

## ARTICLE OPEN



# COPEX: co-evolutionary crystal structure prediction algorithm for complex systems

Xiangyang Liu<sup>1</sup>, Haiyang Niu<sup>1✉</sup> and Artem R. Oganov<sup>2✉</sup>

Crystal structure prediction has been widely used to accelerate the discovery of new materials in recent years. Up to this day, it remains a challenge to predict the stable stoichiometries and structures of ternary or more complex systems due to the explosive increase of the size of the chemical and configurational space. Numerous novel materials with a series of unique characteristics are expected to be found in this virgin territory while new algorithms to predict crystal structures in complex systems are urgently called for. Inspired by co-evolution in biology, here we propose a co-evolutionary algorithm, which we name COPEX, and which is based on the well-known evolutionary algorithm USPEX. Within this proposed algorithm, a few USPEX calculations for ternary systems and multiple for energetically-favored pseudobinary or fixed-composition systems are carried out in parallel, and co-evolution is achieved by sharing structural information on the fittest individuals among different USPEX sub-processes during the joint evolution. We have applied the algorithm to W–Cr–B, Mg–Si–O, and Hf–Ta–C, three very different systems, and many ternary compounds have been identified. Our results clearly demonstrate that the COPEX algorithm combines efficiency and reliability even for complex systems.

*npj Computational Materials* (2021)7:199; <https://doi.org/10.1038/s41524-021-00668-5>

## INTRODUCTION

Exploring and designing novel materials is a major research avenue in materials science. High-throughput computation, where structure prediction methods act as an essential component, has proven its great utility in the area of materials discovery<sup>1</sup>. In the past decade, many important discoveries, e.g., transparent insulating phase of sodium<sup>2</sup>, superhard M-carbon<sup>3,4</sup>, partially ionic high-pressure form of elemental boron<sup>5</sup>, and superconducting sulfur hydride H<sub>3</sub>S with the record high  $T_c$  203 K<sup>6,7</sup>, etc., were triggered by crystal structure prediction methods.

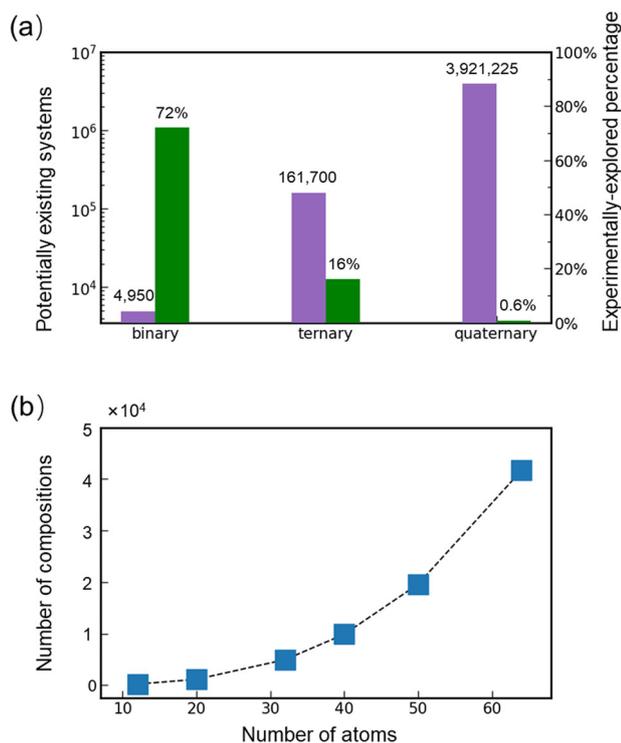
Recently, 244 stable ternary nitrides have been discovered using a data mining structure prediction algorithm, doubling the list of 213 previously known stable ternary nitrides<sup>8</sup>. This discovery further demonstrated the remarkable power of computational materials discovery, indicating that our knowledge of ternary compounds is poor. One might have the impression that a major part of all possible binary, ternary, and even quaternary compounds have already been experimentally discovered. However, as shown in Fig. 1a, only 16% of ternary, and just 0.6% of quaternary systems have been experimentally studied at normal conditions, as demonstrated by P. Villars in 2013<sup>9</sup>. Hence, experimental exploration of multicomponent systems is still very far from complete. With reliable and efficient crystal structure prediction methods, one can imagine that the full exploration of possible ternary or more complex compounds is not beyond our reach anymore. In addition, much attention has been attracted to materials with ternary or more complex compositions due to many important discoveries in the past few years, such as the successful experimental synthesis of possibly the first room-temperature superconductor which is a S–C–H compound formed at high pressure<sup>10</sup>.

Without having the pretense of being rigorous or exhaustive, crystal structure prediction algorithms can be classified into two

categories, one is based on existing knowledge in the form of crystal structure databases, and the other is based on reliable and efficient exploratory algorithms which are capable of making predictions with little or no pre-existing knowledge. When it comes to the former category, data mining<sup>11–16</sup> acts as a representative approach, which predicts the structures of the target compounds when the structure prototypes are already present in the database. In the latter category, evolutionary algorithms<sup>17–20</sup>, basin hopping<sup>21</sup>, minima hopping<sup>22</sup>, metadynamics<sup>23,24</sup>, simulated annealing<sup>25</sup>, and random searching<sup>26</sup> are widely used. However, a big challenge remains when it comes to the exploration of ternary and more complex systems. As shown in Fig. 1b, the number of possible compositions in a ternary system increases dramatically with increasing number of atoms in the unit cell. For instance, with a given unit cell of 50 atoms, 19,600 different compositions are possible in a ternary system. According to our experience, one needs to sample at least hundreds or even thousands of different structures using evolutionary algorithm to locate the structure with global energy minimum of a given composition. Clearly, exploring the enormous chemical space of ternary and more complex systems by doing separate crystal structure prediction calculations for each composition is impossible.

The situation is greatly improved by variable-composition evolutionary searches, which take advantage of the exchange of information between different compositions in the same population, and are capable of predicting all stable compounds in a single calculation<sup>27</sup>. However, while this works well for binary systems, for ternaries it can miss a significant percentage of stable compounds. The cases where USPEX was used for predicting ternary<sup>28</sup> and even quaternary<sup>29</sup> compounds involved very expensive calculations, with careful cross-checks and adding

<sup>1</sup>State Key Laboratory of Solidification Processing, International Center for Materials Discovery, School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, P.R. China. <sup>2</sup>Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 30 bldg. 1 Bolshoy Boulevard, Moscow 121205, Russia. ✉email: haiyang.niu@nwpu.edu.cn; A.Oganov@skoltech.ru



**Fig. 1** Number of potentially existing complex systems and compositions. **a** The number of binary, ternary and quaternary systems and the percentage of experimentally studied systems at normal conditions. **b** The number of compositions as a function of the number of atoms in the unit cell of a ternary system.

results from binary and pseudobinary calculations to the population, and an increased human factor.

Clearly, the key issue of crystal structure prediction for complex systems is the sampling of the high-dimensional chemical space in an effective way. In biology, co-evolution is defined as the evolution of two or more species which mutually affect each other, and it is one of the major forces that produce biodiversity<sup>30</sup>. Recently, co-evolution was used as the basis for an algorithm to predict the optimal material among all possible compounds of all elements<sup>31</sup>. Inspired by the concept of co-evolution, we present a co-evolutionary algorithm to address the problem of compound and crystal structure prediction for complex systems. Here we use USPEX (Universal Structure Predictor: Evolutionary Xtalgraphy)<sup>17,18,32</sup> for every single evolutionary calculation. Therefore we named our co-evolutionary algorithm as COPEX (Co-USPEXs). In our algorithm, each process of USPEX crystal structure prediction can be viewed as evolution of one biological species, and the co-evolution between different USPEX processes (species) is the key feature of our algorithm. We have applied this approach to three representative systems, Mg–Si–O, W–Cr–B, and Hf–Ta–C, where many stable ternary compounds have been identified. Our results show that this approach accelerates the exploration of the chemical space of complex systems.

## RESULTS

### Stability principles of ternary compounds

Stability of a ternary compound can be evaluated by the highest energy of formation in all the combinations of compounds and their own elemental forms. The energy of formation of  $E_{A_xB_yC_{1-x-y}}$  can be expressed as

$$E_f = E_{A_xB_yC_{1-x-y}} - xE_A - yE_B - (1-x-y)E_C, \quad (1)$$

where  $E_A$ ,  $E_B$ , and  $E_C$  refer to the ground-state energy of elemental forms of A, B, and C.  $E_f$  is a function of the composition, and its calculation requires the knowledge of  $E_A$ ,  $E_B$ ,  $E_C$ , and  $E_{A_xB_yC_{1-x-y}}$ . Stable compounds shall have negative energy of formation  $E_f$ , however, this is necessary but not sufficient. The sufficient condition is that the compound be thermodynamically more stable than all of the possible products of decomposition.

As a result, all the thermodynamically stable compounds form a convex hull. As discussed in the introduction section and also demonstrated by Villars and Iwata<sup>9</sup>, there are so many possible stoichiometries that computationally exploring the whole chemical space is impossible. Thus, before discussing our co-evolutionary algorithm, we would like to explain the strategies that we use to navigate the chemical space.

Traditionally, according to the chemical bonding features, molecular, ionic, ionic-covalent, and intermetallic ternary compounds can be divided into four groups, (I) ternary molecular (all non-metals), (II) acids, (III) ternary ionic or ionic-covalent (including oxysalts), and (IV) intermetallics. Among the 45,000 ternary structures from the Crystallography Open Database (COD)<sup>33</sup>, we found (analyzing interatomic distances) that 45% are molecular compounds, 49% are ionic or ionic-covalent, and the remaining 6% are intermetallic compounds. Since our work is focused on crystal structure prediction of inorganic compounds, therefore, the 45% molecular compounds are disregarded in our analysis. In nature, the octet rule governs the formation of ionic or covalent compounds. For instance, by analyzing all the ionic, ionic-covalent compounds and intermetallics, one finds that about 46% of them have both the compositions  $(AB)_k(AC)_l$  (where  $AB_k$  and  $AC_l$  are stable binary compounds in the corresponding systems) and  $A_x(BC)_m$  (in which A is a pure element, and  $BC_m$  is a stable binary compound in the B–C system). Also, about 35% and 11% are found exclusively in  $(AB)_k(AC)_l$  or  $A_x(BC)_m$  category, respectively. It is reasonable that atoms always seek to adopt the most stable electronic configuration. Following this philosophy, one could take advantage of the above principle and explore the chemical space along the pseudobinary joints.

The situation is more complex for predicting ternary intermetallics. Since metallic bonding comes from free electrons, the individual properties of the constituent atoms have less influence on the formation of structures than in the case of ionic and covalent compounds. In general, intermetallics have a higher propensity to form solid solutions<sup>34</sup>, host significant concentrations of defects<sup>35</sup>, and form unusual stoichiometries. Many ternary stable compounds adopt the same structure prototypes as the binary compounds<sup>9</sup>. This situation holds in many cases, but should not be accepted as a universal axiom.

In conclusion, ternary compounds are usually formed by combining or atom-substituting stable binary compounds. Hence, we can summarize that the promising and efficient way to find the thermodynamically stable ternary compounds is focusing on the compositions along the pseudobinary joints or some fixed-composition obtained by atomic substitution in stable binary compounds. However, one cannot rule out the possibility to find stable ternary compounds which do not follow the above rules. In such a case, a fully unconstrained variable-composition ternary structure search is still needed and different exploration strategies are encouraged.

### Co-evolutionary algorithm

The evolutionary structure prediction algorithm mimics Darwinian evolution and employs a natural selection of the fittest in combination with variation operators, which has been proven to be very successful. In multicomponent systems, due to their vast compositional space, many structures/compositions have to be sampled (rendering the searches quite expensive), and still it is possible to skip stable compositions. Thus, in order to explore the

configurational space effectively, one could focus on the compositions of the pseudobinary systems or some fixed compositions. Thus sampling task can be divided into several sub-processes, and each of them tackles one part of the compositional space. In biology, co-evolution proved a major force to create biodiversity. Here we take advantage of this idea to propose the COPEX algorithm to manipulate the multiple crystal structure prediction processes. The working procedure of COPEX is shown in Fig. 2 and the main steps are:

1. Gathering the structural information of all the known elemental, binary (including all the stable and low-energy metastable phases), and experimentally known ternary phases in this or similar systems.
2. Taking the above gathered structures as seeds and performing the COPEX calculation which will initialize two independent ternary variable-composition structure prediction processes at the very beginning.
3. Evaluating the results every ten generations and extracting low-energy structures to form a structure pool.
4. Sharing good structures between different USPEX runs, and applying variation operators to generate new compositions and structures.
5. Constructing the phase diagram of the system, and manipulating USPEX runs, i.e., launching or terminating pseudobinary or fixed-composition USPEX sub-processes.
6. Repeat steps 3–5 until pre-specified halting criteria are achieved.

USPEX has proven to be efficient and reliable for predicting stable structures in fixed-composition and in binary systems<sup>27,36–38</sup>. Thus, the USPEX processes in COPEX can accelerate the process to locate the stable stoichiometry or structure, which in turn will benefit the evolutionary process and generate high-quality structures for the next generation. The phase diagram of the ternary system can be generated using the energies of formation of all structures. Since the phase diagram will change dynamically during the evolutionary process, we set to analyze the phase diagram after every 10 generations and manipulate the

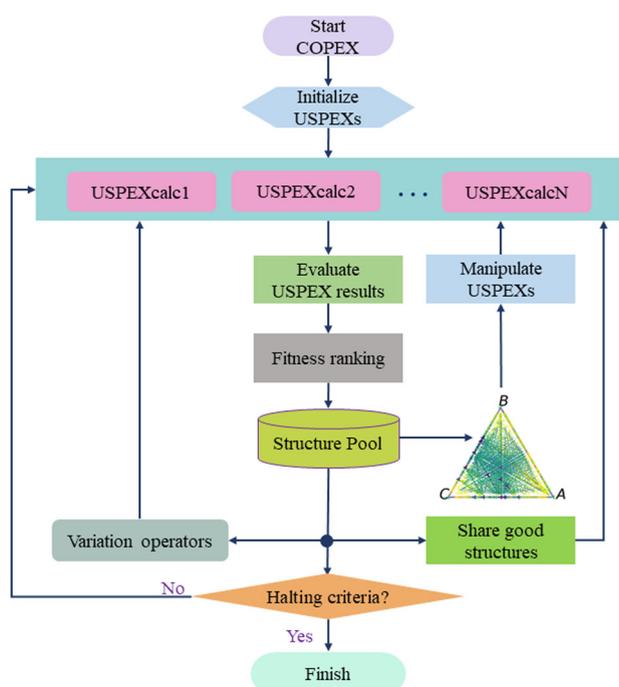
compositions and USPEX processes automatically as illustrated in step 4. The success of the evolutionary algorithm for crystal structure prediction relies largely on the various variation operators, i.e., heredity, softmutation, lattice mutation, permutation, and transmutation<sup>37,39</sup>. Among them, the chemical transmutation operator is designed for variable-composition structure search, which means swapping atoms at randomly selected positions while keeping positions of most atoms unchanged. After every generation of COPEX, we reassign the atomic positions of different elements with closer atomic radii, which turned out to be very helpful to the co-evolutionary process. In the following, we have applied the COPEX algorithm to W–Cr–B, Mg–Si–O, and Hf–Ta–C, three chemically very different systems, and found evidence the high efficiency and reliability of the method.

### Application to the W–Cr–B system

The W–Cr–B system has attracted extensive attention due to the remarkable mechanical properties and chemical inertness of transition metal borides, such as CrB<sub>4</sub><sup>40</sup> and WB<sub>5–x</sub><sup>41</sup>. Besides, tungsten chromium borides, such as M<sub>3</sub>B<sub>2</sub> and M<sub>5</sub>B<sub>3</sub> (M refers to metal atoms) commonly exist in the Ni-superalloy single crystal turbine blades as the predominant precipitate phases. For a long time, the transition metal elements were believed to be distributed randomly over the sublattice M in M<sub>3</sub>B<sub>2</sub> and M<sub>5</sub>B<sub>3</sub><sup>42–44</sup>. However, in our previous work, we have found both from theory and experiment that the transition metal atoms are distributed in an ordered fashion in the M<sub>3</sub>B<sub>2</sub> and M<sub>5</sub>B<sub>3</sub> borides, and are stoichiometric, in fact, W<sub>2</sub>CrB<sub>2</sub> and W<sub>4</sub>CrB<sub>3</sub>, respectively<sup>44,45</sup>. Naturally, one may wonder if any other stable ternary compounds exist in the W–Cr–B system. Thus, we chose COPEX to investigate this system.

Before running COPEX, we need to thoroughly gather the structural information of all the stable phases in the corresponding unary and binary systems. If not, COPEX will need to spend unnecessary efforts to explore these systems. The W–B and Cr–B systems were widely investigated in literature<sup>40,41,46,47</sup>. In the Cr–B system, at ambient pressure five ground-state compounds (Cr<sub>2</sub>B, Cr<sub>5</sub>B<sub>3</sub>, CrB, CrB<sub>2</sub>, and CrB<sub>4</sub><sup>40</sup>) and two metastable compounds (Cr<sub>2</sub>B<sub>3</sub> and Cr<sub>3</sub>B<sub>4</sub>) are found<sup>46</sup>. In the W–B system, five ground-state compounds (W<sub>2</sub>B, W<sub>8</sub>B<sub>7</sub>, WB, WB<sub>2</sub>, and WB<sub>3</sub>) are reported at ambient pressure<sup>48</sup>. In addition, WB<sub>5–x</sub> is also taken into account in our calculation; this phase is stabilized by zero-point vibrational energy<sup>41</sup>. Furthermore, no stable binary compounds exist in the W–Cr system. Taking this prior information, we have performed two independent COPEX runs of the W–Cr–B system at ambient pressure. Besides W<sub>2</sub>CrB<sub>2</sub> and W<sub>4</sub>CrB<sub>3</sub>, we have also predicted three stable ternary compounds, namely, WCrB, WCrB<sub>2</sub>, and WCr<sub>2</sub>B. The crystal structure parameters of the predicted three ternary compounds are given in Supplementary Table 1. Their elastic constants are given in Supplementary Table 2, and we can see that all of them satisfy Born criteria of mechanical stability<sup>49</sup>. Phonon calculations show no imaginary frequencies in these three compounds, which indicates their dynamical stability. The phase diagram and energies of formation of all the sampled structures are given in Fig. 3.

Here we would like to discuss the structural origins and features of the predicted ternary compounds. The crystal structures of the five ternary phases and some related compounds are drawn in Fig. 4. Among the five ternary W–Cr–B compounds, W<sub>2</sub>CrB<sub>2</sub> and W<sub>4</sub>CrB<sub>3</sub> share the same structural features as other earlier reported phases. Namely, the prototypes of W<sub>2</sub>CrB<sub>2</sub> and W<sub>4</sub>CrB<sub>3</sub> are derivatives of V<sub>3</sub>B<sub>2</sub> and Cr<sub>5</sub>B<sub>4</sub> structure types, respectively, while the prototypes of WCrB<sub>2</sub>, WCrB, and WCr<sub>2</sub>B have never been reported before to the best of our knowledge. As shown in Fig. 4c, the structure of W<sub>4</sub>CrB<sub>3</sub> consists of two layers where one layer has the same structure as W<sub>2</sub>CrB<sub>2</sub> (see Fig. 4b) while the other is made



**Fig. 2** The COPEX algorithm. Flowchart of the co-evolutionary algorithm COPEX.

of W-centered square antiprisms, just like in the structure of  $W_2B$  (see Fig. 4a). Therefore,  $W_4CrB_3$  is composed of an alternating array of trigonal prism layer and square antiprism layer. In other words,  $W_4CrB_3$  can be viewed as an intergrowth or a stacking of  $W_2CrB_2$  and  $W_2B$  along the [001] direction, indicating that  $W_2CrB_2$  and  $W_4CrB_3$  are very likely intergrown in practice, which agrees with experimental observations<sup>44,45</sup>.

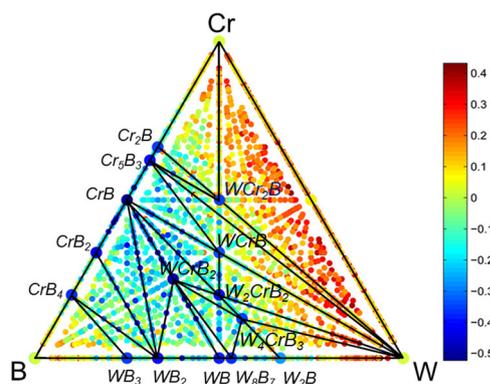
The structure of  $WCrB_2$  has never been reported before, it shares some similarities with  $CrB$  and  $WB$ . As  $CrB$  and  $WB$  adopt the same ground-state structure, here we use  $WB$  as an example. The ground-state phase of  $WB$  crystallizes into the  $I4_1/amd$  structure (also known as  $\alpha$ -WB)<sup>47</sup>. The zigzag chains formed by B atoms in  $\alpha$ -WB are perpendicular to each other along the  $a$ -axis as shown in Fig. 4d. One B atom is surrounded by 7 W atoms, which form a capped trigonal prism. Therefore,  $\alpha$ -WB can also be viewed as being constructed by the perpendicular zigzag chains lined by capped trigonal prisms. What if the zigzag chains are parallel to each other instead of perpendicular? The resulting structure would be the well-known  $\beta$ -WB<sup>48</sup>, which is also constructed by the same capped trigonal prisms.

As shown in Fig. 4e,  $WCrB_2$  is also constructed by the capped trigonal prism, in which one B atom is surrounded by 3 W and 4 Cr atoms. Different from  $\alpha$ - and  $\beta$ -WB, B atoms in  $WCrB_2$  form parallel armchair chains. As shown in Fig. 4f, one B atom in  $WCrB$  is coordinated by 5 W and 4 Cr atoms, resulting in a slightly distorted

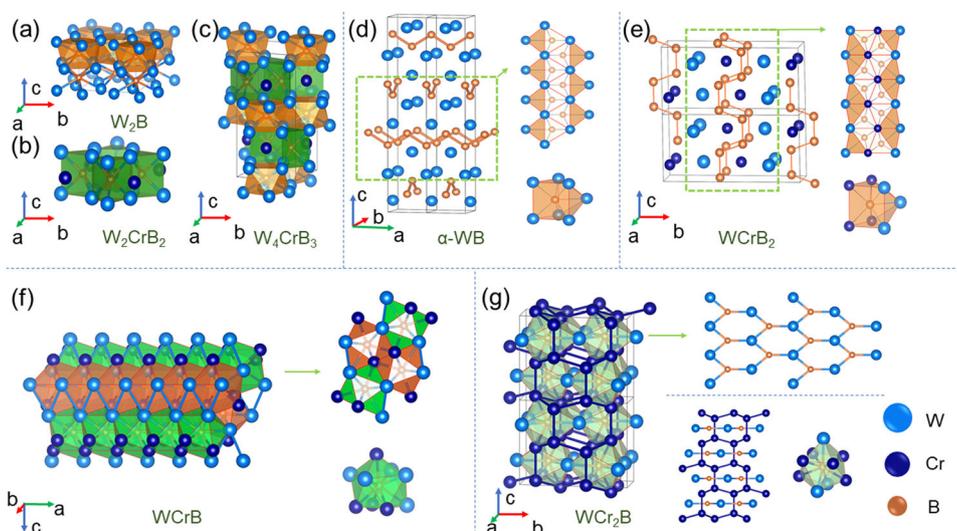
capped square antiprism. The polyhedra are arranged in a zigzag way, with a  $2_1$  screw axis.  $WCr_2B$  has  $Cmcm$  space group. Interestingly, the Cr atoms form a sublattice with the lonsdaleite topology. Boron atoms are in a tricapped trigonal prismatic coordination (with 3 W and 6 Cr neighbors). Furthermore, Cr and B atoms form a graphene-like flat layer, penetrated by the hexagonal framework of Cr atoms, and result in an interesting graphene-diamond hybrid structure.

Ternary tungsten chromium borides were thought to be nonexistent for a long time. Our work, however, has uncovered five stable tungsten chromium borides at ambient pressure. In these phases, we see strong hybridization between  $W-d$ ,  $Cr-d$ , and  $B-p$  states (Supplementary Fig. 3), indicating strong covalent  $W-B$  and  $Cr-B$  bonds.

As shown in the  $W-Cr-B$  phase diagram (Fig. 3),  $WCrB_2$  is located on the  $WB-CrB$  pseudobinary joint. The stability of  $WCrB_2$  can be further investigated by taking end-point compounds  $WB$  and  $CrB$  into consideration. Here we have constructed artificial  $oC16-WB$  and  $oC16-CrB$  by taking the same structure of  $WCrB_2$ , and the energy of formation of  $WCrB_2$  from  $oC16-WB$  and  $oC16-CrB$  is  $-0.2$  eV/atom. When the atomic positions of Cr and W in  $WCrB_2$  are interchanged, the calculated enthalpy of formation is higher than  $WCrB_2$  by 0.194 eV/atom. Although W and Cr belong to the same group, they occupy distinct positions due to their large difference in atomic radius, which makes  $W-Cr$  disorder unfavorable. In summary, five ternary borides have been identified with the help of COPEX.



**Fig. 3 Phase diagram of the W-Cr-B system.** Energies of formation (in  $eV\text{ atom}^{-1}$ ) of the sampled structures of the W-Cr-B system at ambient pressure are shown. The color bar gives the scale of the energies of formation.

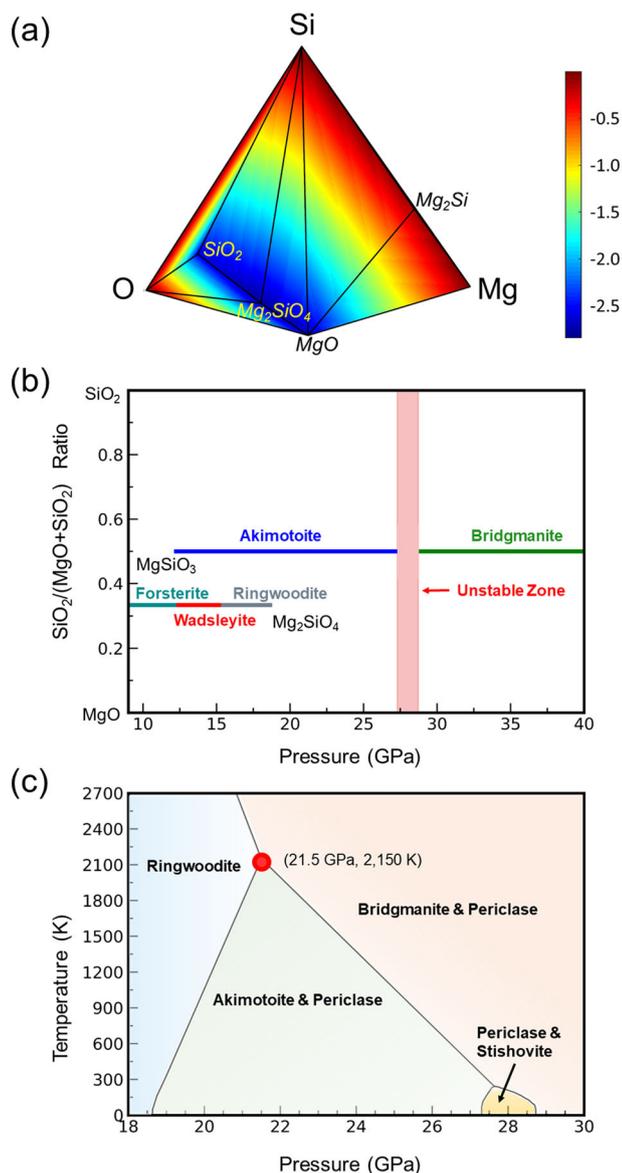


**Fig. 4 Crystal structures of the phases in the W-Cr-B system.** a  $W_2B$ , b  $W_2CrB_2$ , c  $W_4CrB_3$ , d  $\alpha$ -WB, e  $WCrB_2$ , f  $WCrB$ , and g  $WCr_2B$ .

would further demonstrate the performance of the COPEX algorithm in an ionic-covalent ternary system.

By exploring the whole range of compositions in the Mg–Si–O at ambient pressure, we constructed the phase diagram of the Mg–Si–O system as shown in Fig. 5a. The vertices and edges of the convex hull refer to stable phases and phase boundaries, respectively. Any compound located on the convex polyhedron is thermodynamically stable. It can be seen that  $\text{Mg}_2\text{SiO}_4$  is the only thermodynamically stable ternary compound in the Mg–Si–O system, and the most likely decomposition pathway of  $\text{Mg}_2\text{SiO}_4$  is into MgO and  $\text{SiO}_2$ .

Unexpectedly, our calculation shows no thermodynamically stable ternary compound in the Mg–Si–O system at 28 GPa. In order to further investigate this phenomenon, we then



**Fig. 5 Thermodynamics of Mg–Si–O system.** **a** Mg–Si–O convex hull with energies of formation (in eV atom<sup>−1</sup>, with scale shown by color bar) of the sampled structures at ambient pressure. **b** Pressure-composition phase diagram of the MgO–SiO<sub>2</sub> system. **c** Pressure–temperature phase diagram of Mg<sub>2</sub>SiO<sub>4</sub> and its decomposition pathways. The triple point between phases ringwoodite-bridgmanite+periclase-akimotoite+periclase is highlighted by red dot.

constructed the phase diagram of the pseudobinary MgO–SiO<sub>2</sub> system, including the zero-point energy. As shown in Fig. 5b, at zero Kelvin akimotoite (MgSiO<sub>3</sub> with ilmenite-type structure) will decompose into periclase (MgO) and stishovite (SiO<sub>2</sub>) at 27.29 GPa, and then periclase and stishovite will recombine into bridgmanite (perovskite-type MgSiO<sub>3</sub>) at 28.72 GPa. In other words, both akimotoite and bridgmanite are thermodynamically unstable in the pressure range 27.29–28.72 GPa and therefore a ternary compound instability zone exists, which was also observed in another theoretical work<sup>51</sup>.

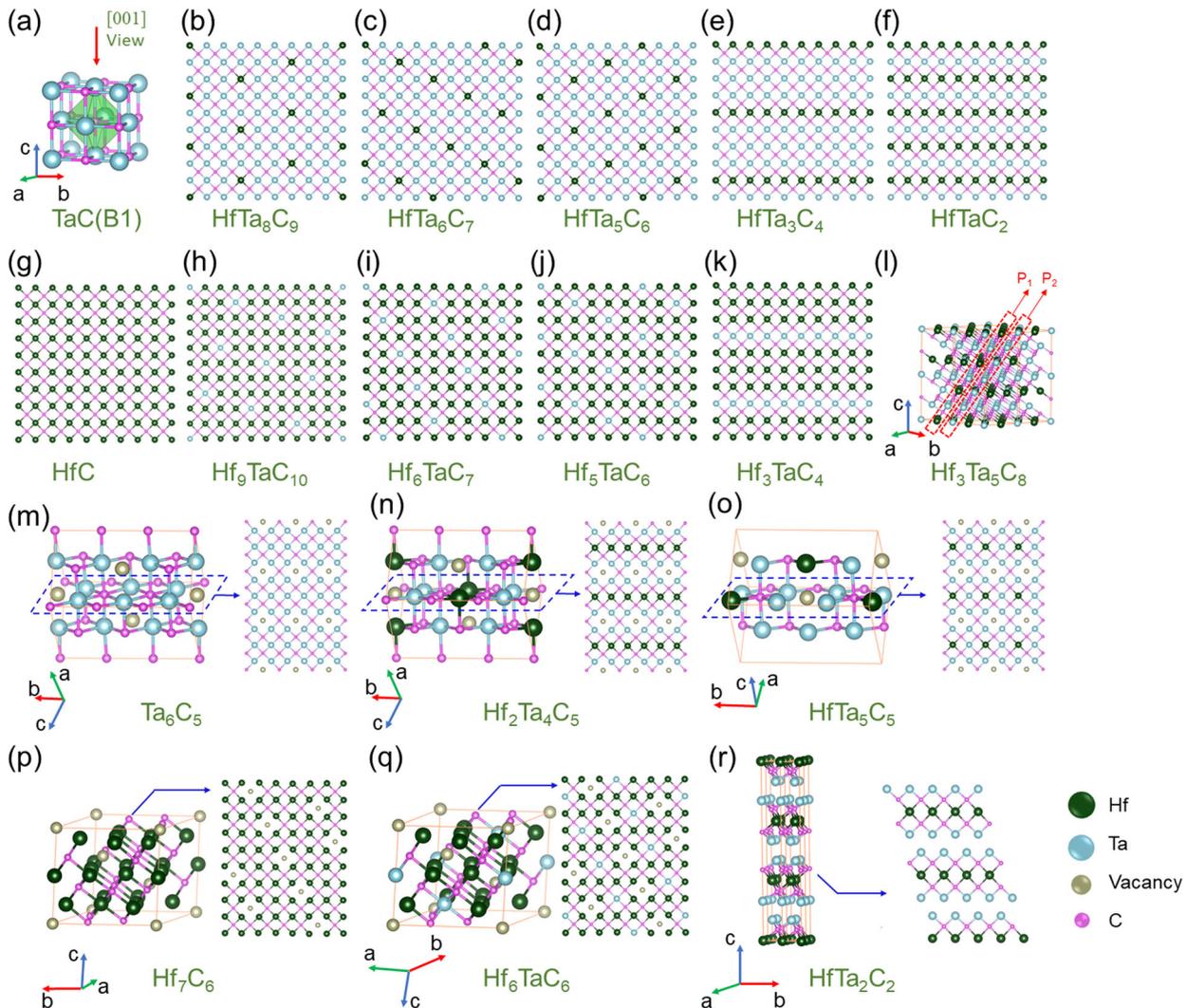
By considering the effects of temperature, we have constructed the temperature–pressure phase diagram of Mg<sub>2</sub>SiO<sub>4</sub>, as shown in Fig. 5c. The zone of instability of ternary compounds narrows as temperature increases, and eventually disappears at 250 K and 27.62 GPa, and akimotoite will transform into bridgmanite directly without decomposition above 250 K. We can also conclude that the vibrational properties of akimotoite and bridgmanite play an essential role to stabilize them in the Earth’s mantle. Even though the zone of instability of ternary compounds disappears above 250 K, we still need to pay attention to it, since the impurity (Fe, Al, et al.) effect<sup>50</sup> may extend it to higher temperatures. It is possible that this zone crosses the geotherm in some planets, and therefore might have significant influence on the internal structures and dynamics of these planets. Furthermore, ringwoodite decomposes into periclase and akimotoite at 18.63 GPa at zero Kelvin. A triple point between phases ringwoodite-bridgmanite+periclase-akimotoite+periclase is found at 21.5 GPa and 2150 K, which is in reasonable agreement with the literature<sup>51,55</sup>. In other words, ringwoodite will decompose into periclase and bridgmanite above this point. This phase transition is the main cause of the discontinuity in the mantle at the depth of 670 km<sup>51,55</sup>. The phase diagram shown in Fig. 5c is essential for determining the temperature distribution in the deep interiors of the Earth<sup>52</sup>.

### Application to the Hf–Ta–C system

High-performance refractory materials are of crucial significance in some extreme applications, such as gas turbines, heat shields for hypersonic vehicles, and more. For instance, transition metal carbides (TMCs) HfC and TaC are the most refractory materials among binary compounds with the highest melting points of nearly 4000 °C<sup>56</sup>. Hf and Ta are located next to each other in the periodic table and have similar chemical properties, and both HfC and TaC possess the B1 rocksalt crystal structure. Mixing the two binary carbides provides an approach to tune the electronic and mechanical properties. One of the questions immediately rising is whether ternary Hf–Ta–C carbides will be ordered stoichiometric compounds or solid solutions with Hf–Ta disorder. In order to answer this question and explore the configurational space of the Hf–Ta–C system, we performed a COPEX calculation.

Herein, structure searches using the COPEX algorithm with up to 40 atoms in the primitive cell were carried out at ambient pressure. As usual, the stable structures of the pure elements and binary compounds should be provided. We have found four (HfC, Hf<sub>7</sub>C<sub>6</sub>, Hf<sub>4</sub>C<sub>3</sub>, and Hf<sub>3</sub>C<sub>2</sub>) and three (TaC, Ta<sub>6</sub>C<sub>5</sub>, and Ta<sub>2</sub>C) stable phases in the Hf–C and Ta–C systems, respectively, at ambient pressure, which is in agreement with the literature<sup>57</sup>. In addition, many low-energy metastable phases like Hf<sub>8</sub>C<sub>7</sub>, Hf<sub>6</sub>C<sub>5</sub>, Hf<sub>2</sub>C, and Ta<sub>5</sub>C<sub>4</sub> were also considered in our calculation<sup>58,59</sup>. Furthermore, no stable compound has been found in the Hf–Ta system. Taking this information into account, we performed a COPEX search and obtained the phase diagram of the Hf–Ta–C system at ambient pressure, as shown in Fig. 6. Eventually, our study identified 14 stable ternary compounds containing HfTa<sub>8</sub>C<sub>9</sub>(P1̄), HfTa<sub>6</sub>C<sub>7</sub>(R3̄), HfTa<sub>5</sub>C<sub>6</sub>(C2/m), HfTa<sub>3</sub>C<sub>4</sub>(C2/m), Hf<sub>3</sub>Ta<sub>5</sub>C<sub>8</sub>(R3̄m), HfTaC<sub>2</sub>(R3̄m), Hf<sub>3</sub>TaC<sub>4</sub>(C2/m), Hf<sub>5</sub>TaC<sub>6</sub>(C2/m), Hf<sub>6</sub>TaC<sub>7</sub>(R3̄), Hf<sub>9</sub>TaC<sub>10</sub>(P1̄), HfTa<sub>2</sub>C<sub>2</sub>(R3̄m), HfTa<sub>5</sub>C<sub>5</sub>(Cm), Hf<sub>2</sub>Ta<sub>4</sub>C<sub>5</sub>(C2/m), Hf<sub>6</sub>TaC<sub>6</sub>(R3̄). In all of these





**Fig. 7** Crystal structures of the phases in the Hf-Ta-C system. **a** Crystal structure of TaC, and the two-dimensional projections (along the [001] direction of the B1 structure) of the structures **b** HfTa<sub>8</sub>C<sub>9</sub>, **c** HfTa<sub>6</sub>C<sub>7</sub>, **d** HfTa<sub>5</sub>C<sub>6</sub>, **e** HfTa<sub>3</sub>C<sub>4</sub>, **f** HfTaC<sub>2</sub>, **g** HfC, **h** Hf<sub>9</sub>TaC<sub>10</sub>, **i** Hf<sub>6</sub>TaC<sub>7</sub>, **j** Hf<sub>5</sub>TaC<sub>6</sub>, **k** Hf<sub>3</sub>TaC<sub>4</sub>. **l** The crystal structure of Hf<sub>3</sub>Ta<sub>5</sub>C<sub>8</sub> is formed by alternation of one layer of HfTaC<sub>2</sub> (denoted as P<sub>1</sub>) and one layer of HfTa<sub>3</sub>C<sub>4</sub> (denoted as P<sub>2</sub>). Crystal structures and their projections along the [001] direction of the B1 structure: **m** Ta<sub>6</sub>C<sub>5</sub>, **n** Hf<sub>2</sub>Ta<sub>4</sub>C<sub>5</sub>, **o** HfTa<sub>5</sub>C<sub>5</sub>, **p** Hf<sub>7</sub>C<sub>6</sub>, **q** Hf<sub>6</sub>TaC<sub>6</sub>, **r** HfTa<sub>2</sub>C<sub>2</sub>.

We have also compared the chemical composition sampling intensity maps obtained by COPEX and USPEX for the Hf-Ta-C system. The results show that COPEX has an enhanced sampling region as highlighted in Fig. 9, in which nine out of the ten predicted stable ternary compounds are located. USPEX shows similar but much lighter sampling intensity than COPEX. In other words, USPEX lays the foundation which ensures the sampling diversity and COPEX further enhances the sampling of the promising regions of the chemical space.

## DISCUSSION

Predicting the crystal structure of ternary and more complex compounds is very challenging, as the chemical and configurational space expands dramatically with increasing number of atomic types and the total number of atoms. Here, we propose an algorithm to address this challenge, developing the algorithm called COPEX, coupled with the powerful evolutionary algorithm USPEX and based on the idea of co-evolution. Within COPEX, multiple USPEX processes are carried out in parallel, where two of them target the whole configurational space while the rest are

designed to focus only on the most promising area of the compositional space. Co-evolution is achieved by sharing structural information on the fittest individuals among the different USPEX processes during the evolution. Through sharing high-quality structures, COPEX can maximize the characteristic of the self-improving evolutionary algorithm and improve the success rate of structure prediction.

In order to demonstrate the performance of the COPEX algorithm, we have chosen three quite different systems, the intermetallic systems W-Cr-B and Hf-Ta-C, and the ionic-covalent system Mg-Si-O, to explore. The atomic radii of Cr and W are significantly different, while the atomic radii of Hf and Ta are almost the same. This explains the different behavior of these two systems. In the W-Cr-B system, besides W<sub>2</sub>CrB<sub>2</sub> and W<sub>4</sub>CrB<sub>3</sub> we have also found three stable ternary compounds, WCrB, WCrB<sub>2</sub>, and WCr<sub>2</sub>B, belonging to new structure types. In the Hf-Ta-C system, we have identified 14 stable ternary compounds, all of which are based on the B1 rocksalt structure type including four compounds with vacancy ordering. For the Mg-Si-O system, besides Mg<sub>2</sub>SiO<sub>4</sub> and MgSiO<sub>3</sub>, no other ternary compound has been found at pressures up to 40 GPa. Importantly, we have found

at low temperature a region of decomposition (27.29–28.72 GPa at zero Kelvin) where no stable ternary compounds exist.

In conclusion, we have applied COPEX to three representative systems W–Cr–B, Mg–Si–O, and Hf–Ta–C, and then evaluated its

performance based on benchmark tests, and analyzed its sampling efficiency using a low-dimensional representation of the complex high-dimensional atomic environments and chemical composition space. The results demonstrate the reliability and efficiency of this methodology and illustrate that COPEX is a powerful tool for predicting new complex materials.

**Table 1.** Benchmark tests using crystal structure prediction algorithms COPEX and USPEX applied to different systems at ambient pressure.

| System  | Algorithm | $N_{\text{pop}}$ | $N_{\text{gen}}$ | $P_{\text{ran}}$ | $N_{\text{sam}}$ | $N_{\text{com}}$ |
|---------|-----------|------------------|------------------|------------------|------------------|------------------|
| W–Cr–B  | COPEX1    | 100              | 171              | $35 \pm 6$       | 19879            | 5                |
|         | COPEX2    | 100              | 173              | $35 \pm 5$       | 20133            | 5                |
|         | USPEX1    | 100              | 178              | $35 \pm 7$       | 20186            | 0                |
|         | USPEX2    | 100              | 180/275          | $35 \pm 3$       | 20150/30053      | 0/1              |
|         | M-USPEXs  | 100              | 179              | $35 \pm 6$       | 20168            | 2                |
| Mg–Si–O | COPEX     | 100              | 32               | $35 \pm 4$       | 3471             | 1                |
|         | USPEX1    | 100              | 34               | $35 \pm 4$       | 3452             | 0                |
|         | USPEX2    | 100              | 34               | $35 \pm 2$       | 3491             | 0                |
|         | M-USPEXs  | 100              | 34               | $35 \pm 4$       | 3571             | 0                |
| Hf–Ta–C | COPEX     | 120              | 163              | $35 \pm 5$       | 21400            | 10               |
|         | USPEX1    | 120              | 167              | $35 \pm 5$       | 21769            | 2                |
|         | USPEX2    | 120              | 171              | $35 \pm 6$       | 21635            | 1                |
|         | M-USPEXs  | 120              | 167              | $35 \pm 5$       | 21711            | 5                |

$N_{\text{pop}}$ ,  $N_{\text{gen}}$ ,  $P_{\text{ran}}(\%)$ ,  $N_{\text{sam}}$ , and  $N_{\text{com}}$  refer to the population size per generation, generation size in every single search, percentage of structures generated randomly, total number of sampled structures, and the number of predicted ternary compounds in each system. All calculations were performed at ambient pressure. Structures with up to 40 atoms in the primitive cell are considered.

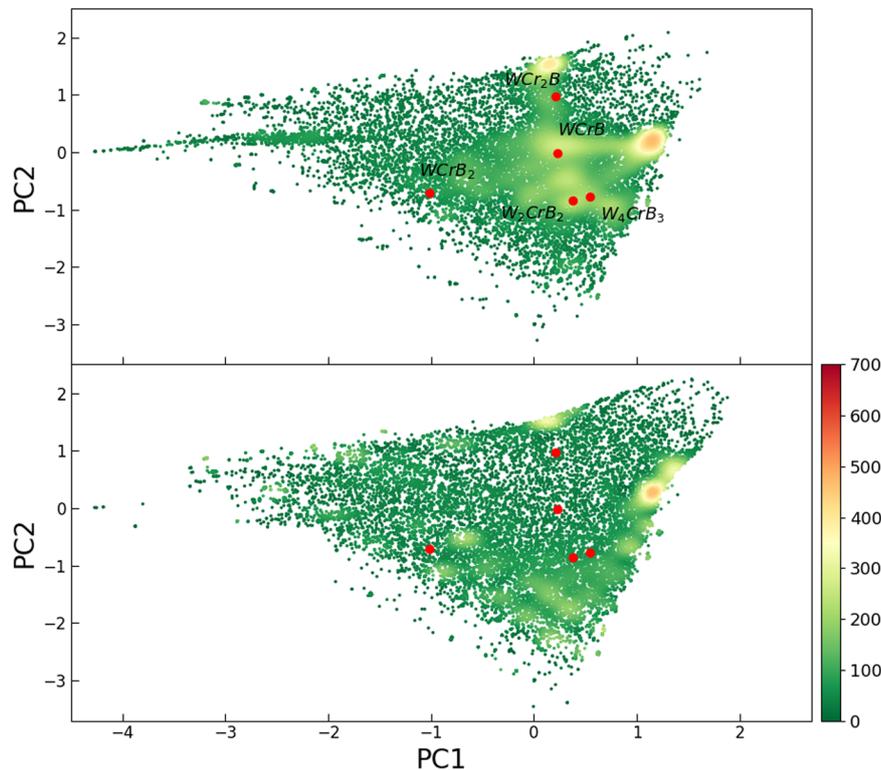
## METHODS

### DFT calculations

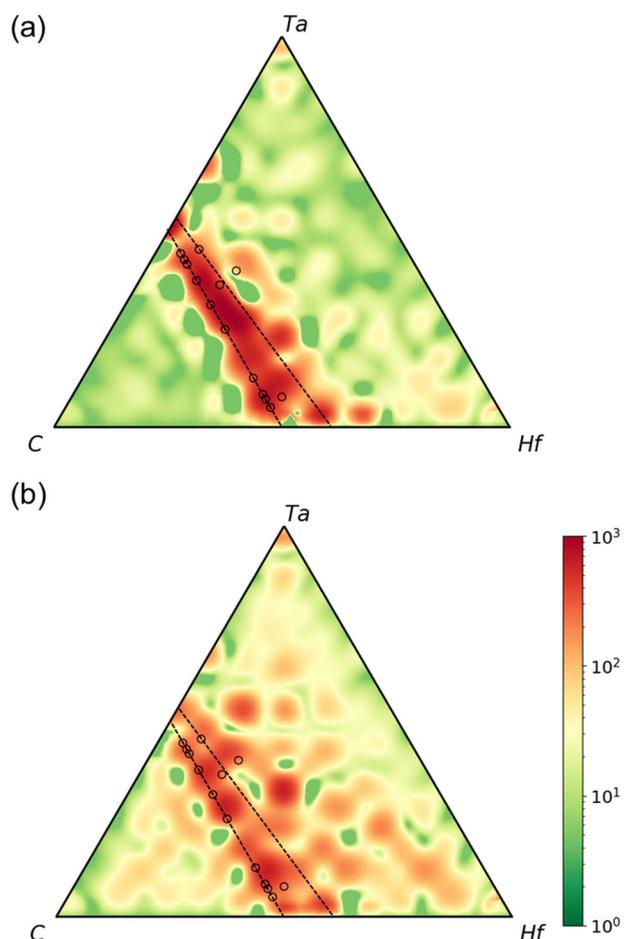
Density functional theory (DFT) calculations within the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation (GGA<sup>66</sup>), implemented in the VASP<sup>67</sup> code, are used for structure relaxations and total energy calculations. The all-electron projector-augmented wave method is used, with a plane-wave kinetic energy cutoff of 600 eV and k-point mesh with the reciprocal-space resolution of  $2\pi \times 0.02 \text{ \AA}^{-1}$ . Phonon calculations were performed using Phonopy<sup>68</sup> code using the force constants calculated by VASP.

### COPEX calculations

The COPEX calculations were performed by following the procedure shown in Fig. 2 and the sub-processes were carried out using the evolutionary algorithm USPEX. In the two USPEX runs that cover the whole ternary chemical space, 100 structures are produced in each generation, while in other USPEX runs, 60–80 structures are produced. In each USPEX run, structures of the first generation were produced randomly. For variable-composition searches, the fractions of structures produced by random structure generator, heredity, softmutation, lattice mutation, and transmutation were 35, 35, 10, 10, and 10% in the second generation. For fixed-composition searches, the fraction of structures produced by random structure generator, heredity, softmutation, lattice mutation, and permutation were 35, 35, 10, 10, and 10% in the second generation. In subsequent generations, the weights of variation operators were adjusted dynamically. Stability of a compound is determined using the thermodynamic convex hull construction.



**Fig. 8** PCA maps of the W–Cr–B system. Structures are taken from the COPEX (upper panel) and USPEX (lower panel), respectively. The map is colored according to the sampling intensity, divided into  $29 \times 26$  bins, and the color bar refers to the number of sampled structures in each bin. The five stable ternary compounds are highlighted by red dots. The PC1 and PC2 refer to the first and second principal components, respectively.



**Fig. 9** Sampling intensity maps of the Hf-Ta-C system in chemical composition space. Data are generated from **a** COPEX and **b** USPEX. The map is colored according to the sampling intensity, divided into  $23 \times 20$  bins, and the color bar refers to the number of sampled structures in each bin. The stable ternary compounds are marked by black circles.

In step 3, 10% fittest structures (i.e., those closest to the convex hull) from each USPEX sub-process were extracted to construct a structure pool, in which identical structures are discarded.

In step 4, structures in the structure pool are manipulated in two ways. First, structures will be distributed to each USPEX sub-process as “seed” structures. Such structures are treated the same as the structures generated by USPEX. Structures which do not exist in the current USPEX sub-process will be selected as the high-quality structural gene carriers to generate child structures. Importantly, the distributed structures must follow the stoichiometry rule of each USPEX sub-process. For instance, the stoichiometry of seed  $A_m B_n C_z$  must satisfy  $m \times C_1 = x$ ,  $n \times C_1 + p \times C_2 = y$ , and  $q \times C_2 = z$ , in which  $A_m B_n$  and  $B_p C_q$  refer to the end-point compounds of the pseudobinary system, and all the variables are positive integers.

Second, we apply the variation operators to structures in the structure pool to generate structures and then remove duplicate structures using fingerprints<sup>69</sup>. The resulting structures will be added to different USPEX sub-processes according to the distribution principle discussed previously. The variation operators used here include heredity, softmutation, lattice mutation, and MTransmutation. Among them, heredity, softmutation, lattice mutation were already implemented in USPEX<sup>32,37</sup>. In addition to these operators, an operator named MTransmutation (Multi-Equivalent positions transmutation) is introduced here. In the MTransmutation operator, atoms of a certain type populating a randomly selected Wyckoff position are replaced by atoms of another type. The heredity and MTransmutation operators are very effective in generating high-quality offspring as demonstrated in Supplementary Figs 7–9.

In step 5, information on all the sampled structures is gathered and the composition-enthalpy phase diagram is constructed. Importantly, edge lines of the convex hull on the phase diagram are the potential searching paths. Searching priority of these pseudobinary systems is ranked by the numbers of low-energy structures along the path inspired by the Bell–Evans–Polanyi principle<sup>70,71</sup> and paper<sup>69</sup>, which demonstrated that low-energy minima are clustered in compact regions of the configurational space. In addition, whenever a ternary compound appears on the current phase diagram, its composition will be added to the list of fixed-composition sub-processes. One can set how many USPEX sub-processes are to be run in parallel, but all the promising pseudobinary and fixed-composition systems are recommended to be searched eventually. COPEX will launch new USPEX sub-processes based on the above criteria. In our work, the top 10% structures based on the rank of fitness according to their enthalpy will be deemed as low-energy structures. For instance, COPEX can easily find out that  $\text{MgO-SiO}_2$ ,  $\text{MgSiO}_3$ , and  $\text{Mg}_2\text{SiO}_4$  are the promising systems that shall be explored. The two ternary variable-composition USPEX sub-processes are kept alive during the whole COPEX search. The pseudobinary or fixed-composition sub-processes are evaluated every 10 generations, and run on subsystems that only produce structures that are far above the updated convex hull, will be terminated. Naturally, COPEX can reactivate such sub-processes as the search continues. One can also launch or terminate sub-processes manually.

#### DATA AVAILABILITY

The data supporting the findings of this study are available within the paper and Supplementary information. Structures of all stable and some low-energy metastable structures are available via Github at: <https://github.com/Dustglaxy/Co-evolutionary-crystal-structure-prediction-algorithm-for-complex-systems>.

#### CODE AVAILABILITY

The program of this study will be available in the upcoming version of USPEX.

Received: 27 April 2021; Accepted: 6 November 2021;

Published online: 09 December 2021

#### REFERENCES

- Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
- Ma, Y. et al. Transparent dense sodium. *Nature* **458**, 182–185 (2009).
- Wang, Y., Panzik, J. E., Kiefer, B. & Lee, K. K. Crystal structure of graphite under room-temperature compression and decompression. *Sci. Rep.* **2**, 520 (2012).
- Li, Q. et al. Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.* **102**, 175506 (2009).
- Oganov, A. R. et al. Ionic high-pressure form of elemental boron. *Nature* **457**, 863–867 (2009).
- Drozdov, A., Erements, M., Troyan, I., Ksenofontov, V. & Shylin, S. I. Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. *Nature* **525**, 73–76 (2015).
- Duan, D. et al. Pressure-induced metallization of dense  $(\text{H}_2\text{S})_2\text{H}_2$  with high-Tc superconductivity. *Sci. Rep.* **4**, 6968 (2014).
- Sun, W. et al. A map of the inorganic ternary metal nitrides. *Nat. Mater.* **18**, 732 (2019).
- Villars, P. & Iwata, S. Pauling file verifies/reveals 12 principles in materials science supporting four cornerstones given by nature. *Chem. Met. Alloy.* **6**, 81–108 (2013).
- Snider, E. et al. Room-temperature superconductivity in a carbonaceous sulfur hydride. *Nature* **586**, 373–377 (2020).
- Ceder, G., Morgan, D., Fischer, C., Tibbetts, K. & Curtarolo, S. Data-mining-driven quantum mechanics for the prediction of structure. *MRS Bull.* **31**, 981–985 (2006).
- Morgan, D., Ceder, G. & Curtarolo, S. High-throughput and data mining with ab initio methods. *Meas. Sci. Technol.* **16**, 296–301 (2004).
- Fischer, C. C., Tibbetts, K. J., Morgan, D. & Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.* **5**, 641–646 (2006).
- Hofmann, D. W. & Apostolakis, J. Crystal structure prediction by data mining. *J. Mol. Struct.* **647**, 17–39 (2003).
- Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. Predicting crystal structures with data mining of quantum calculations. *Phys. Rev. Lett.* **91**, 135503 (2003).

16. Zhang, L., Chen, Z., Su, J. & Li, J. Data mining new energy materials from structure databases. *Renew. Sust. Energ. Rev.* **107**, 554–567 (2019).
17. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704–244704 (2006).
18. Oganov, A. R., Lyakhov, A. O. & Valle, M. How evolutionary crystal structure prediction works and why. *Acc. Chem. Res.* **44**, 227–237 (2011).
19. Falls, Z., Avery, P., Wang, X., Hilleke, K. P. & Zurek, E. The xtalopt evolutionary algorithm for crystal structure prediction. *J. Phys. Chem. C* **125**, 1601–1620 (2020).
20. Wang, Y., Lv, J., Zhu, L. & Ma, Y. Calypso: a method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063–2070 (2012).
21. Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
22. Goedecker, S. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **120**, 9911–9917 (2004).
23. Martoňák, R., Laio, A. & Parrinello, M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
24. Dahm, K.-W. & Marhold, W. Simulation of structural phase transitions by metadynamics. *Z. Kristallogr. Cryst. Mater.* **220**, 489–498 (2005).
25. Doll, K., Schön, J. C. & Jansen, M. Structure prediction based on ab initio simulated annealing. *J. Phys. Conf. Ser.* **117**, 12014 (2008).
26. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter.* **23**, 053201 (2011).
27. Zhang, W. et al. Unexpected stable stoichiometries of sodium chlorides. *Science* **342**, 1502–1505 (2013).
28. Niu, H., Oganov, A. R., Chen, X.-Q. & Li, D. Prediction of novel stable compounds in the Mg-Si-O system under exoplanet pressures. *Sci. Rep.* **5**, 18347 (2015).
29. Naumova, A. S., Lepeshkin, S. V., Bushlanov, P. V. & Oganov, A. R. Unusual chemistry of the C-H-N-O system under pressure and implications for giant planets. *J. Phys. Chem. A* **125**, 3936–3942 (2021).
30. Thompson, J. N. The evolution of species interactions. *Science* **284**, 2116–2118 (1999).
31. Allahyari, Z. & Oganov, A. R. Coevolutionary search for optimal materials in the space of all possible compounds. *NPJ Comput. Mater.* **6**, 55 (2020).
32. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
33. Gražulis, S. et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40**, D420–D427 (2012).
34. Morgen, R. A. The relation between pseudo binary lines and solid solutions in metallic ternary systems. *J. Am. Chem. Soc.* **49**, 39–43 (1927).
35. Bryan, R. F. Modern crystallography. Vol. 2. Structure of crystals edited by BK Vainshtein, VM Fridkin and VL Indenbom. *Acta Crystallogr. B. Struct. Sci.* **51**, 893–894 (1995).
36. Lyakhov, A. O. & Oganov, A. R. Evolutionary search for superhard materials: methodology and applications to forms of carbon and TiO<sub>2</sub>. *Phys. Rev. B* **84**, 092103 (2011).
37. Zhu, Q., Li, L., Oganov, A. R. & Allen, P. B. Evolutionary method for predicting surface reconstructions with variable stoichiometry. *Phys. Rev. B* **87**, 195317 (2013).
38. Reilly, A. M. et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr. B. Struct. Sci., Cryst. Eng. Mater.* **72**, 439–459 (2016).
39. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX-Evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
40. Niu, H. et al. Structure, bonding, and possible superhardness of CrB<sub>3</sub>. *Phys. Rev. B* **85**, 144116 (2012).
41. Kvashnin, A. G. et al. WB<sub>5-x</sub>: synthesis, properties, and crystal structure-new insights into the long-debated compound. *Adv. Sci.* **7**, 2000775 (2020).
42. Beattie, H. J. The crystal structure of an M<sub>3</sub>B<sub>2</sub>-type double boride. *Acta Crystallogr.* **11**, 607–609 (2010).
43. Idowu, O., Ojo, O. & Chaturvedi, M. Microstructural study of transient liquid phase bonded cast Inconel 738LC superalloy. *Metall. Mater. Trans. A* **37**, 2787–2796 (2006).
44. Hu, X. et al. Atomic-scale observation and analysis of chemical ordering in M<sub>3</sub>B<sub>2</sub> and M<sub>5</sub>B<sub>3</sub> borides. *Acta Mater.* **149**, 274–284 (2018).
45. Hu, X., Zhu, Y., Sheng, N. & Ma, X. The Wyckoff positional order and polyhedral intergrowth in the M<sub>3</sub>B<sub>2</sub>- and M<sub>5</sub>B<sub>3</sub>-type boride precipitated in the Ni-based superalloys. *Sci. Rep.* **4**, 7367 (2014).
46. Wang, B., Wang, D. Y., Cheng, Z., Wang, X. & Wang, Y. X. Phase stability and elastic properties of chromium borides with various stoichiometries. *ChemPhysChem* **14**, 1245–1255 (2013).
47. Cheng, X. et al. Interstitial-boron solution strengthened WB<sub>3+x</sub>. *Appl. Phys. Lett.* **103**, 171903 (2013).
48. Cheng, X.-Y., Chen, X.-Q., Li, D.-Z. & Li, Y.-Y. Computational materials discovery: the case of the W-B system. *Acta Crystallogr. C. Struct. Chem.* **70**, 85–103 (2014).
49. Born, M. & Huang, K. *Dynamical Theory of Crystal Lattices* (Oxford University Press, 1954).
50. Allegre, C. J., Poirier, J.-P., Humler, E. & Hofmann, A. W. The chemical composition of the earth. *Earth Planet. Sci. Lett.* **134**, 515–526 (1995).
51. Hernández, E. R., Brodholt, J. & Alfè, D. Structural, vibrational and thermodynamic properties of Mg<sub>2</sub>SiO<sub>4</sub> and MgSiO<sub>3</sub> minerals from first-principles simulations. *Phys. Earth Planet. Inter.* **240**, 1–24 (2015).
52. Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55**, 2083–2110 (1991).
53. Stixrude, L. & Lithgow-Bertelloni, C. Thermodynamics of mantle minerals-I. Physical properties. *Geophys. J. Int.* **162**, 610–632 (2005).
54. Stixrude, L. & Lithgow-Bertelloni, C. Thermodynamics of mantle minerals-II. Phase equilibria. *Geophys. J. Int.* **184**, 1180–1213 (2011).
55. Hernández, E. R., Alfè, D. & Brodholt, J. The incorporation of water into lower-mantle bridgmanites: a first-principles study. *Earth. Planet. Sci. Lett.* **364**, 37–43 (2013).
56. Wuchina, E., Opila, E., Opeka, M., Fahrenholtz, W. G. & Talmy, I. G. Uhtcs: ultra-high temperature ceramic materials for extreme environment applications. *Electrochem. Soc. Interface* **16**, 30–36 (2007).
57. Yu, X.-X., Weinberger, C. R. & Thompson, G. B. Ab initio investigations of the phase stability in group IVB and VB transition metal carbides. *Comput. Mater. Sci.* **112**, 318–326 (2016).
58. Zeng, Q. et al. Prediction of stable hafnium carbides: stoichiometries, mechanical properties, and electronic structure. *Phys. Rev. B* **88**, 214107 (2013).
59. Kostenko, M., Gusev, A. & Lukoyanov, A. Disorder-order and order-order phase transformations in Ta<sub>5</sub>C<sub>4</sub> phases predicted using the evolutionary algorithm and symmetry analysis. *Phys. Chem. Chem. Phys.* **22**, 24116–24132 (2020).
60. Smith, C. J., Yu, X.-X., Guo, Q., Weinberger, C. R. & Thompson, G. B. Phase, hardness, and deformation slip behavior in mixed Hf<sub>x</sub>Ta<sub>1-x</sub>C. *Acta Mater.* **145**, 142–153 (2018).
61. Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics (2001).
62. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
63. Cheng, B. et al. Mapping materials and molecules. *Acc. Chem. Res.* **53**, 1981–1991 (2020).
64. Woodley, S. M. & Catlow, R. Crystal structure prediction from first principles. *Nat. Mater.* **7**, 937–946 (2008).
65. Oganov, A. R., Ma, Y., Glass, C. W. & Valle, M. Evolutionary crystal structure prediction: overview of the USPEX method and some of its applications. *Psi-k. NewsL.* **84**, 142–171 (2007).
66. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
67. Sun, G. et al. Performance of the Vienna ab initio simulation package (VASP) in chemical applications. *J. Mol. Struct. THEOCHEM* **624**, 37–45 (2003).
68. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
69. Oganov, A. R. & Valle, M. How to quantify energy landscapes of solids. *J. Chem. Phys.* **130**, 104504 (2009).
70. Bell, R. P. The theory of reactions involving proton transfers. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **154**, 414–429 (1936).
71. Evans, M. G. & Polanyi, M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* **31**, 875–894 (1935).

## ACKNOWLEDGEMENTS

The authors would like to thank Linshuang Zhang and Yifu Gao for carefully reading the manuscript. This research was supported by The National Natural Science Foundation of China (grant No. 22003050) and the Research Fund of the State Key Laboratory of Solidification Processing (NPU), China (grant No. 2020-QZ-03). The calculations were performed in the International Center for Materials Discovery (ICMD) cluster of NPU. A.R.O. is supported by Russian Science Foundation (grant 19-72-30043).

## AUTHOR CONTRIBUTIONS

H.N. and A.R.O. conceived the research. H.N. and X.L. implemented the COPEX algorithm and performed the calculations. X.L., H.N., and A.R.O. wrote the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-021-00668-5>.

**Correspondence** and requests for materials should be addressed to Haiyang Niu or Artem R. Oganov.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021