Structure prediction drives materials discovery

Artem R. Oganov[1,2,3]*, *Chris J. Pickard*[4,5]*, *Qiang Zhu*[6] *and Richard J. Needs*[7]

Abstract | Progress in the discovery of new materials has been accelerated by the development of reliable quantum-mechanical approaches to crystal structure prediction. The properties of a material depend very sensitively on its structure; therefore, structure prediction is the key to computational materials discovery. Structure prediction was considered to be a formidable problem, but the development of new computational tools has allowed the structures of many new and increasingly complex materials to be anticipated. These widely applicable methods, based on global optimization and relying on little or no empirical knowledge, have been used to study crystalline structures, point defects, surfaces and interfaces. In this Review, we discuss structure prediction methods, examining their potential for the study of different materials systems, and present examples of computationally driven discoveries of new materials — including superhard materials, superconductors and organic materials — that will enable new technologies. Advances in first-principle structure predictions also lead to a better understanding of physical and chemical phenomena in materials.

[1]*Skolkovo Institute of Science and Technology, Moscow, Russia.*

[2]*Moscow Institute of Physics and Technology, Dolgoprudny, Moscow, Russia.*

[3]*International Centre for Materials Discovery, Northwestern Polytechnical University, Xi'an, China.*

[4]*Department of Materials Sciences & Metallurgy, University of Cambridge, Cambridge, UK.*

[5]*Advanced Institute for Materials Research, Tohoku University, Aoba, Sendai, Japan.*

[6]*Department of Physics and Astronomy, High Pressure Science and Engineering Center, University of Nevada Las Vegas, Las Vegas, NV, USA.*

[7]*TCM Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK.*

**e-mail: a.oganov@skoltech.ru; cjp20@cam.ac.uk*

https://doi.org/10.1038/s41578-019-0101-8

New materials have historically been discovered by either trial-and-error processes or serendipity, both of which require labour-intensive and challenging experiments. In the past decade, it has become possible to discover new materials systematically on a computer, and the path to this breakthrough has been paved by the development of crystal structure prediction (CSP) methods[1–3]. There are two largely complementary approaches: one based on existing knowledge and the contents of crystal structure databases (data mining) and the other based on powerful exploratory computer algorithms capable of making predictions with little or no pre-existing knowledge.

Databases such as the Inorganic Crystal Structure Database (ICSD)[4] and the Pauling File[5] are invaluable resources that report experimentally observed structures of inorganic materials. The ICSD currently contains approximately 204,000 peer-reviewed data entries, and the Pauling File has some 335,000 entries; both databases are growing steadily. To date, approximately 159,000 ICSD entries have been assigned to a little over 9,000 distinct structure prototypes. However, the number of structures in the ICSD is small compared with the rapidly increasing number of structures that have been generated using first-principles methods.

Data mining approaches have received much attention[6–9]. In this Review, we focus on the fundamental and reliable non-empirical methods based on powerful exploratory algorithms. The major advantage of such methods is their ability to generate completely new knowledge beyond existing databases and intuition. We discuss these methods, examining the basic concepts and the systems they can be applied to. We then highlight examples of recent discoveries of counterintuitive new materials and phenomena achieved through the use of these methods, ranging from superhard materials to electrides, organic materials and superconductors with the highest known critical temperatures[10–12]. We finally outline the future perspectives for the field, examining the challenges that will need to be overcome, which include working with large systems and taking into account disorder and temperature, predicting synthesizable metastable structures and predicting chemical properties.

Considerations for materials prediction

Structure. Understanding the structure of matter at the atomic level is central to modern materials science. Until recently, experiments offered the only reliable source of crystal structures, but computational methods have emerged as a complementary source. In particular, density functional theory (DFT) methods and modern computing power can be combined to relax thousands of structures to local minima in the Born–Oppenheimer energy surfaces, the potential energy surfaces calculated in the approximation that the nuclear and electronic motions can be treated separately[13]. CSP aims to determine the minimum energy (or, in a more general case, the global minimum or maximum of a property of interest) over all values of the relevant inputs while

identifying the low-lying local minima (the metastable phases). Possible inputs could be, for example, the numbers and types of atoms in the system. A wide variety of systems can be handled by structure prediction, from molecules and clusters to 2D systems (2D crystals, crystalline surfaces or grain boundaries) and crystals (with quasicrystals beyond the scope of this Review). CPS has led to the discovery of many new structures that have subsequently been confirmed experimentally (TABLE 1). The possibility of discovering completely new structures is one of the key advantages of structure prediction as compared with data-driven approaches; data mining, on the other hand, can give reasonable, even if non-exact, solutions at a fraction of the computational cost.

The energy landscape. A very large number of structures can be generated using a source of randomness and then relaxing each of them to their lowest-energy local minimum. The basin of attraction is defined as the set of points within a structure space that leads to a particular minimum-energy configuration by a path of steepest descent on the potential energy surface. These basins of

Table 1 | **Selected examples of materials discovered or structurally characterized with the help of first-principles structure prediction**

Year	Material	Description	Refs
Crystalline materials			
2011	Derivative-DNTT	An organic semiconductor with high hole mobility (12.3–16.0 $cm^2 V^{-1} s^{-1}$)	193
2014	Zr_2Co_{11}	Hard magnet, structure of which was a long-standing puzzle	232
2014–2015	H_3S	Superconductor with one of the highest critical temperatures recorded (T_c = 203 K)	156,157
2011	δ-$Mg(BH_4)_2$	Hydrogen storage material obtained at elevated pressure (2.1 GPa)	48
2014	$[–NH–CO–NH–C_6H_4–]_n$, $[–CO–NH–CO–C_6H_4–]_n$, $[–NH–CS–NH–C_6H_4–]_n$	Polymers with high capacitance	233
2016	ε-Resorcinol	New polymorph with two molecules in the asymmetric unit obtained from melt recrystallization	188
2017	Coumarin (II–V)	New polymorphs with various numbers of molecules in the asymmetric unit (1,2 or 3) obtained from melt recrystallization	190
2017	Glycine dihydrate and new phase of glycene	A previously unknown phase and the first hydrate form of glycine	189
2017	Sr_5P_3	Novel electride material	181
2014	ZrO	Identification of suboxide phase in Zr/ZrO_2 interfaces	234
2014–2015	Li_7Ge_3	Phase formed in the lithiation of Ge anodes	56,57
2008–2009	γ-B	Superhard material, *Hv* = 50 GPa	138
2006–2013	M-carbon	Superhard material	33,131,136
2014–2015	TiN_2	Hard metallic phase (predicted *Hv* = 25.6 GPa)	150,151
2010–2013	FeB_4	Superconducting and controversial superhard material	143,144
2015	β-$Li_{15}Si_4$	Anode for Li battery	59
2016	C_3N_4	Thermodynamically stable tetragonal high-pressure phase of C_3N_4 predicted, synthesized and recovered	235
2018	W_2CrB_2 and W_4CrB_3	Hard precipitates in superalloys	149
2014	$P2_1/c$-MnB_4 and $C2/m$-MnB_3	New hard material MnB_3 (predicted *Hv* = 32.3 GPa). Refined and experimentally confirmed structure for MnB_4 (predicted *Hv* = 40.1 GPa)	148
2017–2019	LaH_{10}	Record high-T_c superconductor (experimental T_c = 250–260 K at 170–200 GPa)	10–12,169
2018	UH_7	New compound, predicted to be superconducting (T_c = 46–66 K, stable above 22 GPa and metastable at 0 GPa)	236
2018	Na_2B_{30}	Chemical composition and crystal structure of a puzzling compound clarified; its near superhardness predicted (*Hv* = 37.4 GPa)	53
Beyond crystal structures			
2014	B_{36}^-	First experimental evidence of a flat planar boron cluster	90
2014	B_{40}^-	First experimental evidence of boron fullerene	91
2014	Rutile-TiO_2-(110)	Elucidation of structures of a catalytically active surface	106
2016	Rutile-TiO_2-(011)	Elucidation of structures of a catalytically active surface	237
2017	2D tellurium	First elemental 2D materials in group VI elements	96
2015	Ni_3InAs	Determination of the composition and structure of nickelide contact material for InAs	238

DNTT, DNA nucleotidylexotransferase; *Hv*, Vickers hardness; T_c, superconducting critical temperature.

attraction cover the Born–Oppenheimer energy surface. Low-energy basins may be found clumped together in deep but smooth funnels, which allows the use of fast global optimization methods. The probability of successfully finding a low-energy structure depends on the shapes and sizes of the hyper-volumes of the basins of attraction and the details of the search[14–17].

A histogram of the energies provides information about the energy landscape of the system[18]. The efficiency of searches can be improved by reducing the size of the structure space of interest. For example, knowledge of existing structures and chemical and physical information can be used to ensure that the searches are initiated with a set of structures that are chemically reasonable while maintaining a large degree of randomness. Various fingerprints (a fingerprint is a set of values permitting comparison of two crystal structures; for example, this could be a histogram of interatomic distances and angles in a structure) can be used to build maps[18] of structural similarity (also known as sketch maps[19]), with similar structures located next to each other on a 2D space, allowing the visualization of multidimensional energy landscapes by introducing energy as the third dimension. It is also possible to build chemical maps, displaying how the various materials structures or properties vary with chemical composition; this includes structure maps[20,21] and maps of nanoparticle stability. The development of fingerprint functions that are sensitive to geometry as well as chemistry has led to the concept of materials cartography[22], which goes well beyond energy landscapes and allows the grouping of materials into classes possessing similar physical properties before the properties are measured or even calculated. These concepts (FIG. 1) are extremely powerful, permitting the analysis and rationalization of large structural data sets.

Structural symmetries. The structure space can be divided into regions according to their symmetry. For a reasonably large system, almost all of the structure space consists of regions of the lowest *P*1 symmetry (the symmetry group that consists only of translations). This is obvious when it is realized that choosing random structures almost inevitably leads to a structure without symmetry. The fraction of the structure space covered by regions of *P*1 symmetry increases approximately exponentially with system size. Yet, nature prefers crystal structures with symmetry, in contrast to usually asymmetric, and metastable, biomolecules. Not only is the asymmetric *P*1 space group almost non-existent in crystals, the distribution of crystal structures over Fedorov space groups is extremely uneven[23–26]. For example, one-third of all inorganic crystals belong to just 6 of the 230 possible space groups, *Pnma*, $P2_1/c$, *Fm*3*m*, *Fd*3*m*, *P*-1 and *I*4/*mmm*[26]. For organic crystals, this unevenness is even greater[23]; this is understood only partially. Pauling's fifth rule[27] (the number of essentially different kinds of constituents in a crystal tends to be small) can be simplistically viewed as implying that nature prefers high-symmetry structures with primitive unit cells containing only a few atoms. Such structures may therefore be found in inexpensive searches. One line of reasoning is that low-energy structures tend to include repeated identical units (atoms, molecules or other units) because each atom in the structure may be most stable in a particular environment. Constraining searches to conform to high-symmetry space groups is a useful strategy. The allowed symmetries can be reduced systematically, which helps in discovering lower-symmetry structures that may sometimes be the most stable. Imposing symmetry constraints can provide enormous reductions in the search space; alternatively, one can use algorithms that automatically focus on the low-energy parts of the energy landscape (such as evolutionary algorithms, metadynamics or minima hopping) and achieve the correct structures with the right symmetries.

Chemistry. The 118 elements in Mendeleev's periodic table can be combined in many different ways to form an enormous number of distinct chemical compounds. There must be many more possible compounds and structures than those that have been reported from over a century of X-ray diffraction experiments. With 118 elements in the Mendeleev table, 7,021 binary and 273,937 ternary systems can be constructed (each potentially with multiple stable compounds), but only 72% of binary, 16% of ternary and just 0.6% of quaternary systems have been fully or (probably more often) partially experimentally studied under normal conditions[28], and far fewer have been studied in extreme conditions. The possibility of discovering completely new structures is one of the key advantages of structure prediction as compared with data-driven approaches. Many of the published results of structure predictions involve crystal structures or topologies that were previously unknown. This situation is even more frequent at non-standard conditions. Chemistry provides crucial understanding of the ways in which atoms combine to form stable materials and of the nature of the chemical reactions by which one material is transformed into another. However, what can we do in a situation in which the standard rules of chemistry do not apply? This can happen, for example, under extreme conditions of high pressure and/or temperature. Nanoparticles and surfaces of crystals provide another setting in which to study unusual chemical phenomena.

Thermodynamics. Thermodynamics plays a central role in determining the structures that might be formed in nature or experiments. In real life, the effects of sample quality, annealing schedules, impurities or applied pressure are also important. Nevertheless, there are many examples in which the lowest-energy structure found at the DFT level of theory turned out to correspond to the experimentally observed structure. Low-energy metastable structures may also be important, as is the case for diamond. Most organic matter is also metastable.

Whereas the global energy minimum usually corresponds to an experimentally achievable stable crystal structure, nature shows some preferences in choosing which local-minimum (metastable) structures can be formed. Among the very large number of low-energy local minima, only a small subset appear

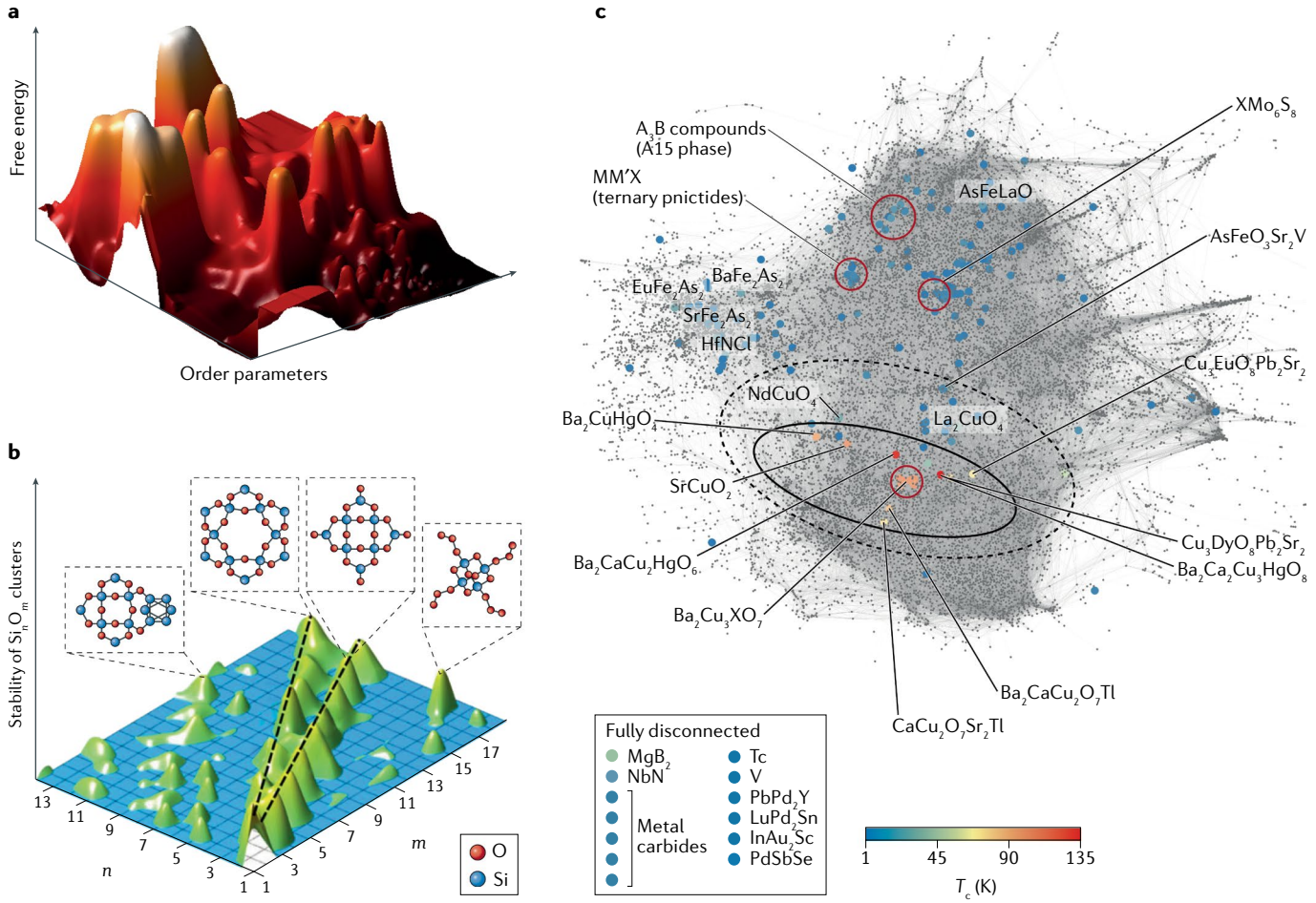


Fig. 1 | **Mapping the materials space. a** | Energy landscape of crystalline Au_8Pd_4, showing how the low-energy structures cluster in one area[39]. **b** | Stability map of Si_nO_m nanoparticles, showing ridges and islands of stability and a sea of instability[89]. **c** | Superconducting materials cartogram[22]. In all these cases, neighbouring points turn out to have similar structures and properties, and materials possessing particular stability or maximal properties are clustered in the same area of the map. Such landscapes with one or a few 'basins of attraction' are highly suitable for evolutionary optimization. Dynamical methods (metadynamics and minima hopping) and random sampling also benefit from such topology of the materials space. Panel **a** is adapted with permission from Oganov, A. R. et al. How evolutionary crystal structure prediction works and why. *Acc. Chem. Res.* **44**, 227–237 (REF. [39]), copyright 2011 American Chemical Society. Panel **b** is adapted with permission from Lepeshkin, S. V. et al. Method for simultaneous prediction of atomic structure and stability of nanoclusters in a wide area of compositions. *J. Phys. Chem. Lett.* **10**, 102–106 (REF. [89]), copyright 2019 American Chemical Society. Panel **c** is adapted with permission from REF. [22], ACS.

to be synthesizable, and it is not generally understood why the others cannot be made. Stevanovic connected the volume of phase space associated with each local minimum with the likelihood of synthesis of the corresponding structure[17], and Sun[29], analysing databases, found that observed metastable phases are usually less than 0.1–0.2 eV atom^{-1} above the ground state (explosives are a prominent exception). Furthermore, Sun[29] hypothesized that observed metastable phases should be thermodynamically stable at some values of pressure, temperature, chemical potentials, electric fields, particle size and other parameters. If true, these hypotheses would hand thermodynamics a decisive role in the prediction of synthesizable metastable polymorphs.

Extending structure prediction to variable compositions using chemical potentials allows the identification of stable compounds. The Maxwell convex hull construction affords a particularly appealing approach that can be used to identify stable and metastable structures and stoichiometries (BOX 1; FIG. 2). A structure on the convex hull is thermodynamically stable, and a structure above the hull is metastable. Note that there are potentially an infinite number of stoichiometries, and for each stoichiometry, there may be an infinite number of structures, although the number of different structures is reduced in systems with periodic cells or clusters with a finite number of atoms. Convex hull construction provides a global view of the relative stabilities of structures and stoichiometries, and can be used for binary, ternary and quaternary systems and so on. Examples of convex hulls for binary (Na–Cl) and ternary (Mg–Si–O) systems are shown in FIG. 2a,b. In the Na–Cl system, the only stable compound at normal conditions is NaCl;

Box 1 | **Crystal structure prediction with multiple solutions**

Crystal structure prediction aims to determine the lowest-energy structures, whereas the discovery of materials with a set of desired properties is aimed at a vast chemical and structural space. In this case, researchers usually do not look for a single solution in crystal structure searches. Instead, they search for multiple solutions.

Convex hull optimization

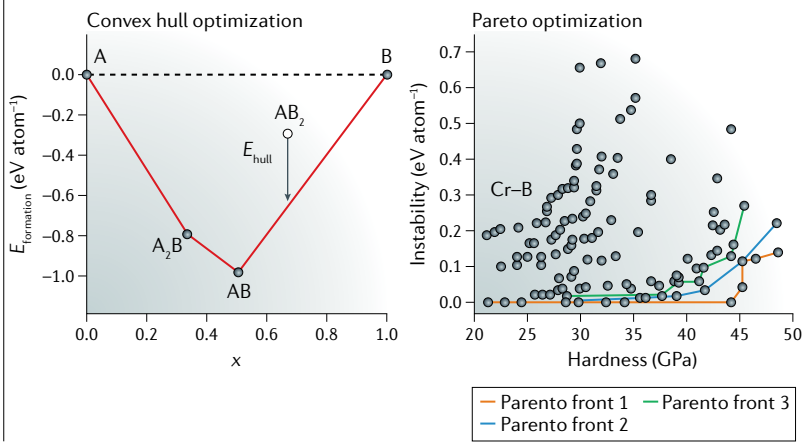
Discovering new materials requires an exploration of all possible stoichiometries for a given chemical system. To take into account the variation of chemical composition, we can devise a strategy to examine a range of compositions of interest for the given chemical system using a suitable fitness function (a function that enables ranking of candidate solutions by their optimality) to evaluate and compare structures with different stoichiometries. For a set of compounds with different stoichiometries, the stability can be measured by the minimum energy of transformation into an isochemical mixture of other phases. Let us take a binary system A–B as an example. The normalized energy of formation of A_xB_{1-x} can be expressed as:

$$\triangle E_{formation} = E(A_xB_{1-x}) - [xE(A) + (1-x)E(B)] \quad (3)$$

where $E(A_xB_{1-x})$ is the energy of the compound (normalized per atom) and the expression in brackets is the energy of the mixture of stable forms of A and B. A negative energy of formation is a necessary but not a sufficient condition of stability. A necessary and sufficient condition is that stable phases form a convex hull, as shown in the plot. A suitable fitness function is the minimum vertical distance from the convex hull (E_{hull}). The generalization of this scheme to multicomponent systems is straightforward.

Pareto optimization

If the aim is to suggest new materials with optimum physical properties, at least two variables need to be optimized: the target physical property (or properties) and stability. The solution of such a multi-objective optimization problem is, in general, not one material but a set of materials forming the so-called first Pareto front. It is good practice to pay attention to not only the first Pareto front but also the few lowest-rank Pareto fronts. The first Pareto front consists of the so-called non-dominated solutions, materials that cannot be beaten on all properties at the same time (but may be beaten on some). Removing the first Pareto front and repeating the same procedure gives the second Pareto front and so on. In the simplest scheme, the rank of the Pareto front can be used as fitness in multi-objective optimization. In practice, Pareto optimization works well when the number of objectives is not large, such as, less than four. In the figure below, the first three Pareto fronts are shown; the data were obtained from Pareto maximization of hardness and minimization of an instability (E_{hull}) for the Cr–B system.



however, under pressure, numerous Na–Cl compounds become stable[30], such as Na_3Cl and $NaCl_3$. Similarly, a very surprising compound, Na_2He, was predicted to be stable at pressures above 100 GPa (REF.[31]). These compounds were synthesized experimentally[30,31], and their structures are shown in FIG. 2c,d. The convex hull of the Mg–Si–O system (this system is a first approximation to the composition of mantles of terrestrial planets) at 500 GPa shows some new compounds that are unknown at normal conditions[32].

Materials discovery and design. The ability to make reliable predictions of likely compositions and their structures means that computational methods can routinely complement experimental efforts in searching for new candidate materials. These searches may target a chosen property (design) or survey a range of possibilities with minimal preconceptions (discovery).

Crystal structure prediction

Following a simple combinatorial argument[33], the number of possible distinct structures for a compound with N atoms in a unit cell of volume V can be estimated as:

$$C = \frac{1}{(V/\delta^3)} \frac{(V/\delta^3)!}{[(V/\delta^3)-N]!\,N!} \quad (1)$$

where δ is a relevant discretization parameter (for instance, 1 Å). Already for small systems ($N \approx 10$–20), C is astronomically large (roughly 10^N using $\delta = 1$ Å and a typical atomic volume of 10 Å^3) and increases exponentially with the number of degrees of freedom d ($d = 3N + 3$): $C \approx \exp(ad)$, where a is some system-specific constant. Clearly, it is not feasible to examine all possible arrangements of the atoms in space — there are just too many of them.

If each structure is relaxed, the number of degrees of freedom will decrease because correlations between atomic positions will emerge (interatomic distances adjust to reasonable values, and unfavourable interactions are avoided). Because the complexity of the problem is exponential in d, this simple trick simplifies the problem greatly, as it reduces the effective d. For example, d is reduced from 39 to 10.9 in Au_8Pd_4, from 99 to 11.6 in $Mg_{16}O_{16}$ and from 39 to 32.5 in $Mg_4N_4H_4$ (REF.[34]). Not surprisingly, all successful structure prediction methods include structure relaxation. Nevertheless, even for relaxed structures, one encounters the 'exponential wall', or non-deterministic polynomial-time (NP)-hard behaviour: the number of possible local minima increases exponentially with the number of atoms in the unit cell (for a rigorous proof, see REF.[35]), though with a reduced exponent.

Among the most popular CSP methods are random sampling[15,36–38], various evolutionary algorithms[39–44], metadynamics[45] and minima hopping[46,47] (BOX 2).

Working with experiment. Combining structure prediction with diffraction or other experimental methods has provided a very successful approach to solving crystal structures. Some diffraction data may be insufficient to permit the solution of a particular structure, and sometimes even the exact stoichiometry is hard to establish, but information on lattice constants is often available, as perhaps are some indications of likely space groups. Knowledge of the lattice constants provides powerful constraints on the possible structures, which greatly simplifies the searching. Low-quality

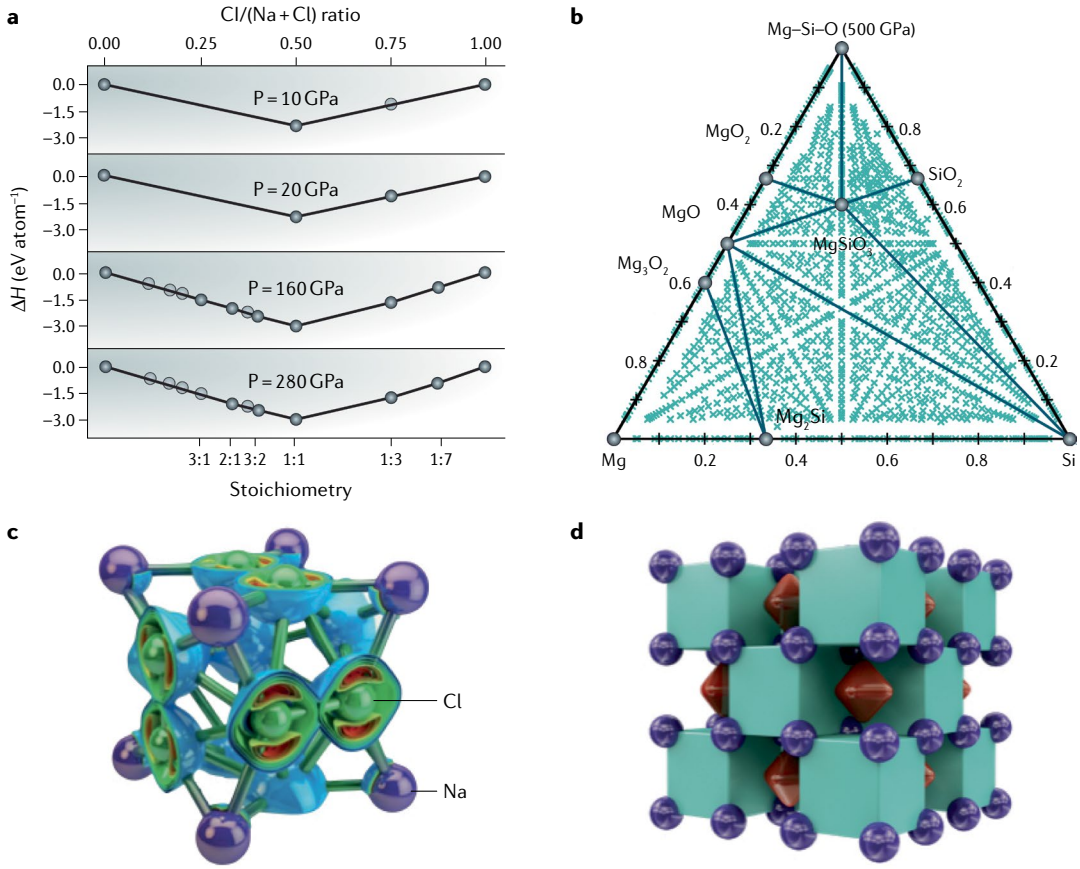


Fig. 2 | **Compound prediction with crystal structure prediction methods. a** | Convex hulls for the binary Na–Cl system at different pressures. **b** | Convex hull of the ternary Mg–Si–O system at 500 GPa. Light blue symbols denote sampled metastable compositions, and circles denote stable phases. **c** | Structure of the $NaCl_3$ compound, with isosurfaces of the electron localization function (with values increasing from blue to red). **d** | Structure of the compound Na_2He, with interstitial maxima of the electron localization function shown as red objects inside empty cubes[31]. ΔH, enthalpy of formation. Panel **a** is adapted with permission from REF.[30], AAAS. Panel **b** is adapted from REF.[32], CC-BY-4.0.

diffraction data may be misinterpreted in experiments, and structure prediction could yield better models for structural refinement[48–50]. Constraints from vibrational data, solid-state NMR, X-ray absorption spectroscopy and other spectroscopies can also be exploited. Once a technologically useful material is predicted, it is important to try to verify it by experiment. Knowledge of the crystal structure is crucial for understanding the performance of materials, and when experimental data are insufficient, structure prediction is helpful. For example, CSP has resolved a long-standing controversy over the structure and stoichiometry of sodium boride: in addition to the undisputed Na_3B_{20}, a phase identified by some researchers as orthorhombic (*Imma*) Na_2B_{30} (REF.[51]) and by others as monoclinic (*C*2) Na_2B_{29} (REF.[52]) has been observed. This intriguing phase has been recently solved by CSP[53], which established that the correct stable stoichiometry is Na_2B_{30} and that neither of the previously proposed structures are global minima. The predicted $I2_12_12_1$ structure matches experiments. Unlike metallic Na_2B_{29} or semi-metallic *Imma*-Na_2B_{30}, this phase is semiconducting and nearly superhard (predicted Vickers hardness 37.4 GPa). The structure of β-NiOOH (REF.[54]), an active component of the highly active catalyst for water oxidation, was fully established thanks to CSP. Likewise, a newly predicted reconstruction of the (110) surface of rutile-type RuO_2 (REF.[55]) has explained the extraordinary pseudocapacitance of RuO_2 and established that it is partly due to a surface redox reaction. A Li_7Ge_3 phase was first proposed in a CSP study[56], and the structure was later observed in experiments[57].

At high pressures, the quality of experimental information is often insufficient for solving crystal structures. In such cases, theoretical input is invaluable. Neutron studies of ammonia monohydrate phase II found a detailed powder diffraction pattern with numerous well-defined peaks. However, the structure could not be solved using the experimental data alone. Ab initio structure searches using the known unit cell parameters and likely space groups as constraints uncovered a structure with 112 atoms in the unit cell, in almost perfect agreement with experiment[50]. Several high-pressure phases of $Mg(BH_4)_2$ were synthesized and apparently convincingly resolved using powder X-ray diffraction data[58]. However, a subsequent CSP study suggested that two new tetragonal structures with space groups *P*-4 and $I4_1/acd$ were lower in enthalpy than the earlier proposed experimental $P4_2nm$ structure by 15.4 kJ mol^{-1} and 21.2 kJ mol^{-1}, respectively[48]. Interestingly, the X-ray powder diffraction (XRD) patterns of all

Box 2 | **Crystal structure prediction**

The crystal structure prediction (CSP) problem is a global optimization problem. Of critical importance are the development of smart global optimization algorithms to navigate the structure space, methods for calculating structural energies and properties and for local optimization of geometries and the increase in available computing power.

Geometry optimization

Efficient structure relaxation requires the calculation of forces and stresses. The simplest, but usually not the most efficient, algorithm is the steepest descent algorithm, in which atomic coordinates and lattice vectors are updated iteratively, moving along the force and stress direction. In this algorithm, only the current positions of atoms and forces are considered. The conjugate gradients algorithm also requires only the forces and stresses (that is, the first derivatives of the energy), but it takes history into account and is usually more efficient. Knowledge of the matrix of second derivatives (Hessian) can greatly accelerate convergence to the local minimum, and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm constructs an approximate numerical Hessian for this purpose. More efficient algorithms based on conventional molecular dynamics with additional velocity modifications and adaptive time steps are also available[216].

Methods not based on global optimization.

Two important approaches belong to this group: data mining[6–8] and cluster expansion[211,212]. Data mining allows very quick predictions of the stable crystal structure and is efficient in searching for materials with optimal properties but, relying on databases of crystal structures, is unable to predict completely new crystal structures. Cluster expansion starts with knowledge of the underlying crystal structure and allows the prediction of the ordering of the atoms and/or magnetic moments as a function of temperature. Both methods require considerable amounts of pre-existing structural information, whereas predicting crystal structures by global optimization requires little or no such information.

Global optimization methods and codes

Global optimization is a very large field in applied mathematics (for an introduction, see REF.[217]), with many methods continually developed. In application to CSP, it is important to remember that no existing method can give a guarantee of finding the global minimum in a finite amount of computational time. We list here some popular methods.

- The simplest search strategy, used since 1993 (REF.[36]), would be to sample structures quasi-randomly[15,36–38]; however, in practice, it is crucial to steer the searches towards finding realistic structures while maintaining structural diversity. This is achieved by imposing constraints on symmetry, interatomic distances, coordination numbers, stoichiometries, dimensionality and structural units.
- Simulated annealing[218,219] is a strategy inspired by annealing of crystals, in which gradual cooling leads to the equilibrium crystal structure. Importantly, structure relaxation is not performed during this process, and it is done only during the analysis of results.
- Basin hopping[220] combines, in a loop, large steps in the configuration space followed by structure relaxation and uses the Metropolis criterion for the accept/reject step at each move.
- Metadynamics[45] requires a collective variable to distinguish between states of the system (for CSP, cell vectors are currently used for this purpose as a pragmatic, but imperfect, choice) and scans the low-energy part of the energy landscape by distorting the landscape by a history-dependent potential, the aim of which is to discourage the system from sampling states that have already been sampled. Metadynamics finds low-energy structures and transition paths between them.
- Minima hopping[46] performs short molecular dynamics runs followed by structure relaxation. If the relaxation leads to a structure that has already been visited, a new molecular dynamics calculation is initiated at a higher temperature, and this process is continued until a new structure is found.
- Evolutionary algorithms come in many types with widely different performance[33,40–44,221–223] (the most popular are the Oganov–Glass evolutionary algorithm[33,78] and Wang's version of particle swarm optimization[42]). The common idea is that a population of structures is evolved, driven by natural selection of lower-energy structures that become parents of a new generation of structures. Recipes for producing offspring from parents (genetic crossover and mutations) are of key importance. Hybrid methods also exist, for example, evolutionary metadynamics[224].

Some publicly available codes for structure prediction (with their main areas of application) include the following: AIRSS (inorganic); CALYPSO (inorganic); CrySPY (inorganic); DMACRYS (organic); GASP (inorganic); GAtor (organic); GRACE (organic); MAISE (inorganic); Molpak (organic); UPack (organic); USPEX (inorganic and organic); and Xtalopt (inorganic).

Computation of structural energies

First-principles structure prediction has been made possible by the development of robust and efficient density functional theory (DFT) codes and high-quality pseudopotentials across the periodic table[225]. DFT calculations become too expensive for large systems, but there is rapid progress in the development of machine learning force fields[226–228], which need to be trained on DFT data (energies, forces and stresses) and may deliver the same accuracy as first-principles methods with a speed up of 2–4 orders of magnitude. Recently, it has been demonstrated that combining CSP and machine learning force fields works well in applications[140,229,230]. In addition to the use of machine learning for constructing force fields, it is possible to develop quantitative structure–property relations[9] and then inexpensively predict (or optimize) properties of interest. Normally, properties should be computed for carefully relaxed structures, but at least semi-quantitatively, they can be computed from just the chemical composition and topology of the structure[231].

three structures are compatible with experiment at ambient pressure. However, $I4_1/acd$ is the true thermodynamic ground state, whereas $P4_2nm$ is not even a local minimum. $Li_{15}Si_4$, a potentially useful anode material for Li-ion batteries, has been found[59] to undergo a phase transition to an unknown structure at a pressure of 7 GPa, but the structure could not be solved experimentally; evolutionary metadynamics calculations starting from the ambient-pressure structure (with 152 atoms in the unit cell) have identified the *Fdd*2 structure, which reproduces the experimental XRD pattern. The newly discovered β-$Li_{15}Si_4$ is recoverable to ambient conditions and may exhibit improved cycling properties (for example, it has a smaller volume change upon delithiation).

Molecular crystals under pressure. The use of molecules, rather than single atoms, as building blocks allows much more complex structures to be predicted. Already the simplest molecular solid, hydrogen turns out to be not so simple. The experimental determination of high-pressure phases of solid hydrogen is very challenging because of the small X-ray scattering cross-section and small sample sizes in high-pressure experiments[60,61]. Raman and infrared spectroscopies have provided a great deal of vibrational data[60,62,63]; however, it was not possible to resolve the structures of phases III and IV using the available vibrational data. Structure searches found that the lowest-enthalpy phases III (theory finds two variants of phase III[64,65]) and IV[62] are layer-like with Raman and infrared spectra in good agreement with experiment.

Experimental evidence for a phase V of solid hydrogen was reported[63], and DFT calculations[66] suggested that it could be a stepping stone towards metallization of hydrogen. Structure searches have thus advanced knowledge of the phase diagram of hydrogen at high pressures and low temperatures.

Simple molecules such as water, ammonia and methane are fundamental to chemistry and make up most of the gas giant planets Uranus and Neptune, though in the liquid state. Their behaviour under pressure is complicated, for example, ice has 17 experimentally identified polymorphs on the phase diagram at pressures up to 210 GPa (REFS[67,68]). Under pressures of 90–331 GPa, ammonia was predicted to form ionic solids consisting of NH_4^+ and NH_2^- ions[69]. Two crystalline ionic forms, *Pma*2 and *Pca*2_1, were confirmed by experiment[70]. The identification of the high-pressure phases of solid methane above 5 GPa provides another example of successful interplay between theory and experiments. Experiments suggested that a new solid phase of methane A above 5 GPa should contain 21 molecules in a pseudo-cubic rhombohedral unit cell, yet the structure could not be solved[71]. On the basis of the experimentally determined cell parameters, a CSP study revealed a ground state of rhombohedral symmetry with icosahedral packing of methane molecules[72], very similar to the results of a recent neutron diffraction experiment[73] except that orientationally ordered molecules were used in refining the diffraction data. Moreover, compressed methane displays an icosahedral structure reminiscent of intermetallics that are prone to the formation of quasicrystals, which leads to the question of whether molecular quasicrystals are possible[74].

Role of databases. The increasing growth of open computational materials databases makes the screening of materials with target properties possible. Exploitation of pre-existing data is very important. Existing databases are largely populated by experimentally determined structures, but the computational discovery of structures also feeds into databases and is increasing the power of data-driven approaches. We are confident that there will always be a need for structure prediction: even if it were possible to computationally capture all of the hidden structures, the need to explore and predict their defects, surfaces and behaviour at non-ambient conditions will keep driving efforts in structure prediction. The capability of first-principles approaches to unlock new discoveries indicates that they will have the same long-lived importance as experiments. The main aims of computational predictions for structures of materials are to predict materials with a desired combination of properties for technological applications; to predict stable and metastable compounds and crystal structures at various conditions, including those difficult to reach experimentally; to discover previously unknown structure types and topologies; and to help in determining the structures adopted by specific materials.

Perhaps the simplest way to increase the number of structures is to create 'chemical mutations' of known structures. For example, starting from an elemental structure composed of several atoms, we can replace some of them by other atomic types, perhaps from the same column of the periodic table, in the same spirit as the substitution method used in data mining[75]. This might lead to the generation of many very similar structures, but it could also lead to new structures.

A high-throughput survey based on both data mining with chemical substitution and evolutionary algorithms has been recently conducted to search for new photoactive semiconductors[76]. Four metastable compounds were identified, and in all of them, evolutionary searches found lower-energy structures than data mining: for $Sn_5S_4Cl_2$ by 24.7 meV atom^{-1}, for Sn_4SF_6 by 5.1 meV atom^{-1}, for Cd_4SF_6 by 0.2 meV atom^{-1} and for $Cd_5S_4Cl_2$ by 33.3 meV atom^{-1}. None of these structures can be found by data mining alone; all, except $Cd_5S_4Cl_2$, have no known structural analogues, and whereas $Cd_5S_4Cl_2$ belongs to the known Li_5BiO_5 structure type, it was not found by the substitution algorithm because of the very unusual coupled substitution required: Bi + O→2Cl.

Very recently, a co-evolutionary method, called Mendelevian search, was developed[77]. It can be viewed as evolution of a population of evolutionary searches, each of which focuses on a particular chemical system; these systems compete and exchange information with each other, leading to progressively better systems being sampled. This method, using no empirical information, found that the hardest possible materials are allotropes of carbon, diamond and lonsdaleite and that the highest possible magnetization at 0 K is achieved in iron. In addition, it predicted a number of interesting materials.

Discovering materials with optimal properties. Metastable materials are often more interesting for applications than stable ones, for example, diamond, glasses and most organic molecules are metastable, but there are an infinite number of possible metastable materials, so a method is needed to predict those with interesting properties that can be synthesized. Much can be learned from a systematic analysis of the rich data produced by CSP, but there are also special methods that target the low-energy part of the energy landscape at the expense of extensive sampling of high-energy structures — among these are metadynamics[45], minima hopping[47] and evolutionary algorithms[33,42,78].

Furthermore, it is possible to search for materials with an optimum value of a physical property (or multiple properties) of interest. We argue that computation-based non-empirical searches offer the most reliable path to discovering materials with superior properties. In general, three types of global optimizations can be imagined.

The first is the minimization of the thermodynamic potential to determine stable structures. The internal energy E must be minimized to find the most stable structure of a given chemical compound at zero pressure and temperature. When working at non-zero pressures, the enthalpy H must be minimized in equilibrium. When working at non-zero pressures and temperatures, the Gibbs free energy G should be minimized.

The second option is the optimization of the physical property of interest (such as the hardness[79,80], density[81], bandgap[82] or the thermoelectric figure of merit[83]).

These should be extremized, or some target value must be approached, for example, for absorption of sunlight, a direct gap as close as possible to 1.34 eV is desirable. This predicts the upper limit of materials performance, but may lead to an ill-defined problem when there is no upper limit or produce structures that are so high in energy that they cannot be synthesized.

The third type of global optimization is the multi-objective (Pareto) optimization, in which two or more properties are simultaneously optimized. In our opinion, this type of optimization is most directly related to practical applications; for example, simultaneously optimizing the stability and physical properties of interest, as was done for superhard materials[84,85] and thermoelectrics[83], leads to the identification of materials that have attractive properties and at the same time can potentially be synthesized. The solution of a multi-objective optimization problem is, in general, not one material but a set of materials forming the so-called first Pareto front (BOX 1).

Beyond crystal structure

Structure searches using first-principles methods are not confined to 3D systems. It is also possible to search for structures of point defects, clusters, solid surfaces and interfaces. The prediction of non-crystalline structures presents challenges, but it is extremely important, and considerable progress has been made.

Nanoclusters. Materials strength is largely determined by microstructure. Nanostructuring can be exploited to obtain extreme materials properties, for example, in thermoelectrics (to decrease thermal conductivity), in catalysts and batteries (to increase the effective surface area) or in structural materials (to improve hardness, exploiting the Hall–Petch effect). Quantum confinement enables tuning the electronic and optical properties of nanoparticles by changing their size. Furthermore, the chemistry of nanoparticles can be very sensitive to the chemical environment, and quite unexpected cluster compositions can be 'magic', that is, particularly stable; for example this was shown for Fe–O and Ce–O clusters[86]. It has been shown that, under normal conditions (300 K and 0.21 atmospheres partial pressure of O_2), the predominant Si_7O_m, Fe_4O_m and Ce_4O_m nanoparticles are oxygen-rich magnetic Si_7O_{19}, Fe_4O_8 and Ce_4O_{14} clusters rather than normal Si_7O_{14}, Fe_4O_6 and Ce_4O_8, respectively[86,87]. The presence of reactive oxygen species, such as ozonide-groups (O_3^{2-}), in such clusters may explain the known carcinogenicity of small silica particles[88]. An evolutionary method[89] was developed for the simultaneous prediction of structures of clusters in wide ranges of chemical compositions and for the automatic prediction of magic clusters; this method automatically produces nanoparticle stability maps similar to those presented in FIG. 1b and showed a 5–50 times speed up compared with a one-by-one study of all cluster compositions.

There has been much interest in making boron nanostructures that are analogous to carbon nanotubes, fullerenes and graphene. Although photoelectron spectroscopy (PES) can yield well-resolved spectra serving as an electronic fingerprint, structure determination relies on input from theory[90,91]. Ground-state structures were found to be very diverse; for example, B_{36}^- has a quasi-planar structure, whereas a fullerene-like cage structure becomes energetically favourable in B_{40}^-. The calculated PES spectra agree satisfactorily with experiments.

2D crystals. The discovery of graphene, with its remarkable physical properties and potential applications ranging from reinforcing structural materials to novel electronics, has initiated tremendous efforts on 2D materials research. Much of the early research focused on obtaining 2D materials from bulk samples, for example, by mechanical exfoliation. This requires the bulk material to have a layered structure with weak interlayer bonding so that a few atomic layers can be exfoliated and hopefully remain stable in air without significant atomic rearrangement. A number of 2D materials beyond graphene were discovered using this strategy, including transition-metal dichalcogenides (MX_2), transition-metal carbides or nitrides (MXenes) and phosphorene, a single layer of black phosphorus. A computational screening (data mining) revealed that 826 stable layered materials could be considered as candidates for the formation of 2D monolayers via exfoliation[92]. An evolutionary method has been developed that can predict viable compositions and structures of free-standing 2D materials using no experimental information[93]. In addition to free-standing 2D materials, there is also great interest in pursuing 2D materials grown on a substrate. Given the various choices of substrate materials and the complex interactions between substrates and 2D materials, the potential search space is massive, making first-principles predictions an invaluable tool. Indeed, the discovery of two allotropes of borophene[94] was stimulated by a theoretical prediction[95]. More recently, 2D tellurium has been proposed[96]. Though challenging to make, this class of material may exhibit advantageous properties. For instance, unlike bulk boron allotropes, borophene shows metallic conductivity, whereas the effective masses of charge carriers in semiconducting 2D tellurium are calculated to be three or four times lower than in MoS_2, a commonly studied 2D semiconductor. 2D materials can also be formed by molecules; for example, 2D ice has been studied extensively under various experimental conditions. Despite their relevance to disparate areas such as cloud microphysics and tribology, the phase transitions in low-dimensional ice are not fully understood. Inspired by the observation of 'square ice' sandwiched between graphene sheets, constrained searches have revealed a rich phase diagram of 2D ice[97–99] (FIG. 3a).

Surfaces. Surfaces of many materials, especially semiconductors, usually have different structures from the bulk, and their reconstructions play a key role in determining their properties and behaviour (such as epitaxial crystal growth or catalysis). The prediction of surface reconstructions is a long-standing problem. Just as for crystals, it is not possible to enumerate all reconstruction models, nor does chemical intuition suffice to produce reliable models, even for elemental systems (consider the complexity of the Si-(111) 7×7 reconstruction[100]).

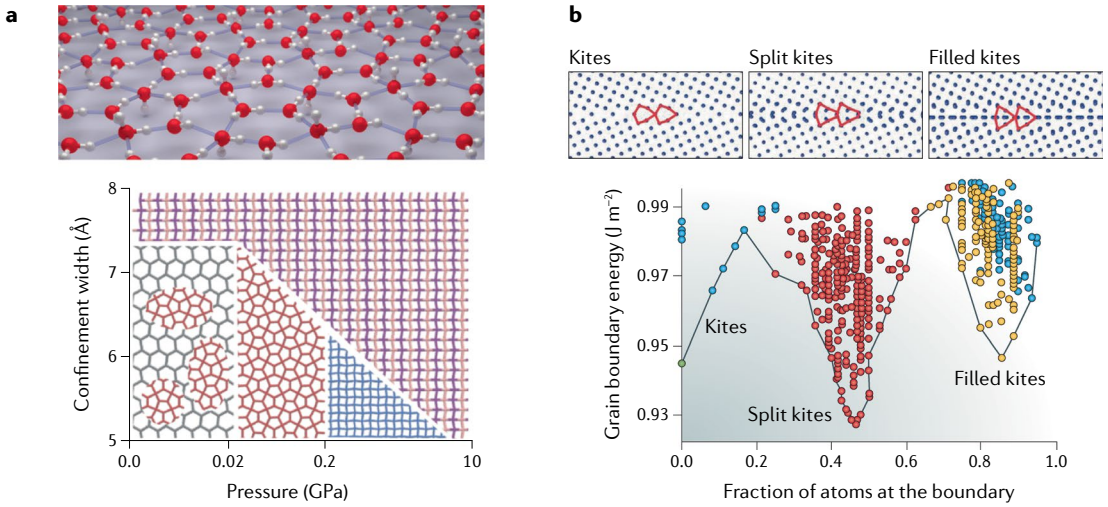


Fig. 3 | **Applications of crystal structure prediction to systems beyond bulk crystals. a** | Phase diagram of monolayer ice with respect to lateral pressure and confinement width[97]. The colours are guides to the eye. The top panel shows a rendering of the 'Cairo' tiling indicated in red on the phase diagram. **b** | Evolutionary search and clustering analysis of grain boundary phases of Cu-5(210)[001] (REF.[114]). Three grain boundary phases are identified (top), corresponding to three minima of the energy as a function of the fraction of atoms at the boundary (bottom). Panel **a** is adapted from REF.[97], CC-BY-3.0. Panel **b** is adapted from REF.[114], CC-BY-4.0.

The extremely large search space makes surface structure prediction difficult, and a number of methods have been proposed[101–105]. An evolutionary technique to predict surface structures based on the evolutionary algorithm USPEX has been developed[101]. It allows automatic exploration of stable and low-energy metastable configurations with variable stoichiometry and variable surface cells in the physically accessible range of chemical potentials[101]. This method has helped to resolve controversies about the reconstructions of the (110) surface of rutile (TiO_2)[106] and, because surface band structure depends on the atomic arrangements, can be used to tune functional materials to obtain better light absorption properties[107]. At different values of the chemical potential of oxygen, four different surface phases were observed for this surface; theoretically predicted models[106] provide invaluable insights into previous experiments. Additional degrees of freedom can also be explored, such as ferroelectric polarization, which was investigated for the $BaTiO_3$-(001) surface[108].

Interfaces and grain boundaries. Grain boundaries can exhibit different phases depending on the conditions of formation; studying the phase diagrams of grain boundaries is an emerging field[109]. In the traditional γ-surface approach, grain boundary models are constructed by joining two misoriented crystals while sampling different translations of the grains parallel to the grain boundary plane. The lowest-energy configuration is then taken to be the ground state. Calculations of silicon twist boundaries showed ordered ground states at 0 K (REF.[110]). A method based on molecular dynamics that allows variations in the atomic density was also proposed[111,112]. Application of these methods to fcc metals (Cu, Ag, Au and Ni) using embedded-atom force fields led to the prediction of new ground states and multiple phases of several [001] symmetric tilt boundaries with different atomic densities and complex periodic units many times larger than those of the bulk crystals[111,112]. Both ab initio random sampling[113] and evolutionary[114,115] approaches have been applied to the first-principles prediction of grain boundaries. The evolutionary approach samples grain boundary structures with varying numbers of atoms and cell sizes; a rich polymorphism of grain boundary structures of symmetric tilt boundaries of copper and of a series of bcc metals were found within the entire misorientation range[114,115] (FIG. 3b). Grain boundary structure prediction has been extended to multicomponent systems, such as stoichiometric interfaces between silicon and alumina[116] and non-stoichiometric grain boundaries in $SrTiO_3$ (REFS[113,117,118]). Grain boundaries can display new physics through quantum confinement effects and unexpected stoichiometries; the observed superconductivity at $LaAlO_3/SrTiO_3$ interfaces[119] is still unexplained. Grain boundary structure prediction will enable the design of polycrystalline and composite materials.

Point defects. Localized defects frequently determine the physical properties of materials. Semiconductors are doped by introducing impurities, and ionic conductivity is controlled by vacancies in the host lattice. The nitrogen-vacancy centre in diamond has been suggested as a source of qubits for future quantum technologies[120]. Until recently, atomistic models of defects were hand-built using a combination of intuition and pre-existing structures in related systems; one example is the model of the silicon tetra-interstitial[121] based on earlier results[122] on the structure of nitrogen platelets in diamond. Using random structure searches[38], interstitial defects in silicon were explored by removing 5 silicon atoms from an atom-centred sphere in a 32-atom supercell. Six silicon atoms were then repeatedly and randomly placed within the void, and the

resulting structures were relaxed using DFT methods and ranked according to total energy. Split-<110> and hexagonal interstitial structures were readily recovered as the most stable configurations. This approach was further developed and applied to a variety of complex defects in silicon[123,124] and zirconolite[125]. An extension of the convex hull construction was proposed to provide a graphical representation of the relative stability of point defects with variable composition[126]. Albeit with a focus on empirical potentials, an evolutionary algorithm has been developed specifically to explore local defect structure[127].

Examples of new materials

Hard and superhard materials. Hard materials are essential for cutting, drilling and machining tools. Pioneering experiments[128] and subsequent studies[129,130] have shown that compression of graphite at room temperature to ~17 GPa creates a new transparent superhard allotrope of carbon. It was found[131] that a monoclinic metastable structure called M-carbon, first reported in REF.[33], matches experimental X-ray diffraction patterns, though other structures were later proposed[132–134] and shown to match the low-resolution experimental data. Later, a transition path sampling calculation suggested that M-carbon has the lowest barrier of formation from graphite at room temperature; thus, it is kinetically the most likely to form among all sp^3 carbon forms[135]. Finally, a higher-resolution X-ray diffraction study produced patterns that were compatible only with M-carbon, supporting its identification as the structure of the metastable superhard carbon allotrope[136].

All boron allotropes are superhard[137], that is, they have Vickers hardness >40 GPa. The structure of its hardest allotrope, γ-boron, was predicted and then confirmed by experiment[138]. γ-Boron has a very wide stability field, in the range of 8–89 GPa, and a hardness of 50 GPa (REF.[139]) and is recoverable to ambient conditions, which means it could in principle be used as a superhard material (the only problem being the need to use high pressure for synthesizing it). Several low-energy metastable structures were predicted[16,140] and may eventually be synthesized.

Transition-metal borides, carbides and nitrides comprise a prominent class of hard materials. Cr–B, Cr–C and Cr–N systems were explored to search for maximally hard and stable phases using Pareto optimization[84]. It was found that CrB_4 is the hardest compound among these systems; its predicted hardness is 48 GPa, making it superhard[84,141]. The W–B system was also studied[142], and among the predicted new stable phases was WB_5, with a predicted hardness of 45 GPa and a very high fracture toughness of 4 MPa $m^{-½}$; this material was predicted to retain excellent mechanical properties even at high temperatures (~2,000 K). Borides may in general be more promising systems for finding superhard materials than carbides or nitrides because electrons donated by metal atoms to carbon or nitrogen will occupy antibonding orbitals, whereas electron-deficient boron, accepting an electron (according to the Zintl–Klemm rule), will behave like a carbon atom; thus, metal borides can be analogous to superhard sp^3 forms of carbon. Within this picture, borides can be harder than pure boron allotropes but cannot exceed the hardness of cubic BN (~60 GPa). A CSP study[143] suggested the existence of semiconducting FeB_2 and superconducting FeB_4. The latter was predicted to be marginally metastable but was synthesized under moderate pressures of 8 GPa (REF.[144]) and found to be superconducting (with a superconducting critical temperature, T_c, < 2.9 K) and controversially superhard (the measured[144] hardness of 62 GPa is most likely a large overestimate[145,146]).

The compositions and structures of stable and low-energy metastable borides of 41 metals have been thoroughly studied in a work[147] that provided a useful broad-brush picture. Furthermore, a $P2_1/c$ structure of MnB_4 (REFS[147,148]) was found to be the ground-state structure, rather than the $C2/m$ structure (reported in an early experimental work but later shown to be dynamically unstable), and a new compound, MnB_3, was predicted[148] to be stable at normal conditions. Predictions of both MnB_3 and the $P2_1/c$ structure of MnB_4 have been confirmed by experiments[148]. Both MnB_3 and MnB_4 were predicted to have very high Vickers hardnesses of 32.3 GPa and 40.1 GPa, respectively. Using a combination of variable-composition compound prediction and scanning transmission electron microscopy, tiny (~200 μm in size) W–Cr–B precipitates in a Ni-based superalloy were explored[149]. CSP searches revealed two stable ordered stoichiometric ternary borides, W_2CrB_2 and W_4CrB_3, the structures of which explain experimental observations. Crystal structure of precipitate phases is important for understanding the precipitation hardening of superalloys. A study on the Ti–N system at pressures up to 60 GPa found several new stable phases[150], the most extraordinary and hardest of which is titanium pernitride $I4/mcm$-TiN_2 (predicted Vickers hardnesses of 25.6 GPa at ambient pressure), which was subsequently synthesized[151].

Superconductors. Superconductivity is used, among other applications, in magnetic resonance imaging (MRI), magnetic levitation, particle colliders and fast electronic switches. Superconductivity can be classified as conventional, based on the phonon-based Bardeen–Cooper–Schrieffer (BCS) mechanism with s-wave electron-pairing, or unconventional, based on a mechanism that is still not theoretically understood. Until 2014, all high-temperature superconductors were cuprates displaying unconventional superconductivity, and the record for the highest T_c was held by $HgBa_2Ca_2Cu_3O_{8+\delta}$, with a T_c of 133 K (REF.[152]) at ambient pressure, increasing to 166 K at 23 GPa (REF.[153]).

An upsurge in searches for high-T_c conventional superconductors was initiated by Ashcroft's prediction of potential high-T_c superconductivity in solid hydrogen[154] and in hydrogen-rich materials[155]. Three properties of metallic hydrides are of particular importance here: the presence of a high hydrogen-derived electronic density of states close to the Fermi level, strong electron–phonon coupling and high phonon frequencies.

Ground-breaking work on hydrogen sulfide at high pressures led to a new record of high-temperature superconductivity in H_3S, with a T_c of 203 K at ~155 GPa.

H_3S was predicted[156] computationally using the USPEX code and then verified experimentally[157]. This is the first example of a previously unknown material predicted to be a high-temperature superconductor that has been confirmed experimentally. H_3S at high pressures was found to be a strongly anharmonic phonon-mediated superconductor exhibiting hydrogen bond symmetrization[158,159]. The predicted high-pressure structure of H_3S was confirmed experimentally[160,161]. Conventional H_2S was found to decompose readily at high pressures, and the energetics of this process were studied using the AIRSS and CALYPSO structure prediction methods[162]. Further investigation by USPEX has produced a more complete phase diagram of the H–S system[163].

The next record of high-T_c superconductivity has been predicted and already confirmed in metal hydrides. We mention predictions that CaH_6 (REF.[164]), YH_6 (REF.[165]), ThH_{10} (REF.[166]), AcH_{10} and AcH_{16} (REF.[167]) all have $T_c > 200$ K at pressures of 100 GPa or above. Even room-temperature superconductivity has been predicted in such systems, albeit at extreme pressures. According to a recent survey of rare-earth hydrides at high pressures[168], structures of the stable H-rich clathrates feature H_{24}, H_{29} and H_{32} cages with weak covalent bonding between hydrogen atoms and rare-earth atoms occupying the centres of the cages. Among these, sodalite-type YH_{10} (with H_{32} cages) was predicted to be a room-temperature superconductor with an estimated T_c of up to 303 K at 400 GPa, as derived from the Eliashberg equation. Independent work[10] predicted the same material, YH_{10}, and isostructural LaH_{10} as room-temperature superconductors, with T_c values of 305–326 K at 250 GPa and 274–286 K at 210 GPa, respectively. LaH_{10} has already been synthesized[169], and there are two experimental reports of a record-breaking T_c of 250–260 K (REFS[11,12]). Systematizing numerous predictions of T_c values for metal hydrides, a strong relationship between T_c and the position of the metal in the periodic table has been observed (FIG. 4): maximum T_c values correspond to hydrides of metals with low-lying empty orbitals (such as d^0 elements like Ca, La, Ac and T_H or d^1 elements like Sc and Y)[167]. Orbital populations for such elements are very sensitive to the local atomic environment, which can lead to strong electron–phonon coupling. This simple principle can guide further searches for high-T_c superconductors.

Electrides. Electrides are a class of unconventional materials with a strong accumulation of valence electron density in void spaces, playing the role of anions. These interstitial electrons, not belonging to any particular atom or bond, correspond to states close to the Fermi level, and because they are so weakly bound, a dramatically reduced work function can be expected. This makes them interesting for applications such as the splitting of carbon dioxide at room temperature and the synthesis of ammonia from atmospheric nitrogen under mild conditions[170]. Although organic electrides were discovered first[171,172], more recently, interest has shifted to inorganic materials owing to their potentially higher thermal and chemical stability. The first electride stable at room temperature was achieved experimentally by removing one oxygen atom per formula unit of the precursor compound, cement-phase mayenite $12CaO{\cdot}7Al_2O_3$ (REF.[173]). Now, computational searches have begun to take the lead. The predicted and then synthesized transparent high-pressure form of sodium[174] and the novel helium compound Na_2He (REF.[31]), as well as some other predicted high-pressure phases[175–177], are electrides. The band narrowing that accompanies electride formation has also been predicted to promote a magnetic state in dense potassium[175]. Initially driven by structural analogies, a number of layered electrides, including alkaline earth nitrides $(Sr, Ba)_2N$ and transition-metal or rare-earth carbides $(Y, Tb, Dy, Ho)_2C$, isostructural to the experimentally known electride Ca_2N, were predicted using substitutional approaches[178]. CSP has been applied to search for new electrides of A_2B and AB stoichiometries on the basis of either energy[179] or electronic localization[180]. However, such searches must consider possible competing phases with different stoichiometries[181]. Indeed, a variable-composition search found two new stable electrides of composition Sr_8P_5 and Sr_5P_3 (as opposed to the earlier[179] proposal of Sr_2P) as the ground state at ambient conditions, and Sr_5P_3 was confirmed by experiment[181].

Organic materials. The fields of inorganic and organic CSP are beginning to converge[182–185]. Computational limitations usually make it impractical to predict structures of organic crystals entirely from first principles, as they often contain hundreds of atoms in the unit cell. The most advanced strategy involves two stages: first, high-level interatomic force fields are employed for a first stage of structural optimization and energy evaluation and then more accurate (free) energy evaluation is used for the short list of candidates at the DFT or post-DFT levels. In practice, structure prediction could be simplified further on the basis of the observation that most organic crystals have either one or a fraction of a molecule in the asymmetric unit ($Z' \leq 1$) and in one of a few space groups such as $P2_1/c$, $P\text{-}1$, $P2_12_12_1$, $P2_1$, $C2/c$ and $Pbca$[183]. A strategy for sampling crystal structures in a reduced configuration space (with only a certain number of space groups and small Z') is powerful for solving crystal structures of molecules in real-life settings, as demonstrated in previous blind tests of organic CSP organized by the Cambridge Crystallographic Data Centre[184]. As a similar example, glycine, the simplest amino acid, is known to have six polymorphs; the structure of the short-lived metastable ζ-phase could not be solved for more than a decade and was eventually determined via evolutionary CSP[186].

A computational screening of the pharmaceutical compound dalcetrapib with ten torsional degrees of freedom led to the discovery of a new form that was successfully synthesized under high pressure[187]. However, the assumption that $Z \leq 1$ may be too restrictive, especially if metastable polymorphs are of interest, as reported in recent works[188–192]. Coumarin, a rather simple molecule, turns out to have five polymorphs, one with three molecules in the asymmetric unit[190]. The structures of the four new coumarin polymorphs were determined computationally and confirmed experimentally[190]. Polymorphism is an emerging design strategy for organic functional materials, and structure determination for metastable polymorphs might gain impetus in

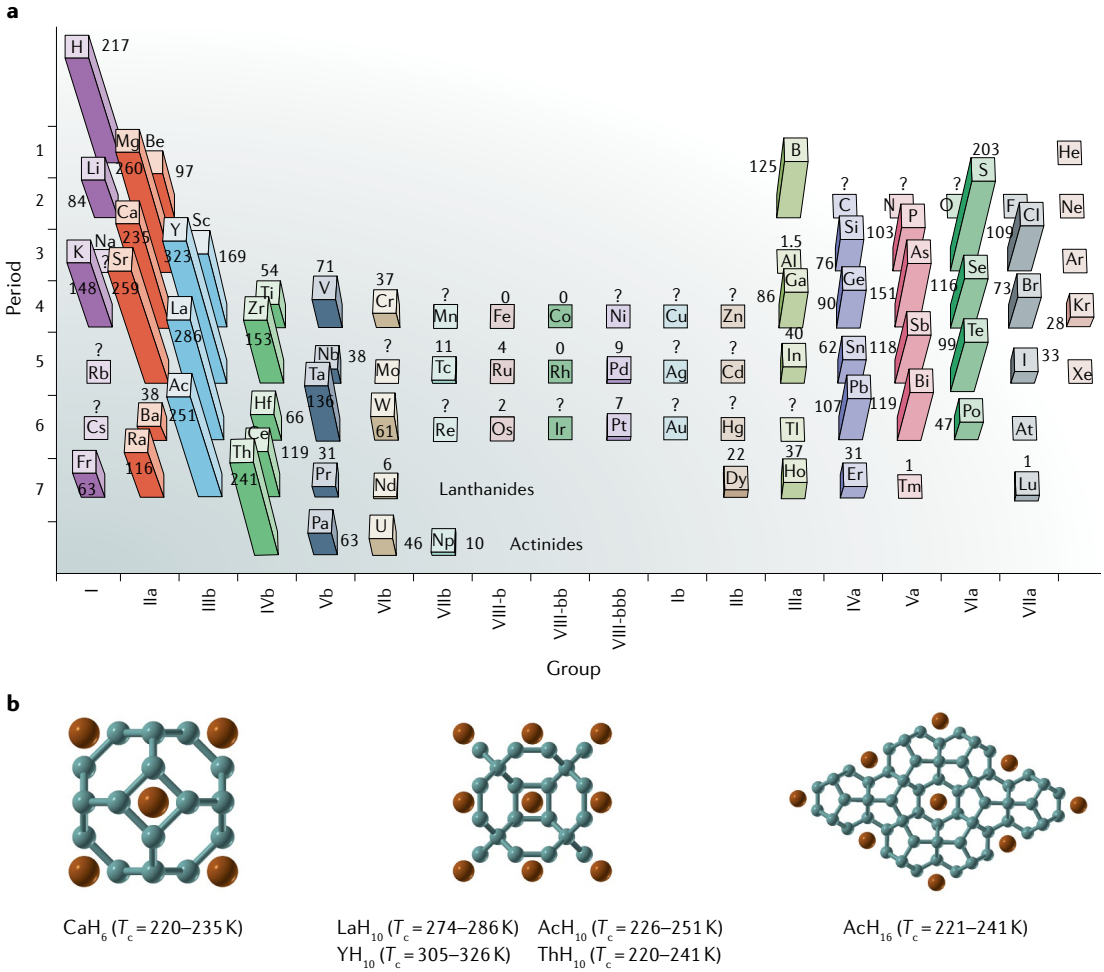


Fig. 4 | **Superconducting materials. a** | Compilation of the computed values of superconducting critical temperature, T_c, of metal hydride superconductors, showing two pronounced belts of high-T_c superconductor-forming metals[167]. For elements marked with a question mark, we do not have data. **b** | Structures of some representative metal hydrides that are predicted to be high-temperature superconductors. Panel **a** is adapted with permission from Semenok, D. et al. Actinium hydrides AcH_{10}, AcH_{12}, AcH_{16} as high-temperature conventional superconductors. *J. Phys. Chem. Lett.* **9**, 1920–1926 (REF.[167]), copyright 2018 American Chemical Society.

the near future. CSP has spread to organic semiconductors, a special class of organic materials with delocalized molecular orbitals in which charge carriers are mobile owing to extended π-conjugation. Organic semiconductors are of considerable interest because they promise fully flexible devices for large-area displays, solid-state lighting and solar cells. Great efforts have been devoted to improving their mobility and stability. In a computational screening, searching for new compounds derivable from the existing molecules, researchers found a new compound with a significantly improved mobility (12.3–16.0 $cm^2 V^{-1} s^{-1}$, whereas mobility in other organic materials rarely exceeds 10 $cm^2 V^{-1} s^{-1}$) with respect to the parent molecules[193]. With the advances in both computational power and CSP methodology, systematic CSP studies of energy-structure-function maps for small hypothetical molecules are now possible[194–196].

A very interesting adaptation of the ideas of evolution was employed to search for porous organic materials with desired shape and size of pores[197]. The researchers started with a library of known self-assembling monomers and a set of simple framework topologies and by applying crossover and conservative mutations to the fittest trial solutions obtained an efficient evolutionary optimizer. They were able to recover experimentally known structures and predict a number of monomers that should self-assemble to form porous materials with very large (16 Å) pores. This work shows how to explore the space of synthesizable organic compounds to predict materials with desired properties.

Photovoltaic materials. Converting solar energy directly into electricity efficiently remains an important long-term goal. Several classes of photovoltaic material are being actively investigated, including single-crystal, amorphous and polycrystalline silicon, III–V compounds, thin-film chalcogenides and organic photovoltaics. Silicon is currently the leading material in the solar cell industry. However, the common cubic diamond form of silicon has an indirect bandgap of 1.17 eV and a minimum direct gap of 3.4 eV, which hinders efficient absorption of solar radiation. There have been intense

efforts to search for materials to replace silicon[82]. An inverse band structure approach for predicting metastable silicon phases was used to find a cubic form with a quasi-direct gap of 1.55 eV. Several more stable candidates were proposed using different structure prediction techniques and are awaiting experimental validation[198–200]. An open-framework allotrope of silicon with a direct bandgap was recently realized experimentally using a two-step synthesis[201]. Furthermore, a low-energy tetragonal T32 structure of silicon with a quasi-direct bandgap of 1.28 eV was predicted to be synthesizable by decompressing the high-pressure Si-II allotrope on the basis of a newly developed evolutionary metadynamics technique[202]. Simultaneously, joint theoretical and experimental work[203] found experimental evidence for T32-Si. Practical applications of T32-Si will be likely if it can be obtained in large quantities.

Conclusions and perspectives

The many examples of successful first-principles structure predictions in the past decade make it clear that these techniques have become central to the study of materials structures and are beginning to emerge as an essential tool in materials discovery (see REF.[3] and, for a more personal perspective, REF.[185]). The computational tools for such work are widely available and are starting to be used directly by experimental groups to design and interpret their experiments.

The approaches discussed in this Review complement data mining, and fruitful combinations of the two are possible. The simplest idea is to use known structure types to create a starting population of crystal structures for an evolutionary search[204]. A more general approach[205] is to start with ideal nets (which are obtained from actual crystal structures by abstraction: disregarding differences between chemical species, removing 1-coordinate and 2-coordinate sites and maximizing symmetry); the 2,500 most common ideal nets describe >70% of all known inorganic structures. De-abstracting ideal nets using group–subgroup relations in a way similar to Bärnighausen trees can generate most of the known inorganic structures and an infinite number of reasonable new structures. The use of this topological structure generator for creating the initial population was shown[205] to significantly (in available tests, by a factor of ~3) accelerate evolutionary structure prediction. Global optimization is capable of discovering entirely new crystal structures and compounds, including those that cannot be related to the initial set of ideal nets. In many cases, such newly predicted compounds have subsequently been confirmed by experiments and can themselves be fed into databases. Knowing about the existence of such hidden ground states is important for understanding fundamental chemistry and may be of technological interest if such states can be prepared.

Structure prediction faces several challenges, including the difficulty of handling large systems, disorder and temperature. There are promising developments that can address each of these, such as building complex structural hierarchies[206], considering ensembles of structures[207] and using efficient and accurate techniques for handling vibrations and temperature[208,209].

The quality of the predictions obtained using the various electronic structure methods is limited by the available computational resources, insights into the chemistry of the system of interest and the skill of the person working on the computation. We expect that the need to describe increasingly complex materials, such as materials with strong electron–electron or electron–phonon interactions and magnetism, will become more and more important.

Whereas pressure is an easy thermodynamic parameter, temperature is not: its proper account requires sampling of typically 10^3–10^7 configurations for each phase, making calculations of the free energy computationally expensive. Fortunately, the number C of local minima of the free energy rapidly decreases with temperature T, possibly by exponential law:

$$C = C_0 exp\left(\alpha \frac{T^* - T}{T^*}\right) \quad (2)$$

where C_0 is the number of local minima at 0 K, α is a constant, and T^* is the characteristic temperature (higher than the melting temperature) at which only one free energy minimum exists[185]. The meaning of equation 2 is that as temperature increases, many neighbouring minima of the potential energy merge into one broad free energy minimum. The most accurate (fully incorporating anharmonic effects) way to compute the vibrational free energy and entropy is molecular dynamics with thermodynamic integration[210]. To study configurational disorder at finite temperatures (for example, in alloys), one can use Monte Carlo simulations with energies evaluated using a cluster expansion effective Hamiltonian[211,212]; this approach is traditionally used for modelling order–disorder phenomena when the underlying structure is known. Although the need for extensive sampling and the variety of physical causes of the entropy greatly complicate finite-temperature structure prediction, equation 2 makes us think that progress can be made soon.

Another area in which we expect rapid progress is the understanding and prediction of metastability. In principle, an infinite number of low-energy metastable crystal structures can be predicted, but only a handful (if any) are synthesizable. Selection rules for synthesizable polymorphs are currently unknown. Sun's hypothesis[29] reduces metastability to thermodynamic stability at some conditions, and this gives a practical recipe to judge about possible synthesizability. The problem can also be approached from a kinetics viewpoint, assuming that to be synthesizable a metastable phase should be easy to assemble from stable building blocks, such as magic nanoclusters[86,89].

The search for materials with optimal properties has so far focused on physical properties, which describe a particular state and often are response functions. Chemical properties are more difficult, as they describe not a single state but a process of changing the state, and their calculation requires extensive and expensive sampling. More accurate reactive force fields (including machine learning force fields), the development of more efficient and automatic sampling methods and growing computing power will allow progress in this area.

We are witnessing not only an avalanche development of new ideas and methods; new types of computing resources are also becoming available. Many ab initio codes are already ported to graphics processing units, and in some cases, this offers big advantages. Virtual machine technology enables almost any architecture to be used; for example, Windows PCs or even cell phones have been used to run such codes (normally in Linux)[213]. This allows distributed computing to be used for high-throughput materials discovery[214]. One day, perhaps quantum computers may also become practically applicable to both the optimization problem and computation of the energies and properties[215].

Published online 4 April 2019

1. Oganov, A. R. (ed.) *Modern Methods of Crystal Structure Prediction* (John Wiley & Sons, 2011).
2. Atahan-Evrenk, S. & Aspuru-Guzik, A. *Topics in Current Chemistry* Vol. 345 (Springer, 2014).
3. Oganov, A. R., Saleh, G. & Kvashnin, A. G. *Computational Materials Discovery* (Royal Society of Chemistry, 2018).
4. Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. The inorganic crystal structure data base. *J. Chem. Inf. Comput. Sci.* **23**, 66–69 (1983).
5. Villars, P. et al. The Pauling file, binaries edition. *J. Alloys Compd.* **367**, 293–297 (2004).
6. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
7. Nosengo, N. Can artificial intelligence create the next wonder material? *Nature* **533**, 22–25 (2016).
8. Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1**, 15004 (2016).
9. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
10. Liu, H., Naumov, I. I., Hoffmann, R., Ashcroft, N. W. & Hemley, R. J. Potential high-Tc superconducting lanthanum and yttrium hydrides at high pressure. *Proc. Natl Acad. Sci. USA* **114**, 6990–6995 (2017).
11. Drozdov, A. P. et al. Superconductivity at 250 K in lanthanum hydride at high pressures. Preprint at *arXiv* https://arxiv.org/abs/1812.01561 (2018).
12. Somayazulu, M. et al. Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures. *Phys. Rev. Lett.* **122**, 027001 (2019).
13. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge Univ. Press, 2004).
14. Wales, D. J. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge Univ. Press, 2004).
15. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).
16. Martiniani, S., Schrenk, K. J., Stevenson, J. D., Wales, D. J. & Frenkel, D. Structural analysis of high-dimensional basins of attraction. *Phys. Rev. E* **94**, 031301 (2016).
17. Stevanovic, V. Sampling polymorphs of ionic solids using random superlattices. *Phys. Rev. Lett.* **116**, 075503 (2016).
18. Oganov, A. R. & Valle, M. How to quantify energy landscapes of solids. *J. Chem. Phys.* **130**, 104504 (2009).
19. Ceriotti, M., Tribello, G. A. & Parrinello, M. Demonstrating the transferability and the descriptive power of sketch-map. *J. Chem. Theory Comput.* **9**, 1521–1532 (2013).
20. Pettifor, D. A chemical scale for crystal-structure maps. *Solid State Commun.* **51**, 31–34 (1984).
21. Villars, P. A. A three-dimensional structural stability diagram for 1011 binary AB_2 intermetallic compounds: II. *J. Alloys Compd.* **99**, 33 (1984).
22. Isayev, O. et al. Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* **27**, 735–743 (2015).
23. Kitaigorodsky, A. I. The close-packing of molecules in crystals of organic compounds. *J. Phys.* **9**, 351–352 (1945).
24. Nowacki, W. Symmetrie und physikalisch-chemische Eigenschaften krystallisierter Verbindungen. I. Die Verteilung der Kristallstrukturen über die 219 Raumgruppen. *Helv. Chim. Acta* **25**, 863–878 (1942).
25. Baur, W. & Kassner, D. The perils of *Cc*: comparing the frequencies of falsely assigned space groups with their general population. *Acta Cryst. B* **48**, 356–369 (1992).
26. Urusov, V. S. & Nadezhina, T. N. Frequency distribution and selection of space groups in inorganic crystal chemistry. *J. Struct. Chem.* **50**, 22–37 (2009).
27. Pauling, L. The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.* **51**, 1010–1026 (1929).
28. Villars, P. & Iwata, S. Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number. *Chem. Met. Alloys* **6**, 81–108 (2013).
29. Sun, W. et al. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).
30. Zhang, W. et al. Unexpected stable stoichiometries of sodium chlorides. *Science* **342**, 1502–1505 (2013).
31. Dong, X. et al. A stable compound of helium and sodium at high pressure. *Nat. Chem.* **9**, 440–445 (2017).
32. Niu, H., Oganov, A. R., Chen, X.-Q. & Li, D. Prediction of novel stable compounds in the Mg-Si-O system under exoplanet pressures. *Sci. Rep.* **5**, 18347 (2015).
33. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
34. Valle, M. & Oganov, A. R. Crystal fingerprint space – a novel paradigm for studying crystal-structure sets. *Acta Cryst. A* **66**, 507–517 (2010).
35. Stillinger, F. H. Exponential multiplicity of inherent structures. *Phys. Rev. E* **59**, 48 (1999).
36. Freeman, C., Newsam, J., Levine, S. & Catlow, C. R. A. Inorganic crystal structure prediction using simplified potentials and experimental unit cells: application to the polymorphs of titanium dioxide. *J. Mater. Chem.* **3**, 531–535 (1993).
37. Schmidt, M. U. & Englert, U. Prediction of crystal structures. *J. Chem. Soc.* **1996**, 2077–2082 (1996).
38. Pickard, C. J. & Needs, R. J. High-pressure phases of silane. *Phys. Rev. Lett.* **97**, 045504 (2006).
39. Oganov, A. R., Lyakhov, A. O. & Valle, M. How evolutionary crystal structure prediction works - and why. *Acc. Chem. Res.* **44**, 227–237 (2011).
40. Deaven, D. M. & Ho, K.-M. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **75**, 288 (1995).
41. Call, S. T., Zubarev, D. Y. & Boldyrev, A. I. Global minimum structure searches via particle swarm optimization. *J. Comput. Chem.* **28**, 1177–1186 (2007).
42. Wang, Y., Lv, J., Zhu, L. & Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **82**, 094116 (2010).
43. Lonie, D. C. & Zurek, E. Xtalopt: an open-source evolutionary algorithm for crystal structure prediction. *Comp. Phys. Comm.* **182**, 372–387 (2011).
44. 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. Matter* **25**, 495401 (2013).
45. Martonak, R., Laio, A. & Parrinello, M. Predicting crystal structures: the Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
46. Goedecker, S. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **120**, 9911 (2004).
47. Amsler, M. & Goedecker, S. Crystal structure prediction using the minima hopping method. *J. Chem. Phys.* **133**, 224104 (2010).
48. Zhou, X.-F., Oganov, A. R., Qian, G.-R. & Zhu, Q. First-principles determination of the structure of magnesium borohydride. *Phys. Rev. Lett.* **109**, 245503 (2012).
49. Meredig, B. & Wolverton, C. A hybrid computational–experimental approach for automated crystal structure solution. *Nat. Mater.* **12**, 123–127 (2013).
50. Fortes, A. D., Suard, E., Lemee-Cailleau, M.-H., Pickard, C. J. & Needs, R. J. Crystal structure of ammonia monohydrate phase II. *J. Am. Chem. Soc.* **131**, 13508 (2009).
51. Naslain, R. & Kasper, J. S. The crystal structure of the phi phase in the boron-sodium system. *J. Solid State Chem.* **1**, 150–151 (1970).
52. Albert, B. A new old: sodium boride: Linked pentagonal bipyramids and octahedra in Na_2B_{20}. *Angew. Chem. Int. Ed.* **37**, 1117–1118 (1998).
53. He, X.-L. et al. Predicting the ground-state structure of sodium boride. *Phys. Rev. B* **97**, 100102 (2018).
54. Li, Y.-F. & Selloni, A. Mosaic texture and double c-axis periodicity of β-NiOOH: insights from first-principles and genetic algorithm calculations. *J. Chem. Phys. Lett.* **5**, 3981–3985 (2014).
55. Zakaryan, H. A., Kvashnin, A. G. & Oganov, A. R. Stable reconstruction of the (110) surface and its role in pseudocapacitance of rutile-like RuO_2. *Sci. Rep.* **7**, 10357 (2017).
56. Morris, A. J., Grey, C. & Pickard, C. J. Thermodynamically stable lithium silicides and germanides from density functional theory calculations. *Phys. Rev. B* **90**, 054111 (2014).
57. Jung, H. et al. Elucidation of the local and long-range structural changes that occur in germanium anodes in lithium-ion batteries. *Chem. Mater.* **27**, 1031–1041 (2015).
58. Filinchuk, Y. et al. Porous and dense magnesium boro-hydride frameworks: synthesis, stability, and reversible absorption of guest species. *Angew. Chem. Int. Ed.* **50**, 11162–11166 (2011).
59. Zeng, Z. et al. A novel phase of $Li_{15}Si_4$ synthesized under pressure. *Adv. Eng. Mater.* **5**, 1500214 (2015).
60. Akahama, Y., Mizuki, Y., Nakano, S., Hirao, N. & Ohishi, Y. Raman scattering and X-ray diffraction studies on phase III of solid hydrogen. *J. Phys. Conf. Ser.* **950**, 042060 (2017).
61. Howie, R. T., Dalladay-Simpson, P. & Gregoryanz, E. Raman spectroscopy of hot hydrogen above 200 GPa. *Nat. Mater.* **14**, 495–499 (2015).
62. Akahama, Y. et al. Evidence from X-ray diffraction of orientational ordering in phase III of solid hydrogen at pressures up to 183 GPa. *Phys. Rev. B* **82**, 060101 (2010).
63. Dalladay-Simpson, P., Howie, R. T. & Gregoryanz, E. Evidence for a new phase of dense hydrogen above 325 Gigapascals. *Nature* **529**, 63–67 (2016).
64. Pickard, C. J. & Needs, R. J. Structure of phase III of solid hydrogen. *Nat. Phys.* **3**, 473–476 (2007).
65. Monserrat, B., Needs, R. J., Gregoryanz, E. & Pickard, C. J. Hexagonal structure of phase III of solid hydrogen. *Phys. Rev. B* **94**, 134101 (2016).
66. Monserrat, B. et al. Structure and metallicity of phase V of hydrogen. *Phys. Rev. Lett.* **120**, 255701 (2018).
67. Bartels-Rausch, T. et al. Ice structures, patterns, and processes: a view across the icefields. *Rev. Mod. Phys.* **84**, 885 (2012).
68. Falenty, A., Hansen, T. C. & Kuhs, W. F. Formation and properties of ice XVI obtained by emptying a type sII clathrate hydrate. *Nature* **516**, 231–233 (2014).
69. Pickard, C. J. & Needs, R. J. Highly compressed ammonia forms an ionic crystal. *Nat. Mater.* **7**, 775–779 (2008).
70. Ninet, S. et al. Experimental and theoretical evidence for an ionic crystal of ammonia at high pressure. *Phys. Rev. B* **89**, 174103 (2014).
71. Nakahata, I., Matsui, N., Akahama, Y. & Kawamura, H. Structural studies of solid methane at high pressures. *Chem. Phys. Lett.* **302**, 359–362 (1999).

72. Zhu, Q., Oganov, A. R., Glass, C. W. & Stokes, H. T. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Cryst. B* **68**, 215–226 (2012).
73. Maynard-Casely, H. et al. The distorted close-packed crystal structure of methane A. *J. Chem. Phys.* **133**, 064504 (2010).
74. Zhou, Z. F. & Harris, K. D. M. Design of a molecular quasicrystal. *ChemPhysChem* **7**, 1649–1653 (2006).
75. Hautier, G., Fischer, C., Ehrlacher, V., Jain, A. & Ceder, G. Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **50**, 656–663 (2010).
76. Davies, D. W. et al. Computer-aided design of metal chalcohalide semiconductors: from chemical composition to crystal structure. *Chem. Sci.* **9**, 1022–1030 (2018).
77. Allahyari, Z. & Oganov, A. R. Coevolutionary search for optimal materials in the space of all possible compounds. Preprint at *arXiv* https://arxiv.org/abs/1807.00854 (2018).
78. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
79. Lyakhov, A. O. & Oganov, A. R. Evolutionary search for superhard materials: methodology and applications to forms of carbon and TiO_2. *Phys. Rev. B* **84**, 092103 (2011).
80. Zhang, X. et al. First-principles structural design of superhard materials. *J. Chem. Phys.* **138**, 114101 (2013).
81. Zhu, Q., Oganov, A. R., Salvado, M. A., Pertierra, P. & Lyakhov, A. O. Denser than diamond: ab initio search for superdense carbon allotropes. *Phys. Rev. B* **83**, 193410 (2011).
82. Xiang, H., Huang, B., Kan, E., Wei, S.-H. & Gong, X. Towards direct-gap silicon phases by the inverse band structure design approach. *Phys. Rev. Lett.* **110**, 118702 (2013).
83. Nunez-Valdez, M., Allahyari, Z., Fan, T. & Oganov, A. R. Efficient technique for computational design of thermoelectric materials. *Comput. Phys. Comm.* **222**, 152–157 (2018).
84. Kvashnin, A. G., Oganov, A. R., Samtsevich, A. I. & Allahyari, Z. Computational search for novel hard chromium-based materials. *J. Phys. Chem. Lett.* **8**, 755–764 (2017).
85. Zhang, Y.-Y., Gao, W., Chen, S., Xiang, H. & Gong, X.-G. Inverse design of materials by multi-objective differential evolution. *Comput. Mater. Sci.* **98**, 51–55 (2015).
86. Yu, X.-H., Oganov, A. R., Zhu, Q., Qi, F. & Qian, G.-R. The stability and unexpected chemistry of oxide clusters. *Phys. Chem. Chem. Phys.* **20**, 30437–30444 (2018).
87. Lepeshkin, S. et al. Super-oxidation of silicon nanoclusters: magnetism and reactive oxygen species at the surface. *Nanoscale* **8**, 1816–1820 (2016).
88. Fubini, B. & Hubbard, A. Reactive oxygen species (ROS) and reactive nitrogen species (RNS) generation by silica in inflammation and fibrosis. *Free Radic. Biol. Med.* **34**, 1507–1516 (2003).
89. Lepeshkin, S. V., Baturin, V. S., Yu. Uspenskii, A. & Oganov, A. R. Method for simultaneous prediction of atomic structure and stability of nanoclusters in a wide area of compositions. *J. Phys. Chem. Lett.* **10**, 102–106 (2019).
90. Piazza, Z. A. et al. Planar hexagonal B_{36} as a potential basis for extended single-atom layer boron sheets. *Nat. Commun.* **5**, 3113 (2014).
91. Zhai, H.-J. et al. Observation of an all-boron fullerene. *Nat. Chem.* **6**, 727–731 (2014).
92. Ashton, M., Paul, J., Sinnott, S. B. & Hennig, R. G. Topology-scaling identification of layered solids and stable exfoliated 2D materials. *Phys. Rev. Lett.* **118**, 106101 (2017).
93. 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* **93**, 054117 (2016).
94. Zhou, X.-F. et al. Semimetallic two-dimensional boron allotrope with massless Dirac fermions. *Phys. Rev. Lett.* **112**, 085502 (2014).
95. Mannix, A. J. et al. Synthesis of borophenes: Anisotropic, two-dimensional boron polymorphs. *Science* **350**, 1513–1516 (2015).
96. Zhu, Z. et al. Multivalency-driven formation of Te-based monolayer materials: a combined first-principles and experimental study. *Phys. Rev. Lett.* **119**, 106101 (2017).
97. Chen, J., Schusteritsch, G., Pickard, C. J., Salzmann, C. G. & Michaelides, A. Two dimensional ice from first principles: Structures and phase transitions. *Phys. Rev. Lett.* **116**, 025501 (2016).
98. Chen, J., Schusteritsch, G., Pickard, C. J., Salzmann, C. G. & Michaelides, A. Double-layer ice from first principles. *Phys. Rev. B* **95**, 094121 (2017).
99. Corsetti, F., Zubeltzu, J. & Artacho, E. Enhanced configurational entropy in high-density nanoconfined bilayer ice. *Phys. Rev. Lett.* **116**, 085901 (2016).
100. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. 7×7 reconstruction on Si (111) resolved in real space. *Phys. Rev. Lett.* **50**, 120 (1983).
101. Zhu, Q., Li, L., Oganov, A. R. & Allen, P. B. Evolutionary method for predicting surface reconstructions with variable stoichiometry. *Phys. Rev. B* **87**, 195317 (2013).
102. Lu, S., Wang, Y., Liu, H., M.-S., Miao & Ma, Y. Self-assembled ultrathin nanotubes on diamond (100) surface. *Nat. Commun.* **5**, 3666 (2014).
103. Chuang, F., Ciobanu, C. V., Shenoy, V., Wang, C.-Z. & Ho, K.-M. Finding the reconstructions of semiconductor surfaces via a genetic algorithm. *Surf. Sci.* **573**, L375–L381 (2004).
104. Sierka, M. et al. Oxygen adsorption on Mo(112) surface studied by ab initio genetic algorithm and experiment. *J. Chem. Phys.* **126**, 234710 (2007).
105. Vilhelmsen, L. B. & Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J. Chem. Phys.* **141**, 044711 (2014).
106. Wang, Q., Oganov, A. R., Zhu, Q. & Zhou, X.-F. New reconstructions of the (110) surface of rutile TiO_2 predicted by an evolutionary method. *Phys. Rev. Lett.* **113**, 266101 (2014).
107. Zhou, R., Qu, B., Li, D., Sun, X. & Zeng, X. C. Anatase (101) reconstructed surface with novel functionalities: Desired bandgap for visible light absorption and high chemical reactivity. *Adv. Func. Mater.* **28**, 1705529 (2018).
108. Chen, P., Xu, Y., Wang, N., Oganov, A. R. & Duan, W. Effects of ferroelectric polarization on surface phase diagram: evolutionary algorithm study of the $BaTiO_3$ (001) surface. *Phys. Rev. B* **92**, 085432 (2015).
109. Harmer, M. P. The phase behavior of interfaces. *Science* **332**, 182–183 (2011).
110. Von Alfthan, S., Haynes, P., Kaski, K. & Sutton, A. Are the structures of twist grain boundaries in silicon ordered at 0 K? *Phys. Rev. Lett.* **96**, 055505 (2006).
111. Frolov, T., Divinski, S., Asta, M. & Mishin, Y. Effect of interface phase transformations on diffusion and segregation in high-angle grain boundaries. *Phys. Rev. Lett.* **110**, 255502 (2013).
112. Frolov, T., Olmsted, D. L., Asta, M. & Mishin, Y. Structural phase transformations in metallic grain boundaries. *Nat. Commun.* **4**, 1899 (2013).
113. Schusteritsch, G. & Pickard, C. J. Predicting interface structures: from $SrTiO_3$ to graphene. *Phys. Rev. B.* **90**, 035424 (2014).
114. Zhu, Q., Samanta, A., Li, B., Rudd, R. E. & Frolov, T. Predicting phase behavior of grain boundaries with evolutionary search and machine learning. *Nat. Commun.* **9**, 467 (2018).
115. Frolov, T. et al. Grain boundary phases in bcc metals. *Nanoscale* **10**, 8253–8268 (2018).
116. Xiang, H., Da Silva, J. L., Branz, H. M. & Wei, S.-H. Understanding the clean interface between covalent Si and ionic Al_2O_3. *Phys. Rev. Lett.* **103**, 116101 (2009).
117. Chua, A. L.-S., Benedek, N. A., Chen, L., Finnis, M. W. & Sutton, A. P. A genetic algorithm for predicting the structures of interfaces in multicomponent systems. *Nat. Mater.* **9**, 418 (2010).
118. Zhao, X. et al. Interface structure prediction from first-principles. *J. Phys. Chem. C* **118**, 9524–9530 (2014).
119. Caviglia, A. et al. Electric field control of the $LaAlO_3$/$SrTiO_3$ interface ground state. *Nature* **456**, 624–627 (2008).
120. Weber, J. et al. Quantum computing with defects. *Proc. Natl Acad. Sci. USA* **107**, 8513–8518 (2010).
121. Coomer, B. J., Goss, J. P., Jones, R., Oberg, S. & Briddon, P. R. Identification of the tetra-interstitial in silicon. *J. Phys. Condens. Matter* **13**, L1–L7 (2001).
122. Humble, P. The structure and mechanism of formation of platelets in natural type Ia diamond. *Proc. R. Soc. A* **381**, 65–81 (1982).
123. Morris, A. J., Pickard, C. J. & Needs, R. J. Hydrogen/silicon complexes in silicon from computational searches. *Phys. Rev. B* **78**, 184102 (2008).
124. Morris, A. J., Pickard, C. J. & Needs, R. J. Hydrogen/nitrogen/oxygen defect complexes in silicon from computational searches. *Phys. Rev. B* **80**, 144112 (2009).
125. Mulroue, J., Morris, A. J. & Duffy, D. M. Ab initio study of intrinsic defects in zirconolite. *Phys. Rev. B* **84**, 094118 (2011).
126. Morris, A. J., Grey, C. P., Needs, R. J. & Pickard, C. J. Energetics of hydrogen/lithium complexes in silicon analyzed using the Maxwell construction. *Phys. Rev. B* **84**, 224106 (2011).
127. Kaczmarowski, A., Yang, S., Szlufarska, I. & Morgan, D. Genetic algorithm optimization of defect clusters in crystalline materials. *Comput. Mater. Sci.* **98**, 234–244 (2015).
128. Aust, R. & Drickamer, H. Carbon: a new crystalline phase. *Science* **140**, 817–819 (1963).
129. Utsumi, W. & Yagi, T. Light-transparent phase formed by room-temperature compression of graphite. *Science* **252**, 1542–1544 (1991).
130. Mao, W. L. et al. Bonding changes in compressed superhard graphite. *Science* **302**, 425–427 (2003).
131. Li, Q. et al. Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.* **102**, 175506 (2009).
132. Umemoto, K., Wentzcovitch, R. M., Saito, S. & Miyake, T. Body-centered tetragonal C_4: a viable sp^3 carbon allotrope. *Phys. Rev. Lett.* **104**, 125504 (2010).
133. Wang, J.-T., Chen, C. & Kawazoe, Y. Low-temperature phase transformation from graphite to sp^3 orthorhombic carbon. *Phys. Rev. Lett.* **106**, 075501 (2011).
134. Niu, H. et al. Families of superhard crystalline carbon allotropes constructed via cold compression of graphite and nanotubes. *Phys. Rev. Lett.* **108**, 135501 (2012).
135. Boulfelfel, S. E., Oganov, A. R. & Leoni, S. Understanding the nature of superhard graphite. *Sci. Rep.* **2**, 471 (2012).
136. Wang, Y., Panzik, J. E., Kiefer, B. & Lee, K. K. Crystal structure of graphite under room-temperature compression and decompression. *Sci. Rep.* **2**, 520 (2012).
137. Oganov, A. R. & Solozhenko, V.L. Boron: a hunt for superhard polymorphs. *J. Superhard Mater.* **31**, 285 (2009).
138. Oganov, A. R. et al. Ionic high-pressure form of elemental boron. *Nature* **457**, 863–867 (2009).
139. Solozhenko, V.L., Kurakevych, O. & Oganov, A. R. On the hardness of a new boron phase, orthorhombic γ-B_{28}. *J. Superhard Mater.* **30**, 428 (2008).
140. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V., Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **99**, 064114 (2019).
141. Niu, H. et al. Structure, bonding, and possible superhardness of CrB_4. *Phys. Rev. B* **85**, 144116 (2012).
142. Kvashnin, A. G. et al. New tungsten borides, their stability and outstanding mechanical properties. *J. Phys.Chem. Lett.* **9**, 3470–3477 (2018).
143. Kolmogorov, A. et al. New superconducting and semiconducting Fe-B compounds predicted with an ab initio evolutionary search. *Phys. Rev. Lett.* **105**, 217003 (2010).
144. Gou, H. et al. Discovery of a superhard iron tetraboride superconductor. *Phys. Rev. Lett.* **111**, 157002 (2013).
145. Zhang, M. et al. Hardness of FeB_4: density functional theory investigation. *J. Chem. Phys.* **140**, 174505 (2014).
146. Wang, Q. et al. Is orthorhombic iron tetraboride superhard? *J. Materiomics* **1**, 45–51 (2015).
147. Van Der Geest, A. & Kolmogorov, A. Stability of 41 metal–boron systems at 0 GPa and 30 GPa from first principles. *Calphad* **46**, 184–204 (2014).
148. Niu, H. et al. Variable-composition structural optimization and experimental verification of MnB_3 and MnB_4. *Phys. Chem. Chem. Phys.* **16**, 15866–15873 (2014).
149. Hu, X. et al. Atomic-scale observation and analysis of chemical ordering in M_2B_2 and M_5B_3 borides. *Acta Mater.* **149**, 274–284 (2018).
150. Yu, S., Zeng, Q., Oganov, A. R., Frapper, G. & Zhang, L. Phase stability, chemical bonding and mechanical properties of titanium nitrides: a first-principles study. *Phys. Chem. Chem. Phys.* **17**, 11763–11769 (2015).
151. Bhadram, V. S., Kim, D. Y. & Strobel, T. A. High-pressure synthesis and characterization of incompressible titanium per-nitride. *Chem. Mater.* **28**, 1616–1620 (2016).

152. Schilling, A., Cantoni, M., Guo, J. & Ott, H. Superconductivity above 130 K in the Hg–Ba–Ca–Cu–O system. *Nature* **363**, 56–58 (1993).
153. Monteverde, M. et al. High-pressure effects in fluorinated $HgBa_2Ca_2Cu_3O_{8+\delta}$. *Europhys. Lett.* **72**, 458–464 (2005).
154. Ashcroft, N. W. Metallic hydrogen: a high-temperature superconductor? *Phys. Rev. Lett.* **21**, 1748 (1968).
155. Ashcroft, N. W. Hydrogen dominant metallic alloys: high temperature superconductors? *Phys. Rev. Lett.* **92**, 187002 (2004).
156. Duan, D. et al. Pressure-induced metallization of dense $(H_2S)_2H_2$ with high-Tc superconductivity. *Sci. Rep.* **4**, 6968 (2014).
157. 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* **525**, 73–76 (2015).
158. Errea, I. et al. Quantum hydrogen-bond symmetrization in the superconducting hydrogen sulfide system. *Nature* **532**, 81–84 (2016).
159. Errea, I. et al. High-pressure hydrogen sulfide from first principles: a strongly anharmonic phonon-mediated superconductor. *Phys. Rev. Lett.* **114**, 157004 (2015).
160. Einaga, M. et al. Crystal structure of the superconducting phase of sulfur hydride. *Nat. Phys.* **12**, 835–838 (2016).
161. Goncharov, A. F. et al. Hydrogen sulfide at high pressure: change in stoichiometry. *Phys. Rev. B* **93**, 174105 (2016).
162. Li, Y. et al. Dissociation products and structures of solid H_2S at strong compression. *Phys. Rev. B* **93**, 020103 (2016).
163. Kruglov, I., Akashi, R., Yoshikawa, S., Oganov, A. R. & Esfahani Davari, M. M. Refined phase diagram of the H-S system with high-Tc superconductivity. *Phys. Rev. B* **96**, 220101 (2017).
164. Wang, H., Tse, J. S., Tanaka, K., Iitaka, T. & Ma, Y. Superconductive sodalite-like clathrate calcium hydride at high pressures. *Proc. Natl Acad. Sci. USA* **109**, 6463–6466 (2012).
165. Li, Y. et al. Pressure-stabilized superconductive yttrium hydrides. *Sci. Rep.* **5**, 9948 (2015).
166. Kvashnin, A. G., Semenok, D. V., Kruglov, I. A., Wrona, I. A. & Oganov, A. R. High-temperature superconductivity in a Th–H system under pressure conditions. *ACS Appl. Mater. Interfaces* **10**, 43809–43816 (2018).
167. Semenok, D., Kvashnin, A. G., Kruglov, I. A. & Oganov, A. R. Actinium hydrides AcH_{10}, AcH_{12}, and AcH_{16} as high-temperature conventional superconductors. *J. Phys. Chem. Lett.* **9**, 1920–1926 (2018).
168. Peng, F. et al. Hydrogen clathrate structures in rare earth hydrides at high pressures: possible route to room-temperature superconductivity. *Phys. Rev. Lett.* **119**, 107001 (2017).
169. Geballe, Z. M. et al. Synthesis and stability of lanthanum superhydrides. *Angew. Chem. Int. Ed.* **57**, 688–692 (2017).
170. Kitano, M. et al. Ammonia synthesis using a stable electride as an electron donor and reversible hydrogen store. *Nat. Chem.* **4**, 934–940 (2012).
171. Ellaboudy, A., Dye, J. L. & Smith, P. B. Cesium 18-crown-6 compounds. A crystalline ceside and a crystalline electride. *J. Am. Chem. Soc.* **105**, 6490–6491 (1983).
172. Dye, J. L. Electrides: Ionic salts with electrons as the anions. *Science* **247**, 663–668 (1990).
173. Matsuishi, S. et al. High-density electron anions in a nanoporous single crystal: $[Ca_{24}Al_{28}O_{64}]^{4+}(4e^-)$. *Science* **301**, 626–629 (2003).
174. Ma, Y. et al. Transparent dense sodium. *Nature* **458**, 182–185 (2009).
175. Pickard, C. J. & Needs, R. J. Predicted pressure-induced s-band ferromagnetism in alkali metals. *Phys. Rev. Lett.* **107**, 087201 (2011).
176. Pickard, C. J. & Needs, R. J. Aluminium at terapascal pressures. *Nat. Mater.* **9**, 624–627 (2010).
177. Miao, M.-S. & Hoffmann, R. High pressure electrides: a predictive chemical and physical theory. *Acc. Chem. Res.* **47**, 1311–1317 (2014).
178. Inoshita, T., Jeong, S., Hamada, N. & Hosono, H. Exploration for two-dimensional electrides via database screening and ab initio calculation. *Phys. Rev. X* **4**, 031023 (2014).
179. Ming, W., Yoon, M., Du, M.-H., Lee, K. & Kim, S. W. First-principles prediction of thermodynamically stable two-dimensional electrides. *J. Am. Chem. Soc.* **138**, 15336–15344 (2016).
180. Zhang, Y., Wang, H., Wang, Y., Zhang, L. & Ma, Y. Computer-assisted inverse design of inorganic electrides. *Phys. Rev. X* **7**, 011017 (2017).
181. Wang, J. et al. Exploration of stable strontium phosphide-based electrides: theoretical structure prediction and experimental validation. *J. Am. Chem. Soc.* **139**, 15668–15680 (2017).
182. Price, S. L. Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* **43**, 2098–2111 (2014).
183. Day, G. M. Current approaches to predicting molecular organic crystal structures. *Crystallogr. Rev.* **17**, 3–52 (2011).
184. Reilly, A. M. et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Cryst. B* **72**, 439–459 (2016).
185. Oganov, A. R. Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss.* **211**, 643–660 (2018).
186. Bull, C. L. et al. ζ-glycine: insight into the mechanism of a polymorphic phase transition. *IUCrJ* **4**, 569–574 (2017).
187. Neumann, M., Van De Streek, J., Fabbiani, F., Hidber, P. & Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat. Commun.* **6**, 7793 (2015).
188. Zhu, Q. et al. Resorcinol crystallization from the melt: a new ambient phase and new riddles. *J. Am. Chem. Soc.* **138**, 4881–4889 (2016).
189. Xu, W., Zhu, Q. & Hu, C. T. The structure of glycine dihydrate: implications for the crystallization of glycine from solution and its structure in outer space. *Angew. Chem.* **129**, 2030–2034 (2017).
190. Shtukenberg, A. G. et al. Powder diffraction and crystal structure prediction identify four new coumarin polymorphs. *Chem. Sci.* **8**, 4926–4940 (2017).
191. Shtukenberg, A. G. et al. The third ambient aspirin polymorph. *Cryst. Growth Des.* **17**, 3562–3566 (2017).
192. Yang, J. et al. DDT polymorphism and the lethality of crystal forms. *Angew. Chem. Int. Ed.* **56**, 10165–10169 (2017).
193. Sokolov, A. N. et al. From computational discovery to experimental characterization of a high hole mobility organic crystal. *Nat. Commun.* **2**, 437 (2011).
194. Campbell, J. E., Yang, J. & Day, G. M. Predicted energy–structure–function maps for the evaluation of small molecule organic semiconductors. *J. Mater. Chem. C* **5**, 7574–7584 (2017).
195. Yang, J. et al. Large–scale computational screening of molecular organic semiconductors using crystal structure prediction. *Chem. Mater.* **30**, 4361–4371 (2018).
196. Musil, F. et al. Machine learning for the structure–energy–property landscapes of molecular crystals. *Chem. Sci.* **9**, 1289–1300 (2018).
197. Berardo, E., Turcani, L., Miklitz, M. & Jelfs, K. E. An evolutionary algorithm for the discovery of porous organic cages. *Chem. Sci.* **9**, 8513–8527 (2018).
198. Wang, Q. et al. Direct band gap silicon allotropes. *Chem. Soc.* **136**, 9826–9829 (2014).
199. Mujica, A., Pickard, C. J. & Needs, R. J. Low-energy tetrahedral polymorphs of carbon, silicon, and germanium. *Phys. Rev. B* **91**, 214104 (2015).
200. Amsler, M., Botti, S., Marques, M. A., Lenosky, T. J. & Goedecker, S. Low-density silicon allotropes for photovoltaic applications. *Phys. Rev. B* **92**, 014101 (2015).
201. Kim, D. Y., Stefanoski, S., Kurakevych, O. O. & Strobel, T. A. Synthesis of an open-framework allotrope of silicon. *Nat. Mater.* **14**, 169–173 (2015).
202. Zhu, Q., Oganov, A. R., Lyakhov, A. O. & Yu, X. Generalized evolutionary metadynamics for sampling the energy landscapes and its applications. *Phys. Rev. B* **92**, 024106 (2015).
203. Rapp, L. et al. Experimental evidence of new tetragonal polymorphs of silicon formed through ultrafast laser-induced confined microexplosion. *Nat. Commun.* **6**, 7555 (2015).
204. Su, C. et al. Construction of crystal structure prototype database: methods and applications. *J. Phys. Condens. Matter* **29**, 165901 (2017).
205. Bushlanov, P. V., Blatov, V. A. & Oganov, A. R. Topology-based crystal structure generator. *Comput. Phys. Commun.* **236**, 1–7 (2019).
206. Ahnert, S. E., Grant, W. P. & Pickard, C. J. Revealing and exploiting hierarchical material structure through complex atomic networks. *NPJ Comput. Mater.* **3**, 35 (2017).
207. Moran, R. F. et al. Hunting for hydrogen: random structure searching and prediction of NMR parameters of hydrous wadsleyite. *Phys. Chem. Chem. Phys.* **18**, 10173–10181 (2016).
208. Monserrat, B., Drummond, N. D. & Needs, R. J. Anharmonic vibrational properties in periodic systems: energy, electron- phonon coupling, and stress. *Phys. Rev. B* **87**, 144302 (2013).
209. Souvatzis, P., Eriksson, O., Katsnelson, M. & Rudin, S. Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. *Phys. Rev. Lett.* **100**, 095901 (2008).
210. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford Univ. Press, 2017).
211. Fontaine, D. D. Configurational thermodynamics of solid solutions. *Solid State Phys.* **34**, 73–274 (1979).
212. Zarkevich, N. A. & Johnson, D. D. Reliable first-principles alloy thermodynamics via truncated cluster expansions. *Phys. Rev. Lett.* **92**, 255702 (2004).
213. National Centre of Competence in Research MARVEL. Download the Quantum Mobile Virtual Machine based on Ubuntu Linux with a collection of quantum simulation codes. *MARVEL* http://nccr-marvel.ch/en/news/communication/2017-12-download-the-quantum-mobile-virtual-machine-based-on-ubuntu-linux-with-a-collection-of-quantum-simulation-codes (2017).
214. Khrapov, N., Roizen, V., Posypkin, M., Samtsevich, A. & Oganov, A. R. Volunteer computing for computational materials design. *Lobachevskii J. Math.* **38**, 926–930 (2017).
215. Cao, Y. et al. Quantum chemistry in the age of quantum computing. Preprint at *arXiv* https://arxiv.org/abs/1812.09976 (2018).
216. Bitzek, E., Koskinen, P., Gahler, F., Moseler, M. & Gumbsch, P. Structural relaxation made simple. *Phys. Rev. Lett.* **97**, 170201 (2006).
217. Michalewicz, Z. & Fogel, D. B. *How to Solve It: Modern Heuristics* (Springer, 2013).
218. Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J. & Caignaert, V. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* **346**, 343–345 (1990).
219. Schon, J. C. & Jansen, M. First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization. *Angew. Chem. Int. Ed.* **35**, 1286–1304 (1996).
220. Wales, D. J. & Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
221. Judson, R. S., Jaeger, E. P., Treasurywala, A. M. & Peterson, M. L. Conformational searching methods for small molecules. II. *Genetic algorithm approach. J. Comput. Chem.* **14**, 1407–1414 (1993).
222. Bush, T., Catlow, C. R. A. & Battle, P. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.* **5**, 1269–1272 (1995).
223. Curtis, F. et al. GAtor: a first-principles genetic algorithm for molecular crystal structure prediction. *J. Chem. Theory Comput.* **14**, 2246–2264 (2018).
224. Zhu, Q., Oganov, A. R. & Lyakhov, A. O. Evolutionary metadynamics: a novel method to predict crystal structures. *CrystEngComm* **14**, 3596–3601 (2012).
225. Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016).
226. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
227. Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **13**, 17930–17955 (2011).
228. Bartok, A. P., Payne, M. C., Kondor, R. & Csanyi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
229. Behler, J., Martonak, R., Donadio, D. & Parrinello, M. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.* **100**, 185501 (2008).
230. Deringer, V. L., J. Pickard, C. & Csanyi, G. Data-driven learning of total and local energies in elemental boron. *Phys. Rev. Lett.* **120**, 156001 (2018).

231. Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
232. Zhao, X. et al. Exploring the structural complexity of intermetallic compounds by an adaptive genetic algorithm. *Phys. Rev. Lett.* **112**, 045502 (2014).
233. Sharma, V. et al. Rational design of all organic polymer dielectrics. *Nat. Commun.* **5**, 4845 (2014).
234. Nicholls, R. J. et al. Crystal structure of the ZrO phase at zirconium/zirconium oxide interfaces. *Adv. Eng. Mater.* **17**, 211–215 (2015).
235. Pickard, C. J., Salamat, A., Bojdys, M. J., Needs, R. J. & McMillan, P. F. Carbon nitride frameworks and dense crystalline polymorphs. *Phys. Rev. B* **94**, 094104 (2016).
236. Kruglov, I. A. et al. Uranium polyhydrides at moderate pressures: prediction, synthesis, and expected superconductivity. *Sci. Adv.* **4**, eaat9776 (2018).
237. Wang, Q., Oganov, A. R., Feya, O. D., Zhu, Q. & Ma, D. The unexpectedly rich reconstructions of rutile TiO_2(011)-(2×1) surface and the driving forces behind their formation: an ab initio evolutionary study. *Phys. Chem. Chem. Phys.* **18**, 19549–19556 (2016).
238. Schusteritsch, G., Hepplestone, S. P. & Pickard, C. J. First-principles structure determination of interface materials: the Ni_xInAs nickelides. *Phys. Rev. B* **92**, 054105 (2015).

Acknowledgements
A.R.O. thanks the Russian Science Foundation (grant 19-72-30043) for generous support of his research. Q.Z. is funded by the National Nuclear Security Administration under the Stewardship Science Academic Alliances Program through the Department of Energy Cooperative Agreement DE-NA0001982. C.J.P. is supported by the Royal Society through a Royal Society Wolfson Research Merit Award. R.J.N. is funded by the Engineering and Physical Sciences Research Council under grant EP/P034616/1.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

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

RELATED LINKS
AIRSS: https://www.mtg.msm.cam.ac.uk/Codes/AIRSS
CALYPSO: http://calypso.cn/
CrySPY: https://tomoki-yamashita.github.io/CrySPY
DMACRYS: http://www.chem.ucl.ac.uk/cposs/dmacrys/index.html
GASP: http://gasp.mse.ufl.edu/
GAtor: http://www.noamarom.com/software/
GRACE: https://www.avmatsim.eu/services/software/
MAISE: http://bingweb.binghamton.edu/~akolmogo/maise/
Molpak: https://sourceforge.net/projects/molpak/
UPack: http://www.crystal.chem.uu.nl/~vaneyck/upack.html
USPEX: http://uspex-team.org/
Xtalopt: http://xtalopt.github.io/