1:2) and to the pair of SIFG with the larger of the two indices on both molecules (e.g., between functional group 3 of s11L 2:3 and functional group 2 of s6R 1:2). Option 0 swaps the assignment of strong interactions: The SIFG with the lower index on one of the two molecules interacts strongly with the SIFG with the higher index on the other molecule, and vice versa.

Note that we do not attempt to simulate co-crystallization of all possible combinations of molecules described above, which would require $\approx 10^5$ independent MD simulations. Instead, we have selected a subset of more than 2600 coformer systems based on their crystallization behavior, as described in the next section.

Molecular Dynamics

To simulate co-crystallization, we use systems of 5184 molecules of two distinct species with a mole fraction of 0.5. Simulations are started from dispersed and mixed configurations in square simulation boxes at a packing fraction of 0.04 molecules/σ^2. Molecules evolve

according to Langevin dynamics at constant volume and temperature, as implemented in HOOMD.[27] Solvent is not modeled explicitly.

We use an automatic temperature protocol to optimize crystallization conditions. The details of this procedure are described in ref.[23] Briefly, simulations proceed in three stages. First, starting from a temperature well above the crystallization temperature, the system is cooled at a constant rate until a cluster of at least 50 molecules is observed for the first time. In the second stage, the temperature is tuned on the fly to grow clusters at a constant rate until a cluster of at least 200 molecules is observed. In the third stage, the system is slowly heated at a constant rate to anneal defects and to obtain highly crystalline clusters. During the last stage of the simulations, the structure of clusters is periodically analyzed, and a crystal quality score (CQ) is assigned based on a comparison of local structural motifs in the cluster with an extensive set of low-energy cocrystal structures (as described in the Methods section). CQ scores range from 0 to 1, where 0 indicates a complete lack of order, and 1 indicates a perfectly ordered bulk crystal.[23] Since crystalline clusters in our simulations typically comprise a few hundred molecules, the largest CQ scores obtained are ≈ 0.85. Clusters with $CQ > 0.4$ display clear order over several unit cells; simulations producing one or more clusters with $CQ > 0.4$ are therefore considered successful crystallization attempts, yielding either a cocrystal, single-component crystals, or rarely both. Simulations that do not rise to this threshold are considered unsuccessful and typically produce clusters with one of several distinct types of disorder, as discussed below. Independent simulation runs of the same system consistently yield similar results (See SI Fig. S5). In previous work,[23] we have also confirmed that simulations employing the automated temperature protocol and simulations of the same model that use a manually optimized *constant* temperature yield clusters of the same crystal structure in almost all cases. The maximum CQ score is typically lower in constant-temperature simulations, which we attribute to the fact that the temperatures that allow for spontaneous nucleation of crystalline to be observed in the simulation time scale can differ substantially from the (higher) temperatures that result in

defect-free growth of established clusters.

We simulated the co-crystallization of over 2600 systems containing two coformers. We selected coformers based on their crystallization behavior, which we established in previous work.[23] Specifically, we used 30 molecules (and their enantiomers) that each readily form large well-ordered crystalline clusters ($CQ > 0.4$) in single-component simulations ("good crystallizers" or GC for short), and 33 molecules (and their enantiomers) that only form poorly ordered aggregates ($CQ < 0.4$) in single-component simulations ("bad crystallizers" or BC for short). All selected coformers are shown in Fig. S1 and S2 in the Supporting Information. We simulated the co-crystallization of all pairs of GC and all pairs of BC, omitting combinations of coformers with the same molecular shape. These choices resulted in a total of 1095 attempted co-crystallization simulations from the GC set, and 1535 simulations in the BC set. Simulations in the GC set are closely related to experimental attempts to create new solid forms from coformers that both have at least one known crystal structure. Simulations in the BC set model the use of co-crystallization to produce ordered solids of coformers that by themselves resist regular crystallization attempts. (As discussed in the SI, approximately one-third of cocrystals reported in the CSD have at least one coformer with no reported single-component crystal structure.)

Crystal Energy Landscapes

We calculated accurate cocrystal energy landscapes for all simulated systems. (Single-component crystal energy landscapes for all coformers were available from recent work.[23,24]) All crystal structures were obtained with POLYNUM, a previously developed program that exploits the geometry of our model molecules to solve an exact cover problem on a 2D hexagonal lattice.[23,24] (See Methods.) By applying the algorithm for a range of different unit cell sizes, different unit cell geometries, and different numbers of molecules per unit cell, all periodic packings can, in principle, be identified (up to a size limit imposed by computational efficacy). POLYNUM makes no reference to crystal symmetry and naturally identifies pack-

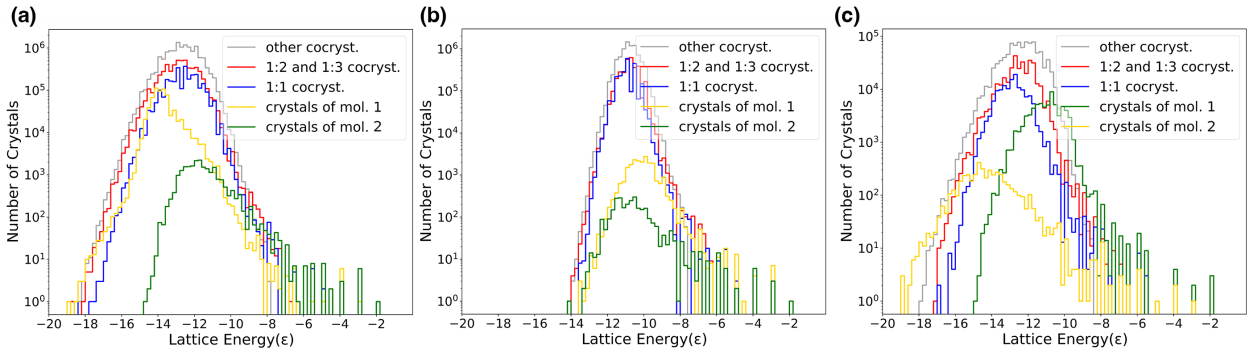

Figure 2: Histograms of lattice energies of crystals identified with POLYNUM for three different coformer systems: (a) s3R 1:4 & s11L 1:2 (BC set); (b) s3L 3:3 & s8R 2:2 (GC set); (c) s4L 2:3 & s7L 1:3 (GC set). Separate histograms are shown for the single-component crystals of the two coformers (yellow and green) and for cocrystals with different compositions (red, blue, and gray). Coformer system (a) forms a cocrystal in MD simulations, system (b) forms an amorphous solid, and system (c) forms a single component crystal.

ings of all symmetries, different numbers of molecules in the asymmetric unit cell (Z'), and different compositions.

Periodic packings of molecules identified with POLYNUM are ranked according to lattice energy in a second step. Lattice energies correspond to energy-minimized lattice geometries and do not account for thermal effects. However, we have previously demonstrated that, because of the rigidity of the molecules and their short-ranged interactions, lattice energy differences reported here are good approximations for differences in *free* energies at the temperature of crystallization.[23] The crystal energy landscapes reported here for our model molecules, therefore, provide more comprehensive (in terms of crystal symmetry and composition) and more accurate thermodynamic information than what is typically available for real molecules through state-of-the-art CSP methods.

Typical examples of crystal energy landscapes obtained with POLYNUM are shown in Fig. 2. POLYNUM generally identifies up to millions of periodic packings for a given molecule. Most of these packings have energies that are too high to be thermodynamically relevant, and many have low densities. (In CSP of real molecules, such packings are either not created in the first place or discarded.) We typically find tens to thousands of packings within what is usually considered the thermodynamically relevant range of lattice

energies that are smaller than $0.9E_0$, where E_0 is the lowest crystal energy found for a given pair of coformers. These numbers are consistent with (co-)crystal energy landscapes of real molecules. Values of E_0 vary substantially depending on the coformer system between -21.0 ϵ/molecule and -11.0 ϵ/molecule. As evident from energy landscapes in Fig. 2, the total number of cocrystal structures identified by POLYNUM by far exceeds the number of single-component structures, typically by two or three orders of magnitude. While cocrystals with a 1:1 stoichiometric ratio are most numerous among reported structures in the CSD (and also among cocrystals found in our simulations, as discussed below), we find equal or greater numbers of cocrystal structures with composition 1:2 and 1:3, and even greater numbers of cocrystal structures with different compositions, which are found exceptionally rarely in experiments and in our simulations. We return to the thermodynamic role of cocrystals with different compositions in a later section.

Overview of Co-Crystallization Results

Among the 1095 distinct pairs of molecules in the GC set for which we simulated co-crystallization, we find that 4.5% form a cocrystal, 76% form single-component crystals (identical to those found in single-component simulations of the same molecules), and 19.5% do not form any well-ordered structures. Among the 1535 systems in the BC set, we find that 15.1% form a cocrystal, and 84.9% form disordered structures. Crystallization outcomes for all simulated systems are summarized in Supporting Information Fig. S3; examples of cocrystals, single-component crystals, and disordered structures are shown in Fig. 3.

There is only limited information about experimental rates of co-crystallization available in the literature. Our observed rates (4.5% and 15.1% in the GC and BC sets, respectively) are lower than the 38% success rate reported in a recent study by Wicker and coworkers.[22] However, these authors selected coformers for their potential to form favorable intermolecular interactions, while molecules were selected without bias in our study. In their experiments, Wicker and coworkers also used neat grinding, a method that has been shown to result in

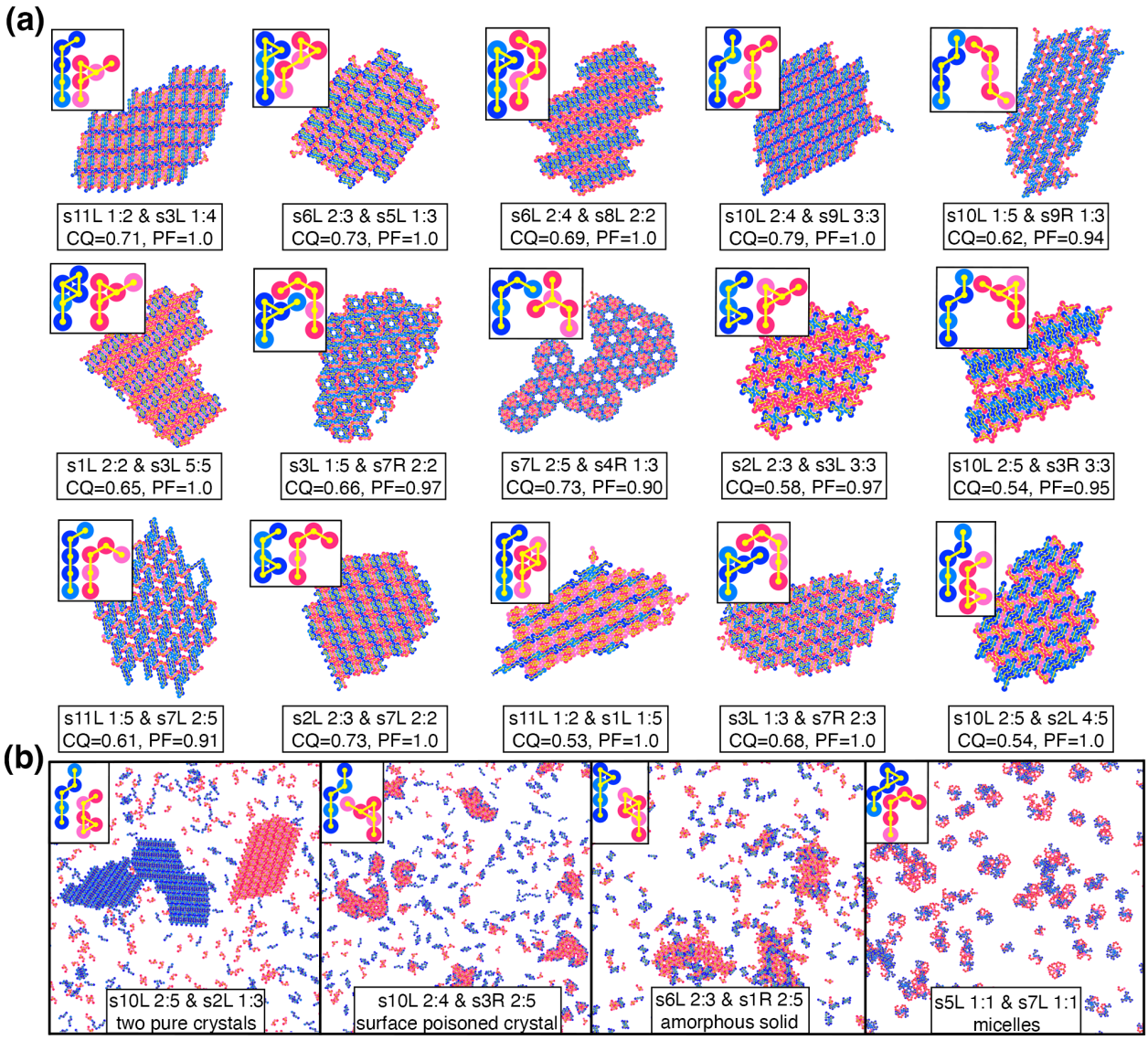

Figure 3: Examples of MD crystallization outcomes. (a) Snapshots of cocrystal clusters. For each cocrystal, insets show space-filling representations of coformers (top left), molecular identifiers of coformers, the CQ score of the cluster, and the packing fraction (PF) of the corresponding crystal structure (bottom). (b) Examples of simulation outcomes that are not cocrystals (from left to right): two single-component crystals, small single-component clusters with coformers acting as surfactants, amorphous clusters, and micelle-like aggregates.

larger co-crystallization rates than solution-based methods.[28,29] Other studies, performed on much smaller sets of coformers, report lower co-crystallization rates consistent with ours.[30] The shorter time scales and larger degree of supersaturation in our simulations compared to experiments might also disfavor the formation of well-ordered structures. Some simulations resulting in cocrystal clusters with low CQ values might yield more well-ordered structures if longer time scales were accessible. Finally, the lack of explicit solvent in our simulations might also influence co-crystallization rates.

Co-crystallization success rates of specific coformers range from zero to 22% in the GC set, and from zero to 32% in the BC set, deviating substantially (and statistically significantly) from the average success rates of 4.5% and 15.1%, respectively, as shown in Fig. S15. No simple correlations between coformer shape, locations of SIFGs, and co-crystallization rates could be identified, consistent with the complex thermodynamic landscapes discussed in the next section.

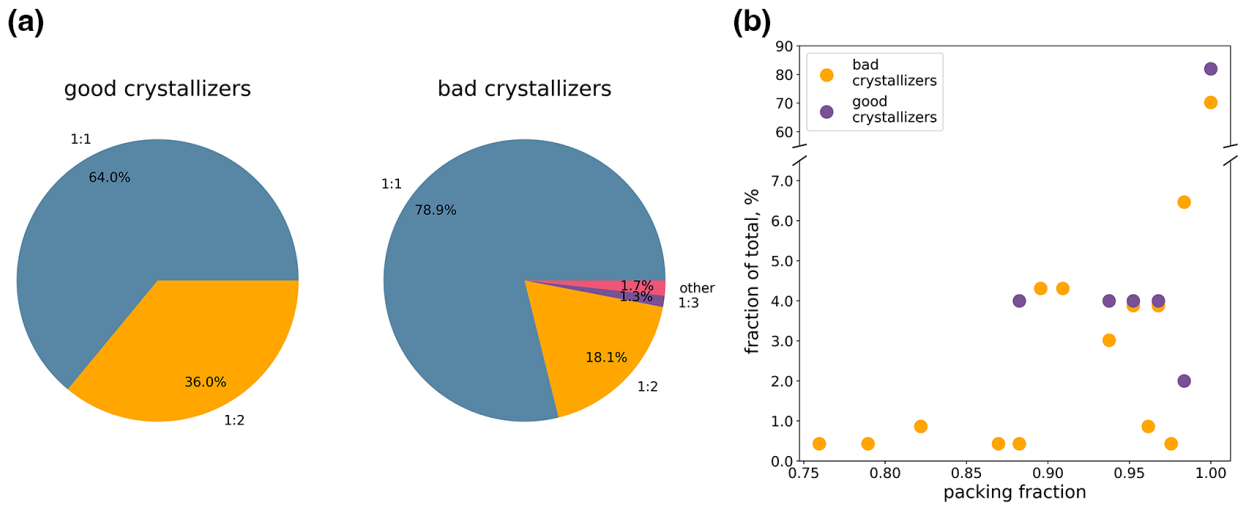

Figure 4: Statistics of composition and density of cocrystals observed in MD simulations. (a) Distributions of coformer composition for cocrystals in the GC and BC set. (b) Histograms of packing fractions of cocrystals.

Cocrystals formed in our simulations have a similar composition distribution to what is reported in the literature.[8,31–34] As illustrated in Fig. 4a, a majority of cocrystals has 1:1 composition. The remaining cocrystals are almost exclusively of 1:2 composition. In the BC

set, we observe a small number of cocrystals with other compositions, including 1:3, 1:4, 2:3, and 1:5, consistent with sporadic reports of such cocrystals in the literature.[35,36] Note that co-crystallization of solutions with an excess of one of the two coformers might facilitate the formation of cocrystals with compositions other than 1:1 but we have not simulated such asymmetric mixtures.[37]

The simple geometry of our molecular model complicates a direct comparison of crystal densities with experiments. Given that functional groups within all molecules are arranged on the sites of a triangular lattice, the largest possible crystal density of our model molecules corresponds to that of a simple close-packed hexagonal lattice of functional groups. We find that 82% and 70% of all co-crystals in the GC and BC set, respectively, are close-packed in this sense, as illustrated in Fig. 4b. (We assign a packing fraction of 1.0 to these crystals). We observe a significant number of cocrystals with smaller packing fractions between 0.9 and 1.0 (corresponding to different numbers of unoccupied sites of the hexagonal lattice in the unit cell), and, in the BC set, a small number of even more "open" crystals. (See Fig. 3a for examples of cocrystals with different packing fractions.) Cocrystal energy landscapes available in the literature similarly show that low-energy cocrystals typically have densities that can deviate by around 10% from the highest density.[38,39]

The molecular models used in this work are only simple caricatures of actual molecules. However, the crystallization behavior of these models resembles that of real molecules in several important aspects, as described above, and is robust with respect to changes in the model. In previous work, we have simulated related models with interactions that are not all pairwise attractive but mimic charged functional groups with both attractive and repulsive interactions. (See Supporting Information of Ref.[23]) While these "charged" models produce more complex (and perhaps more realistic) crystal structures, their overall crystallization propensity and the relative prevalence of crystals with different compositions are consistent with the simpler, purely attractive models we study in this work. This suggests that our models capture some of the key features driving the complex crystallization behavior of real

molecules.

Crystallization Outcomes vs. Crystal Energy Landscapes

The central question addressed in this paper is this: Could the crystallization outcomes described above have been predicted for each case based on the energy landscapes of single-component crystals and cocrystals of the two coformers? For each of our simulated systems of coformers, we calculated ΔE (as defined in Eq. 1) for all low-energy cocrystals. Here, E_{co} is the cocrystal lattice energy, and E_1 and E_2 are the lowest lattice energies of one-component crystals of coformers 1 and 2, respectively, as identified by POLYNUM; c_1 and c_2 are the stoichiometric coefficients of coformers 1 and 2, respectively, in the cocrystal. We then identified the cocrystal with the smallest value of ΔE, which has the largest thermodynamic advantage with respect to single-component crystals. We will refer to this cocrystal as CC* in the following, and to its value of ΔE as ΔE_{min}. Based on thermodynamic arguments alone, if $\Delta E_{\mathrm{min}} < 0$ for a given system of coformers, then a cocrystal (more specifically, CC*) should form. Systems with $\Delta E_{\mathrm{min}} > 0$ should form a single-component crystal.

We first discuss the coformer systems from the set of good crystallizers (GC). The distribution of ΔE_{min}, expressed in percent of the lowest crystal energy E_0, is shown in Fig. 5a together with crystallization outcomes. The majority of all coformer systems in the GC set (59.5%) have $\Delta E_{\mathrm{min}} \leq 0$ and, therefore, the thermodynamic prerequisites for co-crystallization. (The large number of systems with $\Delta E_{\mathrm{min}} = 0$ is an artifact of the simple interactions of our molecules, as discussed in the SI.) The cocrystals we observed in our simulations are indeed all found within this group. This result confirms that a thermodynamic driving force is, in fact, necessary for co-crystallization, as has been reported before.[15,19,40–42] However, only a small fraction of all systems with $\Delta E \leq 0$ actually form a cocrystal (9.0% if only systems with $\Delta E < 0$ are considered, 7.7% if the thermodynamically ambiguous cases with $\Delta E = 0$ are also included). We obtain essentially the same result when we restrict our energy landscapes to single-component crystals with small asymmetric unit cells ($Z' = 1$)

and cocrystals with $Z' = 1$ and a 1:1 composition, which mimics the computational limits of typical CSP studies of real molecules (see Fig. S9 in the SI.) The probability of forming a cocrystal increases with decreasing ΔE: Systems with large negative ΔE are much more likely to co-crystallize than systems with $\Delta E \approx 0$. The low rate of co-crystallization for systems with thermodynamically favorable cocrystals implies a clear kinetic disadvantage of cocrystals over single-component crystals. We furthermore observe kinetic effects even among those systems that do form cocrystals. For 16 % of these systems, the cocrystal that forms is not CC* but a cocrystal with a larger ΔE, as shown in Supplementary Fig. S7a.

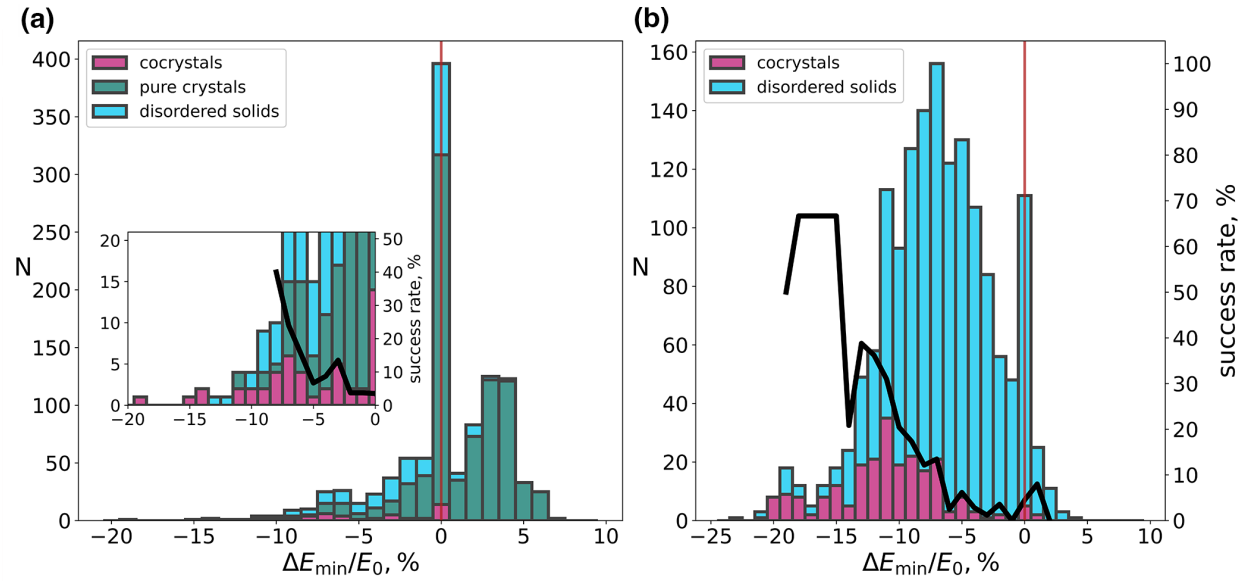

Figure 5: Distributions of ΔE_{min} (expressed in percent of the lowest crystal energy E_0) for coformer systems in the GC set (a) and in the BC set (b). Negative ΔE_{min} indicates a thermodynamically stable cocrystal. Different colors indicate simulation outcomes. The inset in (a) focuses on the part of the distribution with $\Delta E_{\text{min}} < 0$. Thick black lines show the co-crystallization success rate (calculated only for bins with at least 10 coformer systems).

While the majority of all GC systems form single-component crystals, 19.5% form structures with poor crystallinity ($CQ < 0.4$). This is remarkable given that all molecules in the GC set form well-ordered structures ($CQ > 0.4$) when crystallized by themselves. Visual examination indicated that approximately two-thirds of these poorly crystalline systems produce clusters that incorporate significant fractions (20% or larger) of both coformers; the remaining systems are dominated by clusters strongly enriched in one of the two coformers,

including clusters where one coformer acts as surfactant and "poisons" the surface of pure clusters of the other coformer. (See Fig. 3b for examples of disordered clusters with different compositions.)

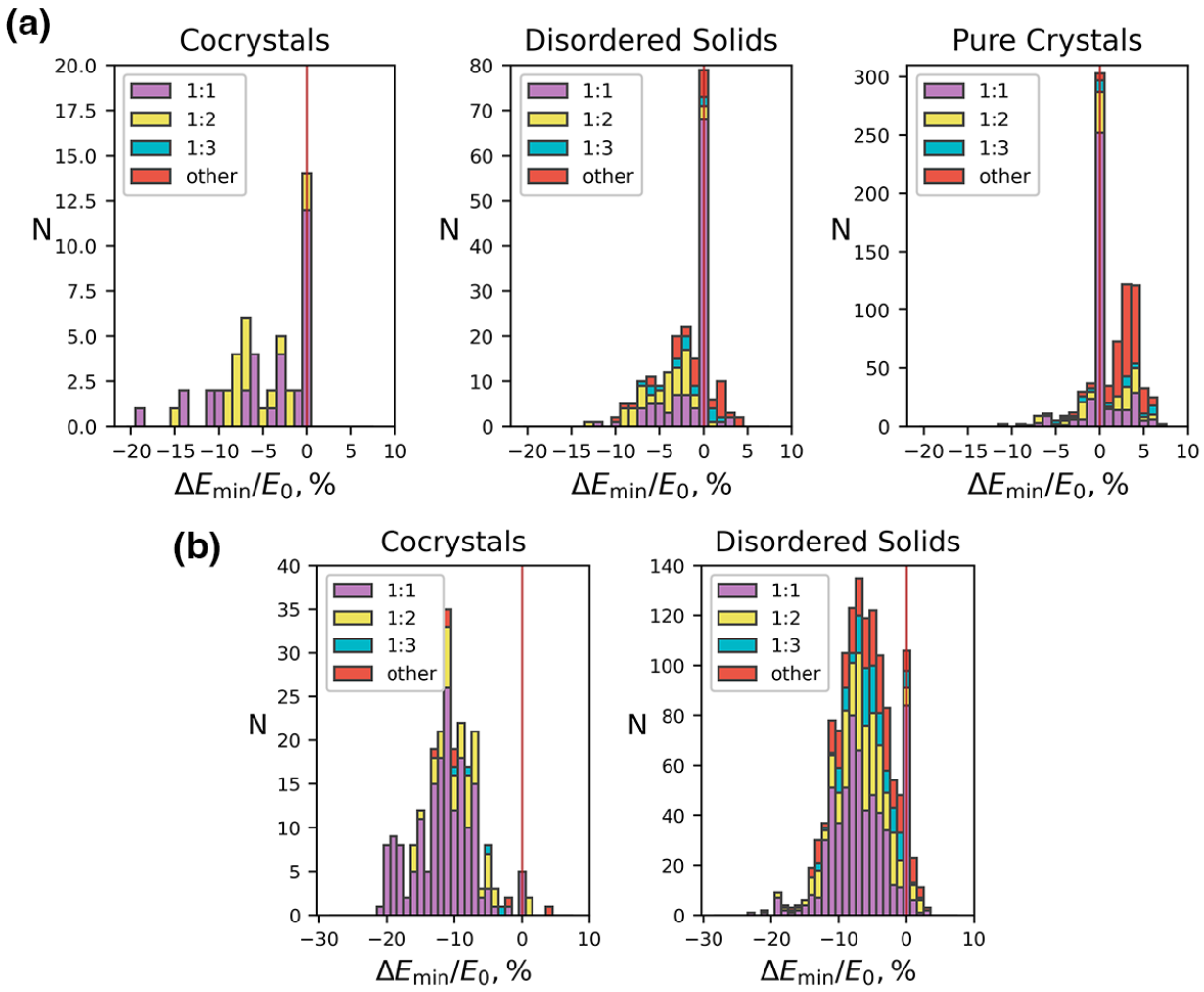

Figure 6: Distributions of ΔE_{min} (expressed in percent of the lowest crystal energy E_0) for coformer systems in the GC set (a) and in the BC set (b). Negative ΔE_{min} indicates a thermodynamically stable cocrystal. Different colors indicate the composition of CC*.

We now turn to the coformer systems from the set of bad crystallizers (BC). A histogram of ΔE for all 1535 systems in the BC set is shown in Fig. 5b. The overall rate of co-crystallization is 15.1% in this set—a value that is more than three times larger than that in the GC set yet still surprisingly small, given the abundance of stable cocrystals for these systems. Almost all systems in this set have thermodynamically stable cocrystals ($\Delta E_{\text{min}} \leq$

0), and many have cocrystals with lower energies than any of the single-component crystals of their coformers (see SI, Fig. S10b). Similar to the GC set, the vast majority of cocrystals observed in the BC set are coformer systems with $\Delta E_{\mathrm{min}} \leq 0$, although we observe a small number of co-crystals that are kinetic products $\Delta E_{\mathrm{min}} > 0$. In addition, kinetic effects among the co-crystallizing system are quite pronounced: For 25 % of these systems, the cocrystal that forms is not CC*, as shown in Supplementary Fig. S7b.

The much larger thermodynamic preference for cocrystals in the BC set compared to the GC set can be rationalized by the fact that single-component crystals have, on average, substantially lower lattice energies in the GC set than in the BC set, as illustrated in Fig. S8. This trend is consistent with simple models of kinetic trapping: Molecules that readily form high-quality crystals (GC) often have access to a few crystal structures with exceptionally low energy, reflecting packing motifs that are energetically particularly favorable. Such low-energy packing motifs are often absent from the crystal landscapes of molecules that tend to form disordered solids, which instead feature a large number of structural motifs with similar energy that compete with one another kinetically.[23,24,43,44] By contrast, the distributions of low-energy cocrystals in the GC and BC set are remarkably similar (Fig. S8), indicating that the lack of particularly favorable single-component crystals of two coformers does not typically translate to their cocrystals. In fact, we observe a quantitatively similar dependence of co-crystallization rate on ΔE in the GC and BC sets (thick black curves in Fig. 5). The higher overall co-crystallization rate in the BC set therefore appears to be due to the larger average thermodynamic advantage of cocrystals in that set. These results suggest that including coformers that by themselves tend to form glasses or poorly crystalline solids in cocrystal screenings can be beneficial and should result in larger success rates and new cocrystals.

Cocrystal composition as an indicator of crystallization kinetics

In recent work, we showed that for more than half of all racemic mixtures of our model molecules, the lowest-energy crystal is neither racemic nor enantiopure but has an unusual composition of enantiomers (e.g., 1:2, 1:3, 2:3, etc.).[23] Crystals with such compositions do not form in our simulations of racemic mixtures and are observed in experiments only extremely rarely. In fact, the vast majority of racemic mixtures with low-energy crystals of unusual composition form disordered solids, not crystals, indicating that the crystallization kinetics are partly encoded in crystal energy landscapes.

For co-crystallization studied in this work, we find a similar correlation between the thermodynamics of cocrystals with unusual composition and co-crystallization rates: Only 3% of cocrystals that form spontaneously in our MD simulations have compositions other than 1:1 and 1:2, but a much larger fraction of coformer systems (12% of systems with $\Delta E_{\text{min}} \leq 0$ in the GC set, and 24% in the BC set) have a CC* with an unusual composition. Among those systems, the vast majority do not co-crystallize in MD simulations, but form disordered solids or single-component crystals, as illustrated in Fig. 6. Examples of systems that have a thermodynamically stable cocrystals of unusual composition but fail to co-crystallize are shown in Fig. 7.

Cocrystals with compositions other than 1:1 are rarely included in CSP because of the computational cost associated with larger unit cells and the apparent experimental irrelevance of these crystals. However, knowledge of their thermodynamic status could provide valuable guidance for solid-state screening. The microscopic details of the complicated crystallization kinetics that prevent systems with low-energy cocrystals of unusual composition from co-crystallizing remain unclear at this point. Based on our simulation results, we speculate that such coformer systems have access to a multitude of energetically equivalent structural motifs with different compositions that form in the early stages of the crystallization process, interfering with the formation of well-ordered cocrystals.

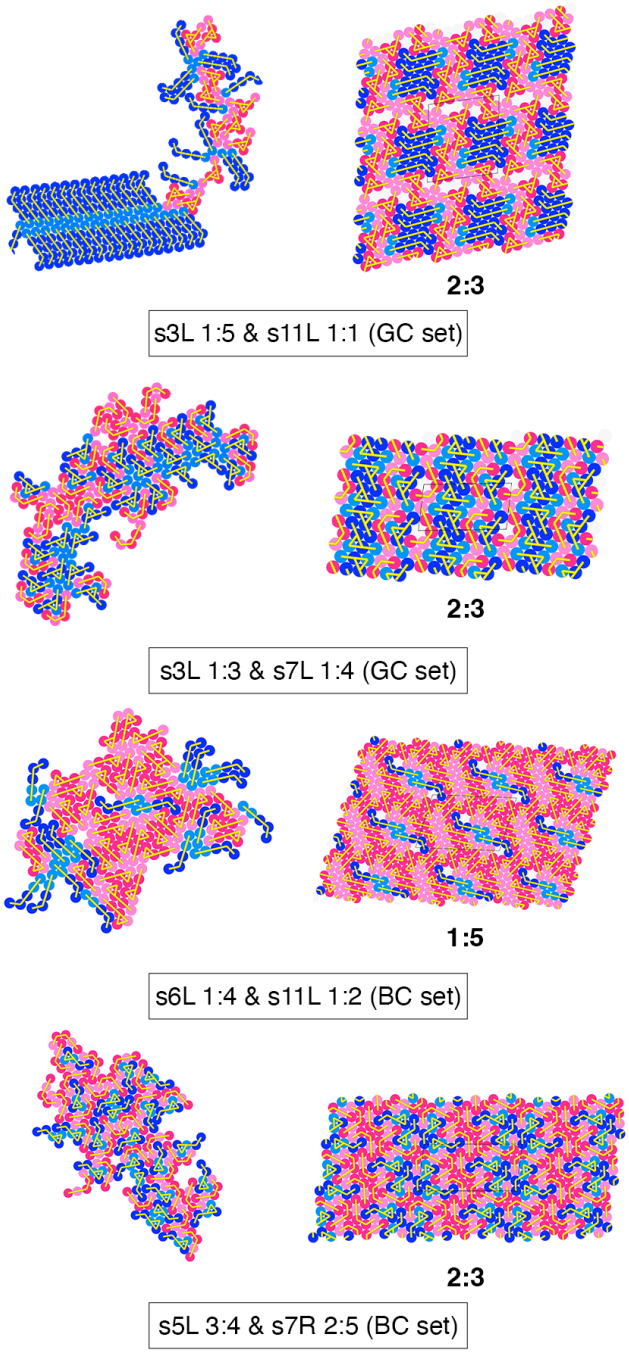

Figure 7: Examples of coformer systems with thermodynamically stable cocrystals of unusual composition. For each of the four examples, typical disordered clusters from MD simulations are shown on the left, and the CC* is shown on the right together with its coformer composition.

Co-crystallization of coformers with disparate crystal energies

Some reports in the literature suggest that choosing coformers whose single-component crystals have similar lattice energy could result in a larger probability of forming a cocrystal.[33,34] Differences in lattice energy are inferred from differences in melting temperatures[45,46] or differences in heat of sublimation. If solvation energies are similar for the two coformers, their relative solubility will therefore be correlated with lattice energy differences. The relative solubility may influence the chance of forming a co-crystal from solution because the component with smaller solubility may rapidly precipitate.[47]

We have tested the correlation between co-crystallization likelihood and single-component lattice energy differences of coformers with our models. Fig. 8 shows the distribution of energy differences between the lowest-energy single-component crystals of the two coformers,

$$\Delta E_{\text{pure}} = \frac{|E_1 - E_2|}{|\min(E_1, E_2)|}. \tag{2}$$

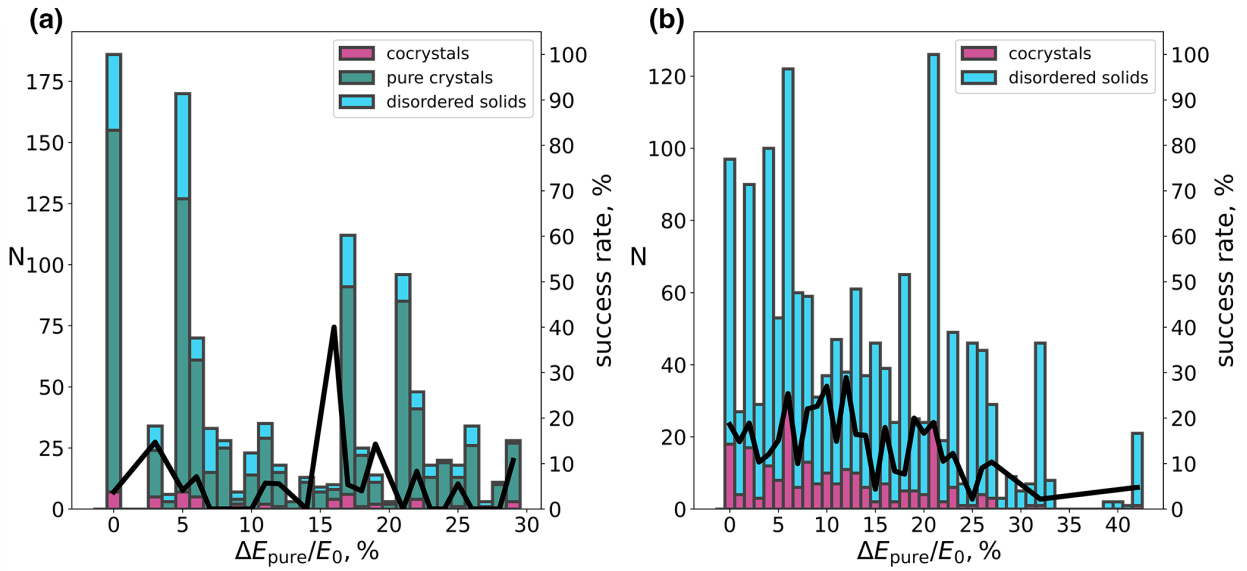

Figure 8: Distribution of energy differences between lowest-energy single-component crystals of coformers in the GC set (a) and the BC set (b).

The range of lattice energy differences in our model molecules is limited to $\approx 40\%$, due to molecules' similar size and interaction energies. Lattice energy differences of real small

organic molecules with comparable sizes are typically modest, too.[48] Within that energy range, we do not observe a strong dependence of the co-crystallization rate on ΔE_{pure}, suggesting that selecting coformers based on similar lattice energy is not an effective strategy for cocrystal discovery, at least not for molecules of similar size.

Conclusions

It is widely assumed that the co-crystallization of (organic) molecules is driven by thermodynamics and that it can, in principle, be predicted from accurate crystal free energy landscapes.[19,21,30,41] In practice, such landscapes are still difficult to obtain due to their extraordinarily large computational cost. In this paper, we employ a schematic molecular model whose computational efficiency allows us, for the first time, to compute highly accurate cocrystal energy landscapes for thousands of different coformer systems and to compare these thermodynamic predictors to simulated co-crystallization outcomes. Our calculations show that thermodynamic stability is an (in most cases) necessary but clearly not sufficient prerequisite for the formation of a cocrystal from a solution of coformers. Close to 90% of all coformer systems with thermodynamically stable cocrystals form either one-component crystals or highly disordered solids. Our simulations demonstrate a marked kinetic disadvantage of cocrystals with respect to other types of solids, whose microscopic origin must lie in the nucleation and growth processes in the early stages of crystal formation. In recent work, we have shown that the simulated crystallization of racemic mixtures can be predicted with near-perfect accuracy when attachment rates of different oligomeric growth units in solution are accounted for.[24] Future work will reveal if the observed kinetic disadvantage of cocrystals can be similarly understood in terms of attachment rates of solution oligomers of different compositions. Our current results show that some kinetic information can be gleaned from the thermodynamic standing of co-crystals with an unusual composition of coformers (i.e., other than 1:1, 1:2, or 1:3). While many coformer systems have thermodynamically stable

co-crystals of this kind, they form only extremely rarely in experiments. Such coformer systems, however, have a strongly increased chance of forming single-component crystals or disordered solids. We hope that our work will renew interest in the kinetic effects of molecular (co-)crystallization and help catalyze efforts to further improve thermodynamics-based crystal structure prediction methods.

Methods

Model and MD simulations

All molecular dynamics simulations were performed with HOOMD.[27] Functional groups of molecules have mass m and diameter σ, which we use as our units of mass and length; the unit of energy is ϵ. Langevin equations of motion for rigid bodies are integrated with a time step of $0.004\sqrt{m\sigma^2/\epsilon}$ and a damping coefficient of $5.0\sqrt{m\epsilon/\sigma^2}$. All simulation snapshots were produced with OVITO.[49]

Intermolecular and non-bonded intramolecular interactions between functional groups are represented by short-ranged pair potentials, $u(r) = u_{\text{att}}(r) + u_{\text{rep}}(r)$, where

$$u_{\text{att}}(r) = \begin{cases} -\epsilon_{\text{att}}, & \text{if } r < \sigma, \\ -\frac{\epsilon_{\text{att}}}{2}\left(\cos\left[\frac{(r-\sigma)\pi}{\omega}\right] + 1\right) & \text{if } \sigma \leq r < \sigma + \omega, \\ 0 & \text{if } r \geq \sigma + \omega, \end{cases}$$

and

$$u_{\text{rep}}(r) = \begin{cases} \epsilon_{\text{rep}}\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon_{\text{rep}} & \text{if } r < \sigma, \\ 0 & \text{else.} \end{cases}$$

We set $\epsilon_{\text{rep}} = 5\epsilon$, and $\omega = 0.2\sigma$. For SIFG, we use $\epsilon_{\text{att}} = 5\epsilon$, all other pairs of functional groups have $\epsilon_{\text{att}} = \epsilon$. These pair potentials are identical to those used in our previous

work.[23,24]

Crystal energy landscapes

For each coformer pair, an extensive polymorph search was performed using a recently developed algorithm (POLYNUM) that solves the exact-cover problem for our rigid molecular shapes in 2D periodic unit cells.[23] POLYNUM produces between 10^5 and 10^7 of packings for a given combination of molecules, including crystals with different compositions and numbers of molecules in the asymmetric unit (Z'). Crystal structures with packing fractions smaller than 1.0 are identified by including empty space as effective particles in POLYNUM. Because the numerical cost of enumerating molecular packings increases rapidly with the number of molecules, we restrict our search to unit cells containing at most 13 molecules in the unit cell and 500,000 structures per unit cell. Despite these restrictions, POLYNUM identified the vast majority of all crystal structures found in our simulations. For a handful of coformer systems in the BC set we have observed crystalline clusters in our simulations that could not be matched with any packing identified by POLYNUM. These are crystals with very large unit cells, as illustrated in Fig. S16. We have manually constructed the unit cells for these structures and included them in our structure analysis. The limited number of molecules in the unit cell also implies limits on the composition of cocrystals that can be identified with POLYNUM. It is therefore possible that the group of cocrystals with unusual composition is substantially larger than suggested by our current results and extends to lower energies.

Analyzing MD simulations

We use molecular order parameters to automatically detect and classify the largest crystalline clusters that appear in our MD simulations. The details of this procedure are described in Ref.[23] In brief, we first categorize clusters of molecules according to composition: clusters containing less than 20% of either of the two species are categorized as single-component clusters, and those with mole fractions between 0.2 and 0.8 as potential cocrystals. In the

next step, the microscopic structure of a given cluster is compared to the low-energy single-component crystals or cocrystals identified with POLYNUM, respectively. (More specifically, we use the 100 low-energy single-component crystals, the 200 low-energy cocrystals with 1:1 composition, and the 200 low-energy cocrystals with composition 1:2 and 1:3 for this comparison.) Based on the similarity of radial distribution functions and relative orientations of nearest-neighbor molecules, the crystal structure P^* that best matches the cluster is identified. The CQ score used to quantify the crystallinity of a given cluster is defined as the fraction of molecules in the cluster whose local environment matches that of a molecule in P^*. Clusters with $CQ > 0.4$ are classified as crystalline, all others as disordered. A well-ordered cluster from MD simulations can in principle, be assigned a low CQ score if its crystal structure has unusually high energy and is therefore not included in the lists of low-energy polymorphs used for structure comparison. However, we have visually inspected all simulation trajectories and confirmed that no clusters were mislabeled as disordered in this way.

Acknowledgement

The research presented in this work was supported by the National Science Foundation under Grant No. CHE-1900626 and by the W. M. Keck Foundation. The authors thank the Center for High Performance Computing at the University of Utah for the technical support and computational resources.

Supporting Information Available

Additional simulation details, figures, data analysis, methods, and extended discussion.

References

(1) Kalepu, S.; Nekkanti, V. Insoluble drug delivery strategies: Review of recent advances and business prospects. *Acta Pharmaceutica Sinica B* **2015**, *5*, 442–453.

(2) Aitipamula, S. et al. Polymorphs, salts, and cocrystals: What's in a name? *Crystal Growth and Design* **2012**, *12*, 2147–2152.

(3) Habgood, M. Analysis of enantiospecific and diastereomeric cocrystal systems by crystal structure prediction. *Crystal Growth and Design* **2013**, *13*, 4549–4558.

(4) Harmsen, B.; Leyssens, T. Dual-Drug Chiral Resolution: Enantiospecific Cocrystallization of (S)-Ibuprofen Using Levetiracetam. *Crystal Growth and Design* **2018**, *18*, 441–448.

(5) Bolton, O.; Simke, L. R.; Pagoria, P. F.; Matzger, A. J. High power explosive with good sensitivity: A 2:1 cocrystal of CL-20:HMX. *Crystal Growth and Design* **2012**, *12*, 4311–4314.

(6) Zheng, Q.; Unruh, D. K.; Hutchins, K. M. Cocrystallization and Thermal Behaviors of the Micropollutants Gemfibrozil, Aceclofenac, and Bisphenol A. *Crystal Growth and Design* **2022**, *22*, 2208–2217.

(7) Kunkel, D. A.; Hooper, J.; Bradley, B.; Schlueter, L.; Rasmussen, T.; Costa, P.; Beniwal, S.; Ducharme, S.; Zurek, E.; Enders, A. 2D Cocrystallization from H-Bonded Organic Ferroelectrics. *Journal of Physical Chemistry Letters* **2016**, *7*, 435–440.

(8) Habgood, M.; Price, S. L. Isomers, conformers, and cocrystal stoichiometry: Insights from the crystal energy landscapes of caffeine with the hydroxybenzoic acids. *Crystal Growth and Design* **2010**, *10*, 3263–3272.

(9) Wood, P. A.; Feeder, N.; Furlow, M.; Galek, P. T.; Groom, C. R.; Pidcock, E.

Knowledge-based approaches to co-crystal design. *CrystEngComm* **2014**, *16*, 5839–5848.

(10) Sarkar, N.; Gonnella, N. C.; Krawiec, M.; Xin, D.; Aakeröy, C. B. Evaluating the Predictive Abilities of Protocols Based on Hydrogen-Bond Propensity, Molecular Complementarity, and Hydrogen-Bond Energy for Cocrystal Screening. *Crystal Growth and Design* **2020**, *20*, 7320–7327.

(11) Sandhu, B.; McLean, A.; Sinha, A. S.; Desper, J.; Sarjeant, A. A.; Vyas, S.; Reutzel-Edens, S. M.; Aakeröy, C. B. Evaluating Competing Intermolecular Interactions through Molecular Electrostatic Potentials and Hydrogen-Bond Propensities. *Crystal Growth and Design* **2018**, *18*, 466–478.

(12) Devogelaer, J. J.; Meekes, H.; Tinnemans, P.; Vlieg, E.; de Gelder, R. Co-crystal Prediction by Artificial Neural Networks**. *Angewandte Chemie - International Edition* **2020**, *59*, 21711–21718.

(13) Heng, T.; Yang, D.; Wang, R.; Zhang, L.; Lu, Y.; Du, G. Progress in Research on Artificial Intelligence Applied to Polymorphism and Cocrystal Prediction. *ACS Omega* **2021**, *6*, 15543–15550.

(14) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. Can the formation of pharmaceutical cocrystals be computationally predicted? 2. Crystal structure prediction. *Journal of Chemical Theory and Computation* **2009**, *5*, 1432–1448.

(15) Chan, H. C.; Kendrick, J.; Neumann, M. A.; Leusen, F. J. Towards ab initio screening of co-crystal formation through lattice energy calculations and crystal structure prediction of nicotinamide, isonicotinamide, picolinamide and paracetamol multi-component crystals. *CrystEngComm* **2013**, *15*, 3799–3807.

(16) Sun, G.; Jin, Y.; Li, S.; Yang, Z.; Shi, B.; Chang, C.; Abramov, Y. A. Virtual Coformer Screening by Crystal Structure Predictions: Crucial Role of Crystallinity in Pharmaceutical Cocrystallization. *Journal of Physical Chemistry Letters* **2020**, *11*, 8832–8838.

(17) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.

(18) Nyman, J.; Day, G. M. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31132–31143.

(19) Taylor, C. R.; Day, G. M. Evaluating the Energetic Driving Force for Cocrystal Formation. *Crystal Growth and Design* **2018**, *18*, 892–904.

(20) Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Accounts of Chemical Research* **2009**, *42*, 117–126.

(21) Issa, N.; Karamertzanis, P. G.; Welch, G. W.; Price, S. L. Can the formation of pharmaceutical cocrystals be computationally predicted? I. Comparison of lattice energies. *Crystal Growth and Design* **2009**, *9*, 442–453.

(22) Wicker, J. G.; Crowley, L. M.; Robshaw, O.; Little, E. J.; Stokes, S. P.; Cooper, R. I.; Lawrence, S. E. Will they co-crystallize? *CrystEngComm* **2017**, *19*, 5336–5340.

(23) Carpenter, J. E.; Grünwald, M. Heterogeneous Interactions Promote Crystallization and Spontaneous Resolution of Chiral Molecules. *Journal of the American Chemical Society* **2020**, *142*, 10755–10768.

(24) Carpenter, J. E.; Grünwald, M. Pre-Nucleation Clusters Predict Crystal Structures in Models of Chiral Molecules. *Journal of the American Chemical Society* **2021**, *143*, 21580–21593, PMID: 34918909.

(25) Edwards, A. J.; Mackenzie, C. F.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A. Intermolecular interactions in molecular crystals: What's in a name? *Faraday Discussions* **2017**, *203*, 93–112.

(26) Liu, G.; Wei, S. H.; Zhang, C. Review of the Intermolecular Interactions in Energetic Molecular Cocrystals. *Crystal Growth and Design* **2020**, *20*, 7065–7079.

(27) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science* **2020**, *173*, 109363.

(28) Karimi-Jafari, M.; Padrela, L.; Walker, G. M.; Croker, D. M. Creating cocrystals: A review of pharmaceutical cocrystal preparation routes and applications. *Crystal Growth and Design* **2018**, *18*, 6370–6387.

(29) Friščič, T.; Jones, W. Recent advances in understanding the mechanism of cocrystal formation via grinding. *Crystal Growth and Design* **2009**, *9*, 1621–1637.

(30) Abramov, Y. A.; Iuzzolino, L.; Jin, Y.; York, G.; Chen, C.-H.; Shultz, C. S.; Yang, Z.; Chang, C.; Shi, B.; Zhou, T.; Greenwell, C.; Sekharan, S.; Lee, A. Y. Cocrystal Synthesis through Crystal Structure Prediction. *Molecular Pharmaceutics* **2023**, *20*, 3380–3392, PMID: 37279175.

(31) Patrick Stahly, G. A survey of cocrystals reported prior to 2000. *Crystal Growth and Design* **2009**, *9*, 4212–4229.

(32) Kuminek, G.; Cavanagh, K. L.; Rodríguez-Hornedo, N. Measurement and mathematical relationships of cocrystal thermodynamic properties. *Pharmaceutical Crystals: Science and Engineering* **2018**, 223–271.

(33) Perlovich, G. L. Thermodynamic characteristics of cocrystal formation and melting

points for rational design of pharmaceutical two-component systems. *CrystEngComm* **2015**, *17*, 7019–7028.

(34) Perlovich, G. Melting points of one- and two-component molecular crystals as effective characteristics for rational design of pharmaceutical systems. *Acta crystallographica Section B, Structural science, crystal engineering and materials* **2020**, *76*, 696–706.

(35) Saikia, B.; Bora, P.; Khatioda, R.; Sarma, B. Hydrogen Bond Synthons in the Interplay of Solubility and Membrane Permeability/Diffusion in Variable Stoichiometry Drug Cocrystals. *Crystal Growth and Design* **2015**, *15*, 5593–5603.

(36) Shemchuk, O.; D'Agostino, S.; Fiore, C.; Sambri, V.; Zannoli, S.; Grepioni, F.; Braga, D. Natural Antimicrobials Meet a Synthetic Antibiotic: Carvacrol/Thymol and Ciprofloxacin Cocrystals as a Promising Solid-State Route to Activity Enhancement. *Crystal Growth and Design* **2020**, *20*, 6796–6803.

(37) Kole, G. K.; Tan, G. K.; Koh, L. L.; Vittal, J. J. Co-crystals of tetrakis-1,2,3,4-(4-carboxyphenyl)cyclobutane with dipyridyl spacers: design and serendipity. *CrystEngComm* **2012**, *14*, 6190–6195.

(38) Habgood, M.; Deij, M. A.; Mazurek, J.; Price, S. L.; Ter Horst, J. H. Carbamazepine co-crystallization with pyridine carboxamides: rationalization by complementary phase diagrams and crystal energy landscapes. *Crystal Growth and Design* **2010**, *10*, 903–912.

(39) Issa, N.; Barnett, S. A.; Mohamed, S.; Braun, D. E.; Copley, R. C.; Tocher, D. A.; Price, S. L. Screening for cocrystals of succinic acid and 4-aminobenzoic acid. *CrystEngComm* **2012**, *14*, 2454–2464.

(40) Shunnar, A. F.; Dhokale, B.; Karothu, D. P.; Bowskill, D. H.; Sugden, I. J.; Hernandez, H. H.; Naumov, P.; Mohamed, S. Efficient Screening for Ternary Molecular Ionic Cocrystals Using a Complementary Mechanosynthesis and Computational Structure Prediction Approach. *Chemistry - A European Journal* **2020**, *26*, 4752–4765.

(41) Oliveira, M. A.; Peterson, M. L.; Davey, R. J. Relative enthalpy of formation for Cocrystals of small organic molecules. *Crystal Growth and Design* **2011**, *11*, 449–457.

(42) Svärd, M.; Ahuja, D.; Rasmuson, Å. C. Calorimetric Determination of Cocrystal Thermodynamic Stability: Sulfamethazine-Salicylic Acid Case Study. *Crystal Growth and Design* **2020**, *20*, 4243–4251.

(43) Hormoz, S.; Brenner, M. P. Design principles for self-assembly with short-range interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 5193–5198.

(44) Gianni, S.; Freiberger, M. I.; Jemth, P.; Ferreiro, D. U.; Wolynes, P. G.; Fuxreiter, M. Fuzziness and Frustration in the Energy Landscape of Protein Folding, Function, and Assembly. *Accounts of Chemical Research* **2021**, *54*, 1251–1259.

(45) Salahinejad, M.; Le, T. C.; Winkler, D. A. Capturing the crystal: Prediction of enthalpy of sublimation, crystal lattice energy, and melting points of organic compounds. *Journal of Chemical Information and Modeling* **2013**, *53*, 223–229.

(46) Perlovich, G. L. Prediction of Sublimation Functions of Molecular Crystals Based on Melting Points: Cocrystal Formation Thermodynamics Application. *Crystal Growth and Design* **2017**, *17*, 4110–4117.

(47) Thakuria, R.; Delori, A.; Jones, W.; Lipert, M. P.; Roy, L.; Rodríguez-Hornedo, N. Pharmaceutical Cocrystals and Poorly Soluble Drugs. *International Journal of Pharmaceutics* **2013**, *453*, 101–125.

(48) Ouvrard, C.; Mitchell, J. B. O. Can We Predict Lattice Energy from Molecular Structure? *Acta Crystallographica Section B: Structural Science* **2003**, *59*, 676–685.

(49) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO-the

Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **2010**, *18*, 015012(7pp).

TOC Graphic