Computational Materials Science 235 (2024) 112802

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full length article

Towards quantitative evaluation of crystal structure prediction performance

Lai Wei [a,1], Qin Li [b,1], Sadman Sadeed Omee [a], Jianjun Hu [a,*]

[a] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, 29201, United States of America
[b] School of big data and statistics, Guizhou University of Finance and Economics, Guizhou, 550001, China

ARTICLE INFO

Dataset link: http://www.materialsproject.org, https://github.com/usccolumbia/CSPBenchMetrics

Keywords:
Benchmark
Materials discovery
Crystal structure prediction
Distance metric
Performance metrics

ABSTRACT

Crystal structure prediction (CSP) is now increasingly used in the discovery of novel materials with applications in diverse industries. However, despite decades of developments, the problem is far from being solved. With the progress of deep learning, search algorithms, and surrogate energy models, there is a great opportunity for breakthroughs in this area. However, the evaluation of CSP algorithms primarily relies on manual structural and formation energy comparisons. The lack of a set of well-defined quantitative performance metrics for CSP algorithms makes it difficult to evaluate the status of the field and identify the strengths and weaknesses of different CSP algorithms. Here, we analyze the quality evaluation issue in CSP and propose a set of quantitative structure similarity metrics, which when combined can be used to automatically determine the quality of the predicted crystal structures compared to the ground states. Our CSP performance metrics can then be utilized to evaluate the large set of existing and emerging CSP algorithms, thereby alleviating the burden of manual inspection on a case-by-case basis. The related open-source code can be accessed freely at https://github.com/usccolumbia/CSPBenchMetrics.

1. Introduction

Deep learning-based AlphaFold has been revolutionizing the field of molecular biology by predicting tens of thousands of protein structures from sequences [1], which can accelerate the understanding of protein structures and functions. In the realm of computational chemistry, quantitative metrics like RMSD(N) which are evaluated using the COMPACK algorithm [2] and X-ray powder diffraction pattern differences (POWDIFF) using through deGelder's cross-correlation function [3] are commonplace for comparing molecular crystal structures. However, a pressing challenge in materials science, specifically in the domain of inorganic crystal structure prediction, remains unresolved. This challenge seeks to determine the stable crystal structure based solely on the composition of an inorganic material. If successfully addressed, it has the potential to significantly expedite the discovery of novel functional materials, as key material properties such as thermal conductivity, band gap, and elastic constants can be conveniently computed using first-principle methods like Density Functional Theory (DFT) with tools like VASP [4].

Traditionally, the inorganic CSP algorithms are mainly based on the DFT calculation of energies combined with global search algorithms, such as the global landscape exploration [5,6], the minima hopping procedure search for minima on the energy landscape [7], and the KLMC method presented by Woodley [8], which has been instrumental in modeling kinetic processes within the realm of crystal structure prediction. However, the complexity and demanding computing resources of DFT make it challenging to develop new CSP algorithms. Nowadays, although the formation energy can be efficiently predicted by graph neural networks (GNNs) [9–11], there is a good amount of accuracy-computing trade-off in this approach. Recent advancements in deep neural network-based energy potentials [12] have showcased the possibility of efficiently predicting formation energy through potential energy functions derived from neural networks [13]. It can be expected that an increasing number of CSP algorithms will emerge, as has happened in the protein structure prediction field with the CSP competitions organized annually since 1994. In that case, large-scale benchmark studies and objective quantitative evaluation of CSP prediction performances will be needed to illuminate the progress and weaknesses of different algorithms, as has been done in CASP history.

In order to obtain new potential crystal structures, we usually utilize three main categories of crystal structure prediction algorithms, including search-based, template-based, and deep learning-based CSP algorithms. The global search-based CSP algorithms such as USPEX and CALYPSO combine search algorithms with DFT energy calculations for structure search. There are also several open sourced algorithms such as CrySPY [14], XtalOpt [15], GASP [16], and AIRSS [17,18]. However, the most widely used and well-established leading software for de novo

* Corresponding author.
E-mail address: jianjunh@cse.sc.edu (J. Hu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.commatsci.2024.112802
Received 11 July 2023; Received in revised form 8 December 2023; Accepted 9 January 2024
Available online 23 January 2024

CSP are GA-based USPEX and particle swarm optimization (PSO)-based CALYPSO. Despite their closed source code, their binary programs can be easily obtained and both come with several advanced search techniques such as symmetry handling, crowding niche, and so on. Global search has also been combined with universal neural potentials for crystal structure prediction as done by the GN-OA algorithm in [13] and AGOX [19]. With the many possible search algorithms [20], the family of such algorithms can keep growing. The second category of CSP algorithms is template-based element substitutions combined with relaxation including TCSP [21] and CSPML [22], in which they use rules and machine learning models to guide the selection of structure templates. For a given composition, the template-based element substitution method first selects a structure with a similar composition or prototype and then replace some of the atoms in the template structure with atoms of different element types in the given composition. E.g., $SrTiO_3$'s structure can be found by replacing the Ba atoms of the $BaTiO_3$ structure with Sr and then doing structure relaxation. The last category of CSP algorithms is based on deep learning-based algorithms inspired by the Alphafold [23].

With these emerging CSP algorithms, it is critical to benchmark and evaluate their performances in predicting structures of varying complexity so that strengths and obstacles can be identified. However, upon a thorough examination of the relevant literature, it is surprising to find that most of the CSP prediction results are manually verified by authors on a case-by-case basis. This verification process typically involves structure inspection, comparison of formation enthalpies, DFT-calculated energy analysis, examination of property distributions, computation of distances between structures, or a combination of these methods. There has been a severe lack of quantitative measures for the objective evaluation of CSP prediction performance. In one of the earliest reports of USPEX (Universal Structure Predictor: Evolutionary Xtallography), which used evolutionary algorithms to identify stable and novel crystal structures that may not be easily discovered through traditional experimental or theoretical methods [24], the authors compared the energy difference of the predicted structures and ground state and then compared the structural similarity by manually inspecting the predicted structures against the experimentally determined structures. Similar approaches have been used for assessing predicted crystal structures in related CSP works [25,26] using evolutionary algorithms. In a related study by Hofmann et al. [27], the authors used the largest distance between the unit cell edges and the nearest grid point of the experimental structure to evaluate CSP performance. Another widely used method for generating predicted crystal structures is CALYPSO (Crystal Structure Analysis by Particle Swarm Optimization), which was developed by Wang et al. [28]. In this work, energy distributions and the distance against distortion for graphite and diamond structures were utilized to evaluate and analysis the predicted structures. Additionally, in the work by Tong et al. [29] on accelerating CALYPSO using data-driven learning of a potential energy surface, the authors employed the evolution of the root mean square errors (RMSEs) of the predicted energy and force by the Gaussian Approximation Potential (GAP) for the testing set to evaluate the CALYPSO structure search for a specific cluster. A vector-type structure descriptor distance has also been used for comparing the predicted structures against the ground states [30].

The metrics used in validating the predicted structures against ground states were usually set by the authors with a certain arbitrariness. So far there is not a set of quantitative indicators of the quality of the predicted crystal structures with easy-to-use calculation code despite the existence of several structure similarity metrics as summarized in [31]. Table 1 provides an overview of the evaluation methods currently used in state-of-the-art CSP works. The abbreviations M-i, M-o, M-e, M-s, and M-d represent manual structural inspection, comparison with experimentally observed structures, comparison of energy or enthalpy values, success rate analysis, and computation of distances between structures, respectively. It is noteworthy that computational methods such as DFT-energy or enthalpy calculations are commonly employed in many studies. However, manual structural similarity inspection methods continue to be widely used even today, which leads the casual reader to wonder how exactly a predicted crystal structure is evaluated in terms of its prediction quality, especially when the predicted structure does not exactly match the ground state. Additionally, energy or enthalpy calculations for structure similarity evaluation using DFT can be time-consuming. Furthermore, performance evaluation methods such as success rates, and ad hoc distance calculations between structures present challenges in standardizing, validating, and comparing the CSP results.

Inspired by the variety of quantitative metrics used in evaluating molecule generation algorithms by the benchmark MOSES [46], here we aim to address the challenges in defining good structure distance/similarity scores to measure the quality of CSP algorithms. We evaluated a series of energy and structure-based performance metrics for crystal structure prediction algorithms. For each metric, we check how their values correlate with the formation energy differences and perturbation deviations between the predicted structures and the ground state structures. We tested their correlations for both random perturbations (applied to each atomic site independently) and symmetric perturbations [43] (applied only to Wyckoff sites without disrupting the symmetry), both of which have been adopted in CSP algorithms. We also showed that while every single metric cannot be used to fully characterize the quality of a predicted structure against the ground state, together they can capture the key structural similarity. Applications of these metrics were additionally used to compare the performance of CSP algorithms based on different search algorithms. We have also used these metrics to visualize the search trajectories of the structures for the GN-OA algorithms and explained their key limitations.

2. Method

2.1. Evaluation metrics

Evaluation metrics play a crucial role in materials science research, as they provide a quantitative way to assess the performance and effectiveness of different material structure prediction algorithms. Currently, there are many evaluation benchmark metrics in the molecule research area, such as RDKit [47] and MOSES [46]. However, in the field of materials informatics, we do not have a unified standard for evaluating the similarity between two crystal structures that arise during the crystal structure prediction process. Here we introduce a set of benchmark metrics for CSP by combining the energy distance along with several common distance metrics, including M3GNet energy distance, minimal Wyckoff RMSE distance, minimal Wyckoff MAE distance, RMS distance, RMS anonymous distance, the Sinkhorn distance, the Chamfer distance, the Hausdorff distance, superpose RMSD distance, edit graph distance, XRD spectrum distance, fingerprint distance to standardize the comparison of material structure prediction algorithms.

The structure similarity in CSP has a unique property: the candidate structure and the ground state structure compared have the same number of atoms within the given unit cell. Then the key step is to match atoms of one structure to the corresponding atoms of the other structure to minimize the MAE error. There are several desirable characteristics for a good structure similarity measure: (1) correlation: the structure difference should correlates well with the distance metric; (2) convergence: when the predicted structures during the CSP search approach to the ground state, the distance metric scores should converge to 0; (3) applicability: the distance metric should be used to not just evaluate very similar structures, but also relative distant intermediate structures, which lacks in the success rate metric. Here we introduce eleven distance metrics that may be used in evaluating the prediction performance of CSP algorithms.

Table 1
Overview of state-of-the-art CSP works for validating predicted structures. Abbreviations M-i, M-e, M-s, M-d stand for different validation methods, where M-i, M-e, M-s, M-d represent manual inspection, comparison of the energy or enthalpy, success rate, and computation of distances between structures.

Author	Algorithm	Year	M-i	M-e	M-s	M-d
J.C. Schön, M. Jansen [32]	Simulated annealing	1995	✓			
H. Putz [33]	Simulated annealing	1999	✓	✓		
Hofmann DWM [27]	Data Mining	2003	✓			
Scott M. Woodley [34]	Evolutionary Algorithm	2004		✓	✓	
R. Oganov [24]	Evolutionary Algorithm	2006	✓	✓		
R. Oganov [25]	Evolutionary Algorithm	2006	✓	✓		
Christopher C. Fischer [35]	Data Mining	2006	✓			
Kuo Bao [36]	Hopping method	2009		✓		
Giancarlo Trimarchi [37]	Evolutionary Algorithm	2009	✓	✓		
R. Oganov [26]	Evolutionary Algorithm	2010		✓		
David C. Lonie [15]	Evolutionary Algorithm	2011		✓		
Yanchao Wang [28]	Particle swarm optimization (PSO)	2012		✓		✓
S Q Wu [38]	Evolutionary Algorithm	2013	✓			
Anton O. Oliynyk [39]	Data-Driven: ML	2017		✓		
Qunchao Tong [29]	Particle swarm optimization (PSO)	2018		✓		✓
Maximilian Amsler [40]	Hopping method	2018		✓		
Asma Nouira [41]	Data-Driven: ML	2018		✓	✓	
Evgeny V. Podryabinkin [42]	Evolutionary Algorithm	2019	✓			
Lai Wei [21]	Template-Based Substitution	2022				✓
Xuecheng Shao [43]	A symmetry-orientated method	2022	✓		✓	
Xiangyang Liu [44]	Evolutionary Algorithm	2022	✓			
Yanchao Wang [45]	Particle swarm optimization (PSO)	2022	✓			
Guanjian Cheng [13]	Data-Driven: ML	2022	✓		✓	

2.1.1. Energy distance (ED)

The formation energy is the energy required to form a material from its constituent elements in their reference states, which can provide information regarding the stability and reactivity of materials. While DFT calculation of formation energy is ideal for its accuracy, it is too slow in many applications. As a result, machine learning-based energy models have become an important topic with significant progress recently for materials discovery and design. Here we use the M3GNet [12], a graph neural network-based surrogate potential model, to calculate the formation energies of the ground state structure and the predicted structure and then their energy distance. The formula is shown in the following equation:

$$\mathrm{ED} = |E_p - E_g|, \tag{1}$$

where E_p is the energy of the predicted structure while the E_g is the ground state structure.

It should be noted that zero of small energy distance of two structures does not mean they have identical structures. For example diamond and graphite have small energy difference but the structures are very different.

2.1.2. Wyckoff position fraction coordinate distance (WD)

The Wyckoff position fraction coordinate distance is used to compare two structures that have the same Wyckoff position configurations and the same space group in the symmetrized structures. It is useful to measure the similarity of a candidate structure and the ground state structure for those CSP algorithms that can search structures while preserving symmetry (space groups). We used both RMSE and MAE to calculate the Wyckoff position fraction coordinate distances.

RMSE stands for Root Mean Square Error, which can be calculated as the square root of the average of the squared differences between the predicted and actual values. Let $X = \{p_1, p_2, \ldots, p_n\}$ and $Y = \{q_1, q_2, \ldots, q_n\}$ be two sets of n data points. Then their RMSE error can be calculated as:

$$\mathrm{WD}_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i - x_i'\right)^2 + \left(y_i - y_i'\right)^2 + \left(z_i - z_i'\right)^2\right)} \tag{2}$$

where (x_i, y_i, z_i) are the coordinates of the ith point p_i from set X. (x_i', y_i', z_i') are the coordinates of the ith point q_i from set Y. n is the number of data points in each set. To calculate this distance, the input structures have to be symmetrized first e.g. using Pymatgen's SpaceGroupAnalyzer module and their space group and Wyckoff position configuration must be the same. The related algorithms FINDSYM, SFND, and RGS can be found in [48–50].

The MAE Distance is the mean absolute error (MAE) distance between two sets of data points. The MAE distance is calculated by taking the absolute difference between corresponding data points in the two sets of data, summing these absolute differences, and dividing by the total number of data points. Let $X = \{p_1, p_2, \ldots, p_n\}$ and $Y = \{q_1, q_2, \ldots, q_n\}$ be two sets of n data points. The mean absolute error (MAE) Wyckoff distance between X and Y is defined as:

$$\mathrm{WD}_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(x_i - x_i'\right)^2 + \left(y_i - y_i'\right)^2 + \left(z_i - z_i'\right)^2} \tag{3}$$

where (x_i, y_i, z_i) are the coordinates of the ith point p_i from set X. (x_i', y_i', z_i') are the coordinates of the ith point q_i from set Y. n is the number of data points in each set.

Note that the above calculations assume that the Wyckoff positions are selected correspondingly for the given two compared structure. Otherwise, one need to calculate the minimal RMSE/MAE distance between each pair of Wyckoff positions, which is the smallest possible RMSE/MAE distance that can be obtained between X and Y by permuting the equivalent data points for each Wyckoff position in X and Y since the Wyckoff coordinates in the cif file can be any point of all the equivalent positions. The WD distance is also applicable for the results of global optimizations that were performed in spacegroup P1 where the symmetries can be assigned afterward optimization using Pymatgen routine.

2.1.3. Adjacency matrix distance (AMD)

The adjacency matrix (AM) is widely used to represent the connection topology of atoms for a given crystal structure. The value of $M(i, j)$ is set to 1 if there exists a bond between the atom i and atom j or set to 0 if otherwise. Here we use the canonical distance as the cutoff distance for a pair of element types to define the connection status. Given two structures S_1 and S_2 with adjacency matrices M_1 and M_2, the adjacency matrix distance is defined as:

$$\mathrm{AMD} = 1 - \frac{2n}{n_1 + n_2} \tag{4}$$

where n is the number of matrix cells that both matrices have the value of 1. n_1 is the number of matrix cells of M_1 with the value of 1 and n_2 is the number of matrix cells of M_2 with the value of 1.

The AMD can be used to measure the topological similarity between two compared structures, which are assumed to have an equal number of atomic sites. However, there have been longstanding discussions in the chemical and crystallographic literature since the 1960s, concerning the choice of a cutoff radius for defining atomic connections. The use of a canonical distance as a uniform cutoff might oversimplify the nuanced nature of interatomic relationships within crystal structures. This oversimplification could potentially impact the accuracy of the adjacency matrix and the subsequent evaluation of topological similarity between structures. We also find that the correlation between the perturbation magnitude and the AMD distance is weak (see Supplementary Figure $S3$ and $S4$). Our findings underscore the potential limitations of relying solely on a canonical distance as a cutoff measure for defining atomic connections within crystal structures, but our primary objective is to enable equitable and precise comparisons among diverse structures, thereby enhancing the overall quality assessment. This standardization not only ensures fairness in our evaluations but also fosters the reliability and robustness of our findings, allowing us to draw meaningful conclusions from our analyses.

2.1.4. Pymatgen RMS distance (PRD) and RMS anonymous distance (PRAD)

We calculate RMS Anonymous Distance using the structure_matcher module of the PyMatGen (Python Materials Genomics) package [51], which allows distinct species in one structure to map to another. It also calculates the root-mean-square error (RMSE) between two structures according to Eq. (2), but its atomic site matching process does not consider the difference of atom types before calculating the RMSD. It is useful in cases where one wants to compare the overall structural similarity between two structures, without being concerned with the differences in atom types.

2.1.5. Sinkhorn distance (SD)

The Sinkhorn Distance (SD) [52] is a distance metric commonly used to compare two probability distributions or point clouds, particularly in high-dimensional spaces. In the context of crystal structures, we can represent them as point clouds by considering their constituent atomic sites. Given two structures as $S_1 = \{p_1, p_2, \dots, p_m\}$ and $S_2 = \{q_1, q_2, \dots, q_n\}$ where p_i and q_i are atomic sites (point clouds) of structure S_1 and S_2, respectively. The Sinkhorn Distance (SD) does not rely on relative coordinates and remains independent of the simulation cell used. Formally, the SD is defined as follows:

$$SD\left(S_1, S_2\right) = \frac{1}{\epsilon}\left(\sum_{i,j} T_{i,j} \log \frac{T_{i,j}}{u_i v_j} - \log C\right) \tag{5}$$

In Eq. (5), S_1 and S_2 correspond to the crystal structures being compared. The terms u and v refer to the marginals of S_1 and S_2, respectively. These marginals represent the total mass assigned to each atomic site within their respective distributions, shedding light on how mass is distributed across the atomic positions in each crystal structure. Additionally, ϵ is a regularization parameter, and C serves as a normalization constant. T is the transport plan which quantifies how mass is efficiently redistributed between atomic sites in S_1 to align it with S_2. The transport plan can be computed using the equation that follows, revealing the optimal mass transfer strategy between the crystal structures.

$$T_{i,j} = \exp(-\epsilon h_{i,j}) u_i v_j \tag{6}$$

In Eq. (6), $h_{i,j}$ denotes the transport cost of transporting one unit of mass from site i in S_1 to site j in S_2.

SD can be thought of as a regularized version of the Earth Mover's Distance (EMD), also referred to as the Wasserstein Distance, where the smoothness of the transport plan is controlled by the regularization parameter ϵ. The transport plan gets quite sparse and the SD approaches the EMD when ϵ is very large. The transport plan becomes very dense and the SD approaches the Euclidean Distance when ϵ is very tiny.

2.1.6. Chamfer distance (CD)

The Chamfer Distance (CD) [53] is defined as the average distance of the summed-up squared distances between two point clouds' nearest neighbor correspondences. Similar to SD, we also represent two crystal structures as atomic sites to define it. Given two structures as $S_1 = \{p_1, p_2, \dots, p_m\}$ and $S_2 = \{q_1, q_2, \dots, q_n\}$ where p_i and q_i are atomic sites of structure S_1 and S_2, respectively, we can formally define CD as:

$$CD\left(S_1, S_2\right) = \frac{1}{m}\sum_{p \in S_1} \min_{q \in S_2} \|p - q\|_2 + \frac{1}{n}\sum_{q \in S_2} \min_{p \in S_1} \|q - p\|_2 \tag{7}$$

In Eq. (7), $\|p-q\|^2$ is the squared Euclidean Distance between sites p and q. It is relatively fast and easy to compute, can handle large point sets, and is less sensitive to outliers and noise in the data. However, it also has some drawbacks, such as being dependent on the metric used to measure distances between points and being insensitive to the relative ordering of the points.

2.1.7. Hausdorff distance (HD)

The Hausdorff Distance (HD) [54] measures the maximum distance between any point in one set and its nearest point in the other set, or vice versa. We represent two crystal structures as atomic sites to define them. Given two structures as $S_1 = \{p_1, p_2, \dots, p_m\}$ and $S_2 = \{q_1, q_2, \dots, q_n\}$ where p_i and q_i are atomic sites of structure S_1 and S_2, respectively, we can formally define HD as:

$$HD(S_1, S_2) = \max\left\{\sup_{p \in S_1} \inf_{q \in S_2} \|p - q\|, \sup_{q \in S_2} \inf_{p \in S_1} \|q - p\|\right\} \tag{8}$$

In Eq. (8), $\|p - q\|$ is the distance between sites p and q, which can be any distance metric such as the Euclidean Distance, the Manhattan Distance, or the Minkowski Distance. The sup function takes the supremum or the least upper bound of the distances overall points in the set, and the inf function takes the infimum or the greatest lower bound of the distances overall points in the other set. The max function takes the maximum of the two supremum, which ensures that the Hausdorff distance is a symmetric metric.

2.1.8. Superpose Distance (SPD)

The Superpose Distance (SPD) is a measure of the structural similarity between two 3D protein structures, which is very similar to the CMPZ algorithm used to compare the similarity of periodic structures [31]. SPD is essentially a variation of the RMSE, which is a commonly used metric for quantifying the structural similarity between two protein structures. We again represent two crystal structures as atomic sites to define them. Given two structures as $S_1 = \{p_1, p_2, \dots, p_n\}$ and $S_2 = \{q_1, q_2, \dots, q_n\}$ where p_i and q_i are atomic sites of structure S_1 and S_2, respectively. We use the Superpose3D [55] package which takes these two structures as input representing two sets of atomic sites of the same length N. It attempts to superimpose them using rotations, translations, and optionally scale transformations while treating them as rigid objects in order to reduce the RMSE across corresponding sites. The RMSE of the paired sites is calculated as SPD following alignment. It can be defined by the following equation:

$$SPD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{s_l}\left|S_{1_{ni}} - \left(\sum_{j=1}^{s_l} hR_{ij}S_{2_{nj}} + T_i\right)\right|^2} \tag{9}$$

In Eq. (9), s_l denotes the dimension of the atomic sites, R_{ij} denotes rotation matrix (a $s_l \times s_l$ sized array representing the rotation), where $|R| = 1$, T_j denotes a translation vector, and h denotes a scalar. One limitation of this distance measure is that their superimposing/alignment does not consider the atomic types of the points.

2.1.9. Graph Edit Distance (GED)

The Graph Edit Distance (GED) [56] is a distance metric used to compare two graphs with possibly different numbers of nodes and edges. It measures the minimum number of operations needed to convert one graph into another. The permitted operations are insertions, deletions, and substitutions of nodes and edges. GED is defined by the following equation:

$$GED(G_1, G_2) = \min_{\zeta}\Big(\sum_{i\in V_1}\alpha_{\zeta}(i) + \sum_{(i,j)\in E_1}\beta_{\zeta}(i,j)\Big) \tag{10}$$

In Eq. (10), $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are the two input graphs to be compared, ζ is a graph edit path that maps nodes and edges from G_1 to G_2, α, and β are the cost of editing a node and an edge, respectively.

Here, we use a measure of the similarity between two crystal structures represented as graphs based on the differences in the connectivity and bonding patterns of the atoms in the structures. We use the atomic number difference as the node substitution cost and the length difference of two edges as the edge substitution cost. Then the GED uses the linear sum assignment algorithm, also known as the Hungarian algorithm [57] to find an assignment of nodes in structure A to nodes of structure B that minimizes the total substitution cost. The algorithm works by iteratively constructing a dual feasible solution, which is then used to generate an alternating path in a bipartite graph. This alternating path is used to update the current assignment and improve the overall cost. We use the implementation in [58] package for this distance calculation.

This GED distance considers the atomic site element types during its site alignment process, which can complement the shortcoming of the Superpose distance SPD. However, we found that the correlation between the perturbation magnitude and the GED distance is weak (see Supplementary Figure S1 and S2).

2.1.10. X-ray diffraction spectrum distance (XD)

X-ray diffraction (XRD) is a technique used to determine the atomic and molecular structure of a material by analyzing the diffraction patterns resulting from X-ray interactions with a crystal. In our investigation, we utilize Pymatgen's diffractionPatternCalculator to compute the XRD features, which manifest as intensity values, for a specified structure. In order to quantitatively assess the resemblance between two material structures based on their XRD features, we calculate the Euclidean distance, referred to as the X-ray diffraction (XRD) Spectrum Distance. Given two crystal structures $S_1 = (p_{x_1}, p_{x_2}, \ldots, p_{x_n})$ and $S_2 = (q_{x_1}, q_{x_2}, \ldots, q_{x_n})$, where p_{x_i} and q_{x_i} represents the XRD features of structures S_1 and S_2, respectively, in an n-dimensional space, the Euclidean XRD spectrum distance XD between the two structures can be represented as follows:

$$XD\left(S_1, S_2\right) = \sqrt{\sum_{i=1}^{n}\left(q_{x_i} - p_{x_i}\right)^2} \tag{11}$$

2.1.11. Orbital Field Matrix distance (OD)

The Orbital Field Matrix (OFM) is calculated for each site in the supercell by considering the valence shell electrons of neighboring atoms. Each crystal structure is transformed into a supercell, which prevents sites from coordinating with themselves. And then the average of the orbital field matrices at each site is found to characterize a structure. The Orbital Field Matrix (OFM) distance is determined by calculating the Euclidean distance between the OFM features [] of two structures.

Given two crystalline structures $S_1 = (p_{o_1}, p_{o_2}, \ldots, p_{o_n})$ and $S_2 = (q_{o_1}, q_{o_2}, \ldots, q_{o_n})$, where p_{o_i} and q_{o_i} represents the OFM features of structures S_1 and S_2, respectively, in an n-dimensional space, the Euclidean OFM distance OD between two structures can be defined as the following equation:

$$OD\left(S_1, S_2\right) = \sqrt{\sum_{i=1}^{n}\left(q_{o_i} - p_{o_i}\right)^2} \tag{12}$$

2.1.12. CrystalNN Fingerprint distance (CFD)

CrystalNN Fingerprint distance (CFD) is defined over the CrystalNN fingerprint representation of crystal structures [59]. It is a type of fingerprint used in materials science and chemistry to represent the local environment around each atom in a crystal structure. It is based on the concept of coordination number, which refers to the number of neighboring atoms that are closest to a particular atom in a crystal structure. The CrystalNN fingerprint calculates the coordination number of each atom in a crystal structure using a nearest-neighbor algorithm called CrystalNN, which is based on the Voronoi tessellation of the crystal lattice. The algorithm takes into account the sizes and charges of the atoms involved and can distinguish between different coordination environments, such as tetrahedral, octahedral, and trigonal bipyramidal. The resulting fingerprint is a vector of 244 numbers that represents the coordination environment around each atom in the crystal structure. The fingerprint can be used to compare different crystal structures and identify similarities and differences in their local environments. It can also be used as input for machine learning algorithms to predict the properties of materials based on their crystal structures.

2.2. Evaluation procedure

We used three ways to evaluate the utility of the selected distance metrics for CSP study including: (1) studying how the structure distance metrics change with the structure perturbation; (2) comparing the CSP performances of three CSP algorithms over a set of test structures; (3) understanding the search dynamics or behavior of the optimization algorithms using the trajectories of the search process.

3. Results

3.1. Evaluation of performance metrics

To evaluate how different performance metrics reflect the actual closeness of the predicted structures to the ground state structures, we use two perturbation methods to generate two sets of perturbed crystal structures with varying magnitudes for a given stable structure. We then calculate how their formation energy differences correlate with the performance metric distances as well as the perturbation magnitudes. The first perturbation method directly changes the coordinates of all sites by a random uniform p percentage (from 1 to 50% with 1000 points) without considering the space group symmetry. So the resulting structures may lose their space symmetry. By perturbing a given stable crystal structure with an increasing series of perturbation magnitudes, we can simulate the search process of CSP algorithms to a certain degree. The second perturbation method comes from the PyxTal package [60], which can do two types of perturbations over a given structure: one is changing the lattice parameters by a given percentage; the other is perturbing the atomic coordinates of the Wyckoff sites with a given magnitude in Å. These perturbation strategies serve as controlled means to assess how performance metric distances correspond with the differences in formation energies, as well as the magnitudes of the perturbations themselves. Throughout these perturbation processes, the periodicity and integrity of the unit cell remain unaffected despite the alterations applied to the atomic coordinates or lattice parameters. Here we focus on this symmetry-preserving Wyckoff site coordinate perturbation with a 2% lattice perturbation.

Fig. 1(a) shows the parity plot of perturbation magnitude and the formation energy distances of the structures compared to the ground state structure $SrTiO_3$. Here the formation energy is predicted using the universal machine learned M3GNet potential [12]. It can be found that as the perturbation percentage goes up, the energy difference between the perturbed structures and the stable structure also goes up. We can also find that the range of energy distances for a given perturbation percentage increases as the perturbation magnitude goes up, indicating the fact that highly disrupted structures tend to have diverse energy

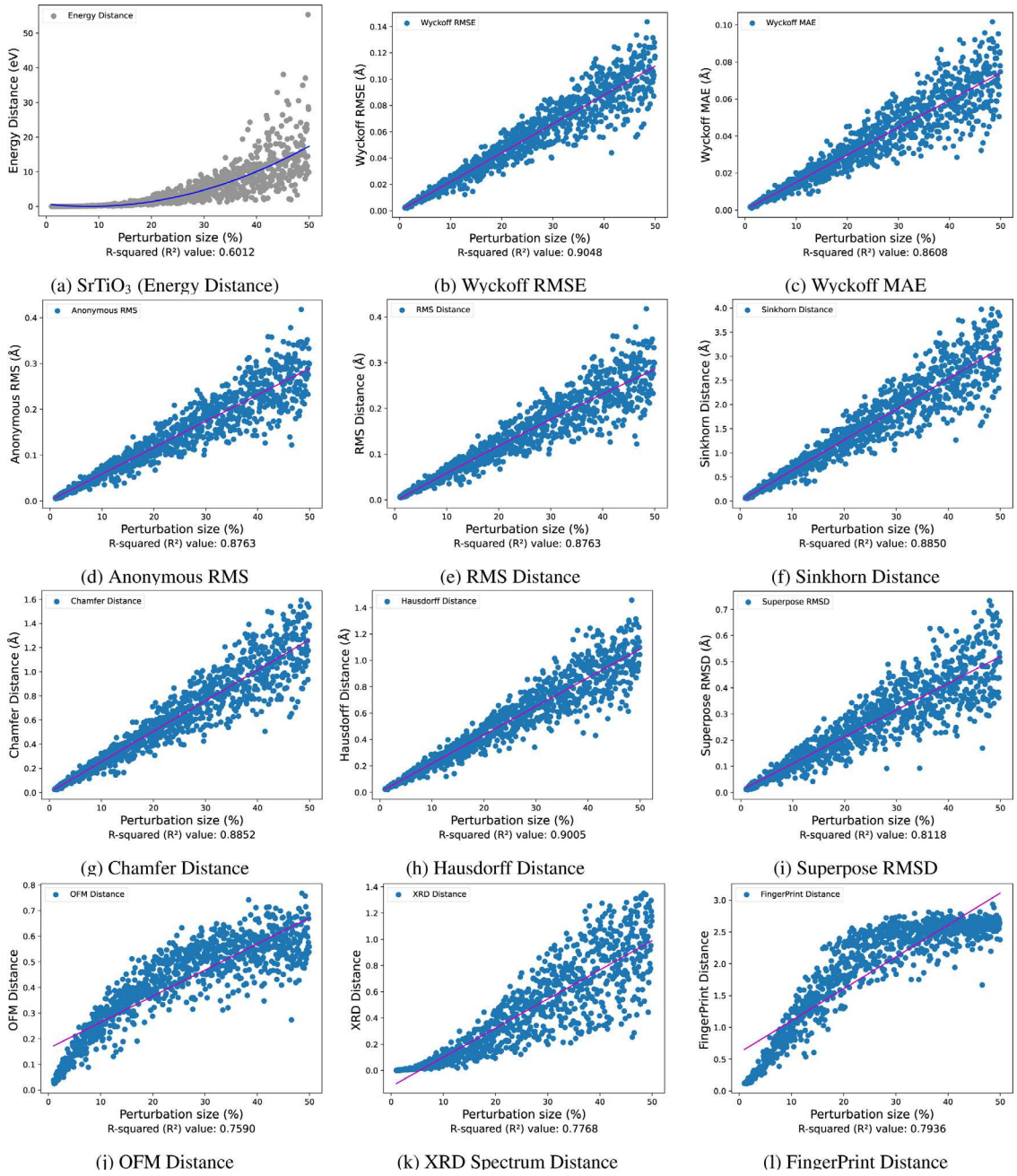


Fig. 1. Structure distances vs perturbation size evaluated over the random perturbation structures of $SrTiO_3$. The units for XRD Spectrum distance and OFM distance are counts per second and the number of valence shell electrons, respectively.

values. Fig. 1(b)–(l) shows the correlation between perturbation magnitude and eleven performance metrics. It has three types of correlations. Fig. 1(b)–(i) show the linear correlation of perturbation with respect to the following distance metrics, including Wyckoff RMSE, Wyckoff MAE, Anonymous RMS, RMS distance, Sinkhorn distance, Chamfer distance, Hausdorff distance, and Superpose RMSD. Out of the eleven metrics, Wyckoff RMSE, RMS distance, and Hausdorff distance (Fig. 1(b)(e)(h)) show a higher degree of linearity. The units for XRD Spectrum distance and OFM distance are counts per second and the number of valence shell electrons, respectively. Relatively speaking, the remaining metrics demonstrate a certain degree of nonlinear correlation, including XRD Spectrum distance, OFM distance, and FingerPrint distance sorted by the degree of nonlinearity. All eleven metrics can be used to measure the similarity between candidate structures and the ground state structure.

While the eleven performance metrics show a good correlation with the structure perturbation in Fig. 1, they are evaluated over the randomly perturbed structures which neglect the symmetry relationships among equivalent Wyckoff atomic sites. However, many efficient CSP algorithms use symmetry-obeying search operators which do not violate the atomic symmetry relationship during the coordinate search. To simulate this situation, we generate a second set of symmetry-preserving perturbed structures from the ground state structure ZrSO. Fig. 2 shows the correlations of performance metrics with respect to the perturbation magnitude. Compared to the random perturbation structures of $SrTiO_3$, symmetrical perturbations, in which the perturbation is applied to the coordinates of Wyckoff sites, may lead to more pronounced structural changes. It can be found that the correlation between perturbation magnitude and performance metrics is weaker (Fig. 2). Because symmetrical structures possess internal repetition

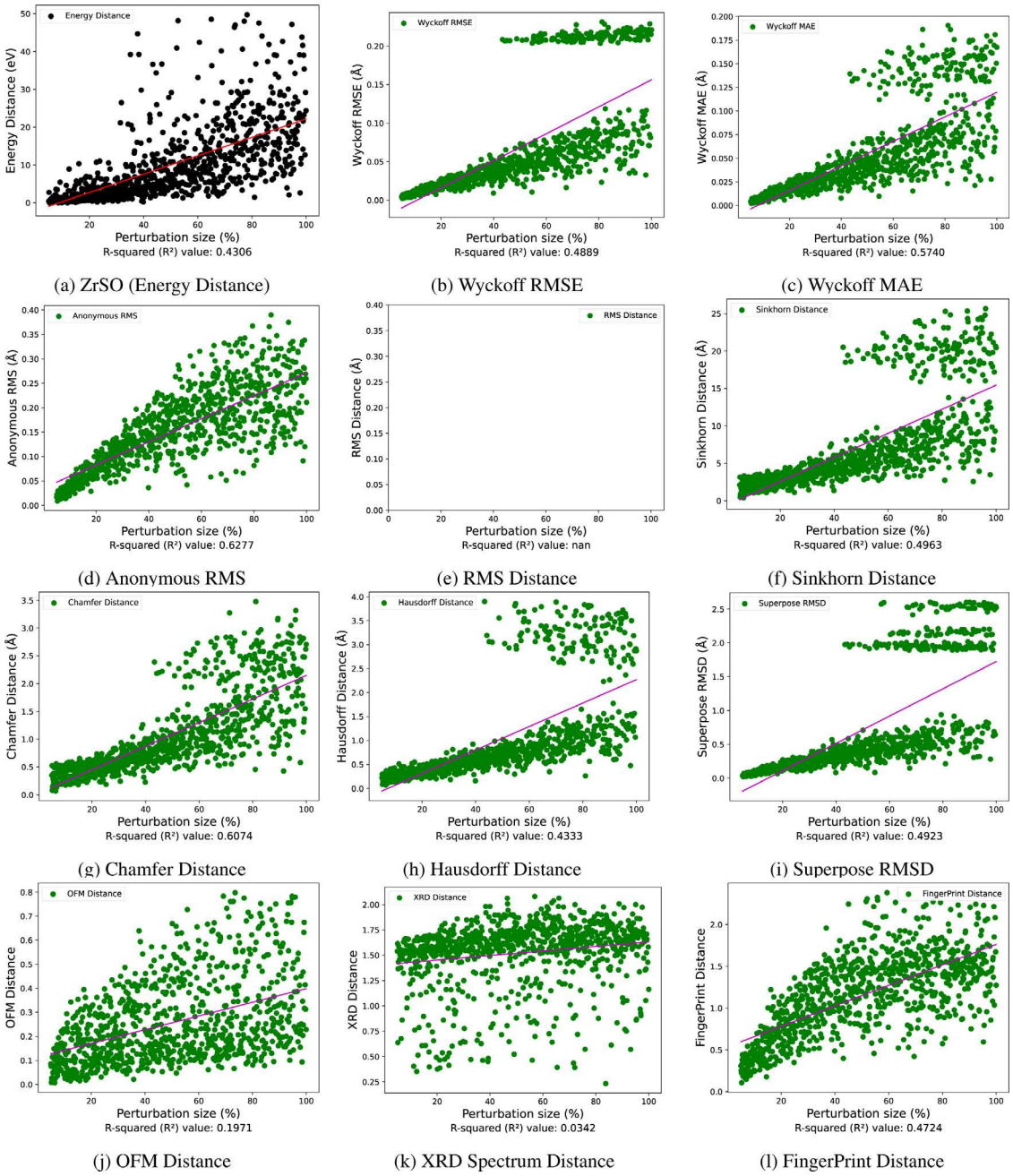


Fig. 2. Structure distances vs perturbation size over the symmetrically perturbation structures of ZrSO. The dataset is generated by applying lattice a/b/c perturbation with small 5% changes while fraction coordinates are perturbed with 2% to 100%. Space group remains unchanged. (e) The units for XRD Spectrum distance and OFM distance are counts per second and the number of valence shell electrons, respectively.

and symmetry, a perturbation at one point can propagate equivalent perturbations in other points. Consequently, accurately describing the perturbation based solely on its magnitude becomes challenging. In such scenarios, the influence of perturbations can extend beyond their immediate vicinity, impacting a broader range of elements or variables within the system. Therefore, to fully comprehend and analyze the effects of symmetrical perturbations, it is crucial to possess a comprehensive understanding of the system's dynamics and interactions. From Fig. 2(a), we can find that the perturbations quickly lead to high variations of the energies of the perturbed structures despite that the pattern is similar to random perturbation when the perturbation magnitude/percentage is small. For the distance metrics in Fig. 2 (b, c, f, g, h, i), we all observe regular patterns at the bottom which are similar to the trends in the random perturbation results (Fig. 1). The top right overhead dots show the impact of symmetric perturbation: a relatively small perturbation can also cause big distance changes. We also find that Figs. 2(j, k, l) have much higher variation than those in Figs. 1(j, k, l) respectively. These results show that our selected distance metrics tend to have higher variation when used to evaluate the structure similarity for symmetrically perturbed structures. As shown in 2(e), due to the relatively large changes in the structure of the symmetric perturbation, the distance metric of the RMS could not be calculated.

In order to better assess and analyze our metrics, we selected a subset of 200 structures with the lowest energy distances to plot the correlation between energy distances and structure distances over the random and symmetrically perturbation structures of $SrTiO_3$ and ZrSO. Fig. 3 shows the correlation of Wyckoff RMSE, Wyckoff MAE, Anonymous RMS, Sinkhorn distance, Chamfer distance and Hausdorff

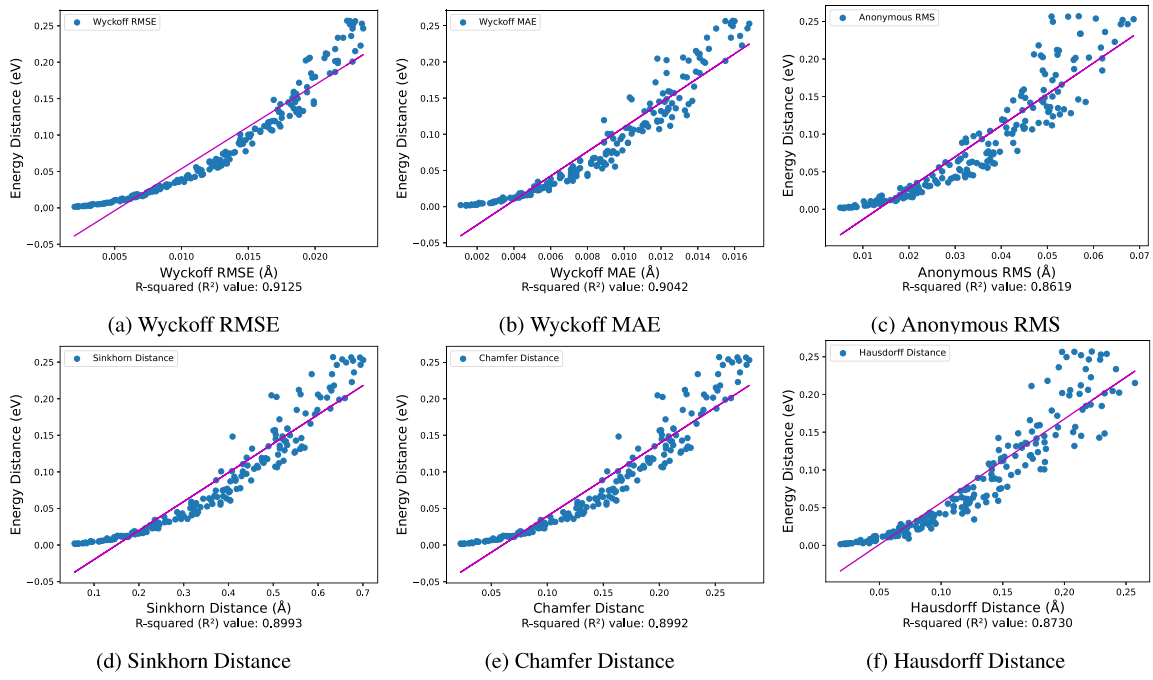


Fig. 3. Energy distances vs structure distances evaluated over the 200 random perturbation structures of $SrTiO_3$.

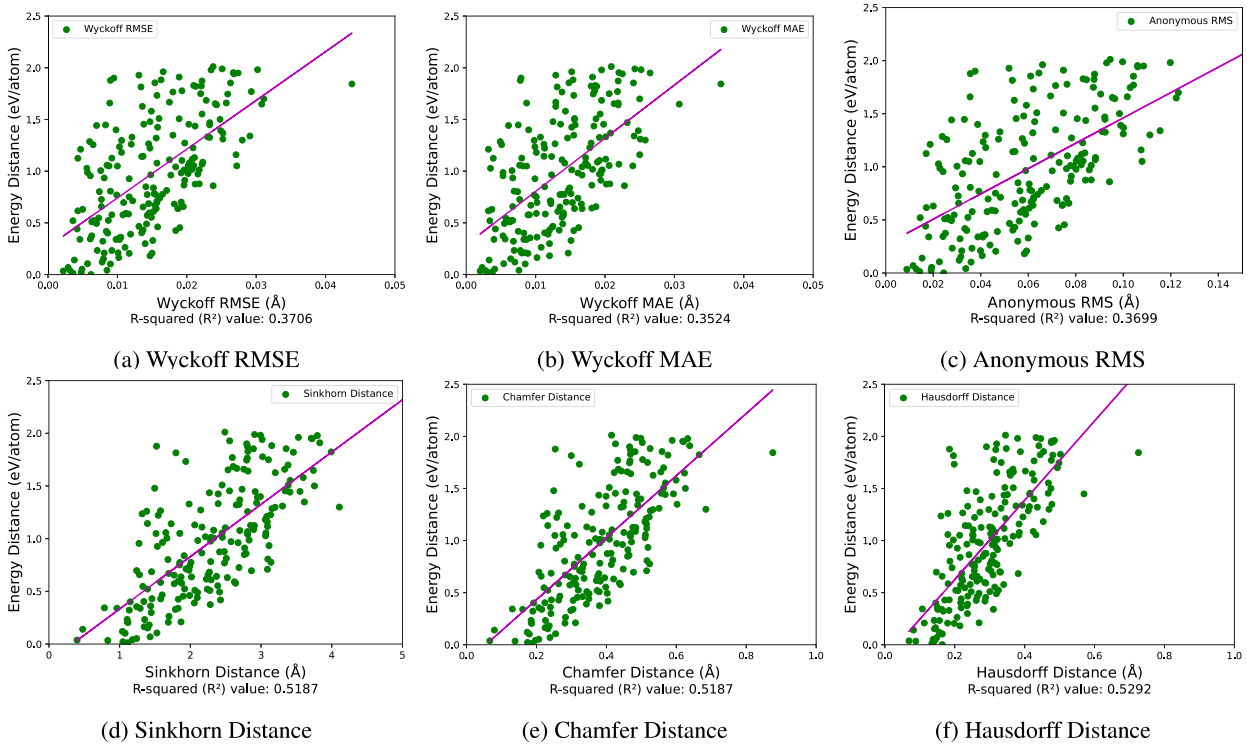


Fig. 4. Energy distances vs structure distances evaluated over 200 symmetrically perturbation structures of ZrSO.

distance of the random perturbation. All R-squared (R^2) values exceed 0.85, indicating a strong correlation between energy distances and structural distances. As depicted in Fig. 4(a)(b)(e), the R^2 values are 0.3706, 0.3524 and 0.3699, respectively. For the Sinkhorn distance, Chamfer distance and Hausdorff distance, the R^2 values are all above 0.51.

3.2. Using distance metrics to compare CSP algorithms

From the Material Project Database [61], we individually choose five crystal structures for each type of composition encompassing binary, ternary, and quaternary, which generates a total of fifteen test structures. We also added Ca_4S_4 (mp-1672) as one additional target.

Table 2
A metrics table generated by comparing the ground state structure of Ca_4S_4 with the best structures predicted by only the global minimum energy found by three different optimization algorithms from GN-OA: Random Acceleration Search (RAS), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). The units for XRD Spectrum Distance and OFM Distance are counts per second and the number of valence shell electrons, respectively.

Ca_4S_4			
Algorithm	RAS	BO	PSO
Formation Energy (eV)	−38.323	**−38.809**	−38.781
Energy Distance (eV)	3.407	**2.922**	2.949
Wyckoff RMSE (Å)	N/A	N/A	N/A
Wyckoff MAE (Å)	N/A	N/A	N/A
Anonymous RMS (Å)	N/A	N/A	N/A
RMS Distance (Å)	N/A	N/A	N/A
Sinkhorn Distance (Å)	3.2478	**2.9730**	2.9888
Chamfer Distance (Å)	**0.8120**	0.7432	0.7472
Hausdorff Distance (Å)	0.6268	**0.5738**	0.5768
Superpose RMSD (Å)	**1.8412**	2.3290	**1.8412**
Edit Graph Distance	**0**	**0**	**0**
FingerPrint Distance	**0**	**0**	**0**
XRD Spectrum Distance	2.8915	2.8651	**2.8636**
OFM Distance	0.1535	**0.1419**	0.1425

To compare the performance of three GN-OA algorithms [13] based on three optimization methods including random searching (RAS), Bayesian optimization (BO), and particle swarm optimization (PSO), we applied these three algorithms (GN-OA-RAS, GN-OA-BO, GN-OA-PSO) to the 16 targets, all of which use MEGNet neural network potential for energy calculation. For RAS and BO, we set the init population size to 200 and the total number of iterations to 20,000. For the PSO algorithm, we set the init population size to be 200 and the generation number is 100. All predicted structures should be locally optimized using a relaxer before comparing with the ground state structures, which is a requisite condition for calculating our proposed performance metrics. However, the GN-OA algorithm does not conduct local optimization for their output structures. So we follow their practice in our evaluation. We use M3GNet potential to evaluate the performance, which means the values of the formation energy and energy distance for Table 2 through Table 5 are computed by M3GNet. The formation energy signifies the M3GNet-predicted energy of the best minimum structure discovered by all three algorithms rather than the exact formation energy. As can be seen from the table and the corresponding crystal diagrams, the smaller the value of formation energy, the higher the similarity between the predicted crystal structure and the ground state structure.

Table 2 shows the distance metrics between the ground state structure of Ca_4S_4 and its predicted structures by three CSP algorithms, two of which (GN-OA-RAS and GN-OA-BO) successfully predict the ground-state structures within 5000 iteration steps. The computed distance metrics demonstrate similar performance across all measures. The energy distances for RAS, BO, and PSO are 3.407 eV, 2.922 eV, and 2.949 eV, respectively. Additionally, the XRD spectrum distance values of 2.8915 counts per second, 2.8651 counts per second, and 2.8636 counts per second, along with the corresponding OFM distance values of 0.1535, 0.1419, and 0.1425 for RAS, BO, and PSO, indicate highly consistent results. Notably, both the edit graph distance and the fingerprint distance exhibit perfect matches with the ground state with the values of 0 for each algorithm. Fig. 5 shows the comparison of the ground state with the predicted crystal structures of Ca_4S_4 by the RAS, BO, and PSO algorithms. Figs. 5(b), 5(c), and 5(d) exhibit a striking similarity in structure to Fig. 5(a).

As shown in Table 3, the distances to the ground structure for both BO and PSO results are nearly identical, with energy distances of 5.695 eV and 5.697 eV, respectively, compared to 68.721 eV for RAS. The structures of BO and PSO in Fig. 6(c) and Fig. 6(d) are more similar than the structure of RAS because they have almost the same

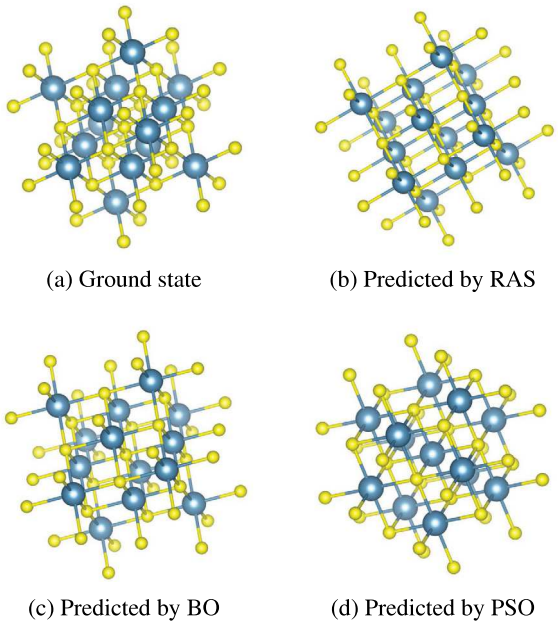


Fig. 5. Comparison of the ground state and predicted crystal structures of Ca_4S_4 by the RAS, BO, and PSO algorithms.

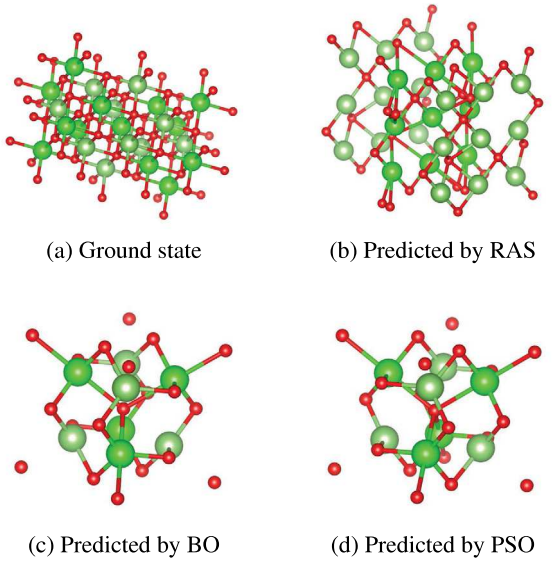


Fig. 6. Comparison of the ground state and predicted crystal structures of $Li_4Zr_4O_8$ by the RAS, BO, and PSO algorithms.

values of formation energy and structure distances. Fig. 6(b) shows a higher symmetry compared to Figs. 6(c) and 6(d), which may also be the reason why the Sinkhorn distance, the Chamfer distance, and the Hausdorff distances indicate better performance for RAS.

For the quaternary material $Li_3Ti_3Se_6O_3$ (Table 4), the BO algorithm achieves better performance in terms of most distance metrics. Among all the distance metrics, BO shows the best performance for the energy distance with a value of 19.227 eV compared to 93.858 eV and 106.608 eV of RAS and PSO respectively. In addition, BO outperforms RAS and PSO in terms of the more challenging indicators including Wyckoff RMSE and Wyckoff MAE, with values of 0.3464 Åand 0.2458 Å, respectively. Similarly, it has the best results in terms of the Sinkhorn distance, the Chamfer distance, and the Hausdorff distance metrics which are 23.6450 Å, 2.8741 Å, and 2.6054 Å. Fig. 7(c) shows the

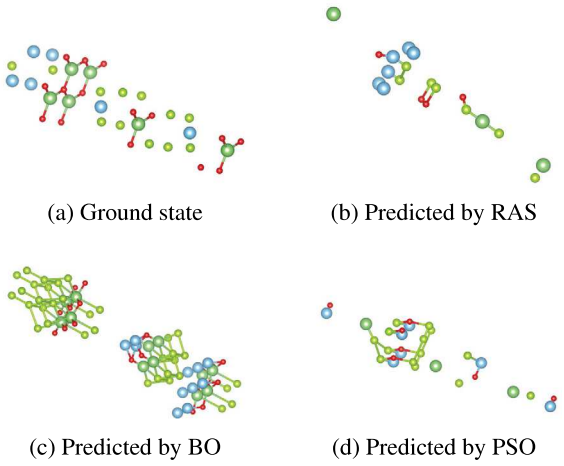


Fig. 7. Comparison of the ground state and predicted crystal structures of $Li_3Ti_3Se_6O_3$ by the RAS, BO, and PSO algorithms.

Table 3
Metric table generated by comparing the ground state structure of $Li_4Zr_4O_8$ with the best structures without relaxation, as predicted by three different optimization algorithms in GN-OA: Random Acceleration Search (RAS), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). The units for XRD Spectrum Distance and OFM Distance are counts per second and the number of valence shell electrons, respectively.

	$Li_4Zr_4O_8$		
Algorithm	RAS	BO	PSO
Formation Energy (eV)	−54.335	**−117.361**	−117.359
Energy Distance (eV)	68.721	**5.695**	5.697
Wyckoff RMSEE (Å)	N/A	N/A	N/A
Wyckoff MAEE (Å)	N/A	N/A	N/A
Anonymous RMSE (Å)	N/A	N/A	N/A
RMS DistanceE (Å)	N/A	N/A	N/A
Sinkhorn DistanceE (Å)	**31.7423**	37.2488	37.0778
Chamfer DistanceE (Å)	**3.0679**	3.5888	3.6062
Hausdorff DistanceE (Å)	**3.5578**	4.4537	4.6748
Superpose RMSDE (Å)	2.9837	**2.8446**	2.9019
Edit Graph Distance	**33**	36	36
FingerPrint Distance	3.5050	**2.8482**	**2.8482**
XRD Spectrum Distance	**1.7977**	2.0243	2.0243
OFM Distance	0.4376	**0.2690**	0.2691

predicted structure of BO, which has a more symmetrical structure compared to Figs. 7(b) and 7(d) predicted by RAS and PSO. The figures and tables presented above effectively demonstrate that our distance metrics accurately capture the differences between crystal structures with the ground states, highlighting that more symmetrical and stable structures tend to allow the CSP algorithms to achieve better performance.

Table 5 shows the distance metrics between the predicted structures by three algorithms and the ground state for the binary target structure $ScBe_5$. Generally, the generated structure needs local optimization relaxation before performance evaluation. Since GN-OA does not include this relaxation step, we showed their results in the Unrelaxed columns while also calculated distance metrics with M3GNet relaxation for final structures. For unrelaxed structures, we find that the PSO algorithm achieves the highest performance for all distance metrics. The energy distances are 12.495 eV, 23.614 eV, and 4.772 eV for RAS, BO, and PSO, respectively. The Sinkhorn distance, Chamfer distance, and Hausdorff distance of PSO are also significantly smaller than those of RAS and BO, which are 2.7904 Å, 0.9301 Å, and 1.2743 Å, respectively. For the predicted structures with M3GNet based relaxation, the quality of most of the structures is improved (see numbers in parentheses), meaning that most of the distance metrics to the ground truth structure are decreased. In particular, the relaxed predicted structure by the BO algorithm shows a decrease in formation energy from −2.633 eV to −25.849 eV, and a decrease in the energy distance from 23.614 eV to 0.397 eV. For the RAS algorithm, the structure after relaxing also shows smaller distance values. The energy distance for RAS is reduced from 12.495 eV to 5.561 eV. It should be noted that without relaxation, PSO showed the best performance by all metrics. However with relaxation, BO algorithm showed the best performance in terms of most metrics except Sinkhorn distance, Chamfer distance, Hausdorff distance and edit graph distance. Performance evaluation metrics of additional the 12 targets are shown in the Supplementary file Table S1 to S12.

Table 4
Metric table generated by comparing the ground state structure of $Li_3Ti_3Se_6O_3$ with the best structures without relaxation, as predicted by three different optimization algorithms in GN-OA: Random Acceleration Search (RAS), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). The units for XRD Spectrum Distance and OFM Distance are counts per second and the number of valence shell electrons, respectively.

	$Li_3Ti_3Se_6O_3$		
Algorithm	RAS	BO	PSO
Formation Energy (eV)	14.146	**−60.487**	26.896
Energy Distance (eV)	93.858	**19.227**	106.608
Wyckoff RMSE (Å)	0.3658	**0.3464**	0.3958
Wyckoff MAE (Å)	0.2669	**0.2458**	0.3240
Anonymous RMSE (Å)	N/A	N/A	N/A
RMS Distance (Å)	N/A	N/A	N/A
Sinkhorn Distance (Å)	31.9493	**23.6450**	30.4329
Chamfer Distance (Å)	3.3169	**2.8741**	3.5282
Hausdorff Distance (Å)	3.2332	**2.6054**	2.8125
Superpose RMSD (Å)	**6.7413**	7.3776	7.1555
Edit Graph Distance	**12**	15	14
FingerPrint Distance	2.6597	**2.0178**	2.1986
XRD Spectrum Distance	1.9977	**1.2185**	1.7125
OFM Distance	0.4486	0.6285	**0.4263**

Fig. 8(a) shows the ground state structure of $ScBe_5$, Fig. 8(b)–(d) are the unrelaxed structures by the algorithms RAS, BO and PSO and Fig. 8(e)–(g) are the predicted structures by algorithms RAS, BO and PSO with M3GNet relaxation, respectively. Fig. 8(d) demonstrates a more symmetrical and similar structure compared to Figs. 8(b) and 8 (c), which implies that the structure in Fig. 8(d) exhibits a more balanced distribution of atoms with reasonable pairwise atomic distances compared to the structures in Fig. 8(b–c). This higher similarity between the structure presented in Fig. 8(d) and the ground state structure 8(a) is reflected in all the distance values in Table 5 for unrelaxed structures. We also compared the structures with relaxation (Fig. 8(e,f,g)) and without relaxation (Fig. 8(b,c,d)). We find that the structures have all become more aggregated: for example, the relaxed structure in Fig. 8(e) to the relaxed structure in Fig. 8(b), which indicates that the structure tend to be fine-tuned into a more stable state after relaxation. The smaller values of Sinkhorn distance, Chamfer distance and Hausdorff distance of the structures predicted by PSO, Figs. 8(d) and 8(g) here indicate the higher similarity of the predicted structures against the ground state structure. Overall, we find it is desirable to add the relaxation operation to the predicted structures by the GN-OA algorithms.

It should be noted that there are situations in there exist multiple low-energy polymorph structures around or even away from the ground state structure for a given composition. In this case, an algorithm may predict a structure that has low energy with high structure distances. In this case, the structure is still a successful prediction. So it is advised that one should use the energy distance along with structure distances to holistically evaluate the quality of a given predicted structure.

3.3. Trajectory studies of the GN-OA search algorithms in CSP of targets without polymorph structures

Here we exploit the multi-dimensional CSP performance metrics to investigate the search behavior of three optimization algorithms used in the GN-OA CSP package [13]. These algorithms incorporate MEGNET global optimizations for structural relaxation. We applied their

Table 5
Metric table generated by comparing the ground state structure of $ScBe_5$ with the best structures without relaxation (Unrelaxed) and with relaxation (local optimization of) by M3GNet (Relaxed), as predicted by three different optimization algorithms in GN-OA: Random Acceleration Search (RAS), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). The units for XRD Spectrum Distance and OFM Distance are counts per second and the number of valence shell electrons, respectively.

	$ScBe_5$					
Algorithm	Unrelaxed			Relaxed		
	RAS	BO	PSO	RAS	BO	PSO
Formation Energy (eV)	−13.752	−2.633	**−21.474**	(−20.685)	**(−25.849)**	(−25.317)
Energy Distance (eV)	12.495	23.614	**4.772**	(5.561)	**(0.397)**	(0.930)
Wyckoff RMSE (Å)	N/A	N/A	N/A	N/A	N/A	N/A
Wyckoff MAE (Å)	N/A	N/A	N/A	N/A	N/A	N/A
Anonymous RMSE (Å)	N/A	N/A	**0.4535**	N/A	**(0.4405)**	0.4558
RMS Distance (Å)	N/A	N/A	**0.4535**	N/A	**(0.4405)**	0.4558
Sinkhorn Distance (Å)	45.1382	47.2978	**2.7904**	(38.7643)	(13.8632)	**(2.2861)**
Chamfer Distance (Å)	9.3233	10.9360	**0.9301**	(7.9629)	(3.4503)	**(0.7620)**
Hausdorff Distance (Å)	13.9499	11.3823	**1.2743**	(11.7346)	(3.2524)	**1.6396**
Superpose RMSD (Å)	1.6628	1.6343	**1.6952**	(1.6014)	**(1.5719)**	1.7038
Edit Graph Distance	7	8	**3**	(6)	(6)	**3**
FingerPrint Distance	2.3392	2.4184	**1.7783**	2.5490	**1.2869**	(1.6103)
XRD Spectrum Distance	2.2394	1.9878	**1.6491**	(2.0843)	**(1.4924)**	1.8258
OFM Distance	0.9898	1.1039	**0.4494**	(0.9719)	**(0.0877)**	(0.3971)

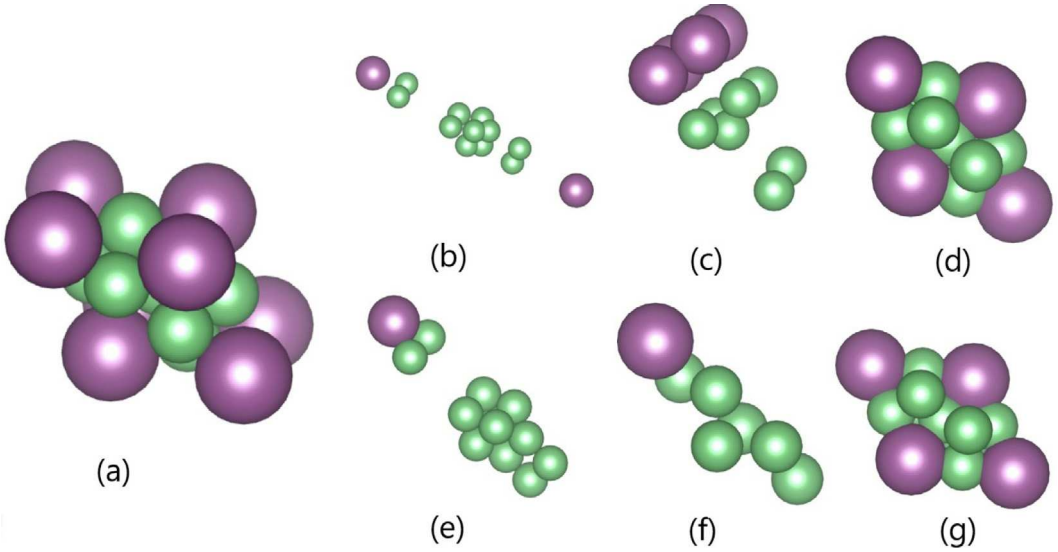
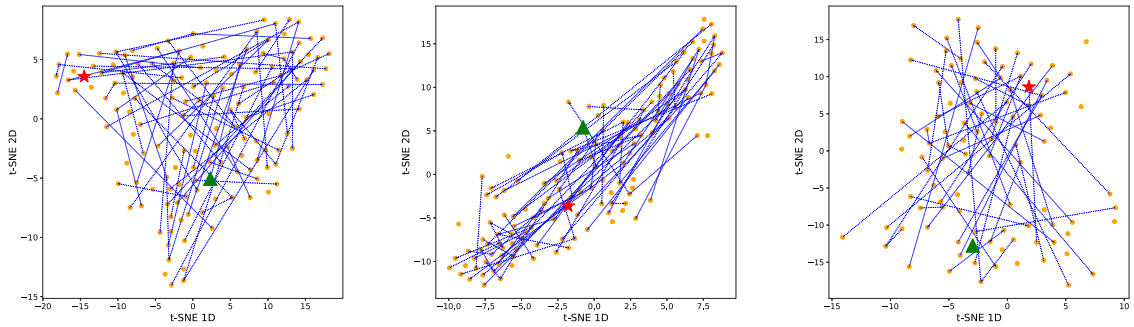


Fig. 8. Comparison of the ground state and predicted crystal structures of $ScBe_5$ by the RAS, BO, and PSO algorithms for Table 5. (a) Ground state structure. (b) Unrelaxed structure by the RAS. (c) Unrelaxed structure by the BO. (d) Unrelaxed structure by the PSO. (e) M3GNet relaxed structure by the RAS. (f) M3GNet relaxed structure by the BO. (g) M3GNet relaxed structure by the PSO.

random search (RAS), Bayesian optimization (BO), and particle swarm optimization (PSO) to search the structures of Ca_4S_4 and $Ba_3Na_3Bi_3$. The readability of a structure by Pymatgen package serves as a crucial determinant of its quality and suitability for further analysis using our neural network potential. Therefore, structures exhibiting overlapping or excessively crowded atomic arrangements may not only present issues in processing with Pymatgen but also render them unsuitable for energy calculations using our neural network model. Therefore, we define the structures that can be read by the Pymatgen package to be called valid structures. For Ca_4S_4, we allocated 5000 structure generations for the random search algorithm which generated 147 valid structures. We allocated 1000 structure generation steps for the BO algorithm which traversed 140 valid structures. For PSO, we used 5000 structure generation steps which created only 100 valid structures over the search process. For all the valid structures during a search, we calculated their distance metrics to the ground state structure and then mapped the distance features using t-SNE [62] to 2D dimension points. The purpose of using t-SNE is to map all data points from the high-dimensional space to the low-dimensional space while preserving the pairwise distances between the points. We then plot the trajectory of the structure search over time by connecting two consecutive points if the newer structure has lower energy than the previous one. The results are shown in Fig. 9. Note the green triangles indicate the starting points while the red stars represent the ground state structures, and the blue lines represent the routes of the search process. The x, y axes in the trajectory figures represent different dimensions (1D and 2D space).

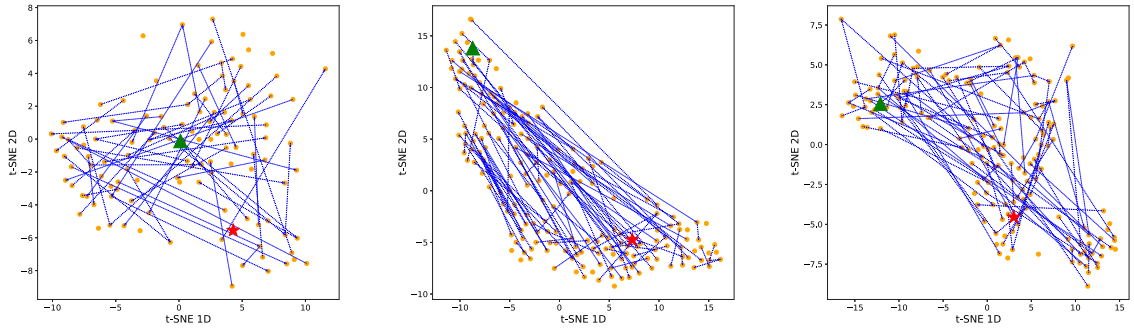
First, we found that for all three algorithms, it is challenging for them to generate valid structures during their search. Both random search and PSO generated less than 150 structures for a total of 5000 structure generations. In comparison, the BO algorithm generated 140 valid structures with only 1000 structure generations. This is consistent with the authors' observation of GN-OA that the BO has better performance in their CSP experiments. From Fig. 9(a) and (c), we found that the Random search and PSO algorithms tend to jump around in a larger design space. In contrast, the BO algorithm is more focused during its search (Fig. 9(b)).

We further applied the three search algorithms to the structure prediction of a ternary compound $Ba_3Na_3Bi_3$, which is more challenging than Ca_4S_4. Fig. 10 shows the three trajectories of the algorithms. First, we found that all three algorithms are much more difficult to generate valid structures, especially for the Random algorithm, which generates only 121 valid structures during its 50,000 tries. In contrast, the BO and



(a) Random search (b) Bayesian optimization (c) Particle swarm optimization

Fig. 9. Trajectories of three search algorithms in crystal structure prediction of Ca_4S_4: (a) Random search with 5000 structure generation steps. 147 valid structures found; (b) Bayesian Optimization with 1000 structure generations. 140 valid structures found; (c) PSO with 5000 structure generations with 100 valid structures. The green triangles indicate the starting points while the red star indicates the ground states, and the blue lines represent the routes of the search process.



(a) Random search (b) Bayesian optimization (c) Particle swarm optimization

Fig. 10. Trajectories of three search algorithms in crystal structure prediction of $Ba_3Na_3Bi_3$: (a) Random search with 50,000 structure generation steps. 121 valid structures found; (b) Bayesian Optimization with 3000 structure generations. 195 valid structures found; (c) PSO with 6000 structure generations with 177 valid structures. The green triangles indicate the starting points while the red star indicates the ground states, and the blue lines represent the routes of the search process.

PSO both generate 195 and 177 valid structures during their 3000 and 6000 structure generation steps, though the success rates of structure generation are still very low. The search trajectory patterns of the Random search and PSO are still more similar by jumping around a large area while the BO algorithm is more focused on their structure search. But compared to Fig. 9(b), the structure range in Fig. 10 is larger due to the higher complexity of the target structure $Ba_3Na_3Bi_3$. Our CSP metrics based trajectory analysis shows that current algorithms need to significantly improve their structure generation success rate to achieve higher efficiency in CSP.

4. Conclusion

Due to the complexity of structural changes during the search process of crystal structure algorithms, it is very difficult to measure the similarity of the candidate structures to the ground state, especially when the algorithms cannot find the exact solution. This is especially challenging when the candidate structure and the target structure can have different spatial symmetries (space groups). We find that it is infeasible to use a single structure similarity measure to describe the CSP prediction quality of different algorithms. By evaluating a set of thirteen structure distance measures (which we call CSPMetrics), we find that using them together allows us to achieve a quantitative method to measure the prediction quality of predicted crystal structures compared to the ground states. Application of our CSPmetric set has allowed us to gain interesting analysis of the structures during the search process of different CSP algorithms. While there are definitely rooms to further improve the metrics so that they can capture the prediction errors happening during CSP algorithm search, we believe our current CSPMetrics can be used as a good starting point to characterize benchmark different CSP algorithms. The availability of the source code additionally makes it easy for such evaluations.

CRediT authorship contribution statement

Lai Wei: Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Qin Li:** Investigation, Writing – original draft. **Sadman Sadeed Omee:** Investigation, Writing – original draft, Methodology. **Jianjun Hu:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code availability

The test crystal structures are downloaded from the Materials Project database at http://www.materialsproject.org. The source code can be found at https://github.com/usccolumbia/CSPBenchMetrics.

Acknowledgments

The research reported in this work was supported in part by National Science Foundation under the grant 2110033. The views, perspectives, and content do not necessarily represent the official views of the NSF. We would especially like to thank the reviewers for detailed constructive suggestions that have significantly improved the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.commatsci.2024.112802.

References

[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.

[2] James Alexander Chisholm, Sam Motherwell, COMPACK: a program for identifying crystal structure similarity using distances, J. Appl. Crystallogr. 38 (1) (2005) 228–231.

[3] R. Alex Mayo, Erin R. Johnson, Improved quantitative crystal-structure comparison using powder diffractograms via anisotropic volume correction, CrystEngComm 23 (40) (2021) 7118–7131.

[4] Jürgen Hafner, Ab-initio simulations of materials using VASP: Density-functional theory and beyond, J. Comput. Chem. 29 (13) (2008) 2044–2078.

[5] J. Christian Schön, Martin Jansen, First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization, Angew. Chem., Int. Ed. Engl. 35 (12) (1996) 1286–1304.

[6] Johann Christian Schön, Nanomaterials-what energy landscapes can tell us, Process. Appl. Ceram. 9 (3) (2015) 157–168.

[7] Stefan Goedecker, Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems, J. Chem. Phys. 120 (21) (2004) 9911–9917.

[8] Scott M. Woodley, Knowledge led master code search for atomic and electronic structures of LaF3 nanoclusters on hybrid rigid ion–shell model–DFT landscapes, J. Phys. Chem. C 117 (45) (2013) 24003–24014.

[9] Tian Xie, Jeffrey C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120 (14) (2018) 145301.

[10] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, Shyue Ping Ong, Graph networks as a universal machine learning framework for molecules and crystals, Chem. Mater. 31 (9) (2019) 3564–3572.

[11] Sadman Sadeed Omee, Steph-Yves Louis, Nihang Fu, Lai Wei, Sourin Dey, Rongzhi Dong, Qinyang Li, Jianjun Hu, Scalable deeper graph neural networks for high-performance materials property prediction, Patterns 3 (5) (2022) 100491.

[12] Chi Chen, Shyue Ping Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. 2 (11) (2022) 718–728.

[13] Guanjian Cheng, Xin-Gao Gong, Wan-Jian Yin, Crystal structure prediction by combining graph network and optimization algorithm, Nature Commun. 13 (1) (2022) 1492.

[14] Tomoki Yamashita, Shinichi Kanehira, Nobuya Sato, Hiori Kino, Kei Terayama, Hikaru Sawahata, Takumi Sato, Futoshi Utsuno, Koji Tsuda, Takashi Miyake, et al., CrySPY: a crystal structure prediction tool accelerated by machine learning, Sci. Technol. Adv. Mater. Methods 1 (1) (2021) 87–97.

[15] David C. Lonie, Eva Zurek, XtalOpt: An open-source evolutionary algorithm for crystal structure prediction, Comput. Phys. Comm. 182 (2) (2011) 372–387.

[16] William W. Tipton, Richard G. Hennig, A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials, J. Phys.: Condens. Matter 25 (49) (2013) 495401.

[17] Chris J. Pickard, R.J. Needs, High-pressure phases of silane, Phys. Rev. Lett. 97 (4) (2006) 045504.

[18] Chris J. Pickard, R.J. Needs, Ab initio random structure searching, J. Phys.: Condens. Matter 23 (5) (2011) 053201.

[19] Mads-Peter V. Christiansen, Nikolaj Rønne, Bjørk Hammer, Atomistic global optimization x: A Python package for optimization of atomistic structures, J. Chem. Phys. 157 (5) (2022) 054701.

[20] Xiangyu Yin, Chrysanthos E. Gounaris, Search methods for inorganic materials crystal structure prediction, Curr. Opin. Chem. Eng. 35 (2022) 100726.

[21] Lai Wei, Nihang Fu, Edirisuriya M.D. Siriwardane, Wenhui Yang, Sadman Sadeed Omee, Rongzhi Dong, Rui Xin, Jianjun Hu, TCSP: a template-based crystal structure prediction algorithm for materials discovery, Inorg. Chem. 61 (22) (2022) 8431–8439.

[22] Minoru Kusaba, Chang Liu, Ryo Yoshida, Crystal structure prediction with machine learning-based element substitution, Comput. Mater. Sci. 211 (2022) 111496.

[23] Jianjun Hu, Yong Zhao, Wenhui Yang, Yuqi Song, Edirisuriya Siriwardane, Yuxin Li, Rongzhi Dong, AlphaCrystal: Contact map based crystal structure prediction using deep learning, 2021, arXiv preprint arXiv:2102.01620.

[24] Colin W. Glass, Artem R. Oganov, Nikolaus Hansen, USPEX—Evolutionary crystal structure prediction, Comput. Phys. Commun. 175 (11–12) (2006) 713–720.

[25] Artem R. Oganov, Colin W. Glass, Crystal structure prediction using ab initio evolutionary techniques: Principles and applications, J. Chem. Phys. 124 (24) (2006) 244704.

[26] Artem R. Oganov, Yanming Ma, Andriy O. Lyakhov, Mario Valle, Carlo Gatti, Evolutionary crystal structure prediction as a method for the discovery of minerals and materials, Rev. Mineral. Geochem. 71 (1) (2010) 271–298.

[27] Detlef W.M. Hofmann, Joannis Apostolakis, Crystal structure prediction by data mining, J. Mol. Struct. 647 (1–3) (2003) 17–39.

[28] Yanchao Wang, Jian Lv, Li Zhu, Yanming Ma, CALYPSO: A method for crystal structure prediction, Comput. Phys. Comm. 183 (10) (2012) 2063–2070.

[29] Qunchao Tong, Lantian Xue, Jian Lv, Yanchao Wang, Yanming Ma, Accelerating CALYPSO structure prediction by data-driven learning of a potential energy surface, Faraday Discuss. 211 (2018) 31–43.

[30] Chang Liu, Hiromasa Tamaki, Tomoyasu Yokoyama, Kensuke Wakasugi, Satoshi Yotsuhashi, Minoru Kusaba, Ryo Yoshida, Shotgun crystal structure prediction using machine-learned formation energies, 2023, arXiv preprint arXiv:2305.02158.

[31] R. Hundt, J.C. Schön, M. Jansen, CMPZ–an algorithm for the efficient comparison of periodic structures, J. Appl. Crystallogr. 39 (1) (2006) 6–16.

[32] J.C. Schön, M. Jansen, Determination of candidate structures for simple ionic compounds through cell optimisation, Comput. Mater. Sci. 4 (1) (1995) 43–58.

[33] H. Putz, J.C. Schön, M. Jansen, Structure prediction for crystalline Ca3SiBr2 using an environment dependent potential, Z. Anorg. Allg. Chem. 625 (10) (1999) 1624–1630.

[34] Scott M. Woodley, Alexey A. Sokol, C. Richard A. Catlow, Structure prediction of inorganic nanoparticles with predefined architecture using a genetic algorithm, Z. Anorg. Allg. Chem. 630 (13–14) (2004) 2343–2353.

[35] Christopher C. Fischer, Kevin J. Tibbetts, Dane Morgan, Gerbrand Ceder, Predicting crystal structure by merging data mining with quantum mechanics, Nat. Mater. 5 (8) (2006) 641–646.

[36] Kuo Bao, Stefan Goedecker, Kenji Koga, Frédéric Lançon, Alexey Neelov, Structure of large gold clusters obtained by global optimization using the minima hopping method, Phys. Rev. B 79 (4) (2009) 041405.

[37] Giancarlo Trimarchi, Arthur J. Freeman, Alex Zunger, Predicting stable stoichiometries of compounds via evolutionary global space-group optimization, Phys. Rev. B 80 (9) (2009) 092101.

[38] S.Q. Wu, Min Ji, Cai-Zhuang Wang, Manh Cuong Nguyen, Xin Zhao, K. Umemoto, R.M. Wentzcovitch, Kai-Ming Ho, An adaptive genetic algorithm for crystal structure prediction, J. Phys.: Condens. Matter 26 (3) (2013) 035402.

[39] Anton O. Oliynyk, Lawrence A. Adutwum, Brent W. Rudyk, Harshil Pisavadia, Sogol Lotfi, Viktor Hlukhyy, James J. Harynuk, Arthur Mar, Jakoah Brgoch, Disentangling structural confusion through machine learning: structure prediction and polymorphism of equiatomic ternary phases ABC, J. Am. Chem. Soc. 139 (49) (2017) 17870–17881.

[40] Maximilian Amsler, Stefan Goedecker, Crystal structure prediction using the minima hopping method, J. Chem. Phys. 133 (22) (2010) 224104.

[41] Asma Nouira, Nataliya Sokolovska, Jean-Claude Crivello, Crystalgan: learning to discover crystallographic structures with generative adversarial networks, 2018, arXiv preprint arXiv:1810.11203.

[42] Evgeny V. Podryabinkin, Evgeny V. Tikhonov, Alexander V. Shapeev, Artem R. Oganov, Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning, Phys. Rev. B 99 (6) (2019) 064114.

[43] Xuecheng Shao, Jian Lv, Peng Liu, Sen Shao, Pengyue Gao, Hanyu Liu, Yanchao Wang, Yanming Ma, A symmetry-orientated divide-and-conquer method for crystal structure prediction, J. Chem. Phys. 156 (1) (2022) 014105.

[44] Xiangyang Liu, Haiyang Niu, Artem R. Oganov, COPEX: co-evolutionary crystal structure prediction algorithm for complex systems, npj Comput. Mater. 7 (1) (2021) 199.

[45] Yanchao Wang, Jian Lv, Pengyue Gao, Yanming Ma, Crystal structure prediction via efficient sampling of the potential energy surface, Acc. Chem. Res. 55 (15) (2022) 2068–2076.

[46] Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, et al., Molecular sets (MOSES): a benchmarking platform for molecular generation models, Front. Pharmacol. 11 (2020) 565644.

[47] Greg Landrum, et al., Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling, Greg Landrum 8 (2013).

[48] Harold T. Stokes, Dorian M. Hatch, FINDSYM: program for identifying the space-group symmetry of a crystal, J. Appl. Crystallogr. 38 (1) (2005) 237–238.

[49] A. Hannemann, R. Hundt, J.C. Schön, M. Jansen, A new algorithm for space-group determination, J. Appl. Crystallogr. 31 (6) (1998) 922–928.
[50] R. Hundt, J. Christian SchoÈn, A. Hannemann, M. Jansen, Determination of symmetries and idealized cell parameters for simulated structures, J. Appl. Crystallogr. 32 (3) (1999) 413–416.
[51] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, Gerbrand Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319.
[52] Marco Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, Adv. Neural Inf. Process. Syst. 26 (2013).
[53] Haoqiang Fan, Hao Su, Leonidas J. Guibas, A point set generation network for 3d object reconstruction from a single image, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 605–613.
[54] Daniel P. Huttenlocher, Gregory A. Klanderman, William J. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 850–863.
[55] GitHub - jewettaij/superpose3d: register 3D point clouds using rotation, translation, and scale transformations, 2022, https://github.com/jewettaij/superpose3d.
[56] Alberto Sanfeliu, King-Sun Fu, A distance measure between attributed relational graphs for pattern recognition, IEEE Trans. Syst. Man Cybern. (3) (1983) 353–362.
[57] Roy Jonker, Ton Volgenant, Improving the Hungarian assignment algorithm, Oper. Res. Lett. 5 (4) (1986) 171–175.
[58] Anthony Gillioz, Kaspar Riesen, Graph matching core, 2023.
[59] Nils E.R. Zimmermann, Anubhav Jain, Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity, RSC Adv. 10 (10) (2020) 6063–6081.
[60] Scott Fredericks, Kevin Parrish, Dean Sayre, Qiang Zhu, PyXtal: A Python library for crystal structure generation and symmetry analysis, Comput. Phys. Comm. 261 (2021) 107810.
[61] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013).
[62] Laurens Van der Maaten, Geoffrey Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).