nature computational science

Article

https://doi.org/10.1038/s43588-025-00775-z

Efficient crystal structure prediction based on the symmetry principle

Received: 16 March 2024

Accepted: 28 January 2025

Published online: 27 February 2025

Check for updates

Yu Han[1,5], Chi Ding[1,5], Junjie Wang[1] ✉, Hao Gao[2,3,4] ✉, Jiuyang Shi[1], Shaobo Yu[1], Qiuhan Jia[1], Shuning Pan[1] & Jian Sun[1] ✉

Crystal structure prediction (CSP) is an evolving field aimed at discerning crystal structures with minimal prior information. Despite the success of various CSP algorithms, their practical applicability remains circumscribed, particularly for large and complex systems. Here, to address this challenge, we show an evolutionary structure generator within the MAGUS (Machine Learning and Graph Theory Assisted Universal Structure Searcher) framework, inspired by the symmetry principle. This generator extracts both global and local features of explored crystal structures using group and graph theory. By integrating an on-the-fly space group miner and fragment reorganizer, augmented by symmetry-kept mutation, our approach generates higher-quality initial structures, reducing the computational costs of CSP tasks. Benchmarking tests show up to fourfold performance improvements. The method also proves valid in complex phosphorus allotrope systems. Furthermore, we apply our approach to the diamond–silicon (111)-(7 × 7) surface system, identifying up to 42 metastable structures within an 18 meV Å^{-2} energy range, demonstrating the efficacy of our approach in navigating challenging search spaces.

Over the past decade, crystal structure prediction (CSP) algorithms combined with density functional theory have evolved into potent instruments for the exploration of novel materials[1–6]. A detailed but not complete summary of various CSP methods and their applications have been discussed in ref. 6. The operational paradigm of a CSP program involves the generation of a multitude of structures, often ranging from several hundred to potentially thousands. Then these trial structures commonly undergo a fitness evaluation, typically based on enthalpy or other properties such as hardness, or their alignment with X-ray diffraction patterns in the case of multi-target searches. The ultimate prediction is then derived by selecting the structure with the highest assessed fitness.

Various approaches exist for enhancing the efficiency of CSP. In recent years, one mainstream direction has been accelerating the computation of 'fitness' by substituting computationally expensive ab initio calculations with surrogate machine learning potentials (MLPs)[7–11], thereby mitigating the overall computational cost of CSPs[6,12–15]. However, the accuracy of MLPs alone cannot guarantee successful CSPs without effective sampling methods, which aim to increase the proportion of relatively reasonable trial structures. For instance, MLP-driven CSP fails in determining complex phosphorus allotropes due to exponentially growing complexity with the number of atoms[16]. Consequently, sampling methods play a crucial role in the performance of CSP approaches. Drawing inspiration from global optimization algorithms, numerous implementations have shown efficacy in searching for the global minima of the potential energy surface (PES). Strategies such as the evolutionary method, involving cutting two searched structures into halves and splicing them into a new one[2,4,6], the particle swarm optimization method, which derives a new set of atom positions through motion equations[5], and Bayesian optimization methods that enhance the efficient selection of structures from an extensive pool of candidates[17,18], have all proven effective in this context.

[1]National Laboratory of Solid State Microstructures, School of Physics and Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing, China. [2]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany. [3]Fisika Aplikatua Saila, Gipuzkoako Ingeniaritza Eskola, University of the Basque Country (UPV/EHU), Donostia/San Sebastián, Spain. [4]Centro de Física de Materiales (CFM-MPC), CSIC-UPV/EHU, Donostia/San Sebastián, Spain. [5]These authors contributed equally: Yu Han, Chi Ding. ✉e-mail: wangjunjie@nju.edu.cn; gaaooh@126.com; jiansun@nju.edu.cn

https://doi.org/10.1038/s43588-025-00775-z

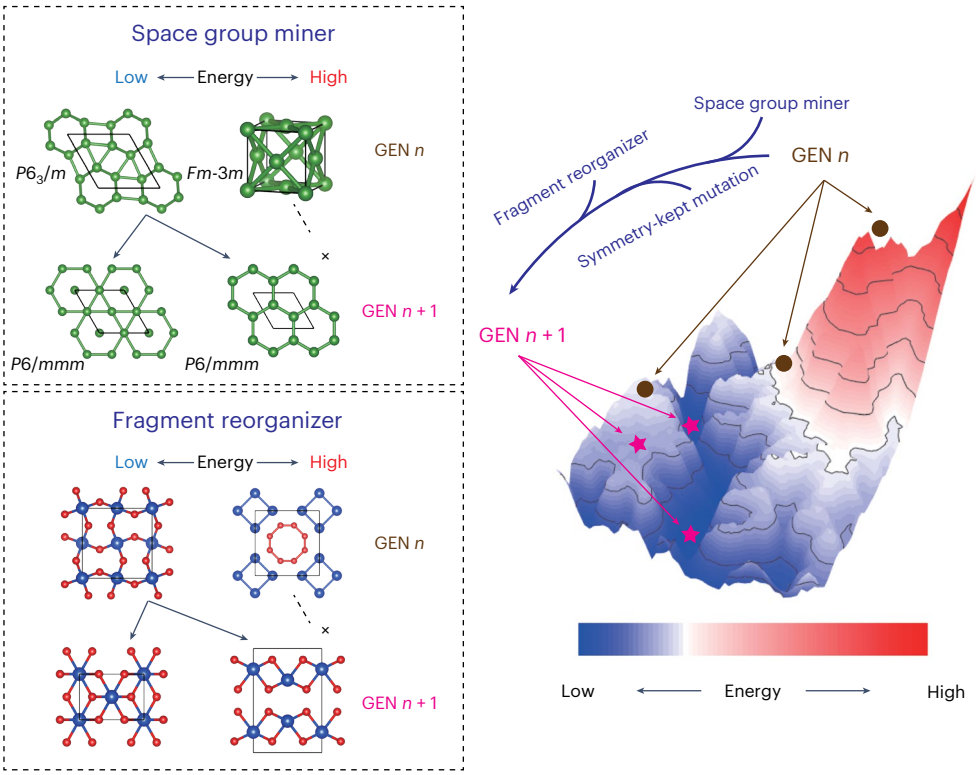


Fig. 1 | Illustration of the concept of the symmetry-principle-guided evolutionary structure generator. The generator integrates a space group miner, which extracts crystal symmetry components from lower-energy structures and incorporates them into offspring, and a fragment reorganizer, which captures local atomic environment features and transfers them to offspring. Augmented by symmetry-kept mutation, it improves the efficiency of PES exploration by focusing sampling in the lower-energy region. 'GEN n' represents any iteration of the evolutionary algorithm, and 'GEN $n+1$' represents the subsequent iteration, highlighting the direction of evolution. The shown PES is from the Matplotlib public dataset[92].

The imposition of physically motivated constraints to prevent the sampling of unphysical and high-energy configurations represents another efficient method for enhancing CSP by reducing the search space. Previous studies have demonstrated the efficacy of straightforward random searching when combined with considerations such as atomic distances or coordination numbers[3]. In addition, there are highly physically motivated molecular-dynamics-based CSP methods such as basin hopping[19] and minima hopping[20], which can effectively steer clear of high-energy configurational regions, albeit at the expense of requiring numerous energy evaluations to escape from minima. Notably, the incorporation of crystal symmetry has been demonstrated as an effective constraint in various CSPs[3,5,6,21–26]. This approach is rooted in observations that stable crystals typically show high symmetries, and among experimentally determined crystal structures, only a tiny fraction possesses the $P1$ space group[27]. However, many existing methods have halted at the superficial level of randomly generating symmetric configurations or maintaining symmetry during the search process. On a deeper level, the symmetry principle[28,29], which offers a robust and general tool, has not been fully leveraged.

In this article, we present an evolutionary structure searching method biased by the symmetry principle, emphasizing not only the importance of symmetry in crystals but also the intricate relationships among them. Furthermore, the limitation of the symmetry principle also underscores the importance of local atom aggregates, prompting the development of a scheme to identify local fragments. Our approach, as illustrated in Fig. 1, involves a structure generator designed to automatically identify both global symmetry and local fragment features within the preceding structure population. This is achieved through combining a space group miner founded on group–subgroup relationships and a fragment reorganizer assisted by graph theory. The generator, complemented by symmetry-kept mutation, serves to substantially improve the quality of trial structures and focus sampling on the lower-energy regions of the PES. Notably, our method shows remarkable benchmark improvements across three distinct testing systems. This advancement is integrated into the latest iteration (version 2.0) of our evolutionary-algorithm-based CSP framework MAGUS (Machine Learning and Graph Theory Assisted Universal Structure Searcher)[6,30–32]. MAGUS has shown efficacy across diverse systems, yielding the discovery of numerous structures[33–37]. Furthermore, we conducted a global structure search within the violet phosphorus and Si (111)-(7 × 7) surface systems, which are known for their complex PES and structural building blocks. The outcomes affirm the enhanced search capabilities of MAGUS.

Results

Symmetry principle-biased evolutionary structure generator

The symmetry principle summarized in refs. 28,29 reveals general relationships in crystal chemistry, highlighting the common presence of symmetry in crystal structures and the group–subgroup connections between their different space groups (see 'The symmetry principle in CSP' in Methods). This observation inspired a biased sampling strategy, increasing the chance of sampling the same space group and supergroups of lower-energy configurations during the CSP process (see 'Symmetry-kept mutation and space group miner' in Methods).

Figure 2 illustrates a representative trajectory of how MAGUS identifies the global-minima phase for $MgAl_2O_4$ accelerated by a space group miner and symmetry-kept mutation. Supposing the preferred space groups are $P2_1(4)$, $C2(5)$, $Pc(7)$ and $Cc(9)$ (as derived in Extended Data Fig. 1a,b), the resultant probabilities for selecting each space group by space group miner is shown in Fig. 2a. Figure 2b(i) is an initially randomly generated structure, with atomic coordinates assigned randomly based on the mined space group symmetry $Ccc2(37)$. This space

https://doi.org/10.1038/s43588-025-00775-z

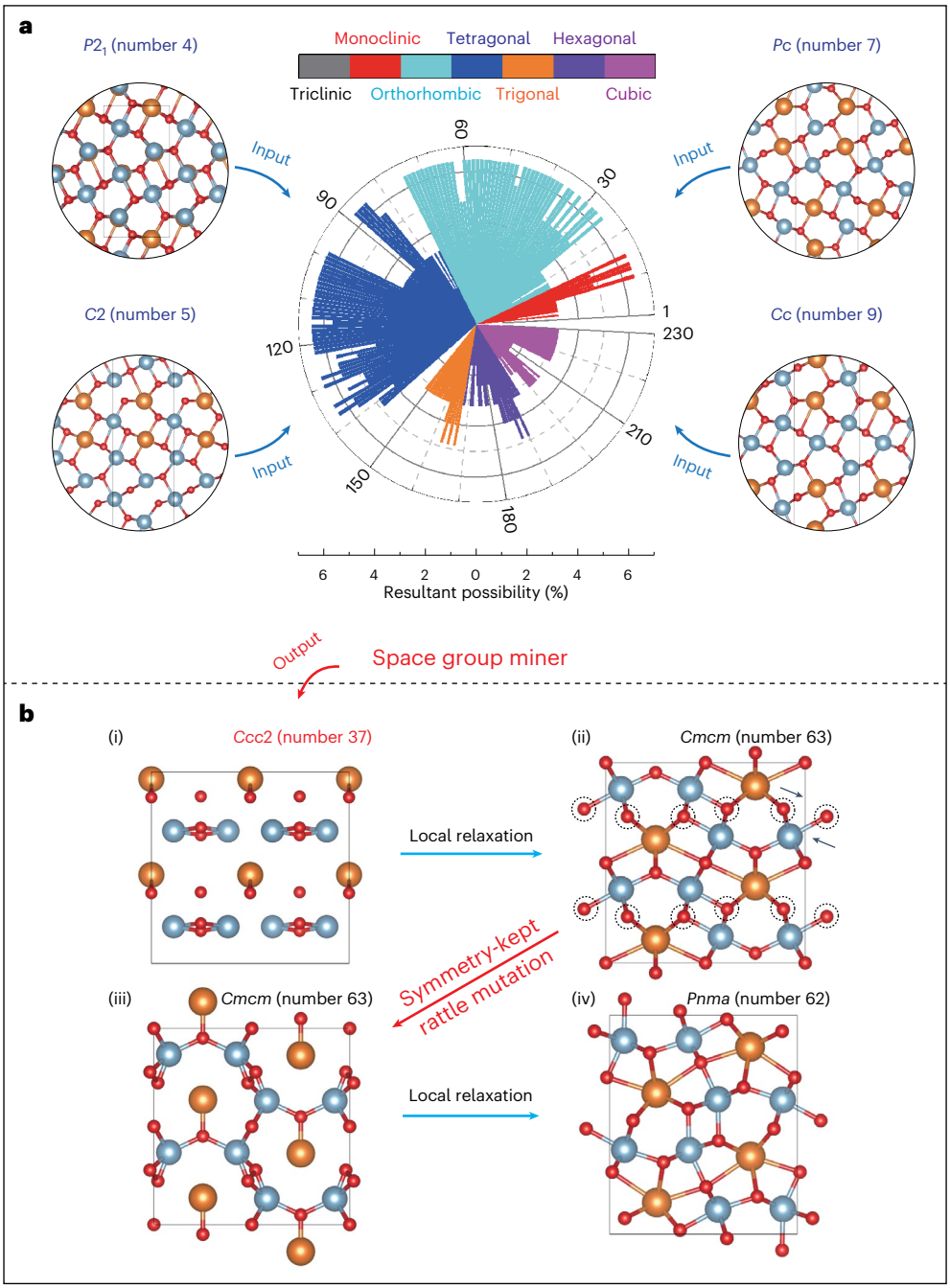


Fig. 2 | Illustration of how MAGUS found the global minima for $MgAl_2O_4$, highlighting the usage of a space group miner and symmetry-kept rattle mutation. a, The resultant probabilities for selecting each space group, assuming the preferred space groups are 4, 5, 7 and 9. **b**, One space group selected by the space group miner, *Ccc*2, along with an initial random structure (i) generated within this symmetry. The structure in (ii) is the local relaxed phase of (i), with *Cmcm* symmetry. The structure in (iii) is the offspring produced by the symmetry-kept rattle mutation strategy applied to (ii), maintaining the parent's symmetry *Cmcm*. The rattled atoms are marked with dashed-line circles in (ii) and are moved in the directions indicated by the arrows. The structure in (iv) is its local relaxed phase *Pnma*, which is the global minima. The red atoms represent O, the orange atoms represent Mg and the cyan atoms represent Al.

group is a supergroup of the above preferred space groups 4, 5, 7 and 9, mining into symmetry components of a base-centered *C* cell type, glide planes and a 2-axes rotational symmetry. Following a local relaxation process, it transforms into a metastable *Cmcm* phase, as shown in Fig. 2b(ii). The base-centered *C* cell type remains unchanged during the local relaxation, while an additional mirror symmetry is induced, resulting in a transformation of the corresponding point group *mm*2 into its supergroup *mmm*. One of the two symmetry components glide planes *c* is disrupted, while the other remains preserved. Figure 2b(iii) shows an offspring structure generated by symmetry-kept rattle mutation applied to the structure in Fig. 2b(ii). This offspring structure maintains the *Cmcm* space group symmetry of its parent but alters the position basis for O atoms, marked with dashed-line circles. Following a local relaxation, it evolves into the ground state *Pnma* (Fig. 2b(iv)), signifying the success of the search process. This evolution path requires a minimal number of structures and markedly reduces the requisite quantity of structures when searching for global minima, as discussed in the 'Benchmark' section.

https://doi.org/10.1038/s43588-025-00775-z

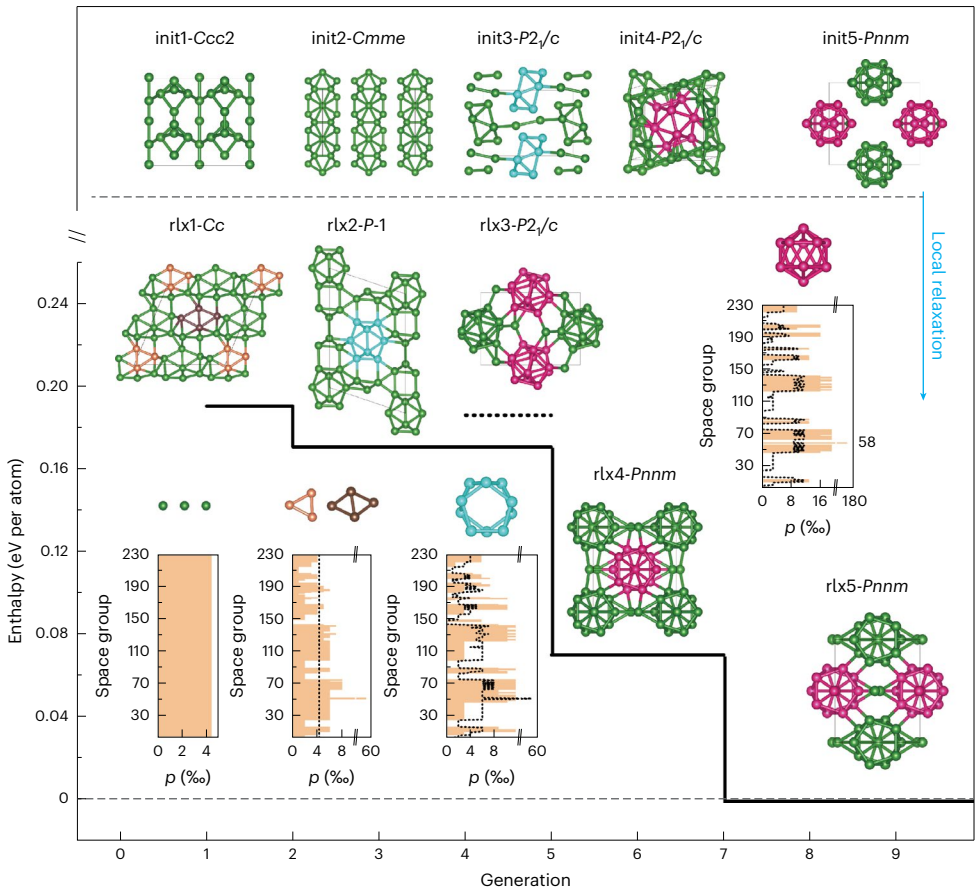


Fig. 3 | Illustration of one trajectory of MAGUS's global search into the γ-B system with an on-the-fly fragments reorganizer and a space group miner. Different fragments (isolated or contained within crystal structures) are highlighted in distinct colors, showing the evolution from isolate (green) → B_3 (orange), B_4 (brown) → B_{10} (blue) → B_{12} (magenta). The structures before and after local relaxation are denoted accordingly (init *n* → rlx *n*). The insets show the probability (*p*) of space group selection for generations 1, 2, 5 and 7 as suggested by the space group miner. The dashed lines are included to guide the eyes for comparison with the distribution of the previous inset.

The tendency of crystal configurations toward symmetry arises from the limited occurrence of energetically favorable environments. To identify these favorable local features, we introduce a structure decomposition scheme (see 'Graph-theory-based structure decomposition' in Methods) and integrate the resulting fragments into offspring structures using a fragment reorganizer (see 'Symmetry-constrained fragments reorganizer' in Methods). The space group symmetry for this process is suggested by the on-the-fly space group miner.

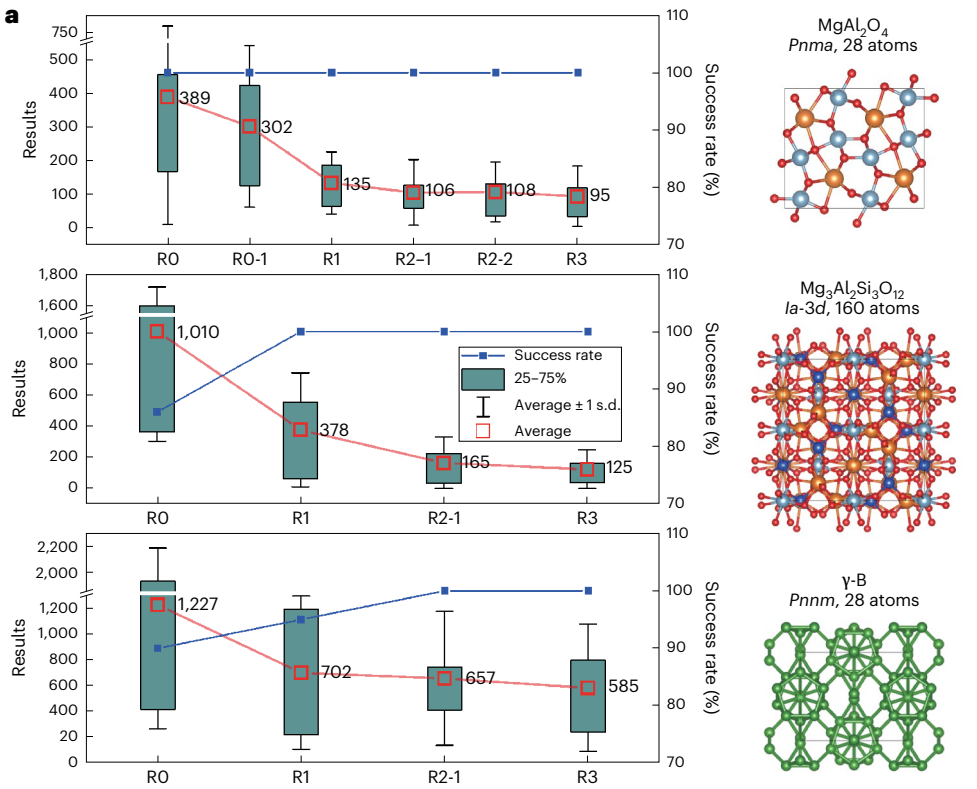
Figure 3 illustrates a trajectory of MAGUS's global search into the γ-boron (γ-B) system comprising 28 atoms[38], which is a widely employed and challenging benchmark system for CSP[21,31,39,40]. In this chosen trajectory, the program is initialized with a uniform distribution of space groups as well as 28 isolate B atoms. In the first generation, the best structure was identified as a *Cc* metastable phase (denoted as rlx1), introducing B_3 triangle fragments (in orange) and a preference bias towards symmetries containing glide planes. In the second generation, a *P*-1 metastable phase (denoted as rlx2) emerged, introducing a double-layered pentagon pattern B_{10} fragment (in blue) into the fragment pool, which is a subset of B_{12} icosahedra. Note that the fragment marked in blue is not a complete icosahedron but appears similar from the selected perspective. In the fourth generation, a $P2_1/c$ initial structure reorganizing on B_{10} (init3) relaxed into a metastable $P2_1/c$ phase (rlx3), and contributed the B_{12} icosahedra fragment pattern (in magenta), identical in shape to B_{12} from α-B in Extended Data Fig. 2a. In the fifth generation, a metastable *Pnnm* phase (rlx4) was discovered by relaxing on another $P2_1/c$ initial structure reorganizing on B_{12} icosahedra (init4), introducing a preference for the specific symmetry *Pnnm* (number 58) and the icosahedra fragment, which can be readily reorganized into the global-minima phase (init5). Note that the trajectory of an evolutionary-algorithm program is inherently stochastic. In practice, a specific low-energy structure can be reached through diverse trajectories, including rattle mutation of other parents, and random initial symmetry structures composed of only isolated atoms or various combinations of different fragments with varying space group symmetries. For an idealized illustration of our concept, the depicted path is purposely selected and low-energy structures obtained through unrelated trajectories were excluded.

The collaborative efficacy of the space group miner and boron fragments, such as triangles, pentagons and icosahedra, in substantially reducing the search space is evident, with further discussion provided in the 'Benchmark' section.

Benchmark

To validate the effectiveness of our methods, we selected three classical systems: $MgAl_2O_4$ under 100 GPa, garnet pyrope $Mg_3Al_2Si_3O_{12}$, and boron under 60 GPa, which have been previously employed as benchmarks in numerous CSP software[6,13,21,25,31,39–41]. Supplementary Section IA provides a detailed comparison of different CSP methods, emphasizing the advantages of our approach. The methods employed for each group are shown in Fig. 4b and the corresponding results are shown in Fig. 4a. In later discussion, we focus on the average number of structures to get the global minima of each group. All tests among

https://doi.org/10.1038/s43588-025-00775-z



b

Methods

	Random ratio	Symmetry-kept mutation	Space group miner	Fragments reorganizer
R0	0.3	×	×	×
R0-1	0.7	×	×	×
R1	0.7	√	×	×
R2-1	0.7	√	√	×
R2-2	0.7	√	×	√
R3	0.7	√	√	√

Fig. 4 | Benchmark comparison for different testing systems. a,**b**, The results obtained (**a**) using different methods (**b**) to isolate and identify their effects. The numbers represent the average number of structures needed to find the ground state. Group R0 represents the group with a random ratio of 0.3 and no additional techniques. Group R0-1 uses a random ratio of 0.7 without any additional techniques. Group R1 introduces the symmetry-kept mutation technique to group R0-1. Group R2-1 adds the space group miner to group R1, while group R2-2 adds the fragments reorganizer into group R1. Group R3 activates all techniques. Statistics for the Mg–Al(–Si)–O systems are based on 50 independent runs, and 20 for the boron system, to mitigate the influence of randomness. The box bounds in the plots represent the 25th to 75th percentiles of the results, with the whiskers showing the average value ± s.d. and the red square marking the average value. Red ticks indicate that a technique is enabled, while blue crosses indicate that a technique is disabled. The red atoms represent O, orange atoms represent Mg, cyan atoms represent Al and blue atoms represent Si.

all the three systems are conducted with the same parameter settings detailed in Supplementary Section IIA.

We performed six groups of independent tests on the $MgAl_2O_4$ system to systematically illustrate the impact of various adaptations to the previous program. The first group sets the ratio of random structures to 0.3 for each new generation, akin to the conventional evolutionary-algorithm approach[21,41], serving as the baseline R0. Notably, a performance boost of 22% was achieved by increasing the random ratio to 0.7. Furthermore, an additional improvement of 55% was achieved by incorporating the symmetry-kept evolutionary-algorithm operation. This adjustment ensures a more symmetric population, establishing the baseline R1 and underscoring the crucial role of symmetry in evolutionary-algorithm-based CSP performance. Our evolutionary generator, featuring the space group miner and fragments reorganizer, independently demonstrated an efficacy enhancement of approximately 20% compared with the R1 baseline. Collectively, the amalgamation of all employed methods resulted in a remarkable fourfold enhancement in performance compared with the traditional evolutionary-algorithm baseline. This outcome (95) also surpasses the results reported in previous studies[25,31,41] for the same system, which yielded results ranging from 358 to 378. The numbers represent the average number of structures needed to find the ground state. Fewer structures means higher efficiency.

The $Mg_3Al_2Si_3O_{12}$ system, consisting of 160 atoms, serves as a widely employed benchmarking system representative of larger systems. We found that the ground state was easily obtained by analyzing symmetry features from the population. Notably, employing the space group miner, the average probability of selecting the target symmetry *Ia*-3*d* in the final generation for groups R2-1 and R3 is 3.487% and 2.909%, respectively. This represents an approximately sevenfold increase compared with randomly choosing from the entire space group list 2–230 (0.437%). In addition, 48 successful attempts in R3 were accomplished by directly performing local relaxation on random initial structures. This result contrasts with the $MgAl_2O_4$ system, where 42 successful attempts in R3 were accomplished by symmetry-kept-mutation-generated structures. This finding is consistent with the previous observation that acquiring high-symmetry structures through mutation poses substantial challenges, underscoring the necessity of mining towards supergroups to enhance the search process.

Regarding the structure of γ-B[38], employing an on-the-fly space group miner and a fragments reorganizer, an enhancement of 58% is observed compared with the baseline R0. Notably, in group R3, five successful attempts (25%) were accomplished by directly performing local relaxation on random initial structures composed of reorganized fragments. Furthermore, the fragment reorganizer shows noteworthy performance in the red phosphorus and the Si (111)-(7 × 7) surface systems, as detailed below.

Fibrous phosphorus and violet phosphorus

Elemental phosphorus is notorious for its numerous allotropes, including red phosphorus phases type I–V[42]. Here we demonstrate the searching capability for this challenging system, specifically targeting type IV fibrous red phosphorus[43], shown in Extended Data Fig. 3a, and type V Hittorf's phosphorus[44,45], depicted in Extended Data Fig. 3b. This system poses substantial challenges for CSP; previous reports[16,46] have indicated that discovering fibrous P through random searches was impractical. However, by employing fragment reorganization and symmetry-kept rattle mutations, our method successfully identified the fibrous and violet P structure.

With the same constraint of symmetry as previous reports[16], we conducted a set of 47 independent runs, each generating up to 10,100 structures at the maximum generation. Detailed settings for these runs are available in Supplementary Section IIB. The success rate is illustrated in Extended Data Fig. 3a, revealing that the number of structures generated to locate fibrous phosphorus and fibrous phosphorene[47] was 158,000 and 67,800, respectively. In addition, we evaluated success rates relative to other reference structures exhibiting features of experimental structures[47–50]. A comprehensive analysis of these results, along with comparisons with previous methods, is available in Supplementary Section IB.

For violet P, a trajectory of MAGUS's global search to identify violet P is depicted in Fig. 5, showcasing the utilization of fragment reorganization and symmetry-kept rattle mutation. As previously mentioned, directly obtaining the target violet P structure configuration from randomly generated symmetric structures is challenging. However, features resembling P tubes composed of pentagons were observed in some of them, such as rlx1 in Fig. 5. From the decomposition of rlx1, four fragments (P5, P6, P7 and P8) were obtained, and three of them that formed init2 are depicted in Fig. 5 by different colors. Particularly noteworthy is the similarity between the P8 fragment and the P8 cage fragment decomposed from fibrous P in Extended Data Fig. 2d. These fragments were subsequently reorganized and further relaxed into rlx2. Following a series of symmetry-kept rattle mutation and subsequent relaxations, the main feature of violet P, consisting of two perpendicular P tubes, emerged in rlx12, and then adjusted through subsequent symmetry-kept rattle mutations in rlx13, achieving the configuration of violet P. The entire trajectory is shown in Supplementary Fig. 1a. This trajectory also bears some resemblance to the fragment-based approach in ref. 16, but notably, we obtained the P-cage fragments from the structure pool built from scratch during program execution. This distinction potentially highlights the advantage of an on-the-fly decomposition and reorganization architecture. Supplementary benchmarks in Extended Data Fig. 3b indicate that the number of structures generated to locate violet phosphorus (phosphorene[51,52]) is 52,000 (20,800). A more detailed discussion, along with predictions of additional distinct structures, is provided in Supplementary Section IB.

Si (111)-(7 × 7) surface

The intricate surface reconstruction model of Si (111)-(7 × 7) is widely believed to adhere to a dimer–adatom–stacking fault (DAS) structure configuration[53], shown in Fig. 6a. Given its intricate nature, various attempts[14,23,26,54,55] have been undertaken to explore the system, detailed in Supplementary Section IC.

We performed crystal structure searches involving 96, 100, 102, 104 and 108 atoms within the surface region, comparing the fitness of structures based on their surface energy. The computational details for surface energy are available in Supplementary Section IIC. We successfully found that the DAS structural model was the lowest-surface-energy model. The corresponding initial random structure, which can be relaxed into DAS form, is shown in Fig. 6b, generated by the shown internal generation parameters, comprises 12 duplicates of a 5-atom fragment (indicated in brown), 6 duplicates of a 4-atom fragment (indicated in green) and 18 single atoms, totaling 102 atoms. While the initial random structure is not stable, it aligns with the conceptualization of the DAS model, where 4-atom fragments represent sp^2 bonds, and 5-atom fragments symbolize T4-site adsorptions, arranging in *p6mm* symmetry. Furthermore, we identified other metastable structures with surface energy less than DAS + 18 meV $Å^{-2}$, each characterized by unique features, and their corresponding surface energies are presented in Extended Data Fig. 4 and Supplementary Section IID. These structures show diverse building blocks such as dimers, stacking faults, different adatoms sites and vacancies. The total count of 42 structures underscores the efficacy of our approach in generating low-energy configurations.

Detailed information regarding the surface reconstruction prediction module via MAGUS is discussed in ref. 32, but traditional evolutionary algorithms struggle with DAS prediction. Our method efficiently navigates this complex space, as detailed in Supplementary Section IC. The impact of the fragment reorganizer on the search process is evident, as illustrated in the supplementary benchmark tests provided below. These tests aim to assess how different fragments influence the success rate of constructing the DAS model. To simplify the assessment, we compared the summary of distances (D), total number of unaccepted atoms (n) and relative energy (E, in eV per 7 × 7 cell) relative to DAS for the generated structures, as detailed in Supplementary Section IC. Using the fragment reorganizer, 20,000 structures were generated across 4 test groups: isolated atoms, 4-atom fragments, T4 adsorptions and T4 adsorptions with sp^2 fragments, with settings detailed in Supplementary Section IC. The success rates are shown in Fig. 6d and the fragments are shown in the insets. The results show that the fragment reorganizer can substantially enhance the performance of the structure generator, achieving 1 exact DAS structure per 5,000 attempts using T4 and sp^2 fragments. In addition, generating one potential parent for further symmetry-kept rattle mutation ($D < 100$ Å) requires, on average, only 833 structures.

Further demonstrations

In the sections above, we primarily demonstrated constrained atom number searches to ensure a fair comparison with existing methods by maintaining the same restrictions, and to reduce computational costs for statistical analysis. However, in real-world scenarios, such

https://doi.org/10.1038/s43588-025-00775-z

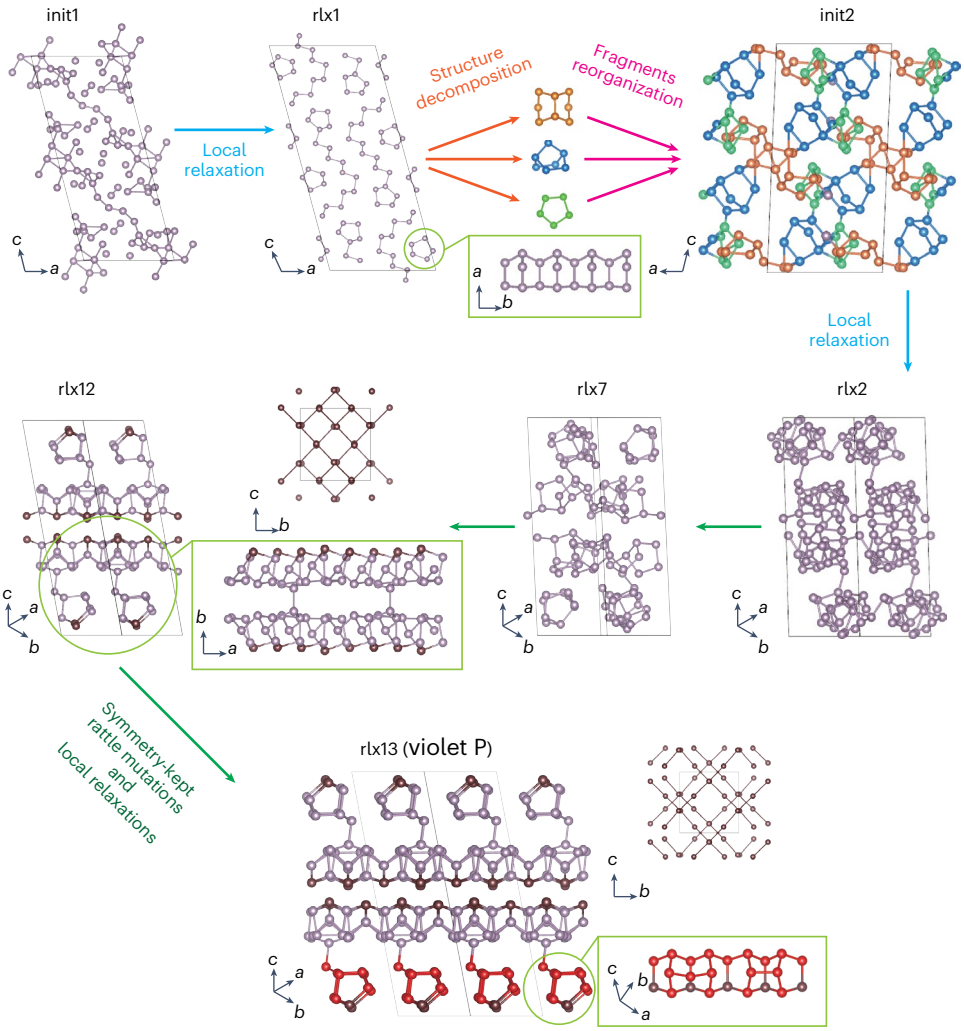


Fig. 5 | A trajectory of MAGUS's global search to identify violet P. In this trajectory, a randomly generated symmetric structure, init1, is relaxed into rlx1, which features P tubes with pentagonal cross-sections. From the decomposition of rlx1, fragments were obtained, shown in different colors, and these fragments were reorganized and further relaxed into rlx2. Through a series of symmetry-kept rattle mutations and subsequent relaxations, the configuration of violet P was achieved in rlx13. The insets enclosed within green frames highlight specific parts of the structure from an alternative perspective. Different atom colors are used to emphasize particular regions of the structure.

parameters are often unknown. It is important to note that our method is fully capable of handling variable atom number (or composition) searches. In Supplementary Section III, we provide 2 additional examples: boron with variable atom numbers at 60 GPa, and borophene with variable atom numbers. Moreover, we showcase a binary material surface search for the GaAs (100) ζ(4 × 2) surface system, which features distinct subsurface dimerization. The unified adaptation of our symmetry principle-based acceleration method across various systems—including crystal structures, two-dimensional materials and surface systems—demonstrates the flexibility and advancements of our approach.

Furthermore, MAGUS offers interfaces to various MLPs, such as GAP[7], Deep Potential[8,56], MTP[9,57], NEP[10,58], MACE[59], CHGNET[60] and HOTPP[11], some of which have already been utilized in our examples, providing additional avenues for CSP acceleration.

Discussion

In this section, we discuss the relations between our method and other CSP works. Our CSP scheme can be viewed as a variant of evolutionary-algorithm CSP, but the usage of symmetry-kept mutation, a space group miner and a fragment reorganizer overcome two major disadvantages of traditional evolutionary-algorithm CSP, corresponding to global and local features of crystals, respectively.

First, in our current method, all unrelaxed structures are generated through symmetry-kept mutation or a random symmetric structure generator, thus preventing configurations with *P*1 symmetry. Traditional evolutionary operators typically fail to preserve the global symmetries of parent low-energy structures, often leading to structures with *P*1 symmetry. As stated by the symmetry principle, atoms reveal a pronounced tendency toward the high symmetry, thus these operators can hinder the searching for stable structures. A recent study, driven by similar motivations and proposed in ref. 25, employs a symmetry-oriented divide-and-conquer method for a systematic integration of symmetry into the system to enhance search performance. This approach successfully predicted structurally complex systems of binary Lennard-Jones mixtures and ice. However, this method ignores the inner transformations of group–subgroup relations between space groups. Group–subgroup relations were utilized for CSP in ref. 41, which implemented a symmetry group decomposition scheme that obtains a family of subgroup from the space group for structure generation. This approach dramatically increased the effectiveness of USPEX method by almost three times. Another phase predication framework

https://doi.org/10.1038/s43588-025-00775-z

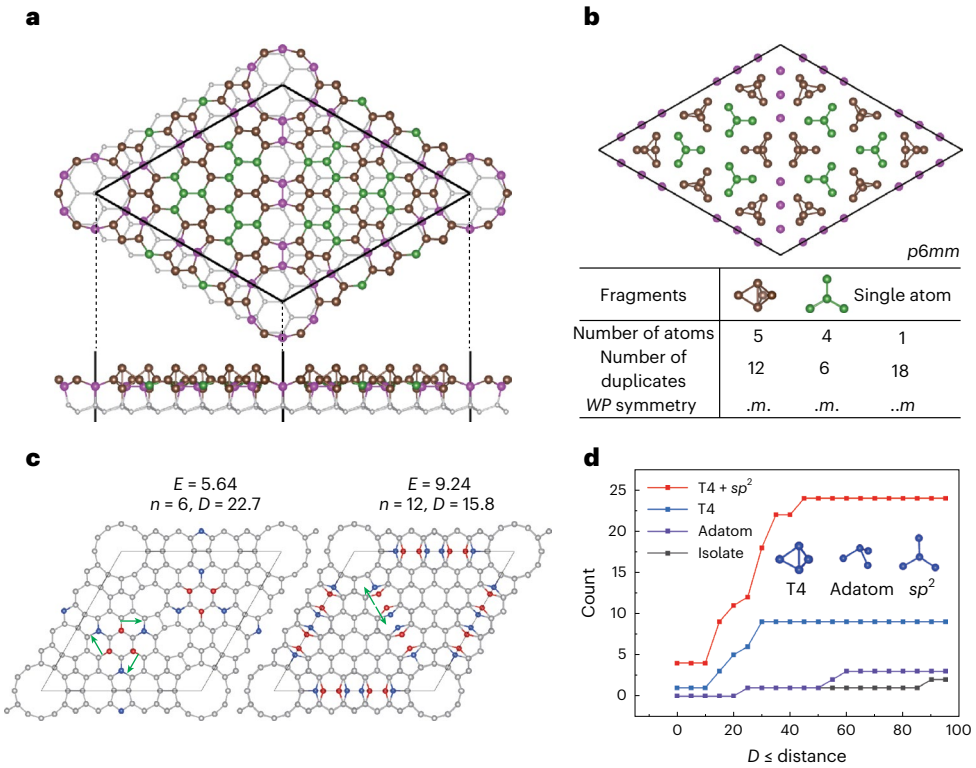


Fig. 6 | Illustration of the target structural model, corresponding initial structure and benchmark results of the Si (111)-(7 × 7) surface system. a, The structure of the Si (111)-(7 × 7) DAS model. Different colors of atoms mark the fragment they belong to. **b**, The initial random structure that could turn into a DAS structure by local relaxation and the corresponding inner generation parameters. **c**, In benchmarking tests for constructing a DAS model, two of successful attempts with their corresponding summary of distances (*D*), total number of unaccepted atoms (*n*) and relative energy (*E*, in eV per 7 × 7 cell) relative to DAS. More structures are shown in Supplementary Fig. 3. **d**, Success count of the fragment reorganizer within 20,000 attempts using different fragments, including isolated atoms, fragments resembling adatoms, T4 fragments and T4 fragments combined with sp^2 fragments.

employing group–subgroup transformation was proposed in ref. 61, which successfully gives ordered ground states of different systems including Li_xCoO_2, Li_xC_6 and $Li_{3x}La_{2/3-x}TiO_3$. However, these methods require prior databases or prototypes and lack the incorporation of information from previous generations.

Second, our structure decomposition method identifies low-energy fragments, and the reorganizer maintains these local features during the search. The generation of structures based on fragments from known structures has proved to be an effective approach in various systems[62–64]. Although traditional real-space crossover and mutation operators have shown success in retaining low-energy local fragments for effective structure searching[2,65,66], they may disrupt local low-energy fragments because these heredity operators are not explicitly designed for their identification. To address this limitation, community detection algorithms-adapted operators have been proposed[31,63,67]. One study[68] employed a structure decomposition approach utilizing community detection method at different resolution levels within the framework of AIRSS, which successfully predicted dense boron[63]. In our previous work[31,67], we introduced two structure decomposition schemes based on graph theory and took the acquired decomposed fragments as a whole in the evolutionary process, which markedly enhanced the performance of MAGUS. These methods are based on decomposition algorithms such as the Girvan–Newman algorithm[69] or Louvain algorithm[70] to find the best solution to the optimization problem of 'modularity' of a crystal graph. However, these algorithms encounter limitations in distinguishing overlapping communities—a common occurrence in crystal structure graphs. For instance, a single atom (node) in graphene is affiliated with three adjacent hexagons, a scenario not accurately identified by prior decomposition schemes. Diverging from the conventional 'modularity'-based decomposition method, the adopted bottom-to-top approach can avoid missing overlap communities during detection. While numerous algorithms, such as the clique percolation method[71], have been specifically designed for detecting overlapping communities in networks, crystal graphs may not always be suitable for treatment as community-based networks. In cases where nodes exhibit equivalent connectivity, as in graphene and diamond, the traditional notion of well-defined communities does not apply, rendering approaches such as the clique percolation method ineffective. However, recognizing the inherent characteristics of crystal structures, such as sp^2/sp^3 hybrid orbitals and hexagonal arrangements, we chose not to employ such overlapping community detection algorithms but rather our current scheme. Moreover, in contrast to concentrating solely on the optimal division of the entire network, a limitation often encountered due to the dependency of resulting communities on distant portions of the network structure[72], the two local indicators we employ focus exclusively on nodes within the neighborhood. Consequently, the extracted local features persist, whether the crystal structure represents a supercell, exhibits point defects or pertains to various structural forms, including three-dimensional periodic systems (bulks), surfaces, nanotubes and clusters.

Most of the previous CSP works mentioned above mainly focus on the transfer of solely global (space group) or local (fragment) features from good candidates to newly generated structures, without considering both types of feature systematically. In our proposed evolutionary framework, we identify both global and local features of low-energy structures from previous iterations. The extracted local fragments serve as fundamental building blocks to generate new candidates, guided by global symmetries derived from elitist structures.

Furthermore, we preserve global symmetry through symmetry-kept mutation during the heredity process. This comprehensive approach substantially enhances CSP performance, as evidenced by the benchmark results. In addition, while most of these aforementioned acceleration methods[25,41,63] mainly focus on searching for bulk systems, and their adaptations to surface systems are not very straightforward, our method provides a unified approach for CSP acceleration across crystal structure searches, surface reconstruction and two-dimensional systems.

The versatility of our approach, demonstrated across a range of systems, highlights the advancements of combining the symmetry principle with structure prediction in a unified framework. While the method may encounter limitations in certain borderline cases, particularly in low-symmetry or disordered systems, or face challenges such as being misdirected or trapped during the evolutionary process, we still hope that these findings will inspire further exploration of symmetry-based strategies in materials discovery pipelines.

Methods

The symmetry principle in CSP

The symmetry principle summarized in refs. 28,29 reveals general relations in crystal chemistry, which contains three aspects.

(1) In crystals, the arrangement of atoms reveals a pronounced tendency toward the highest possible symmetry.
(2) Counteracting factors may prevent the attainment of the highest possible symmetry, but in many cases the deviations from ideal symmetry are only small, and frequently the symmetry reduction corresponds to the smallest possible step.
(3) During solid-state reactions resulting in products of lower symmetry, the higher symmetry of the starting material is often indirectly preserved by the formation of oriented domains.

Here, aspect 1 reveals the preference for symmetry in crystals and there is the physical reason behind the aspect[29,73]: under certain conditions, there is one most stable neighboring environment for atoms with the same species. These atoms tend to attain this environment and the equivalence naturally leads to the symmetry in crystals.

In real systems, complicated interactions, such as covalent bonds, lone electron pairs, the Jahn–Teller effect and so on, can lead to deviations from the highest possible symmetry, as revealed by aspect 2. The competitions between these counteracting factors and the tendency toward the highest possible symmetry induce generally observed group–subgroup relations between space groups in crystals, are exemplified by refs. 28,29. In addition, here we will not discuss aspect 3 as this work is not related to reactions.

Aspect 1 has been widely used in CSP and previous works indicate that the incorporation of symmetry in structure generation or evolution can increase the possibility to hit the global minima[3,5,6,21–25]. But the group–subgroup relations in aspect 2 are rarely studied in CSP. To test the capability of the symmetry principle in CSP, we select the $MgAl_2O_4$ system comprising 28 atoms[74,75], which is a widely employed benchmark system for CSPs[21,25,31,41]. The observations are detailed below.

A typical CSP starts with the generation of random initial structures based on space group symmetry. Following this procedure, we generated 20 random $MgAl_2O_4$ structures for each specific initial space group, amounting to a total of 3,580 structures (fewer than 229 × 20 for some specific space group is incompatible with the composition). Given that these randomly generated atomic coordinates do not inherently satisfy force balance conditions, local relaxations were performed. The distribution of space groups and the corresponding enthalpies of the final states are shown in Extended Data Fig. 1a. Approximately 46% (1,647 structures) fall within the *P*1 space group, which is a common occurrence in evolutionary-algorithm programs. This indicates that, although specific symmetry criteria are met, the generated structures may still lack physical viability, resulting in the inability to maintain their symmetry. Notably, among the remaining structures, an obvious preference is observed for space groups 2, 4, 5, 6, 7, 8, 9, 12, 14, 15, 57, 62, 63 and 166, each with counts exceeding 45 and average enthalpies below −23.10 eV per atom (0.30 eV per atom higher than global minima), lower than average enthalpies of structures with other space groups (−22.89 eV per atom).

Moreover, it has come to our attention that diverse systems show analogous symmetry preferences[76–78], consistent with the symmetry principle, which posits that crystal symmetries are intricately interconnected and related to a conceivable or actually existent higher-symmetry structure. Building on the insights from refs. 28,29, we analyzed group–subgroup relationships among the preferred space groups in Extended Data Fig. 1a.

Group–subgroup relationships among space groups arise from the composition of rotational and translational symmetries within the space group. The former includes 32 point groups, while the latter includes 16 types of glide planes and screw axes, thereby establishing a relationship between them. Further details are provided in Supplementary Section IVA. The relationships among the preferred space groups in Extended Data Fig. 1a are shown in Extended Data Fig. 1b. Connections between each pair of group–subgroup relations with index <4 are indicated by arrows. The preferred space groups show 6 distinct point group symmetries: −1, 2, *m*, 2/*m*, *mmm* and −3*m*. All space groups belong to the first 5 types except for space group 166. Among these, −1, 2 and *m* are subgroups of 2/*m*, and 2/*m* is a subgroup of *mmm*, with these relationships between point groups naturally extending to their corresponding space groups. In addition, glide planes (in space groups 7, 9, 14, 15, 57, 62 and 63) and lattice type *C* (in space groups 5, 8, 9, 12, 15 and 63) are also common symmetry components in the preferred space groups. These observations suggest that the $MgAl_2O_4$ (28 atoms) system shows a preference for mirror symmetry and +1/2 translations. These components align with the exact symmetry attributes of the ground-state symmetry *Pnma*. This observation inspires biased samplings of space groups relative to ones with low energies/enthalpies during the search process, thereby enhancing overall efficiency.

On the basis of the symmetry principle and these observations, we proposed symmetry-kept mutation and a space group miner in our updated CSP scheme.

Symmetry-kept mutation and space group miner

If one low-energy configuration with space group *g* is found during CSP, the group–subgroup relations (aspect 2) recommend following searching in the subgroups and supergroups of *g*. Traditional evolutionary operators usually lower the symmetry of parent structures to *P*1 and lead to inefficient searching in symmetric structures. Symmetry-kept rattle mutation has also been proposed to preserve the symmetry of parent structures during evolutionary-algorithm searching[79]. We have also implemented a symmetry-kept mutation operator in MAGUS (see details in Supplementary Section IVB) and find that this mutation allows for effective searching in subgroups of the space group of the parent structure, but it is not suitable to explore structures with higher symmetry (supergroups).

Extended Data Fig. 1c shows the space group distribution of the resultant structures obtained from applying symmetry-kept rattle mutation to a single metastable *Pnma* phase (with enthalpy of 107.86 meV per atom higher than the global minima) as the parent structure, generating 400 offspring. Approximately 90% (362) of the offspring reverted to the parent phase after local relaxation. Among the remaining 38 new structures, a majority showed lower symmetry. Notably, 63% of these newly generated structures showed subgroup symmetries of the parent phase, while only 3% evolved into supergroups. This observation suggests that a random distortion followed

by relaxation more easily and frequently leads a high-symmetry phase to an adjacent low-symmetry phase than vice versa.

These phenomena can be accounted by typical energy landscapes around high-symmetry phases and the simplest instance is the double-well model widely used for charge wave density and ferroelectric materials. As depicted in Extended Data Fig. 1d, high-symmetry phases are typically surrounded by multiple equivalent low-symmetry phases while the low-symmetry phase has only one path to the high-symmetry phase, the system at the high-symmetry phase local minima can easily drop into two low-symmetry phase local minima with a distortion along either negative or positive directions. But the proper direction from any minimum to the local maximum is only one.

High symmetry is usually relative to very low or high energy minima[80], aligning with the energy distributions observed in symmetric random structure generations[22,79]. If the global minima of the PES have relatively low symmetry, they can be easily reached from neighboring high-energy high-symmetry configurations by mutation. However, if the ground-truth structure in a system has high symmetry, evolutionary operators would like to be inefficient and random generation becomes the primary method to sample the high-symmetry global minima. This assumption is further supported by our benchmarks (see 'Benchmark' section). Moreover, it underscores the importance of manually introducing more supergroups to the current population symmetry.

On the basis of the symmetry principle and observations above, we propose 'space group miner' to bias the choice of space group adaptively during random structure generation based on previous searching results. The workflow of a space group miner is as follows. Initially, the space group symmetry of a specific structure is determined through spglib[81]. Subsequently, the space group miner assembles the space groups g of several best structures from the preceding population, which we take as the 'preferred structures'. The miner then modulates the possibility to choose space group g' to generate new structures using the expression:

$$f(g') = \sum_{g} f_1\delta_{\text{same}}(g,g') + f_2\delta_{\text{super}}(g,g') + f_3\delta_{\text{sub}}(g,g') + f_4\delta_{\text{other}}(g,g')$$

where δ_{same}, δ_{super}, δ_{sub} and δ_{other} equal 1 if g' is the same group, the supergroup, the subgroup and none of the above group relation to group g, and 0 if not. This formulation augments the likelihood of producing new structures sharing the same space group, along with its supergroups, as the preferred structures, while the parameters f_3 and f_4 are set to zero. Although increasing the probability of subgroups is theoretically valid, symmetry tends to break and descend to lower symmetry during relaxation and evolution. Therefore, we assume that configurations with subgroups are effectively explored through symmetry-kept mutations and relaxation processes, as discussed in Extended Data Fig. 1. Thus, a manual preference for subgroup symmetry is unnecessary, leading to $f_3 = 0$, which is further supported by supplementary benchmarking results in Supplementary Section IIIA. Users have the flexibility to define the values of f_1,f_2,f_3 and f_4.

Graph-theory-based structure decomposition

As discussed above, the tendency to the highest possible symmetry is due to the energetically most favorable environment of one element. Thus, identification and preservation of such low-energy environments are important in CSP. However, group–subgroup relations alone cannot identify these low-energy local features in crystals. Here we introduce a structure decomposition method based on both graph theory and group theory and combine it with a space group miner.

Converting a crystal into a graph is a straightforward process. Specifically, by representing each atom as a 'node' and each atomic bond as an 'edge', a crystal graph is formed. The concepts from graph theory finds widespread application, such as the analysis of crystal structure properties[67,82], serving as constraints for structure optimization[24], and more. Notably, the identification of community structures[83], wherein network nodes are tightly knit within groups with looser connections between them. In the context of crystal graphs, a community is likely to represent local bonding environments.

One the basis of previous literatures and our observations, we assume that preferred local fragments in crystals have high symmetry in both the real space and the abstract space of the atomic graph. To this end, we take use of two indicators to rank fragments.

The first indicator for symmetry in the real space is descriptor length inspired by ref. 63 related to the idea from algorithmic information theory. This study proposed that a community with n (larger than 2) symmetry-inequivalent atoms needs $3n - 6$ degrees of freedom to describe the structure of a fragment, and it is $3n - 5$ when there are only 2 symmetry-inequivalent atoms and 0 for a single inequivalent atom. However, this indicator alone cannot distinguish suitable fragments as it does not consider the connectivity between atoms. Thus, the second indicator, uniqueness of node betweenness centrality, is used.

The betweenness centrality[84,85] of a node is defined as the sum of the fraction of all pairs of shortest paths that pass through it. We have found that a preferred community within a crystal graph has a relatively higher uniformity of betweenness centrality. Taking the α-B structure shown in Extended Data Fig. 2a as an example, different fragments taken from it are illustrated in Extended Data Fig. 2f. A B_{12} icosahedra community (Extended Data Fig. 2f(iv)) shows identical betweenness centrality values for each node, resulting in a uniqueness value of one. This uniformity is disrupted by the attachment of any nearby boron atom (Extended Data Fig. 2f(i)–(iii)), subsequently increasing the uniqueness. This indicator tends to favor fragments that are uniformly connected, such as triangles, squares, pentagons, hexagons, tetrahedrons, octahedrons, cubic structures, icosahedra and so forth.

Notably, we did not employ the standard deviation of betweenness centrality as the indicator, despite its capacity to measure dataset uniformity, and it equals zero for the B_{12} icosahedra community, corresponding to first ranking. The reason is, in the case of graphene decomposition (shown in Extended Data Fig. 2e), the standard deviation of betweenness centrality for the 4-atom fragment ([3, 0, 0, 0]) is relatively high compared with other fragments, while its uniqueness ranking is much lower. This type of fragment, representing a center atom with its first neighbor, conveying information on its bonding environments with information on coordination number and bonding angles, is beneficial for structure search. Hence, uniqueness is adopted as a more suitable indicator for such cases.

Fragments with low values on the two indicators can be regarded as good 'genes' for CSP. However, it is impractical to sample all the possible combinations of atoms in crystals. Here we propose an effective iterative way to sample important fragments of a given crystal structure and the computational load can be further reduced by considering space group symmetries.

Our decomposition scheme takes four steps. (1) Identify every inequivalent atom within the structure. (2) Construct a 'neighborhood' structure centered around a chosen atom. (3) Iteratively eliminate atoms from the neighborhood structure and evaluate the 'ranking' of the resultant structure until only three atoms remain. Both steps (2) and (3) are carried out for each inequivalent atom in the lattice cell. (4) Select several fragments with the highest ranking. The detailed description is as follows.

Taking the α-B structure illustrated in Extended Data Fig. 2a as an example, we initially identify the inequivalent atoms within the structure through spglib[81], represented by solid line circles of distinct colors. Upon selecting an atom, a large supercell is constructed, and the chosen atom is positioned at the supercell's center. This step aims to prevent the omission of neighbors that extend beyond periodic boundary conditions. Subsequently, the local environment ('neighborhood') of the atom is built within a cut-off, as depicted in Extended Data Fig. 2b. The selected center atom is marked by a magenta circle and the cut-off is indicated by a magenta dotted circle. By default, the

cut-off is determined by a maximum distance of 5 Å, with the shortest path limited to less than 4 steps. The combination of limitations is adjustable by users and serves to balance the computational cost and the maximum size of the detected community.

Following the construction of the neighborhood structure, atoms are systematically eliminated through an iterative process based on the count of their edges, with those having fewer connections given higher priority for removal. This order is derived from the general principle of community detection[83], which aims to identify groups characterized by dense internal connections and sparser connections between groups. Such an approach ensures that the remaining portion of the structure exhibits a higher density of internal edges, potentially indicating a community structure. In cases where various choices are available, duplicates are generated, and all available choices are exhaustively explored. Then the ranking indicators for all fragment structures are calculated, where the count of inequivalent atoms within a fragment is found by pymatgen[86], and the betweenness centrality of each node is calculated by NetworkX[87].

The fragments exhibiting the highest rankings in the decomposition of α-B, graphene and fibrous red phosphorous[43] along with their corresponding indicators are illustrated in Extended Data Fig. 2a,d,e. In the case of boron fragments, the B_{12} icosahedra is accurately identified with the highest rank, characterized by betweenness centrality uniqueness of 1, and a description length of 0 for its singular symmetry-inequivalent atom. In addition, the B_3 triangle is also identified, as a subset of B_{12} icosahedra. For graphene fragments, the C_6 hexagon is distinguished, aligning with human intuition as a fundamental composition component. Similarly, a C_4 fragment featuring a central atom and its three sp^2-bonded neighbors is identified. For P fragments, pentagons, hexagons and three-coordinate atoms are identified. Notably, the widely known building block [P8] cage, with C_{2v} symmetry, featuring two side views as hexagons and two side views as pentagons, is also recognized. Making use of all these fragments, we can generate more reasonable structures to accelerate the structure search process.

The detection of overlapping communities can be computationally demanding in large networks, especially in real-world scenarios[88]. While our scheme is not exempt from such limitations, crystal graphs typically feature a lower node count, usually less than a hundred in a single cell, and an even smaller number of symmetry-inequivalent atoms, rendering the computational time manageable. For example, an α-B cell with only 36 atoms has merely 2 symmetry-inequivalent atoms, resulting in a substantial reduction in the calculation loop count by a factor of 18. In our current implementation, with a default 'neighborhood' structure cut-off, the process takes 1.53 s on an Intel(R) Xeon(R) Platinum 9242 CPU @ 2.30 GHz. The impact of the cut-off parameter is detailed in Extended Data Fig. 2c. An increased distance/step cut-off leads to a longer computational cost, and vice versa. The fragments in Extended Data Fig. 2a remain detectable as long as they are entirely contained within the neighborhood structure, and reducing the distance or steps too much, surpassing the yellow line mark, leads to the undetection of the B_{12} icosahedra. Besides, the time cost may increase if the crystal graph shows lower symmetry and containing too many symmetry-inequivalent atoms.

Symmetry-constrained fragments reorganizer

In our current scheme, fragments are treated as integral units for generating new crystal structures through a symmetry-constrained molecular structure generation scheme. To preserve the fragments while maintaining the target symmetry, the fragment's symmetry must align with the Wyckoff site symmetry. The program verifies this condition and positions the fragment into a valid Wyckoff site if satisfied. Further details are provided in Supplementary Section IVC and our earlier work[6]. Some other implementations for symmetric molecule crystal generation are also available[79,89,90].

Data availability

Source data for Figs. 2–4 and 6, Extended Data Figs. 1 and 3 are available with this paper. All data were generated using the MAGUS code (version 2.0) and are available from gitlab (https://gitlab.com/bigd4/magus) and on Zenodo at https://doi.org/10.5281/zenodo.14730874 (ref. 91).

Code availability

The MAGUS source code can be accessed from gitlab (https://gitlab.com/bigd4/magus) after registration (https://www.wjx.top/vm/m5eWS0X.aspx), or on Zenodo at https://doi.org/10.5281/zenodo.14730874 (ref. 91).

References

1. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
2. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX—evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
3. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).
4. Lonie, D. C. & Zurek, E. XtalOpt: an open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.* **182**, 372–387 (2011).
5. Wang, Y., Lv, J., Zhu, L. & Ma, Y. CALYPSO: a method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063–2070 (2012).
6. Wang, J. et al. MAGUS: machine learning and graph theory assisted universal structure searcher. *Natl Sci. Rev.* **10**, nwad128 (2023).
7. Bartok, A. P., Payne, M. C., Kondor, R. & Csanyi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
8. Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
9. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn. Sci. Technol.* **2**, 025002 (2021).
10. Fan, Z. et al. Neuroevolution machine learning potentials: combining high accuracy and low cost in atomistic simulations and application to heat transport. *Phys. Rev. B* **104**, 104309 (2021).
11. Wang, J. et al. E(n)-equivariant Cartesian tensor message passing interatomic potential. *Nat. Commun.* **15**, 7607 (2024).
12. Tong, Q., Xue, L., Lv, J., Wang, Y. & Ma, Y. Accelerating CALYPSO structure prediction by data-driven learning of a potential energy surface. *Faraday Discuss.* **211**, 31–43 (2018).
13. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V. & Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **99**, 064114 (2019).
14. Bisbo, M. K. & Hammer, B. Efficient global structure optimization with a machine-learned surrogate model. *Phys. Rev. Lett.* **124**, 086102 (2020).
15. Li, C.-N., Liang, H.-P., Zhang, X., Lin, Z. & Wei, S.-H. Graph deep learning accelerated efficient crystal structure search and feature extraction. *npj Comput. Mater.* **9**, 176 (2023).
16. Deringer, V. L., Proserpio, D. M., Csányi, G. & Pickard, C. J. Data-driven learning and prediction of inorganic crystal structures. *Faraday Discuss.* **211**, 45–59 (2018).
17. Yamashita, T. et al. Crystal structure prediction accelerated by Bayesian optimization. *Phys. Rev. Mater.* **2**, 013803 (2018).
18. Kusaba, A., Kangawa, Y., Kuboyama, T. & Oshiyama, A. Exploration of a large-scale reconstructed structure on GaN(0001) surface by Bayesian optimization. *Appl. Phys. Lett.* **120**, 021602 (2022).

19. Wales, D. J. & Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
20. Amsler, M. & Goedecker, S. Crystal structure prediction using the minima hopping method. *J. Chem. Phys.* **133**, 224104 (2010).
21. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
22. Avery, P. & Zurek, E. RandSpg: an open-source program for generating atomistic crystal structures with specific spacegroups. *Comput. Phys. Commun.* **213**, 208–216 (2017).
23. Wang, S.-W., Hsing, C.-R. & Wei, C.-M. Expedite random structure searching using objects from Wyckoff positions. *J. Chem. Phys.* **148**, 054101 (2018).
24. Shi, X., He, C., Pickard, C. J., Tang, C. & Zhong, J. Stochastic generation of complex crystal structures combining group and graph theory with application to carbon. *Phys. Rev. B* **97**, 014104 (2018).
25. Shao, X. et al. A symmetry-orientated divide-and-conquer method for crystal structure prediction. *J. Chem. Phys.* **156**, 014105 (2022).
26. Brix, F., Verner Christiansen, M.-P. & Hammer, B. Cascading symmetry constraint during machine learning-enabled structural search for sulfur-induced Cu(111)-(43×43) surface reconstruction. *J. Chem. Phys.* **160**, 174107 (2024).
27. Urusov, V. S. & Nadezhina, T. N. Frequency distribution and selection of space groups in inorganic crystal chemistry. *J. Struct. Chem.* **50**, 22–37 (2009).
28. Bärnighausen, H. Group–subgroup relations between space groups: a useful tool in crystal chemistry. *MATCH Commun. Math. Chem.* **9**, 139–175 (1980).
29. Müller, U. *Symmetry Relationships Between Crystal Structures: Applications of Crystallographic Group Theory in Crystal Chemistry* (OUP, 2013).
30. Xia, K. et al. A novel superhard tungsten nitride predicted by machine-learning accelerated crystal structure search. *Sci. Bull.* **63**, 817–824 (2018).
31. Gao, H., Wang, J., Han, Y. & Sun, J. Enhancing crystal structure prediction by decomposition and evolution schemes based on graph theory. *Fundam. Res.* **1**, 466–471 (2021).
32. Han, Y. et al. Prediction of surface reconstructions using MAGUS. *J. Chem. Phys.* **158**, 174109 (2023).
33. Liu, C. et al. Multiple superionic states in helium–water compounds,. *Nat. Phys.* **15**, 1065–1070 (2019).
34. Xia, K. et al. Predictions on high-power trivalent metal pentazolate salts. *J. Phys. Chem. Lett.* **10**, 6166–6173 (2019).
35. Gu, Q., Xing, D. & Sun, J. Superconducting single-layer T-graphene and novel synthesis routes. *Chin. Phys. Lett.* **36**, 097401 (2019).
36. Liu, C. et al. Mixed coordination silica at megabar pressure. *Phys. Rev. Lett.* **126**, 035701 (2021).
37. Ding, C. et al. High energy density polymeric nitrogen nanotubes inside carbon nanotubes. *Chin. Phys. Lett.* **39**, 036101 (2022).
38. Oganov, A. R. et al. Ionic high-pressure form of elemental boron. *Nature* **457**, 863–867 (2009).
39. Ji, M., Wang, C.-Z. & Ho, K.-M. Comparing efficiencies of genetic and minima hopping algorithms for crystal structure prediction. *Phys. Chem. Chem. Phys.* **12**, 11617–11623 (2010).
40. Deringer, V. L., Pickard, C. J. & Csanyi, G. Data-driven learning of total and local energies in elemental boron. *Phys. Rev. Lett.* **120**, 156001 (2018).
41. Bushlanov, P. V., Blatov, V. A. & Oganov, A. R. Topology-based crystal structure generator. *Comput. Phys. Commun.* **236**, 1–7 (2019).
42. Fung, C. M., Er, C. C., Tan, L. L., Mohamed, A. R. & Chai, S. P. Red phosphorus: an up-and-coming photocatalyst on the horizon for sustainable energy development and environmental remediation. *Chem. Rev.* **122**, 3879–3965 (2022).
43. Ruck, M. et al. Fibrous red phosphorus. *Angew. Chem. Int. Ed.* **44**, 7616–7619 (2005).
44. Hittorf, W. Zur kenntniss des phosphors. *Ann. Phys.* **202**, 193–228 (1865).
45. Thurn, H. & Krebs, H. Über struktur und eigenschaften der halbmetalle. XXII. Die kristallstruktur des hittorfschen phosphors. *Acta Crystallogr. B* **25**, 125–135 (1969).
46. Deringer, V. L., Pickard, C. J. & Proserpio, D. M. Hierarchically structured allotropes of phosphorus from data-driven exploration. *Angew. Chem. Int. Ed.* **59**, 15880–15885 (2020).
47. Lu, Y. L. et al. Fibrous red phosphorene: a promising two-dimensional optoelectronic and photocatalytic material with a desirable band gap and high carrier mobility. *Phys. Chem. Chem. Phys.* **22**, 13713–13720 (2020).
48. Yoon, J. Y. et al. Type-II red phosphorus: wavy packing of twisted pentagonal tubes. *Angew. Chem. Int. Ed.* **62**, e202307102 (2023).
49. Scelta, D. et al. Interlayer bond formation in black phosphorus at high pressure. *Angew. Chem. Int. Ed.* **56**, 14135–14140 (2017).
50. Han, W. H., Kim, S., Lee, I. H. & Chang, K. J. Prediction of green phosphorus with tunable direct band gap and high mobility. *J. Phys. Chem. Lett.* **8**, 4627–4632 (2017).
51. Schusteritsch, G., Uhrin, M. & Pickard, C. J. Single-layered Hittorf's phosphorus: a wide-bandgap high mobility 2D material. *Nano Lett.* **16**, 2975–2980 (2016).
52. Zhang, L. et al. Structure and properties of violet phosphorus and its phosphorene exfoliation. *Angew. Chem. Int. Ed.* **59**, 1074–1080 (2020).
53. Takayanagi, K., Tanishiro, Y., Takahashi, S. & Takahashi, M. Structure analysis of Si(111)-7×7 reconstructed surface by transmission electron diffraction. *Surf. Sci.* **164**, 367–392 (1985).
54. Bauer, M. N., Probert, M. I. J. & Panosetti, C. Systematic comparison of genetic algorithm and basin hopping approaches to the global optimization of Si(111) surface reconstructions. *J. Phys. Chem. A* **126**, 3043–3056 (2022).
55. Du, X. et al. Machine-learning-accelerated simulations to enable automatic surface reconstruction. *Nat. Comput. Sci.* **3**, 1034–1044 (2023).
56. Han, J., Zhang, L., Car, R. & Weinan, E. Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* **23**, 629–639 (2018).
57. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
58. Fan, Z. et al. GPUMD: a package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations. *J. Chem. Phys.* **157**, 114801 (2022).
59. Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csányi, G. MACE: higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
60. Deng, B. et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
61. Ran, Y. et al. Towards prediction of ordered phases in rechargeable battery chemistry via group–subgroup transformation. *npj Comput. Mater.* **7**, 184 (2021).
62. Gao, P., Wang, S., Lv, J., Wang, Y. & Ma, Y. A database assisted protein structure prediction method via a swarm intelligence algorithm. *RSC Adv.* **7**, 39869–39876 (2017).
63. Ahnert, S. E., Grant, W. P. & Pickard, C. J. Revealing and exploiting hierarchical material structure through complex atomic networks. *npj Comput. Mater.* **3**, 35 (2017).
64. Yoshikawa, N. & Hutchison, G. R. Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminform.* **11**, 49 (2019).

65. Deaven, D. M. & Ho, K. M. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **75**, 288–291 (1995).
66. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
67. Gao, H., Wang, J., Guo, Z. & Sun, J. Determining dimensionalities and multiplicities of crystal nets. *npj Comput. Mater.* **6**, 143 (2020).
68. Arenas, A., Fernández, A. & Gómez, S. Analysis of the structure of complex networks at different resolution levels. *New J. Phys.* **10**, 053039 (2008).
69. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
70. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
71. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).
72. Fortunato, S. & Newman, M. E. J. 20 years of network community detection. *Nat. Phys.* **18**, 848–850 (2022).
73. Brunner, G. An unconventional view of the closest sphere packings. *Acta Crystallogr. A* **27**, 388–390 (1971).
74. Lewis, G. V. & Catlow, C. R. A. Potential models for ionic oxides. *J. Phys. C* **18**, 1149 (1985).
75. Fang, C. M. & de With, G. Crystal structure and chemical bonding of the high-pressure phase of $MgAl_2O_4$ from first-principles calculations. *Philos. Mag. A* **82**, 2885–2894 (2002).
76. Broholm, C. et al. Quantum spin liquids. *Science* **367**, eaay0668 (2020).
77. Monthoux, P. & Lonzarich, G. G. Magnetically mediated superconductivity: crossover from cubic to tetragonal lattice. *Phys. Rev. B* **66**, 224504 (2002).
78. Paul, R., Hu, S. X. & Karasiev, V. V. Anharmonic and anomalous trends in the high-pressure phase diagram of silicon. *Phys. Rev. Lett.* **122**, 125701 (2019).
79. Fredericks, S., Parrish, K., Sayre, D. & Zhu, Q. PyXtal: a Python library for crystal structure generation and symmetry analysis. *Comput. Phys. Commun.* **261**, 107810 (2021).
80. Wales, D. J. Symmetry, near-symmetry and energetics. *Chem. Phys. Lett.* **285**, 330–336 (1998).
81. Togo, A., Shinohara, K. & Tanaka, I. Spglib: a software library for crystal symmetry search. *Sci. Technol. Adv. Mater. Methods* **4**, 2384822 (2024).
82. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
83. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
84. Brandes, U. On variants of shortest-path betweenness centrality and their generic computation. *Soc. Netw.* **30**, 136–145 (2008).
85. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
86. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
87. Hagberg, A., Swart, P. & Chult, D. S. *Exploring Network Structure, Dynamics, and Function using NetworkX* (Los Alamos National Laboratory, 2008).
88. Xie, J., Kelley, S. & Szymanski, B. K. Overlapping community detection in networks: the state-of-the-art and comparative study. *ACM Comput. Surv.* **45**, 1–35 (2013).
89. Zhu, Q., Sharma, V., Oganov, A. R. & Ramprasad, R. Predicting polymeric crystal structures by evolutionary algorithms. *J. Chem. Phys.* **141**, 154102 (2014).
90. Tom, R. et al. Genarris 2.0: a random structure generator for molecular crystals. *Comput. Phys. Commun.* **250**, 107170 (2020).
91. Han, Y. et al. Source code and demo of MAGUS (Machine Learning and Graph Theory Assisted Universal Structure Searcher) v2.0.0. *Zenodo* https://doi.org/10.5281/zenodo.14730874 (2025).
92. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

Acknowledgements

We thank Z. Fan and Y. Wang for fruitful discussion regarding NEP usage. We gratefully acknowledge the financial support from the National Key R&D Program of China (grant number 2022YFA1403201), the National Natural Science Foundation of China (grants T2495231, 12125404 and 123B2049), the Basic Research Program of Jiangsu (grants BK20233001 and BK20241253), the Jiangsu Funding Program for Excellent Postdoctoral Talent (grants 2024ZB002 and 2024ZB075), the Postdoctoral Fellowship Program of CPSF (grant GZC20240695), the AI & AI for Science program of Nanjing University, and the Fundamental Research Funds for the Central Universities. The calculations were carried out using supercomputers at the High-Performance Computing Center of Collaborative Innovation Center of Advanced Microstructures and the high-performance supercomputing center of Nanjing University.

Author contributions

Y.H. and C.D. implemented the code, collected and analyzed the data, and led the paper preparation. J. Shi, S.Y., Q.J. and S.P. provided feedback throughout the process, and assisted with the paper writing. J. Sun, H.G. and J.W. conceived the project, supervised the research and contributed to securing funding. All authors participated in the discussion of the results and the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

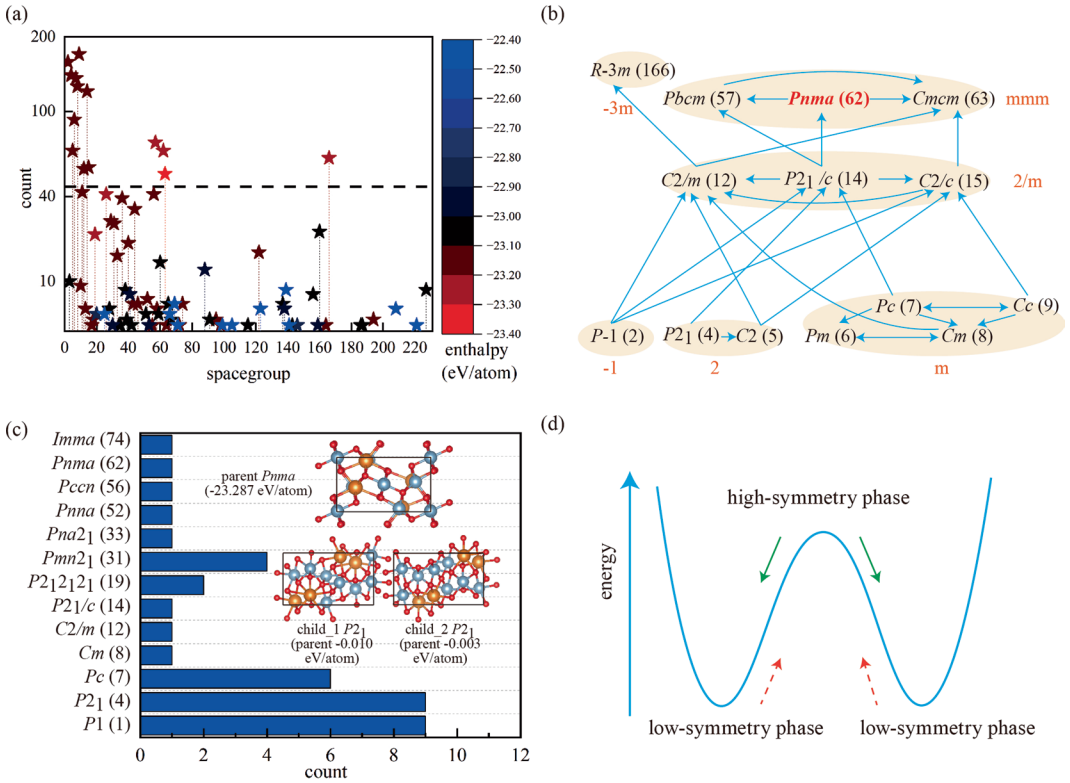
Extended data is available for this paper at https://doi.org/10.1038/s43588-025-00775-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-025-00775-z.

Correspondence and requests for materials should be addressed to Junjie Wang, Hao Gao or Jian Sun.

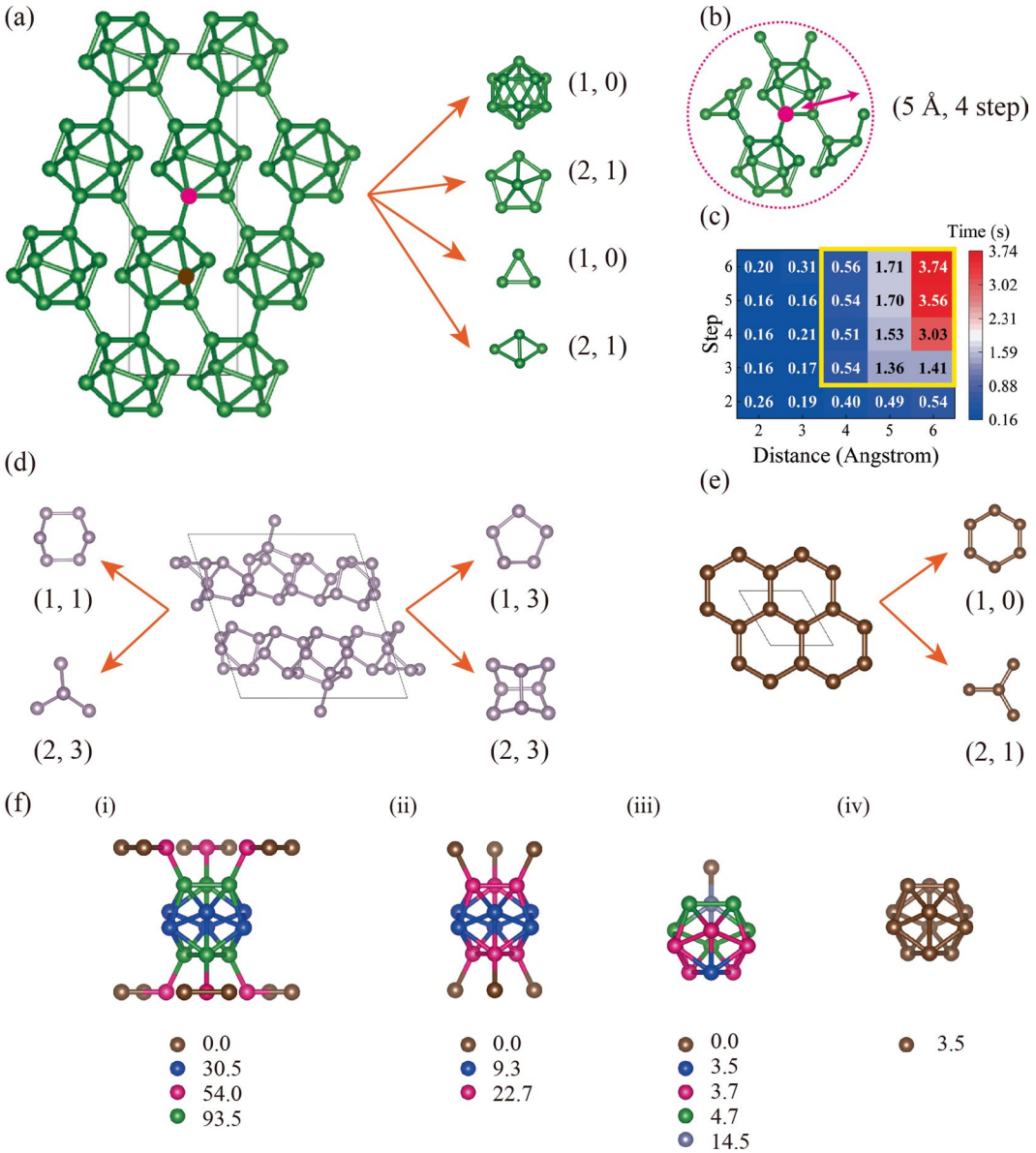
Peer review information *Nature Computational Science* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team.

https://doi.org/10.1038/s43588-025-00775-z



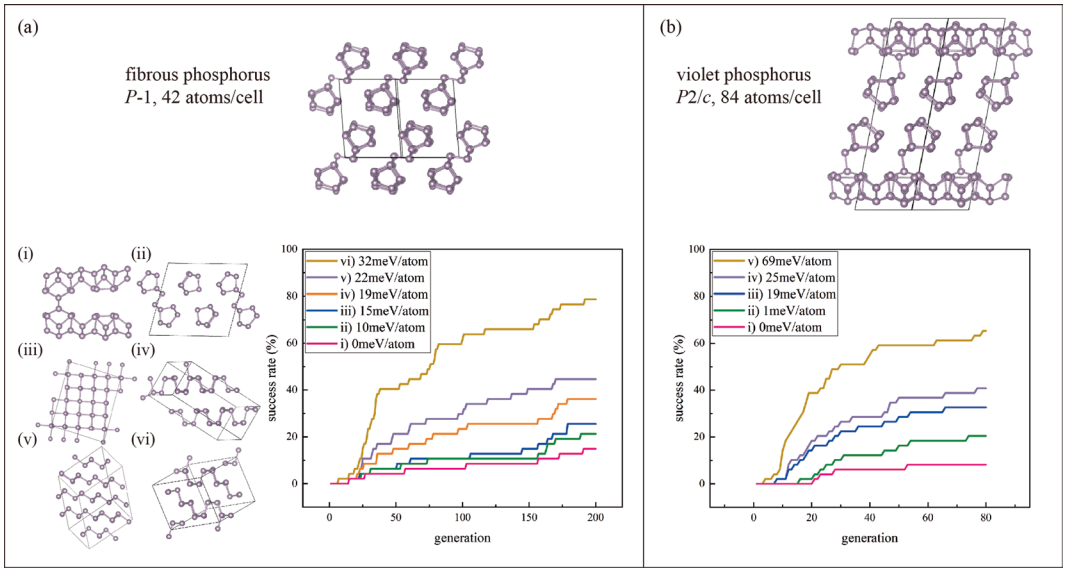
Extended Data Fig. 1 | The symmetry principle in crystal structure prediction for the $MgAl_2O_4$ system. (**a**) Space group distribution (*P*1 structures are excluded) and corresponding enthalpy of 3580 local relaxed structures of $MgAl_2O_4$, whose initial states are 20 random generated structures for each specific space group. The relaxed structures show a preference for certain space group symmetry. (**b**) Group-subgroup relationships between the preferred space groups. Each pair of group-subgroup relations with index<4 is connected by an arrow and the space group of GM is colored in red. (**c**) Spacegoup distribution of resultant structures obtained from applying symmetry-kept rattle mutation to a single metastable *Pnma* phase followed by local relaxation (duplicates of parent structure are excluded). The subfigure shows the view of the parent and two offsprings. It could conclude that most offsprings lowered their symmetry than parent after local relaxation. (**d**) A possible explanation can be derived from the double well model PES. When a high symmetry phase locates surrounding by multi lower energy lower symmetry local minima, its symmetry is prone to breaking. The red atoms represent O, the orange atoms represent Mg, and the cyan atoms represent Al.

https://doi.org/10.1038/s43588-025-00775-z



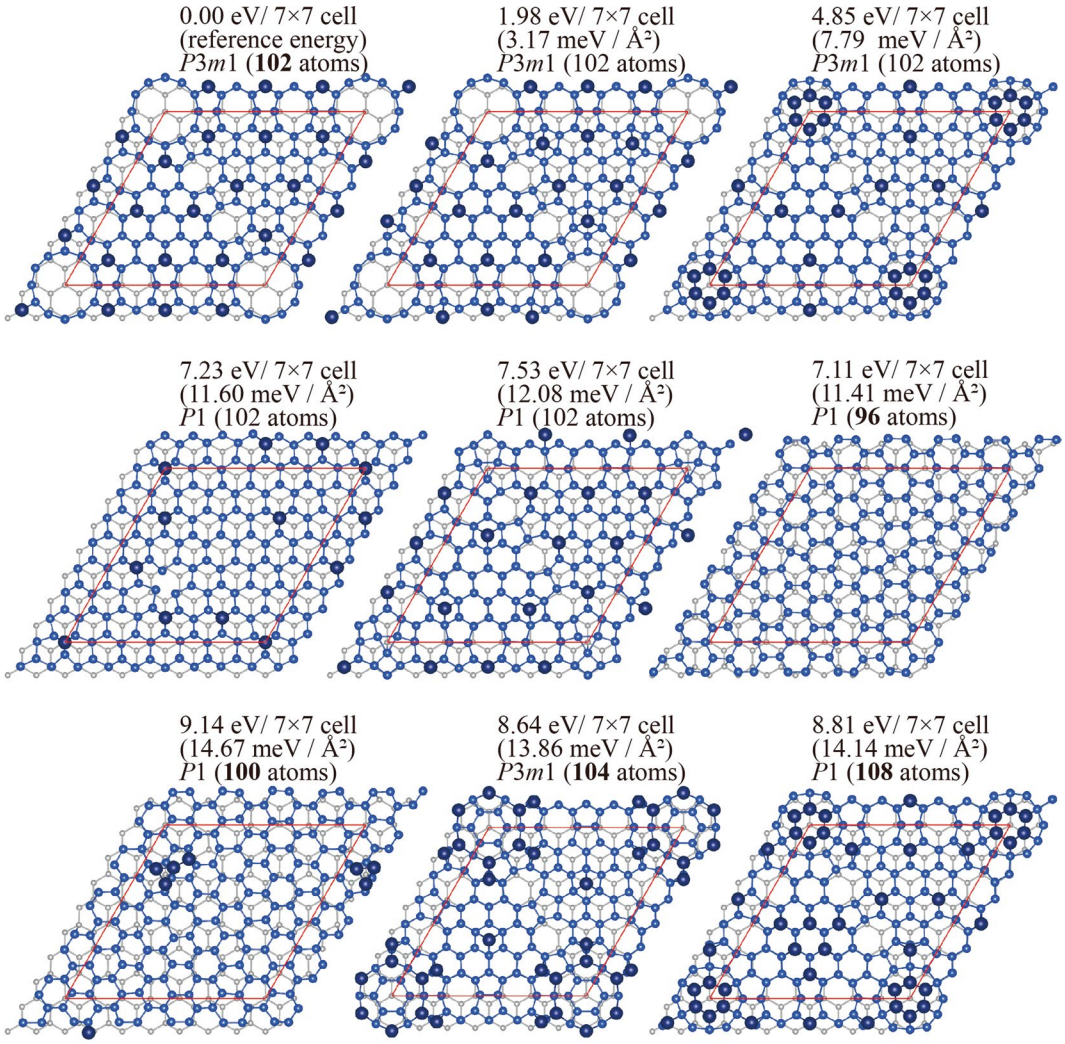
Extended Data Fig. 2 | Illustration of graph-theory-based structure decomposition method. (**a**) Structure of α-B, the boron fragments obtained and their ranking indicators (uniqueness, description length). The two inequivalent atoms in α-B are marked by solid circles of different colors. (**b**) The 'neighborhood' structure within a certain distance and step cutoff of the selected atom marked by magenta. (**c**) Impact of cutoff parameters for building α-B 'neighborhood' structures. As the distance and step cutoff decrease, the time cost decreases as less fragments are identified. The B_{12} icosahedra can only be found with distance cutoff no less than 4 Å and step cutoff no less than 3, marked by yellow line. (**d**) Structure of the fibrous red phosphorous and the decomposed fragments. (**e**) Structure of graphene and the decomposed fragments. (**f**) The B_{12} icosahedra community (iv), has the most uniformly distributed betweenness centrality, and appending any neighbor atom to it (i-iii) will disrupt this uniformity. Therefore, the 'uniqueness' of betweenness centrality is employed as one of the indicators for fragment ranking. Atoms that have same betweenness centrality are same colored, with the specific values of betweenness centrality indicated in the legend at the bottom of the figure.

https://doi.org/10.1038/s43588-025-00775-z



Extended Data Fig. 3 | Structure of the red phosphorus allotropes and the success rate for identifying them. (**a**) Fibrous P structure. Several structures exhibiting characteristics similar to other stable experimental structures are selected and labeled as (ii-vi). The success rate is calculated for identifying structures with machine learning potential energy lower than or equal to that of these reference structures. (**b**) Violet P structure and the corresponding success rates. The reference structures are shown in Supplementary Fig. 1(d).

https://doi.org/10.1038/s43588-025-00775-z



Extended Data Fig. 4 | Different representative metastable Si (111)-(7×7) surface reconstruction models found by MAGUS having 96-108 atoms in the surface region. The surface energy, space group symmetry, and number of atoms in the reconstruction region are indicated. A deeper color and larger atoms represent the upper surface, while lighter color and smaller atoms represent the substrate. The surface energy of reference DAS model is set to 0. More metastable structures are shown in Supplementary Fig. 2.