

FastCSP: Accelerated Molecular Crystal Structure Prediction with Universal Model for Atoms

Vahe Gharakhanyan^{1,*}, Yi Yang^{2,*}, Luis Barroso-Luque¹, Muhammed Shuaibi¹, Daniel S. Levine¹, Kyle Michel¹, Viachaslau Bernat³, Misko Dzamba¹, Xiang Fu¹, Meng Gao¹, Xingyu Liu³, Keian Noori³, Lafe J. Purvis³, Tingling Rao³, Brandon M. Wood¹, Ammar Rizvi¹, Matt Uyttendaele¹, Andrew J. Ouderkirk³, Chiara Daraio³, C. Lawrence Zitnick¹, Arman Boromand^{3,†}, Noa Marom^{2,4,5,†}, Zachary W. Ulissi^{1,†}, Anuroop Sriram^{1,*†}

¹Fundamental AI Research at Meta, ²Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, ³Reality Labs Research at Meta, ⁴Department of Physics, Carnegie Mellon University, Pittsburgh, PA, USA, ⁵Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA

*Equal contribution, †Corresponding authors

Crystal Structure Prediction (CSP) of molecular crystals plays a central role in applications, such as pharmaceuticals and organic electronics. CSP is challenging and computationally expensive due to the need to explore a large search space with sufficient accuracy to capture energy differences of a few kJ/mol between polymorphs. Dispersion-inclusive density functional theory (DFT) provides the required accuracy but its computational cost is impractical for a large number of putative structures. We introduce FastCSP, an open-source, high-throughput CSP workflow based on machine learning interatomic potentials (MLIPs). FastCSP combines random structure generation using Genarris 3.0 with geometry relaxation and free energy calculations powered entirely by the Universal Model for Atoms (UMA) MLIP. We benchmark FastCSP on a curated set of 28 mostly rigid molecules, demonstrating that our workflow consistently generates known experimental structures and ranks them within 5 kJ/mol per molecule of the global minimum. Our results demonstrate that universal MLIPs can be used across diverse compounds without requiring system-specific tuning. Moreover, the speed and accuracy afforded by UMA eliminate the need for classical force fields in the early stages of CSP and for final re-ranking with DFT. The open-source release of the entire FastCSP workflow significantly lowers the barrier to accessing CSP. CSP results for a single system can be obtained within hours on tens of modern GPUs, making high-throughput crystal structure prediction feasible for a broad range of scientific applications.

Code: <https://github.com/facebookresearch/fairchem>

Correspondence: A.S. (anuroops@meta.com), Z.W.U. (zulissi@meta.com), N.M. (nmarom@andrew.cmu.edu), A.B. (aboromand@meta.com)



1 Introduction

Crystal Structure Prediction (CSP) of molecular crystals is a fundamental challenge in materials science [1], with profound implications across several industries including pharmaceuticals [2, 3], organic electronics [4–6], energetic materials [7–9], pigments [10], and agrochemicals [11, 12]. Many molecules are known to form distinct crystal structures, known as polymorphs, depending on crystallization conditions [13–16]. Different polymorphs of the same compound can exhibit widely varying physical properties [17–21]. Therefore, it is desirable to explore the configuration space of structure and properties to optimize the performance of products based on molecular crystals. Computer simulations can greatly accelerate this process compared to experimental polymorph screening and steer crystal growth efforts towards potentially stable polymorphs with desirable properties [6, 22–26].

Different polymorphs are often separated by only a few kJ/mol per molecule in energy [14, 27]. This necessitates very accurate evaluation of the thermodynamic stability of putative structures in order to determine the

most likely form(s) of a given compound. CSP is thus challenging because it requires exploring the vast configuration space of possible crystal structures with very high accuracy. This challenge is represented by the CSP blind tests organized by the Cambridge Crystallographic Data Centre (CCDC), which have been conducted periodically for over two decades to assess the progress of the field [28–35]. In these blind tests, participants aim to predict the crystal structure of target compounds with a varying degree of complexity starting from a chemical “stick diagram”. Over the years, supercomputing capabilities and CSP methods have advanced hand in hand. In the early blind tests only classical force fields were used [28, 29], whereas in more recent blind tests the use of dispersion-inclusive density functional theory (DFT) for final stability ranking has become an established best practice [32, 33]. Unfortunately, the high computational cost of dispersion-inclusive DFT methods limits the scale at which they can be applied. In practice, a hierarchical CSP approach is often applied, in which classical force fields and/or semi-empirical methods are employed in the initial steps of a multi-stage workflow to quickly and coarsely filter structures before final DFT refinement and ranking.

The recent seventh CSP blind test notably saw the first use of machine learning interatomic potentials (MLIPs) for the CSP problem [34, 35]. MLIPs have revolutionized atomistic modeling by enabling accurate predictions of energies and forces at a fraction of the cost of quantum mechanical methods such as DFT [36–47]. These MLIPs are typically trained on large datasets of DFT calculations [48–55]. Until the recent publication of the Open Molecular Crystals 2025 (OMC25) dataset [56], open DFT datasets have been available only for inorganic materials and small isolated molecules. This has hindered the development of MLIPs for molecular crystals. In the seventh blind test, one team (Group 16) developed system-specific AIMNet2 MLIPs [57]. These MLIPs were trained on DFT data for small clusters of molecules extracted from putative crystal structures of each target. They were used for structure relaxations, lattice energy evaluations, and free energy calculations in both the structure generation phase and the ranking phase. In the structure ranking phase, two other teams applied machine learning methods. Group 12 used Gaussian process regression (GPR) trained on density functional tight binding (DFTB) data and Group 15 used ANI-2x MLIPs fine-tuned by transfer learning on the lists of crystal structures provided by the CCDC. Group 16’s system-specific AIMNet2 MLIPs delivered the best performance, on par with dispersion-inclusive DFT. However, system-specific MLIPs require training a new model or fine-tuning an existing one for each molecule of interest, greatly limiting their applicability to high-throughput screening. More recently, several studies employed general-purpose MLIPs, pretrained on many diverse compounds, for CSP without performing system-specific retraining [58–64]. However, owing to remaining issues with the reliability and transferability of general-purpose MLIPs [61, 65–68], most current CSP workflows still rely on a final re-ranking step with DFT, which limits their scalability and use for accelerated materials design and discovery.

In this work, we present the FastCSP framework, an open-source crystal structure prediction workflow that leverages a universal machine learning potential, Universal Model for Atoms (UMA) [38], for geometry relaxation, ranking, and free energy evaluations of random structures generated by Genarris 3.0 [61]. UMA is built on the eSEN architecture [40], an equivariant graph neural network model, and is extended with Mixture of Linear Experts (MoLE) layers to scale to large model sizes without compromising inference speed. Unlike most existing MLIPs, which are typically trained on a single domain (e.g., molecules or inorganic crystals), UMA is trained jointly across multiple domains using shared representations and separate dataset-specific tasks. Here, we use the Open Molecular Crystals (OMC) task [56]. OMC25 includes over 25 million configurations extracted from relaxation trajectories of thousands of putative molecular crystal structures generated by Genarris 3.0 from 50,000 chemically diverse compounds. The crystal structures in OMC25 are labeled with energies, forces, and stress tensors, calculated using dispersion-inclusive DFT. This large, chemically diverse dataset enables UMA to generalize across a wide range of compounds, comprising different elements and capable of forming a variety of intermolecular interactions and crystal packing motifs, without requiring molecule-specific fine-tuning.

We demonstrate the effectiveness of our approach for a set of 28 mostly rigid molecules. Our FastCSP workflow is able to generate the known experimental structures for all of these molecules. In most cases, the experimentally observed structure, or one of the known polymorphs, is ranked as the global minimum, and if not, they are within the top 10 (with the exception of a polymorph of glycine and a polymorph of imidazole known to be less stable) and 5 kJ/mol of the lowest energy structure. We further show that the performance of UMA for geometry optimization, ranking, and free energy calculations is on par with dispersion-inclusive DFT. Thus,

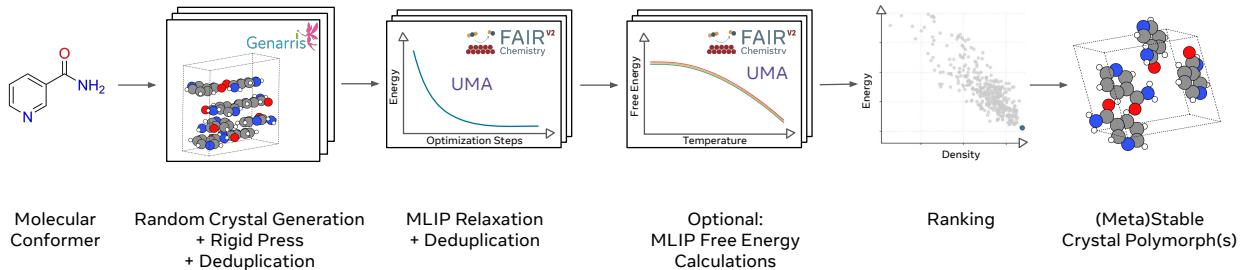


Figure 1 Overview of the FastCSP workflow. Starting from a molecular conformer, Genarris 3.0 generates a diverse set of random crystal structures. The generated structures undergo optimization with Rigid Press and deduplication is performed to remove similar structures. The remaining structures are fully relaxed using the UMA model and undergo another round of deduplication. The final ranking may be based on UMA lattice energy, or optionally, on Helmholtz or Gibbs free energies at finite temperature and pressure, also calculated using UMA.

the accuracy of UMA obviates the need for classical force field pre-screening or DFT-based (re)ranking and enables a CSP workflow that relies exclusively on MLIPs, unlocking results within hours with tens of modern GPUs available. Our entire workflow is released open-source, with the data used to train the models released in a companion paper [56]. This combination of accuracy, speed, and accessibility represents a significant step toward scalable, transparent, and reliable CSP workflows.

2 Methods

2.1 Overview of FastCSP

The FastCSP workflow, illustrated in Figure 1, is designed for efficient, high-throughput crystal structure prediction. It comprises two main stages: random structure generation using Genarris 3.0 [61] followed by geometry relaxation and stability ranking using UMA [38]. For a given input conformer, Genarris 3.0 constructs a large number of molecular packing arrangements across a broad set of compatible space groups. The initial structures undergo compression with the Rigid Press feature of Genarris 3.0, which uses a regularized hard-sphere potential to achieve close-packing. Subsequently, duplicate structures are removed using PYMATGEN’s STRUCTUREMATCHER [69].

The remaining structures are fully relaxed using the OMC task of the Small version 1.1 UMA (UMA-S-1.1) model. The use of a general-purpose MLIP allows us to scale relaxation across thousands of candidate structures with minimal computational cost. On a modern NVIDIA H100 80 GB GPU, each relaxation typically takes around 15 seconds, making our workflow suitable for high-throughput CSP campaigns. Any not-converged structures or those whose molecular connectivity changed during relaxation are discarded. Then, deduplication is performed again using PYMATGEN’s STRUCTUREMATCHER to eliminate redundant structures. Finally, only the structures within a fixed energy window of 20 kJ/mol above the global minimum are retained. This consists the final potential energy landscape. In addition to lattice energy calculations, our MLIPs can optionally be used to compute the Helmholtz and Gibbs vibrational free energies at finite temperatures. This reflects the relative thermodynamic stability of candidate structures under real conditions. Full details of the configuration parameters used for structure generation, energy calculations, and deduplication are provided in SI B.

2.2 Structure Generation

For each target, a single conformer was extracted from the experimental crystal structure in the Cambridge Structural Database (CSD) [70] (for the systems with multiple polymorphs only one form was selected) and relaxed using dispersion-inclusive DFT. The molecular conformation of glycine was extracted from the crystal structure of α -glycine and not relaxed because the zwitterion configuration is not stable in the gas phase [71]. Genarris 3.0 [61] was used to generate a large number of putative crystal structures from the given

molecular conformer. Genarris generates structures with one molecule in the asymmetric unit ($Z' = 1$) in all space groups compatible with the requested number of molecules per unit cell (Z) and the molecular point group symmetry, including cases where molecules occupy special Wyckoff positions [72]. For each target, we generated 500 structures in each compatible space group with $Z \in 1, 2, 3, 4, 6, 8$. For CILJIQ, Form II was not generated with these settings, and structure generation was repeated with 1,000 structures per space group.

Genarris 3.0 generates random molecular crystal structures by sampling unit cell volumes from a Gaussian distribution around a target volume estimated by the built-in PyMoVE model [73]. The estimated volume is initially scaled by a factor of 1.5 to facilitate molecule placement. The first molecule is randomly placed within the unit cell, and the remaining molecules are placed according to the symmetry operations of the selected space group. When the initial placement occurs at a special Wyckoff position, the molecule is aligned to match the site symmetry [72]. Once a molecular crystal structure is generated, it undergoes validation to ensure that interatomic distances (d_{ij}) between atoms i and j from different molecules are physically reasonable. Specifically, the distance d_{ij} must exceed a threshold defined by $s_r \times (r_i^{vdW} + r_j^{vdW})$, where $r_{i/j}^{vdW}$ are the atomic van der Waals radii and s_r is a user-defined fraction with a default of 0.95. Specific intermolecular distance thresholds are assigned for strong hydrogen bonds [72].

The initial generation of molecular crystal structures proceeds until the number of structures requested by the user is reached. Then, geometry optimization with *Rigid Press* [61] is performed. A regularized hard-sphere potential is employed to compress the unit cell, while keeping the molecules' internal geometry rigid. This brings the molecules as close to each other as possible without breaking the space group symmetry to achieve close packing. Next, we deduplicate the generated structures using PYMATGEN's STRUCTUREMATCHER [69] to remove similar crystal structures. For most targets, only about 35% of the generated structures remained after this deduplication step, significantly lowering the number of relaxations to be performed and therefore the overall computational cost. We consider this as an indication that the configuration space was sampled exhaustively.

2.3 Structure Relaxations

UMA-Small version 1.1 model with the OMC task [38] was used for optimizing atomic positions and lattice parameters of all generated structures. Structure relaxations used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer as implemented in the atomic simulation environment (ASE) [74], with a 0.01 eV/Å force threshold, under periodic boundary conditions and without symmetry constraints. Optimizations were limited to 1,000 steps and non-converged structures were discarded.

2.4 Free Energy Calculations

Free energy calculations were conducted for putative structures with relative lattice energies within 10 kJ/mol of the global minimum to better reflect thermodynamic stability under experimental conditions. To obtain Helmholtz free energies at constant volume, we performed phonon calculations within the harmonic approximation (HA), using the supercell method implemented in PHONOPY [75, 76]. To account for thermal expansion, we calculated the Gibbs free energies within the quasiharmonic approximation (QHA). The Gibbs free energy is approximated by calculating Helmholtz free energies for relaxed supercells at a range of fixed volumes. Here, we used volumes ranging between $\pm 6\%$ of the zero-temperature equilibrium volume obtained from our preceding MLIP relaxations. Subsequently, an expression for the Helmholtz free energy as a function of volume at a fixed temperature is obtained by fitting a Vinet equation of state [77] to the calculated free energies at each volume. Finally, we determined the Gibbs free energy of a structure at a given temperature and pressure by identifying the minimum value of the fitted Helmholtz free energy plus the hydrostatic work (PV). Additional details for free energy calculations are given in SI B.

2.5 DFT Calculations

The single molecule geometry was relaxed using the FHI-aims all-electron electronic structure code [78–80]. The PBE0 [81] hybrid functional, which is based on the Perdew–Burke–Ernzerhof (PBE) [82] generalized gradient approximation, was used with the many-body dispersion (MBD) correction method [83–85]. The relaxations were performed with the Tier 2 basis sets of FHI-aims, *tight* numerical settings, and a force

convergence criterion of 0.001 eV/Å. A comparison of the molecular geometry before and after relaxation produces an average root mean squared deviation (RMSD) of atomic positions of 0.086 ± 0.029 Å across the 27 targets, with HURYUQ having the highest RMSD of 0.139 Å.

To benchmark the accuracy of the UMA MLIP, we performed reference dispersion-inclusive DFT calculations for all deduplicated relaxed structures ranked by UMA within 20 kJ/mol from the global minimum of each target. We used Vienna Ab initio Simulation Package (VASP) [86–88] with the same computational settings as those employed in the generation of the OMC25 dataset [56]. The PBE exchange-correlation functional was paired with Grimme’s D3 [89] pairwise dispersion correction. VASP 5.4 projector augmented wave (PAW) PBE pseudopotentials were used with a plane-wave energy cutoff of 520 eV. PYMATGEN’s [69] default Γ-centered k-point grids were used. PBE-D3 relaxations were started from the structures optimized by UMA. The atomic positions and lattice vectors were relaxed until the maximum per-atom residual forces fell below 0.001 eV/Å, or the relaxation process exceeded 1,500 steps.

To benchmark the accuracy of the free energy calculations with the UMA MLIP, we performed DFT free energy calculations for the five predicted polymorphs of glycine. The FHI-aims code was used with Tier 1 basis sets and *light* numerical settings. Free energy calculations within the HA and QHA were performed at the PBE-D3 level using the finite displacement supercell method implemented in PHONOPY [75, 76]. We employed $2 \times 2 \times 2$ supercells and performed harmonic free energy calculations for six fixed volumes, ranging from -6% to 10% of the 0 K equilibrium unit cell volume. All other computational settings matched those of the free energy calculations with UMA. The five polymorphs of glycine were further re-optimized with PBE+MBD using the Tier 2 basis sets and *tight* numerical settings of FHI-aims. These structures were re-ranked based on single-point energy evaluations using the PBE0+MBD using the Tier 1 basis sets of FHI-aims and *light* numerical settings.

3 Results and Discussion

3.1 Benchmark Molecular Dataset

To validate our FastCSP workflow, we curated a diverse set of mostly rigid molecules, shown in Figure 2. We focus on mostly rigid molecules where a single low-energy conformer can be assumed because our goal is to demonstrate the effectiveness of MLIPs for geometry optimization and stability ranking. Extending this approach to molecules with significant conformational flexibility, for which conformer generation must be integrated as a pre-processing step, is left to future work. Our benchmark set includes the 14 compounds categorized in [59] as “Tier 1”, meaning mostly rigid molecules with up to 30 atoms. The set of CSP targets used in [59] only included the 22 most common space groups in the Cambridge Structural Database (CSD) [70, 90] due to the limitations of their structure generation method. Moreover, the MLIP used therein was unable to handle elements, such as Br. We extended the set to 28 compounds in order to explore additional space groups (with Z values of 1, 3, 4, 6, and 8), as well as more chemically diverse compounds capable of forming a variety of intermolecular interactions and packing motifs. CILJIQ was selected from the CSD [70] because it contains bromine, a hydroxyl, and an aldehyde, enabling it to form halogen and hydrogen bonds. CE BYUD and imidazole (IMAZOL) were selected from [91] because they crystallize in space groups $P3_2$ (No. 145) and $Aba2$ (No. 41), respectively, which are not among the top 22 most common space groups. 11 Additional compounds were selected from [62] in order to cover $Z = 1$ and space groups Cc (No. 9), $P4_12_12$ (No. 92), $P\bar{4}2_1c$ (No. 114), and $P6_5$ (No. 170). 5 Compounds in the benchmark set have more than one known polymorph, amounting to a total of 36 crystal structures for the 28 molecules. For Target VIII, Target XII, Target XIII, acetic acid (ACETAC), glycine (GLYCIN), nicotinamide (NICOAM), IHEPUG, and imidazole (IMAZOL), there are multiple entries corresponding to the crystal structure of the same polymorph in the CSD. For Target VIII, Target XII, and Target XIII, we chose the entries that correspond to the blind test reports. After excluding different conformers for nicotinamide and IHEPUG, we followed the procedure of [59] for selecting the most reliable structure(s). Detailed information on the crystal structure(s) of each compound and the corresponding CSD entries is provided in SI Table S1.

Of the compounds included in our benchmark set, Target I, CILJIQ, imidazole, nicotinamide, and glycine have polymorphs. Target I (C_6H_6O , 3,4-cyclobutylfuran) from the first CSP blind test [28] has two known polymorphs: a stable form that crystallizes in the monoclinic space group $P2_1/c$ (No. 14) with $Z = 4$ and

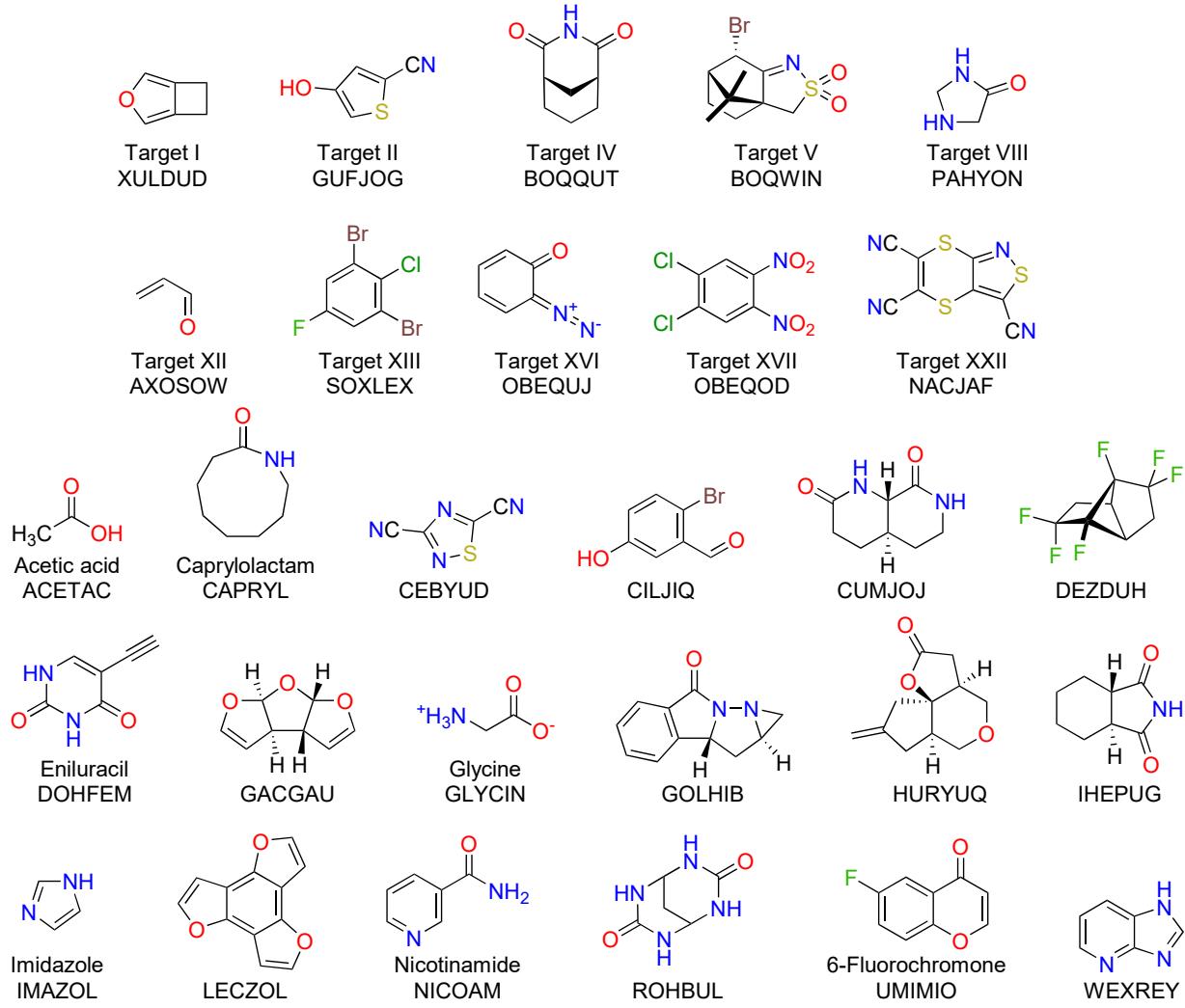


Figure 2 Molecular diagrams, CSD reference codes, and common names of the 28 mostly rigid compounds considered as benchmark systems for FastCSP. More details can be found in SI Table S1.

a metastable form that crystallizes in the orthorhombic space group *Pbca* (No. 61) with $Z = 8$. CILJIQ ($C_7H_5BrO_2$, 2-bromo-5-hydroxybenzaldehyde) has two known polymorphs, both of which have four molecules per unit cell ($Z = 4$). Form I crystallizes in the orthorhombic space group *P2₁2₁2₁* (No. 19) and forms a catemer packing motif of zigzag chains bound by hydrogen bonds between the hydroxyl and aldehyde of neighboring molecules [92]. Form II crystallizes in the monoclinic space group *P2₁/c* (No. 14) with a hydrogen-bonded cyclic dimer packing motif [93]. Their relative stability has not been reported. Imidazole ($C_3H_4N_2$, 1*H*-imidazole) has two known polymorphs. The stable α form crystallizes at ambient conditions in the monoclinic space group *P2₁/c* (No. 14) with $Z = 4$ [94]. The β form recrystallizes at a pressure of 1.2 GPa in the orthorhombic space group *Aba2* (No. 41) with $Z = 8$ [95]. Nicotinamide ($C_6H_6N_2O$, pyridine-3-carboxamide) is well-known as vitamin *B*₃. It is a highly polymorphic compound with nine known polymorphs [96]. Here, we focus on two polymorphs with $Z' = 1$ that have the same molecular conformation: the most stable α form and the ι form that has the lowest melting point. Both crystallize in the monoclinic space group *P2₁/c* (No. 14) with $Z = 4$. Glycine ($C_2H_5NO_2$) is the simplest amino acid. Under ambient conditions, it has three polymorphs. The γ form crystallizes in the trigonal space group *P3₁* (No. 144)/*P3₂* (No. 145) with $Z = 3$ [97], the α form crystallizes in the monoclinic space group *P2₁/n* (No. 14) with $Z = 4$, and the β form crystallizes in the monoclinic space group *P2₁* (No. 4) with $Z = 2$. The γ form is

the most stable, followed by α and β [98–100]. Two additional polymorphs have been obtained under high pressure [101]. The δ form crystallizes in the monoclinic space group $P2_1/a$ (No. 14) with $Z = 4$ via phase transition from the β form and the ε form crystallizes in the monoclinic space group Pn (No. 7) with $Z = 2$, via phase transition from the γ form.

3.2 Crystal Structure Prediction Results

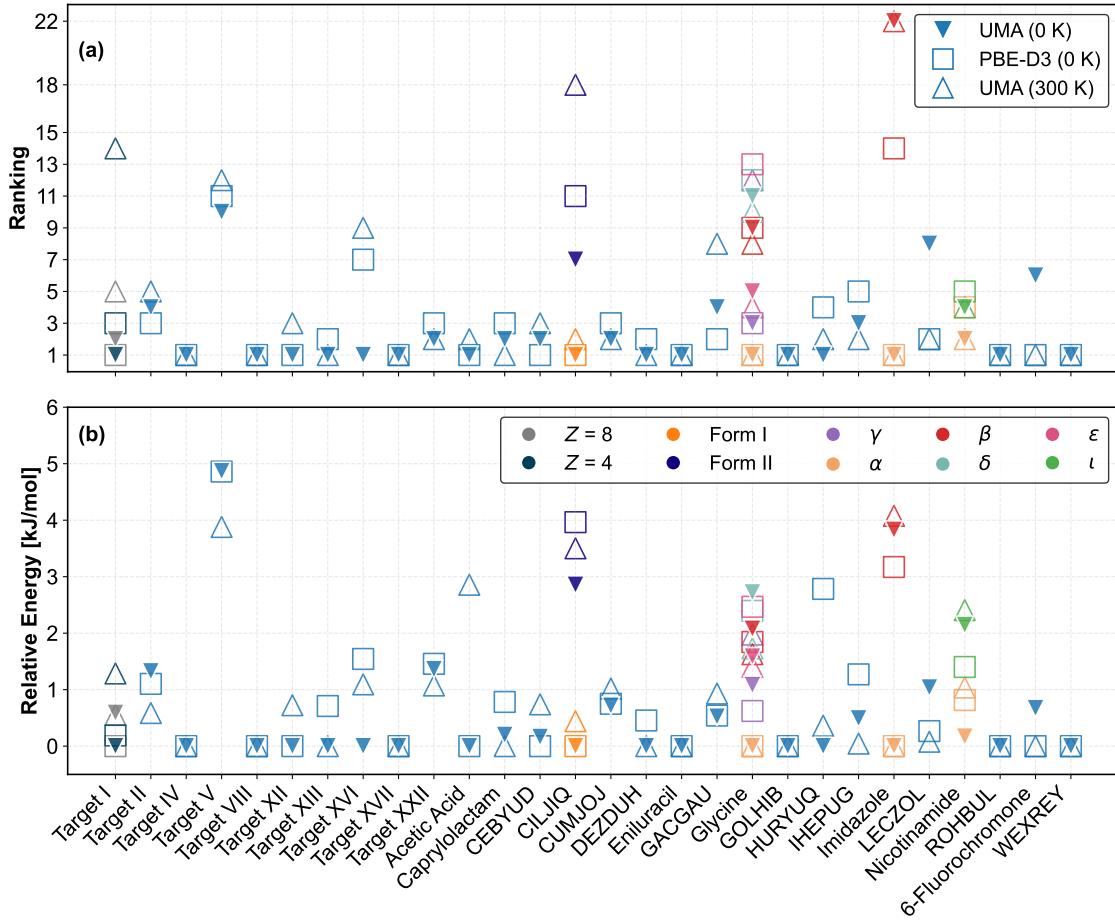


Figure 3 (a) Ranking of the best matches to the experimental structures by the UMA MLIP (at 0 K and 300 K) and PBE-D3 (at 0 K). (b) Relative energy (in kJ/mol) of the best matches to the experimental structures, obtained with the UMA MLIP (at 0 K and 300 K) and PBE-D3 (at 0 K). Each column represents a different molecule and markers in different colors indicate individual known polymorphs. All matches to the experimental structures fall within 5 kJ/mol, and the majority falls within the top 10 lowest energy structures, with many ranked as the global minimum.

The FastCSP workflow successfully generated the experimental structures of all the compounds in the benchmark set. The only exception is Form II of CILJIQ, which was only generated when the number of structures generated by Genarris in each space group was increased from 500 to 1,000. The results presented here for CILJIQ are from the run that sampled 1,000 structures and the results of the run that sampled 500 structures are available in SI Figure S5. For most of the small and mostly rigid compounds studied here, sampling 500 structures per space group was sufficient, however some structures are more difficult to generate and require more attempts [61]. The packing motif of Form II of CILJIQ comprises centrosymmetric cyclic dimers linked by a pair of O–H...O hydrogen bonds [93]. Cyclic dimer is a common packing mode for carboxylic acids [102–104]. Compared to the catemer packing motif of Form I, generating the cyclic dimer motif of Form II may require more sampling to achieve the correct pair of molecules related by an inversion center in the less constrained space group $P2_1/c$ (No. 14). The user may increase the number of structures generated per space group by Genarris, depending on their needs and computational resources. We note that

for caprylolactam two generated structures with $Z = 2$ and $Z = 4$ matched the experimental structure with $Z = 4$ (details of the matching criteria used here are provided in SI B). When loose settings are used for symmetry detection, the $Z = 2$ structure has the same space group, Cc (No. 9), as the experimental form. After deduplication, it remained in the final energy landscape because it had a lower relative lattice energy. A tabulated summary of the CSP results is provided in SI Table S3, and overlays of predicted and experimental structures are shown in SI Table S6.

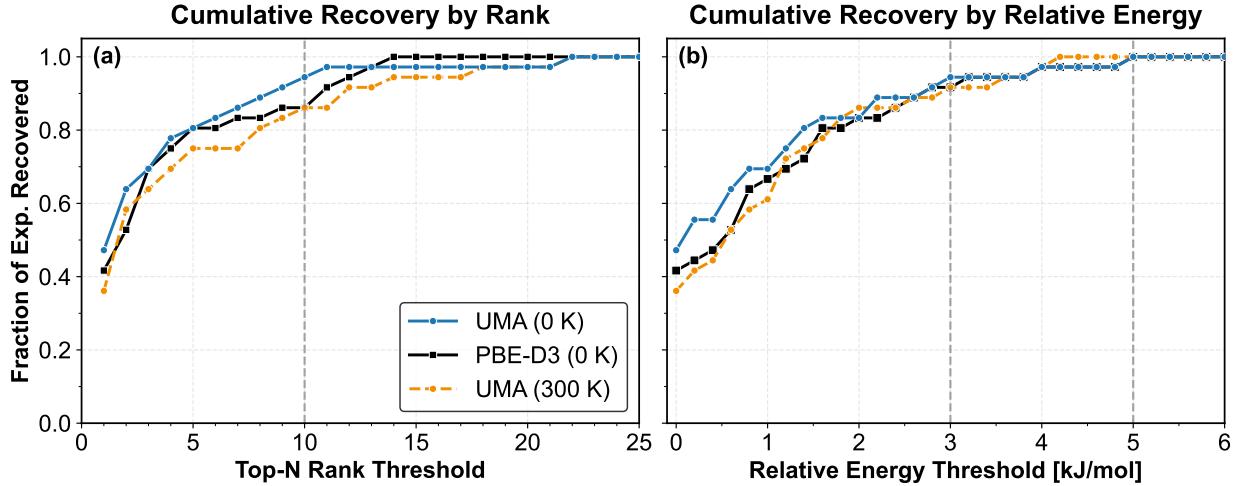


Figure 4 Cumulative recall of experimental structures across the benchmark set. (a) Recall as a function of the top-N ranked predicted structures using ML (UMA MLIP at 0 K and 300 K) and DFT (PBE-D3 at 0 K). (b) Recall as a function of relative energy threshold (in kJ/mol) from the predicted minimum. Recall measures the fraction of known experimental structures that are recovered within a given rank or energy threshold. The UMA model recovers the majority of experimental structures within the top 10 predictions and within 3 kJ/mol of the global minimum.

Figure 3 shows that for 17 of the target molecules the experimental structure, or one of the known polymorphs, is ranked by UMA as the lattice energy (0 K) global minimum. For 8 of the remaining compounds an experimentally observed structure is ranked by UMA within the top 4. The outliers are Target V, LECZOL, and 6-fluorochromone, whose experimental structures are ranked by UMA as #10, #8, and #6, respectively. All experimentally observed structures fall within 5 kJ/mol (≈ 1.20 kcal/mol) of the lattice energy global minimum, with the vast majority falling within 3 kJ/mol (≈ 0.72 kcal/mol) of the global minimum. These results are also reflected in the recall analysis, shown in Figure 4, which measures the fraction of known experimental polymorphs that are successfully recovered within a specified rank or energy threshold. Our UMA-based CSP workflow achieves over 94% recall within the top 10 lowest-energy predictions and 3 kJ/mol of the lattice energy (0 K) global minimum. In most cases, the UMA ranking is in close agreement with PBE-D3. In some cases, UMA even outperforms PBE-D3. For Target XIII, Target XVI, DEZDUH, and HURYUQ, UMA ranks the experimental structure as the global minimum while PBE-D3 does not. For caprylolactam and IHEPUG, UMA ranks the experimental structure closer to the global minimum than PBE-D3. This demonstrates that UMA provides results on par with dispersion-inclusive DFT at a fraction of the computational cost across diverse molecular systems.

It has been argued that CSP should go beyond finding the most stable crystal structure based on 0 K lattice energies, and attempt to predict the most thermodynamically stable crystal structure under finite temperature and pressure conditions by evaluating the Gibbs and Helmholtz free energies [1, 3, 105–109]. The efficiency of MLIPs enables large-scale free energy calculations [57]. Figure S4 in SI shows the change in the ranking and relative energy upon switching from lattice energy at 0 K to free energy at 300 K. In most cases, incorporating free energy corrections at 300 K results in an average relative energy change of ± 2.2 kJ/mol. Because there may be many putative structures within a small energy window, these changes in the relative energy lead to reshuffling of the ranking by ± 23 . The only exception is Target XVI, for which considerable changes in relative energy and ranking are obtained. These results align with previous studies, which showed that including vibrational free energy contributions reversed the relative stability of some of the studied polymorph pairs [27, 109, 110]. It has been reported that for small molecules with limited flexibility,

the room-temperature free energy corrections are often within 2 kJ/mol [27, 110], but can be more significant for large flexible molecules [3]. Figure 3 shows that for Target XXII, caprylolactam, IHEPUG, LECZOL, and 6-fluorochromone the free energy corrections improve the ranking of the experimental structure. For other systems the free energy corrections do not affect the ranking of the experimental structures or make it worse. For example, the Z=4 form of Target I, which is supposed to be the most stable, is considerably destabilized by the free energy correction. The experimental structure of Target XVI is also destabilized by the free energy correction. Figure 4 shows that overall, the free energy corrections make the recall rate slightly worse. This is consistent with the outcomes of the ranking stage of the seventh CSP blind test [35], in which free energy corrections were not always beneficial.

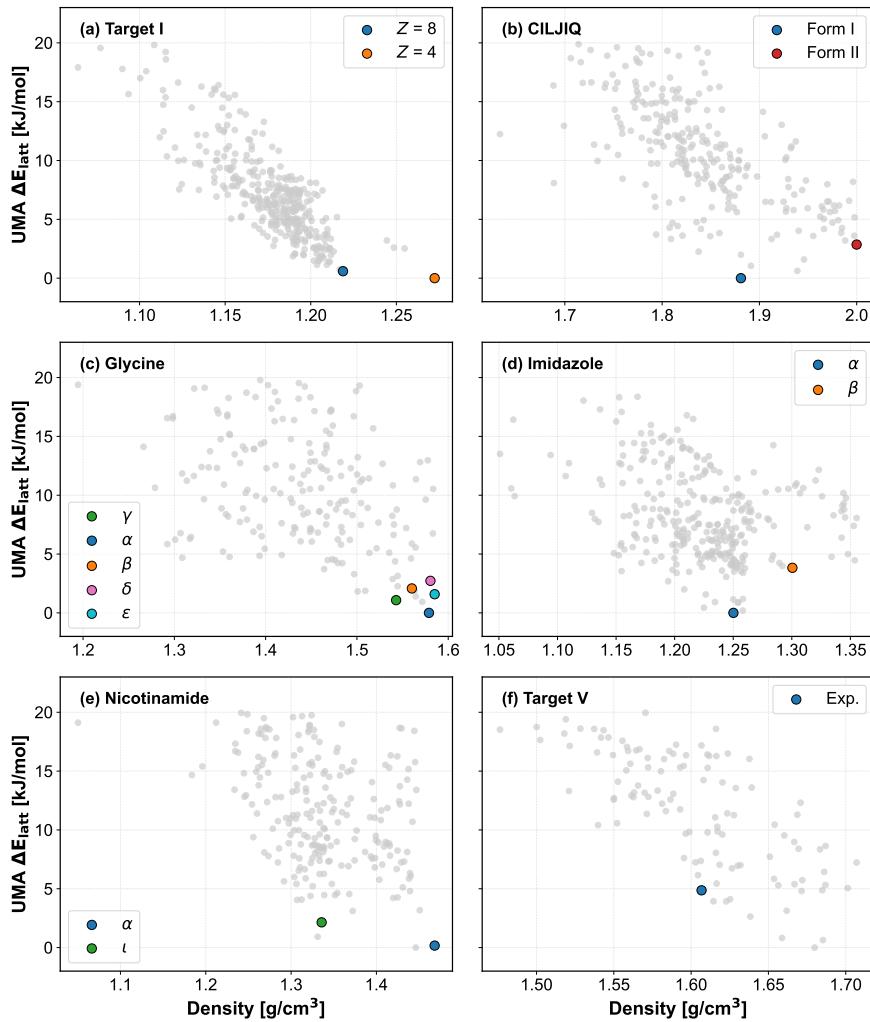


Figure 5 Final energy landscapes obtained with UMA. The relative lattice energy (at 0 K) is plotted as a function of density for (a) Target I, (b) CILJIQ, (c) glycine, (d) imidazole, (e) nicotinamide, and (f) Target V.

Figure 5 shows the final energy landscapes obtained with UMA at 0 K for all targets with multiple polymorphs and for Target V. Results for the remaining targets are presented in SI Figure S1. For comparison, the final energy landscapes obtained with PBE-D3 are presented in SI Figure S2 and the final free energy landscapes obtained with UMA at 300 K are shown in SI Figure S3. For Target I, UMA ranks the $Z = 4$ form as more stable than the $Z = 8$ form by 0.59 kJ/mol at 0 K. This is consistent with the experimental observation that the $Z = 8$ form is metastable [28]. It is noteworthy that in this case UMA performs better than PBE-D3, which incorrectly ranks the $Z = 8$ form as more stable. This is consistent with earlier studies, which showed that the ranking of the two forms of Target I depends on the DFT functional and dispersion method used [59, 111, 112] (the ranking from [59] was extracted from an associated repository and is provided in SI Table S4). As noted above, adding free energy corrections with UMA changes the order of stability, making the $Z = 8$ form more

stable than the $Z = 4$ form by 0.73 kJ/mol. For imidazole, UMA ranks the α form as more stable than the β form by 3.83 kJ/mol at 0 K, in agreement with experimental observations [95] and PBE-D3. Free energy corrections slightly increase the energy difference between the two forms to 4.08 kJ/mol. For nicotinamide, the α form is ranked by UMA as more stable than the ι form by 1.98 kJ/mol at 0 K. This is consistent with experimental observations [96, 113, 114], with our PBE-D3 ranking, and with dispersion-inclusive DFT results reported by others [59, 96] (the ranking from [59] was extracted from an associated repository and is provided in SI Table S4). In this case, adding free energy corrections decreases the energy difference between the two forms to 1.37 kJ/mol. For CILJIQ, Form I is ranked by UMA as more stable than Form II by 2.86 kJ/mol at 0 K in agreement with PBE-D3. Adding free energy corrections changes the energy difference to 3.06 kJ/mol. We were unable to find any experimental studies or other computational studies reporting the relative stability of these polymorphs. For Target V, neither UMA nor PBE-D3 ranks the experimental structure near the global minimum. With UMA, the experimental form is ranked as #10, 4.87 kJ/mol above the lattice energy (0 K) global minimum, which is close to ranking of #11 with PBE-D3. It has been shown previously [59, 112, 115] that lattice energies obtained with various dispersion-inclusive DFT methods misrank the experimental structure of Target V (the ranking from [59] was extracted from an associated repository and is provided in SI Table S4).

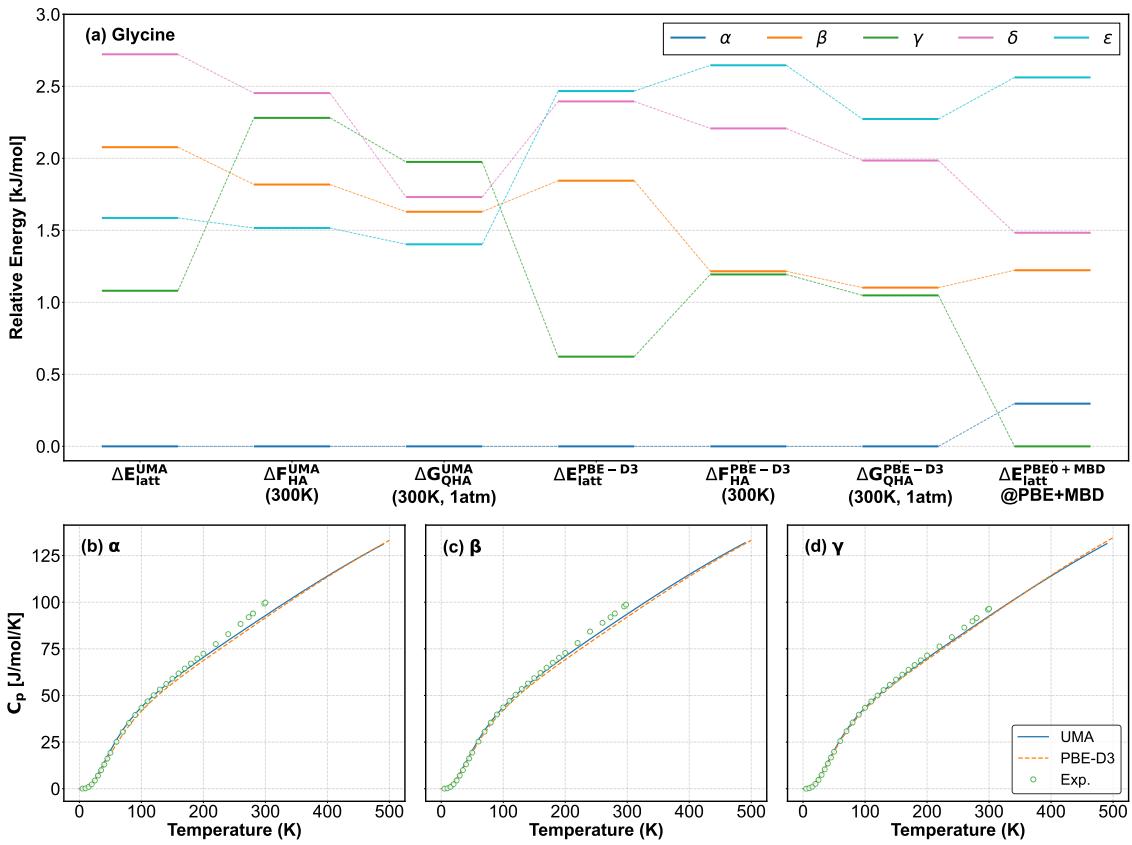


Figure 6 Performance of UMA for the polymorphs of glycine (GLYCIN): (a) Relative energy ranking based on UMA and PBE-D3 lattice energy (ΔE_{latt}), Helmholtz free energy calculated using the HA (ΔF_{HA}), and Gibbs free energy calculated within the QHA at 300 K and 1 atm (ΔG_{QHA}). The lattice energy ranking based on single point energy calculations with PBE0+MBD at the geometry optimized with PBE+MBD is also shown. The constant pressure heat capacity (C_p) as a function of temperature for the (b) α , (c) β , and (d) γ forms of glycine, calculated using UMA and PBE-D3, compared with experimental data [100, 116].

Figure 6a presents a detailed analysis of the ranking of the glycine polymorphs. All five polymorphs are found within 3 kJ/mol of the global minimum. For the three ambient pressure polymorphs, UMA ranks the α form as the most stable one at 0 K, followed by the γ form 1.08 kJ/mol higher in energy, and the β form at 2.08 kJ/mol higher. This ranking is in agreement with our PBE-D3 results and with previous computational studies at the PBE-D3 level [60, 117–119] (the ranking from [60] was extracted from an associated repository

and is provided in SI Table S4). Similar results have been reported in computational studies that employed the PBE functional with other pairwise dispersion corrections [120, 121]. UMA ranks the high-pressure ε form as more stable than the β form, whereas PBE-D3 ranks the ε form as the least stable. With both UMA and PBE-D3, adding free energy corrections destabilizes the γ form with respect to the α form. However, with UMA the γ form becomes the least stable based on Gibbs free energy at 300 K, whereas with PBE-D3 the γ form remains the second most stable. Figure 6b-d shows the temperature dependence of the specific heat of α , β and γ -glycine calculated with UMA and PBE-D3 compared to experimental data [100, 116]. Additional results are provided in SI Figure S11. The UMA results are in close agreement with PBE-D3 and in good agreement with experiment for all three forms of glycine. At 300 K, the specific heat values obtained with both UMA and PBE-D3 are slightly underestimated compared to experiment. This shows that UMA delivers robust performance in free energy calculations, effectively capturing vibrational contributions and thermal expansion. The experimentally observed stability ranking of $\gamma > \alpha > \beta$ is only recovered when re-ranking is performed using the more accurate PBE0 hybrid functional combined with the MBD method for treating dispersion interactions. This is consistent with the results of previous studies, which showed that more accurate DFT methods are needed to correctly rank the glycine polymorphs [59, 121–124] (the ranking from [59] was extracted from an associated repository and is provided in SI Table S4). Because the MLIPs used here were trained on PBE-D3 data, their performance is expected to be comparable to PBE-D3. This level of accuracy is adequate in most cases, but for some systems, such as glycine, re-ranking with more accurate DFT methods may still be needed to resolve very small energy differences between several polymorphs.

The cost of FastCSP workflow, detailed in SI C, highlights its ability to significantly reduce computational costs without compromising on accuracy in CSP. For instance, the end-to-end CSP workflow for glycine without the free energy calculations required only 12 CPU-hours and 205 GPU-hours, a considerable improvement compared to the 2,583 CPU-hours and 187 GPU-hours required by previous methods [59]. By dramatically lowering these computational costs, the FastCSP workflow is expected to accelerate the materials discovery pipeline, enabling industry to reduce the time and expense of developing novel materials.

3.3 MLIP evaluation

We assess the performance of UMA by comparing the relaxed geometries and energy ranking of putative structures to PBE-D3. Further validation of UMA can be found in [56]. The metrics used here reflect the extent to which UMA reproduces the PBE-D3 potential energy surface (PES). If the PBE-D3 results are reproduced reliably, then UMA can completely replace PBE-D3 in CSP workflows with no loss of fidelity.

To assess how well the structures relaxed with UMA match the corresponding structures relaxed with PBE-D3, we calculated RMSD₃₀, using the COMPACK molecular overlay method [125], implemented in the CSD Python API [126]. Each crystal structure is represented by a cluster of 30 molecules. The root mean squared deviation (RMSD) between the two molecular clusters is calculated based on the molecules that match within 35% distance and 35° angle tolerances. This is a community-accepted comparison metric, which was used in the seventh CSP blind test [34, 35]. Two structures are considered as matched if all 30 molecules can be overlaid with RMSD₃₀ < 1 Å. The match rate between structures relaxed with UMA and PBE-D3 is very high with an average of 90% across the benchmark set. A full account is provided in SI Table S5. Overall, the structures relaxed with UMA and PBE-D3 are very similar with an average RMSD₃₀ of 0.22 Å across the benchmark set. Representative RMSD₃₀ histograms are displayed in Figure 7 for glycine, with the lowest mean RMSD₃₀ of 0.09 Å, imidazole, with an average mean RMSD₃₀ of 0.21 Å, and DEZDUH, with the highest average mean RMSD₃₀ of 0.31 Å. Histograms for the remaining targets are provided in SI Figure S6. UMA thus demonstrates excellent relaxation performance for crystal structures of chemically diverse targets with various types of intermolecular interactions and packing motifs.

Figure 8 shows parity plots comparing the relative lattice energies of structures relaxed with UMA to their counterparts relaxed with PBE-D3 for the five polymorphic systems and Target XVI. Plots for the remaining targets, as well as parity plots for single point energies from PBE-D3 obtained at the UMA-relaxed geometries, are provided in SI Figures S7 and S8. Across all targets, UMA produces a mean absolute error (MAE) of 1.16 kJ/mol and a Spearman rank correlation of 0.94. The lowest MAE of 0.51 kJ/mol is obtained for Target II and the highest Spearman rank correlation of 0.98 is obtained for Target VIII and eniluracil. For

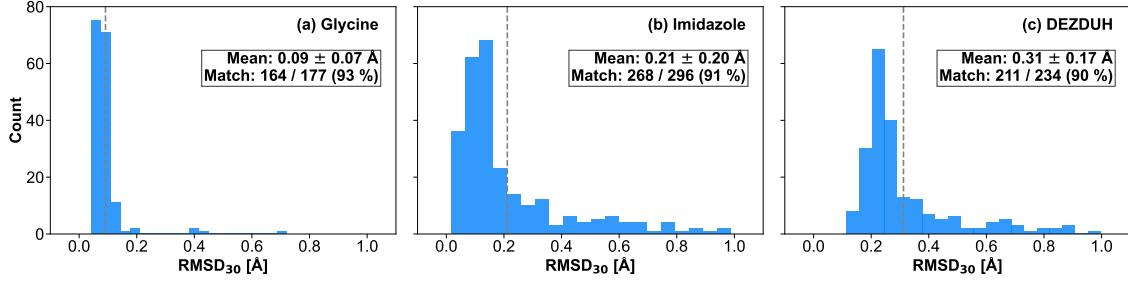


Figure 7 RMSD₃₀ histograms comparing the relaxed crystal structures obtained with the UMA model to those obtained with PBE-D3 for (a) glycine, (b) imidazole, and (c) DEZDUH. The mean RMSD₃₀ is indicated by a vertical dashed line and the match rate is also shown.

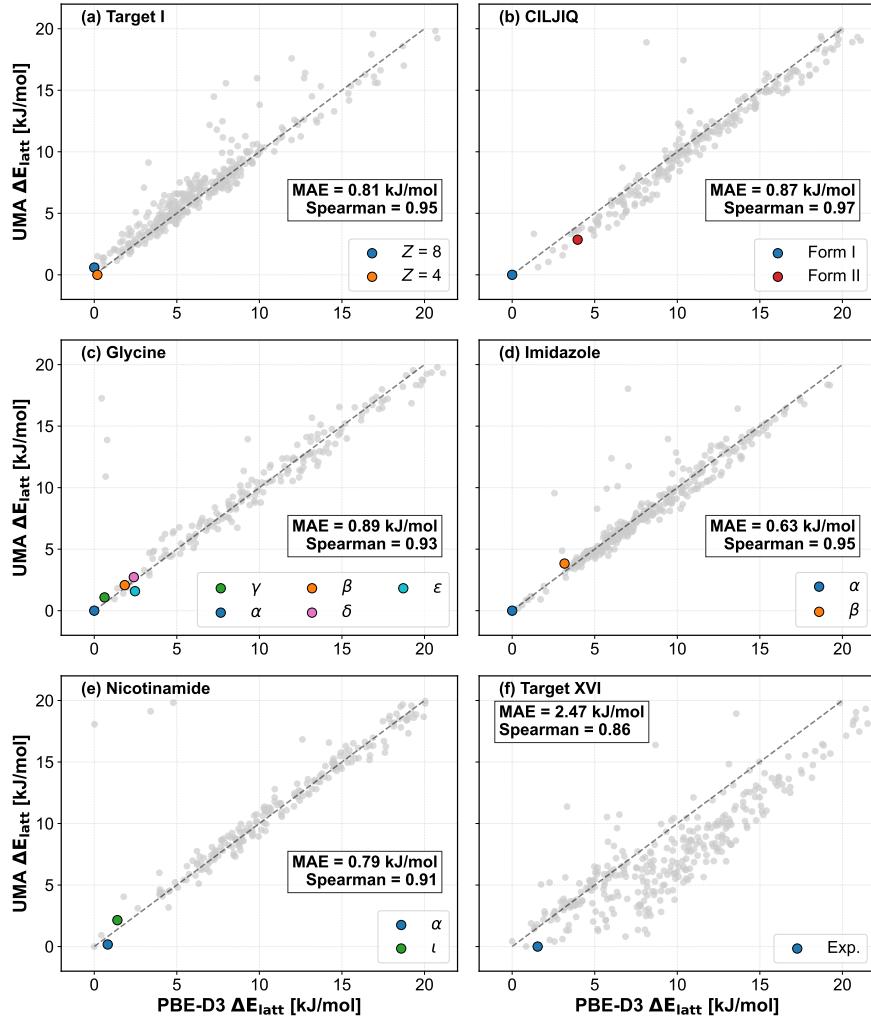


Figure 8 Parity plots comparing the relative lattice energies of structures relaxed with UMA to their counterparts relaxed with PBE-D3 for (a) Target I, (b) CILJIQ, (c) glycine, (d) imidazole, (e) nicotinamide, and (f) Target XVI. The relative energies are referenced to the global minimum of each method.

the polymorphic targets, the stability ordering is for the most part in close agreement with DFT. The only outlier with the lowest Spearman rank correlation, for which significant differences are obtained between UMA and PBE-D3 is Target XVI. The experimental structure of Target XVI from the fifth blind test includes diazide-carbonyl interactions specifically with induction being a problem for many of the empirical potentials and other MLIPs attempted [32, 68]. Moreover, <0.1% of molecules with diazo functional groups are found

in the training split of the OMC25 dataset [56], so this might explain why UMA is not able to capture the energetics of the unique chemical environment in this molecule. Here, the relative energies are referenced to the global minimum of each method, which is consistent with the results presented in Figure 3. We note that the comparison of relative energies can be affected by the choice of reference, in particular in cases that the UMA and DFT global minimum is significantly different, such as Target XVI and HURYUQ, as shown in SI Figures S9 and S10. Based on the results presented here, UMA can be used as an effective and considerably more efficient replacement for PBE-D3 in CSP workflows for mostly rigid compounds.

4 Conclusion

In this work, we have introduced FastCSP, an open-source workflow for molecular crystal structure prediction based on random structure generation followed by geometry optimization and energy ranking by a universal MLIP. We have demonstrated for a chemically and structurally diverse benchmark set of 28 mostly rigid compounds that FastCSP can generate the experimental structures and rank them within the top 10 (with the exception of a polymorph of glycine and a polymorph of imidazole known to be less stable) and 5 kJ/mol of the global minimum. The only generation failure was Form II of CILJIQ, which was not initially generated, but was obtained by increasing the number of structures generated per space group. Target V was the only compound, whose experimental structure was ranked somewhat higher by UMA, as #10 above the global minimum based on lattice energy, however, this ranking was in close agreement with the DFT ranking of #11. The computational efficiency of the UMA MLIPs has enabled us to conduct large-scale free energy calculations for numerous putative structures and perform statistical analysis. Our results indicated that accounting for vibrational and thermal effects typically led to relative energy changes within ± 2.2 kJ/mol compared to the relative lattice energies, and did not necessarily improve the ranking of the experimentally observed structures. Overall, the FastCSP workflow has demonstrated a very robust CSP performance.

Moreover, we have shown that the UMA MLIP reproduces with high fidelity the results of dispersion-inclusive DFT at the PBE-D3 level for relaxation, energy ranking, and free energy calculations at a fraction of the computational cost. The only system, for which the ranking performance compared to PBE-D3 was somewhat worse, was Target XVI, whose unusual chemistry is not well-represented in the OMC25 training dataset. We have thus demonstrated that a single, pretrained MLIP can generalize across chemical space for compounds that vary in their composition, intermolecular interactions, and crystal packing, without requiring system-specific retraining. This effectively eliminates the need for classical force fields in the early stages of the CSP workflow and for final re-ranking with dispersion-inclusive DFT.

The FastCSP workflow is general in the sense that Genarris can generate crystal structures in all space groups and the OMC task of UMA, which was trained on crystal structures of 50,000 chemically diverse molecules containing the elements H, B, C, N, O, F, Si, P, S, Cl, Br, and I, has excellent transferability. It is possible that other UMA tasks or system-specific retraining would be needed for compounds that are not well-represented in the training data because they have an unusual chemical bonding or contain other elements, such as transition metals. In addition, the FastCSP workflow has yet to be tested for flexible molecules with multiple rotatable bonds and for crystals with multiple molecules in the asymmetric unit ($Z' > 1$), co-crystals, salts, hydrates, and solvates. We plan to explore these directions in future work.

The open-source release of FastCSP, along with the Genarris 3.0 structure generator [61], the UMA MLIP [38], and the OMC25 dataset [56], provides the community with the first fully open CSP workflow capable of matching the accuracy of approaches based on dispersion-inclusive DFT and/or system-specific classical force fields or MLIPs. This has far-reaching implications for computational chemistry and materials science because it lowers the barrier to high-throughput screening of molecular solids and enables the scalable exploration of molecular crystal landscapes in a few hours with tens of modern GPUs. We envision our framework being utilized to support the development of products and technologies based on molecular crystals, including pharmaceuticals, agrochemicals, organic electronics, and other domains where polymorph control is critical.

Acknowledgments

N.M. acknowledges support from the National Science Foundation (NSF) Designing Materials to Revolutionize and Engineer our Future (DMREF) program via award DMR-2323749. Y.Y. acknowledges support from the Frontera Computational Science Fellowship awarded by the Texas Advanced Computing Center (TACC).

References

- [1] Beran, G. J. Frontiers of molecular crystal structure prediction for pharmaceuticals and functional organic materials. *Chemical Science* **14**, 13290–13312 (2023).
- [2] Price, S. L., Braun, D. E., and Reutzel-Edens, S. M. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications* **52**, 7065–7077 (2016).
- [3] Yang, M., Dybeck, E., Sun, G., Peng, C., Samas, B., Burger, V. M., Zeng, Q., Jin, Y., Bellucci, M. A., Liu, Y., Zhang, P., Ma, J., Jiang, Y. A., Hancock, B. C., Wen, S., and Wood, G. P. Prediction of the relative free energies of drug polymorphs above zero Kelvin. *Crystal Growth & Design* **20**, 5211–5224 (2020).
- [4] Campbell, J. E., Yang, J., and Day, G. M. Predicted energy–structure–function maps for the evaluation of small molecule organic semiconductors. *Journal of Materials Chemistry C* **5**, 7574–7584 (2017).
- [5] Bhat, V., Callaway, C. P., and Risko, C. Computational approaches for organic semiconductors: from chemical and physical understanding to predicting new materials. *Chemical Reviews* **123**, 7498–7547 (2023).
- [6] Tom, R., Gao, S., Yang, Y., Zhao, K., Bier, I., Buchanan, E. A., Zaykov, A., Havlas, Z., Michl, J., and Marom, N. Inverse design of tetracene polymorphs with enhanced singlet fission performance by property-based genetic algorithm optimization. *Chemistry of Materials* **35**, 1373–1386 (2023).
- [7] Bier, I., O'Connor, D., Hsieh, Y.-T., Wen, W., Hiszpanski, A. M., Han, T. Y.-J., and Marom, N. Crystal structure prediction of energetic materials and a twisted arene with Genarris and GAtor. *CrystEngComm* **23**, 6023–6038 (2021).
- [8] Arnold, J. E., and Day, G. M. Crystal structure prediction of energetic materials. *Crystal Growth & Design* **23**, 6149–6160 (2023).
- [9] O'Connor, D., Bier, I., Tom, R., Hiszpanski, A. M., Steele, B. A., and Marom, N. Ab initio crystal structure prediction of the energetic materials LLM-105, RDX, and HMX. *Crystal Growth & Design* **23**, 6275–6289 (2023).
- [10] Panina, N., Leusen, F. J., Janssen, F., Verwer, P., Meekes, H., Vlieg, E., and Deroover, G. Crystal structure prediction of organic pigments: quinacridone as an example. *Applied Crystallography* **40**, 105–114 (2007).
- [11] Tremayne, M., Grice, L., Pyatt, J. C., Seaton, C. C., Kariuki, B. M., Tsui, H. H., Price, S. L., and Cherryman, J. C. Characterization of complicated new polymorphs of chlorothalonil by X-ray diffraction and computer crystal structure prediction. *Journal of the American Chemical Society* **126**, 7071–7081 (2004).
- [12] Yang, J., Hu, C. T., Zhu, X., Zhu, Q., Ward, M. D., and Kahr, B. DDT polymorphism and the lethality of crystal forms. *Angewandte Chemie* **129**, 10299–10303 (2017).
- [13] Lee, A. Y., Erdemir, D., and Myerson, A. S. Crystal polymorphism in chemical process development. *Annual Review of Chemical and Biomolecular Engineering* **2**, 259–280 (2011).
- [14] Cruz-Cabeza, A. J., Reutzel-Edens, S. M., and Bernstein, J. Facts and fictions about polymorphism. *Chemical Society Reviews* **44**, 8619–8635 (2015).
- [15] Neumann, M. A., van de Streek, J., Fabbiani, F. P., Hidber, P., and Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nature Communications* **6**, 7793 (2015).
- [16] Bernstein, J. *Polymorphism in Molecular Crystals*, vol. 30 (Oxford University Press, 2020).
- [17] 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., Holcombe, F. O., Chiu, Y.-y., and Hussain, A. S. Scientific considerations of pharmaceutical solid polymorphism in abbreviated new drug applications. *Pharmaceutical Research* **20**, 531–536 (2003).
- [18] Coropceanu, V., Cornil, J., da Silva Filho, D. A., Olivier, Y., Silbey, R., and Brédas, J.-L. Charge transport in organic semiconductors. *Chemical Reviews* **107**, 926–952 (2007).

- [19] Censi, R., and Di Martino, P. Polymorph impact on the bioavailability and stability of poorly soluble drugs. *Molecules* **20**, 18759–18776 (2015).
- [20] Chung, H., and Diao, Y. Polymorphism as an emerging design strategy for high performance organic electronics. *Journal of Materials Chemistry C* **4**, 3915–3933 (2016).
- [21] Liu, G., Gou, R., Li, H., and 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* **18**, 4174–4186 (2018).
- [22] Rice, B., LeBlanc, L. M., Otero-de-la Roza, A., Fuchter, M. J., Johnson, E. R., Nelson, J., and Jelfs, K. E. A computational exploration of the crystal energy and charge-carrier mobility landscapes of the chiral [6] helicene molecule. *Nanoscale* **10**, 1865–1876 (2018).
- [23] Yang, J., De, S., Campbell, J. E., Li, S., Ceriotti, M., and Day, G. M. Large-scale computational screening of molecular organic semiconductors using crystal structure prediction. *Chemistry of Materials* **30**, 4361–4371 (2018).
- [24] Schmidt, J. A., Weatherby, J. A., Sugden, I. J., Santana-Bonilla, A., Salerno, F., Fuchter, M. J., Johnson, E. R., Nelson, J., and Jelfs, K. E. Computational screening of chiral organic semiconductors: exploring side-group functionalization and assembly to optimize charge transport. *Crystal Growth & Design* **21**, 5036–5049 (2021).
- [25] Faruque, M. O., Akter, S., Limbu, D. K., Kilway, K. V., Peng, Z., and 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* **24**, 8950–8960 (2024).
- [26] Johal, J., and Day, G. Exploring organic chemical space for materials discovery using crystal structure prediction-informed evolutionary optimisation. *ChemRxiv preprint chemrxiv:2025-v8692* (2025).
- [27] Nyman, J., and Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **17**, 5154–5165 (2015).
- [28] Lommers, 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., Schmidt, M. U., van Eijck, B. P., Verwer, P., and Williams, D. E. A test of crystal structure prediction of small organic molecules. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **56**, 697–714 (2000).
- [29] Motherwell, W. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W., Leusen, F. J., Lommers, J. P., Mooij, W. T., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P., and Williams, D. E. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **58**, 647–661 (2002).
- [30] Day, G. M. *et al.* A third blind test of crystal structure prediction. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **61**, 511–527 (2005).
- [31] Day, G. M. *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, Crystal Engineering and Materials* **65**, 107–125 (2009).
- [32] Bardwell, D. A. *et al.* Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **67**, 535–551 (2011).
- [33] Reilly, A. M. *et al.* Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **72**, 439–459 (2016).
- [34] Hunnisett, L. M. *et al.* The seventh blind test of crystal structure prediction: structure generation methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **80**, 517–547 (2024).
- [35] Hunnisett, L. M. *et al.* The seventh blind test of crystal structure prediction: structure ranking methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **80**, 548–574 (2024).
- [36] Wines, D., and Choudhary, K. CHIPS-FF: Evaluating universal machine learning force fields for material properties. *ACS Materials Letters* **7**, 2105–2114 (2025).
- [37] Loew, A., Sun, D., Wang, H.-C., Botti, S., and Marques, M. A. Universal machine learning interatomic potentials are ready for phonons. *npj Computational Materials* **11**, 1–8 (2025).

- [38] Wood, B. M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J. R., Levine, D. S., Michel, K., Sriram, A., Cohen, T., Das, A., Rizvi, A., Sahoo, S. J., Ulissi, Z. W., and Zitnick, C. L. UMA: A Family of Universal Models for Atoms. *arXiv preprint arXiv:2506.23971* (2025).
- [39] Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, vol. 35, 11423–11436 (2022).
- [40] Fu, X., Wood, B. M., Barroso-Luque, L., Levine, D. S., Gao, M., Dzamba, M., and Zitnick, C. L. Learning smooth and expressive interatomic potentials for physical property prediction. In *International Conference on Machine Learning* (2025).
- [41] Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, vol. 30 (2017).
- [42] Sriram, A., Das, A., Wood, B. M., and Zitnick, C. L. Towards training billion parameter graph neural networks for atomic simulations. In *International Conference on Learning Representations* (2022).
- [43] Gasteiger, J., Becker, F., and Günnemann, S. GemNet: Universal directional graph neural networks for molecules. In *Advances in Neural Information Processing Systems*, vol. 34, 6790–6802 (2021).
- [44] Gasteiger, J., Shuaibi, M., Sriram, A., Günnemann, S., Ulissi, Z. W., Zitnick, C. L., and Das, A. GemNet-OC: Developing graph neural networks for large and diverse molecular simulation datasets. *Transactions on Machine Learning Research (TMLR)* (2022).
- [45] Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 9377–9388 (PMLR, 2021).
- [46] Passaro, S., and Zitnick, C. L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *International Conference on Machine Learning*, 27420–27438 (PMLR, 2023).
- [47] Liao, Y.-L., Wood, B. M., Das, A., and Smidt, T. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations* (2024).
- [48] Levine, D. S., Shuaibi, M., Spotte-Smith, E. W. C., Taylor, M. G., Hasym, M. R., Michel, K., Batatia, I., Csányi, G., Dzamba, M., Eastman, P., Frey, N. C., Fu, X., Gharakhanyan, V., Krishnapriyan, A. S., Rackers, J. A., Raja, S., Rizvi, A., Rosen, A. S., Ulissi, Z., Vargas, S., Zitnick, C. L., Blau, S. M., and Wood, B. M. The Open Molecules 2025 (OMol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762* (2025).
- [49] Sriram, A., Choi, S., Yu, X., Brabson, L. M., Das, A., Ulissi, Z., Uyttendaele, M., Medford, A. J., and Sholl, D. S. The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. *ACS Central Science* **10**, 923–941 (2024).
- [50] Barroso-Luque, L., Shuaibi, M., Fu, X., Wood, B. M., Dzamba, M., Gao, M., Rizvi, A., Zitnick, C. L., and Ulissi, Z. W. Open Materials 2024 (OMat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771* (2024).
- [51] Smith, J. S., Isayev, O., and Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific Data* **4**, 1–8 (2017).
- [52] Smith, J. S., Zubatyuk, R., Nebgen, B., Lubbers, N., Barros, K., Roitberg, A. E., Isayev, O., and Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific Data* **7**, 134 (2020).
- [53] Devereux, C., Smith, J. S., Huddleston, K. K., Barros, K., Zubatyuk, R., Isayev, O., and Roitberg, A. E. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *Journal of Chemical Theory and Computation* **16**, 4192–4202 (2020).
- [54] 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., De Fabritiis, G., and Markland, T. E. SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Scientific Data* **10**, 11 (2023).
- [55] Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **5**, 1031–1041 (2023).

- [56] Gharakhanyan, V., Barroso-Luque, L., Yang, Y., Shuaibi, M., Michel, K., Levine, D. S., Dzamba, M., Fu, X., Gao, M., Liu, X., Ni, H., Noori, K., Wood, B. M., Uyttendaele, M., Boromand, A., Zitnick, C. L., Marom, N., Ulissi, Z. W., and Sriram, A. Open Molecular Crystals 2025 (OMC25) Dataset and Models. *arXiv preprint* (2025).
- [57] Nayal, K. S., O'Connor, D., Zubatyuk, R., Anstine, D. M., Yang, Y., Tom, R., Deng, W., Tang, K., Marom, N., and Isayev, O. Efficient molecular crystal structure prediction and stability assessment with AIMNet2 neural network potentials. *ChemRxiv preprint chemrxiv:2025-ksn4n* (2025).
- [58] Kadan, A., Ryczko, K., Wildman, A., Wang, R., Roitberg, A., and Yamazaki, T. Accelerated organic crystal structure prediction with genetic algorithms and machine learning. *Journal of Chemical Theory and Computation* **19**, 9388–9402 (2023).
- [59] Zhou, D., Bier, I., Santra, B., Jacobson, L. D., Wu, C., Garaizar Suarez, A., Almaguer, B. R., Yu, H., Abel, R., Friesner, R. A., and Wang, L. A robust crystal structure prediction method to support small molecule drug development with large scale validation and blind study. *Nature Communications* **16**, 2210 (2025).
- [60] Weber, J. L., Guha, R. D., Agarwal, G., Wei, Y., Fike, A. A., Xie, X., Stevenson, J., Leswing, K., Halls, M. D., Abel, R., and Jacobson, L. D. Efficient long-range machine learning force fields for liquid and materials properties. *arXiv preprint arXiv:2505.06462* (2025).
- [61] Yang, Y., Tom, R., Wui, J. A., Moussa, J. E., and Marom, N. Genarris 3.0: Generating close-packed molecular crystal structures with rigid press. *ChemRxiv preprint chemrxiv:2025-046zn* (2025).
- [62] Taylor, C. R., Butler, P. W., and Day, G. M. Predictive crystallography at scale: mapping, validating, and learning from 1000 crystal energy landscapes. *Faraday Discussions* **256**, 434–458 (2025).
- [63] Price, L. S., Paloni, M., Salvalaglio, M., and Price, S. L. One size fits all? Development of the CPOSS209 data set of experimental and hypothetical polymorphs for testing computational modeling methods. *Crystal Growth & Design* **25**, 3186–3209 (2025).
- [64] Glick, Z. L., Metcalf, D. P., and Swarthout, S. F. Toward routine CSP of pharmaceuticals: A fully automated protocol using neural network potentials. *arXiv preprint arXiv:2507.16218* (2025).
- [65] Anstine, D. M., and Isayev, O. Machine learning interatomic potentials and long-range physics. *The Journal of Physical Chemistry A* **127**, 2417–2431 (2023).
- [66] Žugec, I., Geilhufe, R. M., and Lončarić, I. Global machine learning potentials for molecular crystals. *The Journal of Chemical Physics* **160**, 154106 (2024).
- [67] Omee, S. S., Fu, N., Dong, R., Hu, M., and Hu, J. Structure-based out-of-distribution (OOD) materials property prediction: a benchmark study. *npj Computational Materials* **10**, 144 (2024).
- [68] Nickerson, C. J., and Johnson, E. R. Assessment of a foundational machine-learned potential for energy ranking of molecular crystal polymorphs. *Physical Chemistry Chemical Physics* **27**, 11930–11940 (2025).
- [69] Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).
- [70] Groom, C. R., Bruno, I. J., Lightfoot, M. P., and Ward, S. C. The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **72**, 171–179 (2016).
- [71] Ding, Y., and Krogh-Jespersen, K. The glycine zwitterion does not exist in the gas phase: results from a detailed ab initio electronic structure study. *Chemical physics letters* **199**, 261–266 (1992).
- [72] Tom, R., Rose, T., Bier, I., O'Brien, H., Vázquez-Mayagoitia, Á., and Marom, N. Genarris 2.0: A random structure generator for molecular crystals. *Computer Physics Communications* **250**, 107170 (2020).
- [73] Bier, I., and 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* **124**, 10330–10345 (2020).
- [74] Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [75] Togo, A., Chaput, L., Tadano, T., and Tanaka, I. Implementation strategies in phonopy and phono3py. *Journal of Physics: Condensed Matter* **35**, 353001 (2023).

- [76] Togo, A. First-principles phonon calculations with phonopy and phono3py. *Journal of the Physical Society of Japan* **92**, 012001 (2023).
- [77] Vinet, P., Rose, J. H., Ferrante, J., and Smith, J. R. Universal features of the equation of state of solids. *Journal of Physics: Condensed Matter* **1**, 1941 (1989).
- [78] Blum, V., Gehrke, R., Hanke, F., Havu, P., Havu, V., Ren, X., Reuter, K., and Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications* **180**, 2175–2196 (2009).
- [79] Ren, X., Rinke, P., Blum, V., Wieferink, J., Tkatchenko, A., Sanfilippo, A., Reuter, K., and 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* **14**, 053020 (2012).
- [80] Zhang, I. Y., Ren, X., Rinke, P., Blum, V., and Scheffler, M. Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar. *New Journal of Physics* **15**, 123033 (2013).
- [81] Adamo, C., and Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **110**, 6158–6170 (1999).
- [82] Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **77**, 3865 (1996).
- [83] Tkatchenko, A., DiStasio Jr, R. A., Car, R., and Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Physical Review Letters* **108**, 236402 (2012).
- [84] Ambrosetti, A., Reilly, A. M., DiStasio, R. A., and Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *The Journal of Chemical Physics* **140**, 18A508 (2014).
- [85] Hermann, J., Stöhr, M., Góger, S., Chaudhuri, S., Aradi, B., Maurer, R. J., and Tkatchenko, A. libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *The Journal of Chemical Physics* **159**, 174802 (2023).
- [86] Kresse, G., and Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558 (1993).
- [87] Kresse, G., and Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54**, 11169 (1996).
- [88] Kresse, G., and Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758 (1999).
- [89] Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**, 154104 (2010).
- [90] CCDC. CSD Space Group Statistics – Space Group Frequency Ordering (2025). Published: 2025-01-01.
- [91] Taniguchi, T., and Fukasawa, R. Crystal structure prediction of organic molecules by machine learning-based lattice sampling and structure relaxation. *ChemRxiv preprint chemrxiv:2025-65l01* (2025).
- [92] Matos Beja, A., Paixão, J., Ramos Silva, M., Alte da Veiga, L., Rocha Gonsalves, A. d., and Serra, A. 2-bromo-5-hydroxybenzaldehyde. *Acta Crystallographica Section C: Structural Chemistry* **56**, 354–355 (2000).
- [93] Silva, M. R., Paixao, J. A., Beja, A. M., Sobral, A. J., and Rocha Gonsalves, A. A new polymorph of 2-bromo-5-hydroxybenzaldehyde. *Acta Crystallographica Section E: Crystallographic Communications* **60**, o84–o85 (2004).
- [94] McMullan, R., Epstein, J., Ruble, J. R., and Craven, B. The crystal structure of imidazole at 103 K by neutron diffraction. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **35**, 688–691 (1979).
- [95] Paliwoda, D., Dziubek, K. F., and Katrusiak, A. Imidazole hidden polar phase. *Crystal Growth & Design* **12**, 4302–4305 (2012).
- [96] Li, X., Ou, X., Wang, B., Rong, H., Wang, B., Chang, C., Shi, B., Yu, L., and Lu, M. Rich polymorphism in nicotinamide revealed by melt crystallization and crystal structure prediction. *Communications Chemistry* **3**, 152 (2020).
- [97] Boldyreva, E., Drebushchak, V., Drebushchak, T., Paukov, I., Kovalevskaya, Y. A., and Shutova, E. Polymorphism of glycine, Part I. *Journal of Thermal Analysis and Calorimetry* **73**, 409–418 (2003).

- [98] Perlovich, G., Hansen, L. K., and Bauer-Brandl, A. The polymorphism of glycine. thermochemical and structural aspects. *Journal of Thermal Analysis and Calorimetry* **66**, 699–715 (2001).
- [99] Chongprasert, S., Knopp, S. A., and Nail, S. L. Characterization of frozen solutions of glycine. *Journal of Pharmaceutical Sciences* **90**, 1720–1728 (2001).
- [100] Drebushchak, V., Kovalevskaya, Y., Paukov, I., and Boldyreva, E. Low-temperature heat capacity of α and γ polymorphs of glycine. *Journal of Thermal Analysis and Calorimetry* **74**, 109–120 (2003).
- [101] Dawson, A., Allan, D. R., Belmonte, S. A., Clark, S. J., David, W. I., McGregor, P. A., Parsons, S., Pulham, C. R., and Sawyer, L. Effect of high pressure on the crystal structures of polymorphs of glycine. *Crystal Growth & design* **5**, 1415–1427 (2005).
- [102] Beyer, T., and Price, S. L. Dimer or catemer? low-energy crystal packings for small carboxylic acids. *The Journal of Physical Chemistry B* **104**, 2647–2655 (2000).
- [103] Das, D., and Desiraju, G. R. Packing modes in some mono-and disubstituted phenylpropionic acids: Repeated occurrence of the rare syn, anti catemer. *Chemistry—An Asian Journal* **1**, 231–244 (2006).
- [104] Aakeröy, C. B., Sinha, A. S., Epa, K. N., Chopade, P. D., Smith, M. M., and Desper, J. Structural chemistry of oximes. *Crystal Growth & Design* **13**, 2687–2695 (2013).
- [105] Hoja, J., Reilly, A. M., and Tkatchenko, A. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **7**, e1294 (2017).
- [106] 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* **211**, 9–30 (2018).
- [107] Hoja, J., and Tkatchenko, A. First-principles stability ranking of molecular crystal polymorphs with the DFT+MBD approach. *Faraday Discussions* **211**, 253–274 (2018).
- [108] O'Connor, D., Bier, I., Hsieh, Y.-T., and Marom, N. Performance of dispersion-inclusive density functional theory methods for energetic materials. *Journal of Chemical Theory and Computation* **18**, 4456–4471 (2022).
- [109] Weatherby, J. A., Rumson, A. F., Price, A. J., Otero de la Roza, A., and Johnson, E. R. A density-functional benchmark of vibrational free-energy corrections for molecular crystal polymorphism. *The Journal of Chemical Physics* **156**, 114108 (2022).
- [110] Nyman, J., and Day, G. M. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics* **18**, 31132–31143 (2016).
- [111] Curtis, F., Li, X., Rose, T., Vazquez-Mayagoitia, A., Bhattacharya, S., Ghiringhelli, L. M., and Marom, N. Gator: A first-principles genetic algorithm for molecular crystal structure prediction. *Journal of Chemical Theory and Computation* **14**, 2246–2264 (2018).
- [112] Whittleton, S. R., Otero-De-La-Roza, A., and Johnson, E. R. Exchange-hole dipole dispersion model for accurate energy ranking in molecular crystal structure prediction. *Journal of Chemical Theory and Computation* **13**, 441–450 (2017).
- [113] Fellah, N., Zhang, C. J., Chen, C., Hu, C. T., Kahr, B., Ward, M. D., and Shtukenberg, A. G. Highly polymorphous nicotinamide and isonicotinamide: solution versus melt crystallization. *Crystal Growth & Design* **21**, 4713–4724 (2021).
- [114] Fellah, N., Tahsin, L., Zhang, C. J., Kahr, B., Ward, M. D., and Shtukenberg, A. G. Efficient polymorph screening through crystallization from bulk and confined melts. *Crystal Growth & Design* **22**, 7527–7543 (2022).
- [115] Price, A. J., Mayo, R. A., Otero-de-la Roza, A., and Johnson, E. R. Accurate and efficient polymorph energy ranking with XDM-corrected hybrid DFT. *CrystEngComm* **25**, 953–960 (2023).
- [116] Drebushchak, V., Boldyreva, E., Kovalevskaya, Y. A., Paukov, I., and Drebushchak, T. Low-temperature heat capacity of β -glycine and a phase transition at 252 K. *Journal of Thermal Analysis and Calorimetry* **79**, 65–70 (2005).
- [117] Xu, W., Zhu, Q., and Hu, C. The structure of glycine dihydrate: implications for the crystallization of glycine from solution and its structure in outer space. *Angewandte Chemie* **129**, 2062–2066 (2017).
- [118] Rybin, N., Novikov, I. S., and Shapeev, A. Accelerating structure prediction of molecular crystals using actively trained moment tensor potential. *Physical Chemistry Chemical Physics* **27**, 5141–5148 (2025).

- [119] Xavier Jr, N. F., da Silva Jr, A. M., and Bauerfeldt, G. F. What rules the relative stability of α -, β -, and γ -glycine polymorphs? *Crystal Growth & Design* **20**, 4695–4706 (2020).
- [120] Rodríguez, J., Costa, G., Da Silva, M., Silva, B., Honório, L., de Lima-Neto, P., Santos, R., Caetano, E., Alves, H., and Freire, V. Structural and optoelectronic properties of the α -, β -, and γ -glycine polymorphs and the glycine dihydrate crystal: A DFT study. *Crystal Growth & Design* **19**, 5204–5217 (2019).
- [121] Marom, N., DiStasio Jr, R. A., Atalla, V., Levchenko, S., Reilly, A. M., Chelikowsky, J. R., Leiserowitz, L., and Tkatchenko, A. Many-body dispersion interactions in molecular crystal polymorphism. *Angewandte Chemie International Edition* **52**, 6629–6632 (2013).
- [122] Reilly, A. M., and Tkatchenko, A. Van der Waals dispersion interactions in molecular materials: beyond pairwise additivity. *Chemical Science* **6**, 3289–3301 (2015).
- [123] Beran, G. J. Modeling polymorphic molecular crystals with electronic structure theory. *Chemical Reviews* **116**, 5567–5613 (2016).
- [124] Kapil, V., and Engel, E. A. A complete description of thermodynamic stabilities of molecular crystals. *Proceedings of the National Academy of Sciences* **119**, e2111769119 (2022).
- [125] Chisholm, J. A., and Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *Applied Crystallography* **38**, 228–231 (2005).
- [126] Sykes, R. A., Johnson, N. T., Kingsbury, C. J., Harter, J., Maloney, A. G. P., Sugden, I. J., Ward, S. C., Bruno, I. J., Adcock, S. A., Wood, P. A., McCabe, P., Moldovan, A. A., Atkinson, F., Giangreco, I., and Cole, J. C. What has scripting ever done for us? The CSD Python application programming interface (API). *Applied Crystallography* **57**, 1235–1250 (2024).
- [127] Ganose, A. M. *et al.* Atomate2: Modular workflows for materials science. *Digital Discovery* **4**, 1944–1973 (2025).
- [128] Callen, H. B. *Thermodynamics and an Introduction to Thermostatistics* (John Wiley & Sons, 2006).

Supplementary Information

A Compounds considered for FastCSP benchmark

Table S1 Summary of the benchmark molecules, listing compound names, corresponding selected CSD reference codes, molecular formula, number of molecules per unit cell (Z), space groups, polymorph details. All structures are measured under ambient pressure except for the two high-pressure glycine polymorphs and β form of imidazole.

No.	Compound	CSD Refcode	Formula	Z	Space Group	Polymorph
1	Target I	XULDUD†	C_6H_6O	8	61 ($Pbca$)	orthorhombic
		XULDUD01		4	14 ($P2_1/c$)	
2	Target II	GUFJOG	C_5H_3NOS	4	14 ($P2_1/n$)	-
3	Target IV	BOQQUT	$C_8H_{11}NO_2$	4	14 ($P2_1/a$)	-
4	Target V	BOQWIN	$C_{10}H_{14}BrNO_2S$	4	19 ($P2_12_12_1$)	-
5	Target VIII	PAHYON01	$C_3H_4N_2O_2$	8	15 ($C2/c$)	-
6	Target XII	AXOSOW01	C_3H_4O	8	61 ($Pbca$)	-
7	Target XIII	SOXLEX01	$C_6H_2Br_2ClF$	4	14 ($P2_1/c$)	-
8	Target XVI	OBEQUJ	$C_6H_4N_2O$	8	61 ($Pbca$)	-
9	Target XVII	OBEQOD	$C_6H_2Cl_2N_2O_4$	4	14 ($P2_1/c$)	-
10	Target XXII	NACJAF	$C_8N_4S_3$	4	14 ($P2_1/n$)	-
11	Acetic Acid	ACETAC03	$C_2H_4O_2$	4	33 ($Pna2_1$)	orthorhombic
12	Caprylolactam	CAPRYL	$C_8H_{15}NO$	4	9 (Cc)	-
13	CEBYUD	CEBYUD	C_4N_4S	3	145 ($P3_2$)	-
14	CILJIQ	CILJIQ†	$C_7H_5BrO_2$	4	19 ($P2_12_12_1$)	Form I
		CILJIQ01		4	14 ($P2_1/c$)	Form II
15	CUMJOJ	CUMJOJ	$C_8H_{12}N_2O_2$	8	114 ($P\bar{4}2_1c$)	-
16	DEZDUH	DEZDUH	$C_8H_6F_6$	4	33 ($Pna2_1$)	-
17	Eniluracil	DOHFEM*	$C_6H_4N_2O_2$	4	14 ($P2_1/n$)	-
18	GACGAU	GACGAU	$C_8H_8O_3$	8	61 ($Pbca$)	-
19	Glycine	GLYCIN16		3	145 ($P3_2$)	γ
		GLYCIN20†		4	14 ($P2_1/n$)	α
		GLYCIN32	$C_2H_5NO_2$	2	4 ($P2_1$)	β
		GLYCIN67		4	14 ($P2_1/a$)	δ [1.9 GPa] [101]
		GLYCIN68		2	7 (Pn)	ε [4.3 GPa] [101]
20	GOLHIB	GOLHIB	$C_{11}H_{10}N_2O$	4	9 (Cc)	-
21	HURYUQ	HURYUQ	$C_{11}H_{14}O_3$	1	1 ($P1$)	-
22	IHEPUG	IHEPUG	$C_8H_{11}NO_2$	8	92 ($P4_12_12$)	-
23	Imidazole	IMAZOL06†	$C_3H_4N_2$	4	14 ($P2_1/c$)	α
		IMAZOL25		8	41 ($Aba2$)	β [1.2 GPa] [95]
24	LECZOL	LECZOL	$C_{12}H_6O_3$	6	170 ($P6_5$)	-
25	Nicotinamide	NICOAM03†	$C_6H_6N_2O$	4	14 ($P2_1/c$)	α
		NICOAM17		4	14 ($P2_1/c$)	ι
26	ROHBUL	ROHBUL	$C_5H_8N_4O_2$	8	61 ($Pbca$)	-
27	6-Fluorochromone	UMIMIO	$C_9H_5FO_2$	1	1 ($P1$)	-
28	WEXREY	WEXREY	$C_6H_5N_3$	4	33 ($Pna2_1$)	-

† The polymorph used to extract the starting molecular conformer.

* DOHFEM is disordered over two sites in a ratio 0.74:0.26. Here we chose the major component.

B Benchmark Setup and Parameters

Conformer Selection. For each target, a single conformer was extracted from the experimental crystal structure in CSD (see Table S1 for the systems with more than one polymorph selected). Except glycine (GLYCIN) due to its zwitterion not existing in the gas phase [71], the conformers were relaxed using the FHI-aims all-electron electronic structure code [78–80] (version 240507) with the PBE0 [81] hybrid functional, which is based on the Perdew–Burke–Ernzerhof (PBE) [82] generalized gradient approximation, combined with the many-body dispersion (MBD) correction method [83–85]. The relaxations were performed using the trust radius enhanced variant of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm as implemented in FHI-aims, with the Tier 2 basis sets of FHI-aims, *tight* numerical settings and a force convergence criterion of 0.001 eV/Å. By measuring RMSD between conformers before and after relaxation, the average RMSD across 27 targets is 0.086 ± 0.029 Å, with HURYUQ having the highest RMSD of 0.139 Å.

```
xc pbe0
spin none
relativistic atomic_zora scalar
occupation_type gaussian 0.01
mixer pulay
n_max_pulay 8
charge 0
charge_mix_param 0.2
sc_accuracy_rho 1e-05
sc_accuracy_eev 1e-03
sc_accuracy_etot 1e-06
sc_accuracy_forces 1e-04
sc_iter_limit 10000
KS_method parallel
empty_states 6
basis_threshold 1e-05
many_body_dispersion
relax_geometry trm 1.e-3
```

Listing S1 The general settings in FHI-aims control.in file used for experimental conformer relaxations.

Structure Generation. For structure generation, we used Genarris 3.0 package [61] with 500 structures generated per space group. Genarris 3.0 automatically identified all space groups compatible with the molecular point group symmetry and the requested number of molecules per unit cell ($Z \in \{1, 2, 3, 4, 6, 8\}$). The target unit cell volume was determined using Genarris 3.0 default setting, scaling the estimated volume by a factor of 1.5 to ensure sufficient placement space. Structures were generated until the desired count per space group (500) was reached. We applied the Symmetric Rigid Press algorithm to compress the unit cell under symmetry constraints and encourage dense molecular packing, using a BFGS optimizer with a convergence tolerance of 0.01. Listing S2 shows the full Genarris configuration file used for our experiments.

```
[master]
name = <target name>
molecule_path = <path to conformer file>
Z = <selected Z value>

[generation]
generation_type = crystal
num_structures_per_spg = 500
spg_distribution_type = standard
specific_radius_proportion = 0.95
max_attempts_per_spg = 100000000
tol = 0.01
unit_cell_volume_mean = predict
volume_mult = 1.5
```

```

max_attempts_per_volume = 10000000
natural_cutoff_mult = 1.2

[symm_rigid_press]
sr = 0.85
method = BFGS
natural_cutoff_mult = 1.2
tol = 0.01
debug_flag = False

```

Listing S2 The base Genarris 3.0 configuration file used for random molecular crystal generation that is run for each Z value.

To reduce redundancy before relaxation, we deduplicated structures using PYMATGEN’s STRUCTURE-MATCHER [69] with relatively loose tolerances: a fractional length tolerance (`itol`) of 0.3, site tolerance (`stol`) of 0.4, and angle tolerance (`angle_tol`) of 5°. On average, this removed nearly 65% of the generated structures, substantially reducing the computational load in the relaxation stage.

MLIP Relaxations. All geometry relaxations were performed using the UMA-Small version 1.1 model with the OMC task [38]. Relaxations were carried out with the BFGS optimizer implemented in ASE and force convergence threshold of 0.01 eV/Å. Both atomic positions and lattice parameters were optimized using ASE’s FRECHETCELLFILTER class without enforcing any symmetry constraints. Each optimization was capped at 1,000 maximum steps, and any structures that did not converge were discarded.

Following geometry relaxations, we applied a second round of structure deduplication using tighter matching criteria (`itol` = 0.2, `stol` = 0.3, `angle_tol` = 5°). We also check for any structures with altered bonding or molecular integrity and discard them, although we found this to be extremely rare. Finally, for downstream analysis, we retained all unique polymorphs within 20 kJ/mol for DFT reranking for validation and 10 kJ/mol for free energy calculations of the lowest-energy structure.

DFT Reranking for Validation. We reranked the UMA-relaxed structures with dispersion-corrected DFT to benchmark the accuracy of the ML-based energy and rank predictions. All DFT calculations used the same computational settings as those employed in the generation of the OMC25 dataset [56], and we refer the reader to the OMC25 paper for full methodological details. Briefly, we used the VASP version 6.3 [86–88] for all calculations with VASP 5.4 PBE pseudopotentials. Input sets were created using RELAXSETGENERATOR from ATOMATE2 [127]. The atomic positions and lattice vectors were relaxed until the maximum per-atom residual forces fell below 0.001 eV/Å, or the relaxation process exceeded 1,500 steps. The total energy convergence tolerance was set to 0.001 meV, and the plane-wave energy cut-off was fixed at 520 eV. The default k-point density in PYMATGEN [69] (using the Γ -centered strategy) was applied.

MLIP Free Energy Calculations. Harmonic free energy calculations were performed using the finite displacement supercell method implemented in PHONOPY [75, 76]. Force constants were calculated using supercells of relaxed structures with lattice vectors of at least 15 Å and atom displacements of 0.01 Å. Γ -Centered meshes, with the number of points determined by the nearest integer to $100 \times a$, where a is the lattice parameter, were used to compute the phonon density of states and thermal properties, including Helmholtz free energy, entropy, and heat capacity at constant volume. Calculations were performed at temperature intervals from 0 to 500 K, in increments of 10 K.

Quasiharmonic free energy calculations were carried out by performing harmonic free energy calculations over seven fixed volumes, ranging from -6% to 6% of the 0K equilibrium volume. For each volume, geometry relaxations of the atomic positions and the cell shape were conducted with a convergence criterion of 0.01 eV/Å, allowing up to 1,000 relaxation steps. The Helmholtz free energies $F(T, V)$ computed from the seven harmonic calculations at different volumes V were used to fit a Vinet equation of state (EOS) [77] to determine the equilibrium volume at each different temperature. Then the Gibbs free energy $G(T, P)$ was approximated by Legendre transformation,

$$G(T, P) = \min_V \{F(T, V) + PV\} \quad (1)$$

Where $F(T, V)$ is the Helmholtz free energy, V is the cell volume, T is temperature, and P is pressure.

Finally, the heat capacity at constant pressure (C_P) was estimated using the following thermodynamic relation,

$$C_P = C_V + \alpha^2 B_T V T \quad (2)$$

Where C_V is the heat capacity at constant volume, α is the thermal expansion coefficient, and B_T is the isothermal bulk modulus [128].

DFT Validation for Glycine. We performed DFT free energy calculations on five matched polymorphs of glycine using FHI-aims (version 240920_2) using the Tier 1 basis sets and *light* species defaults of FHI-aims. The number of k-points (n) was determined by $n \times a \geq 25 \text{ \AA}$, where a is the lattice parameter in a given direction. Geometry relaxations were performed using the trust radius enhanced variant of the BFGS algorithm as implemented in FHI-aims, with a force convergence criterion of 0.01 eV/Å. Harmonic free energy calculations were performed at PBE-D3 level using the finite displacement supercell method implemented in PHONOPY [75, 76]. Force constants were calculated using $2 \times 2 \times 2$ supercells of relaxed structures with atom displacements of 0.01 Å. Γ -Centered meshes, the number of q -points was determined by $n \times a \geq 100$. Calculations were performed at temperature intervals from 0 to 500 K, in increments of 10 K.

Quasiharmonic free energy calculations were carried out by performing harmonic free energy calculations over six fixed volumes, ranging from -6% to 10% of the 0K equilibrium volume. For each volume, volume-constrained geometry relaxations were conducted with a convergence criterion of 0.01 eV/Å. The Helmholtz free energies $F(T, V)$ computed from the six harmonic calculations at different volumes V were used to fit a Vinet equation of state following the procedure outlined above for the MLIP free energy calculations.

The constant pressure heat capacity (C_P) can be calculated by,

$$C_P(T, P) = -T \frac{\partial^2 G(T, P)}{\partial T^2} \quad (3)$$

The five predicted polymorphs of glycine were further re-optimized with PBE, combined with the many-body dispersion (MBD) correction method [83–85] using the Tier 2 basis sets and *tight* species defaults of FHI-aims. These structures were re-ranked based on single-point energy evaluations using the PBE-based hybrid functional, PBE0 [81], paired with the MBD method using the Tier 1 basis sets and *light* species defaults of FHI-aims.

Matching Criteria. To assess the matching criteria between the predicted and experimental crystal structures, we applied the COMPACK crystal packing similarity method [125], implemented in the CSD Python API [126]. Following the conventions of the seventh CSP blind test [34, 35], the similarity measure, RMSD₃₀, is the root mean squared deviation calculated from the atomic positions of matched molecules between two clusters of 30 molecules. Matching molecules were identified using 35% distance and 35° angle tolerances, excluding hydrogen atoms and not allowing molecular differences. Two crystal structures were considered as matching if 30 molecules are overlaid and RMSD₃₀ < 0.5 Å. Where several structures matched the experimentally observed one, we report only the result for the matched structure with the lowest relative energy.

C Runtime Analysis

A runtime analysis of FastCSP is provided in Table S2. CPU and GPU runtimes are given for a random subset of the molecules explored in this work, broken down across the different components: crystal generation, MLIP relaxations, and postprocessing (deduplication + ranking). For glycine, FastCSP required 11.4 CPU-hours and 205.0 GPU-hours (184.0 not accounting for optional free energy calculations), compared to 2583 CPU-hours and 187 GPU-hours in the recent work [59]. While left for future work, further UMA structure relaxation speedups are possible through batching and torch compilation [38]. FastCSP’s speed-ups are a direct consequence of the performance of UMA, not requiring additional DFT relaxations. Although this work used UMA to fully relax all crystal structures, better ways to down-select and filter structures before running a full relaxation could provide even more efficiency gains. NVIDIA V100 16GB and 32GB GPUs were used for all GPU evaluations.

Table S2 FastCSP runtime analysis for several compounds.

No.	Compound	Crystal Generation		MLIP Relaxations		MLIP Free Energy	Postprocess [CPU-hrs]
		Genarris [CPU-hrs]	# of structures	UMA [GPU-hrs]	# of structures		
11	Acetic Acid	13.8	13,608	228.55	182	50.4	1.36
15	CUMJOJ	16.2	26,863	812.88	57	581.7	0.44
19	Glycine	10.6	14,008	183.94	92	21.0	0.77
26	ROHBUL	11.1	26,193	640.10	11	222.1	0.08
28	WEXREY	10.3	22,501	537.91	146	29.4	2.04

D Crystal Structure Prediction Results

Table S3 Summary of the CSP results for the benchmark systems, listing compound names, corresponding CSD reference codes, and performance metrics for experimental matches with different methods and at different stages of FastCSP. Highlighted is the case of CILJIQ.

No.	Compound	CSD Refcode	UMA Relaxed			PBE-D3			UMA Free Energy [T = 300 K]			
			RMSD ₃₀ [\AA]	ΔE [kJ/mol]	Rank	RMSD ₃₀ [\AA]	ΔE [kJ/mol]	Rank	ΔF _H [kJ/mol]	Rank	ΔG _{QH} [kJ/mol]	Rank
1	Target I	XULDUD†	0.20	0.59	2	0.11	0.00	1	0.45	5	0.55	5
		XULDUD01	0.13	0.00	1	0.06	0.19	3	1.38	22	1.29	14
2	Target II	GUFJOG	0.35	1.33	4	0.58	1.10	3	0.26	2	0.59	5
3	Target IV	BOQQUT	0.14	0.00	1	0.20	0.00	1	0.00	1	0.00	1
4	Target V	BOQWIN	0.17	4.87	10	0.22	4.86	11	3.61	11	3.88	12
5	Target VIII	PAHYON01	0.23	0.00	1	0.20	0.00	1	0.00	1	0.00	1
6	Target XII	AXOSOW01	0.24	0.00	1	0.13	0.00	1	0.45	5	0.72	3
7	Target XIII	SOXLEX01	0.09	0.00	1	0.07	0.71	2	0.00	1	0.00	1
8	Target XVI	OBEQUJ	0.15	0.00	1	0.17	1.54	7	0.87	6	1.09	9
9	Target XVII	OBEQOD	0.21	0.00	1	0.05	0.00	1	0.00	1	0.00	1
10	Target XXII	NACJAF	0.19	1.36	2	0.12	1.46	3	1.49	2	1.07	2
11	Acetic Acid	ACETAC03	0.08	0.00	1	0.10	0.00	1	0.00	1	2.86	2
12	Caprylolactam	CAPRYL	0.15	0.20	2	0.18	0.78	3	0.00	1	0.00	1
13	CEBYUD	CEBYUD	0.10	0.16	2	0.04	0.00	1	0.23	2	0.74	3
		CILJIQ ^{†△}	0.26	0.00	1	0.24	0.00	1	0.74	2	0.44	2
14	CILJIQ	CILJIQ01 [△]	0.12	2.86	7	0.13	3.96	11	1.65	5	3.50	18
		CILJIQ ^{†▽}	0.24	0.00	1	0.23	0.00	1	-	-	-	-
		CILJIQ01 [▽]	-	-	-	-	-	-	-	-	-	-
15	CUMJOJ	CUMJOJ	0.16	0.72	2	0.08	0.75	3	1.34	2	1.02	2
16	DEZDUH	DEZDUH	0.20	0.00	1	0.16	0.45	2	0.00	1	0.00	1
17	Eniluracil	DOHFEM*	0.13	0.00	1	0.11	0.00	1	0.00	1	0.00	1
18	GACGAU	GACGAU	0.21	0.52	4	0.15	0.54	2	1.50	8	0.93	8
		GLYCIN16	0.13	1.08	3	0.06	0.62	3	2.28	8	1.98	12
		GLYCIN20†	0.09	0.00	1	0.08	0.00	1	0.00	1	0.00	1
19	Glycine	GLYCIN32	0.11	2.08	9	0.07	1.84	9	1.82	5	1.63	8
		GLYCIN67	0.24	2.72	11	0.20	2.40	12	2.45	11	1.73	10
		GLYCIN68	0.36	1.59	5	0.31	2.47	13	1.52	3	1.40	4
20	GOLHIB	GOLHIB	0.06	0.00	1	0.06	0.00	1	0.00	1	0.00	1
21	HURYUQ	HURYUQ	0.12	0.00	1	0.16	2.78	4	1.22	2	0.36	2
22	IHEPUG	IHEPUG	0.13	0.49	3	0.19	1.27	5	0.56	2	0.05	2
23	Imidazole	IMAZOL06†	0.20	0.00	1	0.13	0.00	1	0.00	1	0.00	1
		IMAZOL25	0.31	3.83	22	0.18	3.17	13	4.53	37	4.08	22
24	LECZOL	LECZOL	0.23	1.04	8	0.15	0.27	2	0.83	4	0.08	2
25	Nicotinamide	NICOAM03†	0.23	0.17	2	0.21	0.82	4	0.66	3	1.04	2
		NICOAM17	0.35	2.15	4	0.28	1.40	5	1.40	4	2.41	4
26	ROHBUL	ROHBUL	0.04	0.00	1	0.08	0.00	1	0.00	1	0.00	1
27	6-Fluorochromone	UMIMIO	0.22	0.67	6	0.08	0.00	1	0.00	1	0.00	1
28	WEXREY	WEXREY	0.29	0.00	1	0.16	0.00	1	0.00	1	0.00	1

† The polymorph used to extract the starting molecular conformer.

* DOHFEM is disordered over two sites in a ratio 0.74:0.26. Here we chose the major component.

△ Obtained from a larger initial pool generated by 1,000 structures per space group.

▽ Obtained from a default initial pool generated by 500 structures per space group.

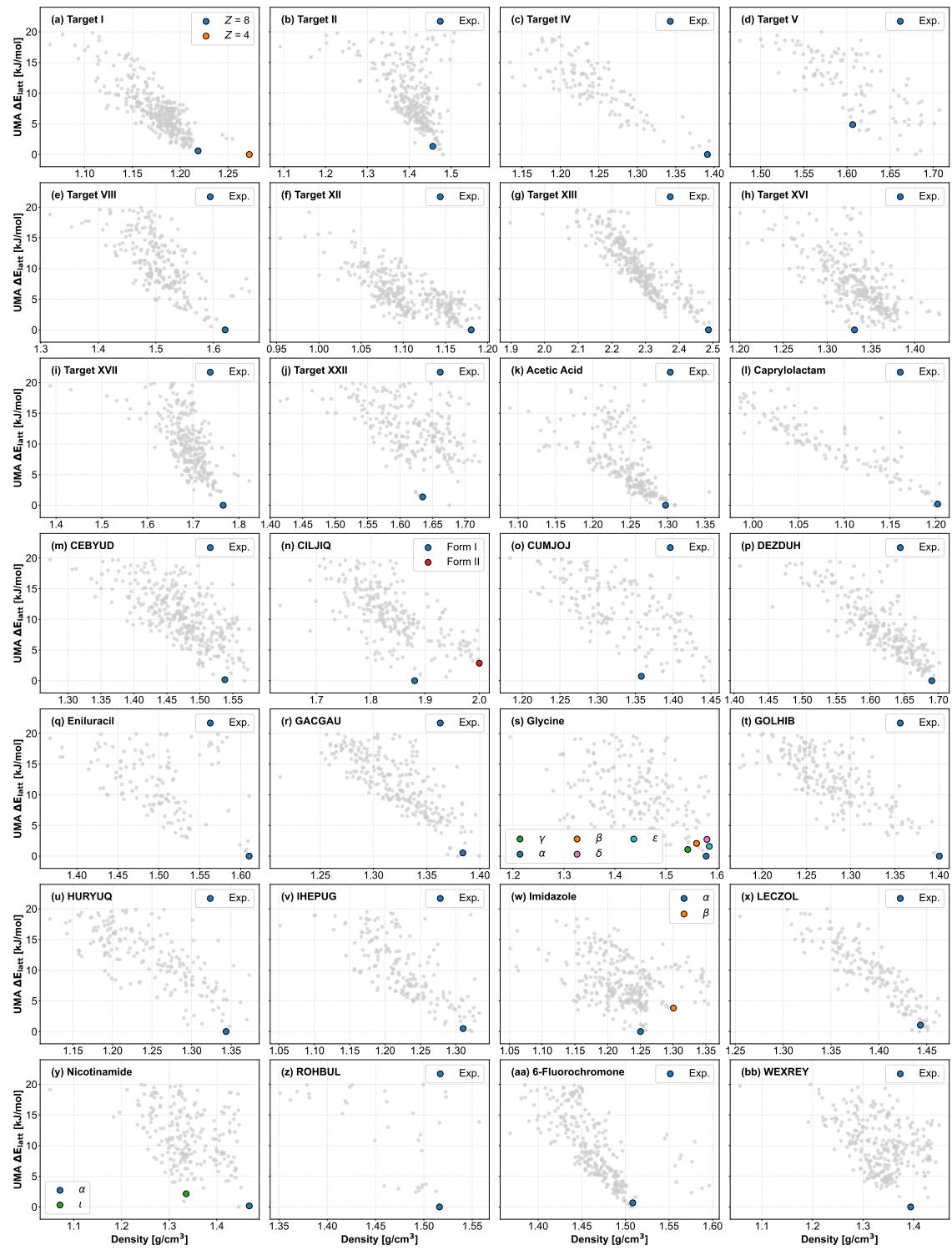


Figure S1 Energy landscapes for all 28 compounds obtained with UMA at 0 K.

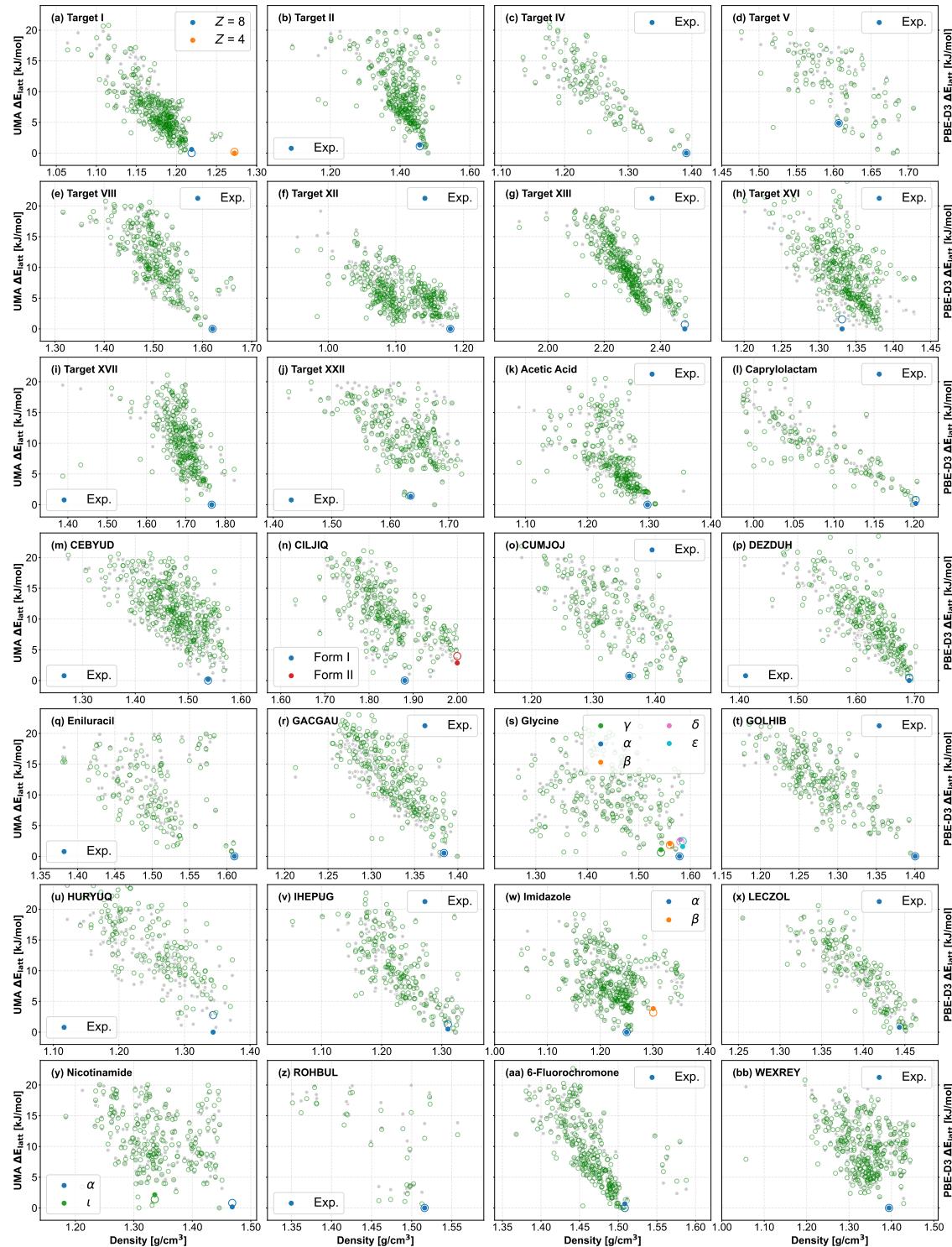


Figure S2 Energy landscapes for all 28 compounds obtained with UMA at 0 K (gray and filled circles) and PBE-D3 at 0 K (green and open circles).

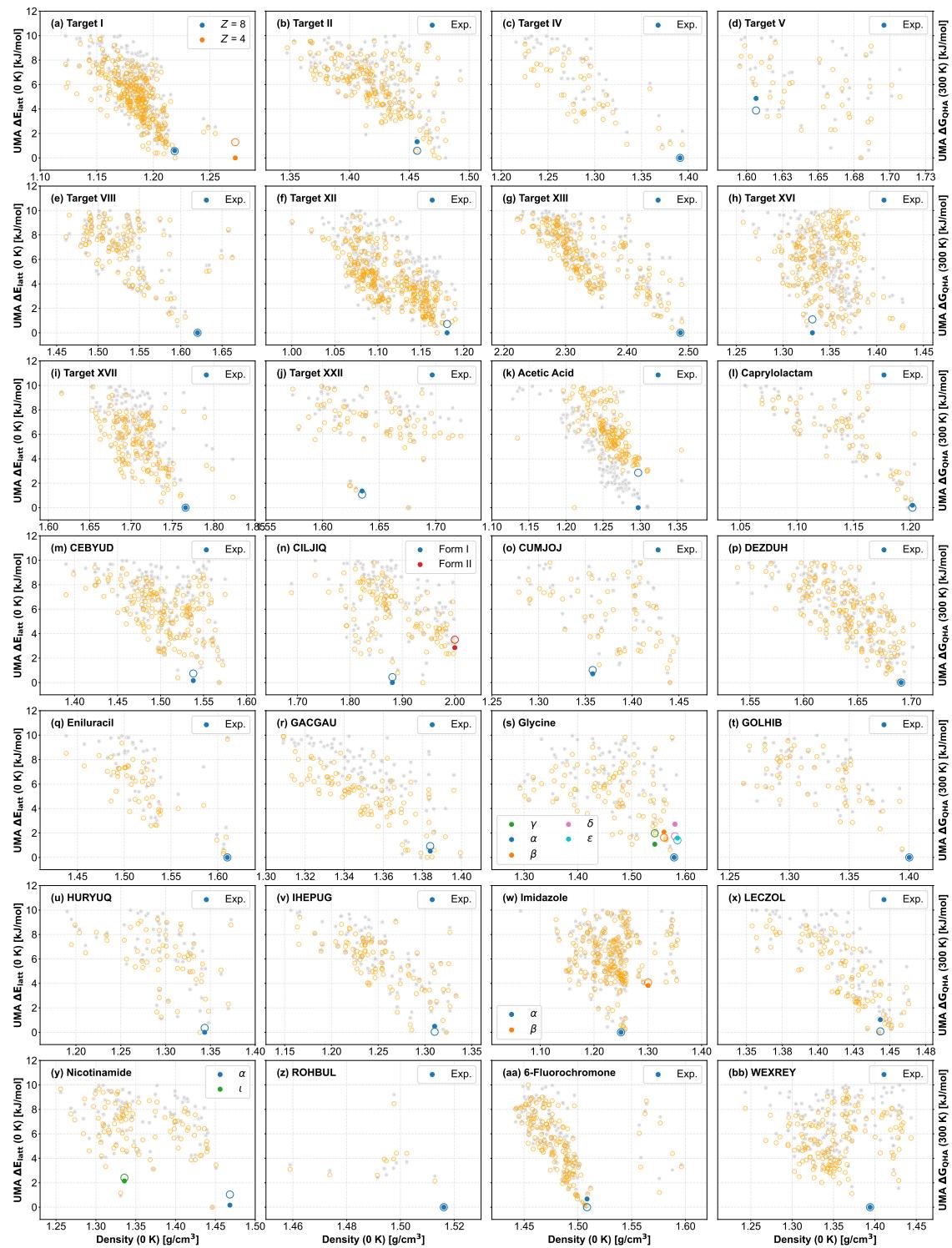


Figure S3 Energy landscapes for all 28 compounds obtained with UMA at 0 K (gray and filled circles) and after applying free energy corrections with UMA at 300 K (orange and open circles).

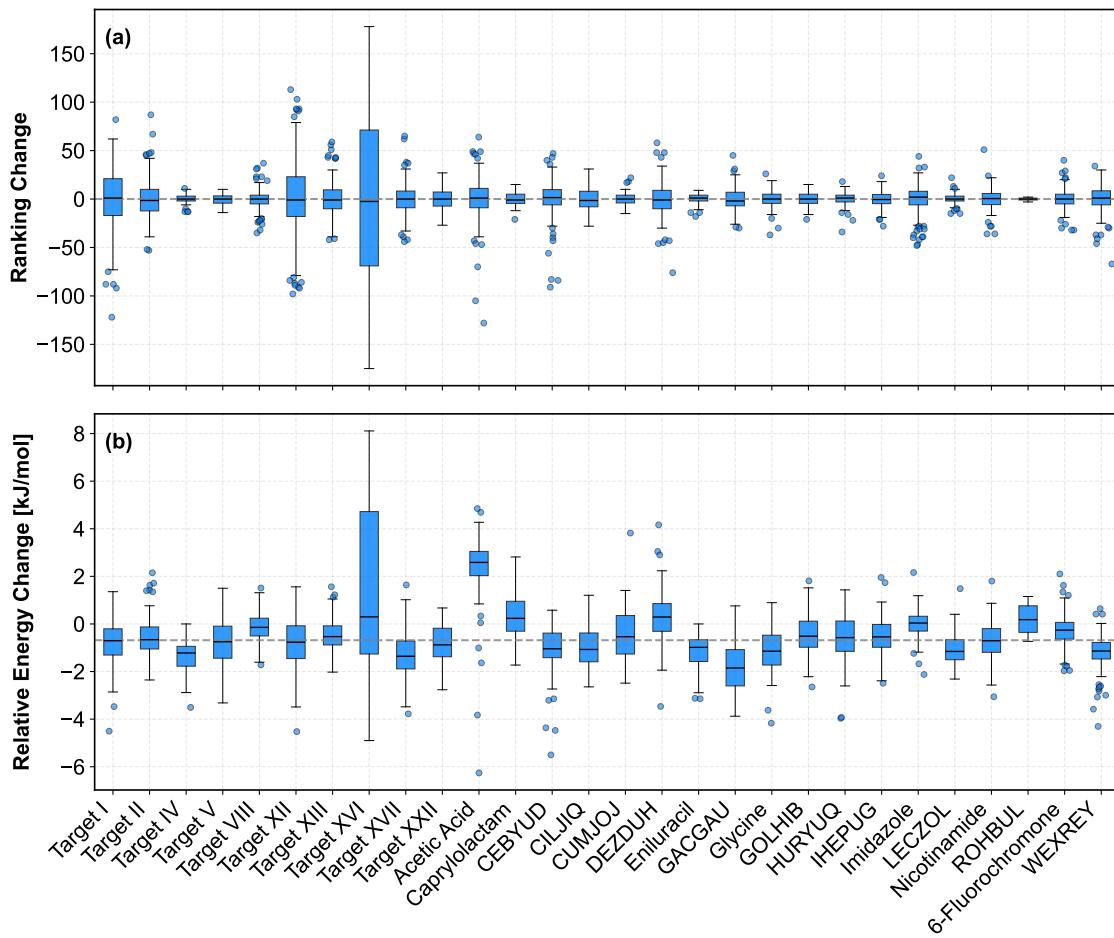


Figure S4 The distribution of (a) ranking changes and (b) relative energy changes between the lattice energy at 0 K and the free energy at 300 K, obtained using UMA. The gray dashed horizontal lines show the median values.

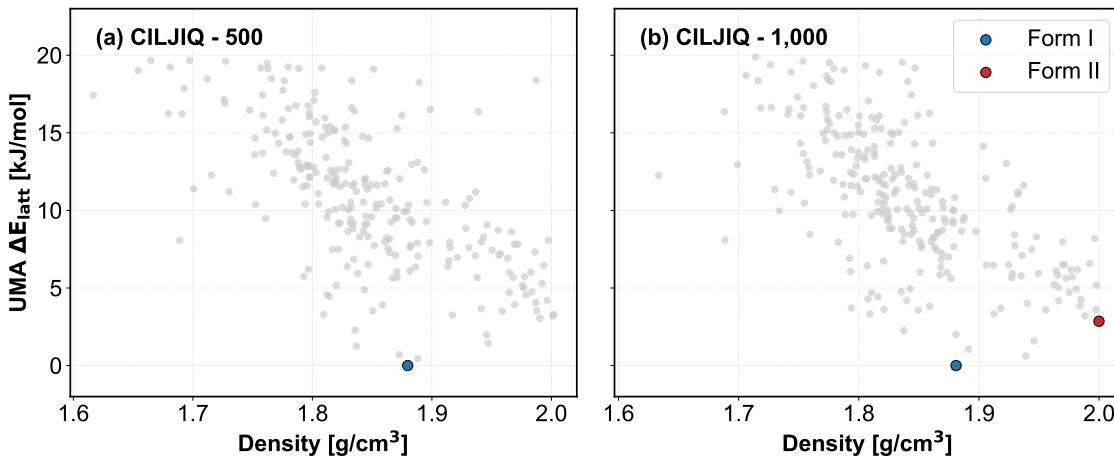


Figure S5 Energy landscapes for CILJIQ with (a) 500 and (b) 1,000 initial structures generated per space group with Genarris.

Table S4 Comparison of the relative energy and ranking obtained with UMA to results reported in repositories associated with [59] and [60], obtained using PBE-D3 and r²SCAN-D3 at the PBE-D3 optimized geometries.

No.	Compound	Form	UMA		PBE-D3 [60]		r ² SCAN-D3 [59]	
			ΔE [kJ/mol]	Rank	ΔE [kJ/mol]	Rank	ΔE [kJ/mol]	Rank
1	Target I	Z = 8	0.59	2	-	-	1.92	2
		Z = 4	0.00	1	-	-	0.00	1
4	Target V	Exp.	4.87	10	-	-	2.76	5
19	Glycine	γ	1.08	3	4.98	37	0.00	1
		α	0.00	1	2.52	9	1.55	2
		β	2.08	9	5.33	44	2.34	3
25	Nicotinamide	α	0.17	2	-	-	0.00	1
		τ	2.15	4	-	-	4.60	18

E UMA Performance

Table S5 Summary of the benchmark for the single point energy and the relaxation tasks. Relative lattice energy (ΔE_{latt} @Method) represents the lattice energy difference between each structure and the global minimum structure as determined by each respective method. Values shown in parentheses indicate performance when using the structure with the lowest DFT energy as the global reference. In cases where no parentheses are shown, the results are unaffected by the choice of method for referencing.

No.	Compound	Single Point Energy				Relaxation			
		MAE [kJ/mol]	R ²	Spearman	Kendall	MAE [kJ/mol]	Spearman	Match	RMSD ₃₀ [\AA]
1	Target I	0.57	0.95	0.98	0.89	0.81 (0.88)	0.95	82 %	0.28
2	Target II	0.37	0.97	0.99	0.94	0.51	0.97	91 %	0.19
3	Target IV	0.98	0.94	0.99	0.91	0.91	0.97	90 %	0.23
4	Target V	0.58	0.98	0.99	0.93	1.09	0.89	96 %	0.20
5	Target VIII	0.44	0.97	0.99	0.94	0.71	0.98	93 %	0.16
6	Target XII	0.72	0.91	0.95	0.81	0.97	0.88	83 %	0.19
7	Target XIII	0.58	0.95	0.98	0.92	0.62 (0.97)	0.97	88 %	0.19
8	Target XVI	2.78 (3.13)	0.53 (0.43)	0.87	0.71	2.47 (2.77)	0.86	87 %	0.22
9	Target XVII	0.91	0.88	0.96	0.89	1.09	0.93	88 %	0.19
10	Target XXII	1.15	0.89	0.98	0.89	0.90	0.97	95 %	0.14
11	Acetic Acid	0.90	0.94	0.98	0.92	1.09	0.94	89 %	0.23
12	Caprylolactam	1.22	0.88	0.94	0.86	1.51	0.87	89 %	0.26
13	CEBYUD	1.17 (1.32)	0.90 (0.88)	0.98	0.90	1.42 (1.57)	0.97	92 %	0.15
14	CILJIQ	0.76	0.96	0.99	0.92	0.87	0.97	88 %	0.20
15	CUMJOJ	0.78	0.96	0.98	0.89	0.93	0.97	98 %	0.23
16	DEZDUH	0.62	0.94	0.98	0.89	2.22 (1.95)	0.90	90 %	0.31
17	Eniluracil	0.80	0.96	0.99	0.94	1.42	0.98	89 %	0.23
18	GACGAU	1.12 (1.15)	0.92 (0.91)	0.96	0.86	1.79 (1.82)	0.93	93 %	0.29
19	Glycine	0.72	0.96	0.98	0.89	0.89	0.93	93 %	0.09
20	GOLHIB	1.29	0.88	0.95	0.85	0.91	0.94	94 %	0.29
21	HURYUQ	0.64	0.97	0.99	0.92	2.55 (1.46)	0.91	96 %	0.23
22	IHEPUG	1.47	0.88	0.96	0.86	1.44	0.94	88 %	0.29
23	Imidazole	0.43	0.97	0.99	0.93	0.63	0.95	91 %	0.21
24	LECZOL	0.59 (1.09)	0.94 (0.91)	0.97	0.90	1.05 (1.56)	0.96	80 %	0.29
25	Nicotinamide	0.72	0.87	0.94	0.88	0.79	0.91	95 %	0.16
26	ROHBUL	0.81	0.97	0.97	0.89	1.28	0.97	81 %	0.15
27	6-Fluorochromone	0.59	0.96	0.98	0.92	0.80 (0.83)	0.97	91 %	0.26
28	WEXREY	0.61	0.93	0.97	0.88	0.89	0.94	89 %	0.23
MEAN		0.87 (0.90)	0.92 (0.91)	0.97	0.89	1.16 (1.17)	0.94	90 %	0.22

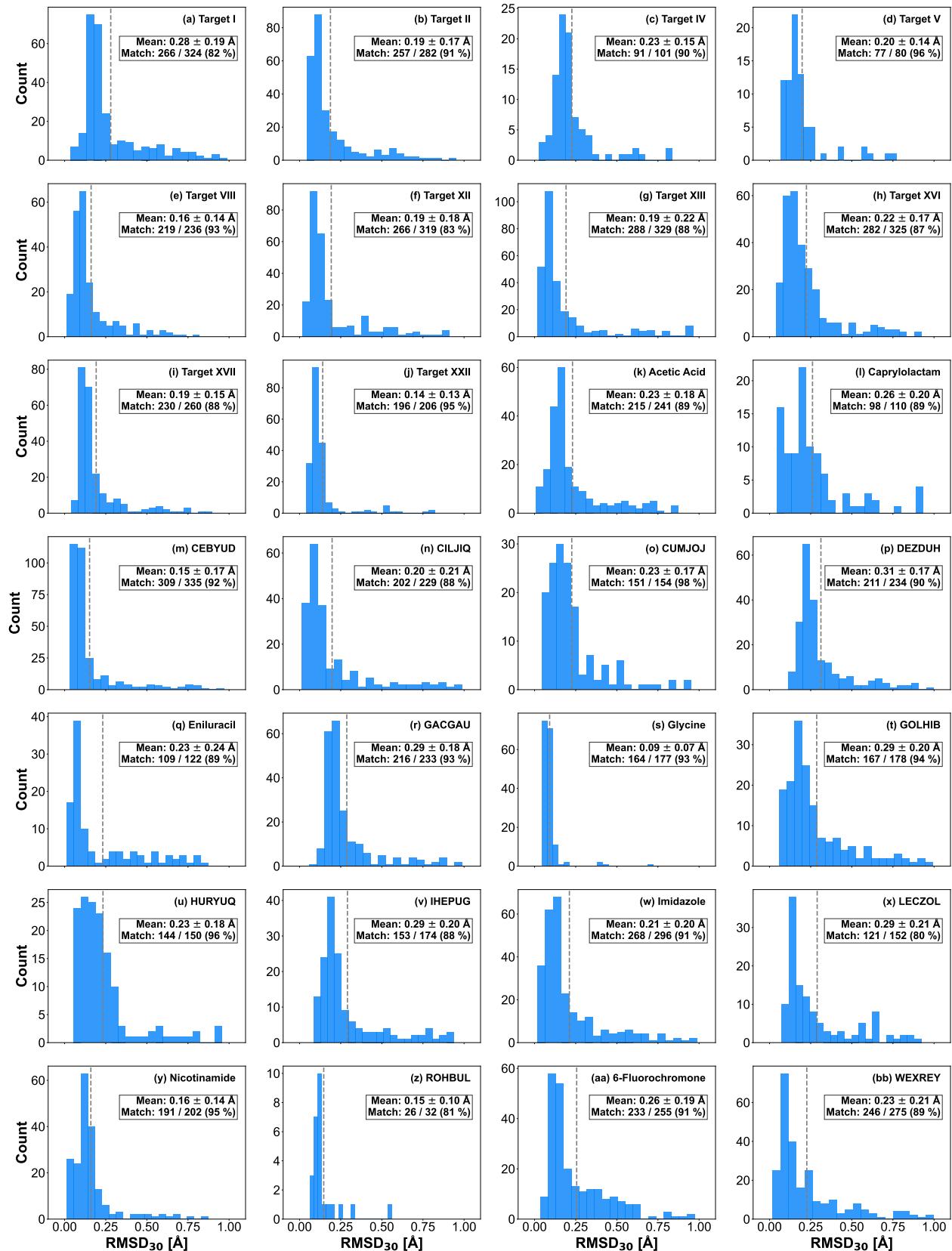


Figure S6 RMSD₃₀ histograms comparing the relaxed crystal structures obtained with the UMA model to those obtained with PBE-D3 for all 28 compounds.



Figure S7 Parity plots of the single point energy task for all 28 compounds. Relative lattice energy (ΔE_{latt} @Method) represents the lattice energy difference between each structure and the global minimum structure of each method.

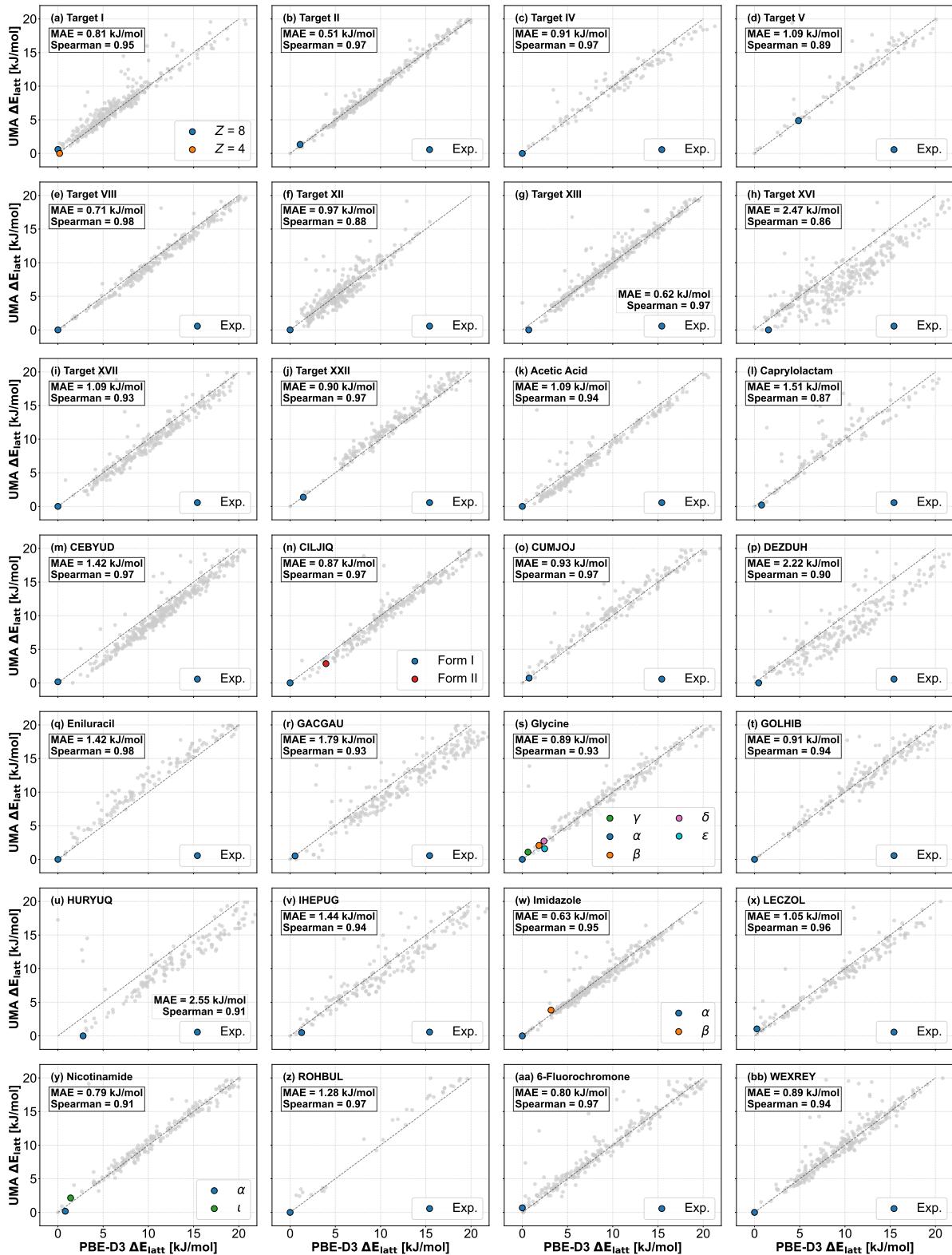


Figure S8 Relaxation performance for all 28 compounds. Relative lattice energy (ΔE_{latt} @Method) represents the lattice energy difference between each structure and the global minimum of each method.

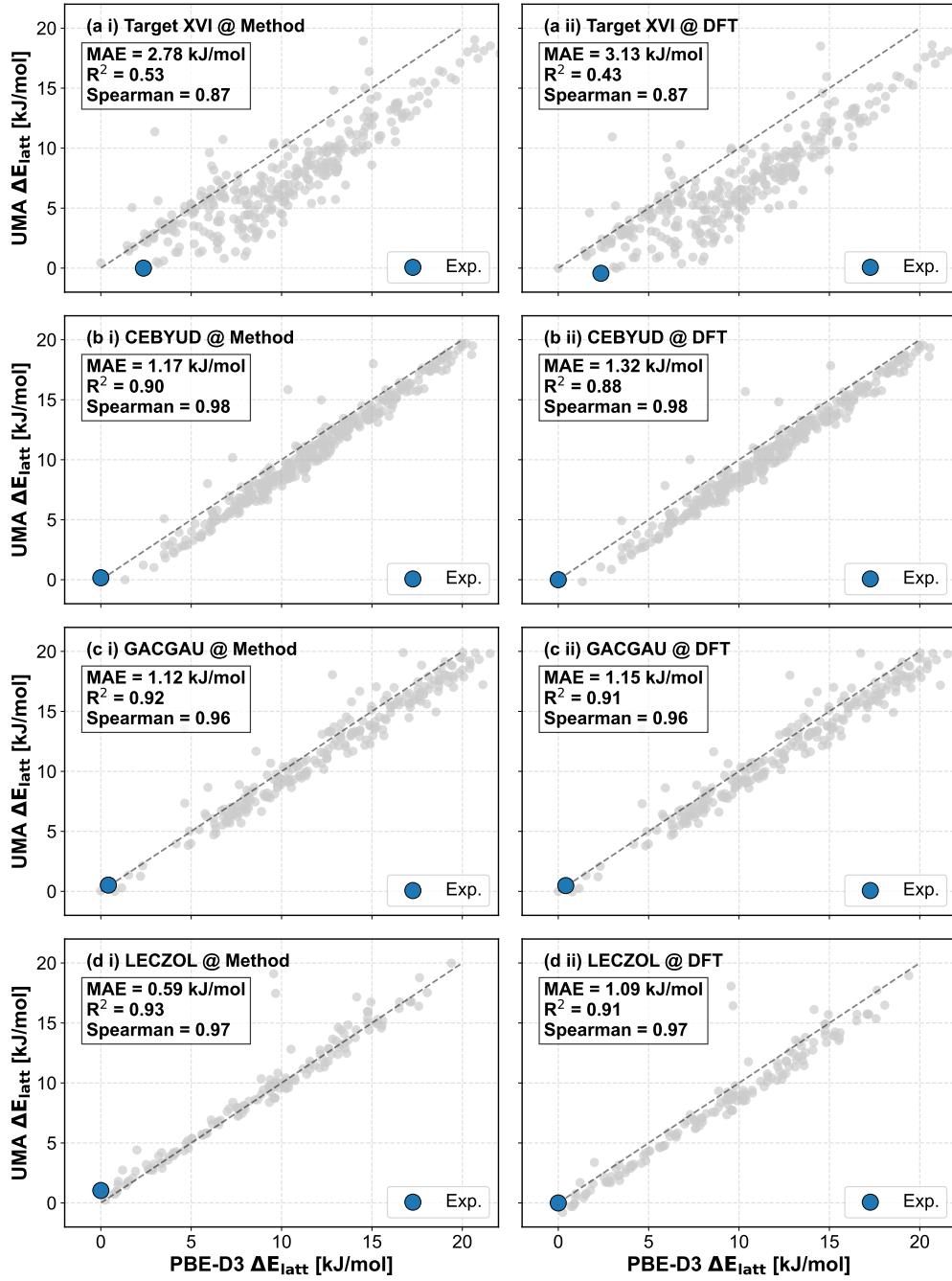


Figure S9 Comparison of different references for four crystal structures in the single point energy task: Target XVI, CEBYUD, GACGAU, and LECZOL. Results are shown for each system using both the global minimum of each method (left column) and the global reference structure with the lowest DFT energy (right column).

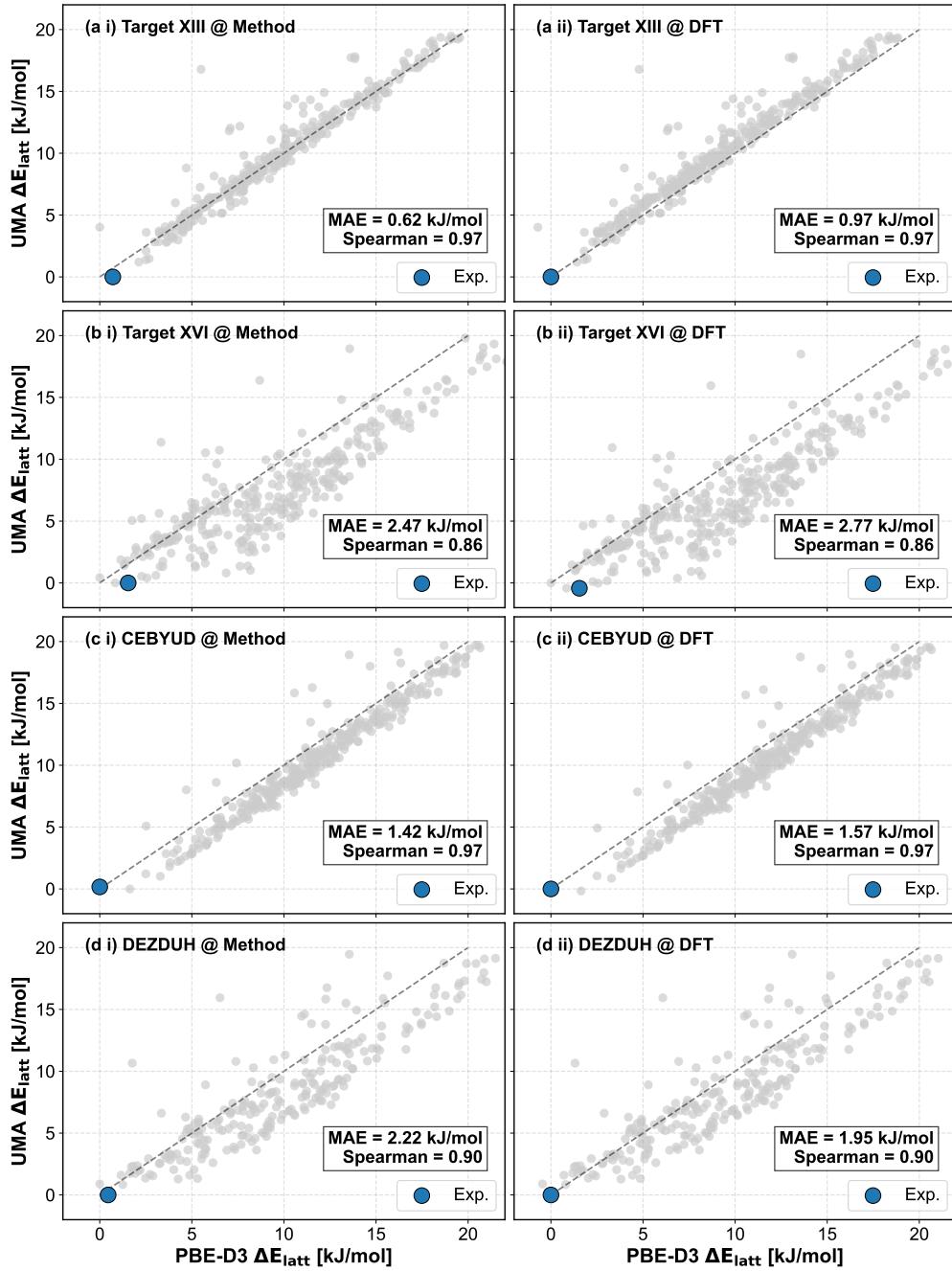


Figure S10 Comparison of different references for four crystal structures in the relaxation task: Target XIII, Target XVI, CEBYUD, and DEZDUH. Results are shown for each system using both the global minimum of each method (left column) and the global reference structure with the lowest DFT energy (right column).

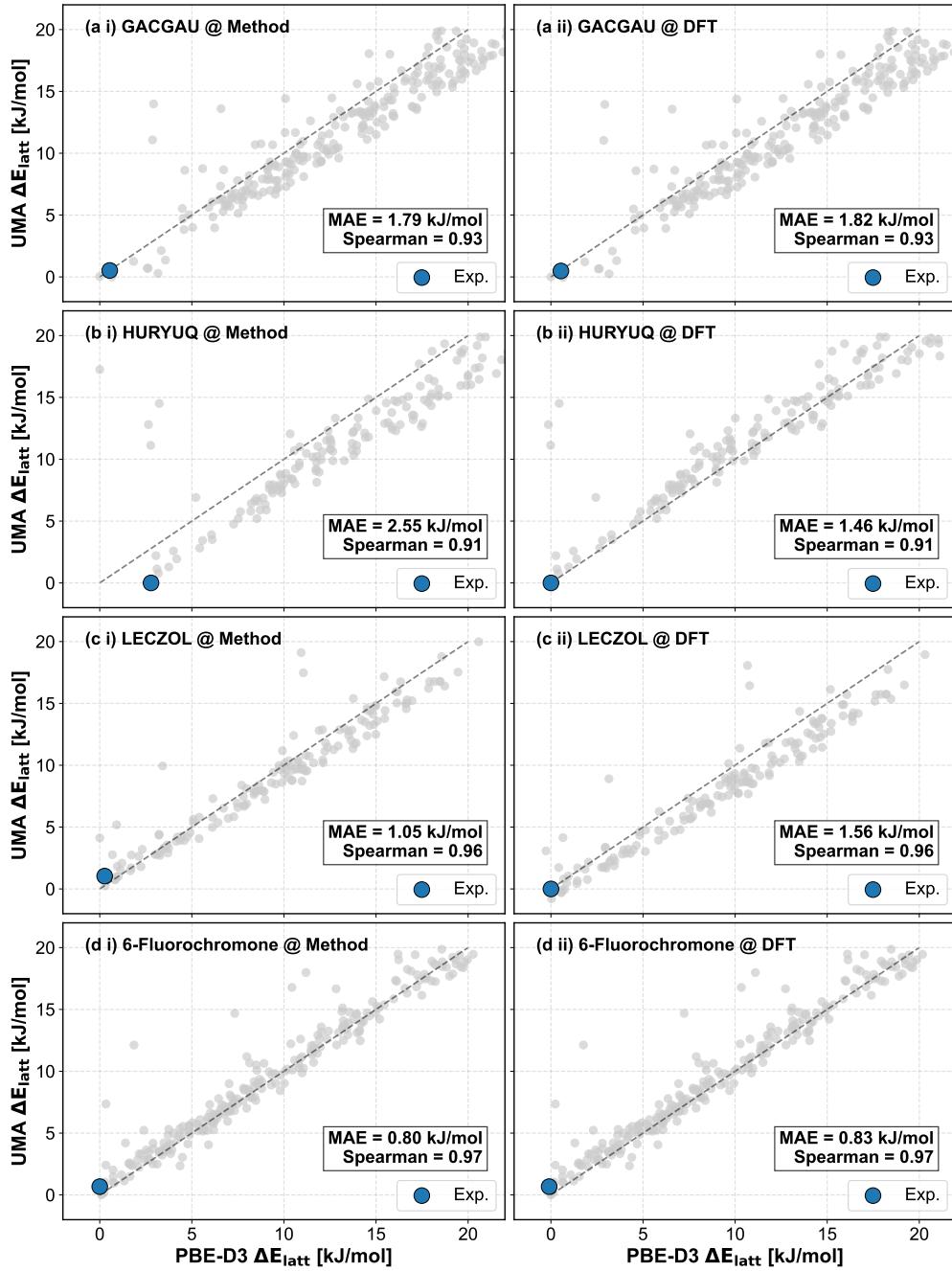


Figure S10 (continued) Comparison of different references for four crystal structures in the relaxation task: GACGAU, HURYUQ, LECZOL, and 6-Fluorochromone. Results are shown for each system using both the global minimum of each method (left column) and the global reference structure with the lowest DFT energy (right column).

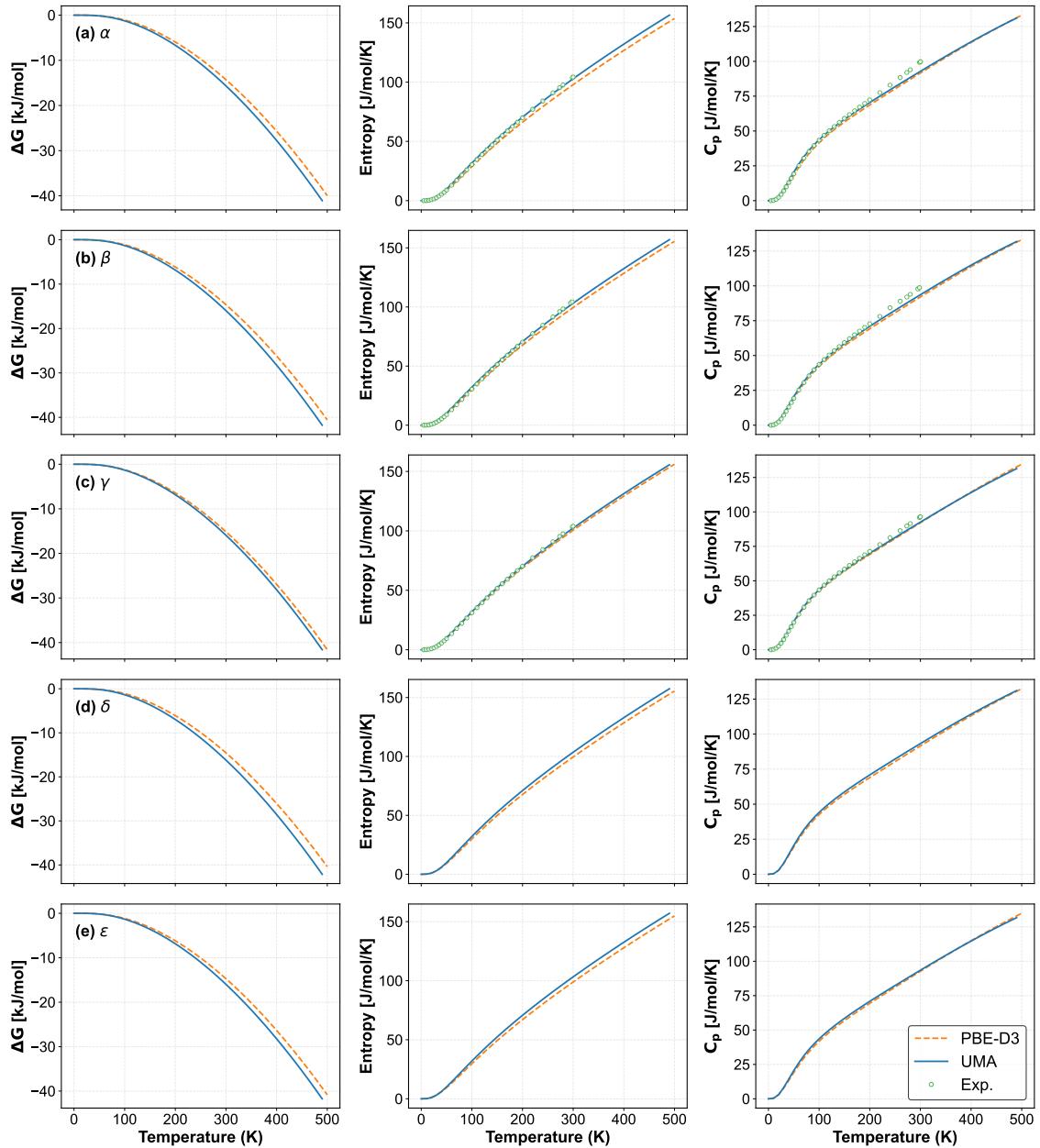
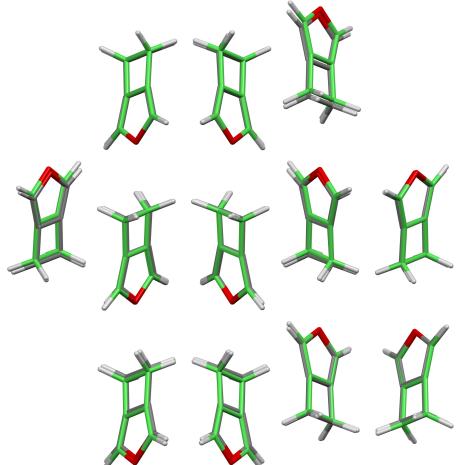
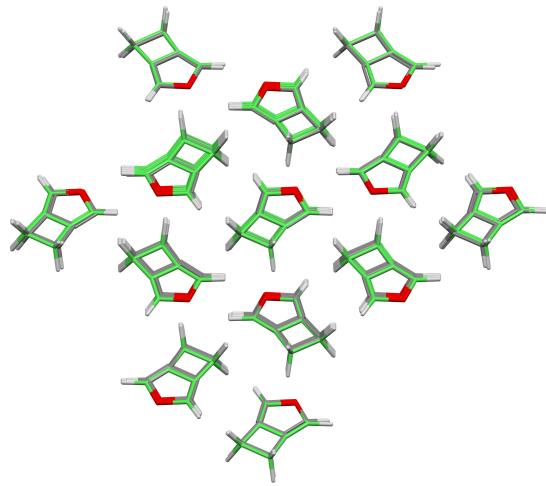


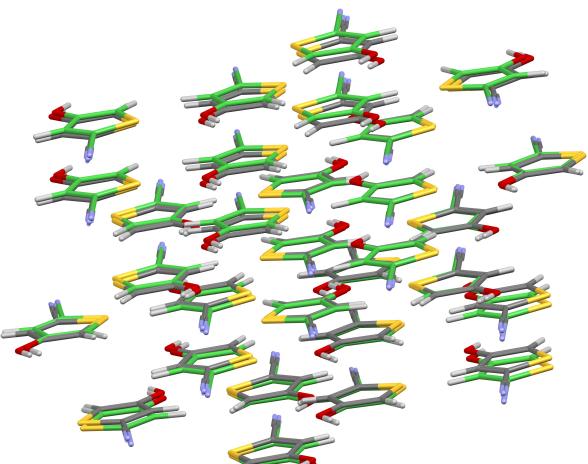
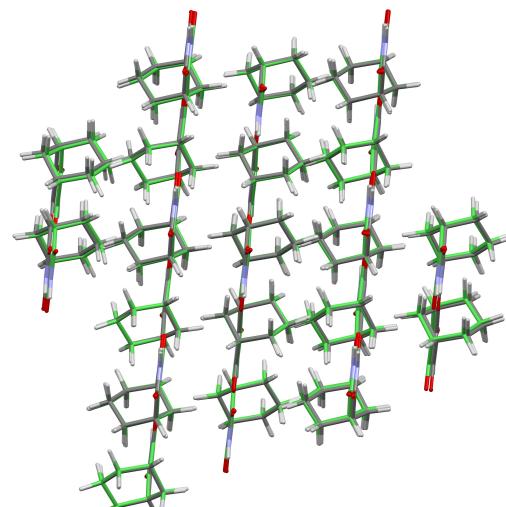
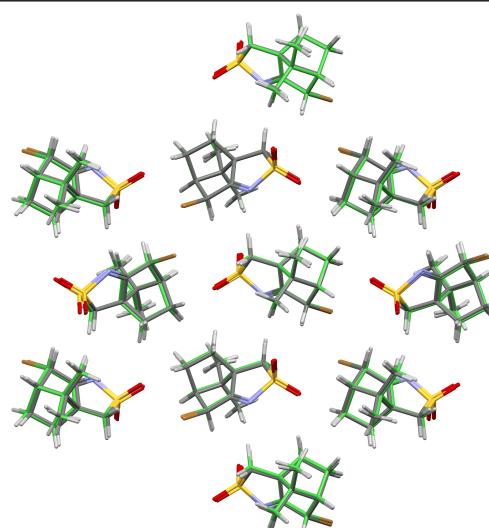
Figure S11 Gibbs free energy, entropy, and heat capacity of (a) α , (b) β , (c) γ , (d) δ , and (e) ϵ -glycine polymorphs obtained from UMA, PBE-D3, and experimental data [100, 116].

F Packing similarity between experimental and matched crystals

Table S6 Overlay of 30-molecule clusters of all considered polymorphs from CSD [70] (atoms colored by element) with the matching predictions found by UMA (green), and the respective RMSD₃₀ values of crystal overlays.

No.	Compound	CSD Refcode	RMSD ₃₀ [\AA]	Packing comparison
1	Target I	XULDUD	0.20	
2	Target I	XULDUD01	0.13	

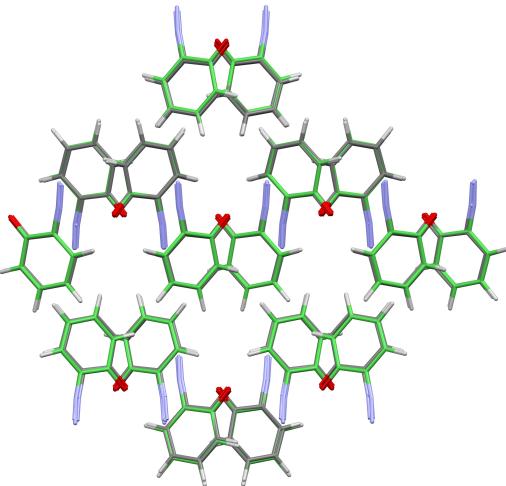
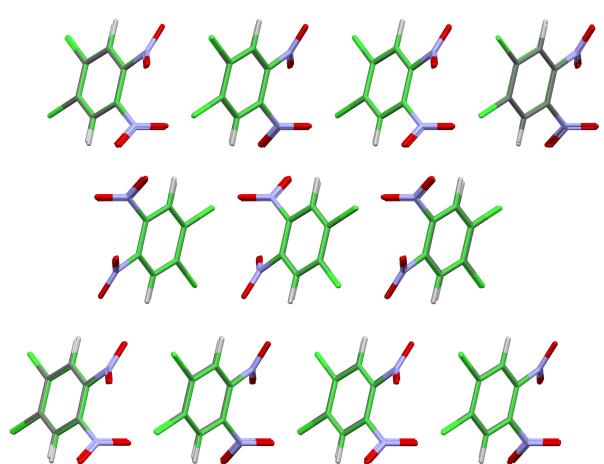
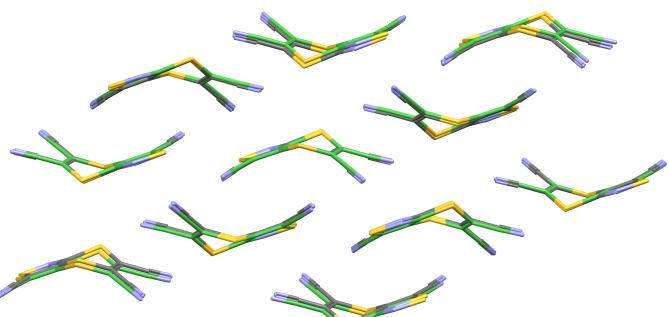
Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
3	Target II	GUFJOG	0.35	
4	Target IV	BOQQUT	0.14	
5	Target V	BOQWIN	0.17	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
6	Target VIII	PAHYON01	0.23	
7	Target XII	AXOSOW01	0.24	
8	Target XIII	SOXLEX01	0.09	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
9	Target XVI	OBEQUJ	0.15	
10	Target XVII	OBEQOD	0.21	
11	Target XXII	NACJAF	0.19	

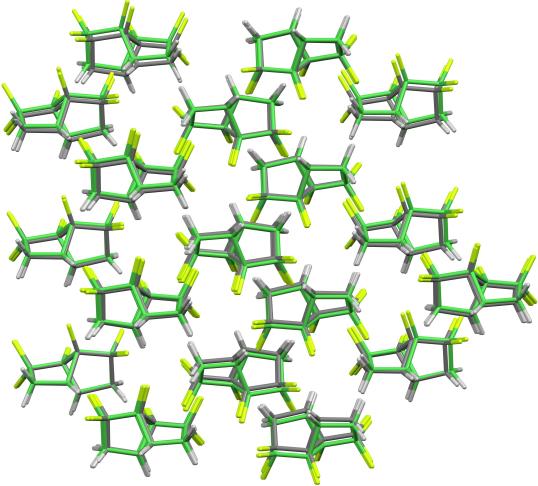
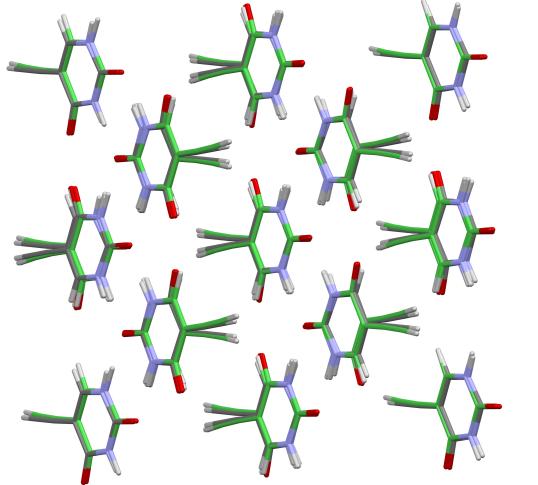
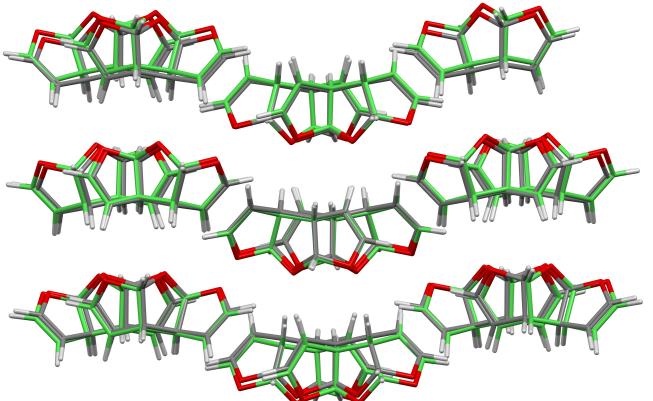
Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
12	Acetic Acid	ACETAC03	0.08	
13	Caprylolactam	CAPRYL	0.15	
14	CEBYUD	CEBYUD	0.10	

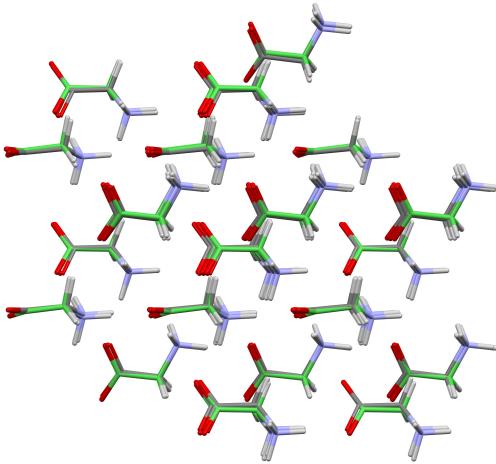
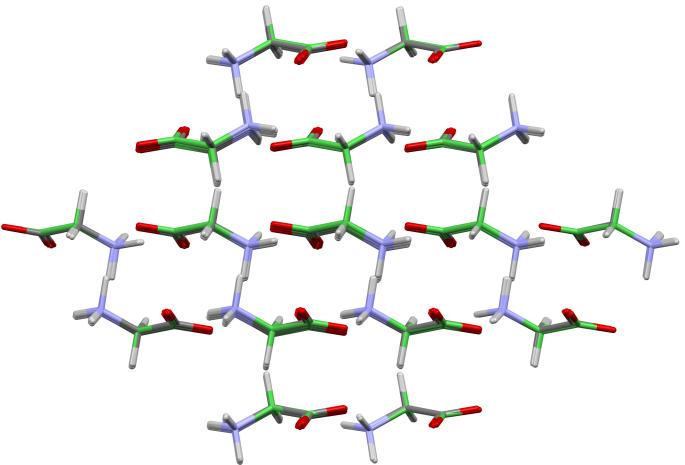
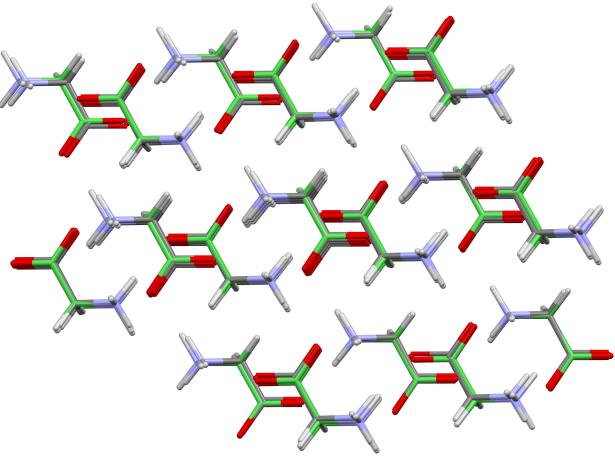
Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
15	CILJIQ Form I	CILJIQ	0.26	
16	CILJIQ Form II	CILJIQ01	0.12	
17	CUMJOJ	CUMJOJ	0.16	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
18	DEZDUH	DEZDUH	0.20	
19	Eniluracil	DOHFEM	0.13	
20	GACGAU	GACGAU	0.21	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
21	Glycine γ	GLYCIN16	0.13	
22	Glycine α	GLYCIN20	0.09	
23	Glycine β	GLYCIN32	0.11	

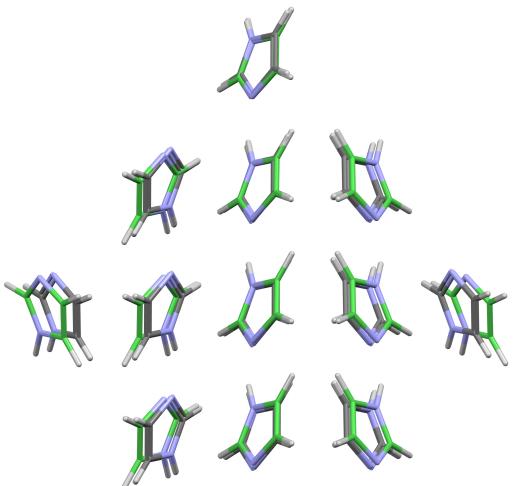
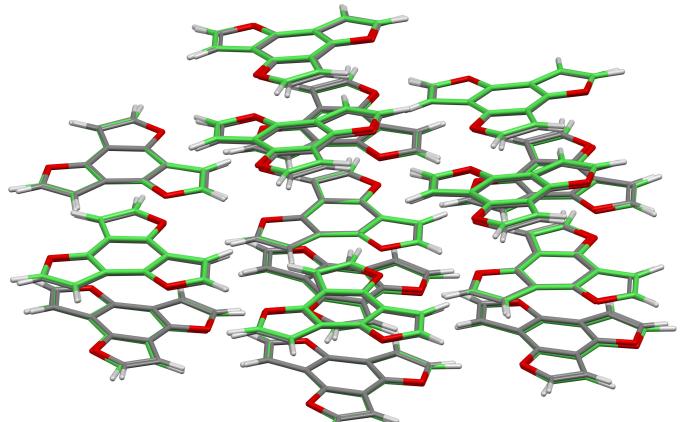
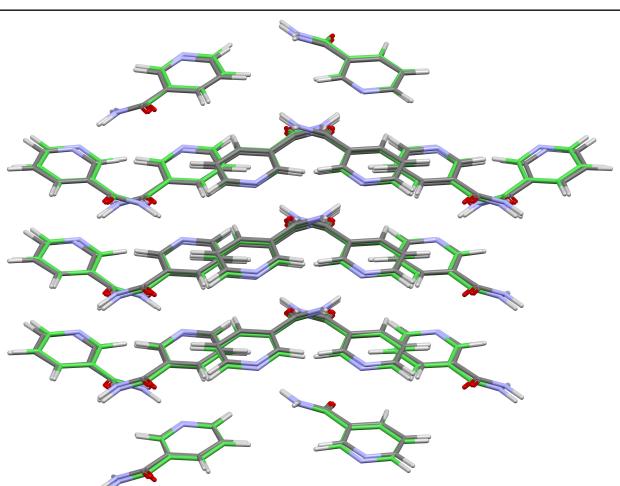
Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
24	Glycine δ	GLYCIN67	0.24	
25	Glycine ε	GLYCIN68	0.36	
26	GOLHIB	GOLHIB	0.06	

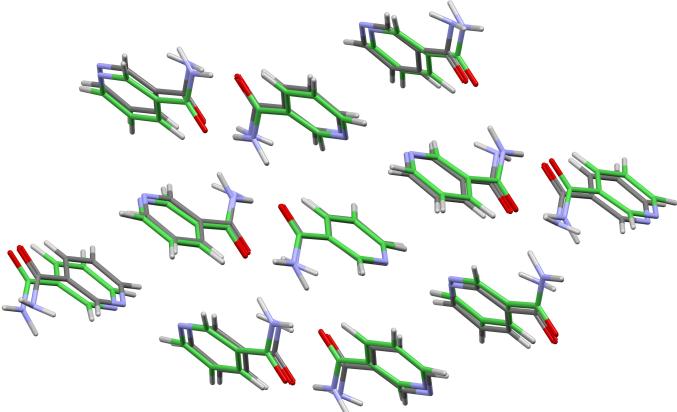
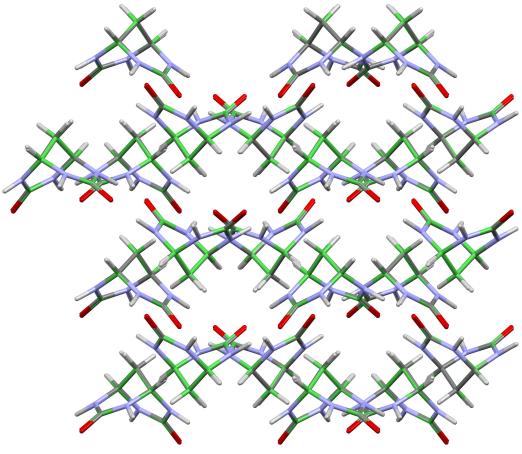
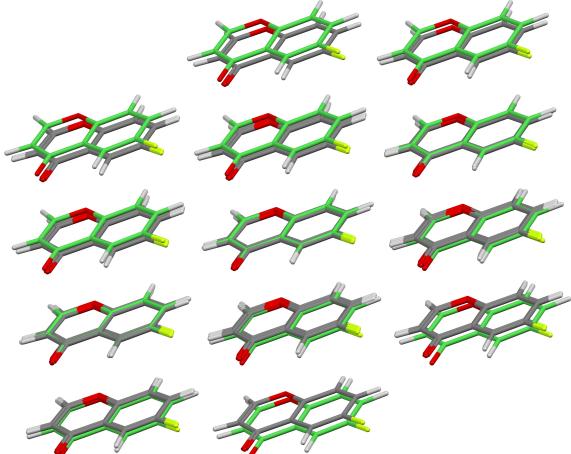
Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
27	HURYUQ	HURYUQ	0.12	
28	IHEPUG	IHEPUG	0.13	
29	Imidazole α	IMAZOL06	0.20	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
30	Imidazole β	IMAZOL25	0.31	
31	LECZOL	LECZOL	0.23	
32	Nicotinamide α	NICOAM03	0.23	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
33	Nicotinamide ι	NICOAM17	0.35	
34	ROHBUL	ROHBUL	0.04	
35	6-Fluorochromone	UMIMIO	0.22	

Continued on next page

No.	Compound	CSD Refcode	RMSD ₃₀ [Å]	Packing comparison
36	WEXREY	WEXREY	0.29	