Crystal structure prediction of materials with high symmetry using differential evolution

Wenhui Yang¹, Edirisuriya M Dilanga Siriwardane², Rongzhi Dong², Yuxin Li¹ and Jianjun Hu²,³,*

¹ School of Mechanical Engineering, Guizhou University, Guiyang 550025, People’s Republic of China
² Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29201, United States of America
E-mail: jianjunh@cse.sc.edu

Received 25 May 2021, revised 23 July 2021
Accepted for publication 13 August 2021
Published 31 August 2021

Abstract
Crystal structure determines properties of materials. With the crystal structure of a chemical substance, many physical and chemical properties can be predicted by first-principles calculations or machine learning models. Since it is relatively easy to generate a hypothetical chemically valid formula, crystal structure prediction becomes an important method for discovering new materials. In our previous work, we proposed a contact map-based crystal structure prediction method, which uses global optimization algorithms such as genetic algorithms to maximize the match between the contact map of the predicted structure and the contact map of the real crystal structure to search for the coordinates at the Wyckoff positions (WP), demonstrating that known geometric constraints (such as the contact map of the crystal structure) help the crystal structure reconstruction. However, when predicting the crystal structure with high symmetry, we found that the global optimization algorithm has difficulty to find an effective combination of WP that satisfies the chemical formula, which is mainly caused by the inconsistency between the dimensionality of the contact map of the predicted crystal structure and the dimensionality of the contact map of the target crystal structure. This makes it challenging to predict the crystal structures of high-symmetry crystals. In order to solve this problem, here we propose to use PyXtal to generate and filter random crystal structures with given symmetry constraints based on the information such as chemical formulas and space groups. With contact map as the optimization goal, we use differential evolution algorithms to search for non-special coordinates at the WP to realize the structure prediction of high-symmetry crystal materials. Our experimental results show that our proposed algorithm CMCrystalHS can effectively solve the problem of inconsistent contact map dimensions and predict the crystal structures with high symmetry.

Keywords: materials, crystal structure prediction, crystal materials, differential evolution, global optimization

(Some figures may appear in colour only in the online journal)
1. Introduction

Discovering novel materials with special properties has great impact over new product design such as high-capacity lithium batteries, superconductors, and solar panel materials with high efficiency. With the advancement of materials genome research, big data, and artificial intelligence, new materials have promoted the development of high-tech [1–3] and emerging industries. To speed up the discovery of new materials, materials informatics has emerged as a new material discovery model that uses data-driven methods [4–6] to build machine learning models for large-scale screening and generation of materials. There are several approaches for discovering new materials including first-principles [7–9] calculation based structural tinkering, inverse materials design [10, 11], generative machine learning models [11–15] and computational crystal structure predictions [1, 16–18].

The structure of a material determines its many properties. In the past, the way to obtain materials’ structures mostly relied on experimental x-ray diffraction [19]. However, the huge composition space of the materials makes it too inefficient to use these methods to design and explore new materials. Materials scientists have made breakthroughs in crystal structure prediction methods based on a combination of quantum mechanics and structure search algorithms [20], which has accelerated the design of new materials. However, the structure search space is still too huge for the structure prediction of complex compounds and existing crystal structure prediction methods are still inefficient. While global search algorithms combined with first-principle free energy calculations have discovered a series of new materials [20, 21], they usually rely on expensive DFT (density functional theory) calculations of the free energies of the sampled structures [22, 23] and can only be used for structure prediction of relatively small systems. To improve the performance of crystal structure prediction, various strategies such as clustering, intelligent mutation operators, and active learning [24, 25] have been proposed to improve sampling efficiency.

Recently, we have developed a generative machine learning model based on generative adversarial networks [12] which can generate new hypothetical inorganic materials formulas by learning the implicit chemical composition rules to form compounds. However, it is important to obtain their structures to evaluate their physicochemical properties. To do that, we propose a new machine learning based crystal structure prediction framework [18, 26], which first predicts the contact map or distance map of hypothetical materials and then use these maps to reconstruct the coordinates of the unit cell atoms. Compared with the crystal structure prediction methods based on global free energies optimization [20, 27], our methods take advantage of a large number of learned, hidden constraints, components and atomic configuration rules and constraints in the known crystal structures, which can make the sampling of the search space more effective.

In our previous work [18] on contact map-based crystal structure prediction, we use global optimization algorithms such as GA (genetic algorithm) and PSO (particle swarm optimization) to maximize the match between the contact map of the predicted structure and the contact map of the real crystal structure to search for the coordinates at the Wyckoff positions (WP) in the unit cell, proving that known geometric constraints (such as the contact map of the crystal structures) can help the crystal structure reconstruction. However, when predicting the crystal structure of materials with high structural symmetry (e.g. fcc crystal structures), we found that the global optimization algorithm usually faces the difficulty of dealing with the inconsistent dimensions of the contact map of the predicted crystal structure and the contact map of the target crystal structure due to the degenerate overlapping atomic positions after symmetric operations. The problem of inconsistent dimensions makes it challenging to predict high symmetric crystal structures. The crystal structures with high symmetry usually contain many special fractional coordinates (0, 1/6, 1/4, 1/3, 1/2, 2/3, 3/4, 5/6, −3/8, −1/4, −1/8, etc), and the multiplicity of these special positions is smaller than the multiplicity of general positions. The global optimization algorithm mainly searches for general positions, and it is difficult to search for these special positions as they require the precise special fractional coordinates. This will cause the WP combination searched by the global optimization algorithm to be incompatible with the chemical formula of the target crystal material, and the dimensions of the contact map are thus inconsistent, making it difficult to calculate the difference/error between the predicted contact map and the given target contact map. In our previous crystal structure prediction based on contact map, we can only predict the crystal structures of low symmetry, requiring that the multiplicity of WP is equal to the number of symmetry transformation operations and the structure does not contain positions with special fractional coordinates that may lead to degenerated positions. The resulting structure reconstruction algorithm thus has very limited coverage. In order to solve this problem, here we propose to use PyXtal [28] (a Python software package that helps crystal structure generation and crystal symmetry analysis) to generate random crystal structures with given symmetry constraints based on information such as chemical formula and space group. In order not to destroy the symmetry of the random crystal structure, we fixed the special coordinate values (0, 1/6, 1/4, 1/3, 1/2, 2/3, 3/4, 5/6, −3/8, −1/4, −1/8, etc), and only optimize the non-special coordinates in the WP using the differential evolution algorithm. Our new algorithm is named CMCrystalHS (contact map based structure prediction of crystals with high symmetry).

Our contribution in this paper can be summarized as follows:

- We identify the challenge of contact map based crystal structure prediction for highly symmetric structures due to the inconsistent dimensionality of the contact map caused by the degenerated atomic positions
- We propose a novel symmetric template generation and differential evolution algorithm based optimization to predict the crystal structures of high symmetry
- We evaluate our algorithms on different types of symmetric structures with extensive experiments to demonstrate
the effectiveness of our algorithm for reconstructing the 
crystal structures of high symmetry from the contact map.

2. Methods

2.1. Contact map based crystal structure prediction

Crystal structure information can be specified by its lattice 
constant, space group and element types and coordinates at the 
WP. Lattice constants are important parameters of crystalline 
materials, which are closely related to the binding energy 
between atoms. Changes in the lattice constants will affect 
the internal composition and force state of the crystal. Space 
group is a collection of all symmetry elements in the internal 
structure of a crystal. The space group type and symmetry 
type reflect the symmetry of the atoms in the internal crys-
tal structure. There are 230 space groups among all crystal 
structures. The WP are used to indicate the symmetry of the 
equivalent atoms in the unit cell. The contact map records the 
connection information between all atoms in the crystal struc-
ture and captures the interaction between atoms. The contact 
map is related to unpaired electrons, ionization energies, dipole 
polarizability, Pauling electronegativity, etc.

In our previous work [18], we developed a contact map 
based crystal structure prediction framework in which the 
space group, the lattice constants, and contact map of atoms 
in the unit cells can all be predicted using data driven machine 
learning models [18, 29–31]. With these information, we have 
demonstrated that it is possible to reconstruct the coordinates 
of the WP using a global optimization algorithm [18]. The 
optimization algorithm reconstructs the crystal structure by 
searching the WP by maximizing the match between the con-
tact map of the predicted structure and the contact map of the 
target crystal structure.

For crystal structures with high symmetry, however, it is dif-
ficult for the optimization algorithm to find an effective com-
bination of WP that satisfy the chemical formula and have the 
same dimension of the target contact map, which leads to the 
failure of the structure prediction. As is shown in figure 1, for 
Ce$_2$As$_2$O$_6$ whose space group number is 11, the dimension of 
the contact map of the predicted crystal structure is 16 $\times$ 16 for 
a candidate structure while the dimension of the real/final con-
tact map is 10 $\times$ 10. The reason is that the target structure has 
special coordinates with high symmetry (multiple symmetry 
operations over the WP may map to the same atomic positions) 
while the structure under optimization has non-special coordi-
nates which will be mapped to different atom positions after 
symmetric operations. The mismatched contact map dimen-
sions will make it difficult to calculate meaningful distance 
between two contact maps.

2.2. Prediction of crystal structures of high symmetry based 
on contact map

Here we propose a novel approach for reconstructing crys-
tal structures of materials with high symmetry. Our predic-
tion framework is shown in figure 2. First, we use the PyXtal 
library to generate 50 random crystal structures of high sym-
metry based on the material formula and space group. Next, a 
suitable random crystal structure of high symmetry is selected 
based on its distance to the target contact map. Finally, the 
selected template structure will be optimized by the differen-
tial evolution algorithm to search for appropriate non-special 
coordinates that can minimize the difference of the predicted 
contact map and the target contact map. The prediction per-
formance will be evaluated using the accuracy of the contact 
map, the root mean square distance (RMSD) and the mean 
absolute error (MAE) between the predicted WP of the crys-
tal structure and the WP of the target structure to evaluate 
the reconstructed crystal structure. The purpose of our experi-
ment is to verify whether our high symmetry crystal structure 
prediction algorithm is feasible, so we use the chemical for-
mula, contact map, space group, lattice constants of the target 
structure. However, in actual situations, the contact map, space 
group and lattice constants are all predicted for a given com-
position or material formula, which themselves may contain 
certain degree of errors, which may affect our reconstruction 
algorithm’s performance.

2.3. Generating random crystal structures

Random structure generation with given symmetry has been 
used to generate candidate structures which are then used to 
find most similar known structures in databases using a radial 
distribution function based fingerprint [32]. PyXtal [28] is a 
Python software package that helps design the material struc-
ture with a certain symmetry constraint, which is useful for 
crystal structure prediction and crystal symmetry analysis. It 
can generate atomic structures for a given symmetry and stoi-
chiometry and do geometric optimization. The PyXtal library 
uses the space group and its WP as a template, inserting atoms 
(or molecules) at the Wyckoff position one at a time until the 
given stoichiometry is met, thereby generating random sym-
metrical crystals. In this way, the correct symmetry can be 
obtained without adjusting the positions of the atoms. During 
this process, it also restricts the atoms from getting too close. 
The following are the steps for PyXtal to generate a random 
crystal, as is shown in the figure 3:

(a) Get user input parameters, such as chemical formula, 
space group, etc. The user can also define the lattice and 
the allowable distance between atoms (suitable for high 
pressure).

(b) Compatibility check of stoichiometry and space group. 
For different space groups, the WP have different mul-
tiplicity. Therefore, some atomic numbers may be incompa-
patible with the space group. PyXtal searches for all possible 
Wyckoff position combinations within the stoichiometric 
range, finds a valid combination from them, and fails to try to 
generate a random crystal structure if it cannot be found.

(c) Generate a random lattice corresponding to the space 
group. The random crystal structure generation of PyX-
tal starts from the selection of the unit cell. The symmetry
group matches a specific type of lattice. To avoid ambiguity, all crystals use conventional unit cell selection. Triclinic cells are the most common case, and other cell types can be obtained by using various constraints.

(d) Start by placing atoms in the WP, inserting one atomic specie at a time. The crystal structure in PyXtal is mainly composed of WP. After determining the space group and lattice, you can grow the structure by inserting one Wyckoff position at a time. PyXtal starts with the largest available WP (general position of the space group). When the required number of atoms is equal to or greater than the size of the regular position, the algorithm continues, otherwise the next largest Wyckoff position (or Wyckoff position set) is selected in descending order.

(e) If the minimum distance of an atom to other atoms at a single Wyckoff position is less than a certain tolerance (based on the atomic species), the Wyckoff position is merged into a smaller special position.

(f) If the newly generated Wyckoff position is too close to the previously generated position, a new generation point is selected within the same original Wyckoff position.

(g) Continue to insert WP and species until the correct stoichiometry is met. If any step in the above steps III–VI fails, the operation will be repeated before the maximum number of attempts is reached.

In our algorithm, after a random crystal structure is successfully generated, we will fix the special coordinates of corresponding WP positions and then try to search the non-special coordinates to maximize the match of its contact map with the target contact map.

2.4. 3D crystal structure reconstruction algorithm

Contact map based crystal structure prediction problem can be mapped as a global optimization problem. Here we select Nevergrad [33], which is an open source platform for derivative-free optimization. It contains various optimizers (such as DE (differential evolution algorithm), PSO (particle swarm optimization), etc), and supports multi-objective optimization and handles constraints. Here, we choose the differential evolution (DE) optimizer to search for the WP guided by a given contact map to perform the prediction of its corresponding crystal structure.

Differential evolution algorithms (DE) [34] are a population-based adaptive global optimization algorithms, which are simple, efficient, and converge quickly. The basic idea of the algorithm is: first, generate a random initial population, and sum the vector difference of any two individuals in the population with the third individual to generate a new individual. If the fitness of the new individual is better than the current individual, replace the old individual with the new individual in the next generation, otherwise the old individual will remain. The optimal solution is approached by continuous iteration.

For all the DE based experiments, we set the population size to 100 and the number of generations to 1000 with a cross-over probability of 0.5 and $F_1 = 0.8, F_2 = 0.8$. 

Figure 1. The dimensions of the predicted contact map and the target contact map are inconsistent due to the special coordinates of WP of highly symmetric crystal structures.

Figure 2. Prediction framework of high-symmetry crystal structures based on contact map.
2.5. Evaluation metrics

The contact map $M$ of the crystal structure is converted from the distance matrix in the pymatgen library according to the distance threshold between atoms in the VESTA software, for each pair of atoms in the unit cell, if their distance is within the range of $[\text{min.length}, \text{max.length}]$, then there is a bond between them and the corresponding contact map position $M[i, j]$ is set to 1, if not it is set to 0. The objective function of the differential evolution algorithm is as follows:

$$\text{fitness}_{\text{opt}} = \frac{2|A \cap B|}{|A| + |B|} \approx \frac{2 \times A \bullet B}{\text{Sum}(A) + \text{Sum}(B)},$$  \hspace{1cm} (1)

where $A$ is the predicted contact map and $B$ is the true contact map of a given composition, both only contain 1/0 entries. $A \cap B$ denotes the common elements between $A$ and $B$, $|g|$ represents the number of elements in a contact map, $\bullet$ denotes dot product, $\text{Sum}(g)$ is the sum of all contact map elements. The fitness essentially measures the overlap of two contact map samples, with values ranging from 0 to 1 with 1 indicating perfect overlap. We call this performance measure as contact map accuracy too.

To evaluate the reconstruction performance of DE, we can use the contact map accuracy as one evaluation criterion. Additionally, we define the RMSD and MAE of two crystal structures as below:

$$\text{RMSD}(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (v_i - w_i)^2}$$

$$= \frac{1}{n} \sum_{i=1}^{n} (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2$$  \hspace{1cm} (2)

$$\text{MAE}(v, w) = \frac{1}{n} \sum_{i=1}^{n} |v_i - w_i|$$

$$= \frac{1}{n} \sum_{i=1}^{n} (|v_{ix} - w_{ix}| + |v_{iy} - w_{iy}| + |v_{iz} - w_{iz}|) ,$$  \hspace{1cm} (3)

where $n$ is the number of independent atoms in the true crystal structure. For symmetrized structures, $n$ is the number of independent atoms of the set of Wyckoff equivalent positions. For regular structures, it is the total number of atoms in the structure. $v_i$ and $w_i$ are the corresponding atoms in the predicted crystal structure and the true crystal structure.

It should be pointed out that in the experiments of this study, the only constraints for the optimization is the contact map, it is possible that the predicted atom coordinates are oriented differently from the target atoms in terms of coordinate systems. To avoid this complexity, we use the spatial position error between the WP of the predicted structure and the WP of the target structure based on the space group. The specific process is: the equivalent positions of predicted WP are obtained by the symmetry operations corresponding to the space group. Since some equivalent positions are outside the unit cell, we will translate the equivalent positions outside the unit cell to the unit cell (the pymatgen library is also processed in this way), the point closest to the corresponding target Wyckoff position is selected from the equivalent positions of each predicted Wyckoff position, and calculate the RMSD and MAE between the selected equivalent positions and the target WP. The calculation method can effectively calculate the relative spatial position error between the predicted structure and the target structure.

3. Results

3.1. Generating and screening random crystal structures with a given symmetry

First we would like to show the quality of random crystal generation given the composition and space group. We selected a set of crystal structures from the materials project database as our test cases with the space group ranging from 42 to 221. We use PyXtal to generate 50 random crystal structures for each chemical formula and its space group of the target crystal material. PyXtal can generate a random crystal lattice consistent with the space group, but the error may be larger. We can set the crystal lattice to that of the target crystal structure when
generating the random crystal structure. Since PyXtal can generate many types of random crystal structures, how to select a suitable crystal structure from the 50 random crystal structures as the template structure for further DE based coordinate optimization guided by the contact map is a key issue. During the experiment, we found that PyXtal can generate many different random crystal structures of WP combinations according to a chemical formula and its space group. The number and multiplicity of WP in some random crystal structures are inconsistent with the target crystal structure, but the number of their total atoms is consistent. In these cases, the structures can also be optimized based on the contact map to improve the accuracy of the contact map, but the random crystal structure WP after optimization may have a larger error compared with the target crystal structure.

We use the following principles to select a suitable random crystal structure from the generated 50 random crystal structures for optimization:

- The number of WPs is the least (that is, the multiplicity of each WP is as large as possible)
- The multiplicity of the WP of each atom is arranged in descending order (meet the requirements of standard cif files)
- The random crystal structure with the largest contact map accuracy is selected from the random crystal structures that meet the above two conditions (in general, the atomic connection of random crystal structure with high contact map accuracy is more consistent with the target structure and easier to optimize)

The WP combination of the random crystal structure screened by this method is likely to be consistent with the WP combination of the target crystal structure.

Figure 4 shows material La$_3$Al$_1$N$_1$ with its real structure and corresponding random structure generated by PyXtal. The random structure have the contact map accuracy of 100% selected out of 50 random crystal structures. The random crystal structure is consistent with the atom connections of the target crystal structure. The random crystal structure’s WP combination of La$_3$Al$_1$N$_1$ is consistent with the target crystal structure, which shows that when WP error is small and it is easier to obtain high contact map accuracy.

3.2. Prediction results of crystal structure based on contact map

To evaluate the crystal structure reconstruction performance of our CMCrystalHS algorithm, we select 10 test cases from the materials project database as shown in table 1. The space group numbers range from 42 to 221 and the number of WP are between 2 and 5. These structures also have different number atoms ranging from 5 to 26 and different number of non-special coordinates from 0 to 13. For all the DE optimizations, the running time ranges from 300 s to 1000 s, depending on the complexity of the structures (the number of non-special coordinates to optimize).

Table 2 shows the crystal structure reconstruction performance of our algorithm. For all the test targets, we report the predicted structure accuracy only using 50 random structures generation (before optimization) and after DE based coordinate optimization. First we find that by using only PyXtal based random structure generation, the contact map accuracy can reach as high as 100% for La$_3$Al$_1$N$_1$, which is a highly symmetric cubic $Pm3n$ structure (space group 221) and contain only special coordinates. For this target, no improvement is possible by our DE algorithm. For the remaining targets, the contact map accuracy range from 0.593 to 0.818 where the lowest accuracy is from In$_2$I$_6$O$_{18}$, which has the largest number of non-special coordinates and the maximum number of atoms in the unit cell. Accordingly, the RMSD errors range from 0.0 to 0.309 and the MAE errors are between 0.0 and 0.183.

In the middle optimized columns, we show the prediction accuracy after applying contact map guided DE based WP coordinate optimization. We find that for four targets(Ce$_5$Se$_8$, Fe$_4$O$_4$F$_4$, K$_7$Ti$_6$Se$_8$, and Li$_2$Cr$_2$O$_4$), we have increased their contact map accuracy to 100% with 25%, 49.93%, 49.93%, and 33.33% improvement respectively, which significantly demonstrates the effectiveness of contact map guided coordinate reconstruction for crystal structure prediction. The other
remaining 5 targets, the contact map accuracy improvements are between 6.67% and 37.44%. At the same time, the WP coordinates’ RMSD errors have also been reduced by 1.05% to 39.67%. In terms of MAE errors, there is only one case that the DE optimization increases the error for In$_2$I$_6$O$_{18}$ by 5.59%. For all other cases, the MAE errors have been reduced by significantly ranging from 6.88% to 40.94%. As shown in table 1, In$_2$I$_6$O$_{18}$ is a very complex structure, with the largest number of WP 5, the largest number of atoms in unit cell 26, and the largest number of non-special coordinates 13. It should be noted that although smaller RMSD and MAE help to obtain higher contact map accuracy in most cases, this is not the decisive factor. From our extensive experiments, we find that there are several factors that affect the crystal structure prediction performance of our algorithm such as the number of WP, the number of atoms in the unit cell, the space group, the number of bonds/topology constraints and etc. Therefore, for In$_2$I$_6$O$_{18}$, although the contact map accuracy of the selected random crystal structure is improved by 37.44% after DE optimization, RMSD is only reduced by 1.05%, and even MAE is increased by 5.59%.

To further demonstrate how the DE-based optimization helps to reconstruct the crystal structures, we show the three case studies in figure 5. Figures 5(a)–(c) shows the prediction results for Fe$_4$O$_4$F$_4$ with a space group 84. The symmetry-guided random structure generation creates a structure with only 66.7% contact map accuracy compared to the true structure with many atoms packed together. After DE-based non-special coordinates optimization guided by the contact map, the contact map accuracy is increased to 100% and its structure is very close to the true structure in (c). The DE-optimization also helps the random structure of K$_1$Ti$_6$Se$_8$ generated by PyXtal in figure 5(d) to be fine-tuned into 100% contact map accuracy by unpacking the clustered atoms in (d). More dramatic structural change is also possible to increase the contact map accuracy as shown in figures 5(g) and (h).

Moreover, we relaxed predicted structures using DFT (density functional theory) to optimize the lattice constants and the atomic positions using the same procedure as in [35] except that the energy cutoff value is set as 400 eV. The DFT fine-tuned structures are shown in figure 6. For Fe$_4$O$_4$F$_4$, the DFT relaxation procedure has found a lower-energy structure compared to our target structure, which leads to low contact map accuracy. For K$_1$Ti$_6$Se$_8$, the DFT fine-tuning has reduced MAE error by 35.29% and increased the RMSD error by 16.67%. Figure 6(b) shows that the fine-tuned structure is more similar to the target structure. For Li$_2$Cr$_2$O$_4$, the DFT relaxation increases the RMSD by 1.65% and MAE by 12.41%.
Figure 5. Experimental results of crystal structure prediction based on contact map.
The small changes indicate that our predicted structure is very close to the DFT fine-tuned structures.

4. Discussion

Starting with the generated highly symmetric random crystal structures, our algorithm takes the contact map as the optimization target, searches for non-special coordinates at the WP, and achieves successful predictions of a set of high-symmetry crystals based on the contact map with high contact map accuracy, and low RMSD and MAE errors between the predicted WP and the real WP. During the experiment, we found that for crystal structures with extremely high symmetry (e.g. space groups 221, 225), the WP of the random crystal structure generated by PyXtal contain many special coordinates. While it is easy to obtain high contact map accuracy for these targets, the space for optimization is small and it is difficult to improve the accuracy of the contact map. One solution to address this issue is to generate more random crystal structures, and then screen out suitable random crystal structures with high contact map accuracy. Another notable issue is that in current experiments, we used the real contact map, space group and other information of the target structures in our crystal structure reconstruction. However, in actual situations, the contact map, space group and other information are all predicted for a given composition or material formula, which themselves may contain certain degree of errors, which may affect our reconstruction algorithm’s performance. At present, the space group [29, 36] and lattice constants [30, 37] of hypothetical crystal structures can be predicted, and the contact map of a given material formula can also be predicted by the deep learning prediction method [31]. While these methods are still in its emerging stage, they have the potential to achieve high precision crystal structure prediction for any given composition when combined with DFT based fine-tuning.

5. Conclusion

In our previous work, we proposed a contact map-based crystal structure prediction framework, which uses global optimization algorithms such as GA and PSO to maximize the match between the predicted contact map of the structure and the contact map of the real crystal structure to search for the coordinates at WP. We showed that geometric constraints such as the contact map or distance matrix of the crystal structure can be utilized for crystal structure reconstruction. However, our contact map-based crystal structure prediction algorithm has difficulty to be used for reconstructing the structures of high symmetry crystal materials due to the challenges of obtaining consistent contact map dimension and the difficulty to find exact special coordinates. Here, we propose a new algorithm CMCrystalHS to use PyXtal to generate a random crystal structure with given symmetry constraints based on information such as chemical formulas and space groups. With the
contact map as the optimization goal, only the non-special coordinates at the WP are optimized to solve the high symmetry crystal structure prediction problem. We use the contact map accuracy, the RMSD and the MAE between the predicted WP of the crystal structure and the WP of the target structure to evaluate the reconstructed high symmetry crystal structures. The experimental results show that by using random crystal structures with given symmetry constraints to search for non-special coordinates with the contact map as the optimization target, our proposed method for predicting highly symmetrical crystal structures can achieve high contact map accuracy and the RMSD and MAE between the predicted WP and the real WP are relatively small. This demonstrates our algorithm can be used to solve high-symmetric crystal structure prediction problems based on predicted contact map. When combined with contact map prediction and composition generation algorithms, it will enable large-scale crystal structure prediction and new materials discovery.

Author’s contribution

Conceptualization, JH; methodology, JH, WY; software, WY, JH; validation, WY, ES, JH; investigation, JH, WY, YL, ES, RD; resources, JH; data curation, JH, and WY; writing—original draft preparation, YZ and JH; writing—review and editing, JH, WY; visualization, WY and RD; supervision, JH; funding acquisition, JH.

Acknowledgments

Research reported in this work was supported in part by NSF under Grants 1940099 and 1905775. The views, perspective, and content do not necessarily represent the official views of NSF.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Jianjun Hu  https://orcid.org/0000-0002-8725-6660

References

[1] Kvashnin A G, Allahyari Z and Oganov A R 2019 Computational discovery of hard and superhard materials J. Appl. Phys. 126 043901
[2] He J, Yao Z, Hegde V I, Naghavi S S, Shen J, Bushick K M and Wolverton C 2020 Computational discovery of stable heteroanionic oxychalcogenides ABXO (A, B = metals; X = S, Se, and Te) and their potential applications Chem. Mater. 32 8229
[3] Brown N, Erhl P, Lewis R, Luksh T, Reker D and Schneider N 2020 Artificial intelligence in chemistry and drug design Chem. Rev. 120 8066–129
[4] Gómez-Bombarelli R et al 2018 Automatic chemical design using a data-driven continuous representation of molecules ACS Cent. Sci. 4 268
[5] Jablonka K M, Ongari D, Moosavi S M and Smit B 2020 Big-data science in porous materials: materials genomics and machine learning Chem. Rev. 120 8066–129
[6] Hoffmann J, Maestri B, Sawada Y, Tung S J, Sellier J M and Bengio Y 2019 Data-driven approach to encoding and decoding 3D crystal structures (arXiv:1909.00949)
[7] Curtis F, Li X, Rose T, Vázquez-Mayagoitia Á, Bhattacharya S, Ghiringhelli L M and Marom N 2018 Gato: a first-principles genetic algorithm for molecular crystal structure prediction J. Chem. Theory Comput. 14 2246
[8] Woodley S M and Catlow R 2008 Crystal structure prediction from first principles Nat. Mater. 7 937
[9] Maddox J 1988 Crystals from first principles Nature 335 201
[10] Zunger A 2018 Inverse design in search of materials with target functionalities Nat. Rev. Chem. 2 0121
[11] Kim B, Lee S and Kim J 2020 Inverse design of porous materials using artificial neural networks Sci. Adv. 6 eaaax9324
[12] Dan Y, Zhao Y, Li X, Li S, Hu M and Hu J 2020 Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials npj Comput. Mater. 6 84
[13] Bradshaw J, Paige B, Kusner M J, Segler M and Hernández-Lobato J M 2019 A model to search for synthesizable molecules Advances in Neural Information Processing Systems pp 7373–49
[14] Noh J, Kim J, Stein H S, Sanchez-Lengeling B, Gregoire J M, Aspuru-Guzik A and Jung Y 2019 Inverse design of solid-state materials via a continuous representation Matter 1 1370
[15] Ren Z et al 2020 Inverse design of crystals using generalized invertible crystallographic representation (arXiv:2005.07609)
[16] Glass C W, Oganov A R and Hansen N 2006 USPEX-evolutionary crystal structure prediction Comput. Phys. Commun. 175 713
[17] Ryan K, Lengyel J and Shatruk M 2018 Crystal structure prediction via deep learning J. Am. Chem. Soc. 140 10158
[18] Hu J, Yang W, Dong R, Li Y, Li X, Li S and Sirivardane E M D 2021 Contact map based crystal structure prediction using global optimization CrystEngComm 23 1765
[19] Oviedo F et al 2019 Fast and interpretable classification of small x-ray diffraction datasets using data augmentation and deep neural networks npj Comput. Mater. 5 60
[20] Oganov A R, Lyakhov A O and Valle M 2011 How evolution of chemical composition space for inverse design of inorganic materials npj Comput. Mater. 6 84
[21] Wang Y, Lv J, Li Q, Wang H and Ma Y 2020 CALYPSO method for structure prediction and its applications to materials discovery Handbook of Materials Modeling: Applications: Current and Emerging Materials (Cham: Springer) p 2729
[22] Zhang L, Wang Y, Lv J and Ma Y 2017 Materials discovery at high pressures Nat. Rev. Mater. 2 17005
[23] Oganov A R, Pickard C J, Zhu Q and Needs R J 2019 Structure prediction drives materials discovery Nat. Rev. Mater. 4 331
[24] Podryabinkin E V, Tikhonov E V, Shafeev A V and Oganov A R 2019 Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning Phys. Rev. B 99 064114
[25] Leaman T, Balachandran P V, Xue D and Yuan R 2019 Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design npj Comput. Mater. 5 21
[26] Hu J, Yang W and Dilanga Siriwardane E M 2020 Distance matrix-based crystal structure prediction using evolutionary algorithms J. Phys. Chem. A 124 10909–19

[27] Lyakhov A O, Oganov A R, Stokes H T and Zhu Q 2013 New developments in evolutionary structure prediction algorithm USPEX Comput. Phys. Commun. 184 1172

[28] Fredericks S, Parrish K, Sayre D and Zhu Q 2021 Pyxtal: a python library for crystal structure generation and symmetry analysis Comput. Phys. Commun. 261 107810

[29] Zhao Y, Cui Y, Xiong Z, Jin J, Liu Z, Dong R and Hu J 2020 Machine learning-based prediction of crystal systems and space groups from inorganic materials compositions ACS Omega 5 3596

[30] Li Y, Yang W, Dong R and Hu J 2020 MLatticeABC: generic lattice constant prediction of crystal materials using machine learning (arXiv:2010.16099)

[31] Hu J, Zhao Y, Yang W, Song Y, Siriwardane E, Li Y and Dong R 2021 Alphacrystal: contact map based crystal structure prediction using deep learning (arXiv:2102.01620)

[32] Lee I H and Chang K J 194 Crystal structure prediction in a continuous representative space Comput. Mater. Sci. 6 11585–94

[33] Rapin J and Teytaud O 2018 Nevergrad—a gradient-free optimization platform https://GitHub.com/FacebookResearch/Nevergrad

[34] Price K, Storn R M and Lampinen J A 2006 Differential Evolution: A Practical Approach to Global Optimization (Berlin: Springer)

[35] Song Y, Siriwardane E M D, Zhao Y and Hu J 2020 Computational discovery of new 2D materials using deep learning generative models (arXiv:2012.09314)

[36] Cruz Cabeza A J, Pidcock E, Day G M, Motherwell W D S and Jones W 2007 Space group selection for crystal structure prediction of solvates CrystEngComm 9 556

[37] Amar M N, Ohriga M A, Seghier M E A B and Ouasr H 2020 Prediction of lattice constant of A$_2$XY$_6$ cubic crystals using gene expression programming J. Phys. Chem. B 124 6037