Check for updates

◆ ANNUAL ISSUE

EARLY CAREER SCHOLARS IN MATERIALS SCIENCE 2023

Organic crystal structure prediction and its application to materials design

Qiang Zhu[1,a] ⓘ, **Shinnosuke Hattori**[2]

[1] *Department of Physics and Astronomy, University of Nevada, Las Vegas, NV 89154, USA*
[2] *Advanced Research Laboratory, R&D Center, Sony Group Corporation, 4-14-1 Asahi-cho, Atsugi-shi 243-0014, Japan*
[a] Address all correspondence to this author. e-mail: qiang.zhu@unlv.edu

In recent years, substantial progress has been made in the modeling of organic solids. Computer simulation has been increasingly shaping the area of new organic materials by design. It is possible to discover new organic crystals by computational structure prediction, based on the combination of powerful exploratory algorithms and accurate energy modeling. In this review, we begin with several key early concepts in describing crystal packing, and then introduce the recent state-of-the-art computational techniques for organic crystal structure prediction. Perspectives on the remaining technical challenges, functional materials screening and software development are also discussed in the end. It is reasonable to expect that, in the near future, accurate predictive computational modeling can be accomplished within a time frame that is appreciably shorter than that needed for the laboratory synthesis and characterization.

Qiang Zhu is currently an associate professor in computational physics at the University of Nevada Las Vegas (UNLV). He received the Ph.D. in Mineral Physics (under the supervision of Prof. Artem R. Oganov) from Stony Brook University in 2013, as well as a B.S. in Materials Science and Engineering from Beihang University of China in 2007. In 2016, he joined UNLV as an assistant professor in the Department of Physics and Astronomy. At UNLV, he is leading a research group in computational physics and materials discovery. Over years, he has been working on the field of computational crystal structure prediction for both inorganic and organic materials. He was the chief developer of the USPEX code between 2012 and 2016. Since 2018, his group has launched several open-source projects in high-throughput materials simulation and machine learning. Together with experimental collaborators, he has investigated crystal polymorphism for a wide range of molecules. He has published over 70 papers in peer-reviewed journals. He is also the recipient of the President's award for distinguished Doctoral Students at Stony Brook University (2013), the NSF CAREER award (2021), the DOE CAREER award (2021), and the Barrick Scholar Award (2022).

Introduction

Molecular solids refer to substances consisting of discrete molecules that are held together by relatively weak intermolecular forces. For centuries, they have been extensively used as key components in chemical industries such as medicine [1], fertilizers, dyes [2], pesticides [3], and high-energy explosives [4]. In addition, some organic solids were found to exhibit

interesting physical properties. In 1921, the first ferroelectric crystal was found in organic Rochelle salt [5]. Organic solids have also received considerable interest in the electronics industry since the discovery of bulk conductivity in polycyclic aromatic compounds in the 1950s [6]. While some molecules can aggregate with no particular order, such as amorphous solids, most organic solids are crystalline, and their physical properties

largely depend on regularly repeating intermolecular packing. For example, the organic semiconductor material rubrene usually exhibits outstanding carrier mobility, but changes in its crystal packing can eliminate that mobility entirely [7]. From a materials perspective, the soft molecular nature of organic crystals combined with other advantages (e.g., light weight, non-toxicity, and low raw material cost) has allowed for new optoelectronic technologies across a wide range of applications, such as fully flexible devices for large-area displays, lighting, electrochemical transistors, and solar cells [8, 9].

Developing new materials with targeted properties through the understanding and controlling of intermolecular interactions in the crystal is the central goal of the field of crystal engineering and design. Nowadays, computer simulation has played an increasingly important role in materials development [10, 11]. Substantial progress has been made in the modeling of organic solids. When the crystal structure information is known, many properties, including geometry, stability, elasticity, ground state electronic structure, and even excited state properties can be reliably estimated from the first-principles calculation. Utilizing computational tools, high-throughput screening of chemical compounds have significantly contributed to organic materials discovery in the recent years [12–17]. More importantly, it has become possible to discover new materials systematically through computation in the past two decades. The path to this breakthrough has been paved by the development of crystal structure prediction (CSP) methods [18]. There are two largely complementary approaches: one based on existing knowledge and the contents of structural databases (data mining) and the other on powerful exploratory computer algorithms capable of making predictions with little to no pre-existing knowledge. In particular, the latter, based on powerful exploratory algorithms, has the ability to generate completely new knowledge, beyond existing databases and intuition [11, 19, 20]. With exciting developments in both methodologies and ever-increasing computational power, it is reasonable to expect that, in the near future, the use of predictive computational modeling can be accomplished within a time frame that is appreciably shorter than that needed to perform the laboratory synthesis and characterization.

The topics on crystal polymorphism and CSP have been reviewed periodically [11, 19, 21, 22] in the past. However, the CSP and its application have progressed significantly in the past five years. Herein, the authors aim to give a timely review on the recent computational methodology developments based on our own expertise. The manuscript will be organized as follows. First, we will introduce the early efforts in developing the fundamentals of organic chemical crystallography. Furthermore, the recent developmental activities in methodology, algorithm, and software in the organic CSP, as well as representative case studies, will be reviewed. Finally, we will discuss the remaining challenges and future outlook in the end.

Early efforts on organic chemical crystallography

Over the last century, small molecule crystallography has received everlasting interest by solid-state chemists since the first determination of organic crystal by X-ray diffraction in the early 1920s [24]. Researchers quickly realized that there exists some fundamental differences between organic crystals and their inorganic counterparts. First, for organic crystals, it is not difficult to picture that molecules can be stacked in many different ways in the three-dimensional (3D) space. Second, the forces holding molecules together in the crystals are much weaker than the regular chemical bonds in the atomic crystals. Although an accurate numerical calculation of intermolecular interaction only became available recently [22], it has been long known that modifying the molecular packing requires much less energy cost than the inorganic crystals. Indeed, crystal polymorphism of organic molecules has been a central subject in chemistry and materials science. Despite its ubiquity, understanding the packing of molecules is a rather elusive subject due to the variety of molecular shapes. Unlike the packing of spheres, geometry analysis on the irregularly shaped molecules appears to be more of an art than a science. In the past, there have been many excellent literature reviews on polymorphism from the perspectives of intermolecular interactions and crystal engineering [19, 21, 22, 25]. In this review, we will focus on two major questions that were proposed in the beginning of contemporary organic chemical crystallography; (i) how to understand the role of symmetry in crystal packing, and (ii) how to map the relation between molecule and crystal structures. To answer these questions, we will go over some early efforts on the development of the fundamental models to describe the intriguing crystal packing behavior.

Close-packing principle

There exist 230 space groups in which the translational symmetry of a unit cell (including lattice centering) can be uniquely combined with the point group symmetry operations of reflection, (improper) rotation, screw axis, and glide plane symmetry operations to allow an object to fill three-dimensional space in a periodic manner. Among all possible 230 space groups, it has been found that some space groups, with the symmetry elements of inversion ($\bar{1}$), the twofold screw rotation (2_1) and glide reflection (g), occur much more often than others. Figure 1(b) shows a most recent survey on the Cambridge Structural Database (CSD) [27] by the authors. Among 1075116 entries with 3D coordinates information, over 87.4% can be accounted by the top 10 space groups
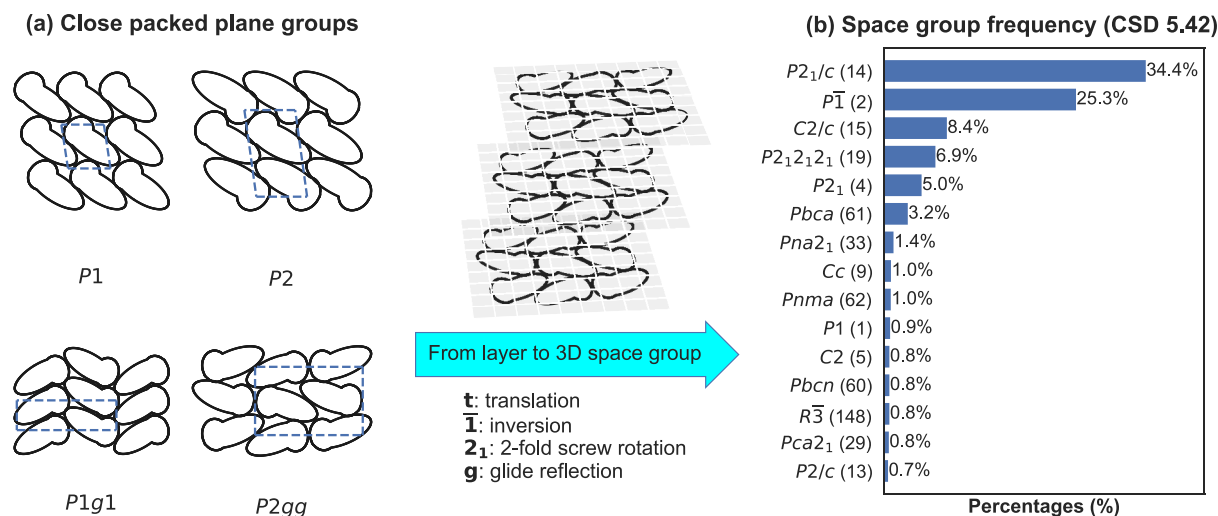
(a) Close packed plane groups

(b) Space group frequency (CSD 5.42)

From layer to 3D space group

t: translation
$\overline{1}$: inversion
2_1: 2-fold screw rotation
g: glide reflection

Space group	Percentage
$P2_1/c$ (14)	34.4%
$P\overline{1}$ (2)	25.3%
$C2/c$ (15)	8.4%
$P2_12_12_1$ (19)	6.9%
$P2_1$ (4)	5.0%
$Pbca$ (61)	3.2%
$Pna2_1$ (33)	1.4%
Cc (9)	1.0%
$Pnma$ (62)	1.0%
$P1$ (1)	0.9%
$C2$ (5)	0.8%
$Pbcn$ (60)	0.8%
$R\overline{3}$ (148)	0.8%
$Pca2_1$ (29)	0.8%
$P2/c$ (13)	0.7%

Percentages (%)

Figure 1: The close-packing theory and symmetry preference of organic crystals. (a) shows four plane groups that can form a close-packed molecular layer in which each molecule touches six neighbors (according to Kitaigorodskii [23]). (b) lists the 15 most frequent space groups extracted from more than 1 million entries in the up-to-date Cambridge Structure Database 5.42 (2021.1). The arrow in the middle illustrates Kitaigorodskii's ideas to extend the close-packing theory from 2D plane to 3D space groups.

($P2_1/c$, $P2_12_12_1$, $P\overline{1}$, $P2_1$, $C2/c$, $Pbca$, $Pna2_1$, Cc, $Pnma$, $P1$), while the contribution from other space groups is much less. In particular, 369,763 (34.39%) entries are found to possess the $P2_1/c$ symmetry. On the contrary, no single space group has more than 10% of the structures for inorganic compounds [28].

To understand the observed strong symmetry preference, people sought to the close-packing idea [23]. Kitaigorodskii and several others analyzed a series of early crystallographic data and found that the packing coefficients for the majority of crystals are between 0.65 and 0.77, which is the same order as the close-packing coefficients of spheres (0.74). Hence, they put forward the hypothesis that the molecular crystal prefers the arrangement that the projection of one molecule fit into the hollows of adjacent molecules to achieve the dense packing. Kitaigorodskii also examined the possibilities in each of the 17 plane groups to form close-packed molecular layer in which each molecule has six neighbors (following the case of close-packed spheres in a 2D plane). For a molecule in an arbitrary shape, only four plane groups ($P1$, $P2$, $P1g1$, and $P2gg$) satisfy this requirement [see Fig. 1(a)]. Furthermore, Kitaigorodskii extended the description to 80 layer groups and identified the close-packed layer group symmetries. Finally, the possible close-packing space group can be derived by considering the ways to stack the close-packed layers through (i) monoclinic displacements (t); (ii) inversion centers ($\overline{1}$); (iii) screw axis (2_1); and (iv) glide planes (g). Kitaigorodskii found that closest packing is attainable in only a few space groups (e.g., $P1$, $P\overline{1}$, $P2_1$, $P2_1/c$, $Pca2_1$, $Pna2_1$, $P2_12_12_1$, .etc.).

If the molecule is centrosymmetric, only $P\overline{1}$, $P2_1/c$, $C2/c$ and $Pbca$ are allowed. This postulate successfully explained the very non-uniform distribution of organic compounds over space groups, and predomination of $P2_1/c$ in organic chemical crystallography [as shown in Fig. 1(b)]. More remarkably, Kitaigorodskii was able to apply his close-packing principle to correct a large number of molecular crystals that were assigned to wrong space groups in the early days when X-ray diffraction was applied to small molecule crystallography.

In addition to his close-packing theory, Kitaigorodskii also introduced an atom–atom potential scheme to construct the energy landscape of molecular crystals. Using this simple potential model, he was able to accurately predict the crystal structures for several simple molecules, such as naphthalene and anthracene. He also extended the model to estimate lattice expansion under finite temperature. These approaches have inspired several further methodology developments to understand the molecular packing [26, 29, 30].

It is important to note that, despite its success in the early days of small molecule crystallography, Kitaigorodskii's model cannot be truly generalized to any case. For instance, it is possible to achieve very high packing density with a coordination number less than six [31]. One also needs to keep in mind that the close-packing theory is a pure geometry analysis; the packing preference is subjective to be complicated by the intermolecular interactions such as hydrogen and halogen bonding. A complete understanding of molecular crystals requires one to fully consider all types of intermolecular interactions and associated energies that sustain molecules in their crystal lattices.

Packing motifs

In addition to understanding the symmetry preference of molecular packing, it is also imperative to recognize, classify, and (even more ambitiously) predict the molecular packing pattern. Obviously, this is a grand challenge in terms of geometry. First, the molecular packing can dramatically change when the molecules adopt a variety of shapes. More importantly, interpreting the relation between molecular structure, packing, and functionality is believed to be an expert system, but it has no universal guidelines. It is visually difficult to capture the pattern even for experienced researchers [32].

To ease the visual challenge, most early works focused on several simple shapes. Among them, Robertson firstly proposed to divide planar aromatic hydrocarbons (PAH) into two categories based on the ratio of molecular area to the thickness [33]. It was found that the disk-like molecules tend to stack together via rigid translation (*stack-promoting*), while the molecules with smaller areas prefer the glide-like reflections (*glide promoting*). Using energetic as well as geometrical criteria, Desiraju and Gavezotti [26] extended the categorization of PAHs into four groups

and used them as a guide for prediction of crystal packing for new molecules. As shown in Fig. 2, they include

1. *herringbone*, in which each molecular column has interactions with the nonparallel neighboring molecules;
2. *sandwich*, consisting of the herringbone motif with sandwich-type diads;
3. γ, which looks like the flattened-our herringbone;
4. β, as a layered structure made up of graphite-like planes.

The classification is advantageous not only because it accounts for the difference in geometry but also is separable from energetics. In a PAH crystal, the intermolecular forces are either from short-distance $C \cdots C$ or $C \cdots H$ interactions. When two parallel molecules are stacked, the main interaction is gained from $C \cdots C$. On the other hand, two molecules also form a close pack via the glide operation [34], in which the main interactions are contributed by the $C \cdots H$ contacts. Therefore, the structures in herringbone or sandwich packing have more $C \cdots H$ contacts while the β and γ structures are characterized by the strong $C \cdots C$ interaction due to the flattened angle between
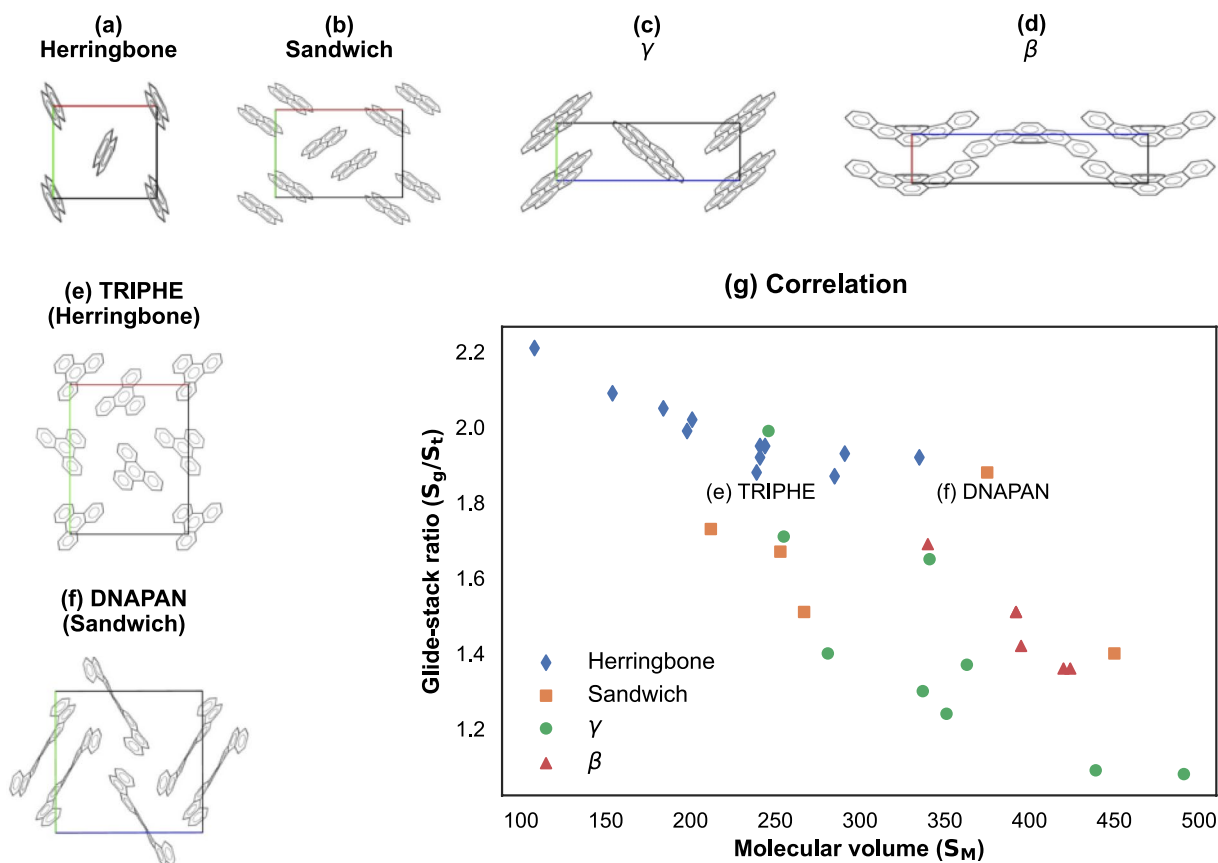

Figure 2: Packing motif analysis on the crystals made of planar aromatic hydrocarbons (PAHs). Desiraju and Gavezotti [26] suggested four representative groups of PAH packing, which are named as (a) herringbone, (b) sandwich, (c) γ and (d) β. (e) and (f) show two structures that suggest the limitation of such classification. (g) shows the overall mapping from molecular to crystal structure for 32 PAH molecules as suggested in Ref. [26].

two adjacent nonparallel molecules. Desiraju and Gavezotti also devised two parameters to measure the overall glide- (S_g) and stack-promoting areas (S_{st}) to quantify the intermolecular contributions from C···H and C···C. By plotting the correlation between the glide-stack ratio (S_g/S_{st}) and the molecular volume (S_M) for about 40 PAH molecules, they suggested a predictive mapping from molecule to crystal structure that can separate four groups of motifs and predict the unknown crystal structures for other PAH molecules [see Fig. 2(g)]. Since this seminal work, many others have utilized and expanded the packing motif concept to more PAHs and to aromatic molecules that have more chemical diversity than PAHs. This packing motif concept has been applied to establish the correlations between these materials' packing and observed physical properties (e.g., charge transport for organic semiconductors [35] and insensitivity for molecular explosives [36]). Recently, several methods have also been developed to automate the assignment of packing motif for the given molecular crystal [35, 37, 38].

Despite its increasing popularity, the definition of four patterns, like many other chemical nomenclatures, lacks a mathematical rigor. Since there exists a gradual shift of the interplanar angle between two adjacent nonparallel molecules from *herringbone* to γ and β, it is visually difficult to define their boundaries. On the other hand, there is no further separation for the sandwich-type packing. In fact, the dinaphthoanthracene crystal [DNAPAN in Fig. 2(f)] clearly shows a distinct geometry compared to the standard sandwich type. In addition, the original proposal of *herringbone*, γ, and β types seemed to be mainly designed for the crystals consisting of two types of molecular alignments. In these cases, the molecules are nonparallel to the nearest layers, but parallel to the second nearest layers. However, assigning the herringbone pattern to a complex structure with more than two alignments [e.g., the triphenylene structure in $P2_12_12_1$ symmetry as shown in Fig. 2(e)] is likely to reduce the model's predictive capability. Therefore, the number of motifs clearly needs to be expanded to characterize more subtle differences due to molecular shapes and symmetry operation.

Going beyond the PAH molecules, the close-packing principle tends to be further compromised by the existence of stronger intermolecular interactions, such as the hydrogen and halogen bonds. The energy of hydrogen (halogen) bonds ranges from 0.2 to 40 kcal/mol, which is much stronger than the typical van der Waals forces (0.1–1.0 kcal/mol). These bonds also have a strong preference for linear geometry. As such, one can easily detect the distinct bonding network if the strong hydrogen or halogen bonds exist in a crystal. Figure 3 shows two representative crystal structures based on (a) the dimers in aspirin and (b) an extended 3D hydrogen bonding networks in resorcinol. Around the 1990s, there have been numerous studies on the formation of hydrogen/halogen units connecting molecules in a crystal structure [39]. Desiraju termed such structural fragments
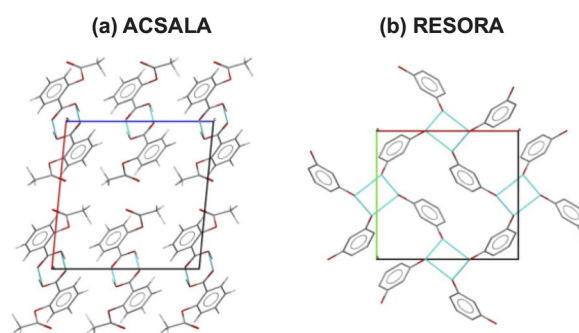

Figure 3: The crystal packing characterized by the strong hydrogen bonds in (a) aspirin form I and (b) α-resorcinol. The dotted lines indicate the intermolecular hydrogen bondings.
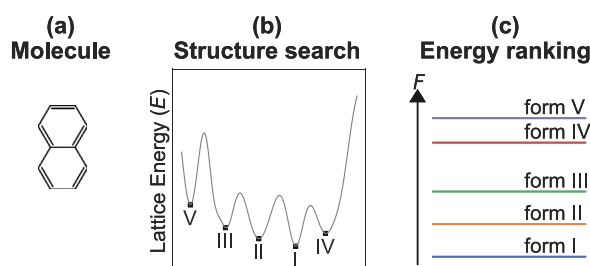

Figure 4: The schematic procedures for a CSP calculation. From the starting molecule (a), CSP needs to go through a complete structure search (b), and then energy ranking will be performed for the screened low-energy minimum structures to decide the final thermodynamic stability.

supramolecular synthons to underpin conceptual similarities between retrosynthetic analyses in conventional organic synthesis and supramolecular chemistry [40]. To develop different synthons for different purposes of applications is an active research area in crystal growth and design [21, 41].

Progress in computational crystal structure prediction

While the aforementioned works provide a big picture regarding the tendencies of molecular packing, it remains challenging to predict the detailed crystal packing before one conducts the experimental work. In the past four decades, researchers have developed a range of statistical and computational tools to enable the accurate prediction of crystal packing through the combination of powerful structure navigation and accurate energy modeling of organic materials.

Figure 4 shows a typical CSP workflow starting from the information of a single molecule. In general, a successful CSP calculation involves two stages of challenges: (i) the search problem of generating all potential low-energy structures and (ii) the energy ranking problem of computing the relative stabilities of

a range of structure candidates. Both of them represent some practical challenges when the system becomes large.

The search space grows exponentially with the number of variables, i.e., the unit cell vectors, symmetry operations, molecular positions, conformation and orientations, as well as the number of molecules per asymmetric unit (Z'). After a trial structural model is built, it needs to be relaxed to the configuration in a local energy minimum of the potential energy surface. Repeating the procedure for many iterations, one expects to identify a complete set of plausible crystal packings that may exist in the real world. Then, their relative thermodynamic (and kinetic) stabilities need to be evaluated by an accurate energy model.

The stage of energy ranking is also challenging for two reasons. First, the energy differences of different crystal packings are extremely small. It is possible to gather several tens to hundreds (or even thousands) of structures within a narrow energy window (say less than 20 kJ/mol) [19]. To accurately describe the tiny difference, high-quality electronic structure methods, instead of classical force fields or low-end electronic structure approaches, need to be employed. Second, computational limitations usually make it impractical to evaluate the lattice or free energy for each candidate structure with the high-cost model. Therefore, it is popular to adopt a multiple stage screening strategy to compromise between efficiency and accuracy.

The progress in CSP has been reflected by the periodically conducted blind tests of organic crystal structure prediction organized by the Cambridge Crystallographic Data Centre (CCDC) [42–47], which have shown that the combination of effective structure generation and energy ranking schemes can predict not only the structure of simple rigid molecules, but also the molecules representing real-life challenges. In the following, we will review the necessary computational techniques that have been developed to address the challenges in both structure searching and energy ranking.

Crystal structure search

Given that molecular packing favors only a few particular symmetry elements, it is natural to utilize these symmetry preferences to reduce the search complexity in a practical CSP calculation. Therefore, most CSP codes pay attention to developing strategies to impose symmetry constraints on all stages of structural search, including the trial structure generation, optimization, and navigation on the entire potential energy surface.

Structure generation

Despite the fact that many CSP programs have their own built-in functions to generate crystals with specific space groups, most of these functions are implemented in the program's main packages and cannot work in a standalone manner. To our knowledge, there exist only a few free packages [49, 50] allowing the generation of random molecular crystals, and they only support molecules occupying the general Wyckoff sites with $Z' = 1$, except for the recent development of Genarris 2.0 [51] which is able to deal with structures having a non-integer value of Z'. According to the report from the 2015 CSP blind test [47], most research groups attempted to reduce the structure generation to a limited range of space-group choices with one molecule in the asymmetric unit ($Z' = 1$). Despite the fact that statistical analysis supports the idea that most organic crystals tend to crystallize with Z' less than or equal to 1, structures with high Z' are not rare [52, 53], they can exist as either the ground state or metastable states [54–57]. Hence, it is risky to restrict the extent of Z' if one aims to do a complete polymorph screening. On the other hand, molecules with high symmetry tend to occupy the special Wyckoff sites with high site symmetry. In fact, a lot of reported crystals with $P2_1/c$ and $C2/c$ symmetry belong to this group. When it comes to CSP, such structures can be found by searching for structure with $Z' = 1$ in the subgroup symmetry (e.g., the naphthalene crystal in $P2_1/c$ symmetry with $Z' = 1/2$ can be represented as a $P2_1$ crystal with the molecules occupying the general Wyckoff site). However, taking the advantage of more symmetry constraints can greatly reduce the searching complexity for more symmetrical molecules, like the Buckyball shape [see Fig. 5(b)] or cage-like molecules [58].

Recently, we started to develop a standalone Python program called `PyXtal` which can be used for customized structure generation for different dimensional systems, including atomic clusters and 1D/2D/3D atomic/molecular crystals [48]. While `PyXtal` is designed for general purposes of crystal symmetry analysis, we emphasize some of its unique features for molecular crystals in Fig. 5. It allows one to generate a trial structure with a customized Z' number. If the molecular symmetry is compatible with the site symmetry of a given Wyckoff site, `PyXtal` also supports the assignment of molecule to the high symmetry site, leading to a fractional Z' in crystallographic description. Furthermore, the same scheme has been extended to generate low-dimensional systems that are described by rod/plane/layer groups [see Fig. 5(c)], which may find some applications for the study of 1D and 2D molecular crystals [59].

In addition to building the purely random crystal from scratch, it is often useful to generate the derivative structure from an existing crystal. When one crystal is converted to another by a phase transition, the symmetries of the crystal structures are usually related. The so-called group-subgroup relation has been well discussed mathematically. It is possible to list all possible subgroup types for every space-group type and to specify the subgroups in a general way by formulae. The international crystallography volume [60], as well as the online Bilbao Crystallographic Server [61], have provided the symmetry relations between a given space group G and its possible maximal
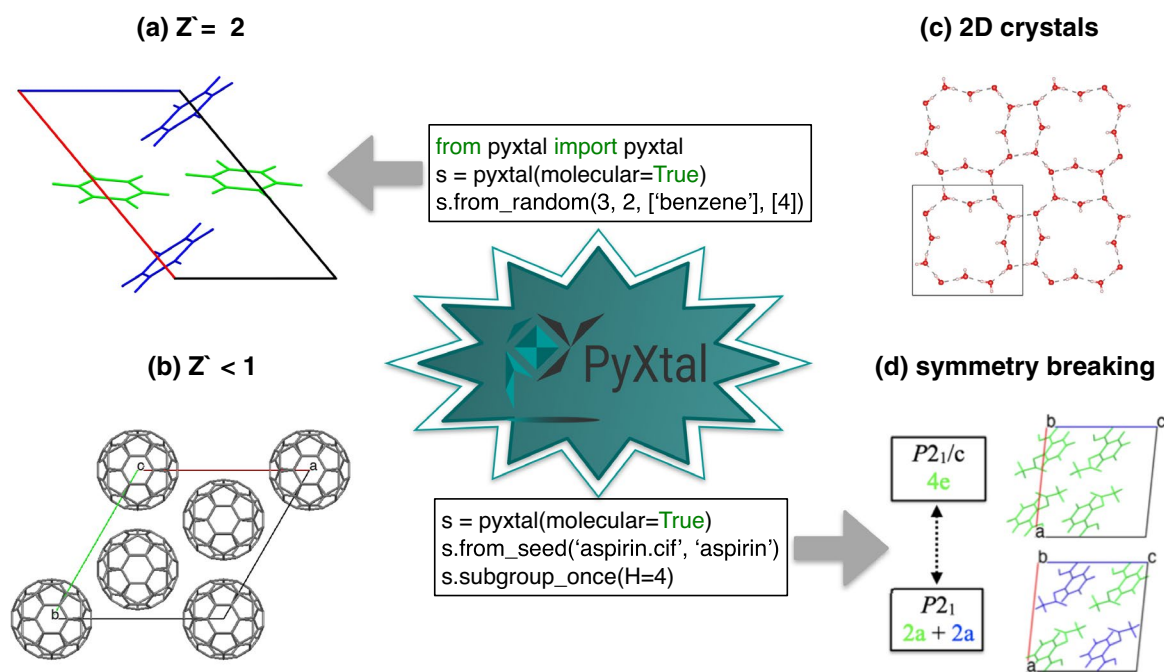
(a) Z` = 2

```
from pyxtal import pyxtal
s = pyxtal(molecular=True)
s.from_random(3, 2, ['benzene'], [4])
```

(c) 2D crystals

(b) Z` < 1

(d) symmetry breaking

```
s = pyxtal(molecular=True)
s.from_seed('aspirin.cif', 'aspirin')
s.subgroup_once(H=4)
```

$P2_1/c$
4e

$P2_1$
2a + 2a

Figure 5: The generation of different based on the open-source `PyXtal` software package [48], (a) a benzene crystal with $Z' = 1$ in $P\bar{1}$ symmetry; (b) a C60 crystal with $Z' = 1/3$ in $P6_3mc$ symmetry; (c) a 2D crystal made of ice; (d) a low symmetry aspirin crystal though symmetry breaking from the original $P2_1$ symmetry. Two short Python scripts are also listed in the text boxes to illustrate the use of `PyXtal` for generating the models in (a) and (d).

translationengleiche (t) and klassengleiche (k) subgroups H. Ideally, to complete the transition from G to H, one needs to know the cell transformation matrix, as well as the Wyckoff splitting scheme. For a given crystal structure, `PyXtal` allows a user to either systematically extract all possible transformations (subject to a cutoff index of symmetry reduction), or randomly pick one possible transformation path between G and H. Figure 5(d) shows an example of transforming a $P2_1/c$ structure in one 4e site to another structure occupying two 2e in $P2_1$ by breaking the inversion symmetry. This module can find its use in at least two applications. First, it can be directly used to study ferroelectric or piezoelectric phase transitions [62]. Second, it can be used as an effective way to generate new child structures in global optimizations that will be discussed in the following subsection.

Geometry optimization

Any generated structure has to be relaxed to the energy minimum configuration. An efficient structure relaxation requires the forces and stresses that can be derived from the energy model. In order to reduce the structural complexity, the optimization is usually applied only to the lattice vectors and molecules in the asymmetric unit. Therefore, the crystal symmetry will be retained after structure relaxation. The simplest algorithm is steepest descent, where the crystal variables (e.g., atomic coordinates and lattice vectors) are iteratively updated, moving along the force/stress direction. In this algorithm, only the current positions of atoms and forces are taken into account. The Conjugate gradients algorithm [63], while still requiring only the forces/stresses (i.e., first derivatives of the energy), follows a set of conjugate vectors toward the target minimum during the line search. The choice of conjugate vectors, instead of the gradient direction, can effectively remedy the slow convergence issue when it approaches the minima. The Newton method takes the knowledge of the second derivatives matrix (Hessian) to greatly accelerate convergence to the local minimum, but the computation of the Hessian is very expensive when the number of variables becomes large. Therefore, a number of quasi Newton methods have been developed to construct an approximate numerical Hessian for the first derivatives. Notable examples include DFP, BFGS, and L-BFGS [63]. More efficient algorithms based on conventional molecular dynamics with additional velocity modifications and adaptive time steps also have been made available [64]. Recently, a hyperspatial optimization idea was also proposed to help avoid having a structure trapped in a shallow minima [65].

While there exist plenty of choices for efficient optimization algorithms, it is important to note that the *no free lunch theorem* still holds. Most of the optimization algorithms, other than steepest descent, assume that the starting point can be connected to the target minimum via a quadratic function. If the quadratic approximation is far from the ideal case, the choice of search direction may become worse than the gradient direction. Figure 6 shows such an example. When the initial position is
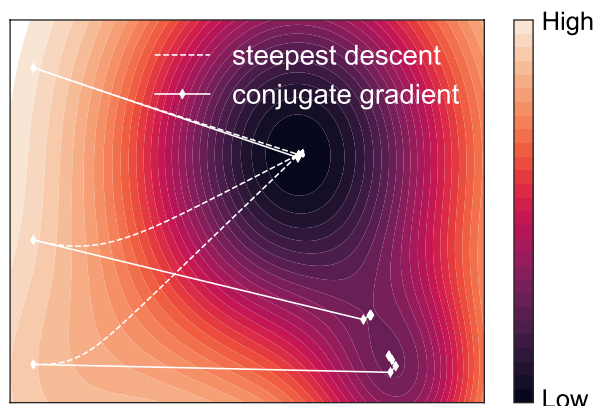
Figure 6: The comparison between steepest descent and conjugate gradient optimization algorithms.

close to the minimum, the conjugate gradients algorithm can effectively identify the minimum with only a few iterations. However, the same algorithm may lead to a worse minimum or even a saddle point when the initial positions become more distant to the target minimum. On the other hand, the steepest descent is more reliable to consistently find the same low-energy minimum as long as the initial position is within the same basin. For the practical application of CSP, it is often the case that the initial structure from a random guess is far from the minimum. Therefore, the simplest steepest descent is highly recommended at the beginning stage of geometry optimization.

Global structure navigation

Combined with global optimization, one can perform an extensive search for the plausible structures. Global optimization is a very large field in applied mathematics, with many methods continually being developed. For the CSP applications, it is important to remember that no existing method can give a guarantee of finding the global minimum in a finite amount of computational time. The following methods have become popular:

1. Random search is the simplest search strategy, starting from the atom–atom potential implementation [66], and then evolving into different versions of ab initio random structure search [67]. In practice, it is crucial to steer the searches toward finding realistic structures while maintaining structural diversity by imposing constraints on symmetry or some pre-defined structural units. When the number of variables is small, it is also possible to sample the configuration space based on the determined quasi-random, low-discrepancy sequences [68, 69]. Such quasi-random sampling has been very successful in a lot of engineering applications [70].

2. Simulated annealing [71, 72] is a strategy inspired by annealing of crystals, where gradual cooling leads to the equilibrium crystal structure. Basin hopping [73] adopts a similar strategy to perturb the crystal by Monte Carlo moves. These two methods mainly differ in whether or not relaxing the structure to the minimum after the perturbation.

3. Metadynamics [74] is designed to overcome large energy barriers through positively biased molecular dynamics simulation. It requires a set of collective variables (CVs) to distinguish between states of the system and scans the low-energy part of the energy landscape by distorting the actual energy landscape by a history-dependent potential, the aim of which is to discourage the system from sampling the states that have already been sampled. In addition to metadynamics, sampling of the CVs can also be achieved using a temperature-accelerated adiabatic free-energy dynamics approach [75]. These approaches are advantageous since they allow direct sampling on the free-energy surface.

4. Evolutionary algorithms are a type of nature inspired method that comes in many flavors [76–78]. The common idea is that a population of structures is evolved, driven by natural selection of lower-energy structures to become parents of a new generation of structures. These types of algorithms have been extremely successful in many different engineering applications. For the applications to the CSP, recipes for producing offspring from parents (genetic crossover and mutations) are of key importance [77].

We note that there also exist a variety of other implementations for predicting inorganic crystals. The fields of inorganic and organic CSP are beginning to converge [79]. For a more complete list of global optimization methods, the readers are recommended to check the Ref. [11]. It is expected more algorithms will be developed in the next few years.

It is also possible to consider more than one target function in global optimization. For instance, one may be interested in solving the crystal structure with unindexed powder x-ray diffraction (PXRD) data. In this case, one needs to simultaneously optimize two quantities, i.e., the minimum lattice energy and the maximum similarity with respect to the reference PXRD pattern [80]. Such multi-objective optimization problems can be handled by Pareto optimization [70]. It is also possible to use this idea to design new materials with optimum physical properties.

Models for intermolecular interactions

From a structure search, we expect to generate many trial structures whose energetic and physical properties need to be evaluated in a high-throughput manner. A popular choice is to use low-cost energy models to perform screening at the initial stage, and then employ more expensive electronic structure

calculations for a short list of structures in the end. Here we will review the models to be used for both stages of structure optimization and final ranking.

Empirical models

Since the early developed atom-atom potential model, there have been plenty of efforts in developing more accurate empirical models to estimate the crystal packing energy. In molecular simulation, the popular choices are classical force fields (FFs) [81–85], distributed multipole models [86], machine-learning potentials (MLP) [87], and the Harris approximation [88]. Among them, the classical FFs and MLPs have been widely used in recent years. Classical FFs are built upon the total energy decomposition into several analytical terms, including 2/3/4-body intramolecular interactions and non-bonded van der Waals (vdW) dispersion. On the other hand, MLPs express the total energy as the machine-learning function of a high-dimensional array that represents each atomic local environment [89, 90]. In general, both approaches involve a parameterization (or training) from the reference data. The MLPs, usually based on neural networks (NN) regression of a set of symmetry-invariant structure descriptors [90, 91], give better accuracy when the configurations are highly correlated with the training data. However, it is also likely to generate nonphysical results if the input data are outside the training domain. Compared to the MLPs, the classical FFs are computationally cheaper and more robust in handling unknown configurations, but they are often insufficient to describe the subtle energy differences in crystal packing.

For an accurate simulation, it is important to know the quality of each energy model. In a structure optimization, a good model should guarantee the reproduction of the same geometry for the input crystal. Figure 7 shows our recent survey of 345 crystal data by comparing the differences between experimental and simulated geometries based on a classical FF (GAFF [81]) and a MLP (ANI-2x [92]). The results suggest that both energy models are sufficient to describe most of the experimental geometries for
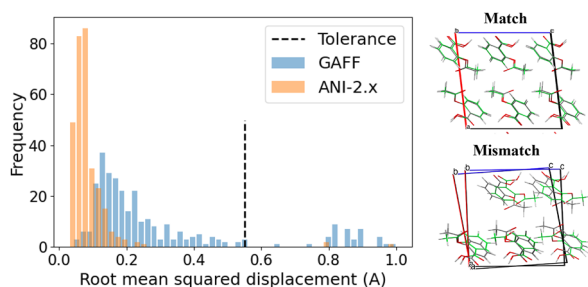

Figure 7: The benchmark of GAFF [81] and ANI-2x [92] for 345 crystals according to the authors' study on the C–H–O–N–S molecular systems.

most molecules. However, some systems cannot be handled. This is a common problem for any kind of generic force field since their training is built upon many different types of data.

A possible solution to the accuracy issue is to develop a customized force field [93]. The basic idea is to start with existing FF parameters to run CSP and then collect a new set of reference data by more accurate electron structure calculation. The FF parameters are then refined based on the new dataset. The tailor-made FF [93] has led to remarkable success in several recent CSP blind tests [45–47]. In principle, this strategy can be adapted to the construction of MLP as well.

Electronic structure theory

To make a final decision about which polymorph is most likely to exist, a short list of structures needs to be evaluated preferably by more accurate electronic structure methods [22]. Crystal polymorphism originates from the competition between intramolecular and intermolecular interactions for different crystal packings. The current state-of-the-art electronic structure methods based on density functional theory (DFT) are insufficient to capture such weak interaction. Various methods have been proposed to explicitly incorporate vdW interactions. One common approach is to add an extra energy term of $-C_6/R^6$ to describe the first term of vdW interactions between two dipoles in a multipole expansion, as implemented in DFT-D [94], DFT-D2 [95], and TS [96]. The C_6 term represents the dipole–dipole dispersion coefficient between the two atoms involved and R is the interatomic distance. More advanced correction includes the additional, dipole-quadrupole, and quadrupole–quadrupole contributions [97–99]. Another approach is to obtain dispersion interactions by designing functionals that explicitly include nonlocal correlations (though still based on pairwise addition) [100]. Furthermore, Tkatchenko and coworkers proposed the many-body dispersion (MBD) method [101] which describes many-body dipolar interactions up to infinite order and also includes electrodynamic response effects. Several test sets have been proposed to serve as the test bed to assess the performance of different DFT models [98, 102]. It was found that the MBD method substantially outperforms the original TS scheme, in particular for molecular crystals.

Although all the aforementioned vdW corrections were originally introduced to fix the shortcoming of a standard DFT calculation, they can also work with other semi-empirical electronic structure methods. For instance, Mortazavi et al recently combined the TS and MBD with third-order density functional tight binding (DFTB3) [103] via a charge population-based method [104]. They found an overall good performance for the X23 benchmark database of molecular crystals, while the entire calculation was 3000 times faster than the full DFT treatment. Recently, various

machine-learning schemes have been introduced to systematically improve approximation of repulsive term in DFTB parameterization, and they showed great promise of achieving the chemical accuracy within the framework of DFTB modeling [105, 106]. Another general-purpose density functional tight-binding method, the semi-empirical extended tight-binding (xTB) model, is also gaining increased popularity due to its extended support for nearly 90 elements and improvements of the underlying theory regarding the treatment of the important electrostatic and dispersion interactions[107]. These exciting developments suggest that such methods can serve as a viable pre-screening tool for CSP in a high-throughput manner.

Successful examples

In the past decades, there have been a wealth of successes in using the CSP for organic crystal structure determination and functional materials design. It is obviously impossible to list each of them. In the following, we will focus on several representative examples in three main areas (i) pharmaceutical polymorph screening; (ii) high Z' compounds and (iii) new organic semiconductors. The choice of these topics are mostly based on the authors' expertise and research interests. For other types of applications, the readers are encouraged to review other literature.

Pharmaceutical compounds

Understanding, predicting, and controlling the structure of crystalline pharmaceuticals materials are a topic of tremendous importance because the crystal packing directly affects properties such as solubility, processability, hygroscopicity, chemical and physical stability, and bioavailability. In recent years, CSP has been increasingly successful in aiding the solid form screening of drugs. For instance, a recent computational screening of the pharmaceutical compound Dalcetrapib with 10 torsional degrees of freedom led to the discovery of a new form which was successfully synthesized under high pressure [108]. Inspired by the CSP calculation results, the search for new polymorphs of galunisertib was expanded to an unusual range of experiments, including melt crystallization under pressure, to work around solvate formation and the thermal instability of the molecule. These works led to the identification of ten unsolvated polymorphs of galunisertib were found [109]. As shown in Fig. 8, both molecules have mass over 350 g/mol and nontrivial topology. These successful examples demonstrate how current state-of-the-art CSP methods can be combined with experimental techniques to assess the polymorph risk in drug development.
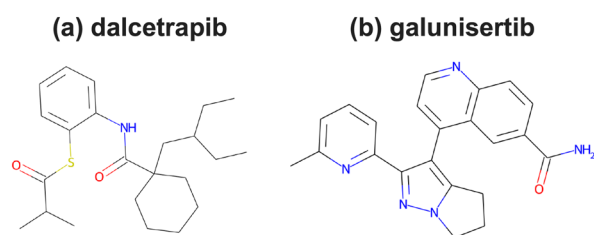

(a) dalcetrapib   **(b) galunisertib**

Figure 8: Two representative drug molecules that CSP has been successfully used to guide the polymorph screening [108, 109].

High Z' crystals

A standard CSP calculation often limits the extent of search to $Z' = 1$. However, an early database survey suggested that over 8% of compounds have high Z' numbers [52]. Therefore, it is necessary to go beyond $Z' = 1$ for future polymorph screening, if metastable polymorphs are of interest, as observed in recent cases [3, 54–56, 110]. For instance, coumarin, a rather simple molecule with only one well-known solid phase, turns out to have five polymorphs, one with three molecules in the asymmetric unit [56]. Structures of the α and β phases of resorcinol, were the first polymorphic pair of molecular crystals solved by X-ray analysis in 1930s [111–113]. A new metastable polymorph was recently observed in a melt crystallization experiment. However, its complex structure ($Z' = 2$, $P2_12_12_1$) was not solved until the aid of CSP efforts [55]. Similar synergy between theory and experiment has also been applied to establish the third ambient polymorph of aspirin [54], which also possesses two molecules in the asymmetric unit with the space-group symmetry of $P2_1/c$ [Fig. 9(a)].

Another remarkable example is the identification of a long term elusive ROY polymorph. For a decade, the molecule 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile has been considered the most polymorphic system [114]. It is called ROY for its red, orange, and yellow crystals. Out of the ten known polymorphs, seven forms have been solved in the 2000s. One of the remaining three unsolved structures, R05, an abbreviation for "red 2005", was only solved recently through the combination of experiment and CSP calculation. This R05 form [Fig. 9(b)] is both the first acentric ROY polymorph, and the first with $Z' > 1$. Since our successful determination of R05, there has been a renewed interest in ROY. Four new ROY polymorphs have been reported since 2019 through the joint efforts between experiment and computation [115–118]. It is expected that more ROY polymorphs await future discovery according to the improved energy ranking method by Beran et al. [119].
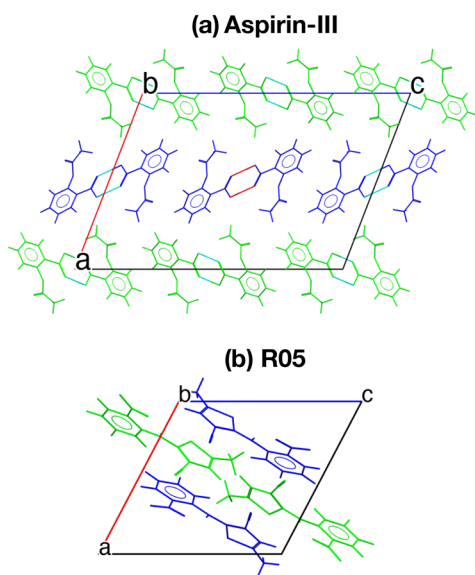
(a) Aspirin-III

(b) R05

Figure 9: Complex packing of high Z' crystals of (a) aspirin [54] and (b) ROY [57]. The molecules in the asymmetric units are colored differently to guide the eye.

New organic semiconductors

Polymorphism has been considered as an emerging design strategy for organic functional materials. Structure determination for metastable polymorphs might gain impetus in the near future. One of the appealing industrial applications is the design of organic semiconductors, a special class of organic materials with delocalized molecular orbitals in which charge carriers are mobile due to extended π-conjunction. Sokolov et al. [120] developed a computational screening scheme to search for new compounds derivable from existing molecules and found a new compound with a significantly improved mobility over the parent molecules. With the advance in both computational power and CSP methodology development, a systematic CSP study of energy-structure-function maps for small hypothetical molecules is now possible [35, 121, 122].

The idea of global optimization is not necessarily limited to the navigation of crystal packing. Instead, it can also be used to identify promising molecules from a vast chemical space. Recently, an evolutionary method, combined with the CSP was developed to explore a user-specified region of chemical space to identify promising molecules for organic semiconductors with high charge carrier mobilities [123]. It has been shown the method can efficiently explore a large space for pentacene- and pyridazine-based molecules having both low reorganization energies and high electron affinities.

Remaining challenges

In recent years, the power of CSP has been demonstrated by many successful examples in materials' development and design. It is clear that CSP has become central to the study of

organic materials. New computational tools for such work are being widely developed by different research groups, as shown by the increasing number of participants in the recent blind tests [42–47]. Not only materials theorists, many experimental groups are actively using the CSP tool to design and interpret their experimental data. That has been said, structure prediction faces several challenges in handling large molecules, high Z' packing, temperature effects, and large chemical space. Among them, some challenges (e.g., free-energy ranking) have been under active consideration. However, some issues may not receive much attention. All issues will be discussed as follows. In particular, we will pay more attention to the issues that have been seldom addressed in other review works.

Overlooked historic wisdom

The symmetry correlation between molecule and crystal packing used to be the central subject in the early days of organic crystallography. However, most research efforts have been shifted to numerical simulations for each concrete molecule and crystal. After the seminal work of Kitaigorodskii, the pursuit of a general geometric model became much less active. Nowadays, most CSP practitioners generate the trial structure from space-group symmetry. This has been shown to be sufficient for small, rigid, and simple-shaped molecules, thanks to the continuous methodology and code developments in the past two decades. Nevertheless, it remains challenging to predict the correct packing even when the space group is known. In fact, Kitaigorodskii's original proposal to study 3D packing from molecular layers may provide a short cut to reduce such search complexity. In the future, it would be worthwhile to revisit those early ideas and check their validity for the new challenges.

In terms of the crystal packing motif, researchers continue to use those highly subjective nomenclatures (as part of historical heritages) to describe the crystal packing and interpret the observed results. For instance, the four groups of PAH crystals [26] are still widely used, despite their apparent limitations as described in the previous section. In fact, understanding the relation between molecular shape and packing is not far from a standard pattern recognition problem in computer vision. The emerging computational techniques in data science, geometry, and computation vision should be imported in a timely manner. However, one must keep in mind that symmetry invariance (e.g., rotation and translation) needs to be taken into account. If the packing pattern can be well grouped, the library of packing motifs should be the next step. Consequently, one can analyze the occurrence of each packing motif in a way similar to the prototype library for inorganic crystals [124, 125]. For a given molecular shape, it is expected that only a small number of *template* packing structures are likely to be populated as probable observable crystal polymorphs [126]. Such *templates* can be
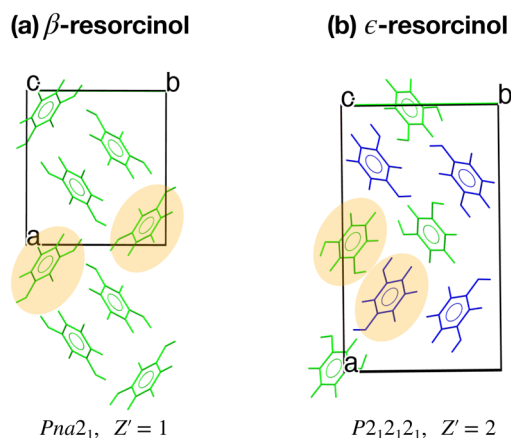
(a) β-resorcinol

(b) ε-resorcinol

$Pna2_1, \ Z' = 1$

$P2_12_12_1, \ Z' = 2$

Figure 10: The comparison between two resorcinol polymorphs (a) β and (b) ε. In (b) two molecules in the asymmetric unit are colored differently. The light blue eclipses are marked to highlight the likely pseudo inversion symmetry.

extracted from the crystal packing library to generate candidate structures for the new molecules [127].

In addition, there has been a small concerted effort to quest the origin of high Z' organic crystals' formation [53, 128–130]. It has been hypothesized that the high Z' crystals may be closely related to a $Z' = 1$ crystal by a pseudosymmetry via small molecular displacements and rotations [53, 129]. Similar phenomenon has been found in several of our studies. For instance, the newly discovered ε-resorcinol ($P2_12_12_1$) was obtained via melt crystallization with a high cooling rate. With slow cooling, a more stable β-resorcinol phase ($Pna2_1$) should be formed. Comparing their crystal packing in Fig. 10, there are some clear hints showing that both polymorphs can be transformed to a common high point group symmetry *mmm* via a pseudo inversion symmetry. Therefore, it is likely that a high symmetry intermediate phase may form in the beginning and then a symmetry lowering occurs after crystal nucleation upon cooling. It would be interesting to check the validity of such a hypothesis through more rigorous characterization in both experiment and simulation. If such high symmetry intermediate phases do exist and they can be reliably predicted, the high Z' crystals can be trivially resolved through the rigorous group-subgroup relation without an extensive CSP search.

Technical challenges

For the standard CSP calculation, the most urgent need is to be able to measure relative stability at finite temperature. Most current CSP techniques should be regarded as *zeroth-order* approaches [131] that consider the most stable crystal structure by evaluating lattice energy only. However, the real world is governed by free energies. It is possible that lattice vibration under finite temperature can substantially change the stability ranking

[132]. Indeed, several groups improved their final polymorph rankings by considering the vibrational free-energy ranking under harmonic, quasi-harmonic, and even anharmonic treatments [47]. More advanced methods have also been proposed [133]. Although the free-energy calculations are still expensive, one should be optimistic that such calculations can be done within a reasonable time frame in the next couple of years with the further improvement in computational methodology.

Another challenge is to enable a more efficient CSP calculation in a high-throughput manner. In a standard CSP run, there is no general mechanism to terminate the search and one usually attempts to generate the trial crystal structures as many as possible. This exhaustive sampling strategy is suitable when one only needs to deal with a few systems. However, the organic materials development may require one to consider many molecules. To enable such high-throughput polymorph screening in a reasonable time frame, we shall develop a more efficient sampling strategy to avoid exhaustive searches. For a given chemical, a large number of hypothetical structures can exist. Relaxing each of them to their lowest-energy local minimum, the set of trajectories then form the basins of attraction on the potential energy surface. The probability of successfully finding a low-energy structure depends on the shapes and sizes of the hyper-volumes of the basins of attraction and the details of the search [67, 134, 135]. It is possible that some molecules may have fuzzy landscapes compared to others [19]. To validate the concept of landscape complexity, we selected three different molecules to run CSP simulations based on our evolutionary algorithms with a population size of 100. We terminated the search at the Nth generation as long as the experimental structure is found at the end of that generation. Such calculations were repeated for 50 times to collect the statistics. Figure 11 shows the distribution of N for each molecule. Clearly, the results indicate that each molecule has a distinct crystal landscape. For the molecules with a simpler crystal landscape, we can confidently set a small number of generations in the CSP simulation. On the other hand, more generations will be needed for the molecules
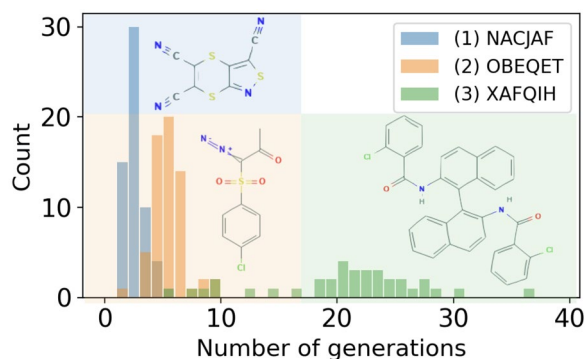

Figure 11: Comparison of CSP complexity for three molecules taken from the previous CCDC blind tests.

with complex landscapes. In the future, a more automated strategy should estimate the packing complexity and guide the high-throughput CSP exploration.

In terms of applications, the recent focus in crystal engineering has progressively shifted from structure to function. In describing the structure–property relation, symmetry plays a deterministic role as it can greatly simplify the calculation of physical quantities, such as elastic constants, vibration modes, and piezo-coefficients. While the relation between symmetry and crystal packing has been widely studied in the past, the symmetry relation between different polymorphs remains largely underexplored. Displacive phase transitions involve the structural changes that follow a low-energy barrier path between two phases with close symmetry relation. Nowadays, symmetry analysis for an organic crystal relies on a lot of human–machine interactions (e.g., extracting the molecular coordinates, looking up the character tables and trying related subgroup conventions). In the future, it is impetus to develop a tool that can automate the symmetry analysis for organic crystals, thus, enabling the studies on organic materials with interesting magnetic [136], ferroelectric [62], piezoelectric, and topological electronic properties [137].

Open software development

In addition to the innovations in methodology and algorithm, it is also worthwhile to discuss the current situation for open-source CSP software developments. In the field of materials' modeling, there is clearly an ongoing paradigm shift driven by the open science movement. In the past decade, we have witnessed substantial growths of open databases for both inorganic [138–140] and organic materials [141–143], as well as the open packages that enable the automated process of materials data and property evaluation [144, 145]. As described in the review, conducting a successful organic crystal prediction requires the integration of several different computational pipelines ranging from force field generation, structure search, and (free) energy ranking. Currently, most researchers rely on subscriptions to commercial license to obtain the data, run simulations, and analyze results. Such heavy software dependence largely discourages young scientists from bringing new expertise and algorithmic innovation to this field. Due to the license restriction, it is also hard to reproduce historically published results, even for an experienced researcher. For the CSP of inorganic materials, there have been plenty of choices (e.g., USPEX [76], AIRSS [67], CALYPSO [146], XtalOpt [147], GASP [148]) that are either completely open source or free for academic researchers. However, the codes [69, 76, 78] for the prediction of organic materials are still limited and none of them are. To promote the open-source activity in organic CSP, we have recently launched the open-source project `PyXtal` [48] to provide easy access

to generating and manipulating crystal structure via different symmetry constraints. We hope that this project can provide a platform that allows more researchers to develop or apply their advanced structure search or energy ranking methods for different CSP-related research activities.

Conclusions

In sum, we have presented a comprehensive review of the recent progress of crystal structure prediction. It is clear that we are witnessing an exciting paradigm shift where computer simulation is playing a more deterministic role in the development of new organic materials. Several notable examples include (i) very complex crystals that represent real-life challenges in pharmaceutical research can be predicted from pure blind tests; (ii) the energy ranking methods have been continuously improved; and (iii) new open-source codes have become available to provide invaluable resources to new practitioners with a minimum learning barrier. These exciting developments have significantly boosted the application of CSP in guiding the screening and design of new organic compounds. While there remain some limitations in handling more complex systems and exploring more diverse chemical space, it is reasonable to expect that, in the near future, the use of predictive computational modeling can be accomplished within a time frame that is appreciably shorter than that needed to perform the laboratory synthesis and characterization.

Acknowledgments

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Open Access

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. A.Y. Lee, D. Erdemir, A.S. Myerson, Crystal polymorphism in chemical process development. Annu. Rev. Chem. Biomol. Eng. **2**, 259–280 (2011)

2. Z. Zhuo, C. Wei, M. Ni, J. Cai, L. Bai, H. Zhang, Q. Zhao, L. Sun, J. Lin, W. Liu, et al., Organic molecular crystal with a high ultra-deep-blue emission efficiency of 85% for low-threshold laser, Dyes. Pigm. 110425 (2022)

3. J. Yang, C. Hu, X. Zhu, Q. Zhu, M.D. Ward, B. Kahr, Ddt polymorphism and the lethality of crystal forms. Angew. Chem. **129**, 10299–10303 (2017)

4. G. Liu, R. Gou, H. Li, C. Zhang, Polymorphism of energetic materials: a comprehensive study of molecular conformers, crystal packing, and the dominance of their energetics in governing the most stable polymorph. Cryst. Growth Des. **18**, 4174–4186 (2018)

5. J. Valasek, Piezo-electric and allied phenomena in Rochelle salt. Phys. Rev. **17**, 475–481 (1921)

6. H. Kallmann, M. Pope, Bulk conductivity in organic crystals. Nature **186**, 31–33 (1960)

7. S. Haas, A.F. Stassen, G. Schuck, K.P. Pernstich, D.J. Gundlach, B. Batlogg, U. Berens, H.-J. Kirner, High charge-carrier mobility and low trap density in a rubrene derivative. Phys. Rev. B **76**, 115203 (2007)

8. Q. Li, Z. Li, Molecular packing: another key point for the performance of organic and polymeric optoelectronic materials. Acc. Chem. Res. **53**, 962–973 (2020)

9. P. Yu, Y. Zhen, H. Dong, W. Hu, Crystal engineering of organic optoelectronic materials. Chemistry **5**, 2814–2853 (2019)

10. A. Jain, Y. Shin, K.A. Persson, Computational predictions of energy materials using density functional theory. Nat. Rev. Mater. **1**, 15004 (2016)

11. A.R. Oganov, C.J. Pickard, Q. Zhu, R.J. Needs, Structure prediction drives materials discovery. Nat. Rev. Mater. **4**, 331–348 (2019)

12. S. Fratini, S. Ciuchi, D. Mayou, G.T. De Laissardière, A. Troisi, A map of high-mobility molecular semiconductors. Nat. Mater. **16**, 998–1002 (2017)

13. P. Friederich, A. Fediai, S. Kaiser, M. Konrad, N. Jung, W. Wenzel, Toward design of novel materials for organic electronics. Adv. Mater. **31**, 1808256 (2019)

14. A. Saeki, K. Kranthiraja, A high throughput molecular screening for organic electronics via machine learning: present status and perspective. Jpn. J. Appl. Phys. **59**, SD0801 (2019)

15. T. Nematiaram, D. Padula, A. Troisi, Bright frenkel excitons in molecular crystals: a survey. Chem. Mater. **33**, 3368–3378 (2021)

16. A. Stuke, C. Kunkel, D. Golze, M. Todorović, J.T. Margraf, K. Reuter, P. Rinke, H. Oberhofer, Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. Sci. Data **7**, 1–11 (2020)

17. C. Kunkel, C. Schober, J.T. Margraf, K. Reuter, H. Oberhofer, Finding the right bricks for molecular legos: a data mining approach to organic semiconductor design. Chem. Mater. **31**, 969–978 (2019)

18. A.R. Oganov, *Modern Methods of Crystal Structure Prediction* (Wiley, New York, 2011)

19. S.L. Price, Predicting crystal structures of organic compounds. Chem. Soc. Rev. **43**, 2098–2111 (2014)

20. A. Gavezzotti, *The Crystalline States of Organic Compounds*, vol. 20 (Elsevier, 2021)

21. M.K. Corpinot, D.-K. Bučar, A practical guide to the design of molecular crystals. Cryst. Growth Des. **19**, 1426–1453 (2018)

22. G.J. Beran, Modeling polymorphic molecular crystals with electronic structure theory. Chem. Rev. **116**, 5567–5613 (2016)

23. A. Kitaigorodskii, Organic chemical crystallography, con Sultants bureau: New York, 1961 (originally published in Russian by the press of the academy of sciences of the USSR, Moscow, 1955); Spek, al, single-crystal structure validation with the program Platon. J. Appl. Crystallogr. **36**, 7–13 (2003)

24. R.G. Dickinson, A.L. Raymond, The crystal structure of hexamethylene-tetramine. J. Am. Chem. Soc. **45**, 22–29 (1923)

25. J.D. Dunitz, A. Gavezzotti, How molecules stick together in organic crystals: weak intermolecular interactions. Chem. Soc. Rev. **38**, 2622–2633 (2009)

26. G.R. Desiraju, A. Gavezzotti, Crystal structures of polynuclear aromatic hydrocarbons. classification, rationalization and prediction from molecular structure. Acta Cryst. B **45**, 473–482 (1989)

27. R. Taylor, P.A. Wood, A million crystal structures: the whole is greater than the sum of its parts. Chem. Rev. **119**, 9427–9477 (2019)

28. W. Baur, D. Kassner, The perils of cc: comparing the frequencies of falsely assigned space groups with their general population. Acta Cryst. B **48**, 356–369 (1992)

29. A. Gavezzotti, Calculation of intermolecular interaction energies by direct numerical integration over electron densities. i. electrostatic and polarization energies in molecular crystals. J. Phys. Chem. B **106**, 4145–4154 (2002)

30. A. Gavezzotti, Calculation of intermolecular interaction energies by direct numerical integration over electron densities. 2. an improved polarization model and the evaluation of dispersion and repulsion energies. J. Phys. Chem. B **107**, 2344–2353 (2003)

31. R.P. Scaringe, A theoretical technique for layer structure prediction, in *Electron Crystallography of Organic Molecules* (Springer, 1991) pp. 85–113

32. G.M. Day, W.S. Motherwell, An experiment in crystal structure prediction by popular vote. Cryst. Growth Des. **6**, 1985–1990 (2006)

33. J.M. Robertson, The measurement of bond lengths in conjugated molecules of carbon centres. Proc. R. Soc. Lond. A **207**, 101–110 (1951)

34. For the planar molecules, the symmetry elements to achieve the close packing can be interpreted as either $\bar{1}$, 2_1, or g. In the text, this operation is broadly defined as glide reflection. As such, many herringbone and sandwich type crystals have the space groups without the explicit g symmetry operator (e.g., $P\bar{1}$ and $P2_1$)

35. J.E. Campbell, J. Yang, G.M. Day, Predicted energy-structure-function maps for the evaluation of small molecule organic semiconductors. J. Mater. Chem. C **5**, 7574–7584 (2017)

36. D. Mathieu, Sensitivity of energetic materials: theoretical relationships to detonation performance and molecular structure. Ind. Eng. Chem. Res. **56**, 8191–8201 (2017)

37. D. Loveland, B. Kailkhura, P. Karande, A.M. Hiszpanski, T.Y.-J. Han, Automated identification of molecular crystals' packing motifs. J. Chem. Inf. Model. **60**, 6147–6154 (2020)

38. D. Ito, R. Shirasawa, S. Iino, S. Tomiya, G. Tanaka, Estimation and prediction of ellipsoidal molecular shapes in organic crystals based on ellipsoid packing. PLoS ONE **15**, e0239933 (2020)

39. M.C. Etter, Hydrogen bonds as design elements in organic chemistry. J. Phys. Chem. **95**, 4601–4610 (1991)

40. G.R. Desiraju, Supramolecular synthons in crystal engineering—a new organic synthesis. Angew. Chem. Int. Ed. **34**, 2311–2327 (1995)

41. G.R. Desiraju, Crystal engineering: from molecule to crystal. J. Am. Chem. Soc. **135**, 9952–9967 (2013)

42. J.P. Lommerse, W.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Gavezzotti, D.W. Hofmann, F.J. Leusen, W.T. Mooij, S.L. Price, B. Schweizer et al., A test of crystal structure prediction of small organic molecules. Acta Cryst. B **56**, 697–714 (2000)

43. W.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W. Hofmann, F.J. Leusen, J.P. Lommerse, W.T. Mooij et al., Crystal structure prediction of small organic molecules: a second blind test. Acta Cryst. B **58**, 647–661 (2002)

44. G.M. Day et al., A third blind test of crystal structure prediction. Acta Cryst. B **61**, 511–527 (2005)

45. G.M. Day et al., Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test. Acta Cryst. B **65**, 107–125 (2009)

46. D.A. Bardwell et al., Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. Acta Cryst. B **67**, 535–551 (2011)

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

48. S. Fredericks, K. Parrish, D. Sayre, Q. Zhu, Pyxtal: A python library for crystal structure generation and symmetry analysis. Comput. Phys. Comm. **261**, 107810 (2021)

49. B.P. van Eijck, J. Kroon, Upack program package for crystal structure prediction: force fields and crystal structure generation for small carbohydrate molecules. J. Comput. Chem. **20**, 799–812 (1999)

50. J.R. Holden, Z. Du, H.L. Ammon, Prediction of possible crystal structures for c-, h-, n-, o-, and f-containing organic compounds. J. Comput. Chem. **14**, 422–437 (1993)

51. R. Tom, T. Rose, I. Bier, H. O'Brien, A. Vazquez-Mayagoitia, N. Marom, Genarris 2.0: A random structure generator for molecular crystals. Comput. Phys. Commun. **250**, 107170 (2020)

52. T. Steiner, Frequency of z prime values in organic and organometallic crystal structures. Acta Cryst. B **56**, 673–676 (2000)

53. C.P. Brock, High-z prime structures of organic molecules: their diversity and organizing principles. Acta Cryst. B **72**, 807–821 (2016)

54. A.G. Shtukenberg, C.T. Hu, Q. Zhu, M.U. Schmidt, W. Xu, M. Tan, B. Kahr, The third ambient aspirin polymorph. Cryst. Growth Des. **17**, 3562–3566 (2017)

55. Q. Zhu, A.G. Shtukenberg, D.J. Carter, T.-Q. Yu, J. Yang, M. Chen, P. Raiteri, A.R. Oganov, B. Pokroy, I. Polishchuk et al., Resorcinol crystallization from the melt: a new ambient phase and new riddles. J. Am. Chem. Soc. **138**, 4881–4889 (2016)

56. A.G. Shtukenberg, Q. Zhu, D.J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko et al., Powder diffraction and crystal structure prediction identify four new coumarin polymorphs. Chem. Sci. **8**, 4926–4940 (2017)

57. M. Tan, A. Shtukenberg, S. Zhu, W. Xu, E. Dooryhee, S.M. Nichols, M.D. Ward, B. Kahr, Q. Zhu, Roy revisited, again: the eighth solved structure. Faraday Discuss. **211**, 477–491 (2018)

58. E.O. Pyzer-Knapp, H.P.G. Thompson, F. Schiffmann, K.E. Jelfs, S.Y. Chong, M.A. Little, A.I. Cooper, G.M. Day, Predicted crystal energy landscapes of porous organic cages. Chem. Sci. **5**, 2235–2245 (2014)

59. F. Yang, S. Cheng, X. Zhang, X. Ren, R. Li, H. Dong, W. Hu, 2d organic materials for optoelectronic applications. Adv. Mater. **30**, 1702415 (2018)

60. H. Wondratschek, U. Müller, *Symmetry Relations Between Space Groups* (International Union of Crystallography, 2006)

61. M.I. Aroyo, J.M. Perez-Mato, C. Capillas, E. Kroumova, S. Ivantchev, G. Madariaga, A. Kirov, H. Wondratschek, Bilbao crystallographic server: I. databases and crystallographic

computing programs. Zeitschrift für Kristallographie-Crystalline Materials **221**, 15–27 (2006)

62. P.-P. Shi, Y.-Y. Tang, P.-F. Li, W.-Q. Liao, Z.-X. Wang, Q. Ye, R.-G. Xiong, Symmetry breaking in molecular ferroelectrics. Chem. Soc. Rev. **45**, 3811–3827 (2016)

63. J. Nocedal, S.J. Wright, *Numerical Optimization*, 2nd edn. (Springer, New York, 2006)

64. E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple. Phys. Rev. Lett. **97**, 170201 (2006)

65. C.J. Pickard, Hyperspatial optimization of structures. Phys. Rev. B **99**, 054102 (2019)

66. M.U. Schmidt, U. Englert, Prediction of crystal structures, J. Chem. Soc., Dalton Trans. 2077–2082 (1996)

67. C.J. Pickard, R. Needs, Ab initio random structure searching. J. Phys. **23**, 053201 (2011)

68. P.G. Karamertzanis, C.C. Pantelides, Ab initio crystal structure prediction—i. rigid molecules. J. Comput. Chem. **26**, 304–324 (2005)

69. D.H. Case, J.E. Campbell, P.J. Bygrave, G.M. Day, Convergence properties of crystal structure prediction by quasi-random sampling. J. Chem. Theory Comput. **12**, 910–924 (2016)

70. M.J. Kochenderfer, T.A. Wheeler, *Algorithms for Optimization* (Mit Press, Cambridge, 2019)

71. J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, V. Caignaert, Prediction of crystal structures from crystal chemistry rules by simulated annealing. Nature **346**, 343 (1990)

72. J.C. Schön, M. Jansen, First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization. Angew. Chem. Int. Ed. **35**, 1286–1304 (1996)

73. A. Banerjee, D. Jasrasaria, S.P. Niblett, D.J. Wales, Crystal structure prediction for benzene using basin-hopping global optimization. J. Phys. Chem. A **125**, 3776–3784 (2021)

74. P. Raiteri, R. Martoňák, M. Parrinello, Exploring polymorphism: the case of benzene. Angew. Chem. Int. Ed. **44**, 3769–3773 (2005)

75. T.-Q. Yu, M.E. Tuckerman, Temperature-accelerated method for exploring polymorphism in molecular crystals based on free energy. Phys. Rev. Lett. **107**, 015701 (2011)

76. A.R. Oganov, C.W. Glass, Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J. Chem. Phys. **124**, 244704 (2006)

77. Q. Zhu, A.R. Oganov, C.W. Glass, H.T. Stokes, Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. Acta Cryst. B **68**, 215–226 (2012)

78. F. Curtis, X. Li, T. Rose, Á. Vázquez-Mayagoitia, S. Bhattacharya, L.M. Ghiringhelli, N. Marom, Gator: a first-principles genetic algorithm for molecular crystal structure prediction. J. Chem. Theory Comput. **14**, 2246–2264 (2018)

79. A.R. Oganov, Crystal structure prediction: reflections on present status and challenges. Faraday Discuss. **211**, 643–660 (2018)

80. S. Habermehl, P. Mörschel, P. Eisenbrandt, S.M. Hammer, M.U. Schmidt, Structure determination from powder data without prior indexing, using a similarity measure based on cross-correlation functions. Acta Cryst. B **70**, 347–359 (2014)

81. J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field. J. Comput. Chem. **25**, 1157–1174 (2004)

82. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, Charmm: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. **4**, 187–217 (1983)

83. S.L. Mayo, B.D. Olafson, W.A. Goddard, Dreiding: a generic force field for molecular simulations. J. Phys. Chem. **94**, 8897–8909 (1990)

84. W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. **118**, 11225–11236 (1996)

85. S. Spicher, S. Grimme, Robust atomistic modeling of materials, organometallic, and biochemical systems. Angew. Chem. Int. Ed. **59**, 15665–15673 (2020)

86. S.L. Price, M. Leslie, G.W.A. Welch, L.S.P.M. Habgood, P.G. Karamertzanis, G.M. Day, Modelling organic crystal structures using distributed multipole and polarizabilities-based model intermolecular potentials. Phys. Chem. Chem. Phys. **12**, 8478–8490 (2010)

87. J.S. Smith, O. Isayev, A.E. Roitberg, Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. Chem. Sci. **8**, 3192–3203 (2017)

88. X. Li, F.S. Curtis, T. Rose, C. Schober, A. Vazquez-Mayagoitia, K. Reuter, H. Oberhofer, N. Marom, Genarris: random generation of molecular crystal structures and fast screening with a Harris approximation. J. Chem. Phys. **148**, 241701 (2018)

89. J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett. **98**, 146401 (2007)

90. J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys. Chem. Chem. Phys. **13**, 17930–17955 (2011)

91. A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments. Phys. Rev. B **87**, 184115 (2013)

92. C. Devereux, J.S. Smith, K.K. Davis, K. Barros, R. Zubatyuk, O. Isayev, A.E. Roitberg, Extending the applicability of the ani deep learning molecular potential to sulfur and halogens. J. Chem. Theory Comput. **16**, 4192–4202 (2020)

93. M.A. Neumann, Tailor-made force fields for crystal-structure prediction. J. Phys. Chem. B **112**, 9810–9829 (2008)

94. S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections. J. Comput. Chem. **25**, 1463–1473 (2004)

95. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J. Comput. Chem. **27**, 1787–1799 (2006)

96. A. Tkatchenko, M. Scheffler, Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. Phys. Rev. Lett. **102**, 073005 (2009)

97. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. J. Chem. Phys. **132**, 154104 (2010)

98. A. Otero-De-La-Roza, E.R. Johnson, A benchmark for noncovalent interactions in solids. J. Chem. Phys. **137**, 054103 (2012)

99. E. Caldeweyher, C. Bannwarth, S. Grimme, Extension of the d3 dispersion coefficient model. J. Chem. Phys. **147**, 034112 (2017)

100. M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B.I. Lundqvist, Van der Waals density functional for general geometries. Phys. Rev. Lett. **92**, 246401 (2004)

101. A. Tkatchenko, R.A. DiStasio, R. Car, M. Scheffler, Accurate and efficient method for many-body van der Waals interactions. Phys. Rev. Lett. **108**, 236402 (2012)

102. A.M. Reilly, A. Tkatchenko, Seamless and accurate modeling of organic molecular materials. J. Phys. Chem. Lett. **4**, 1028–1033 (2013)

103. B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Deshaye, T. Dumitrică, A. Dominguez et al., Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. J. Chem. Phys. **152**, 124101 (2020)

104. M. Mortazavi, J.G. Brandenburg, R.J. Maurer, A. Tkatchenko, Structure and stability of molecular crystals with many-body dispersion-inclusive density functional tight binding. J. Phys. Chem. Lett. **9**, 399–405 (2018)

105. M. Stöhr, L. Medrano Sandonas, A. Tkatchenko, Accurate many-body repulsive potentials for density-functional tight binding from deep tensor neural networks. J. Phys. Chem. Lett. **11**, 6835–6843 (2020)

106. C.H. Pham, R.K. Lindsey, L.E. Fried, N. Goldman, High-accuracy semiempirical quantum models based on a minimal training set. J. Phys. Chem. Lett. **13**, 2934–2942 (2022)

107. C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, Extended tight-binding quantum chemistry methods. WIREs Comput Mol Sci. **11**, e1493 (2021)

108. M. Neumann, J. Van De Streek, F. Fabbiani, P. Hidber, O. Grassmann, Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. Nat. Commun. **6**, 7793 (2015)

109. R.M. Bhardwaj, J.A. McMahon, J. Nyman, L.S. Price, S. Konar, I.D. Oswald, C.R. Pulham, S.L. Price, S.M. Reutzel-Edens, A prolific solvate former, Galunisertib, under the pressure of crystal structure prediction, produces ten diverse polymorphs. J. Am. Chem. Soc. **141**, 13887–13897 (2019)

110. W. Xu, Q. Zhu, C.T. Hu, The structure of glycine dihydrate: implications for the crystallization of glycine from solution and its structure in outer space. Angew. Chem. **129**, 2062–2066 (2017)

111. A. Ubbelohde, J.M. Robertson, A new form of resorcinol. Nature **140**, 239 (1937)

112. J.M. Robertson, A. Ubbelohde, A new form of resorcinol. i. structure determination by x-rays. Proc. R. Soc. Lond. Ser. A **167**, 122–135 (1938)

113. J.M. Robertson, A. Ubbelohde, A new form of resorcinol. ii. thermodynamic properties in relation to structure. Proc. R. Soc. Lond. Ser. A **167**, 136–147 (1938)

114. L. Yu, Polymorphis in molecular solids: an extraordinary system of red, orange, and yellow crystals. Acc. Chem. Res. **43**, 1257–1266 (2010)

115. X. Li, X. Ou, H. Rong, S. Huang, J. Nyman, L. Yu, M. Lu, The twelfth solved structure of roy: single crystals of y04 grown from melt microdroplets. Cryst. Growth Des. **20**, 7093–7097 (2020)

116. A. Levesque, T. Maris, J.D. Wuest, Roy reclaims its crown: new ways to increase polymorphic diversity. J. Am. Chem. Soc. **142**, 11873–11883 (2020)

117. K.S. Gushurst, J. Nyman, S.X. Boerrigter, The po13 crystal structure of roy. CrystEngComm **21**, 1363–1368 (2019)

118. N.P. Funnell, C.L. Bull, C.J. Ridley, S. Capelli, Structural behaviour of op-roy at extreme conditions. CrystEngComm **21**, 4473–4483 (2019)

119. G.J. Beran, I.J. Sugden, C. Greenwell, D. Bowskill, C.C. Pantelides, C. Adjiman, How many more polymorphs of roy remain undiscovered? Chem. Sci. **13**, 1288 (2022)

120. A.N. Sokolov, S. Atahan-Evrenk, R. Mondal, H.B. Akkerman, R.S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S.C. Mannsfeld, A.P. Zoombelt, Z. Bao et al., From computational discovery to experimental characterization of a high hole mobility organic crystal. Nat. Comm. **2**, 437 (2011)

121. J. Yang, S. De, J.E. Campbell, S. Li, M. Ceriotti, G.M. Day, Large-scale computational screening of molecular organic semiconductors using crystal structure prediction. Chem. Mater. **30**, 4361–4371 (2018)

122. F. Musil, S. De, J. Yang, J.E. Campbell, G.M. Day, M. Ceriotti, Machine learning for the structure-energy-property landscapes of molecular crystals. Chem. Sci. **9**, 1289–1300 (2018)

123. C.Y. Cheng, J.E. Campbell, G.M. Day, Evolutionary chemical space exploration for functional materials: computational

organic semiconductor discovery. Chem. Sci. **11**, 4922–4933 (2020)

124. M.J. Mehl, D. Hicks, C. Toher, O. Levy, R.M. Hanson, G. Hart, S. Curtarolo, The aflow library of crystallographic prototypes: part 1. Comput. Mater. Sci. **136**, S1–S828 (2017)

125. D. Hicks, M.J. Mehl, E. Gossett, C. Toher, O. Levy, R.M. Hanson, G. Hart, S. Curtarolo, The aflow library of crystallographic prototypes: part 2. Comput. Mater. Sci. **161**, S1–S1011 (2019)

126. W. Motherwell, Molecular shape and crystal packing: a database study. CrystEngComm **12**, 3554–3570 (2010)

127. Q. Zhu, J. Johal, D.E. Widdowson, Z. Pang, B. Li, C.M. Kane, V. Kurlin, G.M. Day, M.A. Little, A.I. Cooper, Analogy powered by prediction and structural invariants: computationally led discovery of a mesoporous hydrogen-bonded organic cage crystal. J. Am. Chem. Soc. **144**, 9893–9901 (2022)

128. K.M. Steed, J.W. Steed, Packing problems: high z prime crystal structures and their relationship to cocrystals, inclusion compounds, and polymorphism. Chem. Rev. **115**, 2895–2933 (2015)

129. C.P. Brock, Pseudosymmetric layers in high-z prime and p 1 structures of organic molecules. CrystEngComm **22**, 7371–7379 (2020)

130. A.D. Bond, Why do crystal structures waste molecular inversion symmetry? CrystEngComm **12**, 2492–2500 (2010)

131. S.L. Price, Is zeroth order crystal structure prediction (csp 0) coming to maturity? what should we aim for in an ideal crystal structure prediction code? Faraday Discuss. **211**, 9–30 (2018)

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

133. N.S. Abraham, M.R. Shirts, Statistical mechanical approximations to more efficiently determine polymorph free energy differences for small organic molecules. J. Chem. Theory Comput. **16**, 6503–6512 (2020)

134. S. Martiniani, K.J. Schrenk, J.D. Stevenson, D.J. Wales, D. Frenkel, Structural analysis of high-dimensional basins of attraction. Phys. Rev. E **94**, 031301 (2016)

135. V. Stevanović, Sampling polymorphs of ionic solids using random superlattices. Phys. Rev. Lett. **116**, 075503 (2016)

136. M. Deumal, M.J. Bearpark, J.J. Novoa, M.A. Robb, Magnetic properties of organic molecular crystals via an algebraic Heisenberg Hamiltonian. applications to Wilviw, Tolkek, and Kaxhas Nitronyl nitroxide crystals. J. Phys. Chem. A **106**, 1299–1315 (2002)

137. S.S. Borysov, B. Olsthoorn, M.B. Gedik, R.M. Geilhufe, A.V. Balatsky, Online search tool for graphical patterns in electronic band structures. NPJ Comput. Mater. **4**, 1–8 (2018)

138. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder et al., Commentary: The materials project: a materials genome approach to accelerating materials innovation. APL Mater. **1**, 011002 (2013)

139. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L. Hart, S. Sanvito, M. Buongiorno-Nardelli et al., Aflowlib. org: A distributed materials properties repository from high-throughput ab initio calculations. Comput. Mater. Sci. **58**, 227–235 (2012)

140. S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. NPJ Comput. Mat. **1**, 15010 (2015)

141. S.S. Borysov, R.M. Geilhufe, A.V. Balatsky, Organic materials database: an open-access online database for data mining. PLoS ONE **12**, e0171501 (2017)

142. A. Vaitkus, A. Merkys, S. Gražulis, Validation of the crystallography open database using the crystallographic information framework. J. Appl. Cryst. (2021). https://doi.org/10.1107/S1600576720016532

143. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu et al., Pubchem in 2021: new data content and improved web interfaces. Nucleic Acids Res. **49**, D1388–D1395 (2021)

144. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. Comput. Mater. Sci. **68**, 314–319 (2013)

145. A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus et al., The atomic simulation environment—a python library for working with atoms. J. Phys. **29**, 273002 (2017)

146. Y. Wang, J. Lv, L. Zhu, Y. Ma, Crystal structure prediction via particle-swarm optimization. Phys. Rev. B **82**, 094116 (2010)

147. D.C. Lonie, E. Zurek, Xtalopt: An open-source evolutionary algorithm for crystal structure prediction. Comput. Phys. Commun. **182**, 372–387 (2011)

148. B.C. Revard, W.W. Tipton, R.G. Hennig, Genetic algorithm for structure and phase prediction, https://github.com/hennig-group/GASP-python (2018)