Structure prediction of crystals, surfaces and nanoparticles

Scott M. Woodley[1], Graeme M. Day[2] and R. Catlow[1,3]

[1]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

[2]Computational Systems Chemistry, School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK [3]School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, UK

We review the current techniques used in the prediction of crystal structures and their surfaces and of the structures of nanoparticles. The main classes of search algorithm and energy function are summarized, and we discuss the growing role of methods based on machine learning. We illustrate the current status of the field with examples taken from metallic, inorganic and organic systems.

This article is part of a discussion meeting issue 'Dynamic *in situ* microscopy relating structure and function'.

Review

One contribution of 10 to a discussion meeting issue 'Dynamic *in situ* microscopy relating structure and function'.

Subject Areas:
crystallography, nanotechnology

Keywords:
structure prediction, crystals, structural chemistry

Author for correspondence:
R. Catlow
e-mail: c.r.a.catlow@ucl.ac.uk

This paper is dedicated to the memory of Prof. Roy Johnston—a pioneer in structure modelling and prediction in nano-science.

1. Introduction

Structure modelling and prediction are among the perennial challenges in solid state, surface and nano-science. The field is advancing rapidly with the development of new and improved search algorithms, with more accurate energetic models and with the growth in the use of tools and techniques from machine learning; as the predictive capability advances, it is increasingly integrated with experiment. In this article, we chart the development of the field, giving emphasis to the rapid recent progress. We survey first the methodological approaches, then the achievements and challenges, giving examples and case studies; and we conclude by discussing the likely future developments. The illustrations we provide include functional oxides, microporous inorganic crystals, pharmaceutical poly-morphism, functional organic crystals and metal and oxide nanoparticles—systems with wide-ranging

applications, including energy generation and storage, in catalysis, in pharmaceutical science and in gas sorption and separation.

Structure prediction is, of course, a major challenge in other areas of contemporary science, such as molecular biology, and many of the methodological challenges are shared across different fields. For recent developments in the field in bio-sciences, we refer to the review of Kuhlman & Bradley [1]. We also note that our account does not attempt to be comprehensive; for other discussions of the field, we refer to the proceedings of the Faraday Discussion on the topic, especially the papers of Price [2] and Oganov [3] and the discussions in [4–7].

2. Searching the energy landscape

The fundamental challenge in structure prediction is the complexity of the configurational space, with both the coordinates of each particle and, in the case of crystals, the unit cell parameters representing, in principle, independent variables. As discussed by Woodley & Catlow [8], the general approach to this problem has been first to define some 'cost function', which provides a rapidly computable and simple figure of merit of the structure; and then to navigate, using a variety of approaches, the parameter landscape, searching for regions of low-cost function, which are expected to provide plausible possible structures. Having identified one or more such regions, standard minimization procedures are applied to generate the structure of minimum energy (or possibly free energy), with energies calculated using an interatomic potential model or a quantum mechanical technique, which for the latter case in recent work is generally based on density functional theory (DFT). In a system of any complexity, there will be several distinct *local* energy minima and it is hoped that the search is sufficiently comprehensive to identify the lowest, i.e. the *global* energy minimum. The global minimum in energy is assumed to be the most likely observable structure, although the kinetics of nucleation and growth can lead to crystallization of higher energy, metastable structures; the range of experimentally observed metastability has been surveyed in both organic [9,10] and inorganic [11,12] crystals through large-scale computational studies of known polymorphs.

With increases in available computer resources, structures with a larger number of degrees of freedom (number of atoms in a cluster or unit cell) can be targeted and/or more expensive cost functions employed. There is, of course, a trade-off between the chosen quality of the cost function, $E(\mathbf{r}_i)$, and the number of configurations that can be evaluated. Commonly, $\mathbf{r}_i$ are the atomic coordinates. Ideally, the cost function has basins that correspond to regions of interest and their depths match the order of the relative stabilities of the configurations associated with each basin; it may also be beneficial that the basins of interest have a larger width, making them easier to locate in the search. Typically, $E(\mathbf{r}_i)$ is an energy function and only approximate solutions on this landscape are initially required as these can be readily refined on a more accurate energy landscape, $E^{\mathrm{acc}}(\mathbf{r}_i)$, at a later stage of calculations. For the same computer resource, employing a Born model and interatomic potentials [13] as opposed to an electronic structure approach [14] would correspond to being able to evaluate several orders of magnitude more configurations, providing more confidence in identifying all important structures. As with all models, the terms in the Hamiltonian for $E(\mathbf{r}_i)$ should include only what is necessary to capture the important physics, such that, in this case, for each targeted energy minimum of $E^{\mathrm{acc}}(\mathbf{r}_i)$ there needs to be one energy minimum configuration of $E(\mathbf{r}_i)$ that is within the corresponding basin of $E^{\mathrm{acc}}(\mathbf{r}_i)$ and that there is a good match between the minima rankings. As reported below, some approaches apply brute force and search directly on $E^{\mathrm{acc}}(\mathbf{r}_i)$, whereas others try to first develop a suitable Hamiltonian for the system of interest [15–23].

The main conceptual and algorithmic approaches for navigating the energy landscape have been discussed in earlier reviews and book chapters [8,13,24–26], and are as follows.

Simulated annealing—the conceptual basis of which is simple: a Monte Carlo (MC) algorithm coupled with the Metropolis criterion or molecular dynamics (MD)—is used to generate a sequence of configurations that map out a path of a so-called *walker* across the energy landscape. A temperature schedule is followed to simulate the process of annealing: initially starting at a

sufficiently high temperature that the energy barriers between different regions of phase space may be overcome, and upon reducing the temperature low-energy regions may be identified. There are critical temperatures when the walker may become trapped in a high-energy basin. Sufficient sampling of the accessible landscape just before such temperatures is required if lower energy basins are targeted. Increasing the step size between consequent points on the landscape may reduce the number of required steps but is itself constrained as an MC step must not straddle barriers and an MD step requires the local landscape to be sufficiently approximated by a few terms of its Taylor expansion. Each approach is straightforward and robust, with MC sampling many more configurations owing to the application of the Metropolis criterion and MD requiring the tracking of momentum for each atom within the simulation box. However, the search inevitably has some bias from the choice of the initial configuration. Thus, simulations are run several times using different starting configurations. With the current availability of computer resources, the global minimum is no longer the sole final target, but instead the relative stabilities of phases (locally ergodic regions) based on free energies are sought as a function of temperature and pressure, as well as the prediction of their lifetimes [12]. MC and MD can also be employed to gain an insight into the size (number of configurations) of an energy basin and the height of the energy barrier to escape and can help determine the likelihood, or probability, of finding this route of escape; see, for example, the threshold [27,28] and energy lid [28,29] algorithms.

Basin hopping is another branch of global optimizers that effectively maps out the path of a walker across the energy landscape, but with the acceptance/rejection decisions based upon comparing energy minima [30–34]. Each attempted step on the landscape is either a mini-MD run or a standard MC step, followed by a full structural relaxation using a local optimization routine. *Monte Carlo basin hopping (MCBH)* has proved extremely popular for predicting the atomic structures of clusters. The energies of local minima (or possibly other stationary points unless the chosen local optimizer ensures only local minima are found) are typically compared within the Metropolis criterion. If the chosen method generates new trial configurations based only on perturbing optimized configurations, the MCBH step size will need to be much larger than a typical MC step and is adjusted on the fly to ensure a required average success rate of escape from the current basin. Alternatively, a standard MC step size is employed with acquired values of local energy minima fed into the Metropolis criterion: if a new configuration is accepted then the current configuration is replaced by the new non-relaxed configuration and the current value of energy is replaced by the energy minimum found from relaxing the newly accepted configuration. In the latter, the walker explores a landscape of plateaux with many fewer energy barriers than the original continuous energy landscape. It is easy to see why MCBH readily finds the bottom of a superbasin on the energy landscape as it is a simple basin on the landscape of plateaux. If the landscape of plateaux also contains many basins, then this approach will also inevitably have some bias from the choice of the initial configuration. Finally, in the case of *minima hopping* that employs MD, the kinetic energy and run time for each mini-MD run can be dynamically adjusted to achieve a desired average success rate of escaping the current basin and finding a new local minimum structure.

Genetic algorithm techniques [22,24,35] are perhaps the best known of the methods that employ interacting walkers [36], or simultaneously consider more than one point on the energy landscape to find the global minimum, followed by the methods of *particle swarm* [37,38] and *taboo* [39] algorithms. Genetic algorithms or, more generally, evolutionary algorithms start with the creation of a population; typically, a set of random points (e.g. randomly chosen atomic coordinates for each configuration) on the energy landscape. Then, competition is simulated between current members of the population for survival and the chance to procreate. The probability of success in a competition is biased towards selecting lower energy configurations. Procreation, i.e. the creation of new configurations, or children, is simulated by the combining of information (e.g. structural fragments) taken from two or more of the configurations that won in a competition to become a parent. New configurations composed of better/worse features of their parents will typically go on to survive/die, respectively. As with other global optimization schemes, there are many

variants, which is particularly true for evolutionary algorithms. Geometric constraints are often applied during the creation of new configurations, e.g. unphysically short interatomic distances are prevented. The benefits seen for MC techniques of basin hopping are also sought in the design of evolutionary algorithms, with populations being composed of fully relaxed configurations (energy minima). The success of an evolutionary algorithm in locating the global minimum is dependent upon the diversity of the population (having configurations within different basins, or super-basins, as opposed to all within the same basin). Hence, the process of mutation (typically a small MC step) is also applied to new configurations, as well as other strategies like the removal of duplicates and niching [40–43] when determining the survival of configurations in the current population.

Topological procedures, which have a long history in crystallography going back to the early work of Wells [44,45], who developed topological approaches in exploring networks, can be used to rationalize and predict crystal structures. Such methods are most naturally applied to framework structures and have been used to considerable effect in the structural science of zeolites—framework-structured microporous alumino-silicates and silicas—where a powerful topological approach—tiling theory—was employed by Bell and co-workers [46] to enumerate possible network structures from which structures can then be generated by replacing nodes by linking tetrahedra. Lattice energy minimization using Born model interatomic potentials then allows low-energy structures to be identified. Other topological approaches in zeolite science are discussed by, for example, Treacy and co-workers [47,48]; also, as discussed by Arhangelskis *et al.* [49], the approach has been extended to the intensively studied family of metal–organic framework (MOF) structures in which framework structures are generated by linking inorganic structural motifs, for example MO_6 octahedra, with organic linkers.

Generation of random configurations, in which the global optimization, algorithmic approach to explore the energy landscape is replaced by the generation of large numbers of random initial configurations which are then subject to screening of unphysical structures, followed by energy minimization. This approach might initially seem to rely on 'brute force' computational power, as the trial structures are not explicitly directed towards low-energy regions of the energy landscape. However, potential energy surfaces tend to have features that mean that even randomly generated structures find low-energy structures preferentially over high-energy structures [50]. Of particular relevance is the observation that low-energy local minima often have larger basins of attraction— the region of configurational space from which local energy minimization leads to a particular minimum—than high-energy local minima [51]. More recent implementations [51,52] of such an approach have preferred the use of quasi-random, low-discrepancy sequences in place of random numbers; these are deterministic sequences that maintain some of the qualities of random sampling, while ensuring a more uniform sampling of configuration space. The method has enjoyed success in several cases [53–57] and for small, less complex structures a random approach may provide a sufficient sampling of configurational space. It also has the advantage that there is no in-built bias, which can be beneficial in identifying metastable structures, where the application of structure prediction is not only interested in the lowest energy structures and also is perfectly parallelizable because all structures are generated independently. But, as discussed by Woodley & Sokol [58] for larger more complex structures, the approach can become problematic as, even with large numbers of randomly generated configurations, significant regions of configurational space may remain unexplored.

Molecular packing prediction, which has mainly been applied to organic molecular crystals, has employed many of the same methods as inorganic structure prediction, including systematic and random methods [59], genetic algorithms [60–62], simulated annealing [63] and other MC methods [64]. The exception is that topological approaches are less useful for molecular crystals, because their packing is rarely dominated by a particular intermolecular interaction. Instead, the packing of organic molecules is usually determined by a balance of many weak, often competing, interactions. The nearest to a topological approach is that proposed by Gavezzotti [65], in which crystals were built in stages, starting from searching for stable clusters, which were extended

into translationally periodic structures through successive application of space group symmetry operations.

A key difference for molecular packing prediction is that the basic building blocks of generated structures are the constituent molecules, which carry more degrees of freedom than the atoms that are treated in non-molecular systems. Therefore, in addition to the position within the unit cell, the orientation of each molecule must be sampled. Molecular geometries can also be influenced by their environment in a crystal; in particular, exocyclic single bonds can usually rotate with little energetic cost, so that intermolecular interactions in each predicted crystal structure can alter the molecular geometry. This flexibility can be accounted for in several ways. The simplest approach is to generate crystal structures using rigid molecules, constrained to their isolated molecule 'gas-phase' geometry, only introducing molecular flexibility by allowing the molecular geometry to relax at the local energy minimization stage of the procedure. This approach can miss structures where molecular deformation is required for a certain packing, so can be augmented by searches starting from higher energy saddle points on the isolated molecule energy surface [66]. A more comprehensive approach requires sampling of selected intramolecular degrees of freedom during structure generation, alongside the crystal packing variables—the unit cell dimensions, molecular positions and orientations [67].

The most widely applied energy models for molecular crystals are based on accurate interatomic potentials for intermolecular interactions, in which accurate models of electrostatics have been proven to be crucial [68,69], and dispersion-corrected solid-state DFT. The accuracies of these methods have been benchmarked against a reliable set of measured sublimation enthalpies of molecular crystals [70–72]. Although the differences in accuracy between the best methods are small, these must be interpreted in the context of the very small energy differences between structures of molecular organic crystals, which are usually less than 2 kJ mol^{-1} [9]. There has been considerable interest in recent years in including the contributions of lattice vibrations to the free energies of molecular crystal structures [9,72–74], so that relative stabilities can be predicted as a function of temperature [10,75]. Both lattice dynamics and MD methods have been applied to free energy prediction, with the latter also providing information about which structures are likely to interconvert at a given temperature [74,76].

3. The role of machine learning

Machine learning methods have started to find applications across many areas of computational chemistry, which is also true in the area of structure prediction. Machine learning methods have been applied in attempts to improve the accuracy and speed of calculations, as well as adding insight into the analysis of structure prediction results.

The clearest area where machine learning methods have promise in structure prediction is to learn the relationship between structure and energy. As discussed earlier, a computationally efficient cost function is required in the early stages of searching an energy landscape, because of the large number of cost function evaluations or structural optimizations that are required. Thus, interatomic potentials are commonly applied during structure searches. However, the final evaluation of structures usually requires an accurate energy model, which sometimes means that quantum mechanics (QM)-based energy evaluations (such as using DFT) are necessary on large numbers of candidate structures. The computational expense of such calculations can limit applications to simple systems or restrict the methods to researchers with access to very large computing resources. Machine learning has been shown to be able to predict high-level QM energies, either directly from descriptors of atomic structure [77,78] or as a correction to lower level energy calculations [79].

This approach has been applied in crystal structure prediction studies of molecular crystals. As an example, starting from the structures of several pentacene derivatives predicted by interatomic potentials, Gaussian process regression (GPR) was used to learn the energies of these predicted structures at a more accurate and expensive level of theory—in this case, solid-state DFT [80]. DFT calculations were performed on a training subset of the predicted crystal structures. Using

the smooth overlap of atomic positions method [81,82] to quantify the similarities of atomic environments, GPR was then trained to learn the relationship between structure and the DFT energies, providing a model that can be quickly applied to predict the energies of the remaining structures. The results showed that DFT relative energies could be predicted to an accuracy of 1–2 kJ mol^{-1} using training set sizes of only 10%, although errors increased as the complexity of the molecule was increased [80]. The approach has also been applied in a multi-level machine learning model to predict expensive high-level energies of predicted crystal structures using an intermediate, lower cost method to reduce the number of expensive calculations required for training the model [83]. A similar method has been demonstrated using GPR to learn QM energies of fragments of predicted crystal structures, which are summed to give the total lattice energies [84]. The fragment-based method has the advantages of providing more data to the machine learning model, because each crystal structure provides multiple fragments, as well as allowing a wider range of QM methods, which are affordable on fragments, but not entire crystal structures. With the small errors achieved with these approaches, the ranking of predicted structures using the machine-learned energies was shown to reproduce the true QM-based ranking of predicted structures very well.

Deringer and co-workers [85,86] have demonstrated the potential of machine learning in inorganic crystal structure prediction, using GPR to learn interatomic potentials from DFT energies during the structure search. By learning the energies at the same time as searching the energy landscape, the machine-learned energy model can be continually updated and used to drive the exploration for new structures.

The successful application of machine learning models for modelling energies in structure prediction applications demonstrates that the structural descriptors used in these studies capture the features that are correlated with structural stability. This suggests that these descriptors could also form the basis of unsupervised learning approaches to investigate the structural diversity amongst structures. To investigate this, Musil *et al*. [80] applied dimensional reduction methods to project the predicted structures from their inherently high-dimensional space to a low-dimensional representation that can be visualized. The aim of these methods is to maintain as much information as possible on structural similarity in the low-dimensional representation. The results for the landscape of predicted crystal structures of pentacene are shown in figure 1*a*. The projection shows groupings of similar crystal structures, which correspond to different packing motifs. These were confirmed by applying clustering methods to the similarity matrix of structures, which identified seven clusters on the pentacene landscape (coloured in figure 1*a*). Identification of these families of structures would otherwise have required painstaking manual inspection or the development of case-specific heuristic rules, showing that unsupervised learning could be powerful in gaining insight into recurring structural features on a structural landscape. The approach was extended in work by Yang *et al.* [89] to the analysis of 28 isomers of a promising organic semiconductor, showing that the method could identify structural families across the crystal structure landscapes of a series of related molecules (figure 1*b*) and aiding in identifying structure–property relationships within these sets of predicted materials.

Another exciting area of interest for machine learning in structure prediction is the generation of structures themselves, using data-driven approaches in place of simulation-based exploration of the energy landscape. With many developments in generative models for molecules [90,91], this area is likely to move forward for crystals. A simple example is the prediction of possible new inorganic structures based on templates from known structures; Ryan *et al.* [92] trained a neural network model on a database of inorganic crystal structures to predict, based on atomic environment, which chemical compositions are likely to form certain, known structure types. Another recent example in this area was the development of a generative adversarial network trained on a large database of known zeolites that could generate new zeolite structures [93]. An attractive aspect of this work is that the generative model can be trained to produce new structures with targeted properties (e.g. methane heat of adsorption). This type of work could potentially be developed for more complex materials as an alternative to the simulation-based methods discussed above.

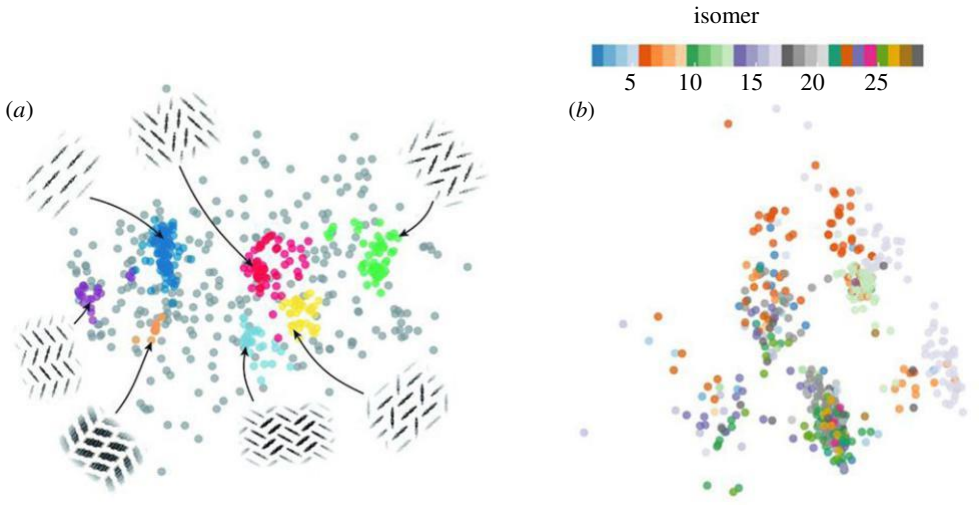

Figure 1. (*a*) Two-dimensional projection, using the sketch-map method [87], of the pentacene crystal structure landscape's similarity matrix [80], showing groupings of predicted crystal packings with similar structures. Each point corresponds to a distinct, low-energy crystal structure, coloured according to cluster analysis using the HDBSCAN* method [88] (grey points are not part of any cluster). Representative crystal packings are shown for each identified cluster. (*b*) Sketch-map representation of the low-energy crystal structure landscapes for 28 isomers of a planar pyrrole-based azaphenacene from screening for potential organic semiconductors [89]. Each point corresponds to a distinct, low-energy crystal structure, coloured according to the molecular isomer. Adapted from [80], published by the Royal Society of Chemistry, and reprinted with permission from [89]. (Online version in colour.)

4. Successes and challenges

We now illustrate the current state of the art in the field by several case studies taking, first, two examples from inorganic structural science, which we follow with an account of the challenges posed by defective oxide surfaces. Oxide and metal nanoparticle structures illustrate the successes and problems of structure prediction in nano-science, including recent work on nanostructures on oxide surfaces; in the final section, the focus shifts to molecular systems, in particular crystals of pharmaceutical molecules and organic porous solids.

(a) Structure prediction in inorganic crystallography

Functional materials chemistry increasingly explores systems of high complexity either in their composition or in their crystal architecture. Here we highlight two areas where structure prediction methods have proved of considerable value in exploring novel functional materials.

Our first example is taken from the work of Collins, Rosseinsky and co-workers [94] concerning the accelerated discovery of a new multicomponent oxide in a highly complex inorganic phase field. The approach is based on the intriguing use of structural motifs (e.g. MO_6 octahedra) rather than individual atoms as the fundamental building blocks of crystal architectures, thereby allowing chemical knowledge and intuition to be integrated into the process. An MC-based simulated annealing algorithm (employing the Extended Module Materials Assembly (EMMA) software) explores different arrangements of the motifs and the 'best' structure is identified. DFT-based energy minimization is then used to generate the predicted structure. A new, interesting and highly complex structure illustrated in figure 2 is generated in the Y-Sr-Ca-Ga-O phase field and the predicted structure was successfully synthesized, with the experimental structure (determined using powder X-ray diffraction) being

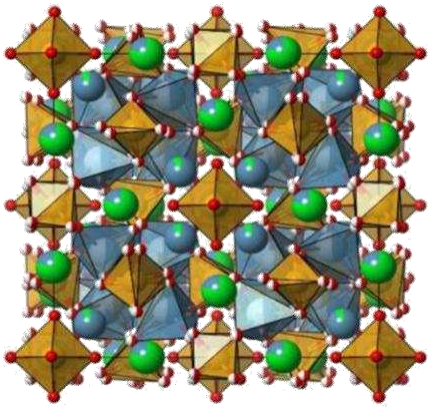

Figure 2. Predicted structure for a novel oxide in the Y-Sr-Ca-Ga-O phase field [94]. Atoms coloured as follows: Ga (brown), Sr (green), Ca (light blue) and O (red). (Online version in colour.)

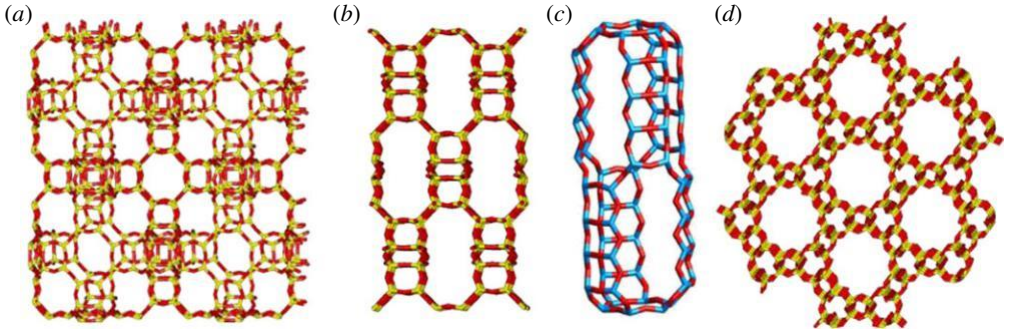

Figure 3. Some microporous structures predicted to be stable as 'silicas' [46,95,96].

close to that predicted, clearly demonstrating the predictive power of current methodologies for this class of material.

Our second case study is from the field of microporous materials, where, as discussed above, topological approaches have proved powerful. As noted, the work of Bell and co-workers [46,95,96] has combined an approach based on combinatorial tiling theory with lattice energy minimization to predict novel microporous structures. Figure 3 illustrates new structures which are found to have low energies as 'silicas' (i.e. structures of composition SiO_2) and which now present a challenge to synthesis. The locations of silicon atoms are typically referred to as T-sites, which are connected via bridging oxygens to form a network of corner-sharing SiO_4 tetrahedra. Interestingly, other structures, illustrated in figure 4*a*,*b*, which are also predicted to be stable as silicas, although not yet synthesized in that composition, are known in other quite different compositions, suggesting that the framework topology has intrinsic stability.

The third example shown in figure 4 illustrates a different mode of structure generation, in this case of a small-pore nano-porous structure, where direct bonding between T-sites greatly reduces the size of the cavities to create the metastable body-centred tetragonal (BCT) structure of ZnS. There are, however, also examples of crystal structure prediction of nano-porous ZnO [100] and SiC [101] that have larger cavities (figure 5) which are constructed from secondary building units [102]. The framework topologies of silicas, or more generally zeolites, have also been exploited in the prediction of MOFs where organic molecules form the bridging units and thus much larger cavities; see, for example, the work on zeolitic imidazolate frameworks where the T-sites contain Zn [103].

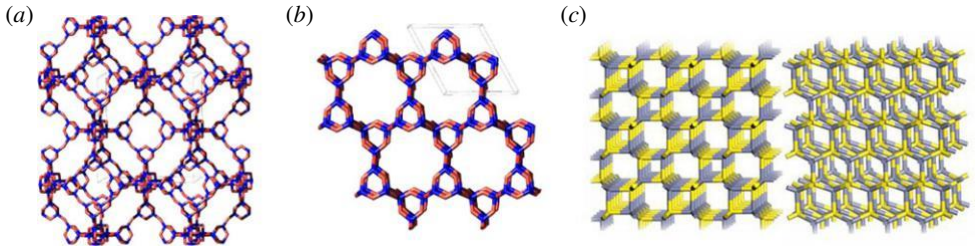

Figure 4. Predicted microporous structures stable in non-silica composition. Left, after Zheng *et al.* [97]; centre, after Correll *et al.* [98]; right, after Hamad *et al.* [99]—two views of the ZnS BCT phase. (Online version in colour.)

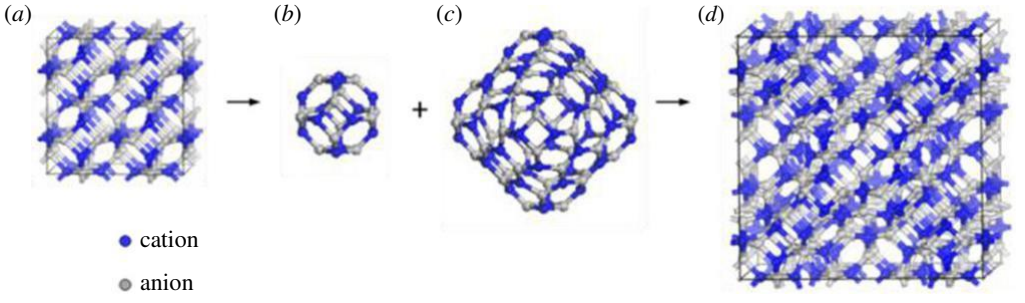

Figure 5. (*a*) Cubic unit cell of the silica sodalite framework represented as a nano-porous network of zinc oxide [100] or silicon carbide [101]; (*b*) structural unit composed of 24 atoms taken from (*a*); (*c*) larger structural unit composed of 96 atoms with the same symmetry, T_h , as that in (*b*); (*d*) predicted nano-porous framework (cubic unit cell) constructed from units shown in (*b*) and (*c*) and replacing T-sites with SiO_4 tetrahedra; this is possibly a new microporous framework. (Online version in colour.)

(b) Oxide surface structures

The surfaces of oxides, even those such as MgO and ZnO which have the simplest of ionic crystal structures, can show remarkable complexity. Surface rumpling, i.e. differential perpendicular displacements of surface ions, occurs even on perfect terraces, owing to differences in the polarizabilities of cations and anions. Moreover, in addition to the well-known surface features, such as steps and corners, surface point defects are commonly present in appreciable concentrations, in some cases due to low formation energies; a special case is provided by 'polar' surfaces, i.e. those which in the surface repeat unit have a dipole moment perpendicular to the surface. Tasker [104] first showed that such surfaces are intrinsically unstable in ionic materials, owing to a divergence in the electrostatic energy. For stability, the surface dipole must be quenched, which may be achieved possibly by electron transfer between bulk and surface regions or by adsorption of polar species but is most commonly effected by the creation of charged surface defects in sufficiently high concentrations to cancel the surface dipole.

The reconstruction of polar oxide surfaces has been widely studied and a particularly intriguing example is provided by ZnO. This wurtzite structured solid has four principal surfaces: two non-polar, i.e. the (1010) and (1120), and two polar, i.e. the zinc-terminated (0001) and the oxygen-terminated (0001). The non-polar surfaces have an interesting structure, as discussed by Whitmore *et al.* [105] and more recently by Mora Fonz *et al.* [106]. Zinc–oxygen vacancy pairs were shown in these computational studies to have low formation energies, leading to extensive surface grooving as observed experimentally. The polar surfaces, however, show remarkable

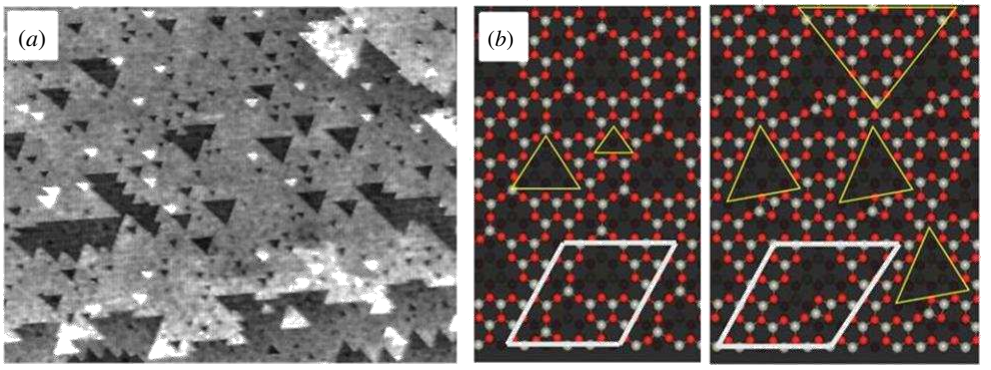

Figure 6. (*a*) Scanning electron microscopy image of the (0001) Zn-terminated surface of ZnO; after Parker *et al.* [107]. (*b*) Two predicted low-energy configurations of the reconstructed Zn-terminated polar surface of ZnO; after Mora Fonz *et al.* [108]. The white and yellow lines indicate the unit cell and observed triangular patterns, respectively. (Online version in colour.)

reconstructions, as illustrated in figure 6*a*, for the case of the Zn-terminated structure, where vacancies cluster to form triangular features.

DFT calculations by Kresse *et al*. [109] demonstrated the stability of such structures and the problem was re-examined in a comprehensive survey by Mora Fonz *et al*. [108]. Their procedure combined a generalized MC approach using the Knowledge Led Master Code (KLMC), developed by Woodley *et al*. [41,57], to explore the large number of vacancy configurations on the surface, with energy evaluations using the General Utility Lattice Program (GULP) [110]. Interestingly, the triangular features observed experimentally fell out from the simulations as the low-energy configurations, as illustrated in figure 6*b*. The calculations were also able to rationalize complex structures on the oxygen-terminated surface and changes of the surface structure with composition. KLMC has also been employed to predict the structure of the surface sublayer of KTaO3 [111]. This ternary oxide adopts the perovskite structure and, given the higher charge of the tantalum cation, it was previously suggested that the 001 surface should terminate with a potassium oxide layer together with vacancies to remove the surface dipole. Using the composition of a partially filled tantalum oxide-terminated 001 surface, KLMC found that the same partially filled potassium layer could be formed, supported by a mixed potassium and tantalum oxide layer that had a more stable surface energy (0.2 eV lower).

Overall, computational and experimental studies have revealed the fascinating variety and complexity of the surface structure of apparently simple oxides and the ability of modelling techniques to rationalize the experimentally observed structures.

(c) Nanostructure prediction

Given that reactions occur on surfaces, it is not surprising to see research specifically based on nanoparticles as the external surface to volume ratio is maximized. Moreover, many properties scale with the size of the nanoparticle so that there is an additional tuning parameter [112] that can be used when seeking a material with a target value of a desired property that improves the performance of an application. For example, to harvest solar power efficiently on the surface of the Earth, it is desirable to have a material with a band gap that is suitable for capturing the most abundant source of photons, i.e. visible light. The atomic structure of a nanoparticle may resemble a cut taken from the bulk phase; typically, for crystalline materials, the equilibrium morphology is determined by the Wulff construction, where the distance from the centre of each faceted surface to the centre of the crystalline particle is proportional to its surface energy. Ignoring the complications that the effects the surrounding environment might have on the structure of the nanoparticle (nanoparticles may be capped with ligands to control their size and to add or

enhance a desired property, or they may be supported as discussed below), the particle may contain grain boundaries and, as a function of particle size, may also undergo a structural phase change. Using MD, Sayle *et al*. [113,114] have simulated the melting of a particle cut of its bulk phase and then the recrystallization of a low-energy configuration, which nicely demonstrated the formation of grain boundaries and, for example, micro-twinning within a particle of MnO_2. Bulk zinc sulfide can readily adopt the sphalerite or wurtzite phase. The former is more stable under ambient conditions; however, the latter becomes more thermodynamically stable for smaller sized nanoparticles. Moreover, as mentioned earlier, simulations have predicted cuts of other phases to become competitive [115] (figure 4*c*) and for the smallest sized clusters, the so-called nanoclusters, the low-energy structures form bubbles [116] and, therefore, no longer resemble a bulk from any known bulk phase. By exploitation of energies of both bulk and non-bulk clusters as a function of size, the critical sizes for 'structural transitions' are predicted; see, for example, a perspective on modelling nanocluster and nucleation [117] or the more recent paper by Bromley *et al*. [118], which both consider particles of zinc oxide. The former also highlights that such studies can help us to understand and gain insight into the atomic mechanisms of nucleation and the early stages of crystal growth.

Nanoparticles, by definition, have a diameter of anything between 1 and 100 nm (10^{-9} metres). To calculate the energy based on classical atomistic methods for the larger particles would currently require modern high-performance computer facilities. For particles of approximately 10 nm, more accurate electronic structure methods can be employed. Only for the smaller sizes and below, i.e. for nanoclusters, can current computer resources be used to sample many configurations on the energy landscape. The global optimization methods described above have had enormous success in predicting the tentative low-energy structures of nanoclusters as a function of size, where size relates to the number of atoms or formula units. Here we restrict ourselves to providing one recent example for a metallic, a covalent and an ionic system.

Strictly speaking all small nanoclusters cannot be metallic as there are not enough electronic states to form bands. In figure 7, the tentative lowest energy configurations for titanium clusters of sizes 1–32 are shown [119]. Their energies and relaxed configurations were obtained using a DFT approach implemented within the all-electron, full potential electronic structure code FHI-aims [120] with the appropriate spin, the PBEsol exchange–correlation functional and a tight basis set equivalent to triple zeta plus polarization. To reduce dramatically the computational cost of finding the configurations shown in figure 7, the study first employed a genetic algorithm, as implemented within KLMC [41], to search for the lowest energy minima on the energy landscape that is defined using a many-body embedded atom method as implemented with GULP [110], which included a combination of a many-body attractive term and a repulsive two-body Born–Mayer interatomic potential. For each size, the best, as determined by GULP, 250 local minimum configurations generated by KLMC were further refined and reranked using FHI-aims. The final configurations suggest a growth mechanism that is based on forming coordination centres by interpenetrating icosahedra, icositetrahedra and Frank–Kasper polyhedra. A better view of these configurations, as well as those shown in figure 8, can be gained by using the website graphical interface of the WASP@N database [121], which contains data on published nanocluster structures. Change the element and a different set of global minima configurations are obtained; see, for example, the review by Baletto & Ferrando [122], who pay special attention to the interplay of energetic, thermodynamic and kinetic factors in the explanation of cluster structures that are observed in experiments. For bimetallic and more generally alloys, see [123,124].

Octahedral bubble-like nanoclusters, with T_h, T_d or T symmetry, which have been predicted for a range of compounds, including zinc sulfide [125], zinc oxide [126], gallium nitride [127] and silicon carbide [101], have recently been reported [128] as a new class of carbon nanostructures which, for the smallest sizes, have a similar stability to the well-known carbon fullerenes. The octahedral carbon clusters contain tetragonal rings, which, despite a common belief, prove to be an energy-efficient means of bending graphene sheets to make three-dimensional spheroid shapes.

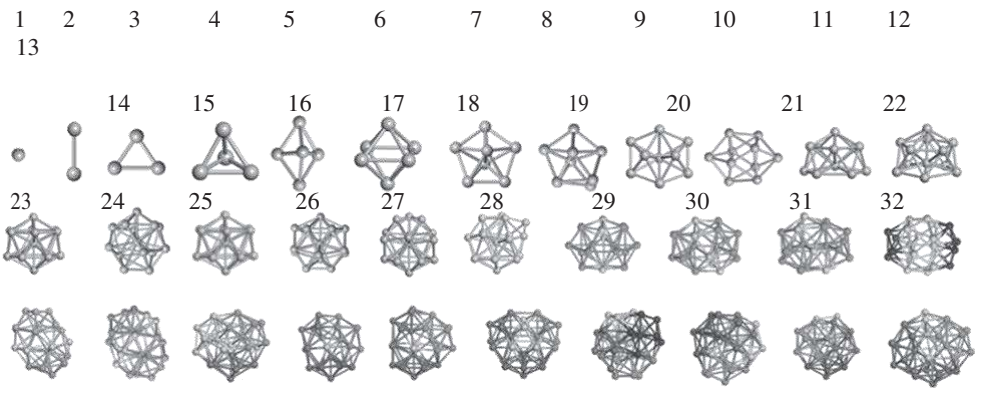

Figure 7. Tentative global minimum atomic structures for clusters of titanium as measured using DFT with the PBEsol functional, as found by Lazauskas *et al*. [119].

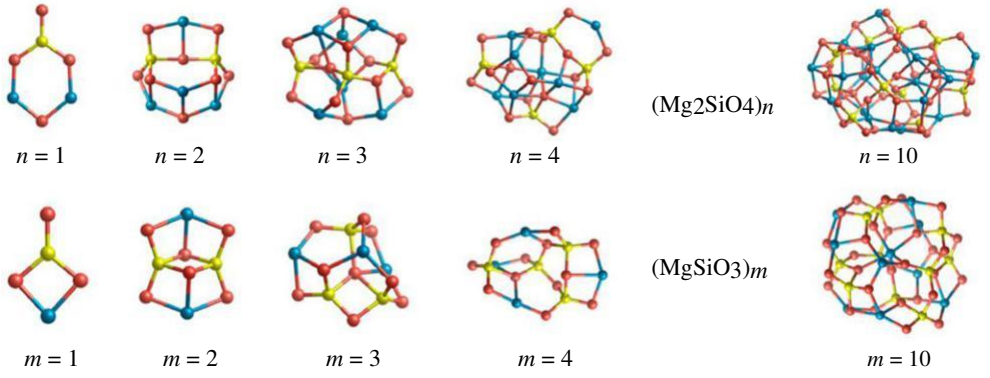

Figure 8. Tentative global minimum atomic structures for clusters of magnesium silicates as measured using DFT with the PBE0 hybrid functional, as found by Escatllar *et al.* [16]. Blue, yellow and red spheres represent magnesium, silicon and oxygen atoms, respectively. (Online version in colour.)

Magnesium-rich silicates are likely to be particularly important for understanding the formation, processing and properties of cosmic dust grains. Although astronomical observations, e.g. from infrared spectra, and laboratory studies have revealed much about such silicate dust, many studies rely on comparisons with the properties of bulk silicates. It is common to assume that a continuous change in a property with respect to the nanoparticle size provides a route for estimating the size and, via a match between simulated and observed data, the composition of a particle. Extrapolating from bulk, under the assumption that the atomic structure of the particles remains bulk-like, right down into the smallest size regime, that of nanoclusters would, however, be a mistake as adding one or two atoms to a nanocluster will change the lowest energy atomic configuration, as already shown in figure 7. This was one of the motivations for predicting the tentative global minima for nanoclusters of magnesium silicates, as shown in figure 8, along with reporting of their infrared vibrational spectra [16].

Given that clusters of compounds with the same stoichiometry can adopt similar atomic structures it may be more efficient to data mine published atomic configurations as opposed to performing a search on the energy landscape [129,130]. Woodley & Bromley [13] provide a more complete description of these and other aspects of computational modelling of nanoparticles.

(d) Supported nanoclusters

Many applications of nanosystems involve nanostructures supported on oxides, which are especially common in catalytic applications. An important example is provided by copper

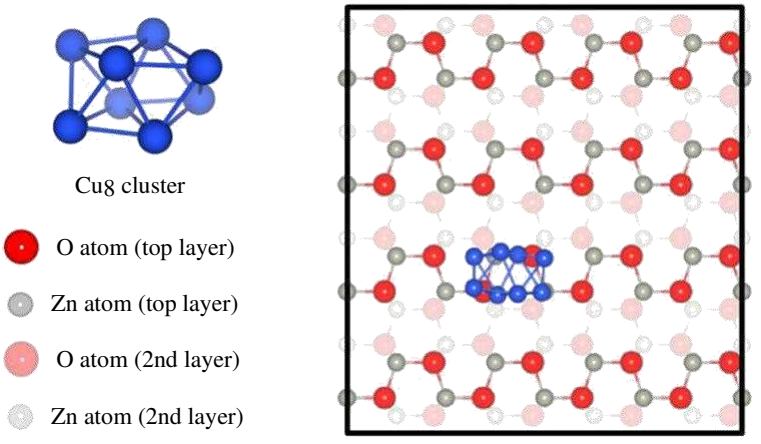

Figure 9. Predicted structure, as viewed from above, of the Cu_8 nanocluster on the (11–20) surface of ZnO (black rectangle indicates the unit cell employed in the simulation) [18]. (Online version in colour.)

supported on zinc oxide, which has been very widely studied owing to its use as an industrial catalyst for the conversion of syngas (CO/H_2) to methanol. Fundamental problems remain, however, regarding both the atomic-level structure of the catalyst and the mechanism of the catalytic reaction.

Work by Mora Fonz *et al*. [18] has addressed the problem of the structure of nano-Cu clusters initially on the non-polar surfaces of ZnO. The first step was to develop an effective interatomic potential for Cu–ZnO interactions, which was achieved by performing a series of DFT calculations on the interaction of Cu atoms and small clusters with the surface; the resulting energy surface was then fitted to a potential model. Using the potential model derived, Cu clusters were 'grown' on the non-polar (11–20) surface by successive additions of single atoms, after which an optimization was performed using the KLMC software. Figure 9 illustrates the resulting structure for a supported Cu_8 nanocluster. Interestingly, compared with the isolated Cu_8 cluster, on the surface the cluster flattens out and effectively 'wets' the surface of the ZnO. Moreover, these configurations are not generated by a procedure in which the cluster is first optimized in isolation and then deposited on the surface and reoptimized. The lowest energy structure is only attained when the atom-by-atom method is used to grow the cluster on the surface. Further work is now in progress on larger clusters and on clusters on the polar surfaces. Once developed, these models will provide a basis for subsequent exploration of the catalytic mechanism.

(e) Organic solids: pharmaceutical solid-form screening

The application of crystal structure prediction methods to organic molecular crystals has, until recently, been dominated by pharmaceutical materials applications. The solid form of active pharmaceutical ingredients (APIs) is important in determining many properties, such as solubility rate, stability, hygroscopicity and compressibility. Therefore, polymorphism— where a molecule adopts multiple crystal structures, either concomitantly or under different crystallization conditions—is important when formulating an API. While polymorphism can be an opportunity for tuning materials properties, the appearance of an unanticipated polymorph of an API represents a significant risk for control of properties [131,132]. The application to polymorph risk assessment for pharmaceuticals [133] has motivated much of the development of structure prediction methods for organic molecular crystals.

A particular challenge of completely sampling the energy landscape of possible crystal structures is the molecular flexibility of typical APIs. Conformational polymorphism, where different molecular conformers are adopted in different crystal structures of a molecule, is

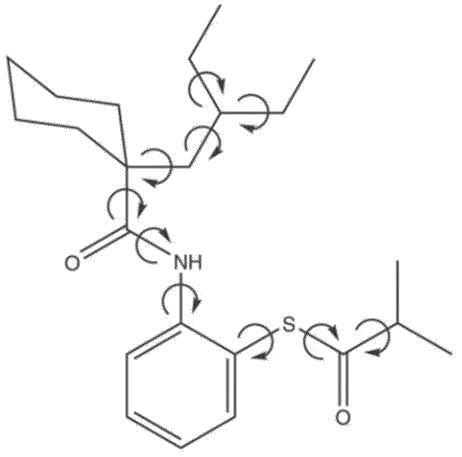

Figure 10. Chemical diagram of dalcetrapib, with arrows highlighting the 10 torsional degrees of freedom. Reproduced from [136].

not uncommon. More than one in three known polymorphic molecules exhibit conformational polymorphism [134]. Furthermore, highly flexible molecules often do not adopt the lowest energy conformer of the isolated molecule in their crystal structures, and intermolecular forces can distort molecules significantly from their gas-phase geometries [135]. Therefore, flexible degrees of freedom within a molecule (e.g. rotation about single bonds) must often be considered along with packing variables, such as the unit cell, molecular positions and orientations. This results in a high-dimensionality search space and predicted crystal structures whose relative stabilities are defined by a balance between many weak intermolecular interactions, relative conformational energies and intramolecular strain. The pharmaceutical application has, therefore, been an excellent application to drive forward structure-searching methods and accurate models for ranking structures.

Neumann *et al.*'s study of dalcetrapib [136], a molecule with 10 rotatable bonds (figure 10), illustrates the value of crystal structure prediction in the pharmaceutical context. Two polymorphs, related by a temperature-induced phase transition, were known after extensive crystallization screening. Structure prediction using solid-state DFT for energy ranking produced structures corresponding to these known polymorphs, with the global energy minimum corresponding to the low-temperature polymorph B. The predictions also produced slightly higher energy, higher density putative structures, which become lower in energy than the known polymorph if pressures approximately above 0.2 GPa are applied in the energy calculations. This suggested an experimental route to this predicted polymorph, which was confirmed by *in situ* high-pressure crystallization from solution, leading to a new crystal corresponding to one of these predicted polymorphs.

Reported successful applications of polymorph prediction applications such as this are becoming more and more frequent, often with the results of structure prediction providing the impetus for further experimental screening. As another example, computational structure prediction was used as part of polymorph screening of the drug molecule galunisertib [137]; the screening revealed an astonishing 10 polymorphs, as well as solvate crystal structures (where solvent molecules crystallize with the organic molecule, forming a two-component crystal). However, despite exploring a wide range of solvent-based crystallizations and more unusual crystallization conditions, the crystal structure corresponding to the global energy minimum has not been realized experimentally. In a final example, structure prediction identified [138] a likely new polymorph of the simple molecule trimesic acid, despite being subjected to considerable attention from the chemical crystallography community for 50 years [139]. The results prompted

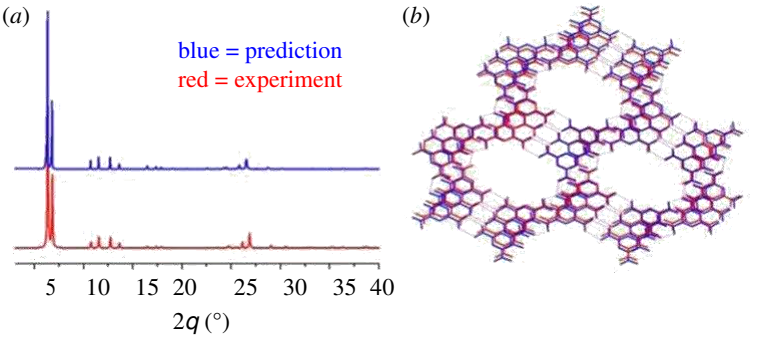

Figure 11. (*a*) Comparison between predicted and experimental powder X-ray diffraction patterns of the computationally predicted new polymorph of trimesic acid; (*b*) structural overlay of the predicted (blue) and the experimental (red) crystal structures. Reproduced from [138]. (Online version in colour.)

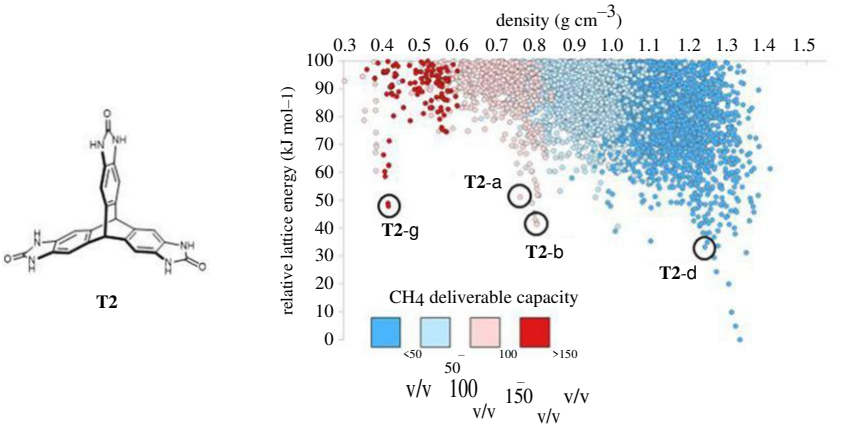

Figure 12. Chemical diagram of benzimidazolone T2 and its energy–structure–function map, showing simulated CH_4 deliverable capacities for all predicted crystal structures [140]. Experimentally observed crystal structures are circled and labelled T2-α to T2-δ. The data point for each predicted crystal structure is colour coded by its CH_4 deliverable capacity (in units of v STP/v; 65–5.8 bar, 298 K). Reproduced from Pulido *et al.* [140]. (Online version in colour.)

a high-throughput robotic crystallization screen of 280 solvent combinations, which led to experimental realization of the new predicted polymorph (figure 11).

(f) Organic solids: functional materials discovery

Molecular organic crystal structure prediction has more recently started to see broader applications in materials discovery. The discovery of materials with targeted properties switches the focus from gaining a detailed picture of the solid-form landscape of a single molecule to an evaluation of the likely crystal structures of larger sets of molecules. As long as the property of interest is readily calculable from the crystal structure, reliable structure prediction can help screen candidate molecules and prioritize synthetic efforts towards the most promising of these. In this area, Day, Cooper and co-workers introduced the concept of an energy–structure–function map [140,141], where the simulated properties of interest are projected onto the energy landscape of predicted crystal structures of a molecule. When such a map is presented visually (figure 12), it can offer a useful visual assessment of which candidate is most likely to lead to the best materials properties for a given application.

An example is Pulido *et al.*'s computationally guided discovery of microporous molecular crystals [140]. Crystal structure prediction was performed on a set of eight molecules chosen as candidates to form porous molecular crystals, with the aim of finding materials with high methane deliverable capacities (e.g. for use in natural gas-powered vehicles). A porous solid must deliver at least 150 volumes of methane at standard temperature and pressure per volume of solid (v STP/v) over a storage and release cycle to be of practical interest for methane storage. Porosity is rare in organic molecular crystals because of the energetic driving force for molecular close packing, so candidate molecules were designed with characteristics that would oppose close packing—awkward, rigid shapes with strong, directional hydrogen bonding interaction sites. The energy–structure–function maps calculated for all eight molecules suggested the triptycene benzimidazolone molecule, **T2**, as the molecule with the most promising crystal structures (figure 12, dark red points). Three new polymorphs of **T2** were subsequently realized (**T2-β**, **T2-γ**, **T2-δ**), all corresponding to predicted structures. These include **T2-γ**, which is the lowest density molecular crystal reported in the Cambridge Structural Database [142] of known organic crystal structures and whose measured surface area and CH_4 capacity match those from the predictions.

The experimentally observed crystal structures in this example are energetically far above the global lattice energy minimum, but occupy so-called 'spikes' on the energy landscape: low-density structures that are much lower in energy than the general energy-density trend on the landscape. These spikes are interpreted as corresponding to isolated, deep regions of the lattice energy surface, so that they can be kinetically trapped under the right crystallization conditions. Here, this corresponds to crystallization from solvent which fills the pores in the structure, but which can be removed from the crystal structure, yielding the activated porous structure. The importance of solvent stabilization in directing crystallization, shown in this example and also known in pharmaceutical crystallization [143], has led to further method development to understand the role of solvent in determining the final structure in microporous molecular crystals [144].

Apart from the area of microporous solids, energy–structure–function maps have also been used in screening for organic molecular semiconductors with high charge carrier mobilities [89,145–147] and studies of molecular organic photocatalysts [148]. These are other application areas where the property of interest depends strongly on crystal packing, so that crystal structure prediction is becoming an enabling technology for computationally led materials discovery.

5. Perspective and future prospects

This review has, we hope, showed the accelerating progress in this already fast-moving field. Structure prediction is now a reality for an increasingly diverse and complex range of compounds and materials. Future developments in the field will include first an increasingly close integration with experiment, with structure prediction techniques being used routinely to guide materials design; second, alongside continuing algorithmic developments, as in, for example [149–151], there will be a rapid growth in the use of machine learning and other approaches from computer science. Third, as we enter the era of exa-scale computing, the horizons and ambitions of the field will expand. Structure prediction will continue to provide challenges for experimentalists, theoreticians and computational scientists.

Data accessibility. This article has no additional data.
Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.
Acknowledgements. We are grateful for collaboration and discussion with Robert Bell, David Mora Fonz, Sally Price, Matt Rosseinsky, Alexey Sokol, Tomas Lazauskas, Andrew Cooper and Michele Ceriotti.

References

1. Kuhlman B, Bradley P. 2019 Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697. (doi:10.1038/s41580-019-0163-x)

2. Price SL. 2018 Is zeroth order crystal structure prediction (CSP_0) coming to maturity? What should we aim for in an ideal crystal structure prediction code? *Faraday Discuss.* **211**, 9–30. (doi:10.1039/C8FD00121A)
3. Oganov AR. 2018 Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss.* **211**, 643–660. (doi:10.1039/C8FD90033G)
4. Addicoat M *et al.* 2018 Structure searching methods: general discussion. *Faraday Discuss.* **211**, 133–180. (doi:10.1039/C8FD90030B)
5. Addicoat M *et al.* 2018 Crystal structure evaluation: calculating relative stabilities and other criteria: general discussion. *Faraday Discuss.* **211**, 325–381. (doi:10.1039/C8FD90031K)
6. Adjiman CS *et al.* 2018 Applications of crystal structure prediction—organic molecular structures: general discussion. *Faraday Discuss.* **211**, 493–539. (doi:10.1039/C8FD90032A)
7. Burger V *et al.* 2018 Applications of crystal structure prediction—inorganic and network structures: general discussion. *Faraday Discuss.* **211**, 613–642. (doi:10.1039/C8FD90034E)
8. Woodley SM, Catlow R. 2008 Crystal structure prediction from first principles. *Nat. Mater.* **7**, 937–946. (doi:10.1038/nmat2321)
9. Nyman J, Day GM. 2015 Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **17**, 5154–5165. (doi:10.1039/C5CE00045A)
10. Nyman J, Day GM. 2016 Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 31 132–31 143. (doi:10.1039/ C6CP05447A)
11. Sun W, Dacek ST, Ong SP, Hautier G, Jain A, Richards WD, Gamst AC, Persson KA, Ceder G. 2016 The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225. (doi:10.1126/sciadv.1600225)
12. Jansen M, Pentin IV, Schön JC. 2012 A universal representation of the states of chemical matter including metastable configurations in phase diagrams. *Angew. Chem. Int. Ed.* **51**, 132– 135. (doi:10.1002/anie.201106220)
13. Woodley SM, Bromley ST. 2018 Introduction to modeling nanoclusters and nanoparticles. In *Frontiers of nanoscience* (eds ST Bromley, SM Woodley), vol. 12, pp. 1–54. Amsterdam, The Netherlands: Elsevier.
14. Buckeridge J, Sokol AA. 2016 One-dimensional nanosystems. In *Computational modeling of inorganic nanomaterials* (eds ST Bromley, MA Zwijnenburg), pp. 45–80. Boca Raton, FL: CRC Press.
15. Wang S, Fan Z, Koster RS, Fang C, van Huis MA, Yalcin AO, Tichelaar FD, Zandbergen HW, Vlugt TJH. 2014 New ab initio based pair potential for accurate simulation of phase transitions in ZnO. *J. Phys. Chem. C* **118**, 11 050–11 061. (doi:10.1021/jp411308z)
16. Escatllar AM, Lazaukas T, Woodley SM, Bromley ST. 2019 Structure and properties of nanosilicates with olivine (Mg_2SiO_4)N and pyroxene ($MgSiO_3$)N compositions. *ACS Earth Space Chem.* **3**, 2390–2403. (doi:10.1021/acsearthspacechem.9b00139)
17. Hou Q, Buckeridge J, Lazauskas T, Mora-Fonz D, Sokol AA, Woodley SM, Catlow CRA. 2018 Defect formation in In_2O_3 and SnO_2: a new atomistic approach based on accurate lattice energies. *J. Mater. Chem. C* **6**, 12 386–12 395. (doi:10.1039/C8TC04760J)
18. Mora-Fonz D, Lazauskas T, Woodley SM, Bromley ST, Catlow CRA, Sokol AA. 2017 Development of interatomic potentials for supported nanoparticles: the Cu/ZnO case. *J. Phys. Chem. C* **121**, 16 831–16 844. (doi:10.1021/acs.jpcc.7b04502)
19. Jiang N, Woodley SM, Catlow CRA, Zhang X. 2015 Applying a new interatomic potential for the modelling of hexagonal and orthorhombic YMnO3. *J. Mater. Chem. C* **3**, 4787–4793. (doi:10.1039/C4TC02759K)
20. Netsianda M, Ngoepe PE, Richard C, Catlow A, Woodley SM. 2008 The displacive phase transition of vanadium dioxide and the effect of doping with tungsten. *Chem. Mat.* **20**, 1764– 1772. (doi:10.1021/cm701861z)
21. Woodley SM, Catlow CRA, Gale JD, Battle PD. 2000 Development of a new force field for open shell ions: application to modelling of $LaMnO_3$. *Chem. Commun.* **0**, 1879–1880. (doi:10.1039/b005802p)
22. Woodley SM, Battle PD, Gale JD, Catlow CRA. 1999 The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.* **1**, 2535–2542. (doi:10.1039/a901227c)
23. Liang T *et al.* 2013 Reactive potentials for advanced atomistic simulations. *Annu. Rev. Mater. Res.* **43**, 109–129. (doi:10.1146/annurev-matsci-071312-121610)

24. Woodley SM. 2004 Prediction of crystal structures using evolutionary algorithms and related techniques. In *Applications of evolutionary computation in chemistry* (ed. RL Johnston), vol. 110, pp. 95–132. Berlin, Germany: Springer.
25. Day GM. 2011 Current approaches to predicting molecular organic crystal structures. *Crystallogr. Rev.* **17**, 3–52. (doi:10.1080/0889311X.2010.517526)
26. Oganov AR (ed.). 2010 *Modern methods of crystal structure prediction*. Weinheim, Germany: Wiley-VCH.
27. Schön JC, Putz H, Jansen M. 1996 Studying the energy hypersurface of continuous systems—the threshold algorithm. *J. Phys. Condens. Matter* **8**, 143–156. (doi:10.1088/0953-8984/8/2/004)

28. Schön JC. 1996 Studying the energy hypersurface of multi-minima systems—the threshold and the lid algorithm. *Berichte der Bunsengesellschaft für physikalische Chemie* **100**, 1388–1391. (doi:10.1002/bbpc.19961000903)
29. Sibani P, van der Pas R, Schön JC. 1999 The lid method for exhaustive exploration of metastable states of complex systems. *Comput. Phys. Commun.* **116**, 17–27. (doi:10.1016/ S0010-4655(98)00176-3)
30. Wales DJ, Doye JPK. 1997 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. (doi:10.1021/jp970984n)
31. Goedecker S. 2004 Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **120**, 9911–9917. (doi:10.1063/1.1724816)
32. Gehrke R, Reuter K. 2009 Assessing the efficiency of first-principles basin-hopping sampling. *Phys. Rev. B* **79**, 085412. (doi:10.1103/PhysRevB.79.085412)
33. White RP, Mayne HR. 1998 An investigation of two approaches to basin hopping minimization for atomic and molecular clusters. *Chem. Phys. Lett.* **289**, 463–468. (doi:10.1016/S0009-2614(98)00431-X)
34. Li ZQ, Scheraga HA. 1987 Monte-Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl Acad. Sci. USA* **84**, 6611–6615. (doi:10.1073/pnas.84.19.6611)

35. Johnston RL. 2003 Evolving better nanoparticles: genetic algorithms for optimising cluster geometries. *Dalton Trans.* **22**, 4193–4207. (doi:10.1039/b305686d)
36. Rossi G, Ferrando R. 2006 Global optimization by excitable walkers. *Chem. Phys. Lett.* **423**, 17–22. (doi:10.1016/j.cplett.2006.03.003)
37. Eberhart RC, Shi YH. 2004 Special issue on particle swarm optimization. *IEEE Trans. Evol. Comput.* **8**, 201–203. (doi:10.1109/TEVC.2004.830335)
38. Wang YC, Lv JA, Zhu L, Ma YM. 2010 Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **82**, 094116. (doi:10.1103/PhysRevB.82.094116)
39. Fournier R. 2010 Density-functional and global optimization study of copper-tin core-shell clusters. *Can. J. Chem. Rev. Can. Chim.* **88**, 1071–1078. (doi:10.1139/V10-066)
40. Curtis F, Rose T, Marom N. 2018 Evolutionary niching in the GAtor genetic algorithm for molecular crystal structure prediction. *Faraday Discuss.* **211**, 61–77. (doi:10.1039/C8FD00067K)

41. Lazauskas T, Sokol AA, Woodley SM. 2017 An efficient genetic algorithm for structure prediction at the nanoscale. *Nanoscale* **9**, 3850–3864. (doi:10.1039/C6NR09072A)
42. Hartke B. 1999 Global cluster geometry optimization by a phenotype algorithm with niches: location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.* **20**, 1752–1759. (doi:10.1002/(SICI)1096-987X(199912)20:16<1752::AID-JCC7>3.0.CO;2-0)
43. Buck U, Pradzynski CC, Zeuch T, Dieterich JM, Hartke B. 2014 A size resolved investigation of large water clusters. *Phys. Chem. Chem. Phys.* **16**, 6859–6871. (doi:10.1039/c3cp55185g)
44. Wells AF. 1984 *Structural inorganic chemistry*, 5th edn. Oxford, UK: Oxford University Press.
45. Wells AF. 1954 The geometric basis of crystal chemistry. *Acta Crystallogr.* **7**, 842–853. (doi:10.1107/S0365110X54002575)
46. Foster MD, Simperler A, Bell RG, Friedrichs OD, Paz FAA, Klinowski J. 2004 Chemically feasible hypothetical crystalline networks. *Nat. Mater.* **3**, 234–238. (doi:10.1038/nmat1090)
47. Treacy MMJ, Rivin I, Balkovsky E, Randall KH, Foster MD. 2004 Enumeration of periodic tetrahedral frameworks. II. Polynodal graphs. *Microporous Mesoporous Mater.* **74**, 121–132. (doi:10.1016/j.micromeso.2004.06.013)

48. Treacy MMJ, Randall KH, Rao S, Perry JA, Chadi DJ. 1997 Enumeration of periodic tetrahedral frameworks. *Z. Kristallogr.* **212**, 768–791.
49. Arhangelskis M *et al.* 2019 Theoretical prediction and experimental evaluation of topological landscape and thermodynamic stability of a fluorinated zeolitic imidazolate framework. *Chem. Mat.* **31**, 3777–3783. (doi:10.1021/acs.chemmater.9b00994)
50. Pickard CJ, Needs RJ. 2011 Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201. (doi:10.1088/0953-8984/23/5/053201)
51. Case DH, Campbell JE, Bygrave PJ, Day GM. 2016 Convergence properties of crystal structure prediction by quasi-random sampling. *J. Chem. Theory Comput.* **12**, 910–924. (doi:10.1021/acs.jctc.5b01112)
52. Karamertzanis PG, Pantelides CC. 2005 Ab initio crystal structure prediction—I. Rigid molecules. *J. Comput. Chem.* **26**, 304–324. (doi:10.1002/jcc.20165)
53. Vasileiadis M, Pantelides CC, Adjiman CS. 2015 Prediction of the crystal structures of axitinib, a polymorphic pharmaceutical molecule. *Chem. Eng. Sci.* **121**, 60–76. (doi:10.1016/j.ces.2014.08.058)
54. Bucar D-K, Day GM, Halasz I, Zhang GGZ, Sander JRG, Reid DG, MacGillivray LR, Duer MJ, Jones W. 2013 The curious case of (caffeine) · (benzoic acid): how heteronuclear seeding allowed the formation of an elusive cocrystal. *Chem. Sci.* **4**, 4417–4425. (doi:10.1039/c3sc51419f)
55. Kazantsev AV *et al.* 2011 Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *Int. J. Pharm.* **418**, 168–178. (doi:10.1016/j. ijpharm.2011.03.058)
56. Tan L, Pickard CJ, Yu K, Sapelkin A, Misquitta AJ, Dove MT. 2019 Structures of CdSe and CdS nanoclusters from ab initio random structure searching. *J. Phys. Chem. C* **123**, 29 370–29 378. (doi:10.1021/acs.jpcc.9b05763)
57. Woodley SM. 2013 Knowledge Led Master Code search for atomic and electronic structures of LaF_3 nanoclusters on hybrid rigid ion-shell model-DFT landscapes. *J. Phys. Chem. C* **117**, 24 003–24 014. (doi:10.1021/jp406854j)
58. Woodley SM, Sokol AA. 2012 From ergodicity to extended phase diagrams. *Angew. Chem. Int. Ed.* **51**, 3752–3754. (doi:10.1002/anie.201109030)
59. van Eijck BP, Kroon J. 2000 Structure predictions allowing more than one molecule in the asymmetric unit. *Acta Crystallogr. Sect. B* **56**, 535–542. (doi:10.1107/S0108768100000276)
60. Kim S, Orendt AM, Ferraro MB, Facelli JC. 2009 Crystal structure prediction of flexible molecules using parallel genetic algorithms with a standard force field. *J. Comput. Chem.* **30**, 1973–1985. (doi:10.1002/jcc.21189)
61. Zhu Q, Oganov AR, Glass CW, Stokes HT. 2012 Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallogr. Sect. B* **68**, 215–226. (doi:10.1107/S0108768112017466)
62. Curtis F, Li X, Rose T, Vázquez-Mayagoitia Á, Bhattacharya S, Ghiringhelli LM, Marom N. 2018 GAtor: a first-principles genetic algorithm for molecular crystal structure prediction. *J. Chem. Theory Comput.* **14**, 2246–2264. (doi:10.1021/acs.jctc.7b01152)
63. Karfunkel HR, Gdanitz RJ. 1992 Ab initio prediction of possible crystal structures for general organic molecules. *J. Comput. Chem.* **13**, 1171–1183. (doi:10.1002/jcc.540131002)
64. Pillardy J, Arnautova YA, Czaplewski C, Gibson KD, Scheraga HA. 2001 Conformation-family Monte Carlo: a new method for crystal structure prediction. *Proc. Natl Acad. Sci. USA* **98**, 12 351–12 356. (doi:10.1073/pnas.231479298)
65. Gavezzotti A. 1991 Generation of possible crystal structures from molecular structure for low-polarity organic compounds. *J. Am. Chem. Soc.* **113**, 4622–4629. (doi:10.1021/ ja00012a034)
66. Day GM, Motherwell WD, Jones W. 2007 A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital. *Phys. Chem. Chem. Phys.* **9**, 1693–1704. (doi:10.1039/b612190j)
67. Sugden I, Adjiman CS, Pantelides CC. 2016 Accurate and efficient representation of intramolecular energy in ab initio generation of crystal structures. I. Adaptive local approximate models. *Acta Crystallogr. Sect. B* **72**, 864–874. (doi:10.1107/S2052520616015122)
68. Coombes DS, Price SL, Willock DJ, Leslie M. 1996 Role of electrostatic interactions in determining the crystal structures of polar organic molecules. A distributed multipole study. *J. Phys. Chem.* **100**, 7352–7360. (doi:10.1021/jp960333b)

69. Day GM, Motherwell WDS, Jones W. 2005 Beyond the isotropic atom model in crystal structure prediction of rigid molecules: atomic multipoles versus point charges. *Crystal Growth Des.* **5**, 1023–1033. (doi:10.1021/cg049651n)
70. Nyman J, Pundyke OS, Day GM. 2016 Accurate force fields and methods for modelling organic molecular crystals at finite temperatures. *Phys. Chem. Chem. Phys.* **18**, 15 828–15 837. (doi:10.1039/C6CP02261H)
71. Otero-de-la-Roza A, Johnson ER. 2012 A benchmark for non-covalent interactions in solids. *J. Chem. Phys.* **137**, 054103. (doi:10.1063/1.4738961)
72. Reilly AM, Tkatchenko A. 2013 Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *J. Chem. Phys.* **139**, 024705. (doi:10.1063/1.4812819)
73. Heit YN, Beran GJO. 2016 How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures? *Acta Crystallogr. Sect. B* **72**, 514–529. (doi:10.1107/S2052520616005382)
74. Dybeck EC, Abraham NS, Schieber NP, Shirts MR. 2017 Capturing entropic contributions to temperature-mediated polymorphic transformations through molecular modeling. *Crystal Growth Des.* **17**, 1775–1787. (doi:10.1021/acs.cgd.6b01762)
75. Cervinka C, Beran GJO. 2018 Ab initio prediction of the polymorph phase diagram for crystalline methanol. *Chem. Sci.* **9**, 4622–4629. (doi:10.1039/C8SC01237G)
76. Dybeck EC, McMahon DP, Day GM, Shirts MR. 2019 Exploring the multi-minima behavior of small molecule crystal polymorphs at finite temperature. *Crystal Growth Des.* **19**, 5568–5580. (doi:10.1021/acs.cgd.9b00476)
77. Bartók AP, Csányi G. 2015 Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057. (doi:10.1002/qua.24927)
78. Smith JS, Isayev O, Roitberg AE. 2017 ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203. (doi:10.1039/C6SC05720A)
79. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA. 2015 Big data meets quantum chemistry approximations: the δ-machine learning approach. *J. Chem. Theory Comput.* **11**, 2087–2096. (doi:10.1021/acs.jctc.5b00099)
80. Musil F, De S, Yang J, Campbell JE, Day GM, Ceriotti M. 2018 Machine learning for the structure–energy–property landscapes of molecular crystals. *Chem. Sci.* **9**, 1289–1300. (doi:10.1039/C7SC04665K)
81. Bartók AP, Kondor R, Csányi G. 2013 On representing chemical environments. *Phys. Rev. B* **87**, 184115. (doi:10.1103/PhysRevB.87.184115)
82. De S, Bartók AP, Csányi G, Ceriotti M. 2016 Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13 754–13 769. (doi:10.1039/C6CP00415F)
83. Egorova O, Hafizi R, Woods DC, Day GM. 2020 Multi-fidelity statistical machine learning for molecular crystal structure prediction. *J. Phys. Chem. A* (doi:10.1021/acs.jpca.0c05006)
84. McDonagh D, Skylaris C-K, Day GM. 2019 Machine-learned fragment-based energies for crystal structure prediction. *J. Chem. Theory Comput.* **15**, 2743–2758. (doi:10.1021/ acs.jctc.9b00038)
85. Deringer VL, Pickard CJ, Csányi G. 2018 Data-driven learning of total and local energies in elemental boron. *Phys. Rev. Lett.* **120**, 156001. (doi:10.1103/PhysRevLett.120.156001)
86. Deringer VL, Proserpio DM, Csányi G, Pickard CJ. 2018 Data-driven learning and prediction of inorganic crystal structures. *Faraday Discuss.* **211**, 45–59. (doi:10.1039/C8FD00034D)
87. Ceriotti M, Tribello GA, Parrinello M. 2011 Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl Acad. Sci. USA* **108**, 13 023–13 028. (doi:10.1073/pnas.1108486108)
88. Campello RJGB, Moulavi D, Zimek A, Sander J. 2015 Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans. Knowl. Discov. Data* **10**, Article 5.
89. Yang J, De S, Campbell JE, Li S, Ceriotti M, Day GM. 2018 Large-scale computational screening of molecular organic semiconductors using crystal structure prediction. *Chem. Mat.* **30**, 4361–4371. (doi:10.1021/acs.chemmater.8b01621)
90. Elton DC, Boukouvalas Z, Fuge MD, Chung PW. 2019 Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **4**, 828–849. (doi:10.1039/C9ME00039A)

91. Sanchez-Lengeling B, Aspuru-Guzik A. 2018 Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365. (doi:10.11 26/science.aat2663)

92. Ryan K, Lengyel J, Shatruk M. 2018 Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**, 10 158–10 168. (doi:10.1021/jacs.8b03913)
93. Kim B, Lee S, Kim J. 2020 Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **6**, eaax9324. (doi:10.1126/sciadv.aax9324)
94. Collins C, Dyer MS, Pitcher MJ, Whitehead GFS, Zanella M, Mandal P, Claridge JB, Darling GR, Rosseinsky MJ. 2017 Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature* **546**, 280–284. (doi:10.1038/nature22374)
95. Foster MD, Delgado Friedrichs O, Bell RG, Almeida Paz FA, Klinowski J. 2003 Structural evaluation of systematically enumerated hypothetical uninodal zeolites. *Angew. Chem. Int. Ed.* **42**, 3896–3899. (doi:10.1002/anie.200351556)
96. Foster MD, Delgado Friedrichs O, Bell RG, Almeida Paz FA, Klinowski J. 2004 Chemical evaluation of hypothetical uninodal zeolites. *J. Am. Chem. Soc.* **126**, 9769–9775. (doi:10.1021/ja037334j)
97. Zheng N, Bu X, Wang B, Feng P. 2002 Microporous and photoluminescent chalcogenide zeolite analogs. *Science* **298**, 2366. (doi:10.1126/science.1078663)
98. Correll S, Oeckler O, Stock N, Schnick W. 2003 $Li_xH_{12-x-y+z}[P_{12}O_yN_{24-y}]Cl_z$—an oxonitridophosphate with a zeolitelike framework structure composed of 3-rings. *Angew. Chem.* **42**, 3549–3552. (doi:10.1002/anie.200351372)
99. Hamad S, Woodley SM, Catlow CRA. 2009 Experimental and computational studies of ZnS nanostructures. *Mol. Simul.* **35**, 1015–1032. (doi:10.1080/08927020903015346)
100. Woodley SM, Watkins MB, Sokol AA, Shevlin SA, Catlow CRA. 2009 Construction of nano-and microporous frameworks from octahedral bubble clusters. *Phys. Chem. Chem. Phys.* **11**, 3176–3185. (doi:10.1039/b902600b)
101. Watkins MB, Shevlin SA, Sokol AA, Slater B, Catlow CRA, Woodley SM. 2009 Bubbles and microporous frameworks of silicon carbide. *Phys. Chem. Chem. Phys.* **11**, 3186–3200. (doi:10.1039/b902603g)
102. Jimenez-Izal E, Ugalde JM, Matxain JM. 2016 Nanocluster-assembled materials. In *Computational modeling of inorganic nanomaterials* (eds ST Bromley, MA Zwijnenburg), pp. 113–148. Boca Raton, FL: CRC Press.
103. Lewis DW, Ruiz-Salvador AR, Gómez A, Rodriguez-Albelo LM, Coudert F-X, Slater B, Cheetham AK, Mellot-Draznieks C. 2009 Zeolitic imidazole frameworks: structural and energetics trends compared with their zeolite analogues. *Crystengcomm* **11**, 2272–2276. (doi:10.1039/b912997a)

104. Tasker PW. 1979 The stability of ionic crystal surfaces. *J. Phys. C Solid State Phys.* **49**, 4977– 4984. (doi:10.1088/0022-3719/12/22/036)
105. Whitmore L, Sokol AA, Catlow CRA. 2002 Surface structure of zinc oxide (1010), using an atomistic, semi-infinite treatment. *Surf. Sci.* **498**, 135–146. (doi:10.1016/S0039-6028(01)01588-6)

106. Mora-Fonz D, Buckeridge J, Logsdail AJ, Scanlon DO, Sokol AA, Woodley S, Catlow CRA. 2015 Morphological features and band bending at nonpolar surfaces of ZnO. *J. Phys. Chem. C* **119**, 11 598–11 611. (doi:10.1021/acs.jpcc.5b01331)
107. Parker TM, Condon NG, Lindsay R, Leibsle FM, Thornton G. 1998 Imaging the polar $(000\bar{1})$ and non-polar $(10\bar{1}0)$ surfaces of ZnO with STM. *Surf. Sci.* **415**, L1046-L1050. (doi:10.1016/S0039-6028(98)00563-9)
108. Mora-Fonz D, Lazauskas T, Farrow MR, Catlow CRA, Woodley SM, Sokol AA. 2017 Why are polar surfaces of ZnO stable? *Chem. Mat.* **29**, 5306–5320. (doi:10.1021/ acs.chemmater.7b01487)

109. Kresse G, Dulub O, Diebold U. 2003 Competing stabilization mechanism for the polar ZnO(0001)-Zn surface. *Phys. Rev. B* **68**, 245409. (doi:10.1103/PhysRevB.68.245409)
110. Gale JD, Rohl AL. 2003 The general utility lattice program (GULP). *Mol. Simul.* **29**, 291–341. (doi:10.1080/0892702031000104887)
111. Deacon-Smith DEE, Scanlon DO, Catlow CRA, Sokol AA, Woodley SM. 2014 Interlayer cation exchange stabilizes polar perovskite surfaces. *Adv. Mater.* **26**, 7252–7256. (doi:10.1002/adma.201401858)

112. Shevlin SA, Woodley SM. 2010 Electronic and optical properties of doped and undoped $(TiO_2)_n$ nanoparticles. *J. Phys. Chem. C* **114**, 17 333–17 343. (doi:10.1021/jp104372j)
113. Sayle TXT, Catlow CRA, Maphanga RR, Ngoepe PE, Sayle DC. 2005 Generating MnO_2 nanoparticles using simulated amorphization and recrystallization. *J. Am. Chem. Soc.* **127**, 12 828–12 837. (doi:10.1021/ja0434073)
114. Sayle D. 2018 From nanoparticles to mesoporous materials. In *Frontiers of nanoscience* (eds ST Bromley, SM Woodley), vol. 12, pp. 129–144. Amsterdam, The Netherlands: Elsevier.
115. Morgan BJ, Madden PA. 2007 A molecular dynamics study of structural relaxation in tetrahedrally coordinated nanocrystals. *Phys. Chem. Chem. Phys.* **9**, 2355–2361. (doi:10.1039/ b701267e)
116. Burnin A, BelBruno JJ. 2002 Zn_nSm^+ cluster production by laser ablation. *Chem. Phys. Lett.* **362**, 341–348. (doi:10.1016/S0009-2614(02)01105-3)
117. Catlow CRA, Bromley ST, Hamad S, Mora-Fonz M, Sokol AA, Woodley SM. 2010 Modelling nano-clusters and nucleation. *Phys. Chem. Chem. Phys.* **12**, 786–811. (doi:10.1039/B9160 69H)
118. Viñes F, Lamiel-Garcia O, Illas F, Bromley ST. 2017 Size dependent structural and polymorphic transitions in ZnO: from nanocluster to bulk. *Nanoscale* **9**, 10 067–10 074. (doi:10.1039/C7NR02818K)
119. Lazauskas T *et al.* 2018 Thermodynamically accessible titanium clusters TiN, N=2–32. *Phys. Chem. Chem. Phys.* **20**, 13 962–13 973. (doi:10.1039/C8CP00406D)
120. Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren XG, Reuter K, Scheffler M. 2009 Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **180**, 2175–2196. (doi:10.1016/j.cpc.2009.06.022)
121. Woodley SM. 2014 Web assisted structure prediction at the nanoscale: database of published atomic structures of nanoclusters. See http://hive.chem.ucl.ac.uk/.
122. Baletto F, Ferrando R. 2005 Structural properties of nanoclusters: energetic, thermodynamic, and kinetic effects. *Rev. Mod. Phys.* **77**, 371–423. (doi:10.1103/RevModPhys.77.371)
123. Johnston RL. 2012 Metal nanoparticles and nanoalloys. In *Metal nanoparticles and nanoalloys*, vol. 3 (eds RL Johnston, JP Wilcoxon), ch. 1, pp. 1–42. Frontiers of Nanoscience. Amsterdam, The Netherlands: Elsevier.
124. Ferrando R. 2016 Introduction. In *Structure and properties of nanalloys*, pp. 1–11. Frontiers of Nanoscience. Amsterdam, The Netherlands: Elsevier.
125. Spano E, Hamad S, Catlow CRA. 2003 Computational evidence of bubble ZnS clusters. *J. Phys. Chem. B* **107**, 10 337–10 340. (doi:10.1021/jp035508d)
126. Al-Sunaidi AA, Sokol AA, Catlow CRA, Woodley SM. 2008 Structures of zinc oxide nanoclusters: as found by evolutionary algorithm techniques. *J. Phys. Chem. C* **112**, 18 860– 18 875. (doi:10.1021/jp805983g)
127. Shevlin SA, Guo ZX, van Dam HJJ, Sherwood P, Catlow CRA, Sokol AA, Woodley SM. 2008 Structure, optical properties and defects in nitride (III-V) nanoscale cage clusters. *Phys. Chem. Chem. Phys.* **10**, 1944–1959. (doi:10.1039/b719838h)
128. Lazauskas T, Sokol AA, Woodley SM. 2019 Are octahedral clusters missing on the carbon energy landscape? *Nanoscale Adv.* **1**, 89–93. (doi:10.1039/C8NA00013A)
129. Woodley SM, Lazauskas T, Illingworth M, Carter AC, Sokol AA. 2018 What is the best or most relevant global minimum for nanoclusters? Predicting, comparing and recycling cluster structures with WASP@N. *Faraday Discuss.* **211**, 593–611. (doi:10.1039/C8FD00060C)
130. Sokol AA, Catlow CRA, Miskufova M, Shevlin SA, Al-Sunaidi AA, Walsh A, Woodley SM. 2010 On the problem of cluster structure diversity and the value of data mining. *Phys. Chem. Chem. Phys.* **12**, 8438–8445. (doi:10.1039/c0cp00068j)
131. Rietveld IB, Céolin R. 2015 Rotigotine: unexpected polymorphism with predictable overall monotropic behavior. *J. Pharm. Sci.* **104**, 4117–4122. (doi:10.1002/jps.24626)
132. Chemburkar SR *et al.* 2000 Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Organic Process Res. Dev.* **4**, 413–417. (doi:10.1021/op000023y)
133. Price SL, Reutzel-Edens SM. 2016 The potential of computed crystal energy landscapes to aid solid-form development. *Drug Discov. Today* **21**, 912–923. (doi:10.1016/j.drudis.2016.01.014)
134. Cruz-Cabeza AJ, Bernstein J. 2014 Conformational polymorphism. *Chem. Rev.* **114**, 2170–2191. (doi:10.1021/cr400249d)

135. Thompson HPG, Day GM. 2014 Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chem. Sci.* **5**, 3173–3182. (doi:10.1039/C4SC01132E)
136. Neumann MA, van de Streek J, Fabbiani FPA, Hidber P, Grassmann O. 2015 Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening. *Nat. Commun.* **6**, 7793. (doi:10.1038/ncomms8793)
137. Bhardwaj RM, McMahon JA, Nyman J, Price LS, Konar S, Oswald IDH, Pulham CR, Price SL, Reutzel-Edens SM. 2019 A prolific solvate former, galunisertib, under the pressure of crystal structure prediction, produces ten diverse polymorphs. *J. Am. Chem. Soc.* **141**, 13 887–13 897. (doi:10.1021/jacs.9b06634)
138. Cui P, McMahon DP, Spackman PR, Alston BM, Little MA, Day GM, Cooper AI. 2019 Mining predicted crystal structure landscapes with high throughput crystallisation: old molecules, new insights. *Chem. Sci.* **10**, 9988–9997. (doi:10.1039/C9SC02832C)
139. Duchamp DJ, Marsh RE. 1969 The crystal structure of trimesic acid (benzene-1,3,5-tricarboxylic acid). *Acta Crystallogr. Sect. B* **25**, 5–19. (doi:10.1107/S0567740869001713)
140. Pulido A *et al.* 2017 Functional materials discovery using energy–structure–function maps. *Nature* **543**, 657–664. (doi:10.1038/nature21419)
141. Day GM, Cooper AI. 2018 Energy–structure–function maps: cartography for materials discovery. *Adv. Mater.* **30**, 1704944. (doi:10.1002/adma.201704944)
142. Groom CR, Bruno IJ, Lightfoot MP, Ward SC. 2016 The Cambridge structural database. *Acta Crystallogr. Sect. B* **72**, 171–179. (doi:10.1107/S2052520616003954)
143. Bhardwaj RM, Price LS, Price SL, Reutzel-Edens SM, Miller GJ, Oswald IDH, Johnston BF, Florence AJ. 2013 Exploring the experimental and computed crystal energy landscape of olanzapine. *Crystal Growth Des.* **13**, 1602–1617. (doi:10.1021/cg301826s)
144. McMahon DP, Stephenson A, Chong SY, Little MA, Jones JTA, Cooper AI, Day GM. 2018 Computational modelling of solvent effects in a prolific solvatomorphic porous organic cage. *Faraday Discuss.* **211**, 383–399. (doi:10.1039/C8FD00031J)
145. Campbell JE, Yang J, Day GM. 2017 Predicted energy-structure-function maps for the evaluation of small molecule organic semiconductors. *J. Mater. Chem. C* **5**, 7574–7584. (doi:10.1039/C7TC02553J)
146. Rice B, LeBlanc LM, Otero-de-la-Roza A, Fuchter MJ, Johnson ER, Nelson J, Jelfs KE. 2018 A computational exploration of the crystal energy and charge-carrier mobility landscapes of the chiral [6]helicene molecule. *Nanoscale* **10**, 1865–1876. (doi:10.1039/C7NR08890F)
147. Cheng CY, Campbell JE, Day GM. 2020 Evolutionary chemical space exploration for functional materials: computational organic semiconductor discovery. *Chem. Sci.* 11, 4922–4933.
148. Aitchison CM *et al.* 2020 Photocatalytic proton reduction by a computationally identified, molecular hydrogen-bonded framework. *J. Mater. Chem. A* **8**, 7158–7170. (doi:10.1039/D0TA00219D)
149. Pickard CJ. 2019 Hyperspatial optimization of structures. *Phys. Rev. B* **99**, 054102. (doi:10.1103/PhysRevB.99.054102)
150. Allahyari Z, Oganov AR. 2020 Coevolutionary search for optimal materials in the space of all possible compounds. *NPJ Comput. Mater.* **6**, 55. (doi:10.1038/s41524-020-0322-9)
151. Yamashita T, Sato N, Kino H, Miyake T, Tsuda K, Oguchi T. 2018 Crystal structure prediction accelerated by Bayesian optimization. *Phys. Rev. Mater.* **2**, 013803. (doi:10.1103/PhysRevMaterials.2.013803)