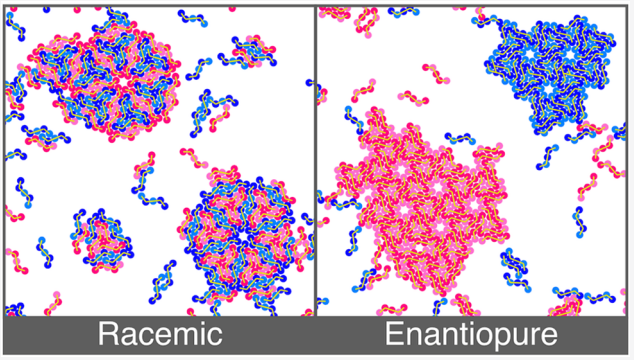
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Heterogeneous Interactions Promote Crystallization and Spontaneous Resolution of Chiral Molecules

John E. Carpenter and Michael Grünwald*

Cite This: J. Am. Chem. Soc. 2020, 142, 10755−10768
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predicting the crystallization of chiral molecules from solution is a major challenge in the chemical sciences. In this paper, we use molecular dynamics computer simulations to study the crystallization of a family of coarse-grained models of chiral molecules with a broad range of molecular shapes and interactions. Our simulations reproduce the experimental crystallization behavior of real chiral molecules, including racemic and enantiopure crystals, as well as amorphous solids. Using efficient algorithms for the packing of shapes, we enumerate millions of low-energy crystal structures for each model and analyze the thermodynamic landscape of polymorphs. In agreement with recent conjectures, our analysis shows that the ease of crystallization is largely determined by the number of competing polymorphs with low free energy. We find that this number and, hence, crystallization outcomes depend on molecular interactions in a simple way: Strongly heterogeneous interactions across molecules promote crystallization and favor the spontaneous resolution of racemic mixtures. Our results help rationalize a number of experimental observations and can provide guidance for the design of molecules and experimental conditions for desired crystallization outcomes.



I. INTRODUCTION

The crystallization of molecules is a necessary but oftentimes complicated step in many chemical processes. While many elements readily crystallize in a single structure, even small molecules (and certainly large biomolecules) can be difficult to crystallize and display pronounced polymorphism.[1,2] Bioactive organic compounds such as drugs, pesticides, and herbicides, need to be available in crystalline forms that are pure, soluble in water, and thermodynamically stable. The chemical industry therefore invests considerable time and effort in the discovery of suitable crystalline forms for new compounds.[3]

A large fraction of organic molecules are chiral, and crystallization can play an important role in the separation of chiral molecules. Chiral compounds are most easily synthesized as racemic mixtures that usually need to be resolved into enantiomers for use in drugs.[4−6] A straightforward and scalable method of chiral separation is the crystallization into physical mixtures of enantiopure crystals, so-called conglomerates.[7] However, this method of spontaneous chiral resolution is rarely applied in industry because chiral compounds preferentially form racemic crystals[8] and extensive searches for conditions that favor formation of enantiopure crystals are seldom performed.

Will a molecule crystallize under given experimental conditions? Will a racemic mixture spontaneously separate or form a racemic cocrystal? Despite much research activity, no definitive answers to these questions have been found.[9−21] In principle, thermodynamics dictates that the crystal with the lowest free energy prevails in equilibrium. Studies of chiral molecules have shown, however, that thermodynamics alone cannot explain the bias toward racemic crystals observed in experiments.[18,19,22] Clearly, kinetic effects of nucleation and growth play an important role in the crystallization of molecules and frequently result in the formation of amorphous solids and of crystals that are only metastable.[23,24]

Nevertheless, much information can be gleaned from a comprehensive list of possible crystal structures and their (free) energies.[25,26] Taking a purely thermodynamic view, modern computational methods of crystal structure prediction (CSP) can provide valuable guidance in the discovery of crystalline forms of organic and inorganic molecules.[27−33] However, while CSP methods can usually find the polymorphs that form in experiments, the majority of predicted crystal structures, even ones with low energies, are never realized in experiments.[13] Can such kinetically inaccessible polymorphs be identified by analyzing a molecule's thermodynamic crystal

Received: February 21, 2020
Published: May 12, 2020

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

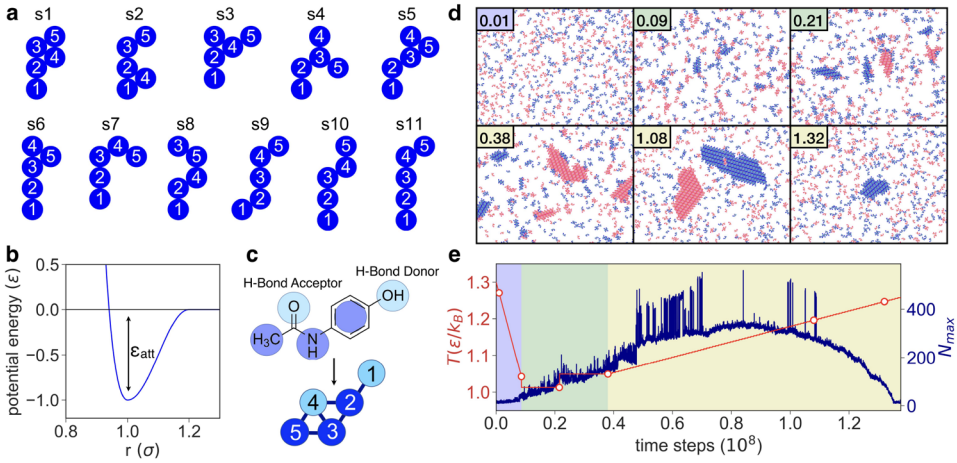


Figure 1. (a) The 11 chiral shapes studied in this work, each consisting of five coarse-grained functional groups. (b) Interaction potential for two functional groups as a function of distance r. (c) Example mapping of acetaminophen to coarse-grained shape s1. Functional groups that interact strongly, in this case via a hydrogen bond, are shown in light blue color. (d) Time series of snapshots from an MD simulation of the crystallization of molecule s2/1:3/5, illustrating the rapid formation of several enantiopure aggregates followed by slow coalescence and growth of high-quality crystals. Labels indicate simulation time in units of 10^8 time steps. Enantiomers are shown in blue and red colors. (e) Temperature (red curve) and number of molecules in the largest cluster (blue curve) as a function of time, for the trajectory illustrated in (d). Background colors indicate the different stages of the applied temperature protocol, as described in the main text. The largest CQ score is obtained after 1.08×10^8 time steps (see the red aggregate visible in the corresponding snapshot in panel (d)). Intermittent fluctuations of the size of the largest cluster indicate transient binding of aggregates.

landscapes? The "confusion principle", developed in the context of glass-forming materials, states that the presence of a large number of structural motifs with similar low energies leads to kinetic competition and a low crystallization propensity.[34,35] This principle has been shown to apply in at least a few inorganic systems[36,37] and has been conjectured to apply much more broadly also for organic molecules.[26] Related ideas have been developed to describe kinetic trapping in protein folding,[38] in the self-assembly of atomistic[39] and colloidal clusters,[40] and in the packing of polyhedral shapes.[41] Nevertheless, no firm guiding principles are currently available that would allow the prediction of crystallization behavior based on crystal energy landscapes or molecular properties.

In this paper, we present a computational study that aims to reveal thermodynamic and kinetic driving forces of chiral crystallization and to establish connections between thermodynamic landscapes and crystallization outcomes. In our efforts, we build on previous computational work that has demonstrated the usefulness of simple molecular models in elucidating structure formation mechanisms of chiral molecules.[16,20,21,42−45] Our study is based on a family of coarse-grained models of chiral molecules. The computational efficiency of these models allows us to simulate the crystallization dynamics of a wide range of molecular shapes and interactions and to fully determine the thermodynamic landscape of crystal structures. Despite the relative simplicity of our model, its analysis reveals general relations between molecular interactions, crystal energy landscapes, and crystallization trends that, we argue, are relevant for the crystallization of real molecules.

Before presenting a detailed discussion of our study, we briefly summarize the most import results below.

1. Molecules with large numbers of competing low-energy polymorphs do not crystallize well but tend to form disordered solids.
2. Most polymorphs with low energies are neither racemic nor enantiopure, but have different compositions of enantiomers. While these polymorphs never form spontaneously in our simulations (consistent with experiments), only molecules with comparably small numbers of such polymorphs crystallize well into racemic or enantiopure structures.
3. The number of competing polymorphs strongly depends on the heterogeneity of intermolecular interactions: Molecules with only a few strongly interacting functional groups have a much smaller number of competing polymorphs and tend to crystallize better than molecules whose functional groups interact with more uniform strength.
4. Heterogeneous interactions favor the spontaneous resolution of racemic mixtures.

II. RESULTS AND DISCUSSION

A. A Simple Model for Chiral Molecules. We represent molecules as rigid bodies consisting of five circular particles in two dimensions, as illustrated in Figure 1a. Constituent particles represent functional groups of the molecule and have a fixed diameter σ. To limit the number of different molecular shapes, we only consider molecules with angles of 60, 120, or 180 deg between triplets of connected functional groups. Furthermore, we consider only chiral shapes, resulting in a total of 11 distinct molecular shapes,[16] which we order according to increasing radius of gyration. Functional groups interact via short-ranged pair potentials that represent effective interactions in solution, as illustrated in Figure 1b (see Methods); solvent effects are included in these interactions, and solvent species are not modeled explicitly. For simplicity, we fix the range of these pair potentials and use the attractive interaction strength ϵ_{att} as the only parameter. We therefore coarsely categorize functional groups according to their strength of attractive interactions with other groups, as

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

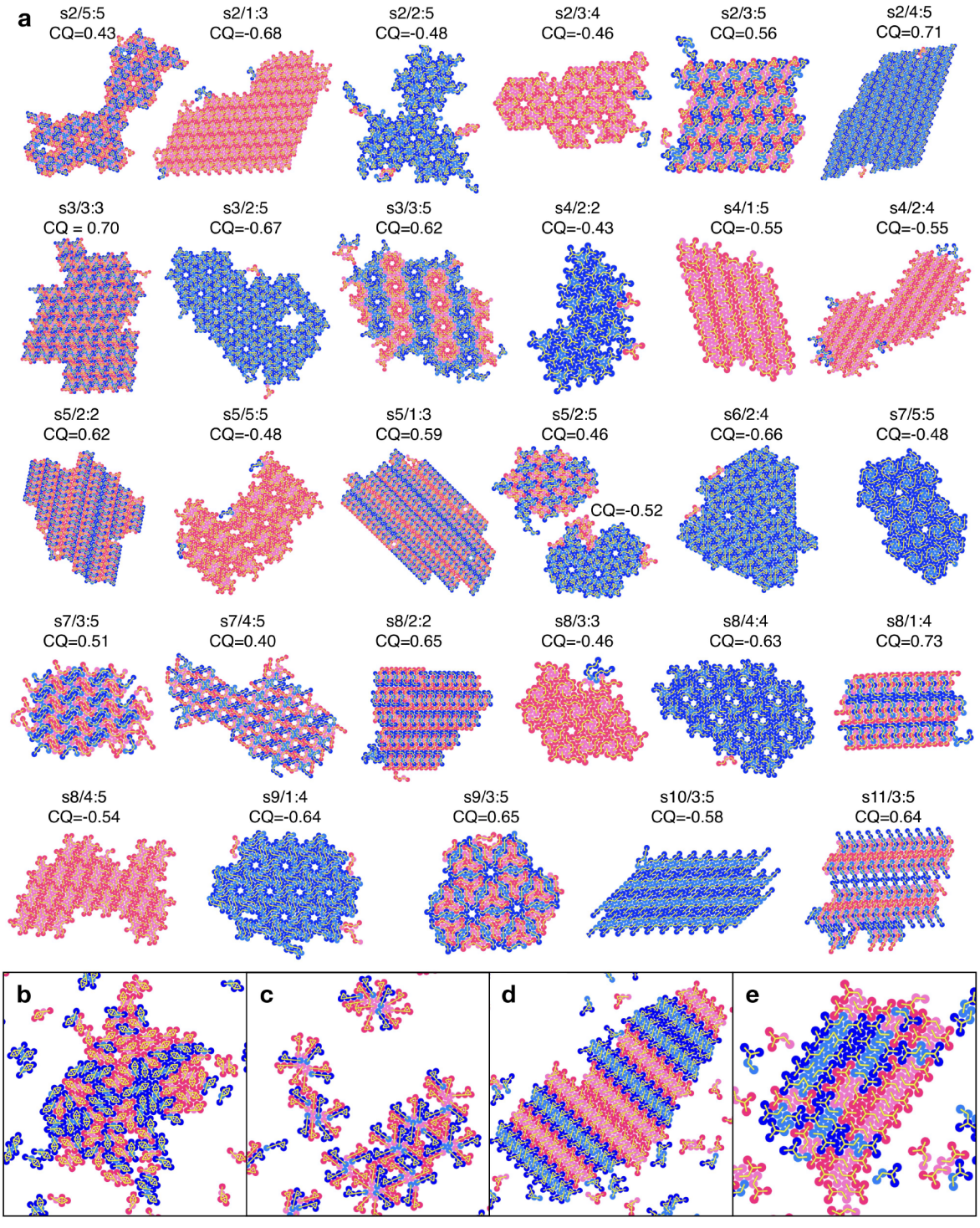
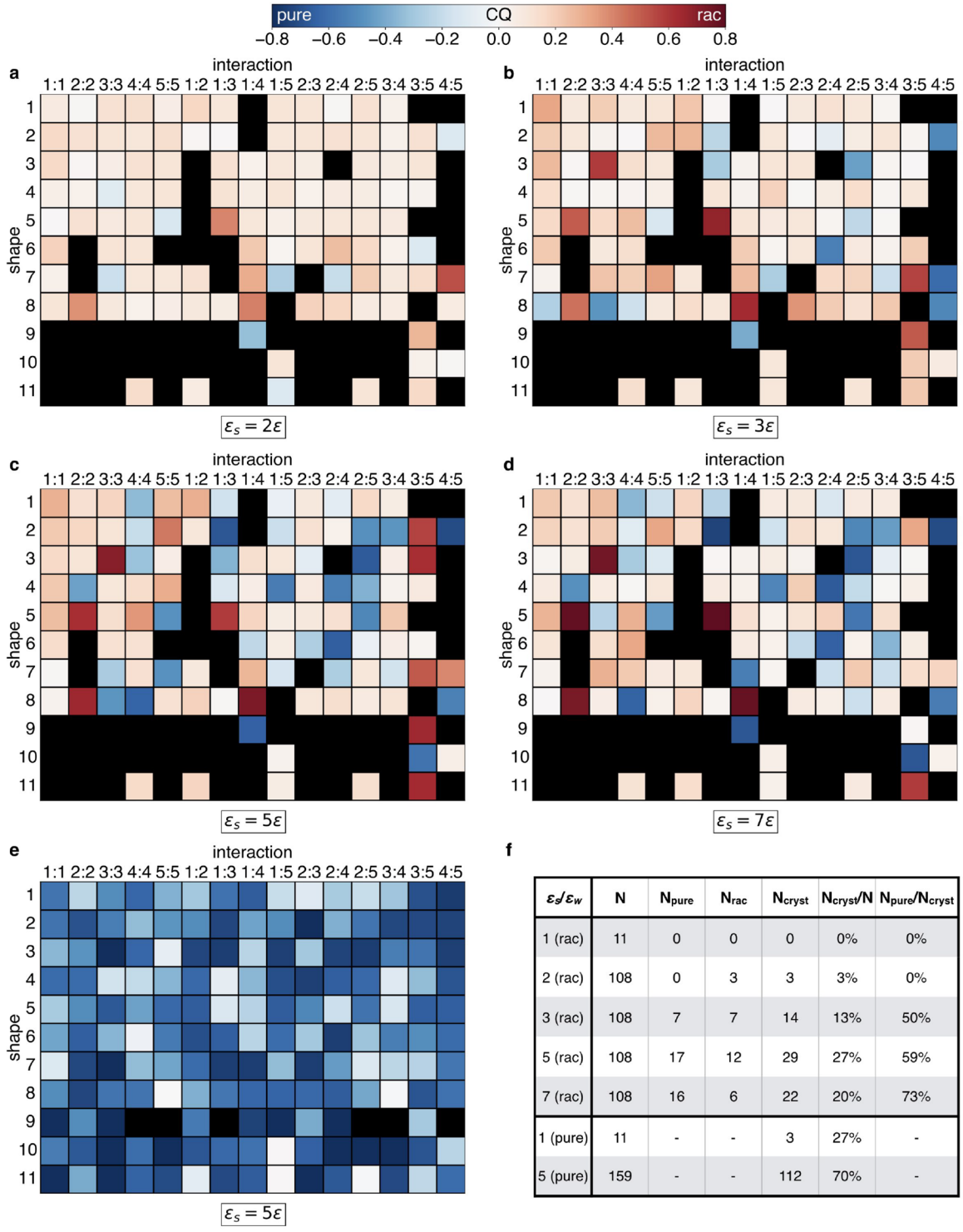


Figure 2. Crystallization outcomes of molecules with $\epsilon_s = 5\epsilon$. (a) Snapshots of all crystals with CQ > 0.4. (b–e) Snapshots of examples of disordered aggregates: (b) amorphous racemic, (c) aggregates of micelles, (d) 1D-racemic aggregates, (e) 0D-racemic (solid solution).

exemplified in Figure 1c. Specifically, we will focus on molecules that have a single pair of functional groups that interact strongly with $\epsilon_{att} = \epsilon_s$ (representing, for instance, a hydrogen bond), while other groups interact with uniform weaker attractions, $\epsilon_{att} = \epsilon$, which we also use as our unit of energy. A molecule can therefore be completely specified by its shape (i.e., 1–11), the pair of functional groups that interact strongly (e.g., 1:4, 3:3, etc.), and the magnitude of strong interactions ϵ_s. For example, we will refer to the coarse-grained molecule shown in Figure 1c with $\epsilon_s = 5\epsilon$ as "s1/1:4/5". For a fixed magnitude of strong interactions ϵ_s, this family of models is composed of a total of 159 distinct molecules that cover a wide space of shapes and interactions.[46] For a given molecular shape, molecules with different interactions can be interpreted either as chemically different molecules with similar shapes in

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755–10768



f

$\varepsilon_s/\varepsilon_w$	N	N_{pure}	N_{rac}	N_{cryst}	N_{cryst}/N	N_{pure}/N_{cryst}
1 (rac)	11	0	0	0	0%	0%
2 (rac)	108	0	3	3	3%	0%
3 (rac)	108	7	7	14	13%	50%
5 (rac)	108	17	12	29	27%	59%
7 (rac)	108	16	6	22	20%	73%
1 (pure)	11	-	-	3	27%	-
5 (pure)	159	-	-	112	70%	-

Figure 3. Crystallization outcomes depend on interaction strength and composition. (a–d) Grid summary of CQ scores for racemic solutions of molecules with different interaction strengths: (a) $\epsilon_s = 2\epsilon$, (b) $\epsilon_s = 3\epsilon$, (c) $\epsilon_s = 5\epsilon$, (d) $\epsilon_s = 7\epsilon$. Each molecule is represented by a cell and identified by its shape (vertical axis) and pair of strongly interacting functional groups (horizontal axis). Colors of cells reflect the highest CQ score achieved during simulation; negative values indicate enantiopure aggregates. Racemic structures are shown in red color, enantiopure structures in blue; larger magnitude indicates better crystalline quality. Black cells indicate molecules that formed 1D- or 0D-racemic disordered aggregates, as discussed in the text. Several molecules with shape s9 are equivalent because of symmetry (i.e., 1:2 and 4:5, 2:2 and 4:4, etc.); redundant molecules are blacked out. (e) Grid summary of CQ scores for enantiopure solutions with $\epsilon_s = 5\epsilon$. Darker colors indicate higher CQ scores. (f) Table summarizing crystallization outcomes. N is the number of simulated systems (those that form 0D-racemic and 1D-racemic structures are not counted), N_{pure} and N_{rac} are the numbers of systems that formed enantiopure and racemic crystals with CQ $\geq$ 0.4, respectively, and N_{cryst} is the total number of systems with CQ $\geq$ 0.4.

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

the same solvent or, alternatively, as identical molecules in different solvent environments.

One potentially important caveat of our model is its focus on attractive interactions of different strength. Longer-ranged interactions, in particular *repulsive* electrostatic interactions between functional groups of the same charge, are not explicitly considered by our model. However, we have performed a limited number of simulations with a modified version of the model that mimics such repulsive interactions. As discussed in detail in section 1.6 of the SI, these charge-like interactions result in crystallization behavior that is consistent with the simple attractive model used throughout this work.

B. Molecular Dynamics of Crystal Formation. We determine the crystallization behavior of a given molecule with molecular dynamics computer simulations in the NVT ensemble at a molecular concentration of 0.04 σ^{-2} (see Methods). Simulations are initiated from dispersed racemic mixtures of 5184 molecules and run for up to 4×10^8 time steps. To facilitate the formation of well-ordered solids, the temperature is adjusted during the simulation by an algorithm that operates in three stages, as illustrated in Figure 1e (also see SI, section 1.1). In the first stage, starting from a temperature that is too high for structure formation, temperature is rapidly decreased until an aggregate containing at least 50 molecules is observed. In the second stage, temperature is automatically adjusted to achieve a growth rate of aggregates of approximately five molecules per 10^6 time steps, until the largest aggregate has reached a size of at least 200 molecules. In the third and final stage, temperature is slowly increased at a rate of $0.2\epsilon/k_B$ per 10^8 time steps, scanning a range of supersaturation values. Throughout the third stage, polymorph-specific order parameters (see Methods) are used to identify the molecular clusters with the highest crystalline quality (CQ) for both enantiopure and racemic structures. CQ is measured on a scale from zero (no discernible order) to one (a perfect bulk crystal); large crystalline aggregates with few defects typically obtain a CQ score of 0.6 or larger, but good crystalline order is evident in all aggregates with CQ $\gtrsim$ 0.4. Snapshots from a typical simulation trajectory that yielded crystalline aggregates are shown in Figure 1d. Using the temperature protocol described above, crystallization is dominated by growth rather than nucleation: In the first and second stages, a substantial number of aggregates rapidly form and grow, typically displaying varying numbers of defects, as illustrated in Figure 1e. In the third stage, as temperature is increased, these defects anneal and well-ordered aggregates can further grow, while disordered clusters fall apart. We find that this simulation protocol yields superior crystals compared to straightforward MD simulations at constant temperature and that independent simulation runs yield consistent values of CQ (see Table S1). We have furthermore confirmed that CQ values are insensitive to a 4-fold reduction of molecular concentration (see Table S2). Similar annealing procedures have been found to optimize self-assembly outcomes of different systems.[47,48]

We compiled a large data set of crystallization trajectories for several different interaction strengths of functional groups as well as for racemic and enantiopure solutions. To give an overview of the various crystallization outcomes, we first discuss results for the family of racemic mixtures with $\epsilon_s = 5\epsilon$, which approximates the relative strength of a strong hydrogen bond with respect to weaker van der Waals interactions.[49] For these interactions, we identify the most crystalline aggregates at temperatures between 0.80 and 1.20 ϵ/k_B, which implies effective interactions of $\epsilon \approx 0.6$ kcal/mol and $\epsilon_s \approx 3.0$ kcal/mol at room temperature. Our model molecules display the same types of crystallization outcomes found in experiments.[8] Specifically, we observe the formation of racemic crystals, enantiopure crystals, and aggregates with various degrees of disorder. Remarkably, all crystals we observe are either racemic or enantiopure; crystals with other mole fractions of enantiomers do not form in our simulations. This result appears to be consistent with experiments, as we were not able to find reports of such crystals in the literature. Examples of the different crystallization outcomes are displayed in Figure 2; results for all molecules with $\epsilon_s = 5\epsilon$ are schematically summarized in Figure 3c. Consistent with experiments,[15] only a minority of molecules (27%) crystallize well, which we define by a CQ score of larger than 0.40. Among these good crystallizers, almost all form a single racemic or enantiopure polymorph, as illustrated in Figure 2a. In one case (s5/2:5/5), we observe the concurrent formation of two polymorphs with high CQ scores: one racemic, the other enantiopure. In experiments, a prevalence of racemic crystals (approximately 9:1) is commonly cited in the literature.[8,17,19] In our 2D simulations, we observe approximately equal numbers of racemic and enantiopure crystals, consistent with experimental accounts of quasi-2D racemic mixtures adsorbed on surfaces.[50] The majority of molecules form aggregates with limited order or no order at all, including racemic amorphous solids (Figure 2b) and aggregates of enantiopure or racemic micelles (Figure 2c).[51] We also observe a large number of racemic disordered aggregates (Figure 2d) that consist of enantiopure linear arrangements stacked on one another without apparent enantiomeric preference. These randomly stacked and partially ordered solids with varying composition of enantiomers, called "1D-racemic", are frequently observed in experiments of quasi-2D systems on surfaces,[52] but also occur in 3D.[53−56] In addition, we find that a few systems form disordered structures that are closely related to solid solutions, which we will refer to as "0D-racemic" (Figure 2e). These structures are composed of enantiopure oligomers that are arranged with spatial but no compositional order. Since both 1D-racemic and 0D-racemic solids do not have well-defined unit cells, assignment of CQ scores (which involves matching of aggregates to a known catalogue of crystal structures, see Methods) to aggregates of this type is ambiguous. We therefore remove them from our analysis and indicate molecules that show this behavior by black-colored cells in Figure 3a−d.

In the following, we first discuss several trends that are apparent from the results of our MD simulations. We then show that these trends can be rationalized and predicted based on the underlying thermodynamic landscape of crystal polymorphs.

Molecular Shape Is Not a Strong Predictor of Crystallization Behavior. In a recent study, Pillong and co-workers characterized the crystallization propensity of 319 organic molecules in different solvents.[15] While an analysis of this data set clearly showed that smaller molecules tend to crystallize better, no strong connections between crystallization and molecular fingerprints sensitive to molecular shape could be identified.[57] The results of our MD simulations of molecules with $\epsilon_s = 5\epsilon$ (Figure 3c) are consistent with these experimental results. We do not find a strong correlation between molecular shape and the quality or composition of crystals. In fact, most molecular shapes can form both racemic and enantiopure

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

crystals for different pairs of strongly interacting functional groups. While some shapes clearly produce better crystals than others (e.g., s2 vs s1), our attempts to find meaningful correlations between CQ scores and simple molecular descriptors of shape, degree of chirality, and position of strongly interacting functional groups on the molecule were unsuccessful. However, we do observe a clear tendency of molecules with particularly elongated shapes (i.e., shapes 9, 10, and 11) to form 1D-racemic aggregates. A similarly reduced crystallization propensity might be expected in experiments for molecules with elongated or planar shapes.

Crystallization and Chiral Separation Are Enhanced When Interactions Are Heterogeneous. We find that crystallization behavior depends markedly on the relative strengths of weak and strong interactions. In addition to the case of $\epsilon_s = 5\epsilon$ discussed above, we have performed systematic MD simulations for smaller ($\epsilon_s = 3\epsilon$, 2ϵ, and 1ϵ) and larger ($\epsilon_s = 7\epsilon$) interaction strengths, as illustrated in Figure 3a−d. With decreasing ϵ_s, the number of good crystallizers decreases sharply (29, 14, and 3 for $\epsilon_s = 5\epsilon$, 3ϵ, and 2ϵ, respectively). For completely uniform interactions, $\epsilon_s = \epsilon$, none of the 11 molecular shapes showed any crystalline tendencies (see Figure S1). Increasing strong interactions from $\epsilon_s = 5\epsilon$ to $\epsilon_s = 7\epsilon$ affects racemic and enantiopure crystals in different ways: while the number of racemic crystals decreases noticeably, the number of enantiopure crystals remains large, resulting in a increase of the fraction of enantiopure crystals to 73% (compared to 50% and 59% for $\epsilon_s = 3\epsilon$ and 5ϵ, respectively). These results are summarized in the table shown in Figure 3f. While we do not have systematic data for even larger values of ϵ_s, preliminary simulations indicate that such interactions favor enantiopure crystals with low densities (see Figure S3).

Enantiopure Solutions Crystallize Markedly Better than Racemic Ones. In addition to the sensitivity to interaction strength discussed above, we find that crystallization behavior also depends strongly on the composition of the solution. As illustrated in Figure 3e, the crystallization propensity and also the quality and size of crystals increases substantially when enantiopure solutions are simulated (at the same molecular concentration and using the same crystallization protocol). Specifically, for $\epsilon_s = 5\epsilon$, 70% of all systems form high-quality crystals, compared to 27% for the racemic solutions. We find the same qualitative dependence on interaction strength for the case of enantiopure solutions, but with overall higher crystallization propensity: For uniform interactions, $\epsilon_s = \epsilon$, 3 out of 11 models crystallize well (27%, see Figure S2), a lower yield than for $\epsilon_s = 5\epsilon$ (70%), but higher than the racemic case, where no crystals formed at all (Figure 3f and Figure S1). Comparing the specific crystal structures obtained from racemic and enantiopure solutions, we find that molecules that form conglomerates from racemic solutions crystallize in the same structures from enantiopure solutions, with comparable or better CQ values. In addition, several molecules that did not produce good crystals from racemic mixtures crystallize well from enantiopure initial conditions. Furthermore, 0D-racemic and 1D-racemic disordered aggregates, which we observe frequently in the racemic case (black cells in Figure 3a−d), cannot form in enantiopure systems. In fact, we find that these molecules tend to form high-quality crystals from enantiopure solutions. Crystalline quality does not, however, improve in all cases; molecules that formed good racemic crystals from racemic solutions do not necessarily form high-quality enantiopure structures (e.g., s5/1:3, s9/3:5, and s11/3:5).

C. Crystal Energy Landscapes. Are the crystals that form in our simulations the lowest (free) energy polymorphs? Are there clear thermodynamic signatures of molecules that form disordered solids? To reveal the thermodynamic landscapes underlying the crystallization of our model molecules, we have enumerated large numbers of polymorphs and calculated their lattice energies. Predicting accurate low-energy crystal structures for real molecules requires computationally expensive methods and significant expertise.[27−33,58,59] In the case of our coarse-grained molecules, low-energy polymorphs can be found more straightforwardly. To this end, we adapted a computationally efficient algorithm for solving the exact cover problem[60,61] to generate dense packings of our model molecules (see Methods). In the following, we will refer to this polymorph enumeration algorithm as POLYNUM. The algorithm allows us to generate between 10^5 and 10^7 polymorphs for each molecule, with unit cells containing up to 14 molecules, with all possible molecular compositions ranging from enantiopure to racemic, and with different packing fractions. We find that the number of polymorphs generally increases rapidly with increasing unit cell size and with decreasing packing fraction, preventing us from identifying potential low-energy polymorphs with very large unit cells or low packing fractions. However, an analysis of the generated crystal polymorphs suggests that we identify the vast majority of crystal polymorphs that have energies low enough to be dynamically relevant (see SI, Section 2.7). In particular, all crystal structures that formed spontaneously in our simulations were also found by POLYNUM (Figure S12). Note that throughout this paper we use zero-temperature crystal energies as estimates of free energies at the temperature of crystal formation. To justify this approximation, we have calculated free energy differences between a small number of polymorphs (see SI, Section 4.2). Because of the short-ranged interactions between functional groups, entropy differences between polymorphs are small, on the order of 0.1ϵ per molecule. We show below that differences of this magnitude do not alter the crystallization behavior of molecules in our simulations substantially.

There Are Many More Racemic Polymorphs than Enantiopure Ones, and Most Polymorphs Have Other Molecular Compositions. We find that the number of low-energy polymorphs varies greatly between molecular shapes, but there are clear general trends. The numbers of polymorphs with different composition loosely follows binomial statistics (see Figure S18), with enantiopure polymorphs forming the smallest group and racemic polymorphs the largest. Over the entire set of polymorphs calculated for all shapes, we find 8 times more racemic polymorphs than enantiopure ones. However, the group of polymorphs with other compositions of enantiomers is 5 times larger still. How are the energies of polymorphs distributed? In Figure 4a,b, we show histograms of the energies of racemic and enantiopure polymorphs for two molecules with the same shape (s2) but different interactions (5:5 and 4:5). Although the different numbers of racemic and enantiopure polymorphs are apparent, both distributions have similar shape, showing an initial exponential increase with increasing energy. Both distributions show a maximum, followed by rapid decline. The positions of these maxima reflect the limitations of POLYNUM in enumerating crystals with very large unit cells and low packing fraction. It is

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

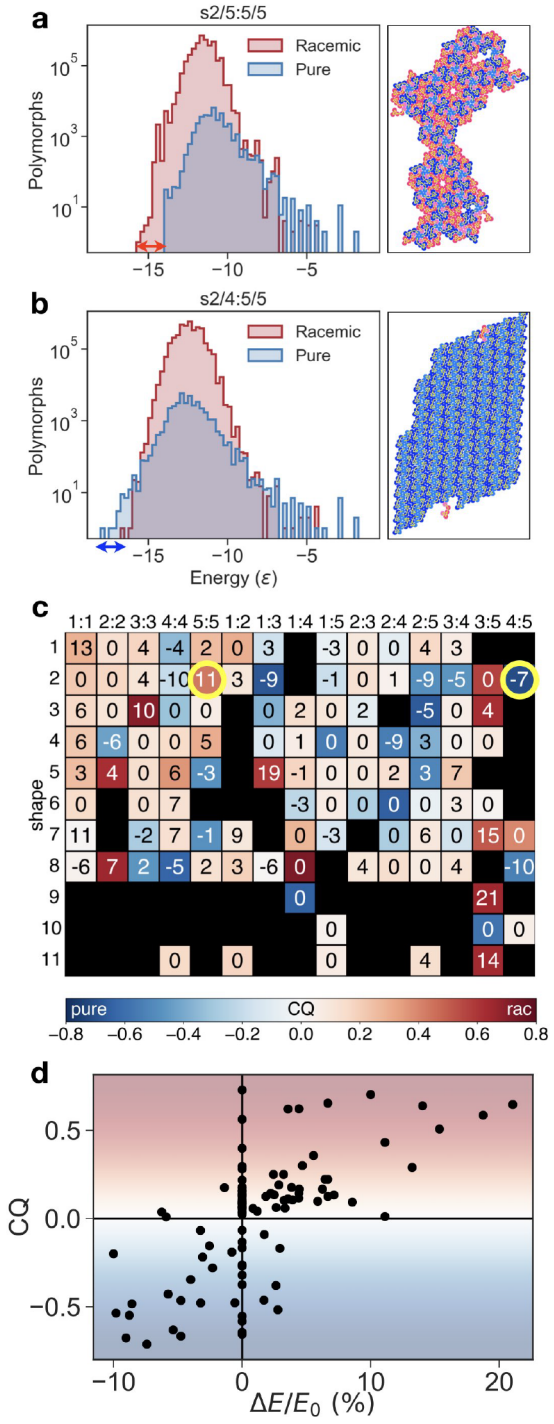


Figure 4. Thermodynamics of chiral crystallization. Histograms of energies per molecule of racemic (red) and enantiopure (blue) polymorphs for (a) s2/5:5/5 and (b) s2/4:5/5. Energy gaps ΔE between lowest energy enantiopure and racemic polymorphs are indicated with a red and blue arrow, respectively. (c) Grid summary of CQ scores for $\epsilon_s = 5\epsilon$. (Compare panel (c) of Figure 3.) Numbers indicate energy gaps ΔE. Systems in panels (a) and (b) are highlighted with yellow circles. (d) Scatter plot of CQ scores and energy gaps ΔE for molecules with $\epsilon_s = 5\epsilon$.

plausible that the true distributions should increase up to energies close to zero because many more periodic structures exist with larger numbers of molecules in the unit cell and larger available volume per molecule. The set of polymorphs enumerated with POLYNUM is therefore incomplete beyond the energies at which the polymorph distributions peak. However, consistent with CSP studies,[62] our MD simulations show that most dynamically relevant polymorphs have energies within 5% of the lowest possible energy. The distributions of polymorphs found with POLYNUM peak at much higher energies. We therefore conclude that the set of polymorphs we analyze includes essentially all dynamically relevant crystal structures.

Large Crystal Energy Differences Predict Crystallization, but Many Molecules Crystallize Well without Strong Thermodynamic Bias. To quantify the thermodynamic bias toward racemic or enantiopure crystals, it is useful to define the energy gap $\Delta E = E_0^{(P)} - E_0^{(R)}$ as the difference between the lowest energies $E_0^{(P)}$ and $E_0^{(R)}$ of any racemic (R) and enantiopure (P) polymorph, respectively. Positive and negative values of ΔE indicate a thermodynamic advantage for racemic and enantiopure crystals, respectively. In Figure 4a,b, the energy gap is indicated by arrows, and it correlates well with crystallization outcomes for these two specific molecules. We find a positive energy gap for s2/5:5, and this molecule indeed forms a racemic crystal of high quality in our MD simulations. By contrast, a negative ΔE and chiral separation are observed for s2/4:5. In Figure 4c, we show values of ΔE, expressed in percent of the lowest crystal energy $E_0 = |\min(E_0^{(P)}, E_0^{(R)})|$ of all racemic and enantiopure crystals of the molecule, for all simulated racemic solutions with $\epsilon_s = 5\epsilon$ together with the CQ scores obtained in MD simulations; the same data are shown as a scatter plot in Figure 4d. As expected from thermodynamic principles and consistent with CSP studies, we observe a clear correlation between ΔE and crystallization outcomes. We find the largest energy gaps (up to 20%) for racemic crystals, which we attribute to the much larger number of racemic crystal structures. Systems with such extreme energy gaps crystallize accordingly. However, in the group of polymorphs with zero to moderately large energy gaps (<10%), which includes the majority of systems, we find a much weaker correlation between ΔE and crystallization outcomes. Many systems with comparably large energy gaps do not crystallize (e.g., s7/1:1, s8/1:3), and some systems form high-quality racemic or enantiopure crystals without any thermodynamic bias (ΔE = 0). For two system (s5/2:5 and s8/3:3) we even observe chiral separation at positive ΔE, that is, against thermodynamic bias. Kinetic effects clearly are important in these cases.

The Likelihood of Crystallization Is (Partly) Encoded in the Energetics of Polymorphs with Unusual Composition. The largest group of polymorphs identified by POLYNUM is crystals with compositions other than racemic or enantiopure, which we refer to as PFCs in the following, short for "polymorphs with fractional composition". The energy distribution of PFCs is illustrated for molecule s2/4:5/5 in Figure 5a. Despite the large number of PFCs, we did not observe the formation of such crystals in any of our simulations. One might therefore expect that PFCs generally have much higher energies and do not compete thermodynamically with racemic and enantiopure polymorphs. To the contrary, we find that in approximately 60% of all systems with $\epsilon_s = 5\epsilon$ the lowest energy crystal is a PFC. We quantify the thermodynamic role of these polymorphs by calculating the energy difference between the lowest energy PFC and the polymorph with the lowest energy among racemic and enantiopure polymorphs, $\Delta E^{(FC)} = E_0^{(FC)} - \min(E_0^{(P)}, E_0^{(R)})$.

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

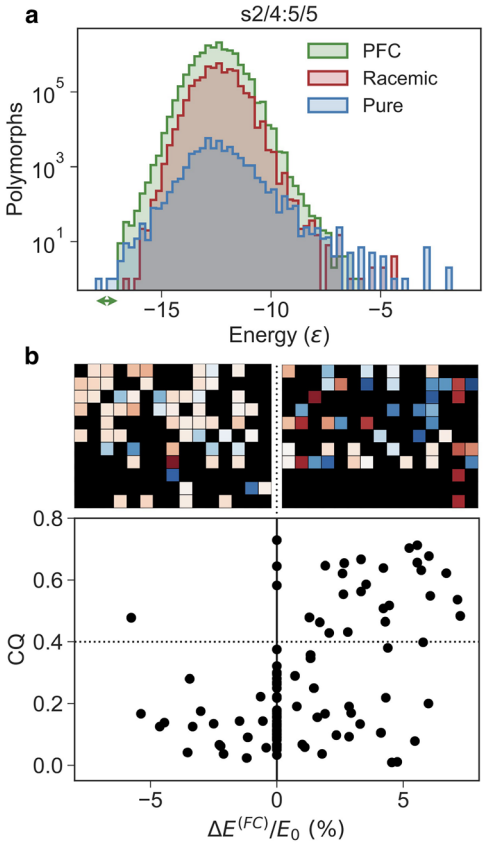


Figure 5. Polymorphs with fractional composition influence crystallization. (a) Histograms of energies per molecule of racemic polymorphs (red), enantiopure polymorphs (blue), and PFCs (green) for s2/4:5/5. The energy gap $\Delta E^{(FC)}$, defined in the text, is indicated with a green arrow. (b) Scatter plot of CQ values and $\Delta E^{(FC)}$ for all simulated systems with $\epsilon_s = 5\epsilon$. Systems with CQ $\gtrsim$ 0.4 (dashed line) are considered to be good crystallizers. CQ grids shown above the plot separately display systems with $\Delta E^{(FC)}/E_0 < 0$ (left) and $\Delta E^{(FC)}/E_0 > 0$ (right), respectively.

Remarkably, we find that $\Delta E^{(FC)}$ is a reasonable predictor for crystallization propensity in our simulations. As shown in Figure 5b, among the systems that have PFCs at the lowest energies ($\Delta E^{(FC)}/E_0 < 0$), very few form good crystals. Almost all good crystallizers are found among the systems that do not have competing PFCs ($\Delta E^{(FC)}/E_0 > 0$). This observation suggests a potentially useful strategy to assess the likelihood of a given molecule to crystallize, by using methods of CSP to evaluate the energies of crystals with fractional composition. Our simulations predict that if PFCs with low energy can be found, crystallization is likely complicated.

Polymorphs with Small Numbers of Competing Polymorphs Crystallize Best. Why do PFCs not crystallize despite their important thermodynamic role? We find that, across all molecules studied here, the likelihood to form a good crystal decreases with the number of thermodynamically competing polymorphs, which we define as the number of polymorphs $N_{0.95}$ found by POLYNUM that have energies within 95% of the lowest crystal energy. Figure 6 demonstrates a clear negative correlation between $N_{0.95}$ and CQ scores obtained in our simulations of racemic and enantiopure solutions. Since PFCs generally constitute the largest group of polymorphs, molecules with low-energy PFCs tend to have large $N_{0.95}$ and low crystallization probability. Furthermore, consistent with

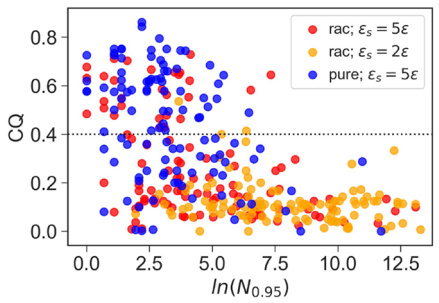


Figure 6. Crystal quality decreases with increasing number of competing polymorphs. A scatter plot of CQ scores, obtained in simulations of racemic and enantiopure solutions, and the number of competing polymorphs $N_{0.95}$. Molecules with CQ $\gtrsim$ 0.4 (dashed line) are considered to be good crystallizers.

crystallization trends illustrated in Figure 3, we find that $N_{0.95}$ depends on the heterogeneity of molecular interactions (ϵ_s) and on the composition of the solution. Compared to racemic solutions with $\epsilon_s = 5\epsilon$, molecules with energetically more uniform interactions ($\epsilon_s = 2\epsilon$) have larger numbers of competing polymorphs on average and crystallize worse (compare Figure 3a). Enantiopure solutions, which crystallize markedly better, have a significantly smaller number of competing polymorphs than racemic ones (namely 28 competing polymorphs on average, compared to 96 for racemic solutions). These trends can be rationalized by considering the geometric features of the space of crystal polymorphs.

D. Polymorph Distributions. In our model, only functional groups that are in immediate contact with one another have meaningful interactions. These contact interactions suggest a simple geometric model of the space of polymorphs.[60] The energy per molecule of a given polymorph p can be written as

$$E_p = -\sum_{i\in\{1:1,1:2,\ldots\}} \epsilon_i c_{p,i}$$

where the sum goes over all types of interactions i (i.e., all pairs of functional groups), $c_{p,i}$ is the number of contacts per molecule of type i in polymorph p, and ϵ_i is the energy per molecule associated with contacts of type i. (In our simulations, $\epsilon_i = \epsilon/2$ for weakly interacting groups, and $\epsilon_i = \epsilon_s/2$ for strongly interacting groups.) We will interpret the numbers $c_{p,i}$ and ϵ_i as the components of the "contact vector" $\vec{c}_p$ of polymorph p and an "interaction vector" $\vec{\epsilon}$, respectively. The energy per molecule of p can then be written as the scalar product, $E_p = -\vec{\epsilon}\cdot\vec{c}_p$. The larger the scalar product of a polymorph's contact vector with $\vec{\epsilon}$, the lower its energy.

What contact vectors are possible? In a close-packed system of molecules on a lattice, all polymorphs have the same total number of contacts per molecule, n^*. (For our models with five functional groups, n^* varies between 9 and 11, depending on the shape.) This condition implies $\sum_i c_{p,i} = n^*$ for close-packed crystals. Geometrically, the contact vectors of close-packed polymorphs are points on a simplex in the space of the different types of contacts between functional groups. (For our models, this space is 15-dimensional.) Non-close-packed polymorphs have fewer total contacts and are thus represented by different simplices defined by $\sum_i c_{p,i} = n < n^*$. Figure 7a illustrates the geometry of this situation for a fictitious system of molecules with only two types of possible contacts, which

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

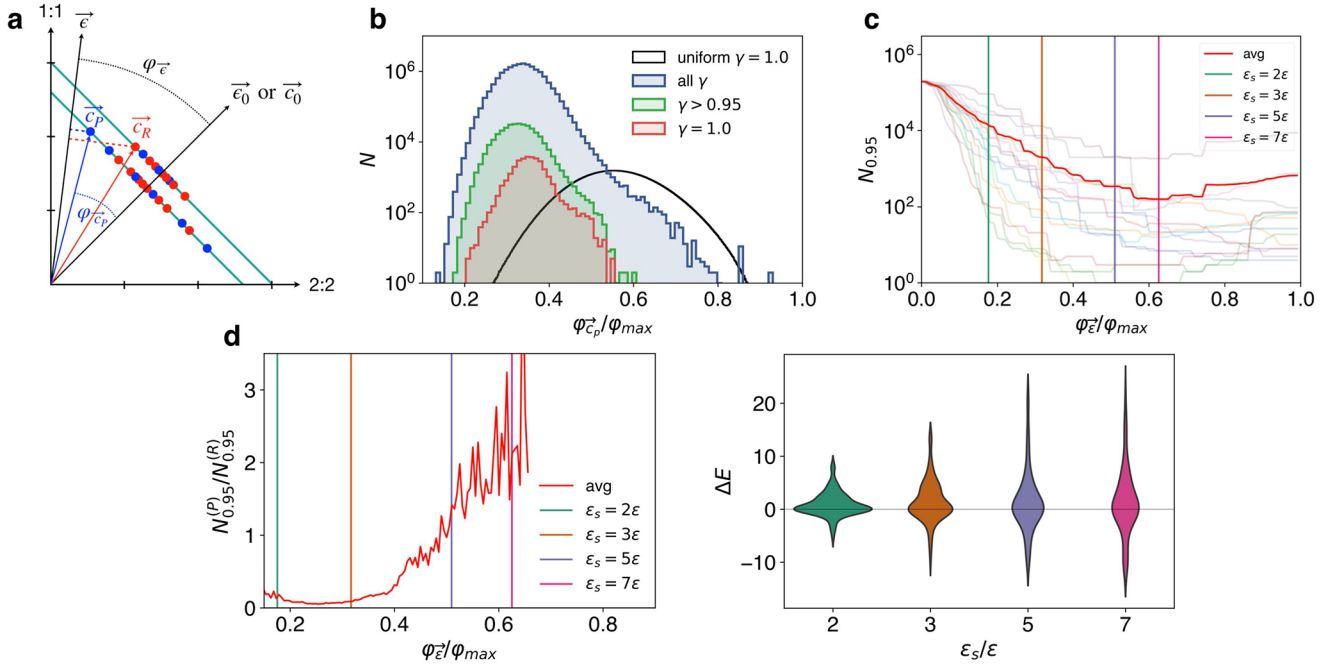


Figure 7. Analysis of the space of polymorphs. (a) Schematic illustration of polymorph space for a fictitious model with two types of contacts between functional groups (1:1 and 2:2). The enantiopure polymorph indicated by $\vec{c}_P$ has the lowest energy because its contact vector has the largest scalar product with $\vec{\epsilon}$, as indicated by dashed lines. (b) Number of polymorphs of shape s2 enumerated by POLYNUM, as a function of $\varphi_{\vec{c}_p}$, for different packing fractions γ: close-packed crystals (γ = 1, red curve), crystals with γ > 0.95 (green curve), and all crystals (blue curve). A uniform distribution of polymorphs on the simplex for γ = 1 is shown in black. (c) Number of competing polymorphs $N_{0.95}$ as a function of $\varphi_{\vec{\epsilon}}$, for shape s2. Separate curves are shown in light colors for different pairs of strongly interacting groups. The average over all interactions is shown as a bold red curve. Vertical lines indicate values of $\varphi_{\vec{\epsilon}}$ corresponding to different interaction strengths ϵ_s used in our MD simulations. (d) Ratio of competing enantiopure and racemic polymorphs as a function of $\varphi_{\vec{\epsilon}}$, averaged over all shapes. Vertical lines are as in panel (c). (e) Violin plot of distributions of energy gaps ΔE for different interaction strengths ϵ_s used in our MD simulations.

we call 1:1 and 2:2 for consistency. Contact vectors of racemic and enantiopure polymorphs are represented as blue and red dots, respectively, on two different simplices (i.e., lines in two dimensions) that represent groups of polymorphs with different packing fractions $\gamma = n/n^*$. Interactions between functional groups are represented by $\vec{\epsilon}$; the direction of $\vec{\epsilon}$ in Figure 7a indicates strong interactions of type 1:1 in this case. The energies of two particular polymorphs, represented by $\vec{c}_P$ and $\vec{c}_R$, are indicated in the figure by the projections of $\vec{c}_P$ and $\vec{c}_R$ onto $\vec{\epsilon}$. As evident from the figure, for a given packing fraction, the polymorph with the largest number of 1:1 contacts must have the lowest energy, as the interactions favor this type of contact.

The number of competing polymorphs depends on the choice of interactions $\vec{\epsilon}$. It is useful to introduce the uniform interaction vector $\vec{\epsilon}_0 = (\epsilon, \epsilon, ..., \epsilon)$, which assigns the same interaction strength to all pairs of functional groups, as illustrated in Figure 7a. A simple measure for the heterogeneity of a given interaction vector $\vec{\epsilon}$ is the angle $\varphi_{\vec{\epsilon}}$ formed between $\vec{\epsilon}$ and the uniform interaction vector $\vec{\epsilon}_0$. The larger $\varphi_{\vec{\epsilon}}$, the larger the discrepancy of interaction strengths of different functional groups. The specific interactions we used in our MD simulations, i.e., $\epsilon_s = \epsilon$, 2ϵ, 3ϵ, 5ϵ, and 7ϵ, correspond to angles $\varphi_{\vec{\epsilon}} = 0°$, 13°, 24°, 38°, and 47°. The largest possible value of $\varphi_{\vec{\epsilon}}$ ($\varphi_{max} \approx 75°$ in our 15-dimensional space) is reached when only a single pair of functional groups has attractive interactions. By the same token, we introduce a contact vector $\vec{c}_0 = (1, 1, ..., 1)$, which describes a fictitious polymorph that has equal numbers of contacts of all types. The angle $\varphi_{\vec{c}_p}$ between the contact vector of a given polymorph $\vec{c}_p$ and $\vec{c}_0$ then describes heterogeneities in the numbers of different types of contacts in p; polymorphs with large numbers of contacts between a few functional groups have large $\varphi_{\vec{c}_p}$.

What choice of interactions between functional groups $\vec{\epsilon}$ minimizes the number of competing polymorphs? It is clear from Figure 7a that uniform interactions (those with small $\varphi_{\vec{\epsilon}}$) result in large numbers of competing polymorphs. For completely uniform interactions $\vec{\epsilon}_0$, *all* polymorphs with the same packing fraction (i.e., all polymorphs on a given simplex) have the same energy. Indeed, our MD simulations of this case ($\epsilon_s = \epsilon$) show no crystallization at all. For more heterogeneous interactions, however, energy differences between polymorphs increase and the number of competing polymorphs decreases. This effect is amplified by the distribution of polymorphs on the simplex. We find that polymorphs are not distributed uniformly on the simplex, as illustrated in Figure 7b: Most polymorphs are found in the vicinity of the center of the simplex, at small values of $\varphi_{\vec{c}_p}$, implying that in most polymorphs different types of contacts appear with similar frequency. (Note that POLYNUM cannot enumerate polymorphs with the smallest values of $\varphi_{\vec{c}_p}$, as these polymorphs have very large unit cells.) Polymorphs with larger numbers of contacts between a few select functional groups (large values of $\varphi_{\vec{c}_p}$) are much rarer and, as evident from Figure 7b and suggested in Figure 7a, also tend to have lower packing fractions γ. Our observation of increased crystallization propensity with increasing interactions strength ϵ_s can

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

therefore be rationalized by the specific distribution of polymorphs. Strong interactions ϵ_s between a small number of functional groups select for the few polymorphs that have large numbers of contacts between these groups, endowing them with low energies and reducing the number of competing polymorphs. (We show in Section 6 of the Supporting Information that the interactions $\vec{\epsilon}$ that result in the largest energy gaps $|\Delta E|$ are mostly heterogeneous and that these interactions consistently produce high-quality crystals in MD simulations.) This point is illustrated in Figure 7c, which shows the number of competing polymorphs $N_{0.95}$ that result from different choices of interaction vectors $\vec{\epsilon}$, characterized by the angle $\varphi_{\vec{\epsilon}}$. $N_{0.95}$ decreases by several orders of magnitude as ϵ_s is increased from ϵ to 5ϵ, which is the interaction strength that results in the best CQ scores in our simulations. Interestingly, for many systems, increasing ϵ_s to even larger values does not lower the number of competing polymorphs significantly. In some case, we even observe an increase in $N_{0.95}$ for larger $\varphi_{\vec{\epsilon}}$, indicating that for these systems there exists an optimal interaction strength that results in the best crystals.

We find that the distributions of enantiopure and racemic crystals are remarkably different. To demonstrate these differences, we plot the fraction of the numbers of competing enantiopure crystals, $N_{0.95}^{(P)}$, and competing racemic crystals, $N_{0.95}^{(R)}$, as a function of $\varphi_{\vec{\epsilon}}$ in Figure 7d. The figure clearly shows that for uniform interactions (small $\varphi_{\vec{\epsilon}}$) most competing polymorphs are racemic. For more heterogeneous interactions, the number of competing polymorphs decreases rapidly (Figure 7c), but the fraction of enantiopure polymorphs *increases*. We find that around $\epsilon_s = 5\epsilon$, the number of competing enantiopure polymorphs becomes equal or even greater than the number of racemic ones, depending on shape and interactions. These trends are directly reflected in the thermodynamic preferences for enantiopure and racemic crystals, as quantified by the energy gap ΔE. As shown in Figure 7e, with increasing ϵ_s we observe broader distributions of ΔE, consistent with an overall decrease of the number of competing polymorphs. In addition, the relative number of large negative ΔE values increases, as enantiopure crystals form an increasing fraction of all competing polymorphs. These observations rationalize our MD results (Figure 3a–d), which show that the fraction of systems that form enantiopure crystals increases with increasing interaction strength ϵ_s.

The observation that strongly heterogeneous interactions among functional groups (i.e., large values of $\varphi_{\vec{\epsilon}}$) result in a larger fraction of enantiopure crystals at low energies indicates that enantiopure systems can form more polymorphs with large numbers of contacts between a small number of functional groups (i.e., polymorphs with large $\varphi_{\vec{\epsilon}_p}$). Why do enantiopure polymorphs have a thermodynamic advantage over racemic ones for large values of $\varphi_{\vec{\epsilon}}$, even though their total number is much smaller compared to racemic polymorphs? We hypothesize that the complementarity of chiral shapes, i.e., the ability of objects of the same handedness to form locally dense arrangements akin to a handshake, favors the formation of multiple contacts between functional groups of the same type. This argument is supported by an analysis of the oligomers (dimers, trimers, etc.) that form in the early stages of our MD simulations. Our choice of a single strong interaction ϵ_s selects for oligomers that have several of those interactions. We indeed find that the majority of oligomers observed in our simulations are enantiopure, as illustrated in Figure S4. However, the frequency of formation of enantiopure polymorphs in our simulations cannot be rationalized by the number of competing polymorphs alone. As shown in Figure 7d, at modest interaction strength $\epsilon_s = 3\epsilon$, the majority of competing polymorphs are racemic. Nevertheless, 50% of crystals that form in our simulations are enantiopure, suggesting that enantiopure crystals have an inherent kinetic advantage, which this study does not address.

III. SUMMARY AND CONCLUSIONS

In this study, we have shown with simple molecular models of chiral molecules that there is a clear correlation between the number of competing low-energy polymorphs and a molecule's likelihood to crystallize in simulations. Enantiopure solutions crystallize markedly better than racemic solutions because of a much smaller number of possible crystal structures. For racemic mixtures, we find that most competing polymorphs are neither racemic nor enantiopure, but PFCs, polymorphs with other ratios of left- and right-handed molecules in the unit cell. These polymorphs have so far been largely overlooked because they never crystallize in experiments (and also not in our simulations). However, we find that only those molecules tend to crystallize well that do *not* have large numbers of competing PFCs. Our results suggest that a reduction in the number of competing polymorphs generally leads to an increase in crystallization likelihood and, for racemic mixtures, an increased likelihood to spontaneously separate via formation of conglomerates.

Of course, care should be taken when generalizing results obtained with 2D models to bulk systems. As one clear difference, crystallization of bulk racemic solutions yields racemic crystals more frequently than enantiopure ones, whereas experiments on surfaces and our 2D simulations result in an approximately equal number of racemic and enantiopure crystals. We nevertheless expect that the general trends observed in our 2D systems should largely apply for bulk crystallization, too. The results of this study are consequences of two statistical properties of polymorphs with different compositions: (1) There are fewer enantiopure than racemic polymorphs and fewer racemic polymorphs than PFCs. (2) Strongly heterogeneous interactions reduce the number of competing low-energy polymorphs and favor enantiopure crystals. Both properties of polymorphs should apply in the case of bulk systems as well.

How does the fraction of good crystallizers observed in our MD simulations compare to experiments? On the one hand, experiments proceed on much longer time scales and can therefore observe crystallization at lower degrees of supersaturation, likely resulting in better crystals. On the other hand, the 2D nature and simplicity of our model likely facilitates crystallization. Interestingly, a recent experimental study of 319 molecules in 18 solvents found crystallization behavior similar to our simulations.[15] We conclude that while a quantitative comparison of crystallization propensities observed in our study and in bulk experiments is complicated, our simulations successfully discern between molecules that crystallize robustly and those that do not, consistent with available experimental data.

There are several parameters that determine the number of competing polymorphs and therefore crystallization yields. We have focused on the interaction strength between functional groups and demonstrated that the presence of a few strong interactions leads to markedly better crystallization outcomes

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

and a larger fraction of conglomerates because it selects for a small number of polymorphs, many of which are enantiopure. The higher propensity of chiral organic salts to form conglomerates compared to neutral versions of the same molecules[9−11] can thus be rationalized by the presence of strong electrostatic interactions. Our results can also provide some guidance for the chemical synthesis of molecules. Whenever chemical modification of a compound is an option, introducing strongly interacting groups should enhance the propensity of molecules to crystallize and separate.

Different numbers of competing polymorphs can help rationalize several other recent observations. A recent study by Gellman and co-workers[50] showed that compared to bulk solutions the formation of enantiopure crystals is enhanced when molecules are adsorbed on surfaces. We argue that the number of polymorphs accessible to such quasi two-dimensional systems is likely much smaller than in bulk solution. Another example is the crystallization of chiral molecules whose orientations are constrained, for instance by an external field. Such molecules crystallize and separate better, as recently demonstrated by Woszczyk and co-workers.[16] We show in Table S4 that the number of low-energy polymorphs decreases dramatically when molecular orientations are constrained. A third example is the use of pressure, as recently demonstrated in the crystallization of B_2O_3, a material that forms a glass from the melt at ambient conditions but readily crystallizes into a single polymorph at elevated pressures. A recent CSP study showed that this behavior can be rationalized by the presence of a large number of low-enthalpy polymorphs at low pressure.[36] These polymorphs share several structural motifs with the glass that forms at ambient conditions. At higher pressures, however, these polymorphs have much higher enthalpies and do not compete any longer.

While an analysis of the thermodynamic landscapes of crystals can provide useful guidance in a statistical sense, it cannot be used in a straightforward way to predict more specifically which polymorph will form. We have shown that the energy gap between racemic and enantiopure crystals correlates well with crystallization outcomes, but energy gaps *within* groups of polymorphs of the same composition are much less predictive. Despite general trends, several molecules with small numbers of competing polymorphs do not form high-quality crystals, as evident from the spread of the data in Figures 5b and 6. Our analysis confirms that while thermodynamics exert a clear bias on crystallization behavior, there are many cases where crystals form or do not form for kinetic reasons. While general trends can be inferred from thermodynamic landscapes, a thorough analysis of the mechanisms and rates of nucleation and growth of different polymorphs is needed to make specific predictions.[63,64] Work in this direction is underway.

IV. METHODS

1. Pair Potential. We use a short-ranged pair potential consisting of a repulsive and an attractive part,

$$u(r) = u_{\mathrm{rep}}(r) + u_{\mathrm{att}}(r)$$

The repulsive part is a WCA potential,[65]

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

The attractive part is given by

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

The potential and its first derivative are continuous at $r = \sigma$. In all MD simulations, we use $\epsilon_{\mathrm{rep}} = 5.0$, $\sigma = 1.0$, and $\omega = 0.2$. For pairs of weakly interacting functional groups we use $\epsilon_{\mathrm{att}} = \epsilon$, which defines our unit of energy. For pairs of strongly interacting groups we use $\epsilon_{\mathrm{att}} \equiv \epsilon_s = \epsilon$, 2ϵ, 3ϵ, 5ϵ, or 7ϵ, as specified in the main text. A plot of the pair potential is shown in Figure 1b.

2. Molecular Dynamics Simulations. We use the software package HOOMD in all our MD simulations.[66,67] We perform rigid-body molecular dynamics using a Langevin thermostat as implemented in HOOMD, with a time step of 0.004 and a damping coefficient of 5.0. (We use the mass m and diameter σ of functional groups as units of mass and length; the unit of energy is ϵ.) Simulation snapshots have been generated with OVITO.[68] Data analysis of our MD simulations utilized the freud Python library.[69]

3. Polymorph Enumeration (POLYNUM). We enumerate polymorphs of our models with a method inspired by algorithms used to solve a variant of the exact cover problem, which can be stated as follows: Given a box and a number of shapes, enumerate all distinct ways to completely fill the box with the shapes. For arbitrary shapes, this problem can be complicated. On a lattice, it can be solved with a fast algorithm called Dancing Links that exploits properties of linked lists.[61] Enumerating crystal polymorphs of our model amounts to solving the exact cover problem on a hexagonal lattice with periodic boundary conditions, as illustrated in Figure S8. Because of the simple geometry of our molecules and the short-ranged nature of interactions, functional groups reside near the vertices of a simple hexagonal lattice in all crystal structures and larger aggregates observed in our simulations. The underlying lattice greatly reduces the search space for crystal polymorphs and provides the connection to the exact cover problem. As evident from Figure 2, not all crystals are close-packed; many have one or more "holes" (i.e., unoccupied lattice sites) in the unit cells. The exact cover algorithm can still be applied by treating holes as distinct "molecules" that occupy a single lattice site. To enumerate polymorphs, we proceed as follows:

1. On a 2D hexagonal lattice, enumerate all unit cells that contain a minimum of 5 and a maximum of M lattice sites. Eliminate geometrically equivalent unit cells and unit cells that are chiral enantiomers.
2. For each unit cell, apply the Dancing Links algorithm to generate all packings of molecules and holes, considering periodic boundary conditions.
3. Eliminate equivalent structures from the set of solutions, based on contacts between functional groups.

The algorithm returns the complete set of unique crystal structures, up to the specified maximum number of lattice cells M. The number of solutions increases rapidly with unit cell size; for our models, we find that we can calculate polymorphs for unit cells containing up to 14 molecules and up to 4 holes ($M = 74$). Our analysis suggests that this limit is sufficient to identify the vast majority of low-energy polymorphs. The energy of crystal polymorphs increases systematically with increasing number of "holes" in the unit cell, as illustrated in Figure S14. Accordingly, the algorithm was able to identify all crystal polymorphs observed in our MD simulations. A detailed description of the algorithm is given in the SI. Note that an efficient method to enumerate 2D packings based on crystal symmetry has been implemented by Hudson and Harrowell.[70]

4. Polymorph Identification and CQ Scores. To determine the CQ score of a molecular cluster C, we first find the polymorph P^* that best matches the structure of the cluster and then compare the local environment of each molecule in C to the environment of molecules in the unit cell of P^*.

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

To find P^* for a given cluster C, we consider as candidates the 100 polymorphs that have the lowest energies, as determined by POLYNUM. For each polymorph P in this set, we determine a score that quantifies its similarity with the cluster, $S_{C,P} = 0.2O_{C,P} + 0.2R_{C,P} + 0.6E_P/E_0$. The similarity score $S_{C,P}$ is determined by three order parameters: a parameter based on radial distribution functions (RDF), $R_{C,P}$, a parameter based on molecular orientation, $O_{C,P}$, and the potential energy per molecule of the polymorph, E_P, divided by the lowest polymorph energy E_0 found by POLYNUM. The potential energy is included in $S_{C,P}$ to ensure that a low-energy polymorph is selected for cases where several structurally similar polymorphs exist. The polymorph with the highest score $S_{C,P}$ is considered the best-matching polymorph for cluster C.

To determine the order parameters $R_{C,P}$, we calculate RDFs $g_i(r)$ for each of the five functional groups i (only considering functional groups of the same type) within a cutoff distance of 6σ. RDFs are averaged over all molecules of the cluster C, giving a set of five RDFs $g_{i,C}(r)$, as well as for a periodically replicated unit cell of a given polymorph P, resulting in an analogous set $g_{i,P}(r)$. We account for thermal expansion and noise present in cluster configurations by calculating $g_{i,P}(r)$ from artificially thermalized bulk crystals, obtained by randomizing positions and orientations of all molecules of polymorph P by small displacements drawn from distributions estimated in MD simulation of a few different bulk crystals close to the temperature of crystal formation. The order parameter $R_{C,P}$, which quantifies structural similarity of cluster C and polymorph P, is then determined from the Pearson correlation coefficients of the RDFs,

$$R_{C,P} = \sum_{i=1}^{5} \langle g_{i,C}(r) g_{i,P}(r)\rangle - \langle g_{i,C}(r)\rangle\langle g_{i,P}(r)\rangle \times [(\langle g_{i,C}^2(r)\rangle - \langle g_{i,C}(r)\rangle^2)(\langle g_{i,P}^2(r)\rangle - \langle g_{i,P}(r)\rangle^2)]^{-1/2}$$

Here, $\langle\cdots\rangle = \frac{1}{6\sigma}\int_0^{6\sigma} \mathrm{d}r \cdots$ implies integration over the full range of the RDFs.

To determine the orientational parameter $O_{C,P}$, we characterize molecular environments by calculating the vectors that connect the center of mass (COM) of a given molecule to the COM of each of the six nearest neighbor molecules. (Nearest neighbors are determined based on COM distance.) These vectors split the unit circle into six sectors. The central angles of these sectors define the components of a vector θ. We find the lexicographically minimal rotation[71] of θ to ensure its invariance with respect to global rotations of the coordinate system. θ can be straightforwardly calculated for all molecules in cluster C. In polymorph unit cells with regular lattice positions, however, several molecules can share the same COM distance from a given central molecule, making the selection of six nearest neighbors ambiguous. For each molecule in a unit cell, we therefore create 10 different vectors θ_m ($m = 1, \ldots, 10$), each independently randomized to approximate effects of thermal noise, as discussed in the last paragraph. Given the vectors $\theta_{j,C}$ for all molecules j in cluster C and the vectors $\theta_{m,k,P}$ for all molecules k in the unit cell of polymorph P, we determine for each molecule j a score that quantifies the similarity of its molecular environment with molecules in P by calculating $s = \min_{m,k}\left(\sum_{l=1}^{6} |\theta_{m,k,P}^{(l)} - \theta_{j,C}^{(l)}|\right)$, where l runs over the six components of θ. If $s < 4°$, molecule j is categorized as "crystalline". The orientational order parameter is defined as the fraction of crystalline molecules in the cluster, $O_{C,P} = N_{\mathrm{cryst}}/N$.

Finally, once the polymorph P^* that best matches the structure of cluster C has been determined by finding the lowest score $S_{C,P} = 0.2O_{C,P} + 0.2R_{C,P} + 0.6\ E_P/E_0$ over the set of 100 low-energy polymorphs P, the CQ score of C is simply defined as the orientational order parameter, $\mathrm{CQ} = O_{C,P^*}$. Note that the CQ score is sensitive to the size of the cluster and to the presence of defects within the cluster, as molecules at the surface or near defects will be categorized as noncrystalline. We have tuned the weights that appear in the scoring function $S_{C,P}$ to optimize the performance of the polymorph identification scheme. Examples of molecular clusters, assigned polymorphs, and associated CQ values are given in Figure S15.

Since only the 100 polymorphs with the lowest energies are considered in the calculation of CQ scores, crystalline clusters with structures that are not included in this set might be assigned low CQ scores. We have therefore visually inspected every simulation trajectory to make sure that clusters with low CQ scores are, in fact, poorly crystalline and that no crystalline cluster went undetected.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.0c02097.

Additional methods, simulation results, and data analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michael Grünwald − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States;* orcid.org/0000-0003-2186-1662; Email: michael.gruenwald@utah.edu

Author

John E. Carpenter − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.0c02097

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Julio Facelli, Christoph Dellago, and Wim Noorduin for useful discussions. J.C. thanks Merlin Carpenter for his assistance with the implementation of POLYNUM. The support and resources of the Center for High Performance Computing at the University of Utah are gratefully acknowledged. This work was supported by the National Science Foundation under Grant No. CHE-1900626.

■ REFERENCES

(1) Bernstein, J. *Polymorphism in Molecular Crystals (Google eBook)*; IUCr Monographs on Crystallography; Clarendon Press, 2002; p 428.

(2) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and fictions about polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619−8635.

(3) Lee, A. Y.; Erdemir, D.; Myerson, A. S. Crystal Polymorphism in Chemical Process Development. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 259−280.

(4) Calcaterra, A.; D'Acquarica, I. The market of chiral drugs: Chiral switches versus de novo enantiomerically pure compounds. *J. Pharm. Biomed. Anal.* **2018**, *147*, 323−340.

(5) Nguyen, L. A.; He, H.; Pham-Huy, C. Chiral drugs: an overview. *Int. J. Biomed. Sci.: IJBS* **2006**, *2*, 85−100.

(6) Smith, S. W. Chiral toxicology: It's the same thing only different. *Toxicol. Sci.* **2009**, *110*, 4−30.

(7) Lorenz, H.; Seidel-Morgenstern, A. Processes to separate enantiomers. *Angew. Chem., Int. Ed.* **2014**, *53*, 1218−1250.

(8) Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolutions*; John Wiley and Sons, 1981.

(9) Jacques, J.; Leclercq, M.; Brienne, M. J. La formation de sels augmente-t-elle la fréquence des dédoublements spontanés? *Tetrahedron* **1981**, *37*, 1727−1733.

(10) Kinbara, K.; Hashimoto, Y.; Sukegawa, M.; Nohira, H.; Saigo, K. Crystal structures of the salts of chiral primary amines with achiral carboxylic acids: Recognition of the commonly-occurring supra-

molecular assemblies of hydrogen-bond networks and their role in the formation of conglomerates. *J. Am. Chem. Soc.* **1996**, *118*, 3441−3449.
(11) Dey, A.; Pidcock, E. The relevance of chirality in space group analysis: A database study of common hydrogen-bonding motifs and their symmetry preferences. *CrystEngComm* **2008**, *10*, 1258−1264.
(12) D'Oria, E.; Karamertzanis, P. G.; Price, S. L. Spontaneous resolution of enantiomers by crystallization: Insights from computed crystal energy landscapes. *Cryst. Growth Des.* **2010**, *10*, 1749−1756.
(13) Price, S. L. Why don't we find more polymorphs? *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2013**, *69*, 313−328.
(14) Wicker, J. G.; Cooper, R. I. Will it crystallise? Predicting crystallinity of molecular materials. *CrystEngComm* **2015**, *17*, 1927−1934.
(15) Pillong, M.; Marx, C.; Piechon, P.; Wicker, J. G.; Cooper, R. I.; Wagner, T. A publicly available crystallisation data set and its application in machine learning. *CrystEngComm* **2017**, *19*, 3737−3745.
(16) Woszczyk, A.; Szabelski, P. Creation of chiral adsorbed structures using external inputs: results from lattice Monte Carlo simulations. *Adsorption* **2016**, *22*, 553−560.
(17) Otero-De-La-Roza, A.; Hein, J. E.; Johnson, E. R. Reevaluating the Stability and Prevalence of Conglomerates: Implications for Preferential Crystallization. *Cryst. Growth Des.* **2016**, *16*, 6055−6059.
(18) Gavezzotti, A.; Rizzato, S. Are racemic crystals favored over homochiral crystals by higher stability or by kinetics? Insights from comparative studies of crystalline stereoisomers. *J. Org. Chem.* **2014**, *79*, 4809−4816.
(19) Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. On the Validity of Wallach's Rule: On the Density and Stability of Racemic Crystals Compared with Their Chiral Counterparts. *J. Am. Chem. Soc.* **1991**, *113*, 9811−9820.
(20) Paci, I.; Szleifer, I.; Ratner, M. A. Chiral separation: Mechanism modeling in two-dimensional systems. *J. Am. Chem. Soc.* **2007**, *129*, 3545−3555.
(21) Latinwo, F.; Stillinger, F. H.; Debenedetti, P. G. Molecular model for chirality phenomena. *J. Chem. Phys. 145* (**2016**), 154503.
(22) Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *J. Am. Chem. Soc.* **2015**, *137*, 11095−11104.
(23) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An extraordinary example of conformational polymorphism. *Pharm. Res.* **2001**, *18*, 859−866.
(24) Llinàs, A.; Goodman, J. M. Polymorph control: past, present and future. *Drug Discovery Today* **2008**, *13*, 198−210.
(25) 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.
(26) Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Acc. Chem. Res.* **2009**, *42*, 117−126.
(27) Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, Á.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246−2264.
(28) Nyman, J.; Reutzel-Edens, S. M. Crystal structure prediction is changing from basic science to applied technology. *Faraday Discuss.* **2018**, *211*, 459−476.
(29) Kim, S.; Orendt, A. M.; Ferraro, M. B.; Facelli, J. C. Crystal structure prediction of flexible molecules using parallel genetic algorithms with a standard force field. *J. Comput. Chem.* **2009**, *30*, 1973−1985.
(30) Oganov, A. R.; Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J. Chem. Phys.* **2006**, *124* , 244704.
(31) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 1−8.
(32) Lund, A. M.; Pagola, G. I.; Orendt, A. M.; Ferraro, M. B.; Facelli, J. C. Crystal structure prediction from first principles: The crystal structures of glycine. *Chem. Phys. Lett.* **2015**, *626*, 20−24.
(33) Avery, P.; Toher, C.; Curtarolo, S.; Zurek, E. XTALOPT Version r12: An open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.* **2019**, *237*, 274−275.
(34) Goodman, C. H. Strained mixed-cluster model for glass structure. *Nature* **1975**, *257*, 370−372.
(35) Greer, A. Confusion by design. *Nature* **1993**, *366*, 303−304.
(36) Ferlat, G.; Seitsonen, A. P.; Lazzeri, M.; Mauri, F. Hidden polymorphs drive vitrification in B 2 O 3. *Nat. Mater.* **2012**, *11*, 925−929.
(37) Wang, R.; Merz, M. D. Non-crystallinity and polymorphism in elemental solids. *Nature* **1976**, *260*, 35−36.
(38) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(39) Wales, D. J. Decoding the energy landscape: Extracting structure, dynamics and thermodynamics. *Philos. Trans. R. Soc., A* **2012**, *370*, 2877−2899.
(40) Hormoz, S.; Brenner, M. P. Design principles for self-assembly with short-range interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5193−5198.
(41) Teich, E. G.; van Anders, G.; Glotzer, S. C. Identity crisis in alchemical space drives the entropic colloidal glass transition. *Nat. Commun.* **2019**, *10*, DOI: 10.1038/s41467-018-07977-2.
(42) Damasceno, P. F.; Karas, A. S.; Schultz, B. A.; Engel, M.; Glotzer, S. C. Controlling Chirality of Entropic Crystals. *Phys. Rev. Lett.* **2015**, *115*, 1−5.
(43) Gantapara, A. P.; Qi, W.; Dijkstra, M. A novel chiral phase of achiral hard triangles and an entropy-driven demixing of enantiomers. *Soft Matter* **2015**, *11*, 8684−8691.
(44) Carmichael, S. P.; Shell, M. S. A simple mechanism for emergent chirality in achiral hard particle assembly. *J. Chem. Phys.***2013**, *139*, 164705.
(45) Lombardo, T. G.; Stillinger, F. H.; Debenedetti, P. G. Thermodynamic mechanism for solution phase chiral amplification via a lattice model. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15131−15135.
(46) Note that because of the symmetry of s9, this shape gives rise to only 9, not 15, unique molecules.
(47) Klotsa, D.; Jack, R. L. Controlling crystal self-assembly using a real-time feedback scheme. *J. Chem. Phys.* **2013**, *138*, No. 094502.
(48) Fullerton, C. J.; Jack, R. L. Optimising self-assembly through time-dependent interactions. *J. Chem. Phys.***2016**, *145*, 244505.
(49) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; Biochemistry; W. H. Freeman, 2002.
(50) Dutta, S.; Gellman, A. J. Enantiomer surface chemistry: Conglomerate: versus racemate formation on surfaces. *Chem. Soc. Rev.* **2017**, *46*, 7787−7839.
(51) "Aggregates of micelles" refers to disordered solids assembled from a population of clusters held together by strong interactions at the center but interacting weakly at the surface of the cluster. The "micelle" classification therefore describes a specific kinetic pathway that leads to an amorphous solid characterized by local, micelle-like motifs.
(52) Zhang, Y. Q.; Lin, T.; Cirera, B.; Hellwig, R.; Palma, C. A.; Chen, Z.; Ruben, M.; Barth, J. V.; Klappenberger, F. One-Dimensionally Disordered Chiral Sorting by Racemic Tiling in a Surface-Confined Supramolecular Assembly of Achiral Tectons. *Angew. Chem., Int. Ed.* **2017**, *56*, 7797−7802.
(53) Gervais, C.; Beilles, S.; Cardinaël, P.; Petit, S.; Coquerel, G. Oscillating crystallization in solution between (+)- and (−)-5-ethyl-5-methylhydantoin under the influence of stirring. *J. Phys. Chem. B* **2002**, *106*, 646−652.
(54) Green, B. S.; Knossow, M. Lamellar twinning explains the nearly racemic composition of chiral, single crystals of hexahelicene. *Science* **1981**, *214*, 795−797.

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768

(55) Torbeev, V. Y.; Lyssenko, K. A.; Kharybin, O. N.; Antipin, M. Y.; Kostyanovsky, R. G. Lamellar Racemic Twinning as an Obstacle for the Resolution of Enantiomers by Crystallization: The Case of Me(All)N + (CH 2 Ph)Ph X - (X = Br, I) Salts. *J. Phys. Chem. B* **2003**, *107*, 13523−13531.
(56) van Eupen, J. T. H.; Elffrink, W. W. J.; Keltjens, R.; Bennema, P.; de Gelder, R.; Smits, J. M. M.; van Eck, E. R. H.; Kentgens, A. P. M.; Deij, M. A.; Meekes, H.; Vlieg, E. Polymorphism and Migratory Chiral Resolution of the Free Base of Venlafaxine. A Remarkable Topotactical Solid State Transition from a Racemate to a Racemic Conglomerate. *Cryst. Growth Des.* **2008**, *8*, 71−79.
(57) Note that Pillong and co-workers also found that molecules with larger degrees of flexibility tend to crystallize worse.[15] Effects of flexibility cannot be probed with our simple rigid models.
(58) Cruz-Cabeza, A. J. Crystal structure prediction: Are we there yet? *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 437−438.
(59) Roy, S.; Matzger, A. J. Unmasking a third polymorph of a benchmark crystal-structureprediction compound. *Angew. Chem., Int. Ed.* **2009**, *48*, 8505−8508.
(60) Nicholls, J.; Alexander, G. P.; Quigley, D. Polyomino Models of Surface Supramolecular Assembly: Design Constraints and Structural Selectivity. *Condensed Matter* **2017**, arXiv:1702.01994.
(61) Knuth, D. E. Dancing links. *Millennial Perspectives in Computer Science: Proceedings of the 1999 OxfordMicrosoft Symposium in Honour of Sir Tony Hoare*; 2000; pp 187−214.
(62) 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.
(63) Coombes, D. S.; Catlow, C. R. A.; Gale, J. D.; Rohl, A. L.; Price, S. L. Calculation of attachment energies and relative volume growth rates as an aid to polymorph prediction. *Cryst. Growth Des.* **2005**, *5*, 879−885.
(64) Li, J.; Tilbury, C. J.; Joswiak, M. N.; Peters, B.; Doherty, M. F. Rate Expressions for Kink Attachment and Detachment During Crystal Growth. *Cryst. Growth Des.* **2016**, *16*, 3313−3322.
(65) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(66) Glaser, J.; Nguyen, T. D.; Anderson, J. A.; Lui, P.; Spiga, F.; Millan, J. A.; Morse, D. C.; Glotzer, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput. Phys. Commun.* **2015**, *192*, 97−107.
(67) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **2008**, *227*, 5342−5359.
(68) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.
(69) Ramasubramani, V.; Dice, B. D.; Harper, E. S.; Spellings, M. P.; Anderson, J. A.; Glotzer, S. C. *freud: A Software Suite for High Throughput Analysis of Particle Simulation Data*. 2019. arXiv:1906.06317
(70) Hudson, T. S.; Harrowell, P. Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. *J. Phys.: Condens. Matter* **2011**, *23*, 194103.
(71) Booth, K. S. Lexicographically least circular substrings. *Information Processing Letters* **1980**, *10*, 240−242.

https://dx.doi.org/10.1021/jacs.0c02097
J. Am. Chem. Soc. 2020, 142, 10755−10768