The XTALOPT Evolutionary Algorithm for Crystal Structure Prediction

Zackary Falls, Patrick Avery, Xiaoyu Wang, Katerina P. Hilleke, and Eva Zurek*

Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260-3000, USA

E-mail: ezurek@buffalo.edu

*To whom correspondence should be addressed

Abstract

Significant progress has been made in the field of *a priori* crystal structure prediction, with a number of recent remarkable success stories. Herein, we briefly outline the methods that have been developed for finding the global minimum structure and interesting local minima without the need for experimental information. Focus is placed on describing the XTALOPT evolutionary algorithm (EA) developed in our group towards this end. XTALOPT is published under well-known open-source licenses, and the EA searches can be analyzed via the AVOGADRO chemical editor and visualizer. We describe new algorithmic developments that have made it possible to predict the structures of ever-more complex crystalline lattices. Benchmark tests, which clearly illustrate how the new developments improve the success rate and accelerate the discovery of the global minimum structure, are carried out. Finally, we describe how XTALOPT has been employed to predict novel ternary hydrides that have the propensity for high-temperature superconductivity under pressure.

Introduction

The last decade has witnessed tremendous advances in program packages that carry out first-principles calculations, spectacular speed-ups in computer hardware, and significant improvements in algorithms for *a priori* crystal structure prediction (CSP).[1–11] These developments have made it now possible to predict the structure of a crystal given only its composition for systems that are not too complicated; i.e. binary or even ternary atomistic systems whose primitive unit cells contain $\lesssim$ 50 atoms. Moreover, nowadays one can compute, with reasonable accuracy, many properties of a solid that can be used to predict its performance in a wide variety of applications, e.g. as a photovoltaic, in energy storage devices, or as a superconducting or superhard material. We are now in an exciting era where materials can be designed rationally *in silico* prior to their experimental synthesis!

This paragraph lists a handfull of the steadily growing examples of CSP-guided theoretical predictions that were later experimentally verified in the area of high-pressure research. The $I4/mmm$ symmetry $BaGe_3$ phase shown in Fig. 1(a), which is composed of a unique germanium lattice of edge-sharing triangular prisms, was predicted to be the most stable structure between 15.6-35.4 GPa.[12] Computations suggested that it is metastable and superconducting at 1 atm with a superconducting critical temperature, T_c, of 5.5 K. Within 1 year of its prediction, this phase was synthesized at 15 GPa and 1300 K, quenched to atmospheric conditions, and experiments showed it was superconducting below 6.5 K.[13] Another superconductor whose high-temperature high-pressure synthesis was inspired via CSP is the $Pnnm$ symmetry FeB_4 structure illustrated in Fig. 1(b).[14,15] First-principles calculations suggested that this material could be a phonon-mediated superconductor with a T_c of 15-20 K.[14] Following the successful synthesis of this material, its T_c was measured to be 2.9 K, and it was also found to be superhard with a nanoindentation hardness of 62 GPa.[16] Perhaps one of the most remarkable examples of materials-by-design is the $Fm\bar{3}m$ symmetry LaH_{10} phase shown in Fig. 1(c) whose hydrogenic lattice resembles a clathrate composed of H_{32} $[4^6 6^{12}]$ polyhedra. CSP calculations predicted it to be stable[17] or metastable[18] under pressure, and electron-phonon calculations suggested it would have a remarkably high T_c. Shortly

after the theoretical predictions were published, a high hydride of lanthanum whose properties were consistent with the $Fm\bar{3}m$ LaH_{10} structure was synthesized at 170 GPa.[19] Moreover, subsequent studies measured T_c values as high as 280 K at 200 GPa,[20] and 250 K at 170 GPa[21] in the lanthanum/hydrogen system under pressure. Another hydride with a nontraditional stoichiometry that was predicted[22] prior to its synthesis[23,24] is the $I4/mmm$ symmetry CaH_4 phase shown in Fig. 1(d), which contains both molecular and atomic hydrogen. The structures and propensity for superconductivity of most binary hydrides have been theoretically studied under pressure.[25–27]

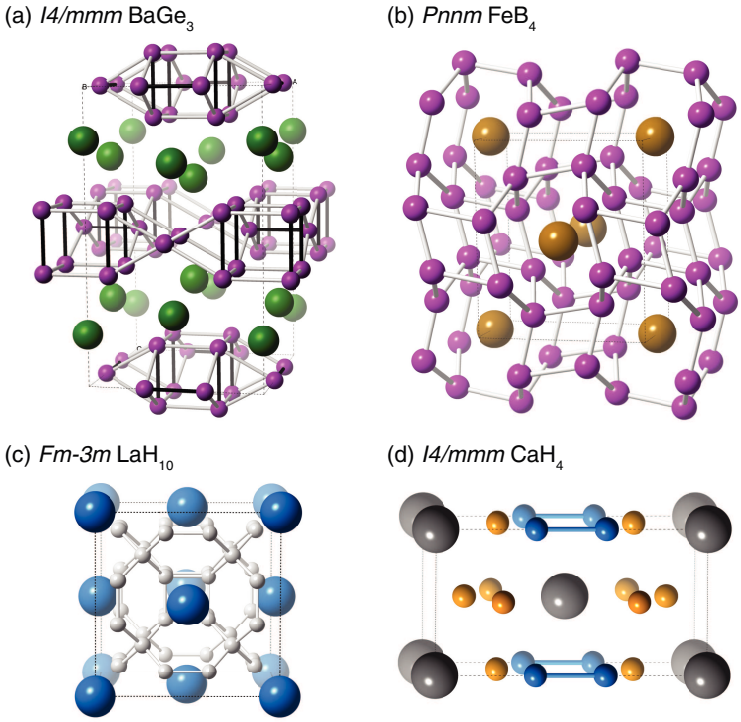


Figure 1: Materials that were first predicted via crystal structure prediction techniques, and later synthesized under high pressure: (a) $I4/mmm$ $BaGe_3$[12,13] (barium/germanium atoms are green/purple), (b) $Pnnm$ FeB_4[14–16] (iron/boron atoms are brown/pink), (c) $Fm\bar{3}m$ LaH_{10}[17–21] (lanthanum/hydrogen atoms are blue/white), (d) $I4/mmm$ CaH_4[22,23] (calcium/hydridic-hydrogen/molecular-hydrogen atoms are grey/yellow/blue).

CSP is extremely important in the high-pressure field[28–32] because there are many difficulties associated with the experimental determination of the atomistic structure of a synthesized phase subject to extreme conditions. For example, it may be impossible to distinguish between two elements with very similar masses via spectroscopic techniques, the positions of hydrogen atoms cannot be determined using X-ray diffraction (XRD), and neutron diffraction capabilities at very

high pressures are not yet available. In these situations, a synergistic feedback loop between experiment and theory is often required to characterize a sample. At times CSP may be the only tool available, for example to study matter at conditions that resemble those inside giant planets. Because the crystal structures and chemical combinations that are stable under pressure are so different than those found at 1 atm,[33–36] CSP techniques that make use of data found in large materials databases[37,38] have not been used for pinpointing the most stable stoichiometries and their crystalline structures in these extreme conditions in the past (however, a new method that uses a linear approximation to the enthalpy has been developed towards this end[39]) .

The main goal of this Feature Article is to describe the XTALOPT algorithm for CSP developed in our group,[40–46] along with its subprograms,[47,48] and to provide examples of some of the fascinating high-pressure materials it has pinpointed. We start with a concise overview of the computational methods that can be used to predict the most stable and important metastable crystalline lattices for a given composition without any experimental information, describe our latest algorithmic developments, and present the results of benchmark calculations that illustrate how these developments have increased the complexity of the structures that XTALOPT can reliably predict.

Computational Details

The benchmark calculations were carried out using XTALOPT release 12[45] along with parameter sets for the pool size, duplicate matching tolerances, and behavior of the evolutionary operators that were optimized in Refs.[40,47] The geometries were relaxed using either the General Utility Lattice Program (GULP),[49] or the Vienna *Ab-initio* Simulation Package (VASP).[50,51] The interatomic potential employed for TiO_2 was introduced in Ref.,[52] and for $SrTiO_3$ we used a modified version[40] of the potential in Ref.[53] The VASP calculations were carried out using density functional theory with the gradient-corrected exchange and correlation functional of Perdew-Burke-Ernzerhof (PBE)[54] coupled with the DFT-D3[55] van der Waals (vdW) correction for $LiBH_4$, and the vdW-DF-optPBE[56–59] functional for NaSi. The projector augmented wave (PAW) method[60] was used

to treat the core states, and a plane-wave basis set with a cutoff of 400 eV was employed for NaSi and 450 eV for $LiBH_4$. The k-point grids were generated using the Γ-centered Monkhorst–Pack scheme, and the number of divisions along each reciprocal lattice vector was chosen such that the product of this number with the real lattice constant was 30 Å in the structure searches. The Li $2s^1$, B $2s^2/2p^1$, H $1s^1$, Na $3s^1$ and Si $3s^2/3p^2$ electrons were treated explicitly using POTCARs provided in the potpaw-PBE.52.tar.gz file from the VASP repository.

Automated Methods for *a priori* Crystal Structure Prediction

A crystalline compound with N atoms in its unit cell can be described by three unit cell vectors and three angles, along with $3N-3$ parameters for the atomic coordinates. Finding the global minimum on the potential energy surface (PES), an example of which is shown in Fig. 2, corresponds to locating the values of the $3N+3$ variables that minimize the (free) energy for a given pressure and temperature. Because it is not computationally feasible to carry out first-principles calculations of the phonon modes for all of the hundreds or thousands of structures optimized in a first-principles CSP search, typically either the internal energy at 0 K, or the enthalpy at a given pressure is minimized, ignoring the zero point energy, and finite temperature contributions.

Since the shape of the PES is unknown, in principle one would need to calculate the energies of all of the minima to determine the one that has the lowest energy/enthalpy. However, because the number of local minima increases exponentially with N,[61] this is only possible for the simplest systems. It has been shown that finding the global minimum of homogeneous,[62] and heterogeneous[63] clusters is a nondeterministic polynomial-time hard (NP-hard problem), meaning that an algorithm scaling as a polynomial in the number of degrees of freedom cannot be found to solve it. In addition, the "no free lunch" theorem proposes that no single CSP algorithm works well for every system.[64] The exponential increase in the number of minima, combinatorial problem for multicomponent systems, and the impossibility of constructing a general-purpose algorithm appears discouraging for CSP.

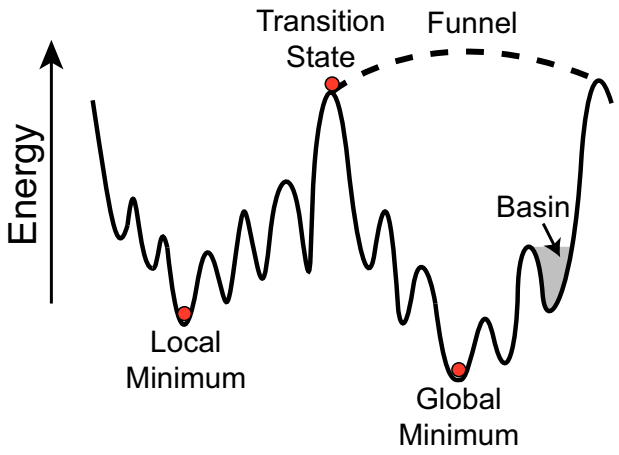


Figure 2: A schematic illustration of a 1 dimensional (1D) PES. It is composed of many local minima that are separated by barriers, and a single global minimum. A basin contains all of the configurations that optimize to a given local minimum. Two funnels comprise this PES, and the transition state between them is shown.

Nonetheless, a number of properties of PES' can be leveraged to construct clever CSP techniques that are able to find good approximations for the global minimum. First of all, a large portion of the multidimensional PES will be very high in energy and possess few local minima. These regions correspond to structures that are not chemically sensible, e.g. they have unrealistic interatomic distances, or assume configurations where atoms of the same type are "clustered" within regions of the unit cell. Therefore, it is desirable that unrealistic structures are removed from a CSP search prior to local optimization. If such a structure is mistakenly kept it will eventually relax into a region of the PES that contains local minima (provided the electronic self-consistent-field steps can converge). Secondly, it has been suggested that basins containing low energy local minima take up the most "space" in the PES.[65] This means that a randomly generated structure is more likely to fall in a low energy basin as compared to one with higher energy. Finally, according to the Bell-Evans-Polanyi principle[66,67] low energy basins are likely to lie close to each other, and the barriers between them are small. As a result small perturbations of a structure will lead to exploration of the basins within a given funnel, and this will ultimately result in the discovery of the lowest energy point. These properties suggest that if a given PES contains only one funnel, finding the global energy minimum should be relatively straightforward. If many funnels located

far away from each other are present the search will be much more difficult. However, a featureless PES that does not contain any regions of attraction is the most difficult to search.

Typically many local minima are discovered during the course of a single CSP search. This can be beneficial, as some CSP searches do not aim to find the global minimum, but rather metastable species with a particular property such as superhard,[68] or magnetic materials.[69,70] If the energies/enthalpies of the metastable structures are not too high, it may be possible to synthesize them by appropriately choosing the pressure, temperature, and starting material, or by employing synthesis techniques for phases far away from equilibrium.[71]

The algorithms that have been adapted towards CSP are well-known metaheuristics designed to find good solutions to diverse optimization problems spanning from circuit design to protein folding. None of them can guarantee that the global minimum has been found for all but the simplest systems. In the sections below we briefly outline automated methods for predicting the structures of 3D-periodic crystals where each atom can be treated individually. Many of these algorithms have been adapted to predict the structures of finite clusters,[72,73] 2D-materials,[74–76] surfaces[77,78] and clusters adsorbed to them,[79] as well as interfaces,[80] and microporous crystal frameworks.[81] Searching for the likely polymorphs of molecular crystals has its own considerations, for further information we refer the reader to Refs.[82–90] We also do not describe CSP methods that make use of information stored in large materials databases either explicitly,[37–39,91,92] or by building machine learning models,[93–96] nor those that require experimental information.[37,97,98]

Random Searching

In the simplest CSP method all $3N+3$ degrees of freedom of a user-defined number of crystals are generated randomly, and the structures are optimized to the nearest local minimum, as illustrated in Fig. 3(a).[9,31,32,99] To ensure that the structures generated are chemically sensible the cell volume, as well as the minimum interatomic distances are constrained. It is typically useful to generate random symmetric structures,[100] and constrain their Bravais lattices or spacegroups, especially if experimental data is available. In some cases it might be appropriate to use clusters or molecules

as the building blocks, and place them randomly within a cell. Pickard and Needs, who have pioneered this method, refer to it as "generating random sensible structures".[31]

In a purely random search every crystal is independent of another, and the algorithm does not learn from its history. To remedy this Pickard and Needs introduced an option to subject particularly stable structures to random mutations, such as atomic displacements or unit cell deformations in their *ab initio* random structure searching (AIRSS) method. This procedure, known as "shaking", makes it possible to explore all of the local minima within a given funnel, and incorporates learning into the search. The AIRSS method has been applied to a wide variety of systems including matter under high pressure,[32] battery materials,[101] and organic molecular solids.[102]

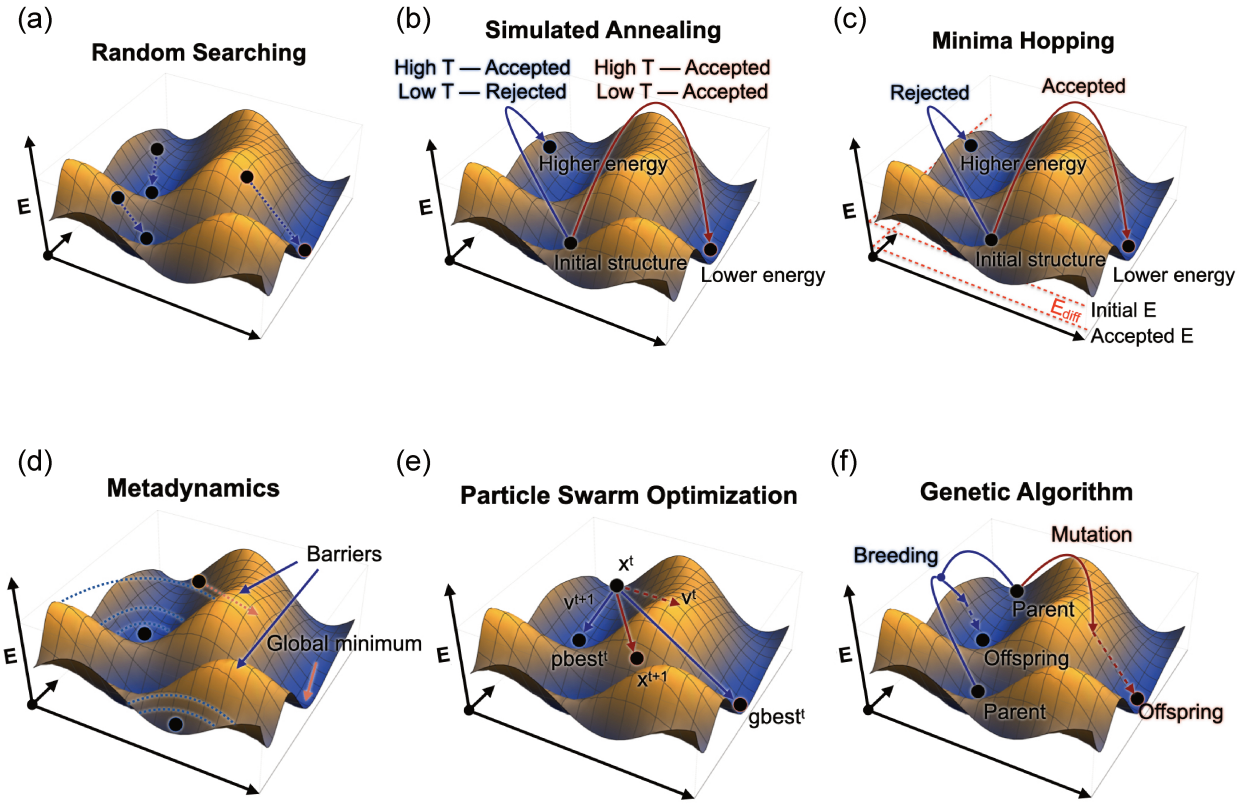


Figure 3: Schematic illustrations of various crystal structure prediction methods in a 2D PES. The dots represent structures, and the dashed arrows local optimizations: (a) random search, (b) simulated annealing, (c) minima hopping, (d) metadynamics, (e) particle swarm optimization, (f) genetic or evolutionary algorithms.

Simulated Annealing

The simulating annealing algorithm, see Fig. 3(b), mimics the metallurgical process of heating a substance until it melts, followed by controlled cooling, and subsequent crystallization.[103] This process removes defects and dislocations from a material, thereby increasing its ductility. Simulated annealing for CSP begins by generating a (typically random) structure whose energy is evaluated. Next, the structure is perturbed by randomly displacing atoms, changing unit cell parameters or permuting atoms of different types, and the energy of the new structure is evaluated. Molecular dynamics can also be employed to move atoms. In the Monte Carlo method of Metropolis the probability of acceptance of the new structure, $P = \exp(-\Delta E/k_{\mathrm{B}}T)$, is calculated using the energy difference between the two structures, ΔE, the Boltzmann constant, k_{B}, and the so-called simulation temperature, T.[104] A random number $0 < \epsilon < 1$ is chosen; if $\epsilon < P$ the structure is accepted, otherwise new structures are generated until one of them is accepted. As a result, T is not a physical temperature, but rather a variable used to control the rejection rate. A run begins with a large T, guaranteeing that most moves are accepted. During the run T is slowly decreased, thereby mimicking the physical annealing process. Finally, a quench run is carried out with $T = 0$ K, so that only downhill steps are accepted, and a local relaxation is performed.

To avoid getting stuck in local minima during the cooling procedure more complicated annealing schemes, which involve periodically raising the temperature, can be employed. Often, full exploration of the PES can only be achieved if numerous runs starting from different initial configurations are performed. Care must also be used in determining the size of the random mutation: if it is too small, the run will not be able to tunnel through barriers between nearby minima, and if it is too big the search essentially becomes random. Doll, Schön and Jansen have pioneered the use of this method, and applied it towards a plethora of crystalline inorganic systems[105] including metal pernitrides,[106] boron nitride,[107] and GeF_2.[108]

Minima Hopping

In Gödecker's minima hopping method[109,110] molecular dynamics is used to explore the PES, and each structure is optimized to the nearest local minimum. A structure's energy is compared with that of its predecessor, and the energy difference corresponding to an allowed move, E_{diff}, is adjusted so that half of the new structures are accepted, as shown in Fig. 3(c). This method keeps track of previously visited minima, and if the algorithm revisits them the kinetic energy employed for the molecular dynamics is increased so as to promote exploration of unchartered regions of the PES. It has been shown that aligning the initial molecular dynamics velocities along the directions of soft phonon modes helps accelerate the search.[111,112] The minima hopping method has been used extensively to predict the structures of materials under pressure such as superconducting $S_xSe_{(1-x)}H_3$ phases,[113] binary intermetallics that are immiscible at 1 atm,[39] and structural candidates for cold compressed graphite.[114] Minima hopping should not be confused with the similarly named basin hopping method, which has been applied extensively to finite clusters,[115] water-ice,[116] and is under active development for more complex systems.[117–119]

Metadynamics

Metadynamics, developed by Parrinello and co-workers, is an enhanced sampling molecular dynamics technique.[120–122] The sampling is accelerated by using a history-dependent bias potential that lifts previously visited areas of the PES. As shown in Fig. 3(d), metadynamics helps overcome barriers and enables the discovery of new minima by depositing a sum of Gaussians along the system's trajectory. This so-called 'basin-flooding' is usually carried out in a space defined by a few selected degrees of freedom, often referred to as the collective variables, such as the interatomic distances, angles, coordination numbers, and unit cell parameters. One drawback of metadynamics is that it may be necessary to understand the physics and chemistry of the process being studied to define the correct collective variables. Another disadvantage is that transition barriers between two or more basins could become flooded during the simulation, thereby preventing the global minimum from being found. Therefore, it may be necessary to carry out more than one simulation

starting from different configurations. Metadynamics has been used to study structural transformations of carbon under pressure,[123] as well as high pressure modifications of carbon-dioxide,[124] and germanium.[125] It has also been employed to simulate a wide range of rare events including structural phase transitions,[126] and chemical reaction mechanisms.[127]

Particle Swarm Optimization

The particle swarm method was inspired by the behavior of large groups of animals such as schools of fish, or flocks of birds.[128] Boldyrev's group was the first to apply this method to clusters,[129] and Ma's to solids using the CALYPSO (Crystal structure AnaLYsis by Particle Swarm Optimization) software package.[130] In the last decade CALYPSO has become one of the most popular tools for predicting the structures of a wide range of chemical systems including clusters, 2D layered materials, solids, interfaces, electrides, and superhard compounds.[30,131–133]

In CALYPSO, the trajectory a single individual takes to traverse the multidimensional PES depends on its position, $x(t)$, and velocity, $v(t)$, as well as the position and velocity of the other structures in the swarm, as illustrated in Fig. 3(e). The position of a structure at a later time, $t+1$, is given by

$$x(t+1) = x(t) + v(t+1) \tag{1}$$

where

$$v(t+1) = \omega v(t) + c_1 r_1 (p_{\text{best}}(t) - x(t)) + c_2 r_2 (g_{\text{best}}(t) - x(t)). \tag{2}$$

In Eq. 2 $p_{\text{best}}(t)$ is the position of the individual after local optimization, and $g_{\text{best}}(t)$ of the global minimum of the swarm at time t. The inertia weight, ω, which ranges from 0.4 to 0.9, is modified during the search. Large values of ω promote global exploration of the PES, and small ones local exploration. The random numbers r_1 and r_2 range between 0 to 1. The coefficients c_1 and c_2 weight the contribution of the individual's position versus that of the global minimum.

A PSO search begins by generating random structures that are locally optimized. Just as in the AIRSS method, the performance of the algorithm can be improved by using sensible choices

for the interatomic distances, and cell volumes, as well as employing symmetry in the random structure generation. Most of the subsequent structures in the search are generated via Eqs. 1 and 2, but random structures are also continuously added to the search to ensure that it does not get stuck in a particular region of the PES. The PSO algorithm is able to broadly sample the PES, while at the same time more thoroughly exploring the most promising areas, and it is able to learn from its history.

Evolutionary/Genetic Algorithms

Evolutionary algorithms (EAs), or genetic algorithms (GAs) adapt concepts from the theory of evolution, such as natural selection, mutation, and reproduction, so they may be employed within CSP. An EA begins by generating chemically sensible random geometries (seeding with specific structures is also possible), followed by optimization to the nearest local minimum by an external program. The computed energy or enthalpy of each structure is employed to determine its fitness, or probability to be chosen as a parent for procreation. This procedure allows the fittest structures to pass on their genetic information (in this case the structural motifs) to future generations. As shown in Fig. 3(f), children can be created by mutations of a single parent, or via combinations of two parents. Mutations, such as distorting the unit cell shape, permuting atoms of different types, or displacement of atoms either via a random-walk or periodic motion (such as a wave), typically perturb a structure only slightly, thereby allowing the algorithm to perform a more thorough local search. The real space cut-and-splice operation, which was introduced by Deaven and Ho for clusters,[134] and later adapted towards crystals by Glass, Oganov and Hansen,[135] is employed for breeding. This two parent operation can significantly perturb a structure, thereby enabling a broad exploration of the PES. Similar to the PSO technique, EAs are able to thoroughly sample the PES, while simultaneously zooming in on the most promising areas, and they can be biased if structural information is known.

In the last fifteen years many groups have released EAs that can be used to predict the structures of atomistic crystals given only their composition. These include USPEX (Universal Structure Pre-

dictor: Evolutionary Xtallography),[8,135–138] GASP,[139,140] MAISE,[14] EVO,[141] as well as algorithms by Trimarchi and Zunger,[142–145] Abraham and Probert,[146] Fadda and Fadda,[147] Woodley and Catlow[52] , Wentcovitch *et al.*,[148] and Hammer *et al.*[149] In what follows we describe the specifics of the XTALOPT EA developed in our group.[40–46]

The XTALOPT Evolutionary Algorithm

The XTALOPT EA, written in C++, was first developed as an extension to the AVOGADRO[150,151] molecular editor, builder and visualizer, and made use of the OPENBABEL[152] chemical toolkit, which has many useful features including conversion between various file formats used to represent chemical data. XTALOPT was initially released as an open-source program under the GNU Public License (GPL),[153] making it freely available for use and collaboration by members of the scientific community. In the latest versions of XTALOPT,[44,45] which have been released under the more flexible 3-clause Berkeley Software Distribution (BSD) license,[154] OPENBABEL is used as an external executable, and AVOGADRO2[155] can optionally be employed to visualize crystals.

XTALOPT generates children via the two-parent cut-and-splice breeding operation,[135] as well as a number of single parent mutations. The latter consist of the hybrid operators "stripple" (strain+ripple), where the shape of the unit cell is modified while simultaneously displacing the atoms in a periodic wave, and permustrain (permutation+strain), in which atoms of two different types are exchanged and the cell's shape is changed. A number of variables in these operators are user-defined, and have been previously described in Ref.[40] and Ref.[2] XTALOPT employs a population-based pool, rather than a generation-based pool, since the former makes better use of a high performance computing environment.[2,40] Searches can be "seeded" with user-defined structures, and geometric constraints can be enforced. The Spglib[156] package is employed for spacegroup identification. Herein, recent developments to the XTALOPT[40,41,157] EA will be described. These modifications have made it possible to discover the global minimum of crystals whose unit cells are larger and more complex than those that could previously be uncovered reliably, enabled searches containing discrete molecular units, and made it possible to predict novel

crystalline materials with a high Vickers hardness.

Our latest developments have improved the graphical user interface (GUI) within XTALOPT to allow greater on-the-fly analysis of the progress of the evolutionary search, and also provided a command line interface for users who do not wish to use a GUI. The initial generation of structures can now be created using custom interatomic distance constraints for atom pairs, and simple molecular units can be employed if desired. These options give the user greater ability to fine-tune the area of the PES to be explored during the course of the EA search. The mitosis method, which enhances local order in large unit cells, and RANDSPG algorithm, which generates random symmetric structures, are two further techniques that can be used in the initial creation of structures, and they can both dramatically speed up the search for more complex systems. Duplicate structures that are removed from the breeding pool are now identified using the XTALCOMP algorithm, which directly compares two structures instead of relying on an indirect fingerprinting scheme. Finally, a method for predicting hard and stable materials using macroscopic hardness models that employs, as input, machine learned values of the shear modulus has been implemented. In the following sections we describe each of the aforementioned developments, and present the results of benchmark calculations that show how well they work. Finally, we give a few examples of some high pressure crystalline lattices that have been predicted using XTALOPT .

New Developments

Graphical vs. Text Based Interface

To make it easy for experts and non-experts alike to set up and carry out CSP searches, inspect generated and optimized individuals on-the-fly, visualize and analyze the results, and modify parameters during the run in real time, XTALOPT was initially paired with an easy-to-use GUI. A number of tabs that are associated with a given function comprise the GUI; some of these are illustrated in Fig. 4. In the 'Structure Limits' tab, Fig. 4(a), information regarding the composition and structural parameters, such as the unit cell size and permitted interatomic distances, can be defined.

Constraints that can be used in the generation of the first random set of individuals, which greatly accelerate the prediction of certain classes of structures, can also be input in this tab. These include the mitosis function, RANDSPG option, and the molecular unit builder, which are described fully in the sections below. In the 'Optimization Settings' panel, the geometry optimization scheme that is to be used during an XTALOPT run can be configured. Supported codes include GULP, which employs interatomic potentials, as well as the first-principles VASP, PWscf, CASTEP, and SIESTA codes, and a generic optimizer option is available since version r12.[45]

The 'Search Settings' panel, Fig. 4(b), specifies options that are employed in the XTALOPT run. This includes the number of random structures that are produced in the first generation, the total number of structures to be created in the search, parameters that determine the behavior of the evolutionary operators, selection of seed structures, and tolerances for symmetry and duplicate structure detection. If multiple formula units are to be investigated, parameters governing crossover, and shared gene pools can be set here. Vickers hardness can also be calculated and used to determine fitness.

The 'Progress' and 'Plot' tabs (Figs. 4(c,d)) can be used to monitor an XTALOPT run in real-time, or to analyze a finished search. In the 'Progress' tab, a continuously updated table of each generated structure including its optimization status, enthalpy, space group, and ancestry (parent structures and operators used to generate it) is found. Clicking on a structure and formula unit will allow the user to replace, kill, or modify it as desired. In addition, a structure's XRD pattern can be generated (Fig. 4(e)). XTALOPT automatically checks to see if AVOGADRO2 is running on the user's computer; if it is the coordinates of the crystal are automatically sent to AVOGADRO2 when the user clicks on the rows in the table, or on the points in the plot (Fig. 4(f)). In the 'Plot' tab, a number of variables, including enthalpy, structure number, and cell volume can be chosen, and the information can be plotted on either axis. Structures can be labeled by their Hermann-Mauguin space group symbol, enthalpy, Vickers hardness, and more. A user may thus track the progress of an XTALOPT run over time, and identify useful trends. For example, if low-enthalpy structures are clustered around the lower bound of permitted volumes, the user might decrease this bound in the

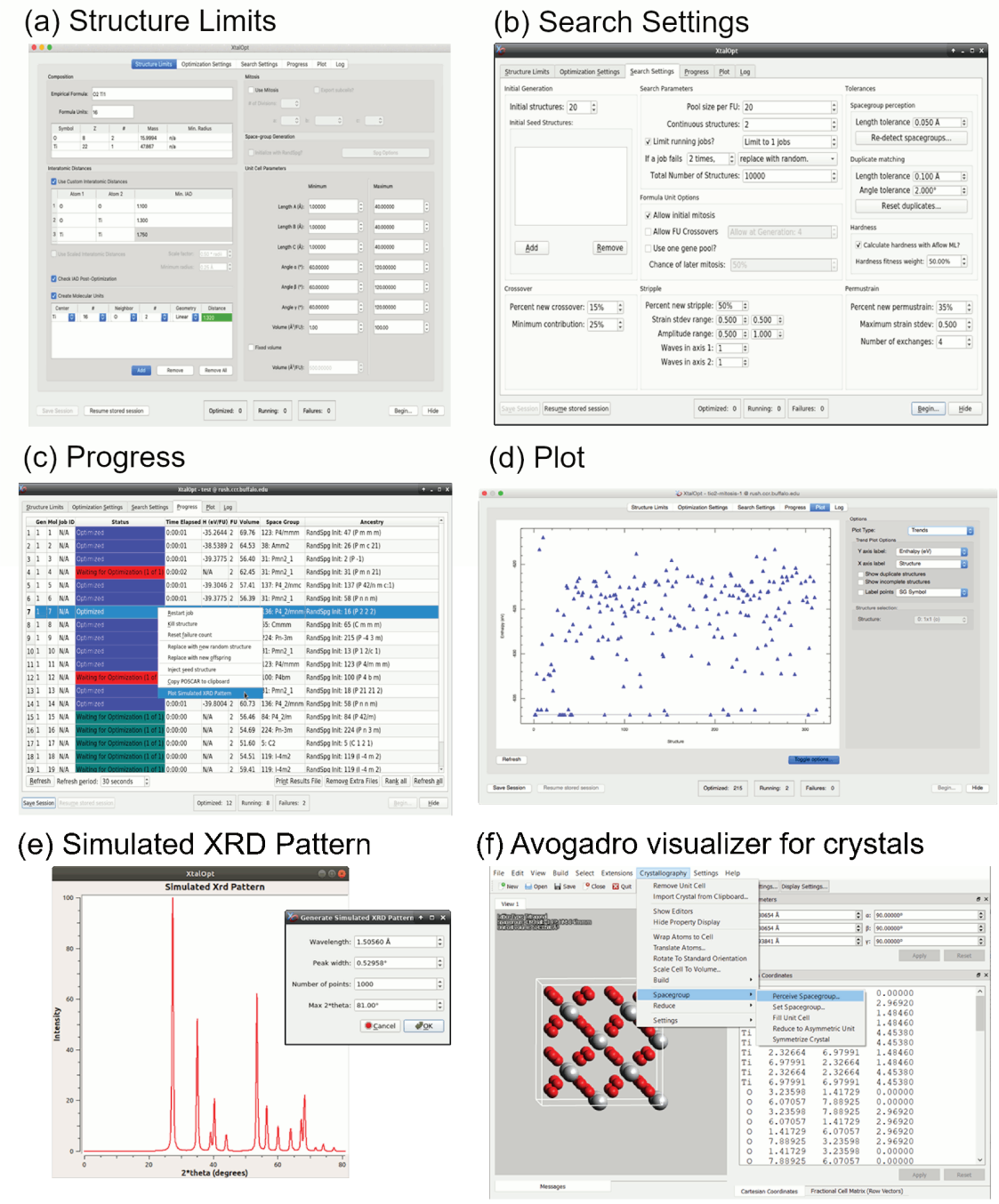


Figure 4: Select tabs from the graphical user interface of XTALOPT showing options for (a) initializing a structure search, (b) choosing the search settings, (c) visualizing its progress, and (d) plotting the results. Additional tabs allow the user to set options for the local relaxations, and compute cluster configuration. (e) The X-ray diffraction patterns for the structures produced can be generated on-the-fly, and (f) they can be visualized and analyzed using AVOGADRO.

'Structure Settings' tab.

XTALOPT can also be run through a command line interface (CLI), first implemented in version

r11.[44] An input file, detailing options found in the GUI, is required along with template files for queue interfaces, and optimization schemes. These options are printed to the terminal window, along with job submissions, progress, and errors. A text file that can be edited at any time to update the desired settings is maintained in the working directory. CLI runs can be imported into the GUI for post-analysis. In both the GUI and CLI versions, results are stored in another continuously updated text file (results.txt), which includes most data found in the 'Progress' tab of the GUI. Typically EA runs carried out with the CLI version are faster than those performed with the GUI. The reason for this is that the regular updating and reloading of the GUI results table requires more processing power as the search space grows. Additionally, the CLI mode removes the need for GUI dependencies, and can be run fully on a remote cluster, thereby reducing network communication processes and further increasing the speed.

Interatomic Distance

To avoid generating non-physical structures with unrealistically small distances between atoms, and hasten exploration of the chemically relevant regions of the PES, minimum interatomic distances (IAD) were initially derived within XTALOPT from sums of covalent radii of two elements extracted from a database. Although the IAD constraints could be changed according to a user-defined scaling factor, this scaling factor was applied equally to all covalent radii in the system, precluding distinct cutoffs for specific pairs of elements. With the inclusion of a new custom IAD option in XTALOPT version r11,[44] each element pair may be assigned an individual IAD restriction through an editable table, as illustrated in Fig. 4(a). In the first random generation, unit cells are created according to the user specified constraints, and subsequently filled with atoms. A check is performed to ensure compliance with the IAD restrictions. If the check fails, the program attempts a different placement, iterating until all atoms are placed within acceptable IAD limits or the structure is discarded following 1000 failed attempts. Minimum IAD values can also optionally be checked following the local optimization steps. If this option is chosen, structures that do not meet the IAD criteria will be marked as failed, and will be excluded from the breeding pool.

Molecular Unit Builder

Many of the well-known molecular CSP algorithms have been tailored towards organic crystals, such as the targets in the Cambridge Crystallographic Data Centre (CCDC) blind tests.[82–87] Discrete units are also ubiquitous in important classes of inorganic materials including complex hydrides,[158,159] expanded metal compounds,[160] hybrid organic-inorganic perovskites,[161,162] and polyhedral clusters such as those present in Zintl compounds,[163] cluster-containing intermetallics,[164] white phosphorus,[165] and various phases of boron.[166] A molecular unit building option, which was implemented in XTALOPT version r10,[43] was designed to facilitate the prediction of inorganic compounds that contain such basic discrete units. This algorithm creates molecules that assume any of the Valence Shell Electron Pair Repulsion (VSEPR) geometries up to octahedral in the first random generation. The stoichiometry, coordination, and geometry assumed by the molecular unit, the identities of the central atom and the atoms at the vertices, as well as the distances between them, need to be defined, as shown in Fig. 4(a). A "no-center" molecular unit may also be specified by choosing "None" for the central unit. This option can be useful for searches on clusters that do not have a central atom, such as polyanion-containing Zintl phases. In some situations, the evolutionary operators can break the discrete units apart. However, because the energies of the resulting structures will typically be high, they will not be added to the breeding pool. This means that the EA search will preferentially explore regions of the PES corresponding to intact molecules. The molecular unit option is compatible with the mitosis operator used to construct large unit cells, but cannot be used in concert with RANDSPG .

Consider, for example a SiH_6 composition with four formula units for which it is desirable to bias the search towards compounds containing $4SiH_4$ tetrahedra and $4H_2$ molecules. All four Si atoms would be chosen as the central atom, each coordinated with four neighboring hydrogen atoms at a user specified distance, required to fall within the IAD criterion. The particular geometric arrangement of the hydrogens surrounding the Si can then be defined – in this case tetrahedral, see-saw, or square planar are all possibilities. Random structures are generated by first placing the central atoms and surrounding neighbors, followed by the remaining atoms (in this case, the eight

hydrogens not involved in the molecular SiH_4 units) randomly throughout the unit cell.

Mitosis

In the original implementation of XTALOPT the placement of atoms within the unit cell in the first generation was done completely randomly, subject to the IAD restraints. However, for large complex cells this method yields structures that are likely to have virtually no local order, as illustrated in Fig. 5(a). Such disordered parents produce offspring with overly high inherent disorder and low diversity, thereby slowing down the identification of more ordered lower enthalpy structures. To overcome this problem, Oganov and co-workers introduced the "cell-splitting" technique and implemented it within USPEX.[137,167] In this method the original set of random structures is generated via replicating smaller subcells to form a full unit cell that is more ordered, as shown in Fig. 5(b). These smaller subcells contain 15-20 atoms and several splittings are employed concurrently. When the stoichiometry of the system is not conducive towards generating identical subcells, atoms are deleted to maintain the correct composition.

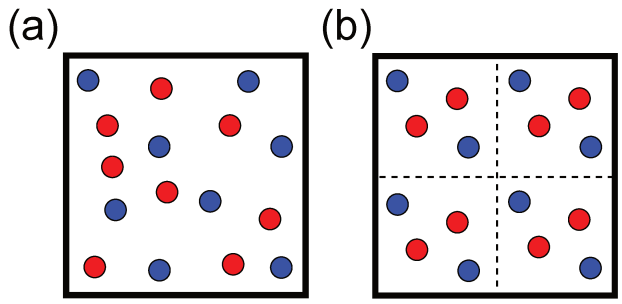


Figure 5: An example of a unit cell that could be generated (a) purely randomly, and (b) by using mitosis. The dashed black lines in (b) highlight the replicated subcells .

A similar technique, dubbed "mitosis", was introduced within XTALOPT version r9.[42] The algorithm determines the number of formula units present. The user defines how many subcells will be replicated, and the number of replications along each lattice vector that will be employed to make the supercell. The code is also able to handle stoichiometries that lack a common denominator and therefore a simple reduction into single formula units. For example, the stoichiometry

Na_3H_{14} can be defined as two units of NaH_7 with an extra Na atom. In this case, the mitosis method will produce a two-cell supercell based on an NaH_7 subcell, and a single Na atom will be added randomly to the supercell, subject to the IAD constraints.

RANDSPG Implementation

Inorganic crystals tend to adopt symmetric structures, and enforcing symmetry in CSP algorithms can speed up finding the global minimum. For instance, a search among the entries in the Inorganic Crystal Structure Database (ICSD) in 2006 revealed that over a quarter of inorganic crystals assume one of the $Pnma$, $P2_1/c$, $Fm\bar{3}m$, and $Fd\bar{3}m$ space groups, while only 1% crystallize in the $P1$ space group.[168] Many CSP codes that employ symmetry in the initial generation place atoms on sets of Wyckoff positions that belong to space groups consistent with the desired stoichiometry. In USPEX atoms are placed on general Wyckoff positions, and the symmetry operations of the space group are applied to determine the equivalent positions.[167] Atoms that wind up too close to one another are merged into a single site with averaged coordinates. The initialization procedure of CALYPSO[169] works similarly, and atoms are placed on combinations of Wyckoff positions compatible with the system stoichiometry, although with the interesting caveat that repeated space groups are disfavored by 80% to enforce coverage of a wide swath of search space.

RANDSPG , an algorithm that can be used as a stand-alone structure generation protocol or in tandem with XTALOPT ,[48] first determines all possible combinations of Wyckoff positions for the given composition, and groups those with the same multiplicities and uniquenesses to speed this search. Atoms are placed in lattices that are commensurate with the space group symmetry according to one of the combinations of Wyckoff positions, subject to the IAD criteria. Because some space groups share Wyckoff positions with smaller multiplicities, the most general Wyckoff position must be used at least once in RANDSPG to ensure the correct space group is generated. However, the option to use the most general Wyckoff position at least once was turned off in XTALOPT to allow smaller formula unit crystals to be used for space groups that have a large multiplicity for the most general Wyckoff position. To compensate for this, SPGLIB is employed

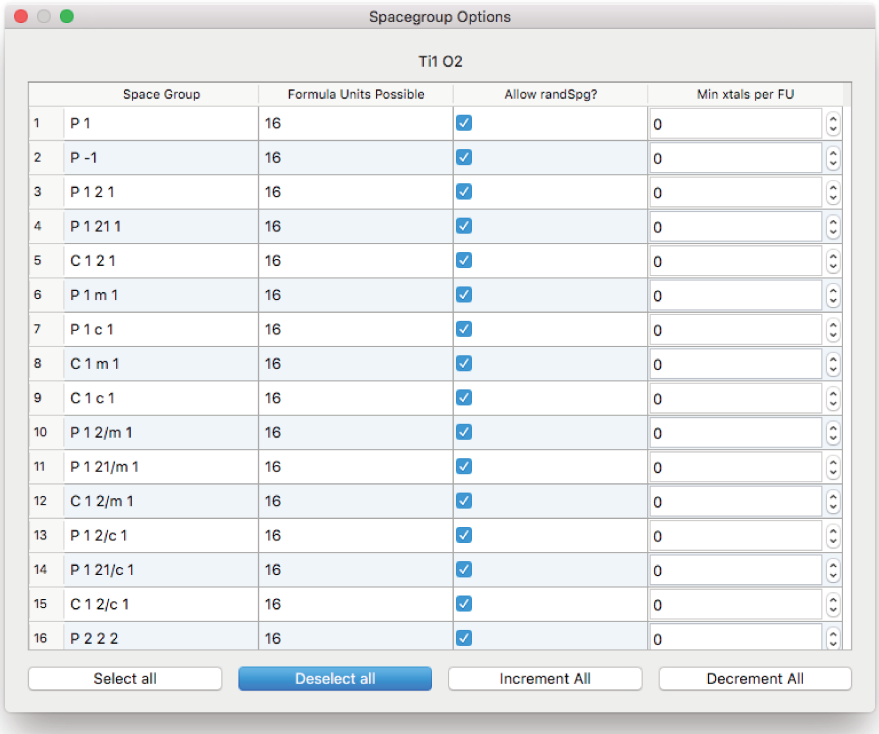


Figure 6: The Spacegroup Options table in XTALOPT .

for symmetry detection to ensure the correct spacegroup is generated. The spacegroups are chosen randomly unless the "Spg Options" button in Fig. 4(a) is checked. In this case, the Table in Fig. 6, which gives the user further control over the search, appears. The table displays the formula units that may produce certain space groups, allows the user to decide how many structures of each spacegroup will be generated, and omits spacegroups that are incompatible with the composition and the formula units specified. If RANDSPG fails to create a requested spacegroup XTALOPT will skip over it. If the number of generated crystals is not enough for the defined initial population size after all requested space groups have been attempted, randomly selected space groups will be made. When initializing an XTALOPT search with RANDSPG , it is recommended that a large initial population is employed because the algorithm samples the PES widely, creating structures that are both very high and very low in enthalpy, which are unlikely to be generated randomly without symmetry constraints.

In the standalone version of RANDSPG the user may choose not to enforce occupation of the most general Wyckoff position, or to specify that a particular element occupies a user defined Wyckoff position. Although these options are not included in the RANDSPG version interfaced with XTALOPT, they can be used to create structures that can serve as "seeds" in a standard EA search.

Niching with XTALCOMP

Maintaining diversity in the gene pool used for CSP is of utmost importance since the search can become biased towards regions of the PES that were initially sampled, leading to convergence on a structure that is not the global minimum. However, enforcing too much diversity is a double-edged sword, as the failure of the most stable structures to proliferate can slow exploration of low lying regions of the PES. Together, this indicates the need for accurate, reliable, and well-tested duplicate identification procedures.

In the initial release of XTALOPT , structural fingerprints based on unit cell volumes, space groups and enthalpies were compared, resulting in many false positives. More accurate fingerprinting methods include those based on pair distribution functions and diffraction patterns as in USPEX,[167] or bond characterization matrices derived from bond angles and lengths as in CALYPSO.[169] Other packages introduce a penalty term for similar structures,[146] or maintain lists of coordinates and lattice parameters against which new structures are continually checked.[139]

Since XTALOPT version r8 the XTALCOMP library[47,170] has been used to identify duplicates via direct mapping of structures onto one another. Two crystals with standardized orientations and Niggli-reduced cells are compared within XTALCOMP . Candidate transforms that map a set of reference vectors from one unit cell onto the lattice vectors of the reference cell are determined. If one of these results in every atom from the transformed cell being mapped onto one in the reference cell, the two are flagged as duplicates. If no such transformation is found, the two are deemed to be unique. The tolerance for the mapping criterion can be edited in the 'Search Settings' tab of XTALOPT shown in Fig. 4(b). Tests have revealed that an EA search using the XTALCOMP

algorithm outperforms one that employs the original approximate duplicate matching scheme.[47] Moreover, it has been shown that XTALCOMP can differentiate between phases differing by the position of a single atom, with only one of the phases being dynamically stable at atmospheric conditions, highlighting the importance of an "exact" duplicate matching scheme.[12]

Searches with Multiple Formula Units

One of the difficulties in CSP is that the number of formula units (FUs) in the global minimum for a given chemical composition is often unknown. To address this problem a number of CSP algorithms including CALYPSO,[169] USPEX[167] and GASP[139] have developed techniques enabling simultaneous searches on multiple FUs. The details of these implementations vary, and herein we describe the method first available in version r10 of XTALOPT.[43] A related problem is that the stoichiometries that are stable and metastable also need to be determined. Towards this end Trimarchi and Zunger introduced a technique to scan the composition space.[144] Subsequently, derivative methods have been introduced in many CSP codes. Because such searches would be extremely computationally demanding this method was not implemented within XTALOPT , and we suggest that a separate run is carried out for each stoichiometry instead.

In the 'Structure Limits' tab, shown in Fig. 4(a), the number of FUs to be considered in the XTALOPT search must be specified. By default, each set of FUs maintains a separate gene pool, although stable structures can seed the gene pools of other FUs if they can be accessed via primitive reduction or supercell generation. Crossovers between structures with different FUs, as shown in Fig. 7, can also be performed if this option is chosen in the 'Search Settings' tab (Fig. 4(b)). In such a situation, the sum of the contributions of the two parent crystals is not required to be 100%, as the FU of the offspring must only be a valid choice, and not necessarily the same as that of the parents. The option to allow FU crossover can be turned on partway into a run, for example at a user-defined generation, to allow individual FUs to first find locally stable structures before being influenced by other FUs.

All FUs can also share a single gene pool, comparing structures based on their enthalpy per

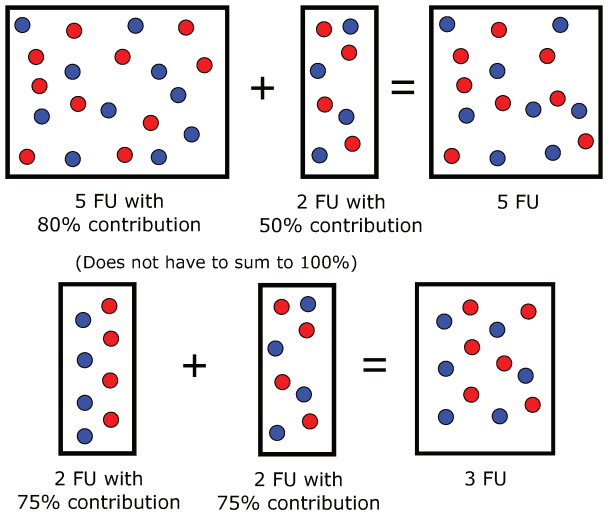


Figure 7: Formula unit crossover in XTALOPT . Top: Generation of 5 FU offspring unit cell from one 5 FU parent crystal with 80% contribution, and 2 FU parent crystal with 50% contribution. Bottom: Two parent crystals with 2 FUs and 75% contribution each produce an offspring structure with 3 FUs.

FU. Because this option biases the search toward FUs that have generated the most stable structures thus far, it is best to turn it on later in the run, once each FU has sufficiently explored its PES. This approach may enable the search to evolve towards the most stable FU. If the 'allow initial mitosis' checkbox in the 'Search Settings' tab is chosen, supercells will be generated from smaller FUs. The main difference between this implementation of mitosis, and the one chosen in the 'Structure Limits' tab, is that here the small FUs are first optimized, and the larger cell is generated from a supercell that has been subject to mutations. Sharing a gene pool also enables the option 'Chance of later mitosis: $n\%$', which results in an $n\%$ chance that a structure selected for procreation will be employed to build a mutated supercell.

Searching for Superhard Materials

In addition to finding particularly stable compounds, CSP algorithms can also be tailored to search for materials with desired properties. Because superhard systems are important in various applications including cutting and polishing, or as abrasives and in armor, the widely used CSP programs USPEX,[68,171] and CALYPSO[172] have implemented a technique for their prediction. Knoops or

Vickers hardnesses are estimated using microscopic models that require as input quantities that can be obtained from the crystal structure, or rapidly computed.[173,174] Instead of solely using a thermodynamic quantity to determine a structure's fitness, the hardness values are also considered. However, the microhardness models used to estimate hardness are known to fail, as in the case of T-carbon, a porous structure that was erroneously predicted to be superhard by such a model.[175] A macroscopic model, on the other hand, gave values that agreed with physical reasoning.[176]

Macroscopic hardness models, which are based on elastic properties such as shear and bulk moduli, are far more robust than microscopic ones. Unfortunately, first principles calculations of the elastic moduli[177,178] are extremely time consuming, because up to 24 geometry optimizations may be necessary for a single structure. To remedy this problem version r12 of XTALOPT[45] has been interfaced[179] with a machine learning (ML) model,[180] trained on materials within the AFLOW (Automatic FLOW) database, to predict a crystal's Voigt-Reuss-Hill average values of the shear, G_{VRH}, and bulk, B_{VRH}, moduli.[181,182] Tests on structures whose hardness values are known, as well as crystals generated during the course of an evolutionary search illustrated that the Teter model[183,184] provides the most robust estimate of the Vickers hardness, via the equation $H_{\text{v}} = 0.151 G_{\text{VRH}}$, when ML values of the moduli were employed.[185]

It is desirable to predict superhard materials that could be made in experiments, meaning that they are local minima whose energies are not too high. Therefore, the new fitness function implemented within XTALOPT considers both the energy/enthalpy of an individual, E_i, as well as its ML Vickers hardness, $H_{\text{v},i}$. The probability, p_i, that a structure is chosen for procreation is then calculated as:

$$p_i = N \left[1 - w \left(\frac{H_{\text{v,max}} - H_{\text{v},i}}{H_{\text{v,max}} - H_{\text{v,min}}} \right) - (1 - w) \left(\frac{E_i - E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}} \right) \right], \tag{3}$$

where $H_{\text{v,max}}$/$H_{\text{v,min}}$, and E_{max}/E_{min} are the highest and lowest Vickers hardnesses, and energies or enthalpies in the breeding pool. The user defined weighting factor, w, can be chosen to favor either stable or hard structures in the search, and the normalization constant, N, is determined so

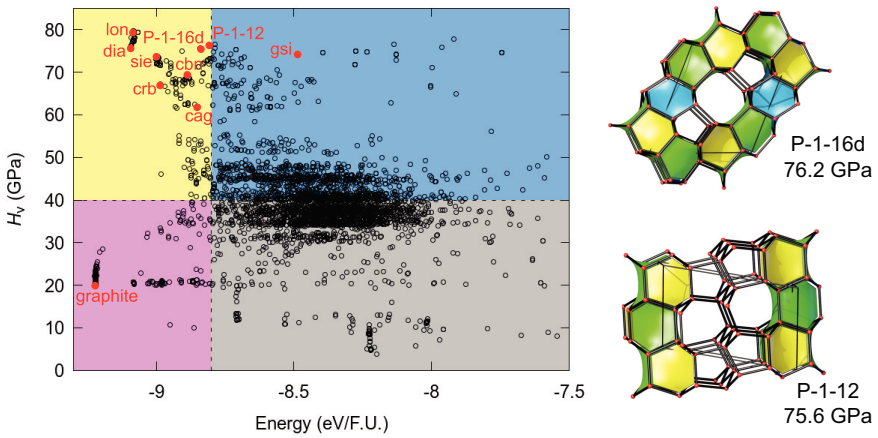


Figure 8: (left) The Vickers hardness, H_v, vs. the energy of structures generated during the CSP searches carried out on the carbon system in Ref.[185] The horizontal dashed line corresponds to $H_\mathrm{v} = 40$ GPa, and the vertical line to $E = -8.80$ eV/atom. The phases in the yellow quadrant are low in energy and superhard. Some experimentally known or previously predicted structures are pointed out: lonsdaleite (lon), diamond (dia), Z-carbon (sie), W-carbon (cnw), M-carbon (cbn), bct-C_4 (crb), Y-carbon (cag), and bc8 (gsi). (right) Two of the newly predicted superhard phases along with their machine-learned H_v. The cages colored in blue are related to diamond, and those in yellow and green are related to lonsdaleite.

that $\sum p_i = 1$. Choosing the option 'Calculate the hardness with AFLOW ML?' in the 'Search Settings' tab will prompt XTALOPT to obtain the ML-derived shear moduli of optimized crystals via a RESTful API[179] so their hardness values may be estimated.

This implementation was tested on the carbon system, and searches were carried out on cells containing 8, 12, 16, and 20 FUs.[185] The enthalpies and Vickers hardness values of all of the optimized crystals are plotted in Fig. 8. Thirty-six previously known structures, and forty-three new, low energy, metastable, superhard phases were identified. Many of the novel phases, two of which are shown in Fig. 8, were composed of bits of diamond, lonsdaleite (hexagonal diamond), or other known superhard materials.

Benchmark Calculations

Quantifying the performance of CSP methods is difficult because of their stochastic nature. Consider an EA search where there is always a finite, albeit potentially small, probability that the

global minimum structure appears in the first random set of individuals, or not at all. Moreover, because various parameters in the evolutionary operators are chosen randomly, as are the parents, each search may explore different areas of the PES. As a result, an EAs performance can only be quantified if many searches using the same parameter set are carried out and the results are analyzed statistically. Unfortunately, these types of benchmarks on large and complex unit cells are too expensive for first-principles methods. Therefore, we have chosen systems where reliable interatomic potentials are available,[52] in particular a 16 FU supercell of TiO_2 (48 atoms), for which the rutile structure is the global minimum, and a 10 FU supercell of the ternary $SrTiO_3$ (50 atoms), which crystallizes in a perovskite structure. These systems have previously been used in benchmarks carried out with various CSP codes.[40,47,48,167,169]

A number of metrics can be employed to quantify how well a particular CSP algorithm performs, with the most obvious being the success rate. Others include the best and average number of structures required to find the global minimum, as well as the standard deviation. It can also be illuminating to determine what percentage of structures correspond to the global minimum when they are generated quasi-randomly, i.e. using the same constraints as those in the CSP search. Comparison with the trajectory of the EA illustrates the advantage of the evolutionary approach, and how the parameters and options employed shape its trajectory.

Towards this end, we randomly generated 9000 structures with the $Ti_{16}O_{32}$ stoichiometry, and carried out 100 XTALOPT searches with 600 total structures per run using the aforementioned developments and interatomic potentials parameters. Table 1 provides the metrics for the six different sets of searches performed. In the "normal" run the shortest interatomic distances were determined using a scaling factor of 0.4 multiplied by the sum of the covalent radii. These results can be compared to searches using custom minimum IAD constraints, and those that employed RANDSPG for the first generation. The mitosis method was tested using three different sized subcells. Previous benchmarks on this system showed that a larger number of structures in the initial generation accelerated the search if it was performed using RANDSPG , but a smaller number of structures was beneficial otherwise.[48] The reason for this is that RANDSPG samples the PES more thoroughly,

whereas large unit cells that are created purely randomly tend to have a high degree of similarity. Therefore, for the "normal", minimum IAD, and mitosis runs 20 random structures were used in the initial generation, whereas for RANDSPG this number was increased to 50. Because rutile does not contain any discrete molecular units, it is unlikely that the molecular unit builder would improve the search, so this development was not tested.

Table 1: Benchmark tests carried out on TiO_2 with 16 formula units in the unit cell.

	Random [a]	Structure Search [b]			
Method	% Min.	% Success	Best Finish	Average Finish	σ
normal [c]	0.28%	100%	6	97	59
min. IAD [d]	1.39%	100%	1	62	64
RANDSPG	3.40%	100%	1	47	55
mitosis – $2 \times 1 \times 1$	0.99%	100%	2	99	76
mitosis – $2 \times 2 \times 1$	10.36%	100%	1	16	24
mitosis – $2 \times 2 \times 2$	28.82%	100%	1	4	3
USPEX – normal [e,f]	—	100%	—	80	69
USPEX – symmetry [e,f]	—	100%	—	77	76
USPEX – cell splitting [e,g]	—	100%	—	41	40
CALYPSO – normal [h,i]	0.00%	100%	—	500	—
CALYPSO – symmetry [h,i]	6.25%	100%	—	220	—

[a] 9000 structures were randomly generated and locally optimized to the nearest stationary point.
[b] 100 separate evolutionary searches were carried out, each with 600 locally optimized structures. The searches that used RANDSPG generated 50 random individuals; all other searches employed 20 random individuals. The volume was fixed to 30.3125 $Å^3$ per FU.
[c] A scaling factor of 0.4 multiplied by the sum of the covalent radii yielded the following minimum distances: Ti-Ti 1.28 Å, O-O 0.52 Å, Ti-O: 0.90 Å.
[d] IAD employed: Ti-Ti 3.00 Å, O-O 2.00 Å, Ti-O: 1.00 Å.
[e] Reference 167.
[f] 100 searches were conducted with up to 400 structures per run.
[g] 73 searches were conducted with "split-factors" of 2 and 4; soft-mutation was employed.
[h] Reference 169.
[i] 3250 structures were locally optimized to benchmark the generation of random structures. The PSO algorithm was benchmarked by carrying out 10 searches until the global minimum was found.

Before discussing the results of the evolutionary runs, let us take a quick look at how well these parameters performed in a quasi-random search. In the "normal" run 25 out of 9000 $Ti_{16}O_{32}$ structures, or 0.28%, corresponded to rutile, suggesting that $\sim$360 individuals would need to be generated to have some assurance the global minimum would be found at least once. Rutile was

generated 125 times (1.39%) using the custom IAD option, clearly illustrating that simple chemically motivated constraints are extremely beneficial within CSP. With RANDSPG the probability of finding rutile increased to 3.40%. It should be noted that this number is highly dependent upon the exact structural constraints applied. In the manuscript introducing the RANDSPG technique a success rate of 1.78% was obtained, but it was also shown that this number can be as high as 23.8% when the exact cell parameters and interatomic distances found in rutile are used.[48] Because the mitosis method generates subcells consisting of 8, 4, and 2 FU and uses them to create $2 \times 1 \times 1$, $2 \times 2 \times 1$ and $2 \times 2 \times 2$ supercells, respectively, it is easy to understand why mitosis dramatically improved the chances of finding rutile randomly as compared to the "normal" run: the problem is effectively reduced to randomly creating the 2 FU primitive cell of rutile within the subcell.

All of the CSP searches identified rutile as the global minimum. For each benchmark the best finish was less than 20, meaning that rutile was found at least once in the randomly generated set of structures. The average finish for mitosis with 8 FU was similar to that of the "normal" run. As compared to the "normal" search, custom IAD constraints decreased the average number of structures needed to find rutile by 36%, and RANDSPG decreased this number by a little over 50%. For mitosis with 4 and 2 FU supercells the average finish was smaller than 20 suggesting that a purely random search is likely sufficient to locate the global minimum using these approaches.

Comparing XTALOPT to another popular evolutionary algorithm, USPEX,[167] shows that custom IAD constraints yielded similar average finishes to a normal USPEX run (which employs these constraints by default). RANDSPG gave significantly lower average finishes (47) than the technique used to generate random symmetric structures in USPEX (77). Finally, "split-factors" of 2 and 4, corresponding to subcells containing 8 and 4 FUs, found the global minimum by 41 structures on average in USPEX , which is not too different than the average taken for the mitosis method using subcells of the same size (58). Comparing to the CALYPSO[169] algorithm, XTALOPT was able to locate the global minimum much more quickly both in a "normal" run and when symmetry constraints were employed (97 vs. 500, and 47 vs. 220 structures, respectively).

As shown in Table 2, for the more complex ternary $SrTiO_3$ with 10 FU and 50 atoms in the

unit cell the global minimum structure could not be found even once when 9000 individuals were generated quasi-randomly using both the "normal" settings, and via RANDSPG . When chemically motivated minimum IAD constraints were employed, the perovskite structure was generated once, and this number increased to six when a 5 FU subcell was used to create a $2 \times 1 \times 1$ supercell. Comparing with the results obtained for the 48 atom TiO_2 supercell clearly shows that the ternary is much more difficult to predict than the binary. Not surprisingly, when a 2 FU subcell was replicated $5 \times 1 \times 1$ times the success rate improved to 19.20% (1741 structures).

Table 2: Benchmark tests carried out on $SrTiO_3$ with 10 formula units in the unit cell.

	Random[a]	Structure Search [b,c]			
Method	% Min.	% Success	Best Finish	Average Finish	σ
normal [d]	0%	69%	208	892	351
min. IAD [e]	0.01%	53%	59	739	392
RANDSPG	0%	78%	44	805	339
mitosis – $2 \times 1 \times 1$	0.07%	81%	10	643	410
mitosis – $5 \times 1 \times 1$	19.20%	100%	1	6	5
USPEX – cell splitting [f,g]	—	94%	—	524	297

[a] 9000 structures were randomly generated and locally optimized to the nearest stationary point.
[b] 100 separate evolutionary searches were carried out, each with 1500 locally optimized structures. All of the searches generated 50 random individuals. The unit cell volumes were constrained to 59-60 $Å^3$ per FU.
[c] The average finish is based upon the completed runs, and did not include the runs that did not find the global minimum.
[d] A scaling factor of 0.4 multiplied by the sum of the covalent radii yielded the following minimum distances: Sr-Ti: 1.42 Å, Sr-O: 1.04 Å, Sr-Sr: 1.56 Å, Ti-O: 0.90 Å, Ti-Ti: 1.28 Å, O-O: 0.52 Å
[e] IAD employed: Sr-Ti: 2.50 Å, Sr-O: 1.50 Å, Sr-Sr: 3.00 Å, Ti-O: 1.20 Å, Ti-Ti: 3.00 Å, O-O: 1.50 Å
[f] Reference 167.
[g] 35 searches were conducted with about 1000 structures per run, and using "split-factors" of 2 and 4.

Turning to the performance of the EA search, the original implementation of XTALOPT found perovskite $SrTiO_3$ in only 7-12% of runs that optimized 1000 structures each.[40] Table 2 shows the results obtained here for searches that generated 50 random structures and optimized a total of 1500 individuals. Clearly, the developments described above dramatically increased the success rate, even though not all of the searches found the global minimum. Therefore, the average finish

and standard deviation columns should be taken with a grain of salt, as the numbers were obtained using only the data from the successful runs. The "normal" search had a slightly higher success rate than when custom IAD restraints were employed (69% vs. 53%), but it had a significantly larger "best finish" (208 vs. 59 individuals). The IAD success rate could be improved by using a different set of minimum IAD parameters, however care must be taken to avoid values that are too restrictive and lead to a large number of individuals being discarded prior to local optimization. Using RANDSPG in the first generation increased the success rate to 78% and decreased the "best finish" to 44, highlighting again the power of random symmetric initialization in a CSP search. The mitosis method was used to create subcells consisting of 5 and 2 FU (containing 25 and 10 atoms, respectively) and replicate them accordingly. The former had a success rate similar to that of RANDSPG , and unsurprisingly the latter was the only method that consistently located the global minimum with a success rate of 100%.

$SrTiO_3$ was also employed to benchmark USPEX , with runs that used "split-factors" of 2 and 4 (subcells containing 25 and $\sim$12 atoms).[167] These parameters closely resemble our mitosis $2 \times 1 \times 1$ and $5 \times 1 \times 1$ settings. USPEX attained a success rate of 94% after 1000 structures were optimized, and this value increased to 100% when when the runs generated 1750 individuals. This suggests that the performance of cell-splitting in USPEX is similar to mitosis within XTALOPT .

Because we were not able to find reliable potential parameters for inorganic materials that contain the types of discrete molecular entities for which the molecular unit builder in XTALOPT was designed, it was necessary to carry out benchmarks using Density Functional Theory (DFT) calculations. Unfortunately, it would be prohibitively expensive to carry out hundreds of DFT-based EA searches, making it impossible to obtain statistical data about the performance of the molecular unit builder. Nonetheless, a comparison of the results of a single EA run carried out using the molecular unit builder, to one without is still able to clearly illustrate how this development accelerates the search.

Our tests showed that the molecular unit builder coupled with DFT optimizations successfully predicted the crystal lattices illustrated in Fig. 9: $Pnma$ $LiBH_4$ ($Z = 4$),[186] and $C2/c$ NaSi

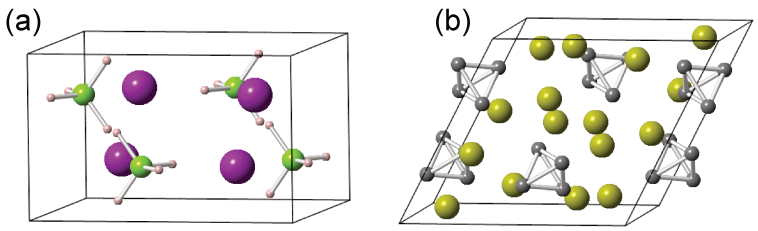


Figure 9: The global minima for XTALOPT searches carried out using the new molecular unit builder implementation: (a) $Pnma$ $LiBH_4$,[186] and (b) $C2/c$ NaSi.[187] Boron/lithium/hydrogen/silicon/sodium atoms are colored green/purple/pink/grey/yellow.

($Z = 16$).[187] When BH_4^- tetrahedra were treated as discrete units in the randomly generated set of individuals, the global minimum was found by the 249^{th} structure. For comparison, the most stable phase located in an atomistic search wherein 630 individuals were optimized was found on the 521^{th} relaxation. This structure has $P1$ symmetry and contains two BH_4^- units, one B_2H_6 unit and an H_2 unit, and it was 56 meV/atom less stable than the global minimum. The prototypical Zintl phase of NaSi was found by XTALOPT by the 583^{rd} structure when molecular units were employed. When they were not, the lowest energy structure obtained after 600 relaxations was still 16 meV/atom higher in energy than $C2/c$ NaSi. These tests show that the molecular unit builder dramatically accelerates the prediction of crystals containing clusters or simple molecules.

Applications of XTALOPT

Because EA based CSP techniques learn from their history, whereas purely random searches do not, the former typically outperform the latter. Metadynamics, simulated annealing, and minima hopping perform best when exploring a limited region of the PES, so multiple runs with different starting configurations may be required to predict a crystal whose structure is totally unknown. The PSO technique and EAs are the methods of choice for predicting novel materials without any experimental information because they are able to broadly sample the PES while at the same time thoroughly exploring the most promising areas. The XTALOPT EA has been designed for predicting the structures of three dimensional crystals where the atoms can be treated independently, or those containing simple single-centered molecular units. When interfaced with a program that

optimizes molecules using first-principles calculations the most complex systems our group has carried out searches on were ternaries with up to $\sim$125 atoms in the unit cell. If experimental lattice parameters are known, the mitosis method can be employed, or if simple molecular units are present, systems that are more complex can potentially be considered. However, for predicting molecular crystals,[87,89] when data from large materials databases can be used to accelerate the search,[37–39] or when searching for emerging ordered structures in a disordered system, for example, during crystal nucleation[188,189] other algorithms may be preferred. Interfacing CSP methods with machine-learning interatomic potentials is sure to increase the complexity of the structures that can be predicted in the near future.[190,191]

With this in mind, currently the XTALOPT EA has successfully been employed to predict the structures of a number of atomistic inorganic crystals, a few recent examples can be found in Refs.[192–203] Our group has mostly applied XTALOPT to study hydrides under conditions of extreme pressures,[157,204–220] because they have the potential to behave as conventional (Bardeen-Cooper-Schrieffer) high-temperature superconductors.[26,27] This line of reasoning can be traced back to Ashcroft who pointed out that hydrogen,[221] and hydrogen-dominant alloys[222] could have all of the properties required for conventional high-temperature superconductivity (large electron-phonon coupling, high density of states at the Fermi level, wide bands and high vibrational frequencies).

Whereas the metallization of hydrogen has been fraught with difficulties, tremendous advances have been made in the search for superconductivity in compressed hydrides. CSP techniques coupled with DFT calculations have predicted the most stable and interesting metastable binary hydrides and studied their propensity for superconductivity.[26] Two regions of the periodic table turn out to be particularly promising. The first are alkaline earth or rare earth polyhydrides, where exceedingly large T_cs have been predicted for compounds that are composed of hydrogenic clathrate lattices; one of these, $Fm\bar{3}m$ LaH_{10}, is illustrated in Fig. 1(c). The other region is found in the p-block elements, with a hydride of sulfur being the binary with the highest T_c (203 K near 150 GPa). This $Im\bar{3}m$ H_3S phase was synthesized[223] at about the same time it was theoretically predicted.[224] As shown in Fig. 10(a) it is formed from two interpenetrating perovskite lattices where each S

atom is octahedrally coordinated by H atoms, and each H atom is bonded to two S atoms.

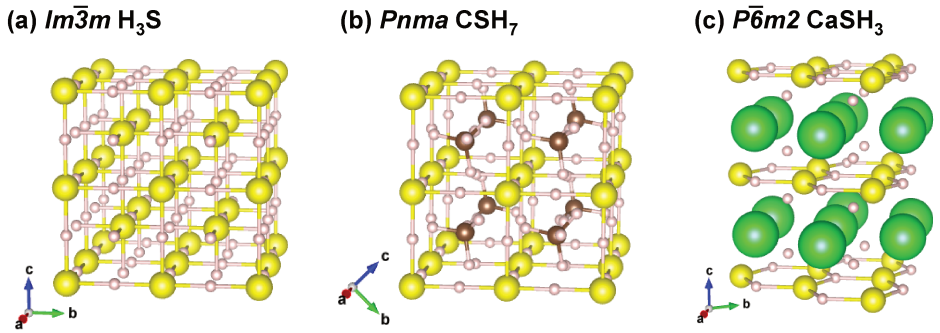


Figure 10: Superconducting high-pressure hydride phases based on H-S lattices: (a) $Im\bar{3}m$ H_3S,[224] (b) $Pnma$ CSH_7,[219] and (c) $P\bar{6}m2$ $CaSH_3$. Sulfur/hydrogen/carbon/calcium atoms are colored yellow/white/brown/green.

Ternary hydrides are the next frontier, and the developments that have been made in CSP are enabling the theoretical exploration of their phase diagrams. Recently, we have uncovered two classes of metastable hydrides related to H_3S as shown in Fig. 10(b,c). Our group,[219] and others[225] have found that a number of CSH_7 stoichiometry phases, differing in the number of formula units comprising the unit cell and the orientations of the methane molecules intercalated within the H_3S framework emerge as being metastable at $\sim$100 GPa. Electron-phonon calculations predicted T_cs as high as 194 K at 150 GPa,[219] and 181 K at 100 GPa.[225] These results suggest that other superconducting materials could be attained by varying the chemical identity of the intercalant, including noble gases[226] or other small molecules. Remarkably, experiments have recently measured room-temperature superconductivity (T_c as high as 288 K at 267 GPa) in one or more hydrogen rich CSH_x phases that have yet to be characterized.[227] Currently, it is not clear if the CSH_7 phases predicted in Refs.[219,225] could account for the experimental results since their computed T_cs decrease with increasing pressure, but experimentally the opposite trend is observed. One reason for the discrepancy might be that the computations were carried out in the static lattice approximation, neglecting the effect of temperature, quantum nuclei, and anharmonic effects, which are known to be important for the properties of high-pressure hydrides.[228]

The second ternary phase we predicted is composed of honeycomb HS layers separated by layers of CaH_2.[220] It was computed to be thermodynamically stable at 300 GPa when the zero-

point energy was included, and remained metastable down to 128 GPa, at which point it achieved its highest calculated T_c of 100 K. We suggested that replacing S by Se or Te, and/or Ca by Mg, Sr or Ba would yield other superconducting materials that might be stable over a wider pressure range.

Conclusions

Spectacular advances in algorithms for *a priori* crystal structure prediction, coupled with developments in computer hardware and software have facilitated inverse-materials design. Recent materials-by-design success stories show a glimpse of what the future may hold. In this Feature Article we have described some of the main computational techniques that can be used for automated crystal structure prediction without using any experimental information. This includes the generation of random lattices, simulated annealing, minima hopping, metadynamics, particle swarm optimization and evolutionary algorithms (EAs). Among the advantages of EAs is that they can both broadly sample the potential energy landscape, as well finely explore the most promising regions. Moreover, they can be biased with crystal prototypes or constrained by experimental observables.

Our focus is on the XTALOPT EA developed in our group, which is available to the scientific community under well-recognized open source licenses (GPL releases 10 or earlier, BSD releases 11 and 12). XTALOPT is cross-platform, can be used on various queuing systems, and has been interfaced with a number of programs designed for periodic lattices (e.g. GULP for interatomic potentials, and VASP, PWscf, CASTEP, SIESTA for first-principles calculations). The XTALOPT EA can be executed either using an intuitive and flexible graphical user interface (GUI) or a command line interface (CLI). The results of each type of run can be analyzed and post-processed via the GUI that is based on the AVOGADRO molecular editor, builder and visualizer.

We describe the GUI features available including functionality to visualize and manipulate crystals, and visualize their calculated XRD patterns within AVOGADRO. Moreover, we outline

new developments, including generating the initial set of structures using custom interatomic distance constraints, random symmetric structures, or discrete molecular units that assume any of the Valence Shell Electron Pair Repulsion geometries up to octahedral. We further show that replicating small subcells to create larger supercells in the first generation can accelerate locating the global minimum of large, complicated unit cells, and discuss how searches can consider multiple formula units simultaneously. The XTALCOMP duplicate matching scheme is described. Finally, we present a novel algorithm that employs a machine-learned estimate of a material's Vickers hardness, in conjunction with the computed energy or enthalpy to search for superhard materials.

Benchmark calculations on binary and ternary compounds with lattices containing $\sim$50 atoms in the unit cell are carried out to test our new developments. The tests clearly show improvements in the success rate, and decrease in the average number of structures that need to be optimized to locate the global minimum. We demonstrate that using molecular units in the first random generation accelerates the discovery of the most stable structures of $LiBH_4$ ($Z = 4$) and the Zintl phase NaSi ($Z = 16$). Finally, we briefly outline our applications of XTALOPT to hydride materials under pressure, and discuss two new families of superconducting ternary compounds that it has discovered.

Acknowledgments

We acknowledge the NSF (DMR-1827815) for financial support, and the Center for Computational Research (CCR) at SUNY Buffalo for computational support.[229] K. H. is thankful to the U.S. Department of Energy, National Nuclear Security Administration, through the Capital-DOE Alliance Center under Cooperative Agreement No. DE-NA0003858 for financial support.

References

(1) Oganov, A. R.; Pickard, C. J.; Zhu, Q.; Needs, R. J. Structure Prediction Drives Materials Discovery. *Nat. Rev. Mater.* **2019**, *4*, 331–348.

(2) Zurek, E. In *Reviews in Computational Chemistry*; Parrill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2016; Vol. 29; pp 274–326.

(3) Woodley, S. M.; Catlow, R. Crystal Structure Prediction from First Principles. *Nat. Mater.* **2008**, *7*, 937–946.

(4) Schön, J. C.; Doll, K.; Jansen, M. Predicting Solid Compounds Via Global Exploration of the Energy Landscape of Solids on the *ab initio* Level Without Recourse to Experimental Information. *Phys. Status Solidi B* **2010**, *247*, 23–39.

(5) Revard, B. C.; Tipton, W. W.; Hennig, R. G. *Structure and Stability Prediction of Compounds with Evolutionary Algorithms*; Topics in Current Chemistry; Springer Berlin Heidelberg, 2014; pp 181–222.

(6) Oganov, A. R. *Modern Methods of Crystal Structure Prediction*; John Wiley & Sons, 2011.

(7) Jansen, M. Conceptual Inorganic Materials Discovery – A Road Map. *Adv. Mater.* **2015**, *27*, 3229–3242.

(8) Oganov, A. R.; Lyakhov, A. O.; Valle, M. How Evolutionary Crystal Structure Prediction Works – and Why. *Acc. Chem. Res.* **2011**, *44*, 227–237.

(9) Pickard, C. J.; Needs, R. J. *Ab initio* Random Structure Searching. *J. Phys.: Condens. Mat.* **2011**, *23*, 053201.

(10) Oganov, A. R. Crystal Structure Prediction: Reflections on Present Status and Challenges. *Faraday Discuss.* **2018**, *211*, 643–660.

(11) Catlow, C. R. A. Modelling and Predicting Crystal Structures. *Interdiscip. Sci. Rev.* **2015**, *40*, 294–307.

(12) Zurek, E.; Yao, Y. Theoretical Predictions of Novel Superconducting Phases of $BaGe_3$ Stable at Atmospheric and High Pressures. *Inorg. Chem.* **2015**, *54*, 2875–2884.

(13) Castillo, R.; Baranov, A. I.; Burkhardt, U.; Cardoso-Gil, R.; Schnelle, W.; Bobnar, M.; Schwarz, U. Germanium Dumbbells in a New Superconducting Modification of $BaGe_3$. *Inorg. Chem.* **2016**, *55*, 4498–4503.

(14) Kolmogorov, A. N.; Shah, S.; Margine, E. R.; Bialon, A. F.; Hammerschmidt, T.; Drautz, R. New Superconducting and Semiconducting Fe-B Compounds Predicted with an *ab initio* Evolutionary Search. *Phys. Rev. Lett.* **2010**, *105*, 217003.

(15) Bialon, A. F.; Hammerschmidt, T.; Drautz, R.; Shah, S.; Margine, E. R.; Kolmogorov, A. N. Possible Routes for Synthesis of New Boron-Rich Fe-B and $Fe_{1-x}Cr_xB_4$ Compounds. *Appl. Phys. Lett.* **2011**, *98*, 081901.

(16) Gou, H.; Dubrovinskaia, N.; Bykova, E.; Tsirlin, A. A.; Kasinathan, D.; Schnelle, W.; Richter, A.; Merlini, M.; Hanfland, M.; Abakumov, A. M. et al. Discovery of a Superhard Iron Tetraboride Superconductor. *Phys. Rev. Lett.* **2013**, *111*, 157002.

(17) Liu, H.; Naumov, I. I.; Hoffmann, R.; Ashcroft, N. W.; Hemley, R. J. Potential High-T_c Superconducting Lanthanum and Yttrium Hydrides at High Pressure. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6990–6995.

(18) Peng, F.; Sun, Y.; Pickard, C. J.; Needs, R. J.; Wu, Q.; Ma, Y. Hydrogen Clathrate Structures in Rare Earth Hydrides at High Pressures: Possible Route to Room-temperature Superconductivity. *Phys. Rev. Lett.* **2017**, *119*, 107001.

(19) Geballe, Z. M.; Liu, H.; Mishra, A. K.; Ahart, M.; Somayazulu, M.; Meng, Y.; Baldini, M.;

Hemley, R. J. Synthesis and Stability of Lanthanum Superhydrides. *Angew. Chem. Int. Ed.* **2018**, *57*, 688–692.

(20) Somayazulu, M.; Ahart, M.; Mishra, A. K.; Geballe, Z. M.; Baldini, M.; Meng, Y.; Struzhkin, V. V.; Hemley, R. J. Evidence for Superconductivity above 260 K in Lanthanum Superhydride at Megabar Pressures. *Phys. Rev. Lett.* **2019**, *122*, 027001.

(21) Drozdov, A. P.; Kong, P. P.; Minkov, V. S.; Besedin, S. P.; Kuzovnikov, M. A.; Mozaffari, S.; Balicas, L.; Balakirev, F. F.; Graf, D. E.; Prakapenka, V. B. et al. Superconductivity at 250 K in Lanthanum Hydride under High Pressures. *Nature* **2019**, *569*, 528–531.

(22) Wang, H.; Tse, J. S.; Tanaka, K.; Iitaka, T.; Ma, Y. Superconductive Sodalite-like Clathrate Calcium Hydride at High pressures. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6463–6466.

(23) Ye, X.; Zarifi, N.; Zurek, E.; Hoffmann, R.; Ashcroft, N. W. High Hydrides of Scandium under Pressure: Potential Superconductors. *J. Phys. Chem. C* **2018**, *122*, 6298–6309.

(24) Wu, G.; Huang, X.; Xie, H.; Li, X.; Liu, M.; Liang, Y.; Huang, Y.; Duan, D.; Li, F.; Liu, B. et al. Unexpected Calcium Polyhydride CaH_4: A Possible Route to Dissociation of Hydrogen Molecules. *J. Chem. Phys.* **2019**, *150*, 044507.

(25) Zurek, E. Pushing Towards Room-temperature Superconductivity. *Physics* **2019**, *12*, 1.

(26) Zurek, E.; Bi, T. High-Temperature Superconductivity in Alkaline and Rare Earth Polyhydrides at High Pressure: A Theoretical Perspective. *J. Chem. Phys.* **2019**, *150*, 050901.

(27) Bi, T.; Zarifi, N.; Terpstra, T.; Zurek, E. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2019; pp 1–36.

(28) Zurek, E.; Grochala, W. Predicting Crystal Structures and Properties of Matter under Extreme Conditions via Quantum Mechanics: The Pressure is on. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2917–2934.

(29) Zhang, L.; Wang, Y.; Lv, J.; Ma, Y. Materials Discovery at High Pressures. *Nat. Rev. Mater.* **2017**, *2*, 17005.

(30) Wang, Y.; Ma, Y. Perspective: Crystal Structure Prediction at High Pressures. *J. Chem. Phys.* **2014**, *140*, 040901.

(31) Pickard, C. J.; Needs, R. J. Structures at High Pressure from Random Searching. *Phys. Status Solidi B* **2009**, *246*, 536–540.

(32) Needs, R. J.; Pickard, C. J. Perspective: Role of Structure Prediction in Materials Discovery and Design. *APL Mater.* **2016**, *4*, 053210.

(33) Grochala, W.; Hoffmann, R.; Feng, J.; Ashcroft, N. W. The Chemical Imagination at Work in *Very* Tight Places. *Angew. Chem. Int. Ed.* **2007**, *46*, 3620–3642.

(34) Zurek, E. In *Handbook of Solid State Chemistry*; Dronskowski, R., Ed.; Wiley-VCH Verlag GmbH & Co: Weinheim, Germany, 2017; Vol. 5; pp 571–605.

(35) Hermann, A. In *Reviews in Computational Chemistry*; Parrill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2017; pp 1–41.

(36) Miao, M.; Sun, Y.; Zurek, E.; Lin, H. Chemistry under High Pressure. *Nat. Rev. Chem.* **2020**, *4*, 508–527.

(37) Meredig, B.; Wolverton, C. A Hybrid Computational-Experimental Approach for Automated Crystal Structure Solution. *Nat. Mater.* **2013**, *12*, 123–127.

(38) Ward, L.; Michel, K.; Wolverton, C. Automated Crystal Structure Solution from Powder Diffraction Data: Validation of the First-Principles-Assisted Structure Solution Method. *Phys. Rev. Mater.* **2017**, *1*, 063802.

(39) Amsler, M.; Hegde, V. I.; Jacobsen, S. D.; Wolverton, C. Exploring the High-Pressure Materials Genome. *Phys. Rev. X* **2018**, *8*, 041021.

(40) Lonie, D. C.; Zurek, E. XtalOpt: An Open-source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2011**, *182*, 372–387.

(41) Lonie, D. C.; Zurek, E. XtalOpt Version r7: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2011**, *182*, 2305–2306.

(42) Falls, Z.; Lonie, D. C.; Avery, P.; Shamp, A.; Zurek, E. XtalOpt Version r9: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2016**, *199*, 178–179.

(43) Avery, P.; Falls, Z.; Zurek, E. XtalOpt Version r10: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2017**, *217*, 210–211.

(44) Avery, P.; Falls, Z.; Zurek, E. XtalOpt Version r11: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2018**, *222*, 418–419.

(45) Avery, P.; Toher, C.; Curtarolo, S.; Zurek, E. XtalOpt Version r12: An Open-Source Evolutionary Algorithm for Crystal Structure Prediction. *Comput. Phys. Commun.* **2019**, *237*, 274–275.

(46) XTALOPT Evolutionary Crystal Structure Prediction. `http://xtalopt.github.io/` (Accessed Nov 9, 2020).

(47) Lonie, D. C.; Zurek, E. Identifying Duplicate Crystal Structures: XtalComp, an Open-source Solution. *Comput. Phys. Commun.* **2012**, *183*, 690–697.

(48) Avery, P.; Zurek, E. RandSpg: An Open-Source Program for Generating Atomistic Crystal Structures with Specific Spacegroups. *Comput. Phys. Commun.* **2017**, *213*, 208–216.

(49) Gale, J. D.; Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **2003**, *29*, 291.

(50) Kresse, G.; Hafner, J. *Ab initio* Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47*, 558–561.

(51) Kresse, G.; Hafner, J. *Ab initio* Molecular-Dynamics Simulation of the Liquid-Metal–Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49*, 14251.

(52) Woodley, S.; Catlow, C. Structure Prediction of Titania Phases: Implementation of Darwinian Versus Lamarckian Concepts in an Evolutionary Algorithm. *Comput. Mater. Sci.* **2009**, *45*, 84–95.

(53) Benedek, N. A.; Chua, A. L. S.; Elsässer, C.; Sutton, A. P.; W., F. M. Interatomic Potentials for Strontium Titanate: An Assessment of their Transferability and Comparison with Density Functional Theory. *Phys. Rev. B* **2008**, *78*, 064110.

(54) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(55) Ehrlich, S.; Moellmann, J.; Reckien, W.; Bredow, T.; Grimme, S. System-Dependent Dispersion Coefficients for the DFT-D3 Treatment of Adsorption Processes on Ionic Surfaces. *ChemPhysChem* **2011**, *12*, 3414–3420.

(56) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(57) Román-Pérez, G.; Soler, J. M. Efficient Implementation of a van der Waals Density Functional: Application to Double-Wall Carbon Nanotubes. *Phys. Rev. Lett.* **2009**, *103*, 096102.

(58) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the van der Waals Density Functional. *J. Phys. Condens. Mat.* **2009**, *22*, 022201.

(59) Klimeš, J.; Bowler, D. R.; Michaelides, A. Van der Waals Density Functionals Applied to Solids. *Phys. Rev. B* **2011**, *83*, 195131.

(60) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(61) Stillinger, F. H. Exponential Multiplicity of Inherent Structures. *Phys. Rev. E* **1999**, *59*, 48–51.

(62) Wille, L. T.; Vennik, J. Computational Complexity of the Ground-state Determination of Atomic Clusters. *J. Phys. A: Math. Gen.* **1985**, *18*, L419–L422.

(63) Greenwood, G. W. Revisiting the Complexity of Finding Globally Minimum Energy Configurations in Atomic Clusters. *Z. Phys. Chem.* **1999**, *211*, 105–114.

(64) Wolpert, D. H.; Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82.

(65) Doye, J. P. K.; Massen, C. P. Characterizing the Network Topology of the Energy Landscapes of Atomic Clusters. *J. Chem. Phys.* **2005**, *122*, 084105.

(66) Bell, R. P. The Theory of Reactions Involving Proton Transfers. *Proc. R. Soc. A* **1936**, *154*, 414–429.

(67) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Trans. Faraday Soc.* **1935**, *31*, 875–894.

(68) Kvashnin, A. G.; Allahyari, Z.; Oganov, A. R. Computational Discovery of Hard and Superhard Materials. *J. Appl. Phys.* **2019**, *126*, 040901.

(69) Flores-Livas, J. A. Crystal Structure Prediction of Magnetic Materials. *J. Phys.: Condens. Mat.* **2020**, *32*, 294002.

(70) Higgins, E. J.; Hasnip, P. J.; Probert, M. I. J. Simultaneous Prediction of the Magnetic and Crystal Structure of Materials Using a Genetic Algorithm. *Crystals* **2019**, *9*, 439.

(71) Parija, A.; Waetzig, G. R.; Andrews, J. L.; Banerjee, S. Traversing Energy Landscapes Away from Equilibrium: Strategies for Accessing and Utilizing Metastable Phase Space. *J. Phys. Chem. C* **2018**, *122*, 25709–25728.

(72) Johnston, R. L. Evolving Better Nanoparticles: Genetic Algorithms for Optimising Cluster Geometries. *Dalton Trans.* **2003**, *22*, 4193–4207.

(73) Sierka, M. Synergy Between Theory and Experiment in Structure Resolution of Low-dimensional Oxides. *Prog. Surf. Sci.* **2010**, *85*, 398–434.

(74) Revard, B. C.; Tipton, W. W.; Yesypenko, A.; Hennig, R. G. Grand-Canonical Evolutionary Algorithm for the Prediction of Two-Dimensional Materials. *Phys. Rev. B* **2016**, *93*, 054117.

(75) Singh, A. K.; Revard, B. C.; Ramanathan, R.; Ashton, M.; Tavazza, F.; Hennig, R. G. Genetic Algorithm Prediction of Two-dimensional Group-IV Dioxides for Dielectrics. *Phys. Rev. B* **2017**, *95*, 155426.

(76) Wang, Y.; Miao, M.; Lv, J.; Zhu, L.; Yin, K.; Liu, H.; Ma, Y. An Effective Structure Prediction Method for Layered Materials Based on 2D Particle Swarm Optimization Algorithm. *J. Chem. Phys.* **2012**, *137*, 224108.

(77) Lu, S.; Wang, Y.; Liu, H.; Miao, M.; Ma, Y. Self-Assembled Ultrathin Nanotubes on Diamond (100) Surface. *Nat. Commun.* **2014**, *5*, 3666.

(78) Jacobsen, T. L.; Jørgensen, M. S.; Hammer, B. On-the-Fly Machine Learning of Atomic Potential in Density Functional Theory Structure Optimization. *Phys. Rev. Lett.* **2018**, *120*, 026102.

(79) Vilhelmsen, L. B.; Hammer, B. A Genetic Algorithm for First Principles Global Structure Optimization of Supported Nano Structures. *J. Chem. Phys.* **2014**, *141*, 044711.

(80) Gao, B.; Gao, P.; Lu, S.; Lv, J.; Wang, Y.; Ma, Y. Interface Structure Prediction via CALYPSO Method. *Sci. Bull.* **2019**, *64*, 301–309.

(81) Woodley, S. M. Engineering Microporous Architectures: Combining Evolutionary Algorithms with Predefined Exclusion Zones. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1070–1077.

(82) Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooij, W. T. M.; Price, S. L.; Schweizer, B. et al. A

Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2000**, *56*, 697–714.

(83) Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M. et al. Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2002**, *58*, 647–661.

(84) Day, G. M.; Motherwell, W. D. S.; Ammon, H. L.; Boerrigter, S. X. M.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. B.; Schweizer, B.; van Eijck, B. P. et al. A Third Test of Crystal Structure Prediction. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2005**, *61*, 511–527.

(85) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J. et al. Significant Progress in Predicting the Crystal Structures of Small Organic Molecules - A Report on the Fourth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2009**, *65*, 107–125.

(86) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerringter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R. et al. Towards Crystal Structure Prediction of Complex Organic Compounds - A Report on the Fifth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2011**, *67*, 535–551.

(87) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R. et al. Report on the Sixth Blind Test of Organic Crystal Structure Prediction Methods. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 439–459.

(88) Cruz-Cabeza, A. J. Crystal Structure Prediction: Are We There Yet? *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 437–438.

(89) Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, A.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246–2264.

(90) Nyman, J.; Reutzel-Edens, S. M. Crystal Structure Prediction is Changing from Basic Science to Applied Technology. *Faraday Discuss.* **2018**, *211*, 459–476.

(91) Fischer, C. C.; Tibbetts, K. J.; Morgan, D.; Ceder, G. Predicting Crystal Structure by Merging Data Mining with Quantum Mechanics. *Nat. Mater.* **2006**, *5*, 641–646.

(92) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12*, 191–201.

(93) Ryan, K.; Lengyel, J.; Shatruk, M. Crystal Structure Prediction via Deep Learning. *J. Am. Chem. Soc.* **2018**, *140*, 10158–10168.

(94) Graser, J.; Kauwe, S. K.; Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chem. Mater.* **2018**, *30*, 3601–3612.

(95) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547–555.

(96) Ishikawa, T.; Miyake, T.; Shimizu, K. Materials Informatics Based on Evolutionary Algorithms: Application to Search for Superconducting Hydrogen Compounds. *Phys. Rev. B* **2019**, *100*, 174506.

(97) Maldonis, J. J.; Xu, Z.; Song, Z.; Yu, M.; Mayeshiba, T.; Morgan, D.; Voyles, P. M. StructOpt: A Modular Materials Structure Optimization Suite Incorporating Experimental and Data and Simulated Energies. *Comput. Mater. Sci.* **2019**, *160*, 1–8.

(98) Gao, P.; Tong, Q.; Lv, J.; Wang, Y.; Ma, Y. X-ray Diffraction Data-assisted Structure Searches. *Comput. Phys. Commun.* **2017**, *213*, 40–45.

(99) Tipton, W. W.; Hennig, R. G. In *Modern Methods of Crystal Structure Prediction*; Oganov, A. R., Ed.; John Wiley & Sons.: Weinheim, Germany, 2010; Chapter 3.

(100) Domingos, R.; Shaik, K. M.; Militzer, B. Prediction of Novel High Pressure H_2O-NaCl and Carbon Oxide Compounds with Symmetry-Driven Structure Search Algorithm. *Phys. Rev. B* **2018**, *98*, 174107.

(101) Mayo, M.; Griffith, K. J.; Pickard, C. J.; Morris, A. J. *Ab initio* Study of Phosphorus Anodes for Lithium- and Sodium-ion Batteries. *Chem. Mater.* **2016**, *28*, 2011–2021.

(102) Zilka, M.; Dudenko, D. V.; Hughes, C. E.; Williams, P. A.; Sturniolo, S.; Franks, W. T.; Pickard, C. J.; Yates, J. R.; Harris, K. D. M.; Brown, S. P. *Ab initio* Random Structure Searching of Organic Molecular Solids: Assessment and Validation Against Experimental Data. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25949–25960.

(103) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680.

(104) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(105) Doll, K.; Schön, J. C.; Jansen, M. Global Exploration of the Energy Landscape of Solids on the *ab initio* Level. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6128–6133.

(106) Kulkarni, A.; Schön, J. C.; Doll, K.; Jansen, M. Structure Prediction of Binary Pernitride MN_2 Compounds (M=Ca, Sr. Ba, La, and Ti). *Chem. Asian J.* **2013**, *8*, 743–754.

(107) Doll, K.; Schön, J. C.; Jansen, M. Structure Prediction Based on *ab initio* Simulated Annealing for Boron Nitride. *Phys. Rev. B* **2008**, *78*, 144110.

(108) Doll, K.; Jansen, M. *Ab initio* Energy Landscape of GeF_2: A System Featuring Lone Pair Structure Candidates. *Angew. Chem. Int. Ed.* **2011**, *50*, 4627–4632.

(109) Goedecker, S. Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.

(110) Amsler, M. In *Handbook of Materials Modeling: Volume 2 Applications: Current and Emerging Materials*; Andreoni, W., Yip, S., Eds.; Springer, Cham: Switzerland, 2018; Vol. 1; pp 1–20.

(111) Sicher, M.; Mohr, S.; Goedecker, S. Efficient Moves for Global Geometry Optimization Methods and Their Application to Binary Systems. *J. Chem. Phys.* **2011**, *134*, 044106.

(112) Roy, S.; Goedecker, S.; Hellmann, V. Bell-Evans-Polanyi Principle for Molecular Dynamics Trajectories and its Implications for Global Optimization. *Phys. Rev. E.* **2008**, *77*, 056707.

(113) Amsler, M. Thermodynamics and Superconductivity of $S_xSe_{1-x}H_3$. *Phys. Rev. B* **2019**, *99*, 060102(R).

(114) Amsler, M.; Flores-Livas, J. A.; Lehtovaara, L.; Balima, F.; Ghasemi, S. A.; Machon, D.; Pailhès, S.; Willand, A.; Caliste, D.; Botti, S. et al. Crystal Structure of Cold Compressed Graphite. *Phys. Rev. Lett.* **2012**, *108*, 065501.

(115) 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* **1997**, *101*, 5111–5116.

(116) Burnham, C. J.; English, N. J. Crystal Structure Prediction via Basin-hopping Global Optimization Employing Tiny Periodic Simulation Cells, with Application to Water-ice. *J. Chem. Theory Comput.* **2019**, *15*, 3889–3900.

(117) Röder, K.; Wales, D. J. Mutational Basin-hopping: Combined Structure and Sequence Optimization for Biomolecules. *J. Phys. Chem. Lett.* **2018**, *9*, 6169–6173.

(118) Calvo, F.; Schebarchov, D.; Wales, D. J. Grand and Semigrand Canonical Basin-Hopping. *J. Chem. Theory Comput.* **2016**, *12*, 902–909.

(119) Sharp, P. M.; Dyer, M. S.; Darling, G. R.; Claridge, J. B.; Rosseinsky, M. J. Chemically Directed Structure Evolution for Crystal Structure Prediction. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18205–18218.

(120) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(121) Martonák, R.; Laio, A.; Parrinello, M. Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Phys. Rev. Lett.* **2003**, *90*, 075503.

(122) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.

(123) Sun, J.; Klug, D. D.; Martoňák, R. Structural Transformations in Carbon Under Extreme Pressure: Beyond Diamond. *J. Chem. Phys.* **2009**, *130*, 194512.

(124) Sun, J.; Klug, D. D.; Martoňák, R.; Montoya, J. A.; Lee, M.-S.; Scandolo, S.; Tosatti, E. High-pressure Polymeric Phases of Carbon Dioxide. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6077–6081.

(125) Selli, D.; Baburin, I. A.; Martoňák, R.; Leoni, S. Novel Metastable Metallic and Semicondcting Germaniums. *Sci. Rep.* **2013**, *3*, 1466.

(126) Plašienka, D.; Martoňák, R. Transformation Pathways in High-Pressure Solid Nitrogen: From Molecular N_2 to Polymeric cg-N. *J. Chem. Phys.* **2015**, *142*, 094505.

(127) Piccini, G. M.; Mendels, D.; Parrinello, M. Metadynamics with Discriminants: A Tool for Understanding Chemistry. *J. Chem. Theory Comput.* **2018**, *14*, 5040–5044.

(128) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. *Proc. IEEE Int. Conf. Neural Networks, 4th* **1995**, *4*, 1942–1948.

(129) Call, S. T.; Zubarev, D. Y.; Boldyrev, A. I. Global Minimum Structure Searches via Particle Swarm Optimization. *J. Comput. Chem.* **2007**, *28*, 1177–1186.

(130) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal Structure Prediction via Particle Swarm Optimization. *Phys. Rev. B* **2010**, *82*, 094116.

(131) Wang, Y.; Lv, J.; Zhu, L.; Lu, S.; Yin, K.; Li, Q.; Wang, H.; Zhang, L.; Ma, Y. Materials Discovery via CALYPSO Methodology. *J. Phys.: Condens. Mat.* **2015**, *27*, 203203.

(132) Wang, H.; Wang, Y.; Lv, J.; Li, Q.; Zhang, L.; Ma, Y. CALYPSO Structure Prediction Method and its Wide Application. *Comput. Mater. Sci.* **2016**, *112*, 406–415.

(133) Wang, Y.; Lv, J.; Li, Q.; Wang, H.; Ma, Y. In *Handbook of Materials Modeling: Volume 2 Applications: Current and Emerging Materials*; Andreoni, W., Yip, S., Eds.; Springer, Cham: Switzerland, 2018; Vol. 1; pp 1–28.

(134) Deaven, D. M.; Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett.* **1995**, *75*, 288–291.

(135) Glass, C. W.; Oganov, A. R.; Hansen, N. USPEX – Evolutionary Crystal Structure Prediction. *Comput. Phys. Commun.* **2006**, *175*, 713–720.

(136) Oganov, A. R.; Glass, C. W. Crystal Structure Prediction Using *ab initio* Evolutionary Techniques: Principles and Applications. *J. Chem. Phys.* **2006**, *124*, 244704.

(137) Lyakhov, A. O.; Oganov, A. R.; Valle, M. How to Predict Very Large and Complex Crystal Structures. *Comp. Phys. Commun.* **2010**, *181*, 1623–1632.

(138) Oganov, A. R.; Glass, C. W. Evolutionary Crystal Structure Prediction as a Tool in Materials Design. *J. Phys.: Condens. Mat.* **2008**, *20*, 064210.

(139) Tipton, W. W.; Hennig, R. G. A Grand Canonical Genetic Algorithm for the Prediction of Multi-component Phase Diagrams and Testing of Empirical Potentials. *J. Phys.: Condens. Mat.* **2013**, *25*, 495401.

(140) Tipton, W. W.; Bealing, C. R.; Mathew, K.; Hennig, R. G. Structures, Phase Stabilities, and Electrical Potentials of Li-Si Battery Anode Materials. *Phys. Rev. B* **2013**, *87*, 184114.

(141) Bahmann, S.; Kortus, J. EVO – Evolutionary Algorithm for Crystal Structore Prediction. *Comput. Phys. Commun.* **2013**, *184*, 1618–1625.

(142) Trimarchi, G.; Zunger, A. Global Space-group Optimization Problem: Finding the Stablest Crystal Structure Without Constraints. *Phys. Rev. B.* **2007**, *75*, 104113.

(143) d'Avezac, M.; Zunger, A. Identifying the Minimum – Energy Atomic Configuration on a Lattice: Lamarckian Twist on Darwinian Evolution. *Phys. Rev. B.* **2008**, *78*, 064102.

(144) Trimarchi, G.; Freeman, A. J.; Zunger, A. Predicting Stable Stoichiometries of Compounds via Evolutionary Global Space-group Optimization. *Phys. Rev. B* **2009**, *80*, 092101.

(145) Trimarchi, G.; Zunger, A. Finding the Lowest-Energy Crystal Structure Starting from Randomly Selected Lattice Vectors and Atomic Positions: First-principles Evolutionary Study of the Au-Pd, Cd-Pt, Al-Sc, Cu-Pd, Pd-Ti, and Ir-N Binary Systems. *J. Phys.: Condens. Matter* **2008**, *20*, 295212.

(146) Abraham, N. L.; Probert, M. I. J. A Periodic Genetic Algorithm with Real-space Representation for Crystal Structure and Polymorph Prediction. *Phys. Rev. B.* **2006**, *73*, 224104.

(147) Fadda, A.; Fadda, G. An Evolutionary Algorithm for the Prediction of Crystal Structures. *Phys. Rev. B* **2010**, *82*, 104105.

(148) Wu, S. Q.; Ji, M.; Wang, C. Z.; Nguyen, M. C.; Zhao, X.; Umemoto, K.; Wentzcovitch, R. M.; Ho, K. M. An Adaptive Genetic Algorithm for Crystal Structure Prediction. *J. Phys.: Condens. Mat.* **2014**, *26*, 035402.

(149) Van den Bossche, M.; Grönbeck, H.; Hammer, B. Tight-Binding Approximation-enhanced Global Optimization. *J. Chem. Theory Comput.* **2018**, *14*, 2797–2807.

(150) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 1–17.

(151) Avogadro Chemistry, Avogadro – Free Cross Platform Molecular Editor. `https://avogadro.cc/` (Accessed Nov 9, 2020).

(152) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*.

(153) Free Software Foundation, The GNU General Public License v3.0. `http://www.gnu.org/licenses/gpl-3.0.html` (Accessed Nov 9, 2020).

(154) Open Source Initiative, The 3-Clause BSD License. `https://opensource.org/licenses/BSD-3-Clause` (Accessed Nov 9, 2020).

(155) Open Chemistry, Avogadro2. `https://www.openchemistry.org/projects/avogadro2/` (Accessed Nov 9, 2020).

(156) Togo, A. Spglib. `https://atztogo.github.io/spglib/` (Accessed Nov 9, 2020).

(157) Lonie, D. C.; Hooper, J.; Altintas, B.; Zurek, E. Metallization of Magnesium Polyhydrides Under Pressure. *Phys. Rev. B* **2013**, *87*, 054107.

(158) Song, Y. New Perspectives on Potential Hydrogen Storage Materials using High Pressure. *Phys. Chem. Chem. Phys.* **2013**, *15*, 14524–14547.

(159) Orimo, S.; Nakamori, Y.; Eliseo, J. R.; Züttel, A.; Jensen, C. M. Complex Hydrides for Hydrogen Storage. *Chem. Rev.* **2007**, *107*, 4111–4132.

(160) Zurek, E.; Wen, X.-D.; Hoffmann, R. (Barely) Solid $Li(NH_3)_4$: The Electronics of an Expanded Metal. *J. Am. Chem. Soc.* **2011**, *133*, 3535–3547.

(161) Zhang, R.; Cai, W.; Bi, T.; Zarifi, N.; Terpstra, T.; Zhang, C.; Valy Verdeny, Z.; Zurek, E.; Deemyad, S. Effects of Non-Hydrostatic Stress on Structural and Optoelectronic Properties of Methylammonium Lead Bromide Perovskite. *J. Phys. Chem. Lett.* **2017**, *8*, 3457–3465.

(162) Manser, J. S.; Christians, J. A.; Kamat, P. V. Intriguing Optoelectronic Properties of Metal Halide Perovskites. *Chem. Rev.* **2016**, *116*, 12956–13008.

(163) Kauzlarich, S. M., Ed. *Chemistry, Structure and Bonding of Zintl Phases and Ions*; VCH-Publishers: New York, 1996.

(164) Simon, A. Metal Clusters Inside Out. *Philos. Trans. R. Soc., A* **2010**, *368*, 1285–1299.

(165) Simon, A.; Borrmann, H.; Craubner, H. Crystal Structure of Ordered White Phosphorus (β-P). *Phosphorus Sulfur Relat. Elem.* **1987**, *30*, 507–510.

(166) Oganov, A. R.; Chen, J.; Gatti, C.; Ma, Y.; Ma, Y.; Glass, C. W.; Liu, Z.; Yu, T.; Kurakevych, O. O.; Solozhenko, V. L. Ionic High-pressure Form of Elemental Boron. *Nature* **2009**, *457*, 863–867.

(167) Lyakhov, A. O.; Oganov, A. R.; Stokes, H. T.; Zhu, Q. New Developments in Evolutionary Structure Prediction Algorithm USPEX. *Comput. Phys. Commun.* **2013**, *184*, 1172–1182.

(168) Urusov, V. S.; Nadezhina, T. N. Frequency Distribution and Selection of Space Groups in Inorganic Crystal Chemistry. *J. Struct. Chem.* **2009**, *50*, 22–37.

(169) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. CALYPSO: A Method for Crystal Structure Prediction. *Comput. Phys. Commun.* **2012**, *183*, 2063 – 2070.

(170) XtalComp: Compare Crystal Structures. `http://xtalopt.openmolecules.net/xtalcomp/xtalcomp.html` (Accessed Nov 09, 2020).

(171) Lyakhov, A. O.; Oganov, A. R. Evolutionary Search for Superhard Materials: Methodology and Applications to Forms of Carbon and TiO_2. *Phys. Rev. B* **2011**, *84*, 092103.

(172) Zhang, X.; Wang, Y.; Lv, J.; Zhu, C.; Li, Q.; Zhang, M.; Li, Q.; Ma, Y. First-Principles Structural Design of Superhard Materials. *J. Chem. Phys.* **2013**, *138*, 114101.

(173) Gao, F.; He, J.; Wu, E.; Liu, S.; Yu, D.; Li, D.; Zhang, S.; Tian, Y. Hardness of Covalent Crystals. *Phys. Rev. Lett.* **2003**, *91*, 015502.

(174) Šimůnek, A.; Vackář, J. Hardness of Covalent and Ionic Crystals: First-Principle Calculations. *Phys. Rev. Lett.* **2006**, *96*, 085501.

(175) Sheng, X.-L.; Yan, Q.-B.; Ye, F.; Zheng, Q.-R.; Su, G. T-Carbon: A Novel Carbon Allotrope. *Phys. Rev. Lett.* **2011**, *106*, 155703.

(176) Chen, X.-Q.; Niu, H.; Franchini, C.; Li, D.; Li, Y. Hardness of T-carbon: Density Functional Theory Calculations. *Phys. Rev. B* **2011**, *84*, 121405.

(177) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Ande, C. K.; van der Zwaag, S.; Plata, J. J. et al. Charting the Complete Elastic Properties of Inorganic Crystalline Compounds. *Scientific Data* **2015**, *2*, 150009.

(178) Toher, C.; Oses, C.; Plata, J. J.; Hicks, D.; Rose, F.; Levy, O.; de Jong, M.; Asta, M.; Fornari, M.; Nardelli, M. B. et al. Combining the AFLOW GIBBS and Elastic Libraries to Efficiently and Robustly Screen Thermomechanical Properties of Solids. *Phys. Rev. Mater.* **2017**, *1*, 015401.

(179) Gossett, E.; Toher, C.; Oses, C.; Isayev, O.; Legrain, F.; Rose, F.; Zurek, E.; Carrete, J.; Mingo, N.; Tropsha, A. et al. AFLOW-ML: A RESTful API for Machine-Learning Predictions of Materials Properties. *Comput. Mater. Sci.* **2018**, *152*, 134–145.

(180) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal Fragment Descriptors for Predicting Properties of Inorganic Crystals. *Nat. Commun.* **2017**, *8*, 15679.

(181) Curtarolo, S.; Setyawan, W.; Hart, G. L.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.;

Wang, S.; Xue, J.; Yang, K.; Levy, O. et al. AFLOW: An Automatic Framework for High-Throughput Materials Discovery. *Comput. Mater. Sci.* **2012**, *58*, 218–226.

(182) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L.; Sanvito, S.; Buongiorno-Nardelli, M. et al. AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput *ab initio* Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.

(183) Teter, D. M. Computational Alchemy: The Search for New Superhard Materials. *MRS Bull.* **1998**, *23*, 22–27.

(184) Chen, X.-Q.; Niu, H.; Li, D.; Li, Y. Modeling Hardness of Polycrystalline Materials and Bulk Metallic Glasses. *Intermetallics* **2011**, *19*, 1275–1281.

(185) Avery, P.; Wang, X.; Oses, C.; Gossett, E.; Proserpio, D.; Toher, C.; Curtarolo, S.; Zurek, E. Predicting Superhard Materials via a Machine Learning Informed Evolutionary Structure Search. *npj Comput. Mater.* **2019**, *5*, 1–11.

(186) Hartman, M. R.; Rush, J. J.; Udovic, T. J.; Bowman Jr., R. C.; Hwang, S.-J. Structure and Vibrational Dynamics of Isotopically Labeled Lithium Borohydride using Neutron Diffraction and Spectroscopy. *J. Solid State Chem.* **2007**, *180*, 1298–1305.

(187) Goebel, T.; Prots, Y.; Haarmann, F. Refinement of the Crystal Structure of Tetrasodium Tetrasilicide, Na_4Si_4. *Z. Kristallogr.* **2008**, *223*, 187–188.

(188) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics Studies of Crystal Nucleation. *IUCrJ* **2015**, *256-266*, 2.

(189) Baidakov, V. G.; Protsenko, K. R. Spontaneous Crystallization of a Supercooled Lennard-Jones Liquid: Molecular Dynamics Simulation. *J. Phys. Chem. B* **2019**, *8103-8112*, 123.

(190) Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R. Accelerating Crystal Structure Prediction by Machine-Learning Interatomic Potentials with Active Learning. *Phys. Rev. B* **2019**, *064114*, 99.

(191) Deringer, V. L.; Proserpio, D. M.; Csányi, G.; J., P. C. Data-Driven Learning and Prediction of Inorganic Crystal Structures. *Faraday Discuss.* **2018**, *211*, 45.

(192) Lin, Y.; Strobel, T. A.; Cohen, R. E. Structural Diversity in Lithium Carbides. *Phys. Rev. B* **2015**, *92*, 214106.

(193) Bethkenhagen, M.; Cebulla, D.; Redmer, R.; Hamel, S. Superionic Phases of the 1:1 Water-Ammonia Mixture. *J. Phys. Chem. A* **2015**, *119*, 10582–10588.

(194) Kohulák, O.; Martoňák, R.; Tosatti, E. High-pressure Structure, Decomposition, and Superconductivity of MoS_2. *Phys. Rev. B* **2015**, *91*, 144113.

(195) Hermann, A.; Guthrie, M.; Nelmes, R. J.; Loveday, J. S. Pressure-induced Localisation of the Hydrogen-bond Network in KOH-VI. *J. Chem. Phys.* **2015**, *143*, 244706.

(196) Martoňák, R.; Ceresoli, D.; Kagayama, T.; Matsuda, Y.; Yamada, Y.; Tosatti, E. High-Pressure Phase Diagram, Structural Transitions, and Persistent Nonmetallicity of $BaBiO_3$: Theory and Experiment. *Phys. Rev. Mater.* **2017**, *1*, 023601.

(197) Eifert, B.; Becker, M.; Reindl, C. T.; Giar, M.; Zheng, L.; Polity, A.; He, Y.; Heiliger, C.; Klar, P. J. Raman Studies of the Intermediate Tin-Oxide Phase. *Phys. Rev. Mater.* **2017**, *1*, 014602.

(198) Kohulák, O.; Martoňák, R. New High-pressure Phases of $MoSe_2$ and $MoTe_2$. *Phys. Rev. B* **2017**, *95*, 054105.

(199) Hermann, A.; Mookherjee, M. High-Pressure Phase of Brucite Stable at Earth's Mantle Transition Zone and Lower Mantle Conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13971–13976.

(200) Plašienka, D.; Martoňák, R.; Tosatti, E. Creating New Layered Structures at High Pressures: SiS_2. *Sci. Rep.* **2016**, *6*, 37694.

(201) Kurzydłowski, D. The Jahn-Teller Distortion at High Pressure: The Case of Copper Difluoride. *Crystals* **2018**, *8*, 140.

(202) Geng, N.; Bi, T.; Zarifi, N.; Yan, Y.; Zurek, E. A First-Principles Exploration of Na_xS_y Binary Phases at 1 atm and Under Pressure. *Crystals* **2019**, *9*, 441.

(203) Zarifi, N.; Liu, H.; Tse, J. S.; Zurek, E. Crystal Structures and Electronic Properties of Xe-Cl Compounds at High Pressure. *J. Phys. Chem. C* **2018**, *122*, 2941–2950.

(204) Zurek, E.; Hoffmann, R.; Ashcroft, N. W.; Oganov, A. R.; Lyakhov, A. O. A Little Bit of Lithium Does a Lot for Hydrogen. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 17640–17643.

(205) Baettig, P.; Zurek, E. Pressure-Stabilized Sodium Polyhydrides, NaH_n ($n > 1$). *Phys. Rev. Lett.* **2011**, *106*, 237002.

(206) Hooper, J.; Zurek, E. Rubidium Polyhydrides Under Pressure: Emergence of the Linear H_3^- Species. *Chem. - Eur. J.* **2012**, *18*, 5013–5021.

(207) Hooper, J.; Zurek, E. Lithium Subhydrides Under Pressure and their Superatom–Like Building Blocks. *ChemPlusChem* **2012**, *77*, 969–972.

(208) Hooper, J.; Zurek, E. High Pressure Potassium Polyhydrides: A Chemical Perspective. *J. Phys. Chem. C* **2012**, *116*, 13322–13328.

(209) Shamp, A.; Hooper, J.; Zurek, E. Compressed Cesium Polyhydrides: Cs^+ Sublattices and H_3^- Three-Connected Nets. *Inorg. Chem.* **2012**, *51*, 9333–9342.

(210) Hooper, J.; Altintas, B.; Shamp, A.; Zurek, E. Polyhydrides of the Alkaline Earth Metals: A Look at the Extremes Under Pressure. *J. Phys. Chem. C* **2013**, *117*, 2982–2992.

(211) Hooper, J.; Terpstra, T.; Shamp, A.; Zurek, E. Composition and Constitution of Compressed Strontium Polyhydrides. *J. Phys. Chem. C* **2014**, *118*, 6433–6447.

(212) Shamp, A.; Zurek, E. Superconducting High-Pressure Phases Composed of Hydrogen and Iodine. *J. Phys. Chem. Lett.* **2015**, *6*, 4067–4072.

(213) Shamp, A.; Terpstra, T.; Bi, T.; Falls, Z.; Avery, P.; Zurek, E. Decomposition Products of Phosphine Under Pressure: PH_2 Stable and Superconducting? *J. Am. Chem. Soc.* **2016**, *138*, 1884–1892.

(214) Zurek, E. Hydrides of the Alkali Metals and Alkaline Earth Metals Under Pressure. *Comments Inorg. Chem.* **2017**, *37*, 78–98.

(215) Shamp, A.; Zurek, E. Superconductivity in Hydrides Doped with Main Group Elements Under Pressure. *Nov. Supercond. Mater.* **2017**, *3*, 14–22.

(216) Bi, T.; Miller, D. P.; Shamp, A.; Zurek, E. Superconducting Phases of Phosphorus Hydride Under Pressure: Stabilization via Mobile Molecular Hydrogen. *Angew. Chem. Int. Ed.* **2017**, *56*, 10192–10195.

(217) Ye, X.; Zarifi, N.; Zurek, E.; Hoffmann, R.; Ashcroft, N. W. High Hydrides of Scandium under Pressure: Potential Superconductors. *J. Phys. Chem. C* **2018**, *122*, 6298–6309.

(218) Zarifi, N.; Bi, T.; Liu, H.; Zurek, E. Crystal Structures and Properties of Iron Hydrides at High Pressure. *J. Phys. Chem. C* **2018**, *122*, 24262–24269.

(219) Cui, W.; Bi, T.; Shi, J.; Li, Y.; Liu, H.; Zurek, E.; Hemley, R. J. Route to High-*Tc* Superconductivity via CH_4 Intercalated H_3S Hydride Perovskites. *Phys. Rev. B* **2020**, *101*, 134504.

(220) Yan, Y.; Tiange, B.; Geng, N.; Wang, X.; Zurek, E. A Metastable $CaSH_3$ Phase Composed of HS Honeycomb Sheets that is Superconducting Under Pressure. *J. Phys. Chem. Lett.* **2020**, *11*, 9629–9636.

(221) Ashcroft, N. W. Metallic Hydrogen: A High-Temperature Superconductor? *Phys. Rev. Lett.* **1968**, *21*, 1748–1749.

(222) Ashcroft, N. W. Hydrogen Dominant Metallic Alloys: High Temperature Superconductors? *Phys. Rev. Lett.* **2004**, *92*, 187002.

(223) Drozdov, A. P.; Eremets, M. I.; Troyan, I. A.; Ksenofontov, V.; Shylin, S. I. Conventional Superconductivity at 203 Kelvin at High Pressures in the Sulfur Hydride System. *Nature* **2015**, *525*, 73–76.

(224) Duan, D.; Liu, Y.; Tian, F.; Li, D.; Huang, X.; Zhao, Z.; Yu, H.; Liu, B.; Tian, W.; Cui, T. Pressure-Induced Metallization of Dense $(H_2S)_2H_2$ with High-T_c Superconductivity. *Sci. Rep.* **2014**, *4*, 6968.

(225) Sun, Y.; Tian, Y.; Jiang, B.; Li, X.; Li, H.; Iitaka, T.; Zhong, X.; Xie, Y. Computational Discovery of a Dynamically Stable Cubic SH_3-like High-Temperature Superconductor at 100 GPa via CH_4 Intercalation. *Phys. Rev. B* **2020**, *101*, 174102.

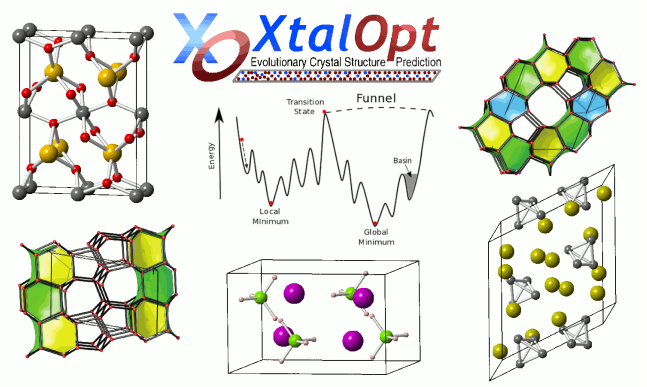
(226) Li, D.; Liu, Y.; Tian, F.-B.; Wei, S.-L.; Liu, Z.; Duan, D.-F.; Liu, B.-B.; Cui, T. Pressure-Induced Superconducting Ternary Hydride H_3SXe: A Theoretical Investigation. *Front. Phys.* **2018**, *13*, 137107.

(227) Snider, E.; Dasenbrock-Gammon, N.; McBride, R.; Debessai, M.; Vindana, H.; Vencatasamy, K.; Lawler, K. V.; Salamat, A.; Dias, R. P. Room-Temperature Superconductivity in a Carbonaceous Sulfur Hydride. *Nature* **2020**, *586*, 373–377.

(228) Errea, I.; Belli, F.; Monacelli, L.; Sanna, A.; Koretsune, T.; Tadano, T.; Bianco, R.; Calandra, M.; Arita, R.; Mauri, F. et al. Quantum Crystal Structure in the 250 Kelvin Superconducting Lanthanum Hydride. *Nature* **2020**, *578*, 66–69.

(229) CCR Facility Description. `https://ubir.buffalo.edu/xmlui/handle/10477/79221` (Accessed Nov 9, 2020).

TOC Graphic


XtalOpt
Evolutionary Crystal Structure Prediction
Funnel
Transition State
Energy
Basin
Local Minimum
Global Minimum

Biographies

Zackary Falls graduated with a BS in Chemistry from Canisius College, Buffalo, NY in 2012. In 2017, he received a PhD in Computational Chemistry under the tutelage of Prof. Eva Zurek at the University at Buffalo. After his PhD, he was the recipient of a National Library of Medicine T15 Postdoctoral Fellowship with which he began research in the Department of Biomedical Informatics at the Jacobs School of Medicine and Biomedical Sciences at the University at Buffalo. There he worked with Prof. Ram Samudrala in the field of drug discovery. He was recently promoted to Research Assistant Professor and continues to research bleeding-edge computational drug discovery techniques and applications.

Patrick Avery completed a BS in Chemistry at Bob Jones University in 2014, and a PhD in Computational Chemistry at the State University of New York at Buffalo in 2019 under the guidance of Dr. Eva Zurek. His research primarily involved developing new methodology for crystal

structure prediction. Since 2019, Patrick has been employed as a research and development engineer with the scientific visualization team at *Kitware*, Inc.

Xiaoyu Wang received his BS degree in chemistry from the University of Science and Technology of China. In 2018 he completed his MS in Chemical Engineering at the University at Buffalo (UB) with Professor Michel Dupuis. Currently he is a third year PhD student in Chemistry at UB, and works with Professor Eva Zurek on the theoretical prediction of superhard and superconducting materials at ambient and high pressures.

Katerina Hilleke received her BS in Chemistry from Harvey Mudd College in Claremont, CA, in 2013. In 2019, she completed her PhD in Chemistry from the University of Wisconsin-Madison, working for Professor Daniel C. Fredrickson. She is now a postdoctoral research associate in the

group of Professor Eva Zurek, where she works on structure search methods for novel materials and properties, especially at high pressures.

Eva Zurek received a BSc in Chemistry and Physics (2000), and an MSc (2002) from the University of Calgary. Eva was awarded a PhD fellowship from the *Max Planck Research School for Advanced Materials* in Stuttgart, and her postdoctoral research was performed at Cornell University with Roald Hoffmann. In 2009 Eva became an Assistant Professor at the University at Buffalo, SUNY, where she was promoted to Full Professor in 2016. Eva has received a number of awards including the *Alfred P Sloan Fellowship*. Her research is geared towards studying the electronic structure, properties and reactivity of a wide variety of materials using first-principles calculations. She is interested in high pressure science, superhard, superconducting, quantum and planetary materials, catalysis, as well as solvated electrons and electrides. Eva develops algorithms for the *a priori* prediction of the structures of crystals, interfaces them with machine learning models, and applies them in materials discovery.