

Crystal structure prediction from first principles

The prediction of structure at the atomic level is one of the most fundamental challenges in condensed matter science. Here we survey the current status of the field and consider recent developments in methodology, paying particular attention to approaches for surveying energy landscapes. We illustrate the current state of the art in this field with topical applications to inorganic, especially microporous solids, and to molecular crystals; we also look at applications to nanoparticulate structures. Finally, we consider future directions and challenges in the field.

SCOTT M. WOODLEY* AND RICHARD CATLOW

are in the Department of Chemistry, University College London, Kathleen Lonsdale Building, Gower Street, London WC1E 6BT, UK.

*e-mail: Scott.Woodley@ucl.ac.uk

Twenty years ago John Maddox published a celebrated News and Views article in *Nature*¹ in which he provocatively wrote: “One of the continuing scandals in the physical sciences is that it remains impossible to predict the structure of even the simplest crystalline solids from a knowledge of their composition.”

We may debate the accuracy of this statement when it was made, and indeed several studies^{2–5} reported before Maddox’s article had demonstrated the viability of modelling if not predicting structures. However, it unquestionably highlights the challenge offered by crystal structure prediction. Indeed, there is a growing need for reliable and robust methods for predicting the atomic-level structures and properties of solids. But how genuinely predictive are our current techniques? Or in other words, how well has the field responded, over the past 20 years, to Maddox’s challenge? This review attempts to answer the question by surveying current approaches and recent applications, and by identifying the challenges and problems that remain.

BACKGROUND

First, we must distinguish between modelling and prediction. The former is now routine. Using either lattice energy minimization (where appropriate interatomic potentials are available) or electronic structure techniques, it is generally possible to model the equilibrium structures of solids with acceptable accuracy^{6,7}. For example, in modelling ionic solids, using standard sets of potential parameters (such as the widely used compilation in ref. 8), errors in lattice parameters are usually within 1%. However, such calculations are not predictions. They may simply involve taking the experimental structure and applying standard minimization methods to demonstrate that the resulting displacements from experiment are small; this procedure may help to validate the methods and potentials used, but it achieves little more. More ambitiously, we may take an approximate structure, whose energy is then minimized, to generate a more accurate model. An elegant example of this approach was provided by Andrews and co-workers⁹, who took an approximate

model for the structure of a new zeolite (zeolite nu-87), which after minimization gave a structure with a different symmetry that more satisfactorily refined the crystallographic data. Such calculations are clearly of value as a tool in crystal structure refinement, but they are still not predictions.

How, then, shall we define prediction in this context? To be genuinely predictive, a procedure must start from no empirically based information on the positions of the atoms in the unit cell. A fully predictive method would not include experimentally known unit-cell dimensions, but we can accept as a sub-class of predictive methods those in which cell dimensions are predefined but no assumptions are made as to atomic coordinates, although chemical composition is implicit by the specification of the atomic contents of the unit cell. Given that minimization methods are generally routine and accurate, the challenge posed by prediction is therefore one of obtaining approximate models for unit-cell structures which may then be subsequently refined by methods using lattice energy or electronic structure techniques. The different classes of predictive technique can be distinguished by the approach adopted in the initial phase of approximate structure generation.

EXPLORING CONFIGURATIONAL SPACE

The generation of approximate structures requires efficient procedures for exploring the configurational space defined by the internal coordinates of the unit cell so as to identify regions corresponding to low-energy structures. In many cases, these procedures will involve exploring the ‘energy landscape’, but in others, rather than evaluating an energy function, some simpler ‘cost function’ or figure of merit (based, for example, on bond lengths and coordination numbers) may be used to assess the viability of a structure. Once an energy function or cost function is defined, the main techniques used in searching the landscape are those set out below.

SIMULATED ANNEALING

The simulated annealing approach¹⁰ is based on simple concepts arising from physical annealing. When a metal melts, the atoms are in a disordered state. On cooling slowly from the molten state, the atoms may crystallize in an ordered manner, with the energy reaching the global minimum. If the temperature is decreased too rapidly (the

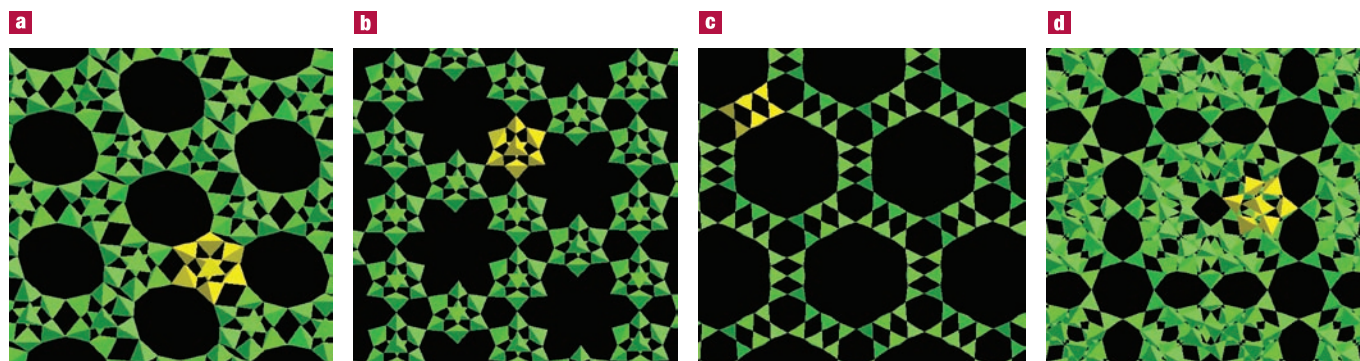


Figure 1 Microporous structures or networks of corner-sharing tetrahedra. These structures were generated¹⁷ using the AASBU approach where the SBU is a double four ring (one unit is highlighted in yellow).

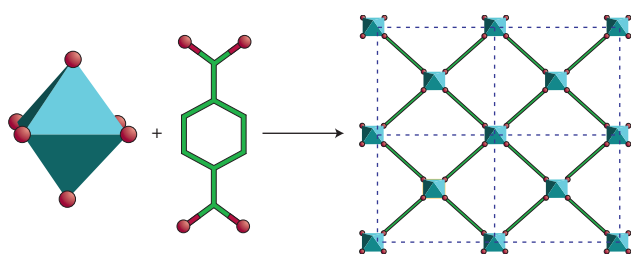


Figure 2 Hybrid inorganic-organic microporous structure. This structure was successfully predicted¹⁸ using the AASBU approach, where the SBUs are terephthalic acid molecules and metal-oxide octahedra.

system is quenched) or if the initial temperature is too low, then the system may freeze into a glass or a crystalline system containing defects such as vacancies and dislocations.

In simulated annealing, the configuration of ions (candidate structure) at each temperature is continually perturbed using either molecular dynamics or a Monte Carlo scheme, where move class operators are applied to perturb a random number of ions within the current candidate structure, and a Metropolis criterion¹¹ is used to determine whether the Monte Carlo move is accepted or rejected. Thus, the initial value of the temperature within the molecular dynamics simulation (or Metropolis criterion) is chosen so that the kinetic energy (or likelihood of accepting a higher-energy candidate) is high enough to allow the system to overcome energy barriers between local minima very easily. The fall in temperature permits a configuration with lower potential energy to be obtained and decreases the probability of jumps between local minima. If this annealing process is continued and carried out slowly enough, the final structure is likely to be the global minimum.

The simulated annealing approach has been used to great effect in assisting structure solution (that is, the determination of structure from experimental data; see for example ref. 12) and in a number of elegant, predictive studies of inorganic solids^{13,14}. Simulated annealing is clearly a straightforward and robust procedure.

Another successful approach that includes the use of simulated annealing was developed and used by Mellot-Draznieks *et al.*^{15,16} and is known as the automated assembly of secondary building units (AASBU) method. The key to this approach is the idea of sticky atom pairs, which can either promote or disfavour SBUs linking together during the simulated annealing or minimization stages. For example, to model aluminophosphates (AlPO_4), SBUs of AlO_4 and PO_4 can

be used whereby the oxygen atoms on the corners of the tetrahedra containing different atom types are encouraged to overlap using simple two-body Lennard-Jones potentials; the correct ratio of cations to oxygen atoms is achieved when a sufficient number of tetrahedra are connected to form the framework. Larger SBUs have been used in the construction of known and hypothetical silicate, aluminophosphate and gallophosphate frameworks¹⁷, some of which are shown in Fig. 1. SBUs are not restricted to inorganic fragments and can be composed of organic molecules or fragments: applying the AASBU approach using terephthalic acid and metal-oxide octahedra, Mellot-Draznieks *et al.*¹⁸ predicted the inorganic-organic hybrid structure shown in Fig. 2, which has been synthesized in the laboratory.

Monte Carlo basin hopping (MCBH¹⁹) is a related powerful technique based on Monte Carlo methods for structure prediction. After applying the Monte Carlo move, the new configuration is immediately relaxed (energy quenched), and the minimized energy of the new candidate is then used in the Metropolis criterion. Thus, MCBH explores a transformed potential energy landscape that is a patchwork of flat plateaux. Importantly, the heights of many energy barriers between polymorphs (local minimum on the original energy landscape) have either vanished or have reduced considerably, and in some cases, transitions are now barrierless. The temperature in the MCBH calculations is typically fixed at 3,000 K, and during the search, the Monte Carlo step size is varied so that the average acceptance of a Monte Carlo move is 50%. MCBH has been used by Wales, who has predicted the global minimum structures for Lennard-Jones clusters containing 1–110 particles²⁰. More exotic schemes have also been applied, for example the hybrid simulated annealing method, developed by Hamad *et al.*²¹ and implemented in the search of global minimum structures of inorganic clusters. Here, between runs, as the molecular dynamics temperature is reduced, the current candidate undergoes several iterations of MCBH. Figure 3 shows a range of cluster structures predicted by using global optimization techniques.

If MCBH and simulated annealing have a weakness, it is that the exploration of the energy landscape necessarily starts from a single point and it is possible that not all low-energy regions may be accessed. To minimize this danger, multiple runs with different initial configurations must be undertaken.

GENETIC ALGORITHM METHODS

Genetic algorithm methods²², or more generally evolutionary algorithm methods, avoid the problems associated with a single starting point by setting up a population of structures. Typically, different random arrangements of atoms make up each candidate structure in the initial population. The algorithm works by mimicking Darwinian (or in some cases Lamarckian) evolution, the basic tenets of

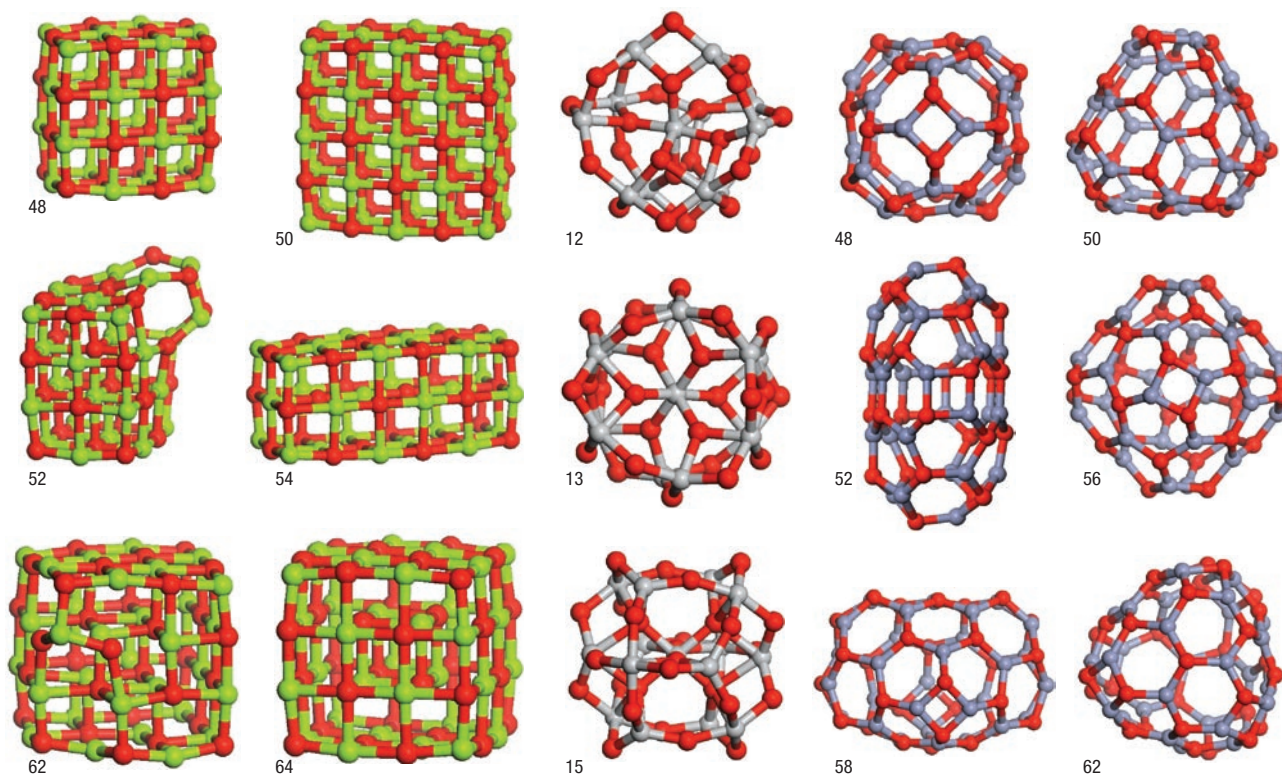


Figure 3 Examples of global minimum structures for oxide nanoparticles generated by using global optimization techniques. Examples, with magnesium, titanium, zinc and oxygen ions coloured green, silver, blue-grey and red, respectively, and grouped from the left, are shown for n units of MgO (ref. 98), TiO_2 (ref. 21) and ZnO (ref. 94).

which are that offspring resemble their parents and that procreation is a reward for success; thus variables determining an organism will, over time (many rounds of procreation), be optimized so that an organism that best fits the environment evolves. When a genetic algorithm is applied to structure prediction, the analogous organism is simply a configuration, or candidate structure. New candidate structures are generated, through procreation, by applying two operators: 'crossover' to pairs of candidate structures, so combining current structural features into new individuals, and 'mutation' to individual candidate structures, so possibly introducing new structural features to the population. Typically, mutation is a Monte Carlo move, a few random ionic displacements, and is important in that it helps to improve and maintain the diversity of the population. The use of different candidates is crucial for the successful application of crossover, where, for example, a random-sized fragment within a candidate structure is switched with a similar-sized (same ionic content) fragment within another candidate to create two new candidate structures. Competition both to 'survive' and be part of the new population, and to 'procreate' in the current population is simulated, where the outcome is based on the relative values of a cost function, of the type mentioned earlier, for the relevant candidate structures. Ultimately, information from the better candidate structures of the current population are statistically more likely to be chosen to create candidate structures for a new population. Thus, candidate structures containing 'good' features will prosper and those with 'bad' features will diminish so that, eventually, a candidate structure should evolve that best fits the target criteria.

A genetic algorithm has many variations, for example in how crossover and mutation routines are actually defined; how competition is simulated within a population when selecting candidates for procreation and/or for survival; how the various modules are linked

together; and ultimately how the population is renewed or updated (see, for example, ref. 23). One popular choice in addressing the first example uses operators based on a phenotype (that is, based on real ionic coordinates), as opposed to genotype (based on analogous genes)^{24–27}. For the second example, simulating a roulette wheel, where the sizes of the slots are based on the quality of all the candidates in the current population, has proved popular for simulating competition. Another common approach uses simple tournaments between two or more randomly selected candidates as they are much faster to perform and are less likely to cause the diversity of the population to fall too quickly. For the third example, both mutation and crossover operators could be used to generate each new offspring or used as two separate procedures for generating a new structure. Finally, in the fourth example, one of the three popular ways to control the population size is to define the new population to be composed only of new candidates plus the current best candidate(s) in the old population (elitism). Killing off duplicate candidates is a variation that has been found to be very beneficial when the chosen population size is small and is generally applied before natural selection in which the better candidates from current and new candidates form the new population. For large populations and when running the algorithm on multiple processors, each new candidate created competes in a tournament with one or a few current candidates, with the loser being eliminated from the population and so keeping the population size fixed.

The viability of a genetic algorithm approach in structure prediction has been shown in a series of studies by, for example, Johnston and co-workers for metallic and alloy clusters²⁵, Woodley and co-workers for both dense and microporous oxide structures²⁸, and Harris and co-workers in solving structures of molecular crystals²⁹. Moreover, the application of the hybrid genetic algorithm,

where all new candidate structures are immediately relaxed, has proved both popular and successful at improving the efficiency of the genetic algorithm^{24–26,30–33}. We note that in one of these studies³³ success was achieved simply by quenching randomly generated structures. A similar approach has been successfully used by Pickard and Needs to several systems including, for example, the prediction of metastable phases of H₂O under pressure³⁴.

TOPOLOGICAL MODELLING METHODS

Topological principles have been used for many decades in enumerating and defining networks that may correspond to real or hypothetical crystal structures, as in the early work of Wells³⁵. The approach has been particularly useful in analysing and predicting frameworks, especially microporous structures, including zeolites. Smith^{36–38} made a substantial contribution to this field in the 1970s. More recently, topological approaches have been applied with considerable success by O’Keeffe and Hyde³⁹, Treacy^{40,41} and by Bell and co-workers⁴². In the latter approach, topologies are generated using advances in combinatorial tiling theory by Dress, Delgado, Friedrichs and others^{43–46}. Their work showed that four-connected (or higher) three-dimensional nets can divide space into polyhedra, so that methods of enumerating polyhedral tilings of Euclidean space may also be used to enumerate the corresponding nets. With the method of Delaney symbols, it can be shown that there are exactly 9, exactly 117 and over 1,300 types of four-connected uninodal, binodal and trinodal nets respectively, which are based on ‘simple’ periodic tilings in which the discrete polyhedral tiles possess only three-connected vertices. It seems likely that, as the number of inequivalent nodes increases, the number of nets enumerated by means of this method will also rapidly increase. The greatest number of topologically distinct T-atoms in a known zeolite framework is 16, in the structure type IWW, and it is conceivable that, using the tiling method, this topology would be discovered among many millions of others. Furthermore, not all zeolites have topologies that correspond to the simple tilings. In fact so far it seems that roughly one-third of the framework types fall into this category, with 7 of the 21 observed uninodal structures and 11 of the 30 known binodals being found among the simple tilings. But by increasing the enumeration to consider ‘quasi-simple’ tilings it is possible to enumerate the 14 remaining uninodal zeolites from among a set of 157. Thus, in principle, the tiling method might be able to generate all possible zeolite frameworks, although a means of automatically filtering the results in order to identify the more likely candidates is needed for the problem to be tractable.

A key feature of the work of Bell and co-workers was the conversion of topologies into structures by placing silicon atoms at all vertices and oxygen atoms halfway along the edges between the vertices followed by the application of energy minimization using standard sets of interatomic potentials for SiO₂. The resulting calculated lattice energies allowed the viability for the structures to be assessed, as discussed in greater detail below.

Zeolite structure prediction is of course a considerable challenge, but the use of topological approaches is, as noted, very widespread in structure prediction and enumeration. Other recent examples are provided by the work of Winkler *et al.*⁴⁷, who developed a systematic approach based on the combination of graph theory with quantum mechanical calculations with applications to hypothetical carbon polymorphs, and Le Bail⁴⁸, who has reported a new code (GRINSP) that explores three-dimensional networks with different connectivities, with wide-ranging applications to framework structured materials.

MOLECULAR PACKING APPROACHES

The prediction of the structures of molecular crystals has been very active in recent years, in part driven by the need in the pharmaceutical

industry to understand and if possible predict polymorphism in crystals of pharmaceutical importance. The problem essentially breaks down into exploring, in a systematic and efficient way, the different modes of packing of the component molecules, including the use of molecular dynamics simulations^{49–51} to allow molecules to find the lowest-energy configurations. A number of grid-search procedures have also been used^{52–54}. One of the most successful has been the MOLPAK method⁵⁵, which has been very effectively combined with energy minimization methods in predicting polymorphic structures^{56–59}. Several other approaches have been used, including simulated annealing techniques using Monte Carlo methods⁶⁰; a review of the current status of the field is given in ref. 61.

Related approaches are based on the concepts of molecular recognition and supramolecular interactions. Dunitz and Gavezzotti⁶² give a detailed discussion and line analysis, and Desiraju⁶³ has pioneered the concept of supramolecular synthons in crystal structure prediction. In addition, a recent very powerful approach has been advanced by Raiteri *et al.*⁶⁴ who used metadynamics to explore energy landscapes for crystalline benzene, with impressive predictions of its different polymorphic phases. More details of this method will be discussed below.

STRUCTURE MODELS BY ANALOGY

An alternative approach to generating approximate structural models is by analogy with known structures of related chemical composition. Although such approaches are not ‘first principles’, they can be very effective. Good examples are given in refs 65 and 66, which address alloy structures. Here, sophisticated data mining approaches are used to identify analogous structures that are then refined using quantum mechanical geometry optimization methods. Data mining has also been applied to developing new atom-pair potentials, which are then used in the search for new molecular crystal structures⁶⁷.

PREDICTIONS AT FINITE TEMPERATURE

At zero temperature, there is a one-to-one correspondence between local minima on the energy or enthalpy (for a defined pressure) landscape and the predicted stable and metastable structures. As temperature increases, more states become accessible, and vibrational contributions to the free energy are important. For finite temperatures, predicted structures are associated with local ergodic regions on the Helmholtz free energy landscape, which increase in size with temperature and merge: statistical approaches are then required. We note that during a molecular dynamics or Monte Carlo simulation at constant temperature, a local ergodic region is simply the set of states where the system remains for longer than the time required to equilibrate and can be observed before escaping to another ergodic region.

There have been two notable approaches in this field: an ergodicity search algorithm, developed by Schön and co-workers in order to predict high-temperature modifications⁶⁸, and the newly developed metadynamics approach for predicting phases at non-zero temperature^{64,69} referred to above. The first approach makes use of Monte Carlo simulations at constant temperature as well as stochastic quenches, and has recently been applied to the free enthalpy landscape of strontium oxide⁶⁸. This study reports the free enthalpy landscape for SrO at pressures of 0 and 160 GPa and eight different temperatures up to 1,850 K as a tree diagram. The structure of these diagrams, where a separate branch represents each local minimum, typically resembles the roots of a tree with the deepest root corresponding to the global minimum. Subsequently, Schön *et al.* computed various thermodynamic and elastic properties for the predicted SrO phases at an *ab initio* level. In earlier work, they derived low-temperature phase diagrams for ternary alkali halide solid solutions from their computed free enthalpies⁷⁰.

The metadynamics approach^{64,69} allows the exploration of the Gibbs free energy as a function of the cell parameters by history-dependent dynamics. Importantly, to discourage successive molecular dynamics runs from exploring previously visited regions the landscape includes penalty functions in the form of Gaussians. By applying steepest descent, the initial metastable, or local minimum, is found for a small unit cell, from which a supercell can be generated for the molecular dynamics runs. Applications of this method include inorganic^{69,71} and organic⁶⁴ crystals referred to earlier. The former investigated the pressure-induced phase transition from diamond to simple hexagonal in silicon, and the structural transformation of an aluminosilicate framework, Li-ABW. The extensive exploration of the phase space allowed a successful prediction of a reconstructive phase transition from anhydrous Li-ABW to eucryptite at the temperature of 920 K, as observed experimentally.

CURRENT STATUS

The methods outlined above have given us a powerful set of tools to respond to the 'Maddox challenge'. But how effective are they? And can we claim that computational crystallography is now truly predictive? We attempt to answer these questions by focusing on three classes of material, namely dense oxides, framework structured materials and molecular solids.

STRUCTURE PREDICTION FOR DENSE MATERIALS

Pannetier *et al.*¹² replied, almost immediately, to Maddox's article with the publication of their implementation of a simulated annealing method and its application to the structure prediction of NbF₄. The cost function, used to assess the quality of a candidate structure, contained two weighted contributions, a Coulomb part and a discrepancy term, which was based on the bond valence model⁷². Their simulated annealing approach, where the unit cell dimensions and contents were fixed, was also used to help solve the structure of the first di-rutile compound LiCoF₄ (ref. 73). Two years after Pannetier *et al.*¹², Freeman and Catlow⁷⁴ published details of their multi-stage simulated annealing approach, which they initially used to generate the structure of rutile. In fact, over the past two decades, numerous groups have made extensive studies in the area of structural prediction, which, importantly, has included the development of 'in-house' software. Typically, even within the same group, the definition of the cost function and the global optimization routines are continually refined. For example, Freeman *et al.*⁷⁵ made modifications to their multi-stage approach before generating the common polymorphs of titania, a system that has become a 'test case' for newly developed global optimizers^{33,74–78}. However, we should note that one set of parameters, for both the algorithm and the cost function used, can generate a large range of different observed structures, as shown by Woodley *et al.* with their multi-stage genetic algorithm approach⁷⁶, and that structure prediction methods have solved previously unknown structures, for example Li₃RuO₄ and, as mentioned already, LiCoF₄ (refs 73,79).

With the available computer power continually increasing, the choice of cost function has evolved from a fitting residual (that is, discrepancy between expected coordination numbers and bond valence), to an energy derived from semi-classical theory and then from density functional theory and Hartree–Fock calculations⁸⁰, and more ambitiously, the free energy⁶⁸. If the cell parameters also formed part of the unknown variables, then all low-energy polymorphs with a defined chemical composition or fixed stoichiometry can be targeted. Schön and Jansen⁸¹ have developed a simulated annealing approach for generating, or finding, all the important (meta)stable candidate structures for a particular chemical formula and for providing information on their stability and eventually guidance on possible synthetic routes. In their earlier work, they successfully

generated the crystal structures formed by pure and mixed noble gases as well as simple ionic compounds¹³, and, moreover, structures that have yet to be synthesized for compounds such as Na₃N and Mg₂OF₂ (refs 82, 83). Further examples include the systems AB₂ (A = Mg, Ca; B = F, Cl)⁸⁴ and high-pressure phases of (M,M')₃N, M₂S (M = Li, Na)⁸⁵. A more complete review of the earlier work of Schön *et al.* can be found in ref. 13.

Unknown reconstructions of surfaces have also been targeted, for example the (103) surface for silicon⁸⁶. Moreover, evolutionary approaches have been applied to predicting the atomic structure of steps on stable crystal surfaces, for example monatomic steps on Si(114)-2×1 (ref. 87).

An application area where global optimization techniques have arguably had the most impact is the prediction of (meta-)stable low-energy structures of nano- and sub-nanoparticles, as observed structural data are very limited. Tiny particles can, for example, be grown in solution^{88–90} or formed by laser ablation of a surface⁹¹. Clusters can be produced with a complete range of sizes and with no one particular size dominant. For some materials, however, clusters with a certain size, so-called 'magic numbers', are formed in abundance. It is widely held that magic numbers indicate particularly stable clusters. In particular, it has been established that for alkali halides, a cuboid fragment, particularly if it has square faces, carved from the bulk, is more stable than the same structure with two ions added or removed⁹². Although the structure of particles whose size is greater than tens or hundreds of nanometres is likely to resemble relaxed fragments cut straight from the bulk phase, the structure of smaller clusters may be remarkably different. For example, the structural motif of the global minima for ZnO and ZnS is typically a cage or bubble-like structure^{93,94}. Moreover, as the size of clusters is increased, for ZnS the first bulk-like particles are predicted to adopt the BCP zeolite phase⁹⁵, as opposed to the observed bulk phases (wurtzite and zincblende). The stable structures for a plethora of systems can be found in the literature, for example Lennard-Jones clusters²⁰, NaCl (ref. 96), AgI (ref. 97), MgO (ref. 98), ZnS (ref. 99), SiO₂ (ref. 100), TiO₂ (ref. 21) and (III–V) nitrides¹⁰¹, as well as metallic¹⁰² and bimetallic or alloy clusters^{103,104}.

FRAMEWORK STRUCTURES

Deem and co-workers^{105–107}, Akporiaye *et al.*¹⁰⁸ and more recently Boisen and co-workers^{109–111}, Mellot-Draznieks *et al.* (see above) and Woodley and co-workers^{26,112} have successfully applied either a simulated annealing or genetic algorithm approach to generating framework materials. Ignoring the differences in the definitions of the cost functions, the different approaches that have been used can be divided into two groups, depending on whether the global optimizer is used to find only the cation sites (tetrahedral or T-sites) or all ion sites. For example, in the successful approach of Deem and Newsam¹⁰⁵, which was published just two years after Maddox's article, the constituent oxygen atoms are added after the locations of the T-sites are found^{105,106}. In contrast, Boisen *et al.* have generated a vast number of candidate framework structures^{109,111} using the geometry and a potential energy function derived from fitting to the disilic acid molecule and starting from a randomized set of *n* silicon and 2*n* oxygen coordinates (*n* = 3,4,6) within the full unit cell. Considering now the prediction of the possible structures that have a 'desired' microporous architecture, Woodley *et al.* have developed an approach whereby the channels or pores can be predefined using exclusion zones, predefined regions within the unit cell. By implementing exclusion zones (where atoms are either forbidden¹¹³ or encouraged to vacate¹¹⁴) within their genetic algorithm, they effectively bias the search towards generating the required metastable framework structures. Within the earlier simulated annealing approach of Deem and Newsam¹⁰⁶, which was developed in order to generate zeolite structures, the locations of the exclusion zones (which were not fixed) were targeted along with the cation positions, or T-sites.

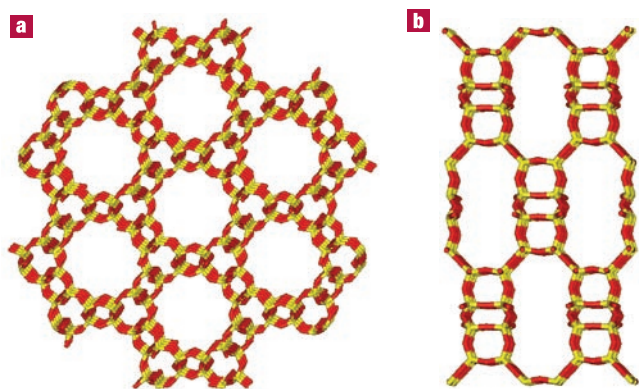


Figure 4 Topological procedures can predict new framework structures. Here we show two hypothetical zeolites derived from tiling theory. **a**, Large-pore structure 1_71 has unidirectional 12-ring channels; **b**, tetragonal structure 1_14 has elongated 12-ring channels, which run in orthogonal directions, intersecting to form 3D pores.

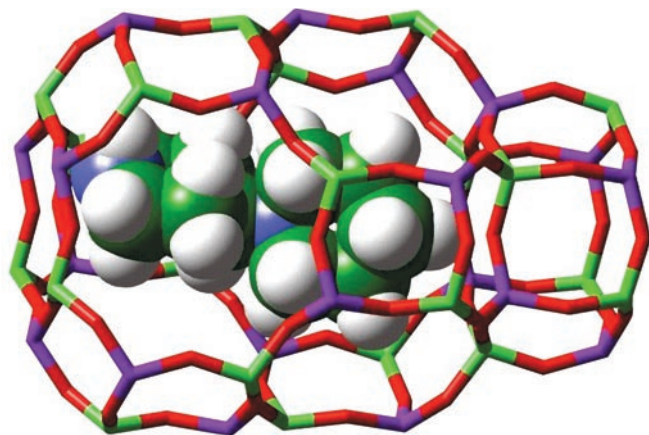


Figure 5 *De novo* procedures have predicted templates for the synthesis of microporous materials. An example is the DAF-5 structure, where for clarity only one of the cages is shown, containing *de novo* designed template molecule 4-piperidinopiperidine.

The success of topological methods in enumerating possible framework structures was highlighted earlier. Moreover, we have argued that the approach becomes still more useful when the energies of the resulting structures are calculated. A particularly effective way of displaying and correlating such calculations was developed by Bell and co-workers⁴², who calculated lattice energies for all silica polymorphs of the uninodal networks derived using the tiling theory of Dress *et al.*^{43,45} discussed above. The resulting energies (with respect to that of alpha quartz (with respect to which all microporous structures are meta-stable)) were plotted against framework density. Also included were the energies of all known microporous topologies, again treated as silica polymorphs. The results are most interesting: the known structures lie close to a band on the plot (with the outliers being systems with atypical chemical composition such as beryllium phosphates). This observation has real predictive value. We can postulate that if a hypothetical structure predicted by the topological/energy minimization procedure is to be synthesized, it must lie

reasonably close to the stability band; and that structures that are distant from this band, although of interest topologically, are unlikely to be of relevance otherwise. Two new structures lying close to the stability band are shown in Fig. 4. They have not yet been synthesized, but there is a clear prediction that their synthesis should be possible. A similar approach can be followed for other chemical compositions known to support tetrahedral frameworks—for example sulphides¹¹⁵, where sulphur has been substituted for oxygen anions in known silicates. Moreover, nanoporous frameworks with alternating cation and anion sites are also predicted to be feasible¹¹⁶.

Microporous materials are synthesized hydrothermally, commonly with the use of organic templates. Lewis *et al.*¹¹⁷ developed a *de novo* computational procedure for designing templates. Their approach uses concepts originally applied in drug design where molecules are ‘grown’ computationally to fit a receptor site. In template design, organic cations are grown to match a target microporous structure. An example is shown in Fig. 5, where the template shown was designed for the microporous aluminophosphate DAF-5. The predicted template succeeded in synthesizing the material rapidly as a pure phase, whereas previous lengthy syntheses had produced impure materials. As a bonus, the predicted site of the template within the host material was verified using microcrystalline single-crystal diffraction techniques¹¹⁸.

MOLECULAR CRYSTAL STRUCTURES

As noted earlier, there are robust and effective procedures for predicting the structures of molecular crystals. The problem is that, except for the simplest systems, the energy landscape is highly complex; hence the widespread occurrence of polymorphism. Moreover, the nature of the polymorph emerging from a particular crystallization procedure may be controlled by kinetic rather than thermodynamic factors. A nice example of this feature is given by recent work of Hulme *et al.*¹¹⁹ and Hamad *et al.*¹²⁰ on 5-fluorouracil, work that also illustrates the predictive nature of the field. Before this study, only one polymorph (Fig. 6a) was known of 5-fluorouracil (an important pharmaceutical compound, used in chemotherapy). Application of the standard packing and minimization procedure predicted a complex energy landscape with a previously unknown structure (Fig. 6b) having a low energy¹¹⁹. The stability of the new structure (polymorph II) is indeed not surprising as it contains ribbons of hydrogen-bonded dimers—a common structural motif in this type of molecular crystal—in contrast to polymorph I which has the unusual feature of short F...F separations. Following the computational predictions, a comprehensive crystallization study was undertaken involving a large number of solvents; polymorph II was eventually crystallized successfully from dry nitromethane, whereas crystallization from an aqueous solution invariably produces polymorph I. This difference was explained by a series of molecular dynamics simulations¹²⁰ in which the early stages of molecular aggregation of 5-fluorouracil were modelled in both water and nitromethane as solvent. In the former, the oxygen atoms of the solute are strongly hydrogen bonded to the solute and aggregation occurs through the association of the hydrophobic F atoms on different solute molecules, whereas solute–solute bonding interactions are weak (see Fig. 7a). In contrast, nitromethane does not form hydrogen bonds with the solute, which is far less solvated by this solvent and consequently forms strong solute–solute hydrogen-bonded structures (Fig. 7b). Thus in the early stages of aggregation in water, the structures formed have short F...F separations, the signature of polymorph I, whereas in nitromethane, hydrogen-bonded dimers are formed, the signature of form II. The effect of the solvent on the polymorphic outcome of the crystallization is explained, following the successful prediction of the new polymorph.

A notable feature within the field of molecular crystal prediction has been the challenge posed by the ‘blind test’ issued by the Cambridge Crystallographic Data Centre¹²¹, in which groups are

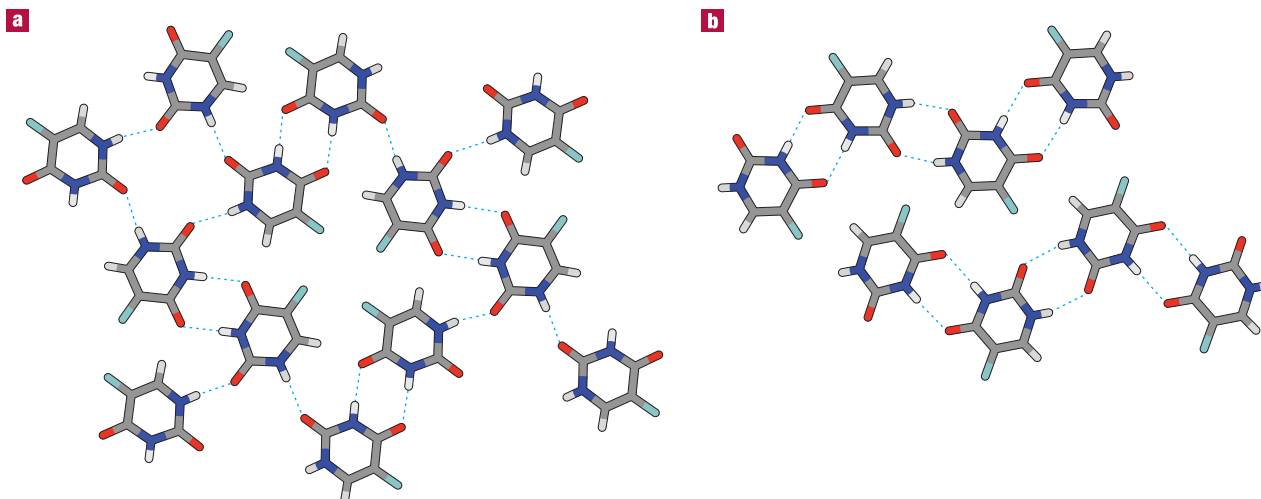


Figure 6 Computational methods were successfully used to predict a new polymorph of 5-fluorouracil. **a**, The known form I; **b**, the newly predicted form II. Hydrogen bonds are represented by dotted lines.

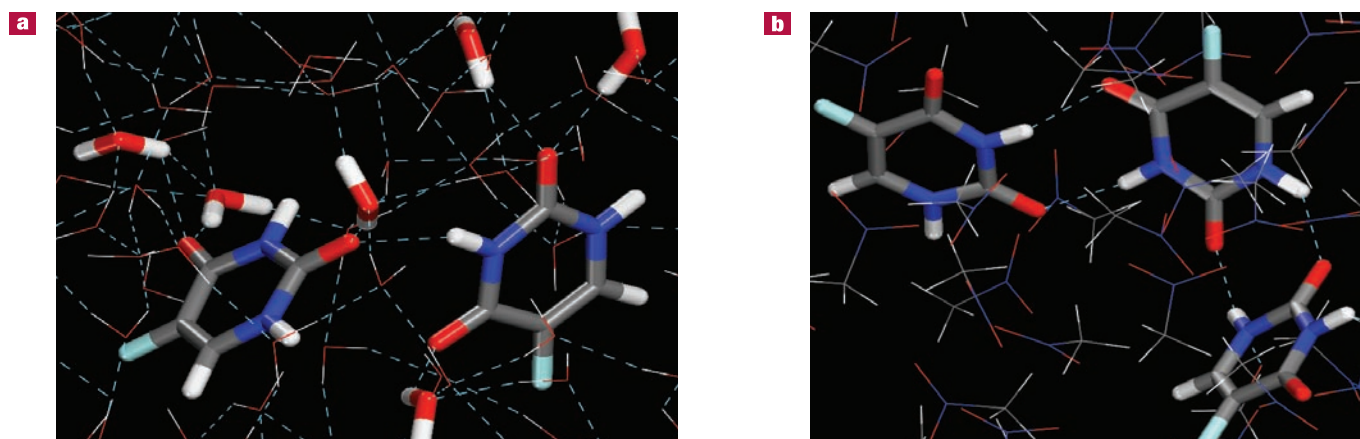


Figure 7 Snapshots from molecular dynamics simulations. **a**, A single hydrogen bond (blue dashed line) formed between 5-fluorouracil molecules in aqueous solution; **b**, a chain of three doubly hydrogen-bonded 5-fluorouracil molecules, formed in pure nitromethane solution.

invited to predict as yet unpublished structures of three or four chosen compounds. A number of such tests have been made and have stimulated worldwide participation. In the first three tests, although some success was achieved, the difficulty of predictive modelling in the field was clear. Greater success was enjoyed in the most recent test in 2007 by a combination of extensive and sophisticated search procedures with the use of accurate and tailor-made potentials (G. M. Day *et al.*, manuscript in preparation). In this context the recent work of Neumann *et al.*¹²² is particularly impressive. They use a hybrid approach to the calculation of lattice energies with standard density functional theory supplemented by empirical van der Waals energies. This approach when combined with an efficient structure generator and minimizer successfully predicted all four compounds in the 2007 blind test. Clearly, molecular structure prediction, if not a solved problem, is one that has made great progress in recent years.

Finally, in the context of predictions relating to molecular crystals, we note that optimization methods have, of course, been very widely applied in the area of macro-molecular structure prediction; and

indeed the literature on protein structure modelling and prediction is probably more extensive than that on crystal structure prediction (see recent reviews^{123–126}). A wide range of techniques are used, many of which have a close affinity with those discussed in this article. Threading or fold recognition and secondary sequence predictions are essentially topological procedures, although very different in character from those used in crystal structure analyses. And there is a long tradition of the use of interatomic potentials or simpler cost functions in conjunction with global optimization techniques in predicting protein structures. A detailed discussion of protein structure optimization is beyond the scope of this article, but it is worth emphasizing that the difficulties and challenges in this field are often very similar to those discussed here.

SUMMARY AND PERSPECTIVE

So, how has the field responded to the ‘Maddox challenge’? Clearly, the horizons have greatly expanded owing to the development of a

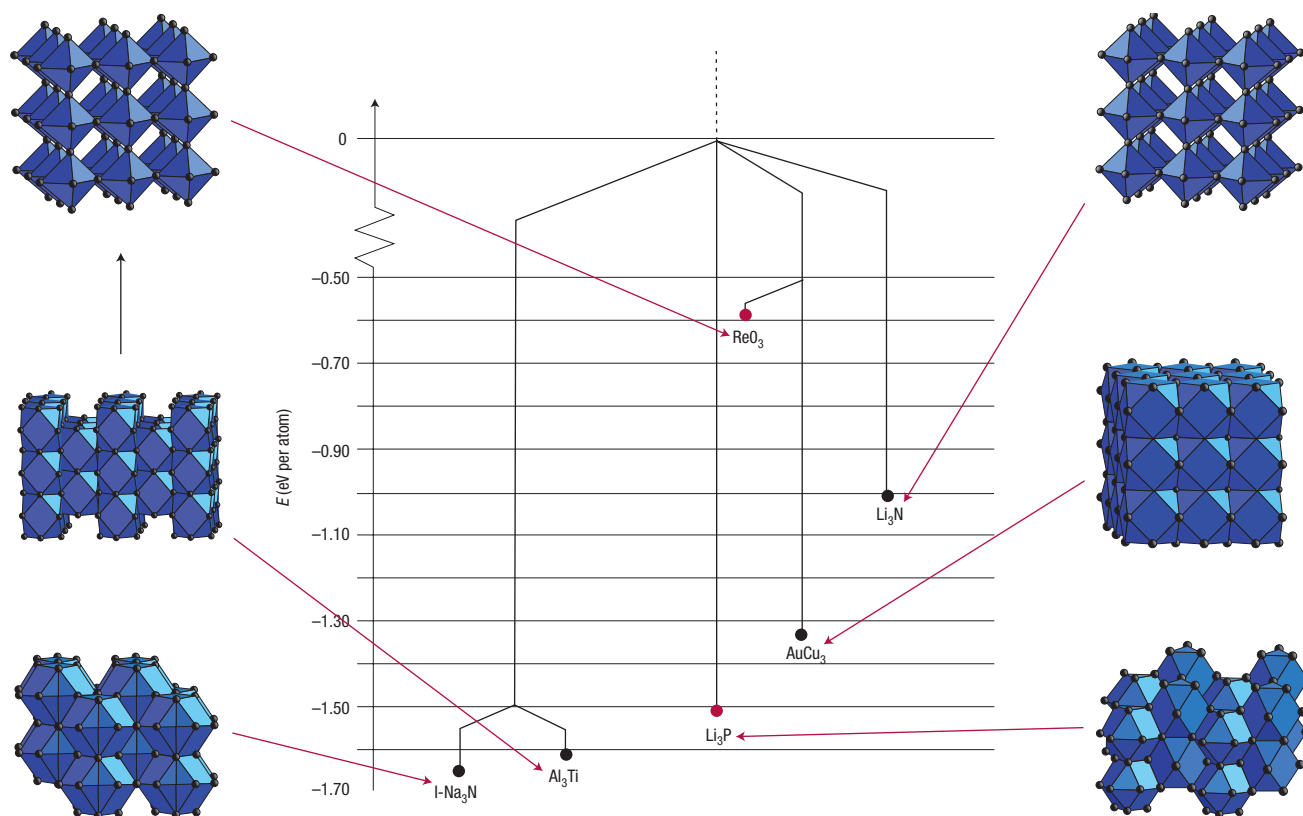


Figure 8 Candidate structures for Na_3N . Structures determined by global optimization are shown together with the barrier structure depicted as a tree graph, where new phases not found in a structural database are labelled with a Roman numeral prefix^{82,85,127}. Each 'root' on this diagram represents a local minimum, with the deepest root corresponding to the global minimum.

much wider range of techniques and more sophisticated and reliable interatomic potentials. The past 20 years have also seen an explosive growth in the application of electronic structure techniques, which may be used both to refine the predictions of methods based on interatomic potentials and to develop improved potential models. And, as in all computational science, the exponential growth in hardware performance means that within a few years, calculations that were quite unfeasible become routine. Structure prediction is not, however, yet routine. Although there has been substantial progress in developing high-quality interatomic potentials, there remain difficulties for several systems. In particular, accurate modelling of the thermodynamics of phase transformations is highly problematic; and there remain the long-standing issues of transferability of potentials from bulk to finite-size clusters. Moreover, not all these difficulties are removed by the use of quantum mechanical techniques to replace energy calculations based on interatomic potentials. There is also an increasing need for potentials and methods that can effectively handle 'hybrid' materials.

Other challenges relate to optimization procedures. For example, although considerable progress has been made in the use of phenotype operators, these have essentially been tailored for specific applications and are not generally applicable. In modelling cluster structures, for example, the phenotype may be biased towards the wrong target if cluster growth results in the emergence of new competing topologies. As already highlighted, there are plenty of techniques for finding the global minimum and many more applications of these methods, but perhaps one can argue that there is a need for algorithms designed specifically for searching for metastable states and saddle points. And we anticipate that there will be more applications of such

approaches and more investigations reported for polymorphs at finite temperatures.

Yet despite these and other difficulties, the use of computational techniques as a tool in structural studies of complex solids is increasingly standard, and genuine predictions are now being reported for several classes of material. Indeed, Jansen, Schön and co-workers^{127–130} (see for example Fig. 8) increasingly use computationally based techniques to guide new solid-state syntheses. Structure prediction will continue to challenge our ingenuity for many years to come.

doi:10.1038/nmat2321

References

1. Maddox, J. Crystals from first principles. *Nature* **335**, 201–201 (1988).
2. Parker, S. C. Prediction of mineral crystal structures. *Solid State Ionics* **8**, 179–186 (1983).
3. Catlow, C. R. A., Thomas, J. M., Parker, S. C. & Jefferson, D. A. Simulating silicate structures and the structural chemistry of pyroxenoids. *Nature* **295**, 658–662 (1982).
4. Ghosh, A., Sarkar, A. K. & Basu, A. N. The breathing shell model calculation of the relative stability of structure of alkali halide crystals. *J. Phys. C* **8**, 1332–1338 (1975).
5. Donnay, G., Donnay, J. D. H. & Takeda, H. Trioctahedral one-layer micas. II. Prediction of the structure from composition and cell dimensions. *Acta Cryst.* **17**, 1374–1381 (1964).
6. Catlow, C. R. A. & Price, G. D. Computer modelling of solid-state inorganic materials. *Nature* **347**, 243–248 (1990).
7. Catlow, C. R. A. *et al.* Computer modelling of inorganic materials. *Annu. Rep. Prog. Chem. A* **101**, 513–547 (2005).
8. Lewis, G. V. & Catlow, C. R. A. Potential models for ionic oxides. *J. Phys. C* **18**, 1149–1161 (1985).
9. Shannon, M. D., Casci, J. L., Cox, P. A. & Andrews, S. J. Structure of the two-dimensional medium-pore high-silica zeolite NU-87. *Nature* **353**, 417–420 (1991).
10. Kirkpatrick, S., Gelatt, J. C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
11. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).

12. 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).
13. Schön, 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. Engl.* **35**, 1287–1304 (1996).
14. Schön, J. C. & Jansen, M. Determination, prediction, and understanding of structures, using the energy landscapes of chemical systems. *Z. Kristallogr.* **216**, 307–325 (2001).
15. Mellot-Draznieks, C., Newsam, J. M., Gorman, A. M., Freeman, C. M. & Férey, G. De novo prediction of inorganic structures developed through automated assembly of secondary building units (AASBU method). *Angew. Chem. Int. Ed.* **39**, 2270–2275 (2000).
16. Mellot-Draznieks, C. *et al.* Computational design and prediction of interesting not-yet-synthesized structures of inorganic materials by using building unit concepts. *Chem. Eur. J.* **8**, 4102–4113 (2002).
17. Mellot-Draznieks, C., Girard, S. & Férey, G. R. Novel inorganic frameworks constructed from double-four-ring (D4R) units: Computational design, structures, and lattice energies of silicate, aluminophosphate, and gallophosphate candidates. *J. Am. Chem. Soc.* **124**, 15326–15335 (2002).
18. Mellot-Draznieks, C., Dutour, J. & Férey, G. R. Hybrid organic–inorganic frameworks: Routes for computational design and structure prediction. *Angew. Chem. Int. Ed.* **43**, 6290–6296 (2004).
19. Wales, D. J. & Scheraga, H. A. Review: Chemistry. Global optimization of clusters, crystals, and biomolecules. *Science* **285**, 1368–1372 (1999).
20. Wales, D. J. & Doyle, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
21. Hamad, S., Catlow, C. R. A., Woodley, S. M., Lago, S. & Mejias, J. A. Structure and stability of small TiO₂ nanoparticles. *J. Phys. Chem. B* **109**, 15741–15748 (2005).
22. Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers* (World Scientific, 1999).
23. Lloyd, L. D., Johnston, R. L. & Salhi, S. Strategies for increasing the efficiency of a genetic algorithm for the structural optimization of nanoscale clusters. *J. Comp. Chem.* **26**, 1069–1078 (2005).
24. Hartke, B. in *Applications of Evolutionary Computation in Chemistry*, 33–53 (Springer, 2004).
25. Johnston, R. L. Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries. *Dalton Trans.* **22**, 4193–4207 (2003).
26. Woodley, S. M. Engineering microporous architectures: Combining an evolutionary algorithm with predefined exclusion zones. *Phys. Chem. Chem. Phys.* **9**, 1070–1077 (2006).
27. Abraham, N. L. & Probert, M. I. J. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys. Rev. B* **73**, 224104 (2006).
28. Woodley, S. M. in *Applications of Evolutionary Computation in Chemistry*, 95–132 (Springer, 2004).
29. Harris, K. D. M., Johnston, R. L. & Habershon, S. in *Applications of Evolutionary Computation in Chemistry*, 55–94 (Springer, 2004).
30. Turner, G. W., Tedesco, E., Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. Implementation of Lamarckian concepts in a genetic algorithm for structure solution from powder diffraction data. *Chem. Phys. Lett.* **321**, 183–190 (2000).
31. Roberts, C., Johnston, R. L. & Wilson, N. T. A genetic algorithm for the structural optimization of Morse clusters. *Theor. Chem. Acc.* **104**, 123–130 (2000).
32. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
33. Woodley, S. M. & Catlow, C. R. A. Structure prediction of titania phases: Implementation of Darwinian versus Lamarckian concepts in an evolutionary algorithm. *Comp. Mater. Sci.* (in the press).
34. Pickard, C. J. & Needs, R. J. When is H₂O not water? *J. Chem. Phys.* **127**, 244503 (2007).
35. Wells, A. F. The geometrical basis of crystal chemistry. 1–4. *Acta Crystallogr.* **7**, 535–554; 842–853 (1954).
36. Smith, J. V. Enumeration of 4-connected 3-dimensional nets and classification of framework silicates. 1. Perpendicular linkage from simple hexagonal net. *Am. Mineral.* **62**, 703–709 (1977).
37. Smith, J. V. Enumeration of 4-connected 3-dimensional nets and classification of framework silicates. 2. Perpendicular and near-perpendicular linkages from 4.82, 3.122 and 4.6.12 nets. *Am. Mineral.* **63**, 960–969 (1978).
38. Smith, J. V. Enumeration of 4-connected 3-dimensional nets and classification of framework silicates. 3. Combination of helix, and zigzag, crankshaft and saw chains with simple 2d-nets. *Am. Mineral.* **64**, 551–562 (1979).
39. O'Keefe, M. & Hyde, B. G. *Crystal Structures I. Patterns and Symmetry* (Mineral. Soc. Am., 1996).
40. Treacy, M. M. J., Randall, K. H., Rao, S., Perry, J. A. & Chadi, D. J. Enumeration of periodic tetrahedral frameworks. *Z. Kristallogr.* **212**, 768–791 (1997).
41. Treacy, M. M. J., Rivin, I., Balkovsky, E., Randall, K. H. & Foster, M. D. Enumeration of periodic tetrahedral frameworks. II. Polynodal graphs. *Micropor. Mesopor. Mater.* **74**, 121–132 (2004).
42. Foster, M. D. *et al.* Chemically feasible hypothetical crystalline networks. *Nature Mater.* **3**, 234–238 (2004).
43. Dress, A. W. M., Huson, D. H. & Molnar, E. The classification of face-transitive periodic 3-dimensional tilings. *Acta Crystallogr. A* **49**, 806–817 (1993).
44. Delgado, O., Huson, D. & Zamorzaeva, E. The classification of 2-isohedral tilings of the plane. *Geometriae Dedicata* **42**, 43–117 (1992).
45. Friedrichs, O. D., Dress, A. W. M., Huson, D. H., Klinowski, J. & Mackay, A. L. Systematic enumeration of crystalline networks. *Nature* **400**, 644–647 (1999).
46. O'Keefe, M. Three-periodic nets and tilings: regular and related infinite polyhedra. *Acta Crystallogr. A* **64**, 425–429 (2008).
47. Winkler, B., Pickard, C. J., Milman, V. & Thimm, G. Systematic prediction of crystal structures. *Chem. Phys. Lett.* **337**, 36–42 (2001).
48. Le Bail, A. Inorganic structure prediction with GRINSF. *J. Appl. Cryst.* **38**, 389–395 (2005).
49. Tajima, N., Tsuzuki, S., Tanabe, K., Aoki, K. & Hirano, T. First principles prediction of crystal structures of CO₂. *Electron. J. Theor. Chem.* **2**, 139–148 (1997).
50. Arikawa, T., Tajima, N., Tsuzuki, S., Tanabe, K. & Hirano, T. A possible crystal-structure of 1,2-dimethoxyethane—prediction based on a lattice variable molecular-dynamics. *Theochem: J. Mol. Struct.* **339**, 115–124 (1995).
51. Hirano, T., Tsuzuki, S., Tanabe, K. & Tajima, N. Totally ab initio prediction of the structures of CO₂ molecular crystal. *Chem. Lett.* **12**, 1073–1074 (1995).
52. Chaka, A. M., Zaniwski, R., Youngs, W., Tessier, C. & Klopman, G. Predicting the crystal structure of organic molecular materials. *Acta Crystallogr. B* **52**, 165–183 (1996).
53. Ammon, H. L., Du, Z. Y., Holden, J. R. & Paquette, L. A. *Acta Crystallogr. B* **50**, 216–220 (1994).
54. Van Eijck, B. P. & Kroon, J. Upack program package for crystal structure prediction: force fields and crystal structure generation for small carbohydrate molecules. *J. Comput. Chem.* **20**, 799–812 (1999).
55. Holden, J. R., Du, Z. Y. & Ammon, H. L. Prediction of possible crystal-structures for C-containing, H-containing, N-containing, O-containing and F-containing organic-compounds. *J. Comput. Chem.* **14**, 422–437 (1993).
56. Aakeroy, C. B., Nieuwenhuyzen, M. & Price, S. L. Three polymorphs of 2-amino-5-nitropyrimidine: Experimental structures and theoretical predictions. *J. Am. Chem. Soc.* **120**, 8986–8993 (1998).
57. Beyer, T. & Price, S. L. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *J. Phys. Chem. B* **104**, 2647–2655 (2000).
58. Price, S. L. & Wibley, K. S. Predictions of crystal packings for uracil, 6-azauracil, and allopurinol: The interplay between hydrogen bonding and close packing. *J. Phys. Chem. A* **101**, 2198–2206 (1997).
59. Beyer, T., Day, G. M. & Price, S. L. The prediction, morphology, and mechanical properties of the polymorphs of paracetamol. *J. Am. Chem. Soc.* **123**, 5086–5094 (2001).
60. Gdanitz, R. J. Prediction of molecular-crystal structures by Monte-Carlo simulated annealing without reference to diffraction data. *Chem. Phys. Lett.* **190**, 391–396 (1992).
61. Price, S. L. From crystal structure prediction to polymorph prediction: interpreting the crystal energy landscape. *Phys. Chem. Chem. Phys.* **10**, 1996–2009 (2008).
62. Dunitz, J. D. & Gavezzotti, A. Molecular recognition in organic crystals: Directed intermolecular bonds or nonlocalized bonding? *Angew. Chem. Int. Ed.* **44**, 1766–1787 (2005).
63. Desiraju, G. R. Crystal engineering: A holistic view. *Angew. Chem. Int. Ed.* **46**, 8342–8356 (2007).
64. Raiteri, P., Martoňák, R. & Parrinello, M. Exploring polymorphism: The case of benzene. *Angew. Chem. Int. Ed.* **44**, 3769–3773 (2005).
65. Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. Predicting crystal structures with data mining of quantum calculations. *Phys. Rev. Lett.* **91**, 135503 (2003).
66. Fischer, C. C., Tibbets, K. J., Morgan, D. & Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nature Mater.* **5**, 641–646 (2006).
67. Hofmann, D. W. M. & Apostolakis, J. Crystal structure prediction by data mining. *J. Mol. Struct.* **647**, 17–39 (2003).
68. Schön, J. C., Čančarević, Ž. P., Hannermann, A. & Jansen, M. Free enthalpy landscape of SrO. *J. Chem. Phys.* **128**, 194712 (2008).
69. Martoňák, R., Laio, A. & Parrinello, M. Predicting crystal structures: The Parrinello–Rahman method revisited. *Phys. Rev. Lett.* **90**, 75503 (2003).
70. Schön, J. C., Pentin, I. V. & Jansen, M. Ab initio computation of the low-temperature phase diagrams of the alkali metal iodide-bromides: MBr_{1-x} (0 ≤ x ≤ 1), where M = Li, Na, K, Rb, or Cs. *J. Phys. Chem. B* **111**, 3943–3952 (2007).
71. Ceriani, C. *et al.* Molecular dynamics simulation of reconstructive phase transitions on an anhydrous zeolite. *Phys. Rev. B* **70**, 113403 (2004).
72. Brown, I. D. *Computer Modelling in Inorganic Crystallography* Ch. 2 (ed. Catlow, C. R. A.) (Academic, 1994).
73. Lacorre, P., Pannetier, J., Hoppe, R., Averdunk, F. & Férey, G. Crystal and magnetic-structures of LiCoF₄—the 1st compound with a dirutile structure. *J. Solid State Chem.* **79**, 1–11 (1989).
74. Freeman, C. M. & Catlow, C. R. A. Structure predictions in inorganic solids. *J. Chem. Soc. Chem. Commun.* 89–91 (1992).
75. Freeman, C. M., Newman, J. M., Levine, S. M. & 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).
76. Woodley, S. M., Battle, P. D., Gale, J. D. & Catlow, C. R. A. The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.* **1**, 2535–2542 (1999).
77. Reinaudi, L., Carbonio, R. E. & Leiva, E. P. M. Inclusion of symmetry for the enhanced determination of crystalline structures from powder diffraction data using simulated annealing. *J. Chem. Soc. Chem. Commun.* 255–256 (1998).
78. Reinaudi, L., Leiva, E. P. M. & Carbonio, R. E. Simulated annealing prediction of the crystal structure of ternary inorganic compounds using symmetry restrictions. *J. Chem. Soc., Dalton Trans.* **23**, 4258–4262 (2000).
79. Bush, T. S., Catlow, C. R. A. & Battle, P. D. Evolutionary programming techniques for predicting inorganic crystal-structures. *J. Mater. Chem.* **5**, 1269–1272 (1995).
80. Doll, K., Schön, J. C. & Jansen, M. Global exploration of the energy landscape of solids on the ab initio level. *Phys. Chem. Chem. Phys.* **9**, 6128–6133 (2007).
81. Schön, J. C. & Jansen, M. Determination of candidate structures for Lennard-Jones-crystals through cell optimization. *Ber. Bunsenges Phys. Chem.* **98**, 1541–1544 (1994).
82. Jansen, M. & Schön, J. C. Structure candidates for the alkali metal nitrides. *Z. Anorg. Allg. Chem.* **624**, 533–540 (1998).
83. Putz, H., Schön, J. C. & Jansen, M. Investigation of the energy landscape of Mg₂OF₂. *Comput. Mater. Sci.* **11**, 309–322 (1998).
84. Wevers, M. A. C., Schön, J. C. & Jansen, M. Determination of structure candidates of simple crystalline AB₂ systems. *J. Solid State Chem.* **136**, 233–246 (1998).
85. Schön, J. C., Wevers, M. A. C. & Jansen, M. Prediction of high pressure phases in the systems Li₃N, Na₃N, (Li₃Na)₃N, Li₃S and Na₃S. *J. Mater. Chem.* **11**, 69–77 (2001).
86. Ciobanu, C. V., Chuang, F. C. & Lytle, D. E. On the structure of the Si(103) surface. *Appl. Phys. Lett.* **91**, 171909 (2007).
87. Briggs, R. M. & Ciobanu, C. V. Evolutionary approach for finding the atomic structure of steps on stable crystal surfaces. *Phys. Rev. B* **75**, 195415 (2007).
88. Kasuya, A. *et al.* Ultra-stable nanoparticles of CdSe revealed from mass spectrometry. *Nature Mater.* **3**, 99–102 (2004).
89. Hamad, S., Cristol, S. & Catlow, C. R. A. Simulation of the embryonic stage of ZnS formation from aqueous solution. *J. Am. Chem. Soc.* **127**, 2580–2590 (2005).
90. Wakisaka, A. Nucleation in alkali metal chloride solution observed at the cluster level. *Faraday Discuss.* **136**, 299–308 (2007).

91. Burnin, A. & Belbruno, J. J. Zn_nS_m^+ cluster production by laser ablation. *Chem. Phys. Lett.* **362**, 341–348 (2002).
92. Whetten, R. L. Alkali-halide nanocrystals. *Acc. Chem. Res.* **26**, 49–56 (1993).
93. Hamad, S., Catlow, C. R. A., Spano, E., Matxain, J. M. & Ugalde, J. M. Structure and properties of ZnS nanoclusters. *J. Phys. Chem. B* **109**, 2703–2709 (2005).
94. Al-Sunaidi, A. A., Sokol, A. A., Catlow, C. R. A. & Woodley, S. M. Structures of zinc oxide nanoclusters: As found by evolutionary algorithm techniques. *J. Phys. Chem. C* (in the press).
95. Hamad, S. & Catlow, C. R. A. Computational study of the relative stabilities of ZnS clusters, for sizes between 1 and 4 nm. *J. Cryst. Growth* **294**, 2–8 (2006).
96. Michaelian, K. Evolving few-ion clusters of Na and Cl. *Am. J. Phys.* **66**, 231–240 (1998).
97. Wootton, A. & Harrowell, P. Inorganic nanotubes stabilized by ion size asymmetry: Energy calculations for AgI clusters. *J. Phys. Chem. B* **108**, 8412–8418 (2004).
98. Roberts, C. & Johnston, R. L. Investigation of the structures of MgO clusters using a genetic algorithm. *Phys. Chem. Chem. Phys.* **3**, 5024–5034 (2001).
99. Woodley, S. M., Sokol, A. A. & Catlow, C. R. A. Structure prediction of inorganic nanoclusters with a predefined architecture using a genetic algorithm. *Z. Anorg. Allg. Chem.* **630**, 2343–2353 (2004).
100. Flikkema, E. & Bromley, S. T. Dedicated global optimization search for ground state silica nanoclusters: $(\text{SiO}_2)_N$ ($N = 6–12$). *J. Phys. Chem. B* **108**, 9638–9645 (2004).
101. Shevlin, S. A. *et al.* Structure, optical properties and defects in nitride (III–V) nanoscale cage clusters. *Phys. Chem. Chem. Phys.* **10**, 1944–1959 (2008).
102. Michaelian, K., Rendón, N. & Garzón, I. L. Structure and energetics of Ni, Ag, and Au nanoclusters. *Phys. Rev. B* **60**, 20003–2010 (1999).
103. Ferrando, R., Fortunelli, A. & Johnston, R. L. Searching for the optimum structures of alloy nanoclusters. *Phys. Chem. Chem. Phys.* **10**, 640–649 (2008).
104. Paz-Borbon, L. O., Johnston, R. L., Barcaro, G. & Fortunelli, A. Structural motifs, mixing, and segregation effects in 38-atom binary clusters. *J. Chem. Phys.* **128**, 134517 (2008).
105. Deem, M. W. & Newsam, J. M. Determination of 4-connected framework crystal-structures by simulated annealing. *Nature* **342**, 260–262 (1989).
106. Deem, N. W. & Newsam, J. M. Framework crystal-structure solution by simulated annealing: test application to known zeolite structures. *J. Am. Chem. Soc.* **114**, 7189–7198 (1992).
107. Falcioni, M. & Deem, M. W. A biased Monte Carlo scheme for zeolite structure solution. *J. Chem. Phys.* **110**, 1754–1766 (1999).
108. Akporiaye, D. E. *et al.* UiO-7: A new aluminophosphate phase solved by simulated annealing and high-resolution powder diffraction. *J. Phys. Chem.* **100**, 16641–16646 (1996).
109. Boisen, M. B., Gibbs, G. V. & Bukowski, M. S. T. Framework silica structures generated using simulated annealing with a potential-energy function-based on an $\text{H}_4\text{Si}_2\text{O}_7$ molecule. *Phys. Chem. Miner.* **21**, 269–284 (1994).
110. Teter, D. M., Gibbs, G. V., Boisen, M. B., Allan, D. C. & Teter, M. P. First-principles study of several hypothetical silica framework structures. *Phys. Rev. B* **52**, 8064–8073 (1995).
111. Boisen, M. B., Gibbs, G. V., O’Keeffe, M. & Bartelme, K. L. A generation of framework structures for the tectosilicates using a molecular-based potential energy function and simulated annealing strategies. *Micropor. Mesopor. Mater.* **29**, 219–266 (1999).
112. Woodley, S. M., Catlow, C. R. A., Battle, P. D. & Gale, J. D. The prediction of close packed and porous inorganic crystal structures. *Acta Cryst. A* **58**, C196 (2002).
113. Woodley, S. M. Prediction of inorganic crystal framework structures. Part II: using a genetic algorithm and a direct approach to exclusion zones. *Phys. Chem. Chem. Phys.* **6**, 1823–1829 (2004).
114. Woodley, S. M., Battle, P. D., Gale, J. D. & Catlow, C. R. A. Prediction of inorganic crystal framework structures. Part I: Using a genetic algorithm and an indirect approach to exclusion zones. *Phys. Chem. Chem. Phys.* **6**, 1815–1822 (2004).
115. Zwijnenburg, M. A., Cora, F. & Bell, R. G. Dramatic differences between the energy landscapes of SiO_2 and SiS_2 zeotype materials. *J. Am. Chem. Soc.* **129**, 12588–12589 (2007).
116. Carrasco, J., Illas, F. & Bromley, S. T. Ultralow-density nanocage-based metal-oxide polymorphs. *Phys. Rev. Lett.* **99**, 235502 (2007).
117. Lewis, D. W., Catlow, C. R. A., Thomas, J. M., Willock, D. J. & Hutchings, G. J. De novo design of structure-directing agents for the synthesis of microporous solids. *Nature* **382**, 604–606 (1996).
118. Sankar, G. *et al.* Structure of templated microcrystalline DAF-5 ($\text{Co}_{0.28}\text{Al}_{0.72}\text{PO}_4\text{C}_{10}\text{H}_{20}\text{N}_2$) determined by synchrotron-based diffraction methods. *Chem. Commun.* **1**, 117–118 (1998).
119. Hulme, A. T., Price, S. L. & Tocher, D. A. A new polymorph of 5-fluorouracil found following computational crystal structure predictions. *J. Am. Chem. Soc.* **127**, 1116–1117 (2005).
120. Hamad, S., Moon, C., Catlow, C. R. A., Hulme, A. T. & Price, S. L. Kinetic insights into the role of the solvent in the polymorphism of 5-fluorouracil from molecular dynamics simulations. *J. Phys. Chem. B* **110**, 3323–3329 (2006).
121. Lommerse, J. P. M. *et al.* A test of crystal structure prediction of small organic molecules. *Acta Crystallogr. B* **56**, 697–714 (2000).
122. Neumann, M. A., Leusen, F. J. J. & Kendrick, J. A major advance in crystal structure prediction. *Angew. Chem. Int. Ed.* **47**, 2427–2430 (2008).
123. Moulton, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **15**, 285–289 (2005).
124. Petrey, D. & Honig, B. Protein structure prediction: Inroads to biology. *Mol. Cell* **20**, 811–819 (2005).
125. Floudas, C. A., Fung, H. K., McAllister, S. R., Monnigmann, M. & Rajgaria, R. Advances in protein structure prediction and de novo protein design: A review. *Chem. Eng. Sci.* **61**, 966–988 (2006).
126. Zhang, Y. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.* **18**, 342–348 (2008).
127. Jansen, M. A concept for synthesis planning in solid-state chemistry. *Angew. Chem. Int. Ed.* **41**, 3746–3766 (2002).
128. Jansen, M. in *Turning Points in Solid-State, Materials and Surface Science* (eds Harris, K. M. & Edwards, P.) 22–50 (Royal Society of Chemistry, 2008).
129. Cancarevic, Z. P., Schön, J. C. & Jansen, M. Stability of alkali metal halide polymorphs as a function of pressure. *Chem. Asian J.* **3**, 561–572 (2008).
130. Liebold-Ribeiro, Y., Fischer, D. & Jansen, M. Experimental substantiation of the ‘Energy landscape concept’ for solids: Synthesis of a new modification of LiBr . *Angew. Chem. Int. Ed.* **47**, 4428–4431 (2008).

Acknowledgements

We thank R. G. Bell, S. Hamad, M. Jansen, S. L. Price, J. C. Schön, A. A. Sokol and J. M. Thomas for discussions, and the EPSRC for financial support via the Portfolio Partnership grant EP/D504872.