

Genarris 3.0: Generating Close-Packed Molecular Crystal Structures with Rigid Press

Yi Yang,[†] Rithwik Tom,[‡] Jose A. G. L. Wui,^{¶,⊥} Jonathan E. Moussa,[§] and Noa Marom^{*,†,‡,||}

[†]*Department of Materials Science & Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[‡]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[¶]*Department of Physics, University of Texas at Austin, Austin, TX 78712, USA*

[§]*Molecular Sciences Software Institute, Virginia Tech, Blacksburg, VA 24060, USA*

^{||}*Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[⊥]*Current address: Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*

E-mail: nmarom@andrew.cmu.edu

Abstract

Polymorphism in molecular crystals influences their properties and performance. Crystal structure prediction (CSP) can help explore the crystal structure landscape and discover potentially stable polymorphs computationally. We present a new version of the Genarris open-source code, which generates random molecular crystal structures in all space groups and applies physical constraints on intermolecular distances. The main new feature in Genarris 3.0 is the “Rigid Press” algorithm, which uses a regularized hard-sphere potential to compress the unit cell and achieve a maximally close-packed structure based on purely geometric considerations without performing

any energy evaluations. In addition, Genarris 3.0 is interfaced with machine-learned interatomic potentials (MLIPs) to accelerate the exploration of the potential energy landscape. We present a new clustering and down-selection workflow that employs the MACE-OFF23(L) MLIPs to perform geometry optimization and energy ranking in the early stages. We use Genarris 3.0 to successfully predict the structure of six targets: aspirin, Target I and Target XXII from previous CSP blind tests, and the energetic materials HMX, CL-20, and DNI. We further analyze the performance of MACE-OFF23(L) compared to dispersion-inclusive density functional theory (DFT) for geometry relaxation and energy ranking. We find significant variability in the performance of MACE-OFF23(L) across chemically diverse targets with particularly poor performance for energetic materials, which is mitigated by our clustering and down-selection procedure. Genarris 3.0 can thus be used effectively to perform CSP and to generate molecular crystal datasets for training ML models.

Introduction

Molecular crystals are used for diverse applications including organic semiconductor devices,¹ energetic materials (EMs),^{2,3} pharmaceuticals,⁴ and agricultural chemicals.⁵ Because molecular crystals are held together by weak van der Waals interactions, they are prone to polymorphism,⁶ which is the ability of the same compound to crystallize into multiple crystal structures. Polymorphism has a far-reaching impact because different polymorphs can have markedly different physical, chemical, and mechanical properties. For example, crystal structure can influence the bioavailability and stability of pharmaceuticals,^{7,8} the sensitivity, detonation velocity, and safety of energetic materials,^{3,9–11} and the charge carrier mobility of organic semiconductors.^{12,13} Consequently, a comprehensive understanding of crystal structure landscapes and screening for polymorphs with desired properties is essential for the development of products based on molecular crystals. It can be time-consuming to perform exhaustive polymorph screening because minor variations in crystallization conditions

can alter the resulting crystal structure and some structures are difficult to crystallize.^{14–17} Computer simulations can provide guidance as to the possible presence of thermodynamically stable polymorphs, which have not yet been experimentally obtained. Indeed, computational crystal structure prediction (CSP) has become an integral part of the pharmaceutical development pipeline.^{18–21} Moreover, computer simulations can further predict the properties of putative crystal structures.^{22–29}

Computational CSP aims to predict all plausible polymorphs of a given compound. Advancements in CSP have been tracked through a series of blind tests organized by the Cambridge Crystallography Data Centre (CCDC).^{30–37} The CSP blind tests have both benchmarked and driven methodological improvements. In addition, they have highlighted the challenges faced by state-of-the-art CSP methods. Over the years, as CSP capabilities have evolved, the complexity of the target systems has increased. The field has progressed from relatively rigid small molecules to more flexible, larger molecules, and from single-component to multi-component crystals. As CSP targets become more complex, the configuration space that needs to be explored grows exponentially.^{38–40} This may require evaluating the relative stability of millions of putative structures. The difficulty is compounded by the fact that the energy differences between polymorphs are usually only a few kJ/mol,^{41,42} requiring high accuracy. The necessary accuracy can be achieved by dispersion-inclusive density functional theory (DFT),^{43–56} albeit at a high computational cost. Some intertwined challenges the CSP community is still grappling with are: predicting stability at finite temperatures,^{53,57–59} the so-called over-prediction problem, where structures corresponding to distinct local minima at 0 K correspond to the same (possibly disordered) structure at finite temperatures,⁶⁰ and crystallographic disorder, caused by multiple molecular conformations, orientations, or atomic positions within the unit cell.³⁶ Addressing these challenges would require going beyond lattice energy evaluations using dispersion-inclusive DFT at 0 K. This calls for the development of ranking methods that are both cost-effective and accurate for optimizing and evaluating the relative lattice energies of millions of candidate crystal structures.

Machine learning interatomic potentials (MLIPs) are considered as a promising route for achieving comparable accuracy to DFT at a significantly lower computational cost.^{61–68} To this end, MLIPs must be trained on large DFT datasets. Most of the available materials datasets are either of inorganic crystals with relatively small unit cells^{62,69–73} or of isolated small organic molecules.^{74–81} MLIPs have limited transferability outside of their training domains.⁸² The lag in the development of MLIPs for molecular crystals may therefore be attributed to the dearth of open datasets for molecular crystals. In order to perform well for molecular crystals, MLIPs must adequately capture intermolecular dispersion interactions. An alternative approach to training directly on molecular crystals is training on molecular datasets that include intramolecular dispersion interactions and/or intermolecular interactions between clusters of molecules. The resulting MLIPs, which capture short-range interactions, are then augmented with dispersion corrections, similar to DFT functionals.^{67,83–86}

The 7th CSP blind test was conducted in two phases, which ran from October 2020 to June 2022. The structure generation phase tested the ability of participants to generate the experimentally observed crystal structure starting from a molecular “stick diagram”.³⁶ The ranking phase tested the ability of participants to relax and rank lists of structures provided by the CCDC.³⁷ Our team (Group 16) used Genarris^{87,88} for crystal structure generation and system-specific AIMNet2^{67,89} MLIPs for geometry relaxation and energy ranking. Random or quasi-random crystal structure generation methods are frequently employed in CSP workflows to explore the potential energy surfaces (PES) of complex molecules with an unbiased sampling of crystal packing.^{90–93} Genarris generates random structures in all space groups compatible with the molecular symmetry and the requested number of molecules per unit cell (Z), including molecules occupying special Wyckoff positions. The target unit cell volume is determined by a machine learned model⁹⁴ and physical constraints are imposed on the intermolecular distances. The version of Genarris that was used in the 7th CSP blind test employed a preliminary implementation of the Rigid Press algorithm, described below, which uses a regularized hard-sphere potential to achieve close packing of molecules in the

unit cell. In the structure generation phase, system-specific AIMNet2 potentials were used to relax and rank millions of structures generated by Genarris. To the best of our knowledge, this was the earliest use (in 2020-2021) of MLIPs for molecular crystal structure prediction. We successfully generated four out of the six possible crystal structures for the targets we attempted, resulting in a success rate of 67%, which was the highest among academic teams and third overall.³⁶ In the ranking phase, our system-specific AIMNet2 potentials attained accuracy on par with dispersion-inclusive DFT methods at a fraction of the computational cost, and exceeded the performance of the MLIPs used by two other teams (Groups 12 and 15).³⁷ A detailed description of the system-specific AIMNet2 potentials and analysis of our results from the 7th CSP blind test is provided elsewhere.⁸⁹ Since the conclusion of the 7th CSP blind test, others have reported incorporating MLIPs for structure optimization and energy ranking in CSP workflows.^{95–97} Generative models^{98,99} and large language models (LLMs)^{100,101} are emerging as promising future approaches to structure generation.

Here, we introduce Genarris 3.0, the latest version of our open-source Python package for molecular crystal structure generation. We provide a detailed description of the Rigid Press algorithm featured in this version. Genarris 3.0 is interfaced with a variety of energy evaluation and relaxation methods via the Atomic Simulation Environment (ASE),¹⁰² providing the user maximal flexibility for choosing their preferred methods. Here, the MACE-OFF⁶¹ MLIPs are employed to accelerate energy evaluations and geometry relaxations. A new workflow for down-selection is presented to gradually reduce the number of candidate structures evaluated with increasingly computationally expensive and more accurate methods. The modular and extensible design of Genarris facilitates the integration of advanced methods for structure generation, optimization, and energy evaluations, as well as the implementation of user-defined workflows, thereby enhancing its capabilities in CSP.

To demonstrate the performance of Genarris 3.0, we have selected six diverse targets, shown in Figure 1. Aspirin (2-acetoxybenzoic acid) is a representative example of a hydrogen bonded crystal. It has two polymorphs, Form I and Form II (CSD reference codes ACSALA

and ACSALA17).^{103,104} Both forms have four molecules per unit cell ($Z = 4$) and crystallize in the monoclinic space group $P2_1/c$ (No. 14). Target I from the first CSP blind test (3,4-cyclobutylfuran)³⁰ has no strong intermolecular interactions. It has two known polymorphs: a stable form that crystallizes in the monoclinic space group $P2_1/c$ (No. 14) with $Z = 4$ and a metastable form that crystallizes in the orthorhombic space group $Pbca$ (No. 61) with $Z = 8$ (CSD reference codes XULDUD01 and XULDUD). Here, we focus on the structure with $Z = 8$, because its higher complexity and larger unit cell size provide a stringent test case for demonstrating the capability of our method to generate crystal structures with higher molecular packing complexity. Target XXII (tricyano-1,4-dithiino[*c*]-isothiazole) from the sixth CSP blind test³⁵ has unusual intermolecular interactions involving C, S, and N atoms. It crystallizes in the monoclinic space group $P2_1/n$ (No. 14) with $Z = 4$ (CSD reference code NACJAF).

In addition, we have selected three energetic materials (EMs). EMs are characterized by exceptionally dense crystal structures and strong intermolecular interactions between nitrogen-containing moieties.⁵⁴ Given that experiments on EMs are inherently risky, CSP represents a valuable approach for safely and effectively exploring their landscapes.^{10,11,105} CL-20 (2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazatetracyclo[5.5.0.0^{3,11}.0^{5,9}]dodecane) has several known polymorphs.^{106–108} Here, we focus on the most stable form, ε -CL-20 (CSD reference code PUBMUU02), which possesses the highest density, greatest detonation velocity, and superior impact stability.¹⁰⁹ The ε -CL-20 form crystallizes in the monoclinic space group $P2_1/n$ (No. 14) with four molecules per unit cell ($Z = 4$). HMX (1,3,5,7-tetranitro-1,3,5,7-tetrazocane) is highly polymorphic and exhibits multiple conformers across its four known forms.^{11,110–113} Here, we focus on δ -HMX (CSD reference code OCHTET03), which crystallizes in the hexagonal space group $P6_1$ (No. 169) to demonstrate structure generation with six molecules per unit cell ($Z = 6$). DNI (2,4-dinitroimidazole, CSD reference code TEVHEH01) has excellent detonation properties, lower sensitivity, and higher thermal stability compared to CL-20 and HMX.¹¹⁴ It crystallizes in the orthorhombic space group $Pbca$

(No. 61) with eight molecules per unit cell ($Z = 8$).¹¹⁵

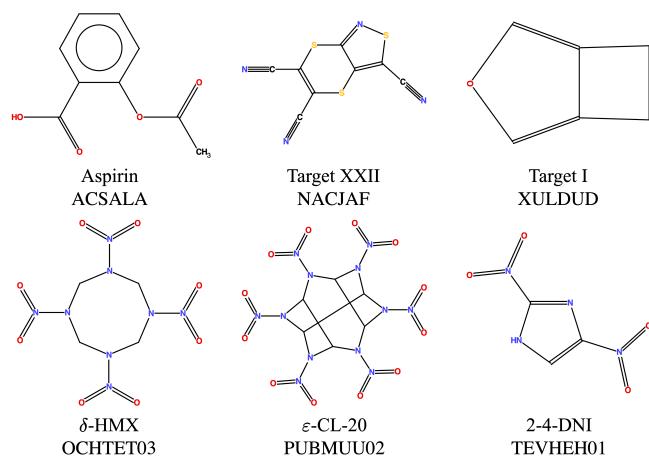


Figure 1: 2D molecular diagrams, common names, and CSD reference codes of the six CSP targets used here.

The experimental structures of all six targets are successfully generated by Genarris 3.0 and retained through the steps of the clustering and down-selection workflow. In the final stage of ranking with dispersion-inclusive DFT, the experimentally observed structures of all targets are ranked as the global minimum or the second lowest-energy structure. We find that MACE-OFF23(L) delivers variable performance for geometry relaxation and energy ranking across chemically diverse compounds. The performance for the energetic materials and Target XXII, whose chemistry is not well-represented in the training data, is worse than for aspirin and Target I. The new clustering and down-selection workflow implemented in Genarris 3.0 is able to mitigate the inconsistent performance of MACE-OFF23(L). This makes Genarris 3.0 a versatile, robust, and efficient code for CSP and for generating molecular crystal datasets for MLIPs training.

Methods

Workflow Overview

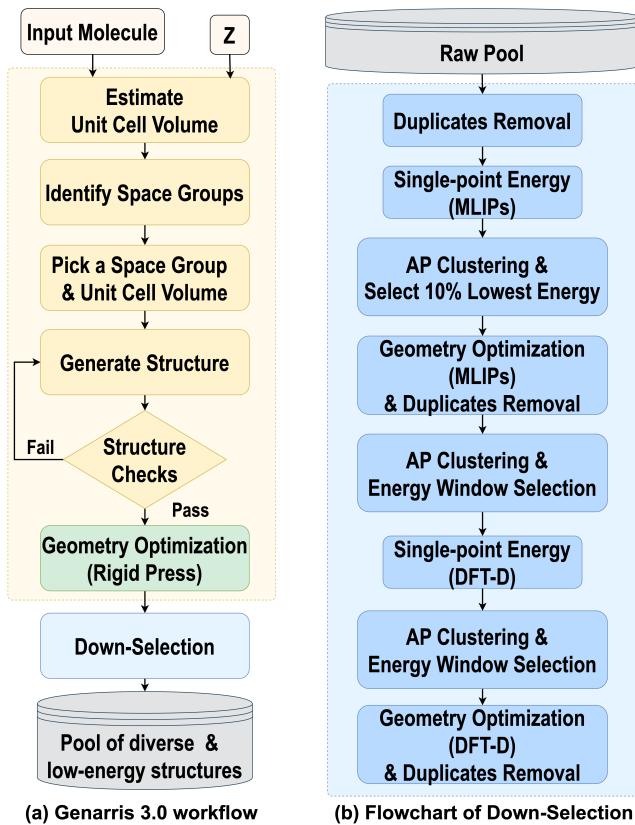


Figure 2: Schematic illustration of the workflow of Genarris 3.0: (a) The workflow of structure generation and (b) the down-selection workflow used here.

Figure 2a shows an overview of the CSP workflow used in this study. Genarris 3.0 starts from a molecular structure provided by the user. Genarris 3.0 does not perform conformational sampling. For flexible molecules, Genarris 3.0 can be used with an ensemble of conformers, as we have demonstrated in the 7th CSP blind test.⁸⁹ Here, we used the molecular conformation extracted from the CSD entry, relaxed using dispersion-inclusive DFT. Genarris identifies all space groups compatible with the requested number of molecules per unit cell (Z) and the molecular point group symmetry, including space groups with molecules occupying special Wyckoff positions.⁸⁸ Currently, Genarris 3.0 generates structures only with one molecule in the asymmetric unit ($Z' = 1$). The number of structures requested by the user is generated

in each compatible space group.

Structure generation starts by generating a unit cell with a volume within a normal distribution around a target value. Previously, Genarris 2.0 employed the target volume estimated by the PyMoVE machine learned model.⁹⁴ When using the Rigid Press algorithm (described below), the initial volume estimate is scaled by a factor of 1.5 to facilitate molecule placement. Molecules are placed in the unit cell as described in Ref.⁸⁸ The first molecule is randomly placed, while the remaining molecules are generated based on space group symmetries. If a molecule occupies a special Wyckoff position, it is aligned with the site symmetry. The generated structure is then checked to ensure that the interatomic distance d_{ij} between atoms i and j from different molecules is not less than $s_r \times (r_i^{vdW} + r_j^{vdW})$, where $r_{i/j}^{vdW}$ are the atomic van der Waals radii and s_r , is a user-defined fraction. Here, we set $s_r = 0.95$ to provide sufficient distance for subsequent Rigid Press optimization. Special intermolecular distance settings are applied to strong hydrogen bonds.⁸⁸ Structures that fail the proximity check are discarded. Structure generation continues until the targeted number of structures is reached. In this work, 4,000 crystal structures were generated in each compatible space group, forming the so-called “raw pool” of structures. All structures in the raw pool are initially optimized with Rigid Press. Subsequently, duplicate removal is performed within each space group by calculating the similarity via the Python Materials Genomics (pymatgen)¹¹⁶ `StructureMatcher` class, using 0.5 fractional length tolerance, 0.5 site tolerance, and 10° angle tolerance. These loose tolerances are used to efficiently discard many similar configurations, significantly reducing redundancy and computational cost in subsequent screening steps.

Next, a series of user-defined screening steps can be executed using increasingly more accurate and computationally expensive methods to gradually reduce the number of structures in the pool. The down-selection workflow may be varied depending on the user’s objective. For example, a workflow intended for generating data to train MLIPs may differ from a CSP workflow. The CSP workflow used here is shown in Figure 2b. The sequence

of clustering and selection steps is designed to balance considerations of structural diversity and energetic stability. For the structures remaining after Rigid Press optimization and duplicate removal, single-point energy (SPE) calculations are performed using MACE-OFF23(L). Afterward, affinity propagation (AP) clustering¹¹⁷ is performed with the target number of clusters set to 10% of the current structure pool. Genarris automatically adjusts the preference hyperparameter within the AP algorithm to achieve the desired number of clusters.⁸⁸ The lowest-energy structure from each cluster is selected. The selected structures are fully relaxed with MACE-OFF23(L), followed by an additional round of duplicate removal. Subsequently, AP clustering is performed again to produce 100 clusters. Up to 5 most stable structures within a 10 kJ/mol energy window are selected from each cluster. For the remaining structures, SPE evaluations are performed using dispersion-inclusive DFT. Then, AP clustering is performed to produce 100 clusters again, and all structures within a 10 kJ/mol energy window are selected from each cluster. Finally, the remaining structures are fully relaxed using dispersion-inclusive DFT and another round of duplicate removal is performed. This comprises the final pool of diverse and low-energy structures.

Genarris 3.0 incorporates significant code improvements. Enhanced modularity is achieved through Python’s Abstract Base Class (`abc`) module. This modular design simplifies the integration of new algorithms, energy evaluation and optimization methods without extensive modifications to the existing code base, thereby increasing flexibility and usability. Additionally, Genarris 3.0 features optimized multiprocessing capabilities, robust support for saving task checkpoints and restart functionality, enhanced process logging for improved monitoring and troubleshooting, and compatibility with GPU-accelerated MLIPs. These advancements substantially improve the computational efficiency and performance during the structure generation and ranking tasks.

Rigid Press

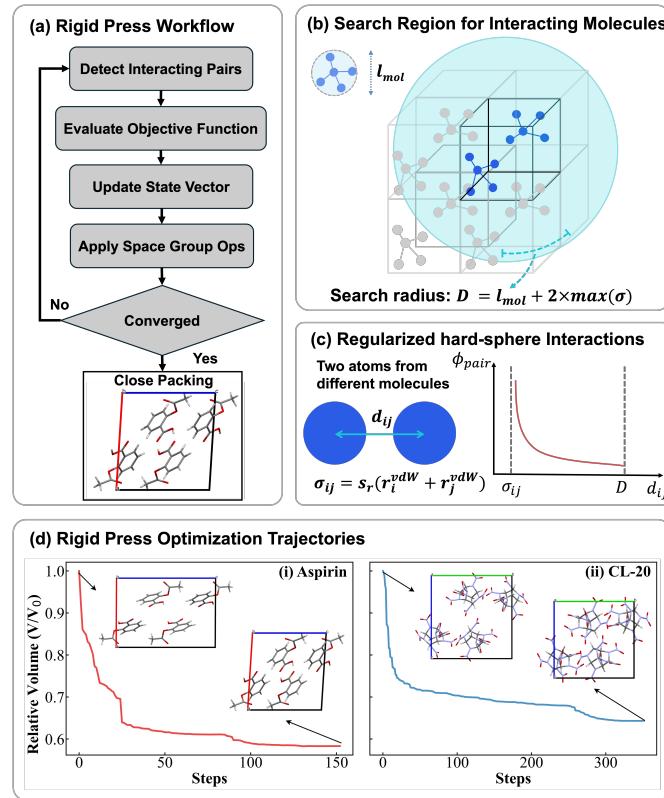


Figure 3: The Rigid Press algorithm: (a) Overall workflow; (b) Identification of interacting molecules; (c) The regularized hard-sphere interaction model; and (d) Representative optimization trajectories for aspirin Form I and ε -CL-20. The ratio of the unit cell volume, V , to the initial volume, V_0 , is plotted as a function of the number of optimization steps. The initial and final structures are also shown.

On the one hand, it may take a very large number of attempts to randomly generate close-packed molecular crystal structures while avoiding unphysical intermolecular contacts, which may lead to significant time spent on generating, checking, and discarding structures. On the other hand, increasing the target unit cell volume facilitates molecule placement, but significantly increases the time spent on relaxation of loosely packed molecular crystal structures. To address this challenge, we have developed the “Rigid Press” algorithm. Rigid Press uses a regularized hard-sphere potential to compress the unit cell based on purely geometric considerations without performing any energy evaluations. The workflow of Rigid Press is illustrated in Figure 3a. First, all the molecule pairs that are within the search radius to be

considered are identified (Figure 3b). Then, an objective function formulated to minimize the unit cell volume while maintaining physical intermolecular distances is evaluated (Figure 3c). The inherently non-differentiable hard-sphere interaction model is transformed into a smooth, differentiable function suitable for standard numerical optimization algorithms. The algorithm keeps the internal molecular geometry frozen (hence the name “Rigid Press”) as it simultaneously optimizes the molecular positions and orientations and the crystal lattice vectors to minimize the unit cell volume, while preserving the space group symmetries. Figure 3d shows representative Rigid Press optimization trajectories of aspirin Form I and ϵ -CL-20 (Rigid Press optimization trajectories for all other CSP targets are shown in the SI). The trajectories are characterized by a rapid initial volume reduction, indicating an effective compaction process from the initial structure (generated with an expanded volume) to the maximally close-packed final structure. The computational cost of Rigid Press optimization is lower by up to two orders of magnitude than relaxation using MLIPs.

Within Rigid Press, a molecular crystal is represented by a state vector \mathbf{s} , constructed to preserve the crystal’s space group symmetry during optimization. The state vector comprises independent lattice vectors, \mathbf{L} , according to the crystal system, the position of the asymmetric unit’s center of geometry, \mathbf{r}_{cog} , and the orientation of the asymmetric unit represented by Euler angles, θ . The objective function, $F(\mathbf{s})$, optimized by the Rigid Press algorithm is defined as:

$$F(\mathbf{s}) = V(\mathbf{s}) + P_{contact}(\mathbf{s}) \quad (1)$$

where $V(\mathbf{s})$ denotes the unit cell volume and $P_{contact}(\mathbf{s})$ is the contact penalty function, which is calculated by summing over atomic penalties from all interacting molecular pairs within a specified cutoff distance:

$$P_{contact} = \sum_{(A,B) \in \mathcal{N}} \sum_{i \in A} \sum_{j \in B} \phi_{pair}(d_{ij}, \sigma_{ij}) \quad (2)$$

Here, the set \mathcal{N} represents all molecular pairs that are sufficiently close for their interactions

to be considered, and $d_{ij} = \|\mathbf{r}_i^A(\mathbf{s}) - \mathbf{r}_j^B(\mathbf{s})\|$ is the distance between atoms i and j belonging to different molecules A and B . σ_{ij} is the hard-sphere diameter for the atom pair (i, j) , defined as a fraction (s_r) of the sum of their van der Waals radii. Here, the default value of s_r is 0.85, with specialized s_r values applied for hydrogen bonds. These s_r values were determined by statistical analyses of experimental structures in CSD.⁸⁸ To compute Eq. (2), all the molecule pairs within the interaction distance D need to be identified. A molecule can interact with other molecules in the unit cell, any of their periodic images or even its own periodic image. As shown in Figure b, the search can be limited to all the cells that are at an interacting distance D from the central cell. This is precomputed for a given state to reduce computational cost.

The pairwise interaction penalty $\phi_{pair}(d, \sigma)$ is defined in a piecewise manner to ensure differentiability:

$$\phi_{pair}(d, \sigma) = \begin{cases} \infty & \text{if } d \leq \sigma \\ w \cdot \frac{D-d}{d-\sigma} & \text{if } \sigma < d < D \\ 0 & \text{if } d \geq D \end{cases} \quad (3)$$

The maximum interaction distance, D , is defined as $D = l_{mol} + 2 \times \max(\sigma_{ij})$, where l_{mol} is the diameter of the smallest sphere enclosing the molecule, defined as twice the maximum distance from the molecular center of geometry to any atom, and $\max(\sigma_{ij})$ is the maximum interaction radius among all atom pairs. This is illustrated in Figure 3b. $w = k/N_{atoms}^2$ is a scaling factor that normalizes the contact penalty by the square of the number of atoms per molecule (N_{atoms}) to ensure appropriate scaling, irrespective of molecular size. The constant k controls the relative importance of the contact penalty in the objective function in Eq. (1). The default value is $k = 0.1$, which was determined empirically to provide a balanced contribution from interaction penalties. Users may adjust this value as needed based on specific use cases.

The final numerical optimization employs the The Broyden–Fletcher–Goldfarb–Shanno

(BFGS) algorithm¹¹⁸ implemented in the SciPy `optimize` class.¹¹⁹ Space group symmetries are preserved by reconstructing the full crystal coordinates from the optimized state vector \mathbf{s} after each optimization step. Specifically, symmetry operations corresponding to the space group are applied to the optimized asymmetric unit's position and orientation parameters, represented within the state vector \mathbf{s} , to generate the complete crystal structure, thus enforcing symmetry constraints. The iterative optimization continues until reaching predefined convergence criteria, with a default tolerance of 0.01 for the gradient norm, or a maximum iteration limit of 5,000. Both criteria can be customized by the user. Upon successful completion, the crystal structure is updated to reflect the optimized close-packed molecular arrangement. We have additionally implemented a faster version of the Rigid Press algorithm without symmetry constraints in C, which is interfaced with Python through Simplified Wrapper and Interface Generator (**SWIG**). Users may select the appropriate version based on their specific requirements.

Computational Details

We have interfaced Genarris 3.0 with various methods for geometry optimization and energy evaluation via the Atomic Simulation Environment (ASE).¹⁰² All dispersion-inclusive DFT calculations were performed using the FHI-aims all-electron electronic structure code^{120–122} (version 240507). For each target, the single molecule geometry was extracted from the experimental crystal structure in CSD. The single molecule geometry was then relaxed using the PBE0¹²³ hybrid functional, which is based on the Perdew–Burke–Ernzerhof (PBE)¹²⁴ generalized gradient approximation, combined with the many-body dispersion (MBD) method.^{125–127} Single point energy (SPE) evaluations with PBE+MBD were performed using the Tier 1 basis sets of FHI-aims and *light* numerical settings. Unit cell relaxations of the final structures using PBE+MBD were performed with the Tier 2 basis sets of FHI-aims and *tight* numerical settings. A $3 \times 3 \times 3$ k-points grid was used to sample the Brillouin zone.

The MLIPs employed here is the MACE-OFF⁶¹ pre-trained transferable organic force

field (OFF). MACE-OFF has three variants trained on the same SPICE 1.0 dataset:⁸¹ small (MACE-OFF23(S)), medium (MACE-OFF23(M)), and large (MACE-OFF23(L)), which differ mainly in the number of hyperparameters. Additionally, the large (L) variant employs an extended cutoff radius of 5 Å, utilizes more chemical channels ($k = 192$), and incorporates a higher maximum equivariant messages ($\max L = 2$). These enhancements enable the MACE-OFF23(L) model to better capture complex many-body effects and long-range interactions to achieve superior accuracy. However, this increased accuracy comes with a higher computational cost. Here, we selected MACE-OFF23(L) because benchmark tests on the X23b dataset¹²⁸ have indicated that it provides predictions comparable in accuracy to dispersion-inclusive DFT.

All geometry relaxations with MLIPs and DFT were performed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm implemented in ASE, with a force convergence criterion of 0.01 eV/Å. We also employed the ASE `FrechetCellFilter` to simultaneously adjust atom positions and the unit cell, along with the ASE constraint `FixSymmetry` to preserve space group symmetry.

To assess the similarity between the predicted and experimentally known crystal structures, we used the COMPACK molecular overlay method,¹²⁹ as implemented in the Crystal Packing Similarity features of the CSD Python API.¹³⁰ A crystal structure is represented by a cluster of N molecules comprised of a central reference molecule and (N-1) nearest-neighbor molecules. The root mean squared deviation (RMSD) between two molecular clusters is calculated based on the molecules that match within the specified tolerances. Here, we calculated the RMSD in the atomic positions for clusters of 30 molecules, labeled as $RMSD_{30}$. To this end, the number of matching molecules between shells of 30 molecules is extracted from the two crystal structures being compared, within 35% distance and 35° angle tolerances, excluding hydrogen atoms. This is the same comparison metric that was used in the seventh CSP blind test.^{36,37}

Results and Discussion

CSP Results

Table 1 summarizes the number of matches to the experimental structure out of the total number of generated structures in the pool at each stage of the CSP workflow. Figures 4 and 5 present the corresponding distributions of unit cell volume, lattice parameters, and space groups obtained at each stage. Similar figures for all other targets are provided in the SI.

Initially, structures are generated across all compatible space groups for each target, as indicated by the uniform space group distributions in Figures 4 and 5. For example, aspirin structures with $Z = 4$ were generated across 26 compatible space groups and DNI structures with $Z = 8$ were generated across 63 compatible space groups. Because the initial generation is performed with an increased target volume, the unit cell volume histograms are significantly overestimated compared to the experimental values at this stage. For most targets, no matches are found after initial generation. For Target XXII, one match is found out of 104,000 structures. For δ -HMX, 268 matches are found out of 52,573 generated structures. This higher match rate is likely because the highly constrained space group symmetry ($P6_1$, No. 169) limits the degrees of freedom for the molecular positions and orientation.

After optimization with Rigid Press, the unit cell volume histograms in Figures 4 and 5 are closer to the experimental values and matches are found for all targets. The space group histograms are unchanged because Rigid Press preserves the space group symmetry. For most targets, only a handful of matches are found out of $\sim 10^5$ generated structures. For Target I, 22 matches are found out of 248,000 structures. The higher number of matches may be attributed to the molecule's rigidity and the common $Pbca$ (No. 61) space group, which facilitates good packing. For δ -HMX, the number of matches increases to 1,430. There is a clear distinction between structures that are difficult to generate, and are generated very

rarely, such as DNI with a single match out of 232,025 structures, compared to structures that are easy to generate and are generated frequently, such as δ -HMX.

Table 1: Summary of the number of matches to the experimental structure out of the total number of structures in the pool at each step of the crystal structure prediction workflow for all six targets. For aspirin, matches to both polymorphs are counted.

CSP Workflow	Aspirin	Target XXII	Target I	δ -HMX	ε -CL-20	DNI
Initial generation	0/100,004	1/104,000	0/248,000	268/52,573	0/92,000	0/232,025
Rigid Press	4/100,004	4/104,000	22/248,000	1,430/52,573	6/92,000	1/232,025
Duplicate removal	2/18,528	1/11,916	2/12,356	3/2,767	1/11,860	1/24,065
AP clustering @MACE-OFF23 SPE	2/1,817	1/1,119	1/1,152	3/270	1/1,219	1/2,391
Relaxation @MACE-OFF23 & duplicate removal	2/1,567	1/1,031	1/914	1/212	1/1,118	1/1,759
AP clustering @MACE-OFF23	2/218	1/184	1/193	1/122	1/310	1/197
AP clustering @PBE+MBD SPE	2/128	1/108	1/98	1/103	1/143	1/124
Relaxation @PBE+MBD & duplicate removal	2/127	1/106	1/89	1/89	1/137	1/116

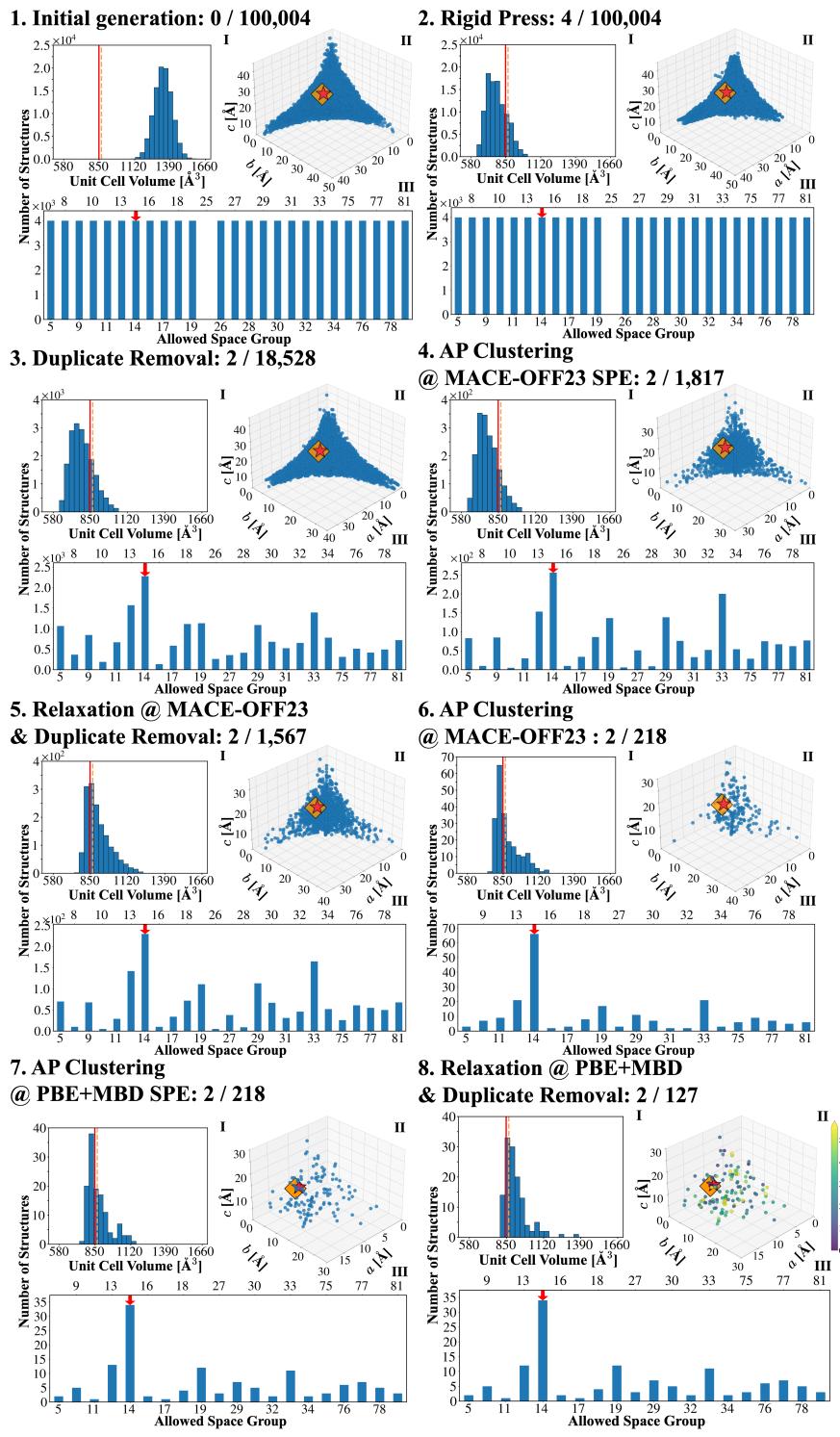


Figure 4: Distributions of unit cell volume, lattice parameters, and space groups, obtained at each step of the Genarris 3.0 workflow for aspirin with $Z = 4$. The experimental unit cell volume of Form I is indicated by a solid vertical red line and Form II is indicated by a dashed vertical orange line. The experimental lattice parameters of Form I and Form II are indicated by a red star and an orange square, respectively, and the experimental space group is indicated by a red arrow.

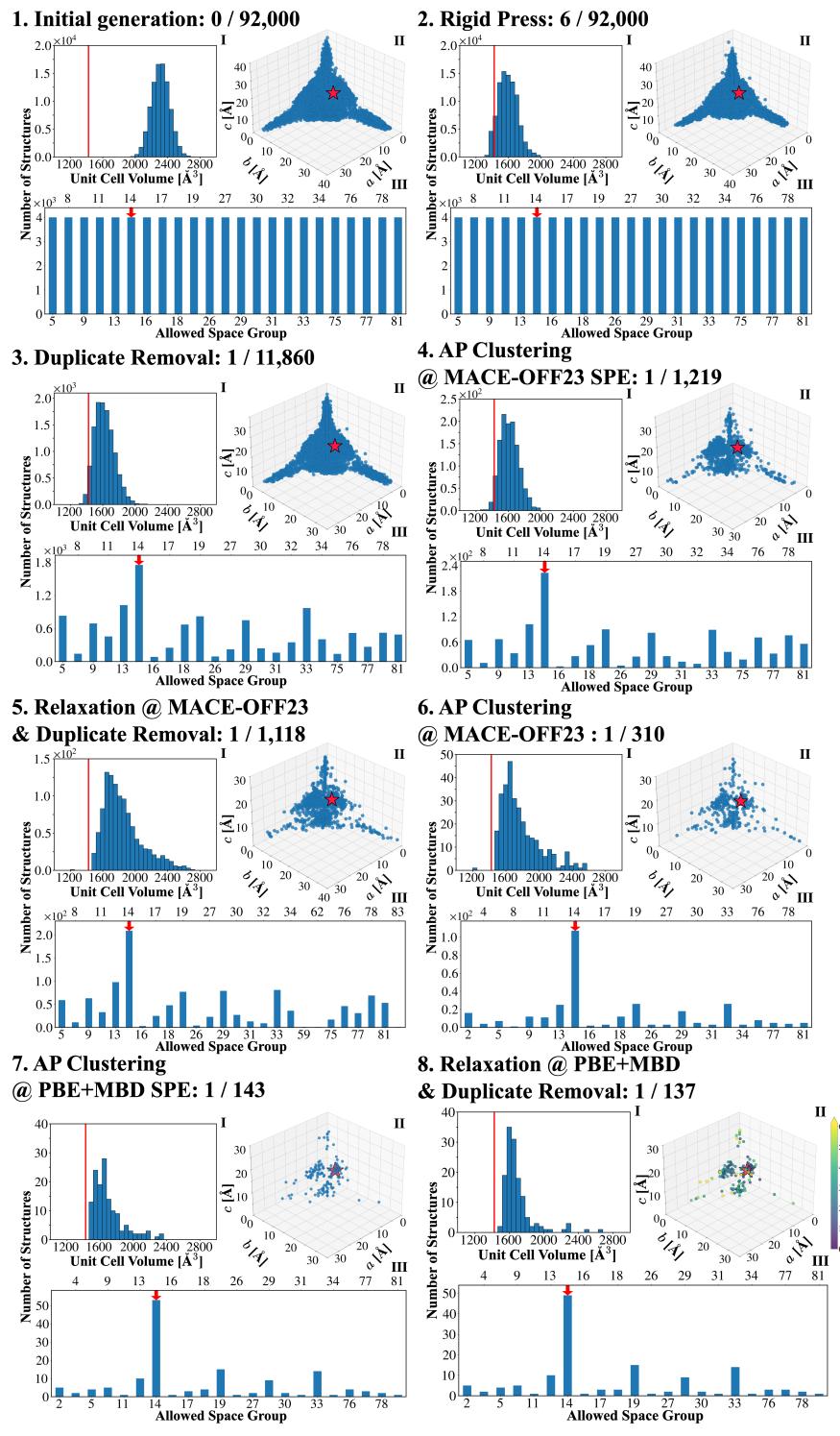


Figure 5: Distributions of unit cell volume, lattice parameters, and space groups, obtained at each step of the Genarris 3.0 workflow for ε -CL-20 with $Z = 4$. The experimental unit cell volume is indicated by a vertical red line, the experimental lattice parameters are indicated by a red star, and the experimental space group is indicated by a red arrow.

At this point, duplicate removal drastically reduces the number of structures in the pool without significantly changing the volume distributions. For Target XXII, ε -CL-20 and DNI, the reduction is by a factor of 8-10. For aspirin, the reduction is by a smaller factor of 5. The greatest reductions are for Target I and δ -HMX by a factor of 19-20. We consider a large number of duplicates as an indication that the configuration space is exhaustively sampled. To reduce the number of duplicates, the user can reduce the number of structures generated in each space group. After this step, the space group distributions are no longer uniform because more duplicates are generated in some space groups than in others. Certain space groups, such as $P2/m$ (No. 10) and $P222$ (No. 16), include more special Wyckoff sites, limiting the number of available general positions. As a result, when Genarris attempts to place molecules on the general Wyckoff position of these space groups fewer unique arrangements are possible. Additionally, tetragonal and orthorhombic crystal systems with higher-symmetry space groups (e.g., Nos. 75-81) impose stricter symmetry constraints, leading to fewer unique crystal packing arrangements. Symmetry elements such as mirror planes and inversion centers greatly increase the multiplicity of equivalent positions and thereby increase the number of duplicates generated. For instance, space group Cm (No. 8) is C -centered, causing each initial placement to generate multiple symmetry-equivalent structures. In all cases, a large number of structures are retained in the space group of the experimental structure(s). After duplicate removal only one match to the experimental structure remains for Target XXII, ε -CL-20, and DNI. For aspirin, one match remains for each polymorph. For Target I and δ -HMX, two and three matches are left, respectively.

After AP clustering and selection based on MACE-OFF23(L) SPE, the number of structures is reduced to 10% while retaining the matches to experiment for all targets. After full unit cell relaxation with MACE-OFF23(L), the unit cell volume histograms in Figures 4 and 5 shift to higher values. Duplicate removal further reduces the number of structures only slightly. This is an indication that the structures remaining after the first clustering and selection step are already unique and structurally diverse. In the two subsequent clustering

and selection steps, the number of remaining structures varies between targets, depending on the number of structures within the 10 kJ/mol energy window (see further discussion below). With each clustering and down-selection step, the volume distributions become narrower, while retaining a large number of structures in the experimental space groups.

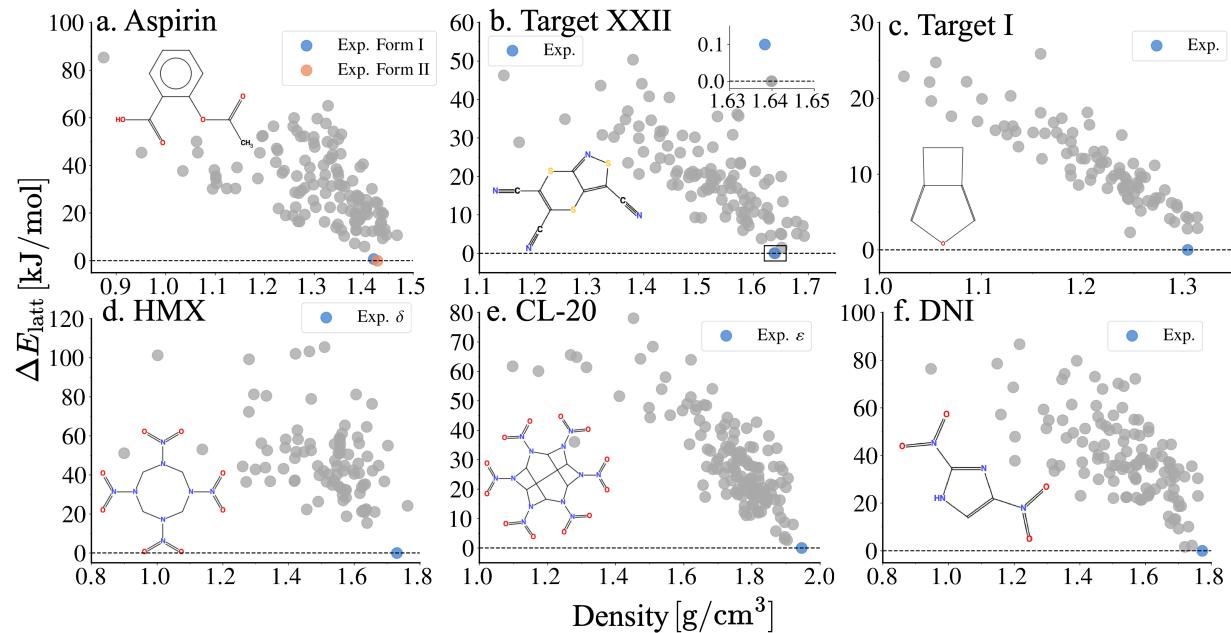


Figure 6: Energy as a function of density, calculated at the PBE+MBD level of theory for six benchmark targets: (a) Aspirin, (b) Target XXII, (c) Target I, (d) HMX, (e) CL-20, and (f) DNI. Experimentally observed polymorphs are highlighted in color, and putative crystal structures are shown in gray. The molecular structures are also shown.

Figure 6 shows the final potential energy landscapes obtained with PBE+MBD for all six targets. For all targets, except for Target XXII, the experimentally observed forms are ranked as the global minimum. For Target XXII, it has been shown previously that the experimental structure is ranked as the global minimum only with PBE0+MBD.¹³¹ For aspirin, Form II is predicted to be more stable than Form I by 0.69 kJ/mol. Experimental observations suggest that Form I is more stable than Form II at 300 K.¹³²⁻¹³⁴ Previous computational studies using dispersion-corrected DFT^{132,135} and fragment-based hybrid quantum classical methods^{136,137} have reported that the two polymorphs are very close in energy. These studies have also shown that the relative stability of Form I and Form II depends on the choice of method

and whether free energy corrections are applied. Figure S2 in the SI shows that PBE+MBD free energy at 300K, calculated using the quasi-harmonic approximation, as described in Ref.,⁵⁴ predicts Form I to be more stable than Form II by 1.10 kJ/mol. For the energetic materials CL-20 and DNI, several putative low-energy, high-density structures were found, with lattice energy 2.69 kJ/mol above the predicted ε form of CL-20 and 1.69 kJ/mol above the experimentally observed form of DNI.

MACE-OFF Performance

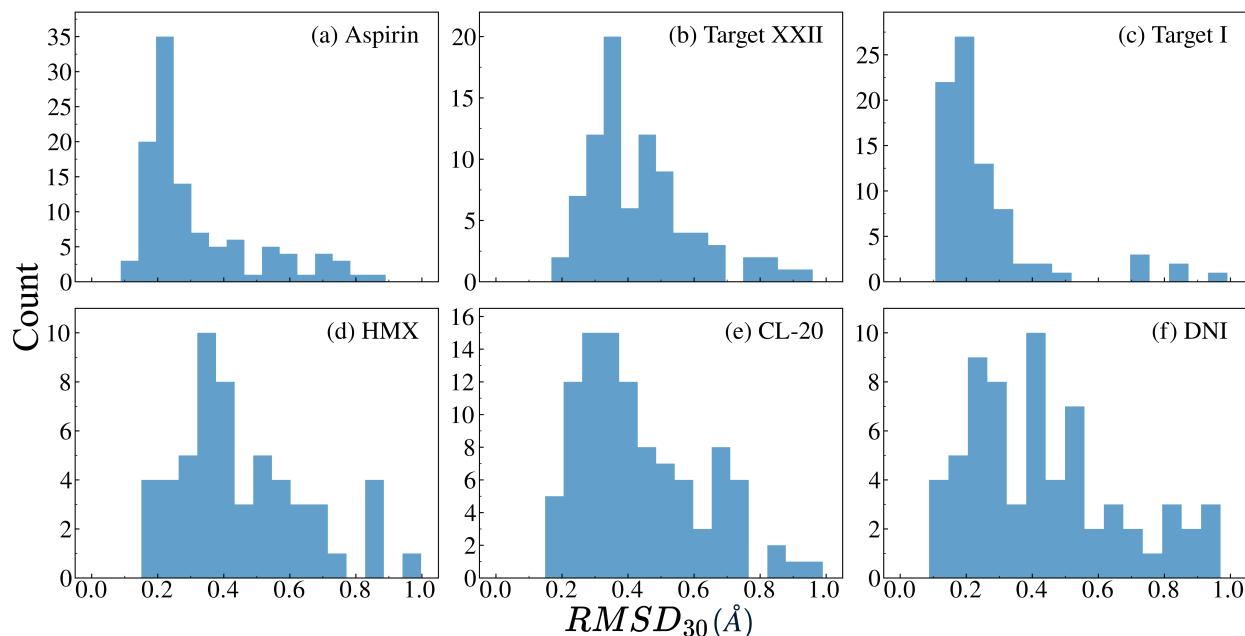


Figure 7: $RMSD_{30}$ histograms of the relaxed crystal structures obtained with the MACE-OFF23(L) model compared to those obtained with PBE+MBD, starting from the same initial configuration, for (a) Aspirin, (b) Target XXII, (c) Target I, (d) δ -HMX, (e) ε -CL-20, and (f) DNI.

In the following, we assess the performance of MACE-OFF23(L) for geometry relaxation and energy ranking by comparing the results with PBE+MBD, which we treat as the ground truth for putative crystal structures. Figure 7 shows the distributions of $RMSD_{30}$ values obtained by comparing the structures relaxed with MACE-OFF23(L) to the structures relaxed with PBE+MBD, starting from the same initial configuration. We consider structures as matching

if 30 molecules are overlaid and $RMSD_{30} < 1 \text{ \AA}$, indicating that both methods produce similar relaxed configurations. Only structures that match their DFT-relaxed counterparts based on the $RMSD_{30}$ values are included in our analysis. We use this as a metric for assessing how closely MACE-OFF23(L) reproduces the PBE+MBD potential energy surface (PES). If the PBE+MBD PES is reproduced well, then we expect MACE-OFF23(L) to arrive at the same local minimum structure. We note that in the ranking stage of the seventh CSP blind test, it was considered a failure if some of the relaxed structures obtained with some methods no longer matched the initial structures provided by the CCDC.³⁷ In particular, with some of the MLIPs used therein, the experimental structures of some of the targets could no longer be matched. In contrast, relaxation failures did not occur with any of the dispersion-inclusive DFT methods used therein.

There is significant variation in the relaxation performance of MACE-OFF23(L) across targets. For aspirin and Target I, the structures relaxed with MACE-OFF23(L) are largely in excellent agreement with PBE+MBD. The $RMSD_{30}$ histograms peak around 0.2 \AA and most structures have an $RMSD_{30}$ below 0.3 \AA . For aspirin, 17 mismatches occurred out of 127 structures, corresponding to 86.6 % match rate. For Target I, there were 8 mismatches out of 89 structures, a 91.0 % match rate. For Target XXII, the relaxation performance of MACE-OFF23(L) is somewhat worse. Its $RMSD_{30}$ histogram peaks around 0.35 \AA , with the majority of structures possessing $RMSD_{30}$ values below 0.6 \AA . Target XXII also has a somewhat lower match rate than aspirin and Target I, with 21 mismatches out of 106 structures amounting to 80.2 %.

For the three energetic materials, the relaxation performance of MACE-OFF23(L) is markedly worse. For δ -HMX, ε -CL-20, and DNI, the $RMSD_{30}$ distributions are broader, peak around $0.3 - 0.4 \text{ \AA}$, and a significant number of structures have $RMSD_{30}$ values above 0.6 \AA . The worse relaxation performance also manifests in a significantly lower match rates for these targets. For δ -HMX there were 34 mismatches out of 89 structures (61.8 %), for ε -CL-20 there were 36 mismatches out of 137 structures (73.7 %), and for DNI there

were 50 mismatches out of 116 structures (56.9 %). Across all targets, we observe a weak correlation between the relaxation performance and the relative lattice energies, where more stable structures tend to have lower $RMSD_{30}$ values, as shown in the SI.

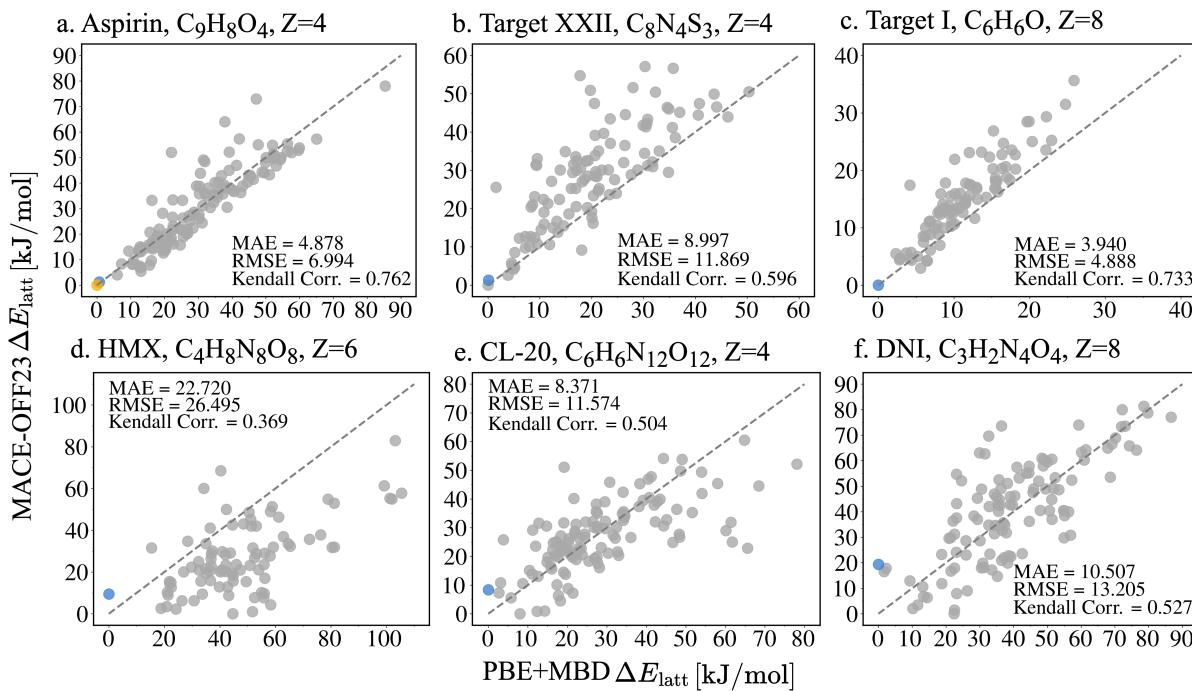


Figure 8: Relative lattice energies ΔE_{latt} obtained with the MACE-OFF23(L) model, compared to those calculated using PBE+MBD for (a) aspirin, (b) Target XXII, (c) Target I, (d) δ -HMX, (e) ε -CL-20, and (f) DNI. The experimentally observed structures are indicated in color. The mean absolute error (MAE), root mean squared error (RMSE), and Kendall correlation score are also shown.

In Figure 8, the performance of MACE-OFF23(L) in stability ranking is assessed by comparing the relative lattice energies of structures relaxed with MACE-OFF23(L) against those relaxed with PBE+MBD, which serve as the reference. For aspirin and Target I, MACE-OFF23(L) performs well. The MAE and RMSE values are below 5 kJ/mol and the Kendall ranking correlation score is above 0.7. It is also apparent in Figure 8a,c that the data points are concentrated quite close to the parity line. For both of these targets, MACE-OFF23(L) ranks the experimentally observed structures as the lowest in energy. For Target XXII, ε -CL-20, and DNI the performance of MACE-OFF23(L) is significantly worse, with

MAE and RMSE values ranging between 9-13 kJ/mol and Kendall ranking correlation scores of 0.5-0.6. It is also evident in Figure 8b, e, f that the data points are scattered farther away from the parity line compared to aspirin and Target I. The worst performance is found for δ -HMX with MAE and RMSE values above 20 kJ/mol and a Kendall ranking correlation score below 0.4. For Target XXII, the experimental structure is ranked as #2 in agreement with PBE+MBD. For the three energetic materials the experimental structures are ranked quite poorly by MACE-OFF23(L), as #12 for δ -HMX, #8 for ε -CL-20, and #19 for DNI. Similar performance trends are also evident from the comparison of the relative energy vs. density landscapes obtained with MACE-OFF23(L) to PBE+MBD, shown in the SI.

The variable performance of MACE-OFF23(L) can be attributed to the similarities and differences between our target molecules and the compounds contained in the SPICE training dataset. The SPICE 1.0 dataset mainly comprises drug-like molecules. This explains the good performance of the MACE-OFF23(L) model for the pharmaceutical target, aspirin, and Target I. The chemistry of Target XXII and, to a greater extent, the energetic molecules is very different from typical pharmaceutical compounds. Energetic materials, which feature a high concentration of nitrogen-containing groups, are underrepresented in the SPICE dataset. Our findings are in agreement with a recent study,¹³⁸ which also reported poor performance of MACE-OFF23(L) for molecules whose chemistry differs from the SPICE dataset, including Target XXII. These results highlight the limitations in the transferability of the MACE-OFF MLIPs across chemically diverse compounds.

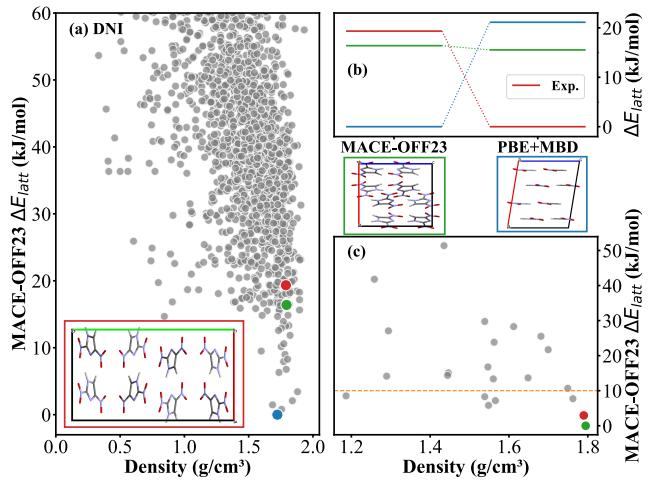


Figure 9: The clustering and down-selection workflow for DNI: (a) Relative lattice energies computed using MACE-OFF23(L) as a function of crystal density after relaxation with MACE-OFF23(L). The experimental structure (red), the MACE-OFF23(L) lowest energy structure (blue) and the lowest energy structure in the cluster containing the experimental structure (green) are highlighted. (b) Comparison of relative lattice energies computed with MACE-OFF23(L) and PBE+MBD for the three structures, which are also shown. (c) MACE-OFF23(L) relative energy as a function of crystal density for the cluster containing the experimental structure. The orange dashed line indicates the 10 kJ/mol energy threshold.

The clustering and down-selection workflow used here can mitigate to some extent the limitations of MACE-OFF23(L), as illustrated in Figure 9 for DNI. Figure 9a shows the landscape of relative energy as a function of density obtained after relaxation with MACE-OFF23(L) and duplicate removal (step 5 in Figure S8). The best match to the experimental structure (colored in red) is ranked as #79, 19.3 kJ/mol above the MACE-OFF23(L) lowest-energy structure. Single-point energy calculations using PBE+MBD on the MACE-OFF23(L) relaxed structures reveal significant changes in the relative energy ranking, as shown in Figure 9b. The experimental structure becomes the global minimum and the MACE-OFF23(L) lowest energy structure (colored in blue) is 21.13 kJ/mol higher in energy. This highlights that inaccuracies in the MACE-OFF23(L) energy ranking could potentially lead to the loss of important structures in the early stages of hierarchical CSP workflows that employ energy cutoffs to pass structures from one stage to the next. Here, the experimental structure is retained thanks to our clustering and down-selection approach. Figure

9c shows the MACE-OFF23(L) relative energy as a function of density for the cluster that contains the experimental structure after the AP clustering step. The experimental structure is ranked second in its cluster, after the structure colored in green. Because our procedure is to select up to 5 structures within a 10 kJ/mol window, rather than selecting only the most stable structure out of each cluster, the experimental structure is retained, despite the limitations of MACE-OFF23(L). This selection method enhances the robustness of our down-selection workflow. Similar analysis for all other targets is provided in the SI, showing that the clustering and down-selection procedure is particularly beneficial for the other two energetic targets δ -HMX and ε -CL-20, whose experimental structures are poorly ranked by MACE-OFF23(L).

Conclusion

In summary, we have presented a new version of our open-source molecular crystal structure generator, Genarris 3.0. In this version, we have implemented the Rigid Press algorithm, which efficiently generates close-packed molecular crystal structures by using a regularized hard-sphere potential to compress the unit cell, while preserving the space group symmetries. In addition, we have interfaced Genarris 3.0 through ASE with a variety of methods for geometry relaxation and energy evaluation, including DFT and MLIPs, offering the user maximal flexibility. We have introduced a new CSP workflow of clustering and down-selection to gradually reduce the number of structures evaluated with increasingly accurate and more computationally expensive methods. For demonstration purposes, we employed the MACE-OFF general-purpose MLIPs in the early stages of the workflow.

Genarris 3.0 successfully generated the experimentally observed crystal structures of the pharmaceutical aspirin, the two past blind test targets, Target I and Target XXII, and the three energetic materials δ -HMX, ε -CL-20, and DNI. The best matched structures were retained throughout the clustering and down-selection workflow. The MACE-OFF23(L)

delivered variable performance for relaxation and energy ranking across chemically diverse compounds. The performance for Target XXII and the energetic materials, whose chemistry is not well-represented in the SPICE 1.0 dataset, was worse than for aspirin and Target I. This has highlighted some limitations in the transferability of general-purpose MLIPs. We have demonstrated that our clustering and down-selection workflow was able to mitigate the inaccuracy of MACE-OFF23(L), especially for the energetic materials, whose experimental structures were significantly misranked.

Our results emphasize that although general-purpose MLIPs, such as MACE-OFF, can significantly accelerate early-stage CSP workflows, dispersion-inclusive DFT remains indispensable for accurate final ranking. Based on our findings, we suggest exercising caution when using general-purpose MLIPs for CSP. We recommend careful validation of the performance of general-purpose MLIPs on a case-by-case basis, especially if the chemistry of the materials of interest is significantly different than the materials represented in the training data. If their out-of-the-box performance is inadequate for the materials of interest, alternative solutions, such as system-specific AIMNet2 potentials⁸⁹ may be considered.

In conclusion, Genarris 3.0 is a versatile and robust open-source code for molecular crystal structure generation. Genarris 3.0 is able to generate structures in all space groups, including with structures occupying special Wyckoff positions. It offers the user maximal flexibility in the choice of method for relaxations and energy evaluations and in the design of CSP workflows. For flexible molecules, Genarris 3.0 may be started with an ensemble of conformers, as we had previously demonstrated within the seventh CSP blind test. Future improvements include generating structures with more than one molecule in the asymmetric unit. Genarris 3.0 may be used to perform CSP by random sampling, to generate initial structure pools for other CSP methods, and to generate datasets for MLIPs training.

Acknowledgement

This work was funded by the National Science Foundation (NSF) via award DMR-2131944. Y.Y acknowledges support from the Frontera Computational Science Fellowship awarded by the Texas Advanced Computing Center (TACC). R.T acknowledges support from the Molecular Sciences Software Institute (MolSSI) Fellowship. Computational resources were provided through the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) MAT210006 program with an allocation of 5 million CPU hours.

Code and Data availability

Genarris 3.0 package is available on GitHub (<https://github.com/Yi5817/Genarris>) and through the website (<https://www.noamarom.com/software/genarris/>) under the BSD-3-Clause license. The predicted candidate structures relaxed with MACE-OFF23(L) and PBE+MBD for 6 target molecules in this study are publicly available at (<https://github.com/Yi5817/Genarris>).

Supporting Information Available

Rigid Press optimization trajectories; Additional details on unit cell volumes, lattice parameters, and space groups distributions at each step of the CSP workflow in Genarris 3.0; Ranking of low-energy structures with different dispersion-inclusive DFT methods and thermal corrections; Tabulated lattice parameters and $RMSD_{30}$ for polymorphs obtained by MACE-OFF23(L) and PBE+MBD; Correlation analysis between $RMSD_{30}$ values and relative lattice energies; Density and relative lattice energy comparison between MACE-OFF23(L) and PBE+MBD; Additional analysis of the clustering and down-selection workflow mitigate limitations of MACE-OFF23(L).

References

- (1) Wang, C.; Dong, H.; Jiang, L.; Hu, W. Organic semiconductor crystals. *Chemical Society Reviews* **2018**, *47*, 422–500.
- (2) Zlotin, S. G.; Churakov, A. M.; Egorov, M. P.; Fershtat, L. L.; Klenov, M. S.; Kuchurov, I. V.; Makhova, N. N.; Smirnov, G. A.; Tomilov, Y. V.; Tartakovskiy, V. A. Advanced energetic materials: novel strategies and versatile applications. *Mendeleev Communications* **2021**, *31*, 731–749.
- (3) Liu, G.; Gou, R.; Li, H.; Zhang, C. Polymorphism of energetic materials: a comprehensive study of molecular conformers, crystal packing, and the dominance of their energetics in governing the most stable polymorph. *Crystal Growth & Design* **2018**, *18*, 4174–4186.
- (4) Lee, A. Y.; Erdemir, D.; Myerson, A. S. Crystal polymorphism in chemical process development. *Annual review of chemical and biomolecular engineering* **2011**, *2*, 259–280.
- (5) Yang, J.; Hu, C. T.; Zhu, X.; Zhu, Q.; Ward, M. D.; Kahr, B. DDT polymorphism and the lethality of crystal forms. *Angewandte Chemie* **2017**, *129*, 10299–10303.
- (6) Bernstein, J. *Polymorphism in molecular crystals*, 2nd ed.; Oxford University Press: Oxford, 2020; Vol. 30.
- (7) Yu, L. X.; Furness, M. S.; Raw, A.; Outlaw, K. P. W.; Nashed, N. E.; Ramos, E.; Miller, S. P.; Adams, R. C.; Fang, F.; Patel, R. M., et al. Scientific considerations of pharmaceutical solid polymorphism in abbreviated new drug applications. *Pharmaceutical research* **2003**, *20*, 531–536.
- (8) Censi, R.; Di Martino, P. Polymorph impact on the bioavailability and stability of poorly soluble drugs. *Molecules* **2015**, *20*, 18759–18776.

- (9) Michalchuk, A. A.; Trestman, M.; Rudić, S.; Portius, P.; Fincham, P. T.; Pulham, C. R.; Morrison, C. A. Predicting the reactivity of energetic materials: an ab initio multi-phonon approach. *Journal of Materials Chemistry A* **2019**, *7*, 19539–19553.
- (10) Bier, I.; O'Connor, D.; Hsieh, Y.-T.; Wen, W.; Hiszpanski, A. M.; Han, T. Y.-J.; Marom, N. Crystal structure prediction of energetic materials and a twisted arene with Genarris and GAtor. *CrystEngComm* **2021**, *23*, 6023–6038.
- (11) Arnold, J. E.; Day, G. M. Crystal Structure Prediction of Energetic Materials. *Crystal Growth & Design* **2023**, *23*, 6149–6160.
- (12) Coropceanu, V.; Cornil, J.; da Silva Filho, D. A.; Olivier, Y.; Silbey, R.; Brédas, J.-L. Charge transport in organic semiconductors. *Chemical reviews* **2007**, *107*, 926–952.
- (13) Chung, H.; Diao, Y. Polymorphism as an emerging design strategy for high performance organic electronics. *Journal of Materials Chemistry C* **2016**, *4*, 3915–3933.
- (14) Neumann, M.; Van De Streek, J.; Fabbiani, F.; Hidber, P.; Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nature Communications* **2015**, *6*, 7793.
- (15) Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J., et al. Functional materials discovery using energy–structure–function maps. *Nature* **2017**, *543*, 657–664.
- (16) Mortazavi, M.; Hoja, J.; Aerts, L.; Quéré, L.; van de Streek, J.; Neumann, M. A.; Tkatchenko, A. Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine. *Communications Chemistry* **2019**, *2*, 70.
- (17) Gui, Y.; Jin, Y.; Ruan, S.; Sun, G.; López-Mejías, V.; Yu, L. Crystal energy landscape

of nifedipine by experiment and computer prediction. *Crystal Growth & Design* **2022**, *22*, 1365–1370.

- (18) Bhardwaj, R. M.; Price, L. S.; Price, S. L.; Reutzel-Edens, S. M.; Miller, G. J.; Oswald, I. D.; Johnston, B. F.; Florence, A. J. Exploring the experimental and computed crystal energy landscape of olanzapine. *Crystal Growth & Design* **2013**, *13*, 1602–1617.
- (19) Braun, D. E.; McMahon, J. A.; Koztecki, L. H.; Price, S. L.; Reutzel-Edens, S. M. Contrasting polymorphism of related small molecule drugs correlated and guided by the computed crystal energy landscape. *Crystal growth & design* **2014**, *14*, 2056–2072.
- (20) Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications* **2016**, *52*, 7065–7077.
- (21) Bannan, C. C.; Ovanesyan, G.; Darden, T. A.; Graves, A. P.; Edge, C. M.; Russo, L.; Copley, R. C.; Manas, E.; Skillman, A. G.; Nicholls, A., et al. Crystal Structure Prediction of Drug Molecules in the Cloud: A Collaborative Blind Challenge Study. *Crystal Growth & Design* **2025**,
- (22) Rice, B.; LeBlanc, L. M.; Otero-de-la Roza, A.; Fuchter, M. J.; Johnson, E. R.; Nelson, J.; Jelfs, K. E. A computational exploration of the crystal energy and charge-carrier mobility landscapes of the chiral [6] helicene molecule. *Nanoscale* **2018**, *10*, 1865–1876.
- (23) Yang, J.; De, S.; Campbell, J. E.; Li, S.; Ceriotti, M.; Day, G. M. Large-scale computational screening of molecular organic semiconductors using crystal structure prediction. *Chemistry of Materials* **2018**, *30*, 4361–4371.
- (24) Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine learning for the structure–energy–property landscapes of molecular crystals. *Chemical science* **2018**, *9*, 1289–1300.

- (25) Schmidt, J. A.; Weatherby, J. A.; Sugden, I. J.; Santana-Bonilla, A.; Salerno, F.; Fuchter, M. J.; Johnson, E. R.; Nelson, J.; Jelfs, K. E. Computational screening of chiral organic semiconductors: exploring side-group functionalization and assembly to optimize charge transport. *Crystal Growth & Design* **2021**, *21*, 5036–5049.
- (26) Tom, R.; Gao, S.; Yang, Y.; Zhao, K.; Bier, I.; Buchanan, E. A.; Zaykov, A.; Havlas, Z.; Michl, J.; Marom, N. Inverse design of tetracene polymorphs with enhanced singlet fission performance by property-based genetic algorithm optimization. *Chemistry of Materials* **2023**, *35*, 1373–1386.
- (27) Bhat, V.; Callaway, C. P.; Risko, C. Computational approaches for organic semiconductors: from chemical and physical understanding to predicting new materials. *Chemical Reviews* **2023**, *123*, 7498–7547.
- (28) Faruque, M. O.; Akter, S.; Limbu, D. K.; Kilway, K. V.; Peng, Z.; Momeni, M. R. High-Throughput Screening, Crystal Structure Prediction, and Carrier Mobility Calculations of Organic Molecular Semiconductors as Hole Transport Layer Materials in Perovskite Solar Cells. *Crystal Growth & Design* **2024**, *24*, 8950–8960.
- (29) Johal, J.; Day, G. Exploring organic chemical space for materials discovery using crystal structure prediction-informed evolutionary optimisation. **2025**,
- (30) Lommerse, J. P.; Motherwell, W. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W.; Leusen, F. J.; Mooij, W. T.; Price, S. L.; Schweizer, B., et al. A test of crystal structure prediction of small organic molecules. *Structural Science* **2000**, *56*, 697–714.
- (31) Motherwell, W. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W.; Leusen, F. J.; Lommerse, J. P.; Mooij, W. T., et al. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 647–661.

- (32) Day, G.; Motherwell, W.; Ammon, H.; Boerrigter, S.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; Van Eijck, B., et al. A third blind test of crystal structure prediction. *Acta Crystallographica Section B: Structural Science* **2005**, *61*, 511–527.
- (33) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J., et al. Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test. *Acta Crystallographica Section B: Structural Science* **2009**, *65*, 107–125.
- (34) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R., et al. Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. *Acta Crystallographica Section B: Structural Science* **2011**, *67*, 535–551.
- (35) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R., et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016**, *72*, 439–459.
- (36) Hunnisett, L. M.; Nyman, J.; Francia, N.; Abraham, N. S.; Adjiman, C. S.; Aitipamula, S.; Alkhidir, T.; Almehairbi, M.; Anelli, A.; Anstine, D. M., et al. The seventh blind test of crystal structure prediction: structure generation methods. *Structural Science* **2024**, *80*.
- (37) Hunnisett, L. M.; Francia, N.; Nyman, J.; Abraham, N. S.; Aitipamula, S.; Alkhidir, T.; Almehairbi, M.; Anelli, A.; Anstine, D. M.; Anthony, J. E., et al. The sev-

enth blind test of crystal structure prediction: structure ranking methods. *Structural Science* **2024**, *80*.

- (38) Hawkins, P. C. Conformation generation: the state of the art. *Journal of chemical information and modeling* **2017**, *57*, 1747–1756.
- (39) Zhu, Q.; Hattori, S. Organic crystal structure prediction and its application to materials design. *Journal of Materials Research* **2023**, *38*, 19–36.
- (40) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C., et al. CREST—A program for the exploration of low-energy molecular chemical space. *The Journal of Chemical Physics* **2024**, *160*.
- (41) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.
- (42) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and fictions about polymorphism. *Chemical Society Reviews* **2015**, *44*, 8619–8635.
- (43) Otero-De-La-Roza, A.; Johnson, E. R. A benchmark for non-covalent interactions in solids. *The Journal of chemical physics* **2012**, *137*.
- (44) Marom, N.; DiStasio Jr, R. A.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A. Many-body dispersion interactions in molecular crystal polymorphism. *Angewandte Chemie International Edition* **2013**, *52*, 6629–6632.
- (45) Reilly, A. M.; Tkatchenko, A. Seamless and accurate modeling of organic molecular materials. *The journal of physical chemistry letters* **2013**, *4*, 1028–1033.
- (46) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *The Journal of chemical physics* **2013**, *139*.

- (47) Beran, G. J. Modeling polymorphic molecular crystals with electronic structure theory. *Chemical reviews* **2016**, *116*, 5567–5613.
- (48) Whittleton, S. R.; Otero-De-La-Roza, A.; Johnson, E. R. Exchange-hole dipole dispersion model for accurate energy ranking in molecular crystal structure prediction ii: Nonplanar molecules. *Journal of Chemical Theory and Computation* **2017**, *13*, 5332–5342.
- (49) Hermann, J.; DiStasio Jr, R. A.; Tkatchenko, A. First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications. *Chemical Reviews* **2017**, *117*, 4714–4758.
- (50) Hoja, J.; Reilly, A. M.; Tkatchenko, A. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, *7*, e1294.
- (51) Hoja, J.; Tkatchenko, A. First-principles stability ranking of molecular crystal polymorphs with the DFT+ MBD approach. *Faraday discussions* **2018**, *211*, 253–274.
- (52) Dolgonos, G. A.; Hoja, J.; Boese, A. D. Revised values for the X23 benchmark set of molecular crystals. *Physical Chemistry Chemical Physics* **2019**, *21*, 24333–24344.
- (53) Hoja, J.; Ko, H.-Y.; Neumann, M. A.; Car, R.; DiStasio Jr, R. A.; Tkatchenko, A. Reliable and practical computational description of molecular crystal polymorphs. *Science Advances* **2019**, *5*, eaau3338.
- (54) O'Connor, D.; Bier, I.; Hsieh, Y.-T.; Marom, N. Performance of dispersion-inclusive density functional theory methods for energetic materials. *Journal of Chemical Theory and Computation* **2022**, *18*, 4456–4471.
- (55) Beran, G. J.; Sugden, I. J.; Greenwell, C.; Bowskill, D. H.; Pantelides, C. C.; Ad-

- jiman, C. S. How many more polymorphs of ROY remain undiscovered. *Chemical science* **2022**, *13*, 1288–1297.
- (56) Price, A. J.; Otero-de-la Roza, A.; Johnson, E. R. XDM-corrected hybrid DFT with numerical atomic orbitals predicts molecular crystal lattice energies with unprecedented accuracy. *Chemical Science* **2023**, *14*, 1252–1262.
- (57) Price, S. L. Is zeroth order crystal structure prediction (CSP_0) coming to maturity? What should we aim for in an ideal crystal structure prediction code? *Faraday discussions* **2018**, *211*, 9–30.
- (58) Yang, M.; Dybeck, E.; Sun, G.; Peng, C.; Samas, B.; Burger, V. M.; Zeng, Q.; Jin, Y.; Bellucci, M. A.; Liu, Y., et al. Prediction of the relative free energies of drug polymorphs above zero kelvin. *Crystal Growth & Design* **2020**, *20*, 5211–5224.
- (59) Firaha, D.; Liu, Y. M.; van de Streek, J.; Sasikumar, K.; Dietrich, H.; Helfferich, J.; Aerts, L.; Braun, D. E.; Broo, A.; DiPasquale, A. G., et al. Predicting crystal form stability under real-world conditions. *Nature* **2023**, *623*, 324–328.
- (60) Francia, N. F.; Price, L. S.; Nyman, J.; Price, S. L.; Salvalaglio, M. Systematic Finite-Temperature Reduction of Crystal Energy Landscapes. *Crystal Growth & Design* **2020**, *20*, 6847–6862.
- (61) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Pu, Y.; Kapil, V.; Witt, W. C.; Magdau, I.-B.; Cole, D. J., et al. Mace-off: Short-range transferable machine learning force fields for organic molecules. *Journal of the American Chemical Society* **2025**,
- (62) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **2023**, *5*, 1031–1041.

- (63) Liao, Y.-L.; Wood, B.; Das, A.; Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059* **2023**,
- (64) Liao, Y.-L.; Smidt, T.; Shuaibi, M.; Das, A. Generalizing denoising to non-equilibrium structures improves equivariant force fields. *arXiv preprint arXiv:2403.09549* **2024**,
- (65) Yang, H.; Hu, C.; Zhou, Y.; Liu, X.; Shi, Y.; Li, J.; Li, G.; Chen, Z.; Chen, S.; Zeni, C., et al. Mattersim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967* **2024**,
- (66) Neumann, M.; Gin, J.; Rhodes, B.; Bennett, S.; Li, Z.; Choubisa, H.; Hussey, A.; Godwin, J. Orb: A fast, scalable neural network potential. *arXiv preprint arXiv:2410.22570* **2024**,
- (67) Isayev, O.; Anstine, D.; Zubaiuk, R. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *Chemical Science* **2025**,
- (68) Fu, X.; Wood, B. M.; Barroso-Luque, L.; Levine, D. S.; Gao, M.; Dzamba, M.; Zitnick, C. L. Learning smooth and expressive interatomic potentials for physical property prediction. *arXiv preprint arXiv:2502.12147* **2025**,
- (69) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G., et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials* **2013**, 1.
- (70) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, 1, 1–15.

- (71) Barroso-Luque, L.; Shuaibi, M.; Fu, X.; Wood, B. M.; Dzamba, M.; Gao, M.; Rizvi, A.; Zitnick, C. L.; Ulissi, Z. W. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771* **2024**,
- (72) Schmidt, J.; Cerqueira, T. F.; Romero, A. H.; Loew, A.; Jäger, F.; Wang, H.-C.; Botti, S.; Marques, M. A. Improving machine-learning models in materials science through large datasets. *Materials Today Physics* **2024**, *48*, 101560.
- (73) Žugec, I.; Geilhufe, R. M.; Lončarić, I. Global machine learning potentials for molecular crystals. *The Journal of chemical physics* **2024**, *160*.
- (74) Blum, L. C.; Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society* **2009**, *131*, 8732–8733.
- (75) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of chemical information and modeling* **2012**, *52*, 2864–2875.
- (76) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data* **2014**, *1*, 1–7.
- (77) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific data* **2017**, *4*, 1–8.
- (78) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data* **2020**, *7*, 134.
- (79) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the applicability of the ANI deep learning molecular po-

tential to sulfur and halogens. *Journal of chemical theory and computation* **2020**, *16*, 4192–4202.

- (80) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of chemical physics* **2020**, *153*.
- (81) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y., et al. Spice, a dataset of drug-like molecules and peptides for training machine learning potentials. *Scientific Data* **2023**, *10*, 11.
- (82) O mee, S. S.; Fu, N.; Dong, R.; Hu, M.; Hu, J. Structure-based out-of-distribution (OOD) materials property prediction: a benchmark study. *npj Computational Materials* **2024**, *10*, 144.
- (83) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical science* **2018**, *9*, 2261–2269.
- (84) Unke, O. T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation* **2019**, *15*, 3678–3693.
- (85) Westermayr, J.; Chaudhuri, S.; Jeindl, A.; Hofmann, O. T.; Maurer, R. J. Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic–inorganic interfaces. *Digital Discovery* **2022**, *1*, 463–475.
- (86) Anstine, D. M.; Isayev, O. Machine learning interatomic potentials and long-range physics. *The Journal of Physical Chemistry A* **2023**, *127*, 2417–2431.

- (87) Li, X.; Curtis, F. S.; Rose, T.; Schober, C.; Vazquez-Mayagoitia, A.; Reuter, K.; Oberhofer, H.; Marom, N. Genarris: Random generation of molecular crystal structures and fast screening with a Harris approximation. *The Journal of Chemical Physics* **2018**, *148*.
- (88) Tom, R.; Rose, T.; Bier, I.; O'Brien, H.; Vázquez-Mayagoitia, Á.; Marom, N. Genarris 2.0: A random structure generator for molecular crystals. *Computer Physics Communications* **2020**, *250*, 107170.
- (89) Nayal, K. S.; O'Connor, D.; Zubatyuk, R.; Anstine, D. M.; Yang, Y.; Tom, R.; Deng, W.; Tang, K.; Marom, N.; Isayev, O. Efficient Molecular Crystal Structure Prediction and Stability Assessment with AIMNet2 Neural Network Potentials. *ChemRxiv* **2025**, DOI: <https://doi.org/10.26434/chemrxiv-2025-ksn4n>.
- (90) Sugden, I. J.; Adjiman, C. S.; Pantelides, C. C. Accurate and efficient representation of intramolecular energy in ab initio generation of crystal structures. II. Smoothed intramolecular potentials. *Structural Science* **2019**, *75*, 423–433.
- (91) Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. Convergence properties of crystal structure prediction by quasi-random sampling. *Journal of chemical theory and computation* **2016**, *12*, 910–924.
- (92) Van Eijck, B. P.; Kroon, J. Structure predictions allowing more than one molecule in the asymmetric unit. *Structural Science* **2000**, *56*, 535–542.
- (93) Fredericks, S.; Parrish, K.; Sayre, D.; Zhu, Q. PyXtal: A Python library for crystal structure generation and symmetry analysis. *Computer Physics Communications* **2021**, *261*, 107810.
- (94) Bier, I.; Marom, N. Machine learned model for solid form volume estimation based on packing-accessible surface and molecular topological fragments. *The Journal of Physical Chemistry A* **2020**, *124*, 10330–10345.

- (95) Kadan, A.; Ryczko, K.; Wildman, A.; Wang, R.; Roitberg, A.; Yamazaki, T. Accelerated Organic Crystal Structure Prediction with Genetic Algorithms and Machine Learning. *Journal of Chemical Theory and Computation* **2023**, *19*, 9388–9402, PMID: 38059458.
- (96) Taylor, C. R.; Butler, P. W.; Day, G. M. Predictive crystallography at scale: mapping, validating, and learning from 1000 crystal energy landscapes. *Faraday Discussions* **2025**, *256*, 434–458.
- (97) Zhou, D.; Bier, I.; Santra, B.; Jacobson, L. D.; Wu, C.; Garaizar Suarez, A.; Almaguer, B. R.; Yu, H.; Abel, R.; Friesner, R. A., et al. A robust crystal structure prediction method to support small molecule drug development with large scale validation and blind study. *Nature Communications* **2025**, *16*, 2210.
- (98) Xie, T.; Fu, X.; Ganea, O.-E.; Barzilay, R.; Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197* **2021**,
- (99) Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Shysheya, S.; Crabbé, J.; Sun, L.; Smith, J., et al. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687* **2023**,
- (100) Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-tuned language models generate stable inorganic materials as text. *arXiv preprint arXiv:2402.04379* **2024**,
- (101) Antunes, L. M.; Butler, K. T.; Grau-Crespo, R. Crystal structure generation with autoregressive large language modeling. *Nature Communications* **2024**, *15*, 1–16.
- (102) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C., et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

- (103) Wheatley, P. 1163. The crystal and molecular structure of aspirin. *Journal of the Chemical Society (Resumed)* **1964**, 6036–6048.
- (104) Chan, E.; Welberry, T.; Heerdegen, A.; Goossens, D. Diffuse scattering study of aspirin forms (I) and (II). *Structural Science* **2010**, *66*, 696–707.
- (105) O'Connor, D.; Bier, I.; Tom, R.; Hiszpanski, A. M.; Steele, B. A.; Marom, N. Ab Initio Crystal Structure Prediction of the Energetic Materials LLM-105, RDX, and HMX. *Crystal Growth & Design* **2023**, *23*, 6275–6289.
- (106) Nielsen, A. T.; Chafin, A. P.; Christian, S. L.; Moore, D. W.; Nadler, M. P.; Nissan, R. A.; Vanderah, D. J.; Gilardi, R. D.; George, C. F.; Flippen-Anderson, J. L. Synthesis of polyazapolycyclic caged polynitramines. *Tetrahedron* **1998**, *54*, 11793–11812.
- (107) Russell, T.; Miller, P.; Piermarini, G.; Block, S. Pressure/temperature phase diagram of hexanitrohexaazaisowurtzitane. *The Journal of Physical Chemistry* **1993**, *97*, 1993–1997.
- (108) Millar, D. I.; Maynard-Casely, H. E.; Kleppe, A. K.; Marshall, W. G.; Pulham, C. R.; Cumming, A. S. Putting the squeeze on energetic materials—structural characterisation of a high-pressure phase of CL-20. *CrystEngComm* **2010**, *12*, 2524–2527.
- (109) Foltz, M. F.; Coon, C. L.; Garcia, F.; Nichols III, A. L. The thermal stability of the polymorphs of hexanitrohexaazaisowurtzitane, Part I. *Propellants, Explosives, Pyrotechnics* **1994**, *19*, 19–25.
- (110) Henson, B.; Smilowitz, L.; Asay, B.; Dickson, P. The β – δ phase transition in the energetic nitramine octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine: Thermodynamics. *The Journal of chemical physics* **2002**, *117*, 3780–3788.

- (111) Cobbley, R. E.; Small, R. The crystal structure of the δ -form of 1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetraazacyclooctane (δ -HMX). *Structural Science* **1974**, *30*, 1918–1922.
- (112) Zhang, L.; Jiang, S.-L.; Yu, Y.; Long, Y.; Zhao, H.-Y.; Peng, L.-J.; Chen, J. Phase transition in octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine (HMX) under static compression: an application of the first-principles method specialized for CHNO solid explosives. *The Journal of Physical Chemistry B* **2016**, *120*, 11510–11522.
- (113) Gao, D.; Huang, J.; Lin, X.; Yang, D.; Wang, Y.; Zheng, H. Phase transitions and chemical reactions of octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine under high pressure and high temperature. *RSC advances* **2019**, *9*, 5825–5833.
- (114) Mei, M.; Ji, J.; Sun, Z.; Zhu, W. Theoretical studies on dynamic properties and intermolecular interactions of 2, 4-dinitroimidazole crystals with different impurity defects. *CrystEngComm* **2024**, *26*, 1234–1244.
- (115) Duddu, R.; Dave, P. R.; Damavarapu, R.; Gelber, N.; Parrish, D. Synthesis of N-amino-and N-nitramino-nitroimidazoles. *Tetrahedron Letters* **2010**, *51*, 399–401.
- (116) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, *68*, 314–319.
- (117) Frey, B. J.; Dueck, D. Clustering by passing messages between data points. *science* **2007**, *315*, 972–976.
- (118) Nocedal, J.; Wright, S. J. *Numerical Optimization*; Springer series in operations research and financial engineering; Springer: New York, NY, 2006.
- (119) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J., et al. SciPy 1.0:

fundamental algorithms for scientific computing in Python. *Nature methods* **2020**, *17*, 261–272.

- (120) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications* **2009**, *180*, 2175–2196.
- (121) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New Journal of Physics* **2012**, *14*, 053020.
- (122) Zhang, I. Y.; Ren, X.; Rinke, P.; Blum, V.; Scheffler, M. Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar. *New Journal of Physics* **2013**, *15*, 123033.
- (123) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of chemical physics* **1999**, *110*, 6158–6170.
- (124) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.
- (125) Tkatchenko, A.; DiStasio Jr, R. A.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Physical review letters* **2012**, *108*, 236402.
- (126) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *The Journal of chemical physics* **2014**, *140*.
- (127) Hermann, J.; Stöhr, M.; Góger, S.; Chaudhuri, S.; Aradi, B.; Maurer, R. J.;

Tkatchenko, A. libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *The Journal of Chemical Physics* **2023**, *159*.

- (128) Kholtobina, A.; Lončarić, I. Exploring elastic properties of molecular crystals with universal machine learning interatomic potentials. *Materials & Design* **2025**, *114047*.
- (129) Chisholm, J. A.; Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *Applied Crystallography* **2005**, *38*, 228–231.
- (130) Sykes, R. A.; Johnson, N. T.; Kingsbury, C. J.; Harter, J.; Maloney, A. G.; Sugden, I. J.; Ward, S. C.; Bruno, I. J.; Adcock, S. A.; Wood, P. A., et al. What has scripting ever done for us? The CSD Python application programming interface (API). *Applied Crystallography* **2024**, *57*, 1235–1250.
- (131) Curtis, F.; Wang, X.; Marom, N. Effect of packing motifs on the energy ranking and electronic properties of putative crystal structures of tricyano-1, 4-dithiino [c]-isothiazole. *Structural Science* **2016**, *72*, 562–570.
- (132) Reilly, A. M.; Tkatchenko, A. Role of dispersion interactions in the polymorphism and entropic stabilization of the aspirin crystal. *Physical review letters* **2014**, *113*, 055701.
- (133) Varughese, S.; Kiran, M.; Solanko, K. A.; Bond, A. D.; Ramamurty, U.; Desiraju, G. R. Interaction anisotropy and shear instability of aspirin polymorphs established by nanoindentation. *Chemical Science* **2011**, *2*, 2236–2242.
- (134) Dichi, E.; Sghaier, M.; Guiblin, N. Pharmaceutical phase diagram: aspirin-caffeine-paracetamol. *Journal of Thermal Analysis and Calorimetry* **2023**, *148*, 6107–6118.
- (135) LeBlanc, L. M.; Otero-de-la Roza, A.; Johnson, E. R. Evaluation of shear-slip transitions in crystalline aspirin by density-functional theory. *Crystal Growth & Design* **2016**, *16*, 6867–6873.

- (136) Wen, S.; Beran, G. J. Accidental degeneracy in crystalline aspirin: New insights from high-level ab initio calculations. *Crystal growth & design* **2012**, *12*, 2169–2172.
- (137) Huang, Y.; Shao, Y.; Beran, G. J. Accelerating MP2C dispersion corrections for dimers and molecular crystals. *The Journal of Chemical Physics* **2013**, *138*.
- (138) Nickerson, C. J.; Johnson, E. R. Assessment of a foundational machine-learned potential for energy ranking of molecular crystal polymorphs. *Physical Chemistry Chemical Physics* **2025**, *27*, 114047.