



Cite this: DOI: 10.1039/d6dd00016a

# Integrating machine learning interatomic potentials with batched optimization for crystal structure prediction

Chengxi Zhao,<sup>†\*a</sup> Zhaojia Ma,<sup>†bc</sup> Dingrui Fan,<sup>†a</sup> Siyu Hu,<sup>bc</sup> Leping Wang,<sup>bc</sup> Feng Hua,<sup>bc</sup> Weile Jia,<sup>bc</sup> En Shao,<sup>\*bc</sup> Guangming Tan,<sup>\*bc</sup> Jun Jiang<sup>†ad</sup> and Linjiang Chen<sup>†ae</sup>

Molecular crystal structure prediction (CSP) faces a persistent computational bottleneck: it requires exhaustive sampling of vast packing landscapes while resolving energy differences of only a few kJ mol<sup>-1</sup>. We introduce BOMLIP-CSP, an open-source Python framework that integrates machine learning interatomic potentials (MLIPs) with a tailored batched optimization strategy, enabling rapid, unbiased structure prediction across the full crystal density range. By introducing tailored parallelism into modern MLIPs, BOMLIP-CSP achieves a ~2.1–2.3× acceleration in large-scale CSP searches without compromising accuracy. In benchmarks covering 34 experimental structures from six CSP blind tests, more than 50% of the experimental crystal structures can be recovered using foundation MLIPs when the correct space group and Z' are included in the search, with more than 70% of the experimental structures recovered by at least one of the tested MLIPs under the present benchmark conditions. Importantly, our results suggest that MLIPs with comparable equilibrium-energy accuracy can yield strikingly different CSP outcomes, indicating that predictive success may depend not only on local energy fidelity but also on how the MLIP energy surface is organised. Together, these results establish BOMLIP-CSP as a broadly accessible platform for accelerated CSP and provide new insight into the interplay between MLIP characteristics and crystal structure discovery.

Received 15th January 2026

Accepted 2nd April 2026

DOI: 10.1039/d6dd00016a

rsc.li/digitaldiscovery

## Introduction

Computational chemistry has become an essential tool for exploring high-performance materials and elucidating structure–property relationships.<sup>1–4</sup> However, the limited availability of experimentally known materials constrains large-scale screening efforts. With advances in high-accuracy computational methods, the field has shifted from primarily interpreting experiments to enabling predictive discoveries.<sup>5–8</sup> For molecular solids, crystal structure prediction (CSP), which seeks to determine the potential crystal structures of a molecular system using only its chemical composition and molecular

formula, is a critical and challenging step that must precede property evaluation in the search for high-performance materials.

CSP has attracted significant attention in materials science due to its importance and the challenges posed by its complex workflow and high accuracy requirements.<sup>9–12</sup> It involves an extensive search for local minima on a vast crystal lattice energy landscape, followed by ranking candidate structures whose energy differences are often within only 1 to a few kJ mol<sup>-1</sup>.<sup>13–15</sup> This extreme sensitivity imposes stringent accuracy requirements, while the vast configurational space demands high computational efficiency, thereby creating a persistent bottleneck that has long constrained the field. Furthermore, despite its potential and significance, organic molecular CSP remains dominated by proprietary software, with only a handful of open-source implementations available,<sup>9,11,16</sup> limiting accessibility and reproducibility for many research groups worldwide.

The blind tests organized by the Cambridge Crystallographic Data Centre (CCDC) over the past two decades have witnessed remarkable progress in CSP capabilities.<sup>17–24</sup> Recently, machine learning (ML) has emerged as a versatile tool for addressing high-dimensional chemistry problems.<sup>25,26</sup> Beyond aiding decision-making across the CSP workflow,<sup>23</sup> machine learning interatomic potentials (MLIPs), which capable of achieving near

<sup>a</sup>State Key Laboratory of Precision and Intelligent Chemistry, Hefei National Research Center for Physical Sciences at the Microscale, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, China. E-mail: Chengxi.zhao@ustc.edu.cn; jiangj1@ustc.edu.cn; linjiangchen@ustc.edu.cn

<sup>b</sup>State Key Laboratory of Processors, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. E-mail: shaoen@ict.ac.cn; tgm@ict.ac.cn

<sup>c</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>d</sup>Hefei National Laboratory, University of Science and Technology of China, Hefei, China

<sup>e</sup>School of Chemistry, School of Computer Science, University of Birmingham, Birmingham, UK

<sup>†</sup> These authors contributed equally: C. Z., Z. M., & D. F.



*ab initio* accuracy at computational costs comparable to classical force fields,<sup>27–29</sup> offer a promising route for CSP tasks that demand both accuracy and efficiency. More recently, the emergence of foundation models<sup>30–32</sup> has further transformed this landscape by reducing or even eliminating the need for time-consuming data generation, model training, and active learning cycles.

Several studies have begun exploring the use of MLIPs to perform prediction problems.<sup>33–35</sup> Such research typically focuses on evaluating the accuracy of the overall CSP strategy, with the structures often first pre-optimized using other methods.<sup>16,36</sup> While these tests are undoubtedly important, the community still lacks a clear assessment of how different MLIPs directly influence CSP outcomes when all other steps in the workflow are held constant. In particular, it remains unclear how these MLIPs perform when the initial trial structures—owing to the lack of prior knowledge of the experimental structure—are largely displaced from their local minima, a scenario that is more representative of practical structure prediction tasks. This also raises another critical issue: the need for a software framework that, tailored to the characteristics of MLIPs, can perform CSP calculations in a simple and fast manner.

Here, we have developed an open-source code, Batched Optimization with Machine Learning Interatomic Potentials Crystal Structure Prediction (BOMLIP-CSP). This code addresses the issue that CSP codes are largely dominated by commercial licenses, leaving researchers in many materials-focused groups with very few accessible options for performing CSP calculations. This code is designed to perform the crystal structure prediction workflow across the full range of crystal densities using MLIPs with ease. We developed a batched structural optimization module tailored to MLIP-based CSP calculations. In the present BOMLIP-CSP implementation, this module markedly accelerates computations and enhances efficiency in large-scale tasks, delivering an observed overall  $\sim 2.1\text{--}2.3\times$  speed-up. This strategy balances computational speed and accuracy, delivering excellent precision under fast calculations, with performance that can be flexibly tailored according to the chosen MLIPs. Two widely adopted foundation atomistic models have been preconfigured within the accelerated batch optimization module for immediate use, and the framework is readily extensible to other methods.

Furthermore, we conducted a benchmark study by applying different MLIPs to optimize crystal structures generated from the same trial structure database. The test set was based on 26 target systems from the first to the sixth CSP blind tests, along with their polymorphs sharing the same asymmetric unit found in the CCDC database yielding a total of 34 crystal structures. When the correct space group and  $Z'$  are present in the search space, even relatively small MLIPs recover more than 50% of these experimental structures at least once, and in many cases the corresponding hits are ranked near the bottom of the energy list. Across the tested MLIPs, more than 70% of the experimental structures were recovered under the present benchmark conditions. Instead, the aim of this benchmark study is not to highlight the ultimate capability of BOMLIP-CSP or to serve as

a definitive comparison of atomistic models, but to provide a systematic evaluation of the impact of different MLIPs on CSP under identical workflow conditions. Within this controlled setting, the results suggest that our framework can achieve a favourable balance between computational efficiency and predictive accuracy, and that judicious MLIP selection may help reduce the number of trial structures required in practice.

## Results

### The BOMLIP-CSP framework

We present BOMLIP-CSP, an open-source software package designed to perform crystal structure prediction and explore the energy landscape of a given molecule. To achieve this, the software aims to generate a large number of possible molecular packing modes, forming trial crystal structures that sample most local minima on the crystal lattice energy landscape. This process, as with most CSP approaches,<sup>23</sup> comprises three main steps: (i) conformer search for the given molecule, (ii) generation of trial crystal structures, and (iii) structural optimization followed by energy ranking. The workflow of BOMLIP-CSP is shown in Fig. 1a.

### Generation of trial structures within the BOMLIP-CSP framework

The Sobol sequence,<sup>37</sup> a quasi-random, low-discrepancy, and uniformly distributed sampling method, was employed to generate vectors for constructing trial crystal structures. This sequence has been widely applied in CSP, such as in quasi-Monte Carlo sampling schemes.<sup>38</sup> Notably, Graeme M. Day and co-workers<sup>9</sup> reported employing Sobol sequences for trial structure generation, which inspired the strategy adopted in the present study.

In our code, an initial coarse crystal structure with a large unit cell length was generated, with its unit cell parameters, the centroid of the asymmetric unit, and its orientation determined by a vector sampled from the Sobol sequence. Then, the space group symmetry operations were applied to the asymmetric unit to construct the complete crystal structure. After that, each unit cell length was gradually reduced until the intermolecular distances within the cell dropped below a predefined threshold. To ensure structural reasonableness, all generated structures must satisfy three criteria, which are enforced or evaluated at different stages of the generation process. Collectively, these criteria constitute what we refer to as the triple-radius validation scheme (Fig. 1b).

First, a covalent radius<sup>39</sup> factor of  $1.15\times$  was applied to ensure that each molecule is complete, with all covalent bonds correct. Second, van der Waals (vdW)<sup>40</sup> radii scaled by  $0.9\times$  were used to identify and exclude structures with excessively close intermolecular contacts. Third, to eliminate physical unreal structures, a spherical radius equal to  $1.5\times$  the vdW radius of the largest atom was applied around each atom; a valid trial structure must form a three-dimensional framework within this criterion. This ensures that unphysical slab-like arrangements (Fig. S1), which either waste optimization time or lead to



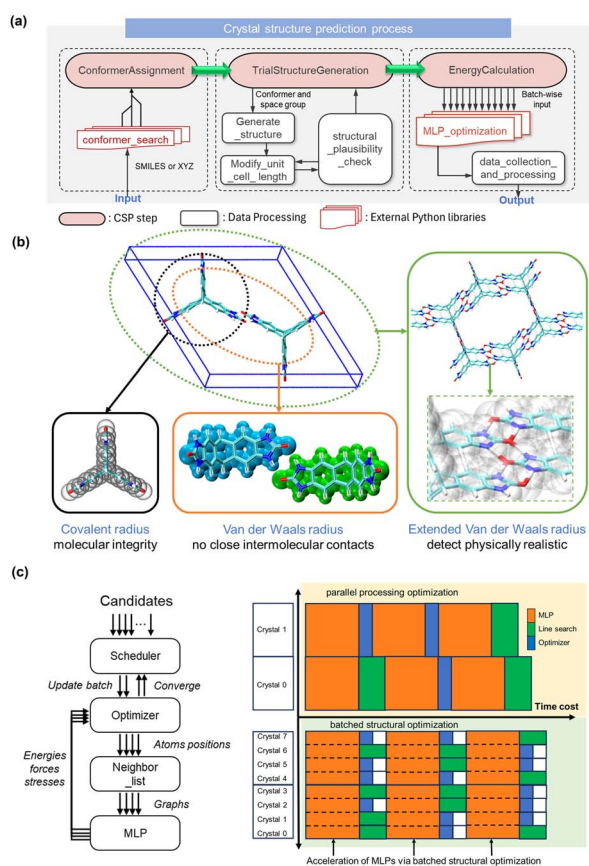


Fig. 1 Illustration of the BOMLIP-CSP workflow and key implementation details. (a) Workflow diagram of the crystal structure prediction process. Each stage is modularized to allow easy integration of additional conformers or structures into the data pool. (b) Three criteria that a valid trial structure must satisfy. Using T2- $\gamma$  as an example, the diagram illustrates: molecular integrity check (black dashed circles), intermolecular contact detection (orange), and assessment of overall physical realism (green). (c) Left: the workflow of batched structural optimization method for MLIPs in BOMLIP-CSP. Right: schematic comparison of time costs between conventional parallel optimization and our implemented batched structural optimization method.

meaningless results, are excluded, while preserving potentially porous yet realistic structures that may correspond to experimentally observed crystals or be essential to a complete energy landscape. This process ensures that each generated trial crystal is of high structural quality, enabling efficient and comprehensive exploration of the energy landscape without wasting computational resources. Moreover, it reduces the risk of missing packing motifs due to sampling bias like porous structure. For example the well-known molecule T2,<sup>7</sup> exhibit porous structures that are challenging to predict using density-based close-packing generation methods.

**A CSP framework integrating MLIP backends and enhanced parallel computation.** In the organic molecule CSP process, efficiently optimizing a large number of generated structures within a reasonable time while maintaining an acceptable level of energy accuracy remains a key challenge. To address this

issue, we integrated MLIPs into our software and tailored their use to the specific demands of CSP, particularly the optimization of large sets of structurally similar candidates with identical atom counts and compositions, thereby enabling targeted acceleration strategies (Fig. 1c).

Traditional parallelization approaches, such as multiprocessing, handling individual structures concurrently, suffer from inherent inefficiencies. In this paper, we present a high-performance CSP relaxation framework specifically designed to integrate multiple mainstream MLIP backends efficiently. The core innovation of our framework is performing fine-grained batch processing strategies by grouping multiple distinct candidate structures into a computing batch. Thus, the neighbour list updating, the quasi-Newton method optimizing, and the energy/force evaluation can be performed simultaneously on all candidates within a batch using tensor operations. This leverages the massive parallel processing capabilities of GPUs and other accelerators most effectively. Besides, our framework employs an intelligent continuous batching scheduling strategy to maintain a global compute pool dynamically. If structures within a batch finish the relaxation early, which means these structures successfully converge to their minimum-energy, the scheduler immediately replaces and pushes unprocessed candidates (until the target atom count threshold is reached) to saturate the utilization of computational resources and maximize the relaxation throughput (see supplementary methods algorithms in SI for details). Our framework demonstrates performance gains in practice, enabling the high-throughput relaxation process for large-scale CSP.

To address this limitation, the developed batched structural optimization method for MLIPs consists of three components: (1) a batch algorithm based on the quasi-Newton method; (2) a batched neighbour list update algorithm compatible with periodic boundary conditions; and (3) a structure optimization scheduling algorithm. The details of these methods are provided in the SI method.

### Evaluating the accuracy and efficiency of BOMLIP-CSP in polymorph-rich systems

To evaluate the effectiveness of the overall strategy, we conducted a test on the T2 molecule, a well-known system that have five polymorphs and whose energy landscape has been extensively studied in prior research.<sup>5,7,41,42</sup> 500 trial structures were generated in each of the following common space group: *Pbca* (61), *P2<sub>1</sub>/c* (14), *P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>* (19), *P1* (2), *P21* (4), *C2/c* (15), *Pna2<sub>1</sub>* (33), *Cc* (9), *Pca2<sub>1</sub>* (29), *C2* (5), *P42* (77). The energy landscape, constructed from all optimized crystal structures, is shown in Fig. S2 and is consistent with previously reported energy landscapes. All experimentally observed polymorphs were successfully predicted in this test and are located at energy spikes, a characteristic commonly associated with experimentally accessible porous structures.<sup>5,7</sup> No physically unrealistic structures were identified in the resulting landscape. This indicates that the strategy effectively filters out unphysical structures without excluding energetically reasonable porous structures.

The summary of the speedup effect is shown in Table 1. This CSP process across 11 space group ranging in size from 46 to 368 atoms, achieving speedups of 1.95–2.15 $\times$ , with batch



Table 1 Summary of speed-up test results for T2 using BOMLIP-CSP

Number of atoms	Number of crystals	Baseline (s)	Batchopt (s)	Speed-up factor
92	1000	550.71	259.01	2.13
184	3500	3786.80	1764.76	2.15
368	1000	3355.27	1723.71	1.95
Mixed	5500	7976.74	3835.32	2.08

optimization delivering benefits across all problem sizes. Subsequently, we performed an end-to-end test by combining all six datasets in the table, achieving an average speedup of approximately  $2.08\times$ . We further evaluated more common space groups with different atom counts to monitor the speedup factor of the batched optimization method. This additional test included datasets ranging in size from 46 to 736 atoms, yielding speedups of  $1.78\text{--}2.57\times$  and an average speedup of  $2.31\times$  (Table S1).

### Benchmark systems and evaluation criteria

In this study, all 26 target systems from the first to the sixth CSP blind tests, together with unique polymorphs sharing the same asymmetric unit found in the CCDC database (CSD version 5.45; corresponding refcodes listed in Table S2), comprise a total of 34 distinct crystal structures that constitute the test set, as shown in Fig. 2. The experimentally obtained structure from the CCDC was separately added to the trial structure pool for relative energy ranking purposes. However, the added experimental structure was used solely to assess whether the foundation atomistic models could maintain its stability, a capability expected of any reliable method, and was therefore excluded from the hit count. All the trial structures were generated using the same method with a fixed random seed, followed by relaxation using different MLIPs as described in the methods section. This ensured that the optimization processes with different MLIPs

were conducted starting from exactly the same initial trial structures, making their results directly comparable.

Four fundamental challenges were selected to evaluate the performance of each MLIP in the CSP process of a given system. These challenges serve both as a benchmark to compare the effectiveness of different MLIPs within our workflow, and as a case study of applying MLIPs to the specific task of CSP. To the best of our knowledge, benchmarking studies involving a large number of structures far from equilibrium with the same chemical composition and similar energies generated during CSP have not been widely investigated.

These four evaluation criteria include the following: the first criterion is whether the experimental structure could remain stable during optimization. A reliable computational method should be able to preserve the packing mode of the experimental structure; otherwise, accurate crystal structure prediction would not be possible, regardless of how well other parts of the workflow perform. Second, the number of times the experimental structure is successfully predicted—referred to as the hit count—within a fixed number of trial structure optimizations. It directly reflects whether the CSP process has been successful. At the same time, a higher hit count indicates that the algorithm is more likely to locate the experimental structure at an early stage. In searches with a fixed total number of generated trial structures, achieving a higher hit count reduces the risk of missing the experimental structure,<sup>6,7,41,43</sup> whereas monitoring search convergence can also be effective but is more complex. The third criterion is the relative lattice energy of the experimental structures, and the fourth is the energy ranking of the experimental structures after duplicate removal. Both criteria evaluate the accuracy of the potential energy calculations and reflect the likelihood of identifying plausible structures through theoretical prediction before experimental validation. Although exceptions exist,<sup>5,7</sup> in most close-packed systems the experimental structures or polymorphs lie near the global minimum in lattice energy, typically ranking among the lowest in the energy ordering when evaluated with high-accuracy methods.<sup>15,24</sup> Moreover, previous studies have shown that the energy differences between most polymorphs are usually within a few  $\text{kJ mol}^{-1}$ .<sup>13,14</sup>

### Overall performance analysis

MACE-OFF-small and SevenNet-0-D3 were used to optimize all generated trial structures. These two MLIPs are computationally faster than other models, and both have implemented batched structural optimization in our code, resulting in further speed improvements. As a result, they can efficiently handle the evaluation of hundreds of thousands of structures.

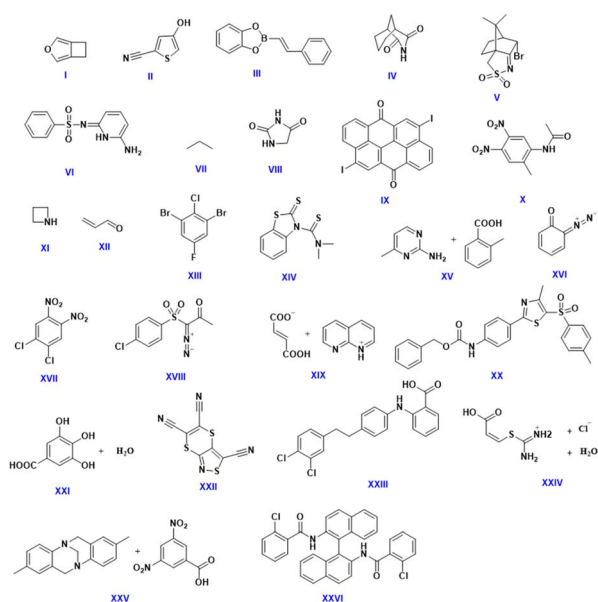


Fig. 2 Chemical diagrams of all the systems under study.



As shown in Fig. 3 (a detailed summary of all the results is provided in Table S2), most experimental crystal structures were successfully stabilized by the MLIPs during geometry optimization, except for two out of the 34 structures that could not be stabilized by any model, see Fig. 3a. These are system III, which contains a boron atom, and system XXII, which may be challenging due to its rare composition without any hydrogen atoms. Each foundation atomistic models was able to stabilize over 85% of the experimental structures (29/34 and 30/34, respectively), with certain systems being stabilized by only one of the two models.

The frequency with which the experimental structure is predicted is a critical criterion, as the energy ranking of candidate structures can be further refined in subsequent calculations using high-accuracy methods such as DFT. However, DFT is unlikely to recover a packing arrangement absent from the initial set of structures generated by the MLIP. In the test involving 2500 structures generated from up to four conformers within the experimental space group, both MLIPs achieved a success rate exceeding 50% in predicting the experimental structure (Fig. 3b). Thirteen experimental structures were predicted by both MLIPs, while twenty-four were predicted by at least one. Eleven experimental structures were predicted by only

one of the two MLIPs; in certain cases, despite being stabilizable by both and optimized from exactly the same set of generated trial structures, an experimental structure was recovered by only one model. This demonstrates that the choice of MLIP can affect prediction outcomes. For smaller molecules, the experimental structure hit count is comparatively higher than that of larger molecules due to the significantly smaller packing space. The number of successful predictions for co-crystals as well as flexible molecules with more rotatable bonds is comparatively lower, which is consistent with conventional understanding in CSP. These systems typically require a larger number of trial structures to be adequately sampled.

The relative lattice energy and energy ranking of the experimental structures for each system are shown in Fig. 3c and d. Nearly all experimental structures lie within 25 kJ mol<sup>-1</sup> of the global minimum. For MACE-OFF-small, the average relative energy is 8.40 kJ mol<sup>-1</sup>, with 20 structures below 10 kJ mol; the highest value is 60.11 kJ mol<sup>-1</sup> (system XVII), followed by 18.33 kJ mol<sup>-1</sup>. For SevenNet-0-D3, the average relative energy is 9.28 kJ mol<sup>-1</sup>, with 15 structures below 10 kJ mol; the highest value is 23.53 kJ mol<sup>-1</sup> (system XXIII). For the energy-ranking criterion, 24 and 22 experimental structures rank within the top 100 for MACE-OFF-small and SevenNet-0-D3, respectively.

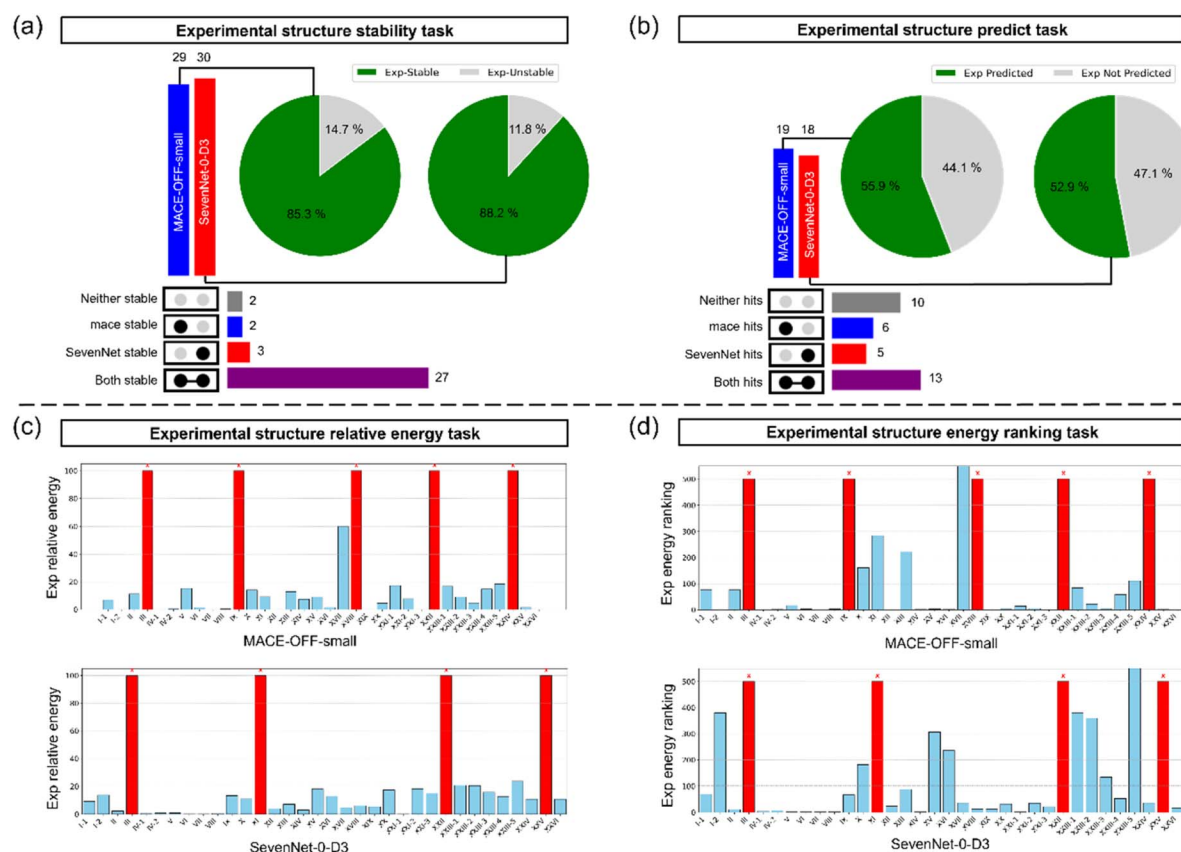


Fig. 3 Benchmarking crystal-structure prediction on all systems across four tasks. (a) Stability test: UpSet plot and pie charts show how many experimental structures are stabilized upon relaxation by MACE-OFF-small (blue), SevenNet-0-D3 (red), or by both models. (b) Structure-prediction test: UpSet plot and pie charts show how many experimental structures are correctly recovered by the CSP workflow using each model. The accompanying bar charts show the detailed breakdown of prediction outcomes. (c) Relative lattice energy of experimental structures calculated using the two MLIPs for each system. (d) Relative lattice energy ranking of experimental structures after duplicate removal. Structures that failed to stabilize during optimization are marked in red in the bar chart.



Moreover, 17 and 7 structures rank within the top 5 for the two models. Noted that experimentally observed structures are not guaranteed to appear in the top 5 of the potential-energy ranking, as polymorphism and free-energy corrections under specific external conditions can substantially alter their relative stability.

Overall, these performances are relatively solid in exploring the energy landscape of CSP, and they can also serve as a key component of the structure generation stage in a hierarchical strategy, which is widely used<sup>15,44,45</sup> and typically follows with DFT or higher-accuracy calculations for pre-experimental prediction.

Guided by the Exp. Stability metric (Fig. 3a and Table S2), failures can be broadly attributed to two scenarios. First, when the experimental structure is stable under a given MLIP but no trial structures sufficiently close to the experimental packing are generated, the failure mainly reflects a limited number of trial structures. Second, when the experimental structure is not stable under that MLIP, the failure is primarily due to deficiencies in the MLIP potential-energy surface rather than the trial-structure generation stage. It should be noted that this does not represent the full capability of our CSP strategy, as more experimental structures could be predicted with a larger number of trial structures or by using other models. For example, in our additional tests, BOQQUT01 and HMTIZ01 could be successfully predicted using MACE-OFF-small if a larger number of trial structures were generated, while XATMIP, XATMOV, SOXLEX01, OBEQET, XATJOT, and NACJAF could be recovered when using MACE-OFF-large.

### In-depth case studies

We next examine two systems (VI and VII) in detail to gain deeper insights into the overall results. After removing

duplicates, the 200 lowest-energy structures from each MLIPs were selected, yielding a total of 400 structures. The Smooth Overlap of Atomic Positions (SOAP)<sup>46</sup> descriptor was used to encode all crystal structures, and the average kernel<sup>47</sup> was applied to quantify the similarity between each pair of structures. This similarity matrix was then used as a precomputed input for dimensionality reduction using the Uniform Manifold Approximation and Projection (UMAP) algorithm,<sup>48</sup> in order to capture the high-dimensional distance relationships. In the resulting SOAP-space projection, each point represents a crystal structure, and the distance between points reflects structural similarity—closer points correspond to more similar structures in SOAP space.

System VI represents a relatively straightforward case in which the experimental structure is correctly predicted by both MLIPs, with its energy ranking reproduced within the top three. The SOAP-space distributions generated by the two models are largely consistent, with only minor differences in the coverage of certain regions. An additional UMAP plot, color-coded by crystal density, is provided in Fig. S3. Crystal packing similarity tests for several pairs of nearby points in the projection are shown in the inset of Fig. 4, where the experimental structures predicted by both models appear close to each other. Furthermore, we evaluated how many structures among these are commonly predicted by both models. To evaluate structural overlap, we examined the 200 lowest-energy predictions from both models and found 37 shared structures, each achieving a 15/15 match (see the Methods section for the definition of a match) in the crystal packing similarity test. Other nearby pairs may also exhibit similar packing modes, and the absence of a full 15/15 match in some cases is primarily attributable to slight differences in the predicted unit cell lengths and volumes, as illustrated in the plot.

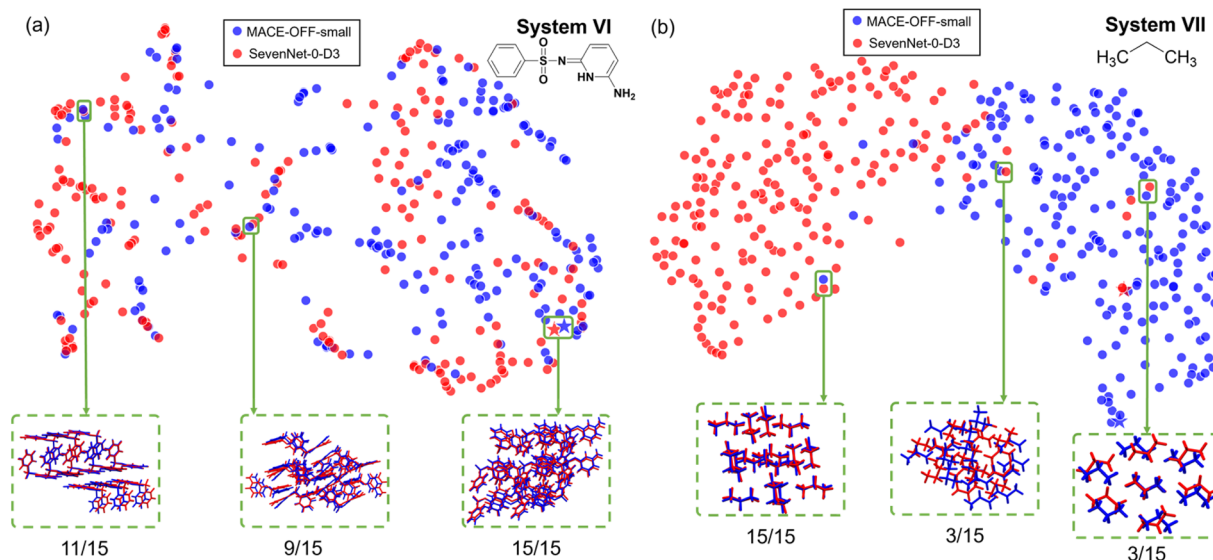


Fig. 4 Analysis of CSP results involving two different foundation atomistic models for selected systems. Two-dimensional UMAP embedding of the SOAP feature space for (a) system VI and (b) system VII. Predicted structures optimized with MACE-OFF-small are shown in blue, while those optimized with SevenNet-0-D3 are shown in red. Structures matching the experimental crystals are marked with star symbols. A randomly selected subset of points was analyzed for packing similarity against neighboring structures, and their corresponding crystal packings are displayed to give readers a clearer understanding of the 3D arrangements.



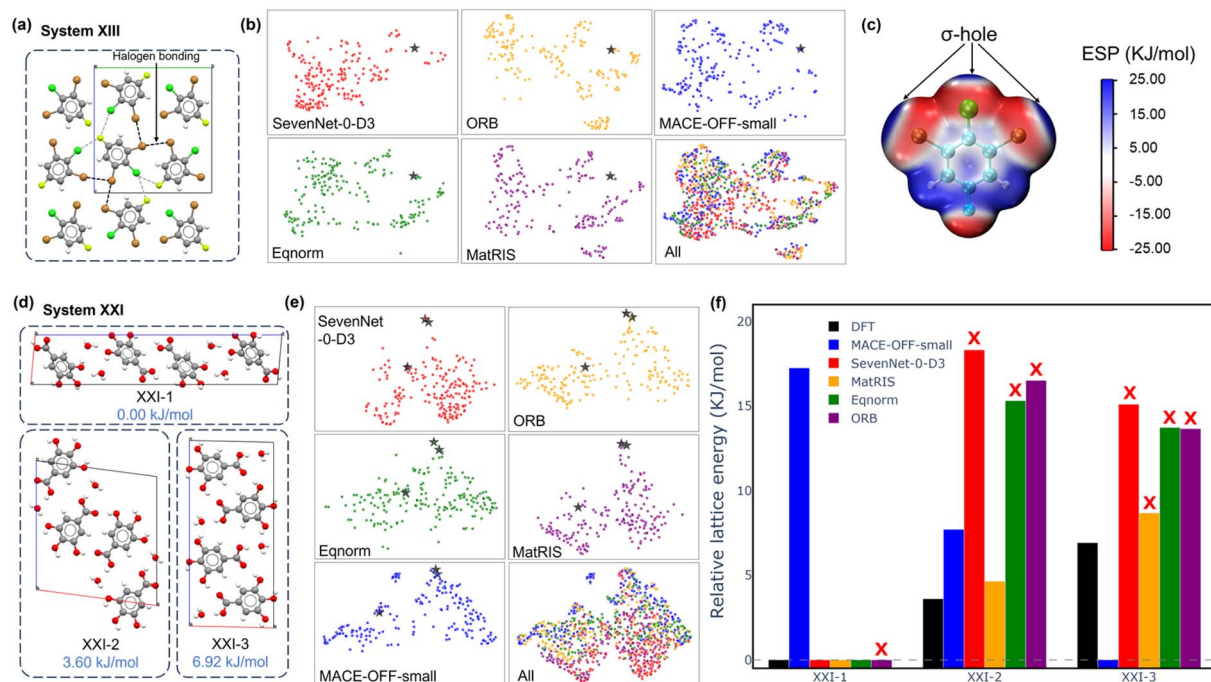
In contrast, system VII presents a more complex case. Although the experimental structure is the global minimum for both atomistic models, one predicts it ten times while the other fails entirely, despite using the same initial trial structures. Dimensionality-reduction reveals that the experimental structure and several points from SevenNet-0-D3 overlap with the MACE-OFF-small cloud, but most SevenNet-0-D3 points are clearly separated—showing that the two models explore distinct regions of structural space. In SevenNet-0-D3, a packing mode close to the experimental form was repeatedly found, but propane conformers were flipped (Fig. S4), producing a local minimum that blocked further optimization. Among the 200 lowest-energy predictions, 111 were shared by both models, reflecting the smaller packing space of small molecules. The packing similarity test and the SOAP descriptor probe different aspects—the former global packing and the latter local atomic environments—making them complementary rather than contradictory in analyzing the results.

### Evaluation of additional MLIPs on challenging systems

Next, we applied additional foundation atomistic models to more challenging systems. The results are summarized in Table S3. System XIII consists of 1,3-dibromo-2-chloro-5-fluorobenzene, a molecule characterized by a rich distribution of halogen atoms along its outer edge. These atoms, particularly chlorine and bromine, can act as both halogen bond donors

and acceptors. As shown in Fig. 5c, the electrostatic potential (ESP) surface of the molecule reveals three positive regions along the C–X bond extensions, indicating the presence of  $\sigma$ -holes<sup>49</sup> and thereby enabling directional intermolecular interactions. This anisotropic charge distribution poses a significant challenge for classical force fields,<sup>50</sup> making the system difficult for some traditional CSP methods to handle.

As illustrated in the UMAP embedding of the SOAP feature space for all models (Fig. 5b), distinct patterns emerge in the low-energy regions across models. Nevertheless, all MLIPs are capable of stabilizing the experimental structure, except for the MACE-OFF-small model, which failed to predict it. This failure may be due to its comparatively less accurate energy predictions for this system, as reflected by the higher energy ranking assigned to the experimental structure. Larger models such as Eqnorm, ORB,<sup>51</sup> and MatRIS exhibit superior performance for this small system—not only in terms of the relative lattice energy of the experimental structure but also in the frequency of successful predictions—despite requiring substantially longer computation times to complete the CSP process compared to the other two models. The energy rankings of the experimental structures obtained by Eqnorm, ORB, and MatRIS are 1, 1, and 2, respectively, indicating excellent performance in crystal structure prediction tasks. Fig. S5 shows a packing similarity analysis of structures optimized by different MLIPs, originating from seed IDs for which at least one MLIP successfully recovers the experimental structure.



**Fig. 5** Analysis of CSP results across multiple MLIPs. (a) 3D crystal structure of system XIII as observed experimentally. Representative halogen bonds are highlighted. (b) Two-dimensional UMAP embedding of the SOAP feature space for system XIII, with separate plots for each foundation atomistic model and a combined overview. The star represents the experimentally observed structure. (c) Electrostatic potential surface of the molecule in system XIII (isovalue: 0.001 electrons per Bohr<sup>3</sup>), with  $\sigma$ -holes on halogen-bond donors marked. (d) 3D crystal structure of system XXI as observed experimentally, with the corresponding DFT relative energy annotated. (e) Two-dimensional UMAP embedding of the SOAP feature space for system XXI, with separate plots for each foundation atomistic model and a combined overview. The star represents the experimentally observed structure. (f) Relative lattice energies predicted by different MLIPs, including the reference DFT result. A red cross above the bar indicates that the corresponding structure was not successfully predicted.



System XXI represents another challenging case. It contains both 3,4,5-trihydroxybenzoic acid monohydrate and an additional water molecule, creating numerous possibilities for diverse hydrogen-bonding networks. The high flexibility of the hydroxyl groups further increases the complexity of the intermolecular packing landscape. As a result, this system exhibits three known polymorphs with the same asymmetric unit composition (XXI-1, XXI-2, and XXI-3; CCDC refcodes: KONTIQ, KONTIQ01, KONTIQ03).

The dimensionality-reduction plots for the five foundation atomistic models also reveal distinct patterns. For MACE-OFF-small and MatRIS, more points are clustered around the experimentally observed structure (denoted by the star symbol) compared with the other models, and both exhibit superior performance in locating experimental structures, as shown in Table S3. The relative lattice energies and rankings of these three known polymorphs were evaluated using DFT as the reference in order to test whether the MLIPs can reproduce the DFT polymorph energy ordering. This level of theory provides a relatively reliable preliminary standard for energy ranking, although more accurate levels of electronic-structure calculation may lead to slight changes in the ordering.<sup>15</sup> When we compared the MLIP results with the DFT rankings, we found that MACE-OFF-small—the only foundation atomistic model having predicted all experimental structures within 5000 trial structures—incorrectly identified XXI-3 as the most stable form, in complete contradiction to the DFT reference results. By contrast, the MatRIS model successfully predicted two of the experimental structures, and its relative energy predictions for all polymorphs are in good agreement with DFT. Surprisingly, some MLIPs larger than MACE-OFF-small failed to predict any, or only one or two, of the experimental structures. This may be because models with a larger number of parameters tend to produce a more rugged crystal lattice energy landscape, requiring a greater number of trial structures to adequately converge and sample the broader packing space.

## Discussion

By systematically evaluating CSP results derived from different foundation atomistic models using an identical set of trial crystal structures, we establish a direct benchmark for assessing their influence on the CSP workflow. This approach reveals a reality of CSP: in realistic scenarios, where the target structure is unknown and initial guesses may be far from the training domain, success depends not only on local energy accuracy near equilibrium but also on the global shape and ruggedness of the potential energy surface. Overall, we find that two state-of-the-art foundation atomistic models, MACE-OFF-small and SevenNet-0-D3, are capable of successfully predicting the experimental structure in over half of the tested systems using a limited number of trial structures. Moreover, in most cases these experimental structures also lie in the low-energy region, making them an excellent database for more accurate energy-ranking methods, such as DFT calculations, to achieve more reliable predictions. For more than 85% of the experimental structures, the models correctly stabilize them, suggesting that

increasing the number of trial structures would likely lead to successful predictions. These results demonstrate that even MLIPs with relatively small parameter counts can be effectively applied to CSP tasks, offering both good accuracy and high computational efficiency.

Using the energy difference ( $\Delta E$ ) between two MLIPs as a proxy for energy accuracy (the closer the experimental structure's relative energy is to zero, the better), we conducted a sign-consistency test between energy advantage and hit advantage. Among systems allowing a fair comparison, 12 cases favored the lower-energy model with more hits, whereas 8 favored the higher-energy model—indicating only a weak association between energy accuracy and hit count. Surprisingly, some experimental structures were successfully predicted by one MLIP but completely missed by another (see Fig. S5), despite both being applied to the same initial trial structures and both being capable of stabilizing the experimental geometries. Moreover, the energy prediction accuracy of an MLIP does not necessarily translate into greater advantages in CSP, especially in the crystal-structure search stage. Considering that larger models can be an order of magnitude more time-consuming than smaller ones, model choice remains an important practical consideration, although developing a robust prospective selection criterion lies beyond the scope of the present work. This discrepancy may arise because, beyond accuracy near local minima, differences in the MLIP-predicted lattice energy surface for a given system can also play a complex role in determining the final CSP outcome. Although some foundation models are incompatible with certain elements, the three steps of BOMLIP-CSP—conformer search, trial structure generation, and optimization—are independent. This modular design allows the workflow to be adapted by replacing the foundation model when necessary, for example with a model that supports the relevant elements. Training a model for target systems of particular interest is also a feasible strategy, similar to the customized force fields used in traditional approaches, although the associated time cost needs to be carefully considered. In addition, proton-transfer events may occur when using MLIPs. However, this does not always act in a purely detrimental way: for some systems, the hydrogen positions in the crystal may indeed involve genuine proton transfer. A detailed investigation of such effects lies beyond the scope of the present work, and here we simply note their possible presence as an additional source of complexity. In addition, evaluating a larger number of systems would further strengthen this assessment, and this will be pursued in future work.

## Conclusion

In this study, we present an open-source, extensible CSP framework that employs MLIPs for structure optimization and energy evaluation. A batched optimization strategy tailored for CSP enables large-scale calculations and delivers about 2.1–2.3× speedup in tests based on BOMLIP-CSP, thereby allowing efficient prediction of multiple systems within limited time. The framework is readily adaptable to other MLIPs with minimal code modifications, enhancing flexibility and broad applicability in CSP.



The results demonstrate that, when the correct space group and  $Z'$  are included in the search, the framework can readily recover over 50% of the experimental crystal structures during structure generation. When MLIPs are chosen appropriately, the success rate exceeds 70% under the present benchmark conditions, with most of these experimental structures also appearing among the lower-energy candidates. This number could be even higher with more extensive sampling. Importantly, all these accuracies are obtained using relatively small MLIPs and they have been substantially accelerated.

The benchmark also shows that even with the same trial database and MLIPs stabilizing experimental structures, different MLIPs can yield very different CSP outcomes, underscoring that not only equilibrium accuracy but also the shape of the crystal lattice energy landscape and connectivity of energy basin are critical.

Looking ahead, our framework offers substantial flexibility for integrating MLIPs into CSP workflows—for example, enabling structure generation using different MLIPs while applying a unified energy ranking method. Such capabilities will accelerate the application of AI-driven approaches in chemical materials discovery, allowing computational chemistry to play a more autonomous role in solid materials, and addressing key challenges in the field of molecular materials science.

## Methods section

### Crystal structure prediction

In all systems, for each molecule, a conformer search was first conducted, followed by energy evaluation using the UFF. Up to four of the lowest-energy conformers were then selected for subsequent trial structure generation. For each target, the search space was constrained to the space group and  $Z'$  of the corresponding experimental structure, while the conformers were generated as described above. For each conformer, 2500 trial crystal structures were generated, resulting in up to 10 000 trial structures per target system. All structures were then subjected to relaxation using a batched structural optimization protocol consisting of two rounds: the first round was performed under an applied pressure of 0.1 GPa, followed by a second round of optimization with the pressure released. In our previous tests, this two-step strategy produced improved results without artificially eliminating porous structures due to the initial pressure. After optimization, a very small number of structures in certain systems exhibited clear molecular collapse and unphysically low energies (around  $-1 \times 10^{15}$  to  $-1 \times 10^{25}$  kJ mol<sup>-1</sup>), likely due to instabilities in the MLIP. Therefore, all structures with absolute energies greater than  $1 \times 10^{15}$  kJ mol<sup>-1</sup> were removed. For the low-energy structures, a duplicate removal process was performed using the CCDC API. Starting from the structure with the lowest energy, each structure was sequentially examined and added to the unique group only if it was determined to be dissimilar to all existing members in unique group. To assess structural matching and identify duplicates, a cluster of 15 molecules was extracted from each structure, and two structures were considered duplicates if

their clusters exhibited a 15/15 match in the CCDC packing similarity test with a distance tolerance of 0.2 and an angle tolerance of 20°, excluding hydrogen atoms from the comparison. The same CCDC packing-similarity criterion was used for matching predicted structures to the experimental reference and for duplicate identification.

### Conformer search for the given molecule

Conformer search for the target molecule was conducted using the ETKDG<sup>52</sup> method as implemented in the open-source library RDKit library.<sup>53</sup> Subsequently, an energy evaluation was performed using the Universal Force Field (UFF),<sup>54</sup> and the four lowest-energy conformers were selected for use in the subsequent stages.

### Visualization of the SOAP space

The UMAP technique was used for dimensionality reduction for mapping high-dimensional data to 2D representations, while preserving both global and local topological structures of the data in the high-dimensional space as much as possible. That is, the points are arranged spatially such that the closer the two points are on the 2D plot, the more similar the two molecules are, as described by the encoding descriptors. SOAP descriptors were generated for all atoms in the crystal structure, using the DDescribe package.<sup>55</sup> The average kernel was used to measure global similarity between crystal structures from SOAP-encoded local atomic environments.

### Density functional theory calculations

Periodic DFT calculations were carried out within the plane-wave pseudopotential formalism, using the Vienna *ab initio* simulation package (VASP) code version 6.4.3.<sup>56</sup> Projector augmented-wave (PAW) method was applied to describe the electron-ion interactions.<sup>57</sup> Generalized gradient approximation (GGA) with the Perdew–Burke–Ernzerhof (PBE) exchange correlation functional was adopted to treat electron interaction energy.<sup>58</sup> Grimme's semi-empirical DFT-D3 scheme with Becke–Johnson damping functions was used here to give a better description of interactions.<sup>59,60</sup>

A kinetic-energy cut-off of 600 eV was used to define the plane-wave basis set. Electronic Brillouin zone was integrated with the smallest allowed spacing between  $k$ -points (KSPACING) being  $0.4 \text{ \AA}^{-1}$  and the generated grid is centred at the  $\Gamma$ -point. Convergence threshold of self-consistency was set to  $10^{-6}$  eV during total energy and force calculation and the Hellmann–Feynman force convergence criterion on each atom was set to smaller than  $0.01 \text{ eV \AA}^{-1}$  when geometry optimizations are needed.

## Author contributions

L. C., J. J., E. S., and G. T. conceived and supervised the project. C. Z. developed the CSP code and performed result interpretation. Z. M. and L. W. optimized the code performance and implemented the batched optimization part. D. F. optimized the code performance, fixed bugs, and carried out all the



calculations. S. H. and W. J. contributed to the installation and test of all foundation model. F. H. assisted with testing during the revision stage. C. Z. and L. C. led the manuscript preparation with input from all authors.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data supporting the findings of this study are available within the paper and its supplementary information (SI) files. The experimentally determined crystal structures used to define the benchmark were obtained from the Cambridge Structural Database and can be retrieved using the refcodes listed in Table S2. Code and scripts developed and used in this study are available at <https://github.com/pic-ai-robotic-chemistry/BOMLIP-CSP>. The archived version of the code is available on Zenodo at: <https://doi.org/10.5281/zenodo.19343560>. The latest version of the software is available *via* the Zenodo concept DOI at: <https://doi.org/10.5281/zenodo.19343559>. Supplementary information is available. See DOI: <https://doi.org/10.1039/d6dd00016a>.

## Acknowledgements

The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences (CAS). L. C. acknowledges the University of Science and Technology of China (USTC) Startup Program for funding. J. J. acknowledges the CAS Strategic Priority Research Program (Grant XDB0450302), the National Natural Science Foundation of China (Grants 22025304, 22033007), and the CAS Project for Young Scientists in Basic Research (Grant YSBR-005) for funding. W. J. and G. T. acknowledge the National Natural Science Foundation of China (Grants T2125013, 92270206, 62032023) for funding, and G. T. acknowledge Funding of Institute of Computing Technology, Chinese Academy of Sciences (Grant No. E461030).

## References

- C. B. Meier, R. S. Sprick, A. Monti, P. Guiglion, J.-S. M. Lee, M. A. Zwijnenburg and A. I. Cooper, Structure-property relationships for covalent triazine-based frameworks: the effect of spacer length on photocatalytic hydrogen evolution from water, *Polymer*, 2017, **126**, 283–290.
- D. Firaha, Y. M. Liu, J. van de Streek, K. Sasikumar, H. Dietrich, J. Helfferich, L. Aerts, D. E. Braun, A. Broo and A. G. DiPasquale, Predicting crystal form stability under real-world conditions, *Nature*, 2023, **623**(7986), 324–328.
- X. Wang, L. Chen, S. Y. Chong, M. A. Little, Y. Wu, W.-H. Zhu, R. Clowes, Y. Yan, M. A. Zwijnenburg and R. S. Sprick, Sulfone-containing covalent organic frameworks for photocatalytic hydrogen evolution from water, *Nat. Chem.*, 2018, **10**(12), 1180–1189.
- J. Jia, J. Liu, Z.-Q. Wang, T. Liu, P. Yan, X.-Q. Gong, C. Zhao, L. Chen, C. Miao and W. Zhao, Photoinduced inverse vulcanization, *Nat. Chem.*, 2022, **14**(11), 1249–1257.
- C. Zhao, L. Chen, Y. Che, Z. Pang, X. Wu, Y. Lu, H. Liu, G. M. Day and A. I. Cooper, Digital navigation of energy–structure–function maps for hydrogen-bonded porous molecular crystals, *Nat. Commun.*, 2021, **12**(1), 817.
- M. O'Shaughnessy, J. Glover, R. Hafizi, M. Barhi, R. Clowes, S. Y. Chong, S. P. Argent, G. M. Day and A. I. Cooper, Porous isorecticular non-metal organic frameworks, *Nature*, 2024, **630**(8015), 102–108.
- A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo and C. J. Stackhouse, Functional materials discovery using energy–structure–function maps, *Nature*, 2017, **543**(7647), 657–664.
- G. J. Beran, I. J. Sugden, C. Greenwell, D. H. Bowskill, C. C. Pantelides and C. S. Adjiman, How many more polymorphs of ROY remain undiscovered?, *Chem. Sci.*, 2022, **13**(5), 1288–1297.
- D. H. Case, J. E. Campbell, P. J. Bygrave and G. M. Day, Convergence properties of crystal structure prediction by quasi-random sampling, *J. Chem. Theory Comput.*, 2016, **12**(2), 910–924.
- R. Tom, T. Rose, I. Bier, H. O'Brien, Á. Vázquez-Mayagoitia and N. Marom, Genarris 2.0: a random structure generator for molecular crystals, *Comput. Phys. Commun.*, 2020, **250**, 107170.
- Q. Zhu and S. Hattori, Automated high-throughput organic crystal structure prediction *via* population-based sampling, *Digital Discovery*, 2025, **4**(1), 120–134.
- M. A. Neumann, Tailor-made force fields for crystal-structure prediction, *J. Phys. Chem. B*, 2008, **112**(32), 9810–9829.
- S. L. Price, From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape, *Phys. Chem. Chem. Phys.*, 2008, **10**(15), 1996–2009.
- J. Nyman and G. M. Day, Static and lattice vibrational energy differences between polymorphs, *CrystEngComm*, 2015, **17**(28), 5154–5165.
- J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio Jr and A. Tkatchenko, Reliable and practical computational description of molecular crystal polymorphs, *Sci. Adv.*, 2019, **5**(1), eaau3338.
- V. Gharakhanyan, Y. Yang, L. Barroso-Luque, M. Shuaibi, D. S. Levine, K. Michel, V. Bernat, M. Dzamba, X. Fu and M. Gao, FastCSP: Accelerated Molecular Crystal Structure Prediction with Universal Model for Atoms, *arXiv*, 2025, preprint, arXiv: 2508.02641, DOI: [10.48550/arXiv.2508.02641](https://doi.org/10.48550/arXiv.2508.02641).
- J. P. Lommerse, W. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. Hofmann, F. J. Leusen, W. T. Mooij, S. L. Price and B. Schweizer, A test of crystal structure prediction of small organic molecules, *Acta Crystallogr. Sect. B Struct. Sci.*, 2000, **56**(4), 697–714.
- W. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. Hofmann,



- F. J. Leusen, J. P. Lommerse and W. T. Mooij, Crystal structure prediction of small organic molecules: a second blind test, *Acta Crystallogr. Sect. B Struct. Sci.*, 2002, **58**(4), 647–661.
- 19 G. Day, W. Motherwell, H. Ammon, S. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer and B. Van Eijck, A third blind test of crystal structure prediction, *Acta Crystallogr. Sect. B Struct. Sci.*, 2005, **61**(5), 511–527.
- 20 G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti and J. Jose, Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test, *Acta Crystallogr. Sect. B Struct. Sci.*, 2009, **65**(2), 107–125.
- 21 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle and G. R. Desiraju, Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test, *Acta Crystallogr. Sect. B Struct. Sci.*, 2011, **67**(6), 535–551.
- 22 A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell and R. Car, Report on the sixth blind test of organic crystal structure prediction methods, *Acta Crystallogr. Sect. B Struct. Sci.*, 2016, **72**(4), 439–459.
- 23 L. M. Hunnisett, J. Nyman, N. Francia, N. S. Abraham, C. S. Adjiman, S. Aitipamula, T. Alkhidir, M. Almehairbi, A. Anelli and D. M. Anstine, The seventh blind test of crystal structure prediction: structure generation methods, *Struct. Sci.*, 2024, **80**(6), 517–547.
- 24 L. M. Hunnisett, N. Francia, J. Nyman, N. S. Abraham, S. Aitipamula, T. Alkhidir, M. Almehairbi, A. Anelli, D. M. Anstine and J. E. Anthony, The seventh blind test of crystal structure prediction: structure ranking methods, *Struct. Sci.*, 2024, **80**(6), 548–574.
- 25 Z. Ye, N. Wang, J. Zhou and D. Ouyang, Organic crystal structure prediction *via* coupled generative adversarial networks and graph convolutional networks, *The Innovation*, 2024, **5**(2), 100562.
- 26 X. Li, Y. Che, L. Chen, T. Liu, K. Wang, L. Liu, H. Yang, E. O. Pyzer-Knapp and A. I. Cooper, Sequential closed-loop Bayesian optimization as a guide for organic molecular metallophotocatalyst formulation discovery, *Nat. Chem.*, 2024, **16**(8), 1286–1294.
- 27 A. P. Bartók and G. Csányi, Gaussian approximation potentials: a brief tutorial introduction, *Int. J. Quantum Chem.*, 2015, **115**(16), 1051–1057.
- 28 J. Behler, Four generations of high-dimensional neural network potentials, *Chem. Rev.*, 2021, **121**(16), 10037–10072.
- 29 J. Behler, First principles neural network potentials for reactive simulations of large molecular and condensed systems, *Angew. Chem., Int. Ed.*, 2017, **56**(42), 12828–12840.
- 30 T. Plé; O. Adjoua; A. Benali; E. Posenitskiy; C. Villot; L. Lagardère; J.-P. Piquemal, *A Foundation Model For Accurate Atomistic Simulations In Drug Design*, 2025.
- 31 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon and W. J. Baldwin, A Foundation Model For Atomistic Materials Chemistry, *J. Chem. Phys.*, 2025, **163**(18), 184110.
- 32 J. Riebesell, R. E. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain and K. A. Persson, A framework to evaluate machine learning crystal stability predictions, *Nat. Mach. Intell.*, 2025, **7**(6), 836–847.
- 33 D. McDonagh, C.-K. Skylaris and G. M. Day, Machine-learned fragment-based energies for crystal structure prediction, *J. Chem. Theory Comput.*, 2019, **15**, 2743–2758.
- 34 S. Wengert, G. Csányi, K. Reuter and J. T. Margraf, A hybrid machine learning approach for structure stability prediction in molecular co-crystal screenings, *J. Chem. Theory Comput.*, 2022, **18**, 4586–4593.
- 35 K. S. Nayal, D. O'Connor, R. Zubatyuk, D. M. Anstine, Y. Yang, R. Tom, W. Deng, K. Tang, N. Marom and O. Isayev, Efficient Molecular Crystal Structure Prediction and Stability Assessment with AIMNet2 Neural Network Potentials, *Cryst. Growth Des.*, 2025, **25**, 9092–9106.
- 36 C. J. Nickerson and E. R. Johnson, Assessment of a foundational machine-learned potential for energy ranking of molecular crystal polymorphs, *Phys. Chem. Chem. Phys.*, 2025, **27**(22), 11930–11940.
- 37 I. M. Sobol, Uniformly distributed sequences with an additional uniform property, *USSR Comput. Math. Math. Phys.*, 1976, **16**(5), 236–242.
- 38 R. G. Della Valle, E. Venuti, A. Brillante and A. Girlando, Inherent structures of crystalline pentacene, *J. Chem. Phys.*, 2003, **118**(2), 807–815.
- 39 B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán and S. Alvarez, Covalent radii revisited, *Dalton Trans.*, 2008, (21), 2832–2838.
- 40 A. v. Bondi, van der Waals Volumes and Radii, *J. Phys. Chem.*, 1964, **68**(3), 441–451.
- 41 Q. Zhu, J. Johal, D. E. Widdowson, Z. Pang, B. Li, C. M. Kane, V. Kurlin, G. M. Day, M. A. Little and A. I. Cooper, Analogy powered by prediction and structural invariants: computationally led discovery of a mesoporous hydrogen-bonded organic cage crystal, *J. Am. Chem. Soc.*, 2022, **144**(22), 9893–9901.
- 42 M. Feng, C. Zhao, G. M. Day, X. Evangelopoulos and A. I. Cooper, A universal foundation model for transfer learning in molecular crystals, *Chem. Sci.*, 2025, **16**(28), 12844–12859.
- 43 P. Cui, D. P. McMahon, P. R. Spackman, B. M. Alston, M. A. Little, G. M. Day and A. I. Cooper, Mining predicted crystal structure landscapes with high throughput crystallisation: old molecules, new insights, *Chem. Sci.*, 2019, **10**(43), 9988–9997.
- 44 A. Mattei, R. S. Hong, H. Dietrich, D. Firaha, J. Helfferich, Y. M. Liu, K. Sasikumar, N. S. Abraham, R. Miglani Bhardwaj and M. A. Neumann, Efficient crystal structure prediction for structurally related molecules with accurate and transferable tailor-made force fields, *J. Chem. Theory Comput.*, 2022, **18**(9), 5725–5738.



- 45 J. Nyman, L. Yu and S. M. Reutzel-Edens, Accuracy and reproducibility in crystal structure prediction: the curious case of ROY, *CrystEngComm*, 2019, **21**(13), 2080–2088.
- 46 A. P. Bartók, R. Kondor and G. Csányi, On representing chemical environments, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2013, **87**(18), 184115.
- 47 S. De, A. P. Bartók, G. Csányi and M. Ceriotti, Comparing molecules and solids across structural and alchemical space, *Phys. Chem. Chem. Phys.*, 2016, **18**(20), 13754–13769.
- 48 J. Healy and L. McInnes, Uniform manifold approximation and projection, *Nat. Rev. Methods Primers*, 2024, **4**(1), 82.
- 49 G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, The halogen bond, *Chem. Rev.*, 2016, **116**(4), 2478–2601.
- 50 C. Zhao, H. Liu, D.-H. Qu, A. I. Cooper and L. Chen, A machine learned potential for investigating single crystal to single crystal transformations in complex organic molecular systems, *Chem. Sci.*, 2025, **16**(5), 2363–2372.
- 51 B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin; J. Godwin, T. Duignan and M. Neumann, Orb-V3: Atomistic Simulation At Scale, *arXiv*, 2025, preprint, arXiv:2504.06231, DOI: [10.48550/arXiv.:2504.06231](https://doi.org/10.48550/arXiv.2504.06231).
- 52 S. Riniker and G. A. Landrum, Better informed distance geometry: using what we know to improve conformation generation, *J. Chem. Inf. Model.*, 2015, **55**(12), 2562–2574.
- 53 G. Landrum, *RdKit: Open-Source Cheminformatics Software*, 2016, , vol. 149, 150), p. 650, <https://www.rdkit.org/>, <https://github.com/rdkit/rdkit>.
- 54 A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard III and W. M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.*, 1992, **114**(25), 10024–10035.
- 55 L. Himanen, M. O. Jäger, E. V. Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, Dscribe: Library of descriptors for machine learning in materials science, *Comput. Phys. Commun.*, 2020, **247**, 106949.
- 56 G. Kresse and J. Furthmüller, *Vienna Ab-Initio Simulation Package (VASP)*, Vienna, Vienna University, 2001.
- 57 P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B*, 1994, **50**(24), 17953.
- 58 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**(18), 3865.
- 59 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu, *J. Chem. Phys.*, 2010, **132**(15), 154104.
- 60 A. D. Becke and E. R. Johnson, A density-functional model of the dispersion interaction, *J. Chem. Phys.*, 2005, **123**(15), 154101.

