# Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications

Qiang Zhu,[a]* Artem R. Oganov,[a,b] Colin W. Glass[c] and Harold T. Stokes[d]

[a]Department of Geosciences, Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York, USA, [b]Geology Department, Moscow State University, Moscow, Russia, [c]High Performance Computing Center Stuttgart (HLRS), Germany, and [d]Department of Physics and Astronomy, Brigham Young University, Provo, Utah, USA

Correspondence e-mail:
qiang.zhu@stonybrook.edu

Evolutionary crystal structure prediction proved to be a powerful approach for studying a wide range of materials. Here we present a specifically designed algorithm for the prediction of the structure of complex crystals consisting of well defined molecular units. The main feature of this new approach is that each unit is treated as a whole body, which drastically reduces the search space and improves the efficiency, but necessitates the introduction of new variation operators described here. To increase the diversity of the population of structures, the initial population and part ($\sim 20\%$) of the new generations are produced using space-group symmetry combined with random cell parameters, and random positions and orientations of molecular units. We illustrate the efficiency and reliability of this approach by a number of tests (ice, ammonia, carbon dioxide, methane, benzene, glycine and butane-1,4-diammonium dibromide). This approach easily predicts the crystal structure of methane A containing 21 methane molecules (105 atoms) per unit cell. We demonstrate that this new approach also has a high potential for the study of complex inorganic crystals as shown on examples of a complex hydrogen storage material $Mg(BH_4)_2$ and elemental boron.

## 1. Introduction

Structure is the most important piece of information about a material as it determines most of its physical properties. It was long believed that crystal structures are fundamentally unpredictable (Maddox, 1988; Gavezzotti, 1994). However, the situation began to change dramatically in the last decade. As the stable structure corresponds to the global minimum of the free energy, several global optimization algorithms have been devised and used with some success for crystal structure prediction (CSP) – for instance, simulated annealing (Pannetier et al., 1990; Schon & Jansen, 1996), metadynamics (Martonák et al., 2003), evolutionary algorithms (Oganov & Glass, 2006; Oganov et al., 2011), random sampling (Freeman et al., 1993), basin hopping (Wales & Doye, 1997), minima hopping (Goedecker, 2004) and data mining (Curtarolo et al., 2003). For inorganic crystals, in many cases it is already now possible to predict the stable structure at arbitrary external pressure. Towards the ambition of designing novel materials prior to their synthesis in the laboratory, reliable and efficient prediction of the structure of more complex (in particular, molecular) crystals becomes imperative.

Molecular crystals are extremely interesting because of their applications as pharmaceuticals, pigments, explosives and metal-organic frameworks (Price, 2004; Baburin et al., 2008). The periodically conducted blind tests of organic crystal structure prediction, organized by the Cambridge Crystal-

lographic Data Centre (CCDC), have been the focal point for this community and they reflect steady progress in the field (Lommerse *et al.*, 2000; Motherwell *et al.*, 2002; Day *et al.*, 2005, 2009; Bardwell *et al.*, 2011). The tests show that it is now possible to predict the packing of a small number of rigid molecules, provided there are cheap force fields accurately describing the intermolecular interactions. In these cases, the efficiency of the search for the global minimum on the energy landscape is not crucial. However, if one has to use expensive *ab initio* total energy calculations or study systems with a large number of degrees of freedom (many molecules, especially if they have conformational flexibility), the number of possible structures becomes astronomically large and efficient search techniques become critically important.

In addition, the nature of weak chemical interactions means that commonly molecules have a wide variety of ways of packing with lattice energies within a few kJ mol$^{-1}$ of the most stable structure. Thus, prediction of such large structures is certainly a challenge, especially if the number of trial struc-

tures has to be kept low to enable practical *ab initio* structure predictions. Recent pioneering works (Kim *et al.*, 2009; Raiteri *et al.*, 2005; Day, 2011), in particular using metadynamics (Raiteri *et al.*, 2005), offer inspiring examples of this.

Compared with other methods, evolutionary algorithms have a special advantage. Exploring the energy surface, such algorithms arrive at the global minimum by a series of intelligently designed moves, involving self-learning and self-improvement of the population of crystal structures (Oganov *et al.*, 2011). Our USPEX (Universal Structure Predictor: Evolutionary Xtallography) code (Oganov & Glass, 2006, 2008; Glass *et al.*, 2006; Oganov & Valle, 2009; Lyakhov *et al.*, 2010) proved to be extremely efficient and reliable for atomic crystals, and here we present an extension of this algorithm to complex crystals composed of well defined units. In the following sections we will mainly discuss molecular crystals. Crystals containing complex ions and clusters can be equally well studied using the methodology developed here, as we show by two tests on challenging systems.
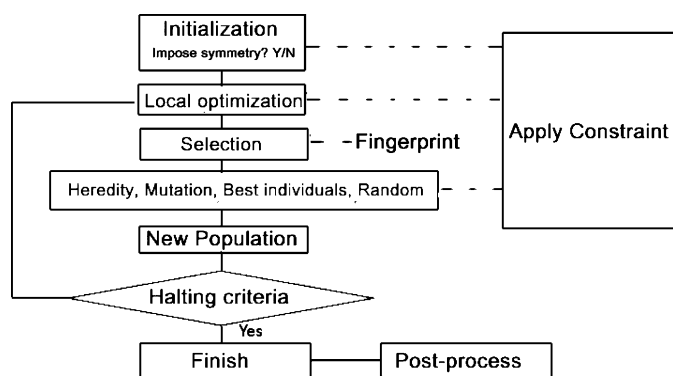
## 2. Methodology

Compared with the prediction of atomic structures, there are several additional considerations to be taken into account for molecular crystals:

(i) A typical unit cell contains many more atoms than a usual inorganic structure, which means an explosion of computing costs if all of these atoms are treated independently.

(ii) Molecules are bound by weak forces, such as the van der Waals (vdW) interactions, and the intermolecular distances are typically larger than those in atomic crystals, which leads to the availability of large empty space.

(iii) Most of the molecular compounds are thermodynamically less stable than the simpler molecular compounds from which they can be obtained (such as $H_2O$, $CO_2$, $CH_4$,
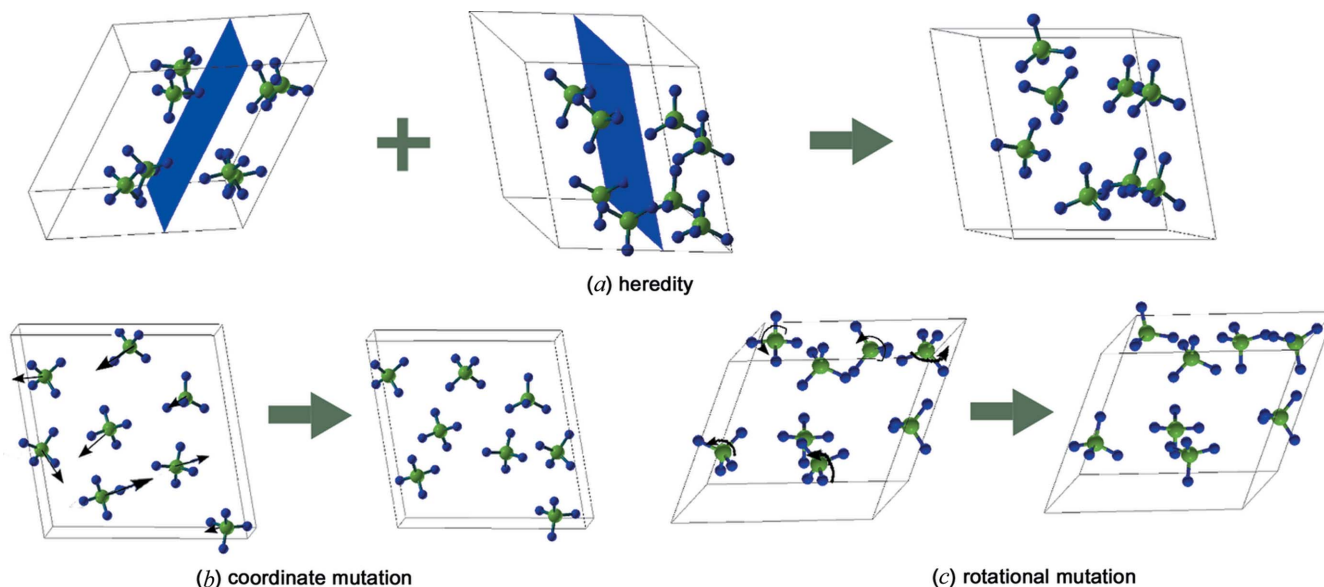


**Figure 1**
Illustration of the constrained evolutionary algorithm.



*(a)* heredity

*(b)* coordinate mutation

*(c)* rotational mutation

**Figure 2**
Illustration of the variation operators: (*a*) heredity; (*b*) coordinate mutation; (*c*) rotational mutation.

NH$_3$). This means that a fully unconstrained global optimization approach in many cases will produce a mixture of these simple molecules, which are of little interest to the organic chemist. To study the packing of the actual molecules of interest, it is necessary to fix the intramolecular connectivity.

(iv) Crystal structures tend to be symmetric, and the distribution of structures over symmetry groups is extremely uneven (Brock & Dunitz, 1994). For example, 35% of inorganic and 45% of organic materials have the point group $2/m$. Compared with inorganic crystals, there is a stronger preference of organic crystals to a smaller number of space groups. Over 80% of organic crystals are found to possess space groups: $P2_1/c$ (36.59%), $P\bar{1}$ (16.92%), $P2_12_12_1$ (11.00%), $C2/c$ (6.95%), $P2_1$ (6.35%) and $Pbca$ (4.24%; Baur & Kassner, 1992).

The first two points indicate that the search space is huge. If we start to search for the global minimum with randomly generated structures it is very likely that most of the time will be spent on exploring uninteresting disordered structures far away from the global minimum. Fortunately, the last two points suggest a way to improve the efficiency. Point (iii) implies that the true thermodynamic ground state corresponding to most organic compositions is a mixture of simpler molecules, which is of little interest. The truly interesting problem, packing of the pre-formed molecules, can be solved by *constrained global optimization* – finding the most stable packing of molecules with fixed bond connectivity. This will not only make the global optimization process meaningful, but at the same time will simplify it, leading to a drastic reduction
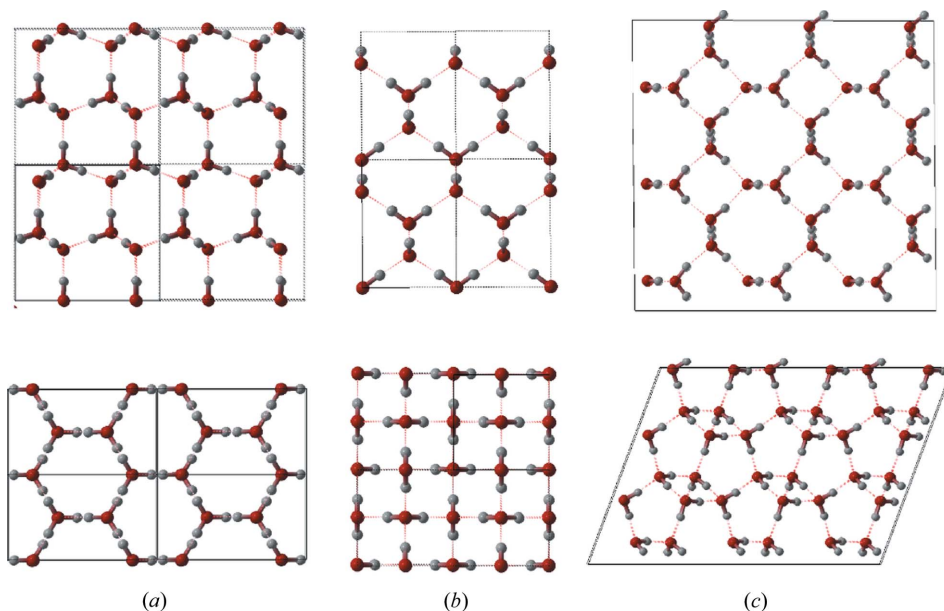


**Figure 4**
Prediction of the crystal structure of ice at 0 GPa. The lowest energy at each generation is shown relative to the ground state. Each generation contains 30 structures. The ground-state structure ice XI was found at the fourth generation. We also found cubic ice and bct4-like ice in the same calculation.

of the number of degrees of freedom and of the search space. Structure prediction (global optimization) must involve relaxation (local optimization) of all structures, and fixing intramolecular bond connectivity has the added benefit of making structure relaxations cheaper and more robust. Depending on their chemical nature, these molecules shall be treated as fully or partly rigid bodies during the action of evolutionary variation operators and local optimization.

Another improvement of the efficiency is achieved by using symmetry in the random generation of new structures – a population of symmetric structures is usually more diverse than a set of fully random (often disordered) structures. Diversity of the population of structures is essential for the success and efficiency of evolutionary simulations.

We have successfully implemented the adapted evolutionary algorithm in the USPEX code. Briefly, our procedure is as follows (as shown in Fig. 1).

(i) The initial structures are usually generated randomly, with randomly selected space groups. First, we randomly pick one of 230 space groups, and set up a Bravais cell according to the prespecified initial volume with random cell parameters consistent with the space group. Then one molecule is randomly placed on a general Wyckoff position,



**Figure 3**
Ice polymorphs at 1 atm found by USPEX. (*a*) Ice XI, derived from ice Ih, space group $Cmc2_1$, $a = 4.338$, $b = 7.554$, $c = 7.094$ Å, O1(0,0.6651,0.0623), O2(0.5,0.8328,−0.0622), H1(0,0.6638,0.2045), H2(0,0.5373,−0.0191), H3(0.6865,−0.2329,−0.0140); (*b*) tetragonal phase, derived from ice Ic, space group $I4_1md$, $a = 4.415$, $c = 6.008$ Å, O(0,0.5,0.0006), H(0.1839,0.5,0.1000); (*c*) bct-4 like ice, space group $Cm$, $a = 4.472$, $b = 10.451$, $c = 5.744$ Å, $\beta = 111.3°$, O1(0,0,0), O2(0.3691,0,0.7197), O3(0.6647,0.3192,0.3605), H1(0.7683,0,0.8871), H2(0.1317,0,0.8908), H3(0.4705, 0.2691, 0.3567), H4(0.0923,0.8787,0.2133), H5(0.3018,0.0751,0.6030).
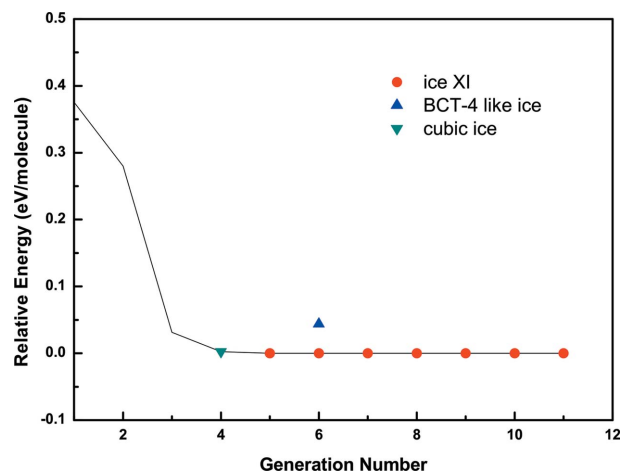
and is multiplied by space-group operations. If two or more symmetry-related molecules are found close to each other, we merge them into one molecule that sits on a special Wyckoff position that has averaged coordinates of the molecular center and averaged orientational vectors (or random, when averaging gives zero). Adding new molecular sites one by one, until the correct number of molecules is reached, we obtain a random symmetric structure. During this process we also make sure that no molecules overlap or sit too close to each other. All produced unit cell shapes are checked and, if necessary, transformed to maximally orthogonal shapes (Oganov & Glass, 2008).

(ii) Structure relaxation is carried out stepwise from low to high precision. At the initial stages we employ the SIESTA code (Soler *et al.*, 2002) for first-principles simulations, which allows the constrained geometry relaxation. As an option we can use the DMACRYS code (Price *et al.*, 2010) for classical calculations. We note that SIESTA provides $Z$-matrix representation for the molecules (Hoft *et al.*, 2006), enabling the specification of the molecular geometry and its internal degrees of freedom (important when dealing with conformationally flexible molecules). For the final stages of relaxation,

we can keep the molecules fully or partly rigid, or allow their complete relaxation (in the latter case, such codes as GULP, Gale & Rohl, 2003, and VASP, Kresse & Furthmuller, 1996, are also supported in USPEX). It is a good strategy to relax the structures in SIESTA with constrained molecular geometry at the beginning stage and then fully relax them using VASP, and here we adopt this strategy for all studied systems. It is well known that DFT within local and semilocal approximations, such as the LDA or GGA, cannot describe vdW dispersion interactions well (*e.g.* Li *et al.*, 2010) and we therefore used the GGA + D approach that includes a damped dispersion correction (Grimme, 2006); this approach is known to work well for molecular crystals.

(iii) At the end of each generation, all structures in the generation are compared using their fingerprints (Oganov & Valle, 2009) and all non-identical structures are ranked by their (free) energies or (if the calculation is done at $T = 0$ K, as we do here) enthalpies. There is an important technical aspect: intramolecular contributions are identical for all different packings of the same molecule and thus decrease the discriminatory power of the fingerprint function. Therefore, we remove the intramolecular distances from the computation of the fingerprint function when dealing with crystals made of molecules with no conformational flexibility.

A certain percentage of higher-energy structures in the population are discarded and the rest participate in creating the next generation using the variation operators detailed below.

To ensure properly constrained global optimization, we not only generate the structures made of the desired molecules, but also check that the bond connectivity has not changed after relaxation – structures with altered connectivity graphs are discarded.

(iv) Child structures (new generation) are produced from parent structures (old generation) using one of the following variation operators: (*a*) heredity, (*b*) permutation, (*c*) coordinate mutation are the same as in atomic crystal structures (Oganov & Glass, 2006; Lyakhov *et al.*, 2010), with the only difference that variation operators act on the geometric centers of the molecules and their orientations, *i.e.* whole molecules rather than single atoms are considered as the minimum building blocks. Since molecules cannot be considered as spheri-
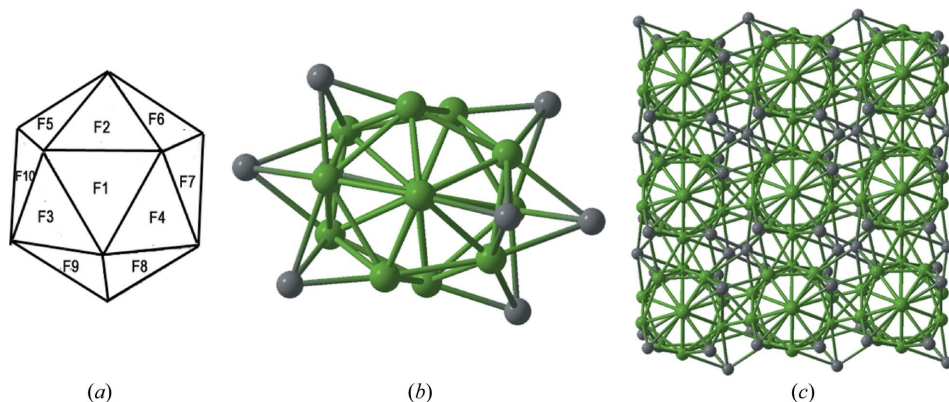


**Figure 5**
Structures of methane: (*a*) illustration of possible sites around the icosahedra, (*b*) 21-molecule rhombohedral methane, with F1, F2, F3, F4 sites occupied; (*c*) view of the icosahedron packing in the rhombohedral methane (space group: $R\bar{3}$). Two C sublattices are marked by different colors in a 21-molecule cell (non-icosahedral, green; icosahedral, grey).
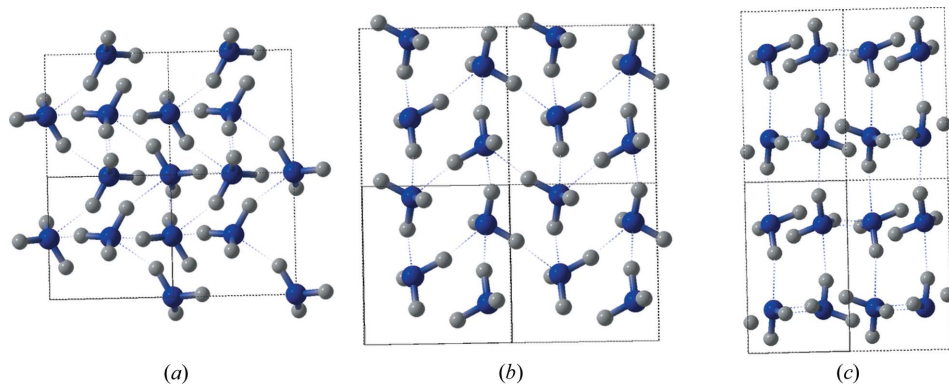


**Figure 6**
Crystal structures of ammonia. (*a*) $P2_13$ phase (stable at 1–6 GPa, $Z = 4$); (*b*) $P2_1/c$ phase (stable at 6–8.5 GPa, $Z = 4$); (*c*) $P2_12_12_1$ phase (stable at 8.5–60 GPa, $Z = 4$).

cally symmetric point particles, additional variation operators must be introduced: (*d*) rotational and conformational mutation of the whole molecules, (*e*) molecular softmutation – a hybrid operator of coordinate and rotational mutation. Fig. 2 shows how variation operators work in our algorithm. Below we describe how these variation operators were used in our tests.

**Heredity**: This operator cuts planar slices from each individual and combines these to produce a child structure. In heredity, each molecule is represented by its geometric center (Fig. 2*a*) and orientation. From each parent we cut (parallel to a randomly selected coordinate plane of the unit cell) a slab of random thickness (within bounds of 0.25–0.75 of the cut lattice vector) at a random height in the cell. If the total number of molecules of each type obtained from combining the slabs does not match the desired number of molecules, a corrector step is performed: molecules in excess are removed while molecules in shortage are added; molecules with a higher local degree of order have higher probability to be added and lower probability to be removed. This is equivalent to our original implementation of heredity for atomic crystals (Oganov & Glass, 2006)

**Permutation**: this operator swaps chemical identity in randomly selected pairs of molecules.

**Coordinate mutation**: All the centers of molecules are displaced in random directions, the distance for this movement for molecule *i* being picked from a zero-mean Gaussian distribution with $\sigma$ defined as

$$\sigma_i = \sigma_{max} \frac{\Pi_{max} - \Pi_i}{\Pi_{max} - \Pi_{min}}, \qquad (1)$$

where $\Pi$ is the local order of the molecule. Thus, molecules with higher order are perturbed less than molecules with low order (Fig. 2*b*). We calculate the local order parameter of each molecule from its fingerprint (Oganov & Valle, 2009) in the computation of which only the centers of all the molecules are
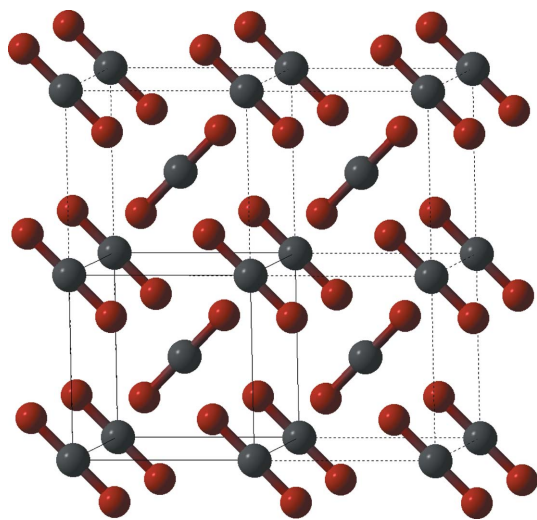


**Figure 7**
Crystal structure of $CO_2$-(II) (space group: $P42/mnm$, $Z = 2$).

used. In the tests described here $\sigma_{max}$ represents of the order of a typical intermolecular distance.

**Rotational mutation**: A certain number of randomly selected molecules are rotated by random angles (Fig. 2*c*). For rigid molecules there are only three varibles to define the orientation of the molecules. For flexible molecules we also allow the mutation of torsional angles of flexible groups. A large rotation can have a marked effect on global optimization, helping the system to jump out of the current local minimum and find optimal orientational ordering.

**Softmutation**: This powerful operator, introduced first for atomic crystals (Lyakhov *et al.*, 2010), involves atomic displacements along the softest mode eigenvectors, or a random linear combination of the softest eigenvectors. In the context of molecular crystals one must operate with rigid-unit modes and this operator becomes a hybrid operator, combining rotational and coordinate mutations. In this case the eigenvectors are calculated first and then projected onto the translational and rotational degrees of freedom of each molecule, and the resulting changes of molecular positions and orientations are applied thus preserving the rigidity of the fixed intramolecular degrees of freedom. To calculate efficiently the normal modes we construct the dynamical matrix from bond hardness coefficients (Lyakhov *et al.*, 2010). The same structures can be softmutated many times, each time along the eigenvector of a new mode.

At the end of the selection, the best individuals in the last generation (usually up to 5) are kept. To maintain diversity of the population, some fraction (usually 15–30%) of the population is randomly generated with symmetry. While simple random generation does not improve the diversity, the use of symmetry does allow a diverse set of structures to be produced.

(v) The simulation is stopped once a predefined halting criterion is met. The lowest-energy structures found in USPEX are then carefully relaxed with higher precision using the same level of theory: the all-electron projector-augmented wave (PAW) method (Blochl, 1994), as implemented in the VASP code (Kresse & Furthmuller, 1996), at the level of generalized gradient approximation (GGA; Perdew *et al.*, 1996) for inorganic systems; or dispersion-corrected GGA + D (Grimme, 2006) approximation for organic crystals. We used the plane-wave kinetic energy cutoff of 550 eV and the Brillouin zone was sampled with a resolution of $2\pi \times 0.07$ Å$^{-1}$, which showed excellent convergence of the energy differences, stress tensors and structural parameters.

## 3. Tests and applications

Here we discuss tests of USPEX on systems with well known stable phases and also show how our method finds hitherto unknown structures. The test cases (including ice, methane, ammonia, carbon dioxide, benzene, glycine and butane-1,4-diammonium dibromide *etc.*) cover a wide range of systems with different molecular shapes (tetrahedral, linear, bent, planar and small biomolecules) and chemical interactions

# feature articles

(vdW dispersion, ionic, covalent and metallic bonding, strong/weak hydrogen bonding, $\pi$–$\pi$ stacking, organic and inorganic molecular systems etc.). All the calculations discussed below were performed in the framework of DFT or DFT + D. Driven by the USPEX code, the structures were initially relaxed in SIESTA with constrained molecular geometry and then fully relaxed in VASP at the final stages.

## 3.1. Ice

Ice ($H_2O$) is an archetypal hydrogen-bonded molecular crystal. The orientations of hydrogen bonds locally obey the well known ice rules, that is, each O atom is tetrahedrally bonded to four H atoms by two strong covalent intramolecular bonds and two much weaker intermolecular bonds (hydrogen bonds). Given the enormous number of possibilities of placing and orienting (even under ice rules) water molecules, prediction of the ice structure is a complex task: according to Maddox (1988), it is *still thought to lie beyond the mortals' ken.*

The normal crystalline form of ice, ice Ih, is disordered and has hexagonal symmetry, with O atoms arranged in a hexagonal diamond motif (a cubic diamond-type ice Ic is also known experimentally) with randomly oriented hydrogen bonds. In the experiment (Fukazawa et al., 1998; Leadbetter et al., 1985), ice XI (ordered version of ice Ih), was found to be the most stable polymorph at 1 atm and low temperatures, but the transformation from disordered ice Ih to ordered ice XI is kinetically hindered and this is why special approaches are needed for the experimental preparation of ice XI (Fukazawa et al., 1998).

With variable-cell USPEX simulations for a four-molecule cell at 1 atm we indeed identified ice XI as the most stable polymorph (Fig. 3a). This structure was found within just four generations, after relaxing ∼ 160 structures. Fig. 4 shows how the lowest energy changed from generation to generation in our calculation. This purely quantum-mechanical calculation required less than 1 day on eight cores of a Dell XPS desktop

PC. Apart from ice XI we found several remarkable structures in the same run.

An ordered version of ice Ic (Murray et al., 2005), a tetragonal phase with a cubic diamond-type oxygen sublattice (Fig. 3b), was found to be energetically competitive with ice XI. At both the GGA and GGA + D levels of theory, its energy is only 2 meV per molecule above that of ice XI. We have also found an interesting low-energy metastable polymorph (Fig. 3c), where the oxygen sublattice has the topology of the hypothetical bct4 allotrope of carbon (Umemoto et al., 2010; Zhou et al., 2010). The bct4-like structure of ice was also found from molecular dynamics simulation of the water's adsorption on the surface of hydroxylated $\beta$-cristobalite (Yang et al., 2005). Proton ordering lowers its symmetry from $I4/mmm$ to $Cm$.

## 3.2. Methane

Methane, the simplest of the saturated hydrocarbons, is an important constituent of the gas-giant planets Uranus and Neptune (Hubbard et al., 1991). The high-pressure behavior of methane is of extreme importance for fundamental chemistry, as well as for understanding the physics and chemistry of planetary interiors.

The tetrahedral $CH_4$ molecules interact practically only by vdW dispersion forces with each other. In spite of the simplicity of the molecule, the phase diagram of methane is still not well established (Bini & Pratesi, 1997; Maynard-Casely et al., 2010; Nakahata et al., 1999; Sun et al., 2009). Different experiments on methane were conducted during the last few decades, resulting in a complex phase diagram drawn by Bini & Pratesi (1997). Of the nine solid phases in the diagram, only the structures of phases (I), (II) and (III) have been determined, while phases (II), (III), (IV), (V) and (VI) only exist below 150 K and at moderate pressures. $CH_4$ is expected to become chemically unstable and decompose at megabar pressures (Gao et al., 2010).

The high-pressure phases of solid methane above 5 GPa have been the subject of numerous experimental and theoretical studies, however, understanding is still incomplete. Bini & Pratesi (1997), based on IR and Raman data, proposed a tetragonal crystal structure for plastic phase A, while high-pressure X-ray powder diffraction experiments suggested that the unit cell contains 21 molecules in the pseudocubic rhombohedral unit cell (Nakahata et al., 1999).

We performed structure prediction simulation for $CH_4$ using experimental cell parameters (Sun et al., 2009) at 11 GPa. We indeed found the best structure to possess a rhombohe-
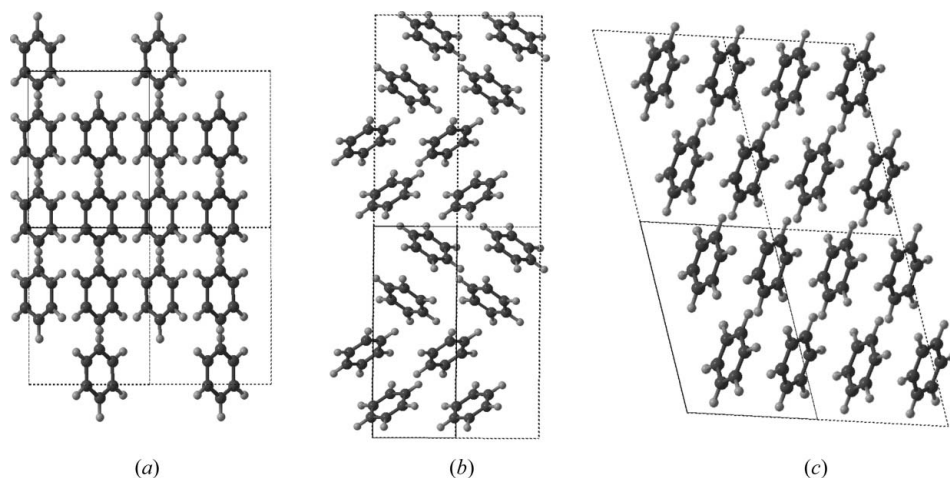


**Figure 8**
Crystal structures of benzene (a) orthorhombic phase (I) (*Pbca*, Z = 4); (b) tetragonal phase (II) (*P4₃2₁2*, Z = 4); (c) monoclinic phase (*P2₁/c*, Z = 2).

dral symmetry, and this structure was found within eight generations and is characterized by the icosahedral packing of methane molecules. This packing fully explains the unusual number (21) of molecules in the cell: 1 molecule is located in the center of the unit cell, 12 molecules around it form an icosahedron, and the remaining 8 molecules are located above the triangular faces of the icosahedron (Fig. 5). A rhombohedral model, very similar to ours, was recently proposed on the basis of neutron scattering experiments (Maynard-Casely *et al.*, 2010): the only difference is that our model has orientationally disordered molecules (as is also most likely to be the case in reality: furthermore, this model has a lower energy), while Maynard-Casely *et al.* (2010) assumed orientationally ordered molecules. The essential icosahedral character of the structure was not mentioned by Maynard-Casely *et al.* (2010), but can be clearly seen on close inspection of their results.

### 3.3. Ammonia

Bonding in $NH_3$ is intermediate between the hydrogen-bonded tetrahedral structure of $H_2O$ and the vdW-bonded close-packed structure of $CH_4$. Weak hydrogen bonding between neighboring ammonia molecules results in a pseudo-close-packed arrangement in the solid state (Fortes *et al.*, 2001). It is extremely interesting to understand the nature of hydrogen bonding in crystalline ammonia; properties of ammonia under pressure are of fundamental interest, as compressed ammonia has a significant role in planetary physics (Hubbard *et al.*, 1991).

At room temperature, ammonia crystallizes at 1 GPa in a rotationally disordered, face-centered cubic phase [phase (III); Fortes *et al.*, 2001; Datchi *et al.*, 2006; Loveday *et al.*, 1996]. X-ray and neutron studies have yielded information about the equation of state and structures of solid ammonia.
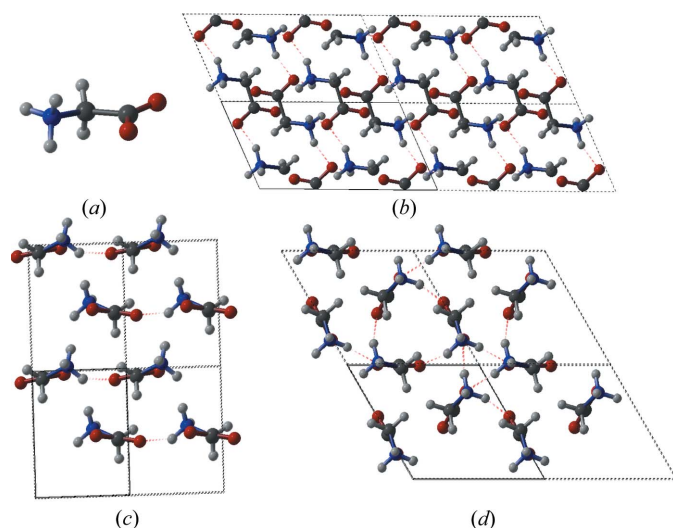
The low $P$, $T$ phase (I) of ammonia undergoes a first-order phase transition into phase (IV) at ∼ 3–4 GPa and then into phase (V) at ∼ 14 GPa. Phase (I) has a cubic structure with the space group $P2_13$, while phase (IV) has been identified as the orthorhombic structure with space group $P2_12_12_1$. Phase (V) might have the same space group as phase (IV) (an isosymmetric phase transition).

We carried out variable-cell structure prediction calculations at 5, 10, 25 and 50 GPa. At low pressures (5 GPa) we found the $P2_13$ structure to be stable (Fig. 6a), in good agreement with the experiment. At high pressures, USPEX without applying symmetry in the initialization still easily found the $P2_13$ structure, however, failed to obtain the ground-state structure $P2_12_12_1$ phase in a simulation with up to 20 generations. The energies of whole-molecule rotation are very small compared with intramolecular bonding energies, thus making the process of finding correct molecular orientations extremely difficult. This indicates that the energy landscape of ammonia is actually very flat. To enhance the searching efficiency we initialized the first generation using random symmetric structures. Also, to retain the diversity of the population 30% of each new generation was produced by a random symmetric mechanism. In this case the ground-state structure (Fig. 6c) appeared within six generations (∼ 210 structural relaxations). In addition, we also found the $P2_1/c$ phase (Fig. 6b) reported before (Pickard & Needs, 2008).

### 3.4. Carbon dioxide

The $CO_2$ molecule has a special significance because it is very abundant in nature and is a model system involving $\pi$-bonding and $sp$-hybridization of C atoms. Similar to methane, carbon dioxide is a vdW crystal with strong (weak) intramolecular (intermolecular) interactions at low pressures (Santoro & Gorelli, 2006). At room temperature and 1.5 GPa, $CO_2$ crystallizes as dry ice with a cubic $Pa3$ structure. At pressures between 12 and 20 GPa, $CO_2$-(I) transforms to the ortho-



**Figure 9**
Glycine polymorphs found by USPEX. (*a*) Representation of glycine zwitterion; (*b*) $\alpha$-glycine at 2 GPa ($Z = 4$, $a = 5.390$, $b = 5.911$, $c = 10.189$ Å, $\beta = 113.2°$); (*c*) $\beta$-glycine at 0.4 GPa ($Z = 2$, $a = 5.372$, $b = 6.180$, $c = 5.143$ Å, $\beta = 111.9°$); (*d*) $\gamma$-glycine at 1 GPa ($Z = 3$, $a = b = 7.070$, $c = 5.490$ Å).
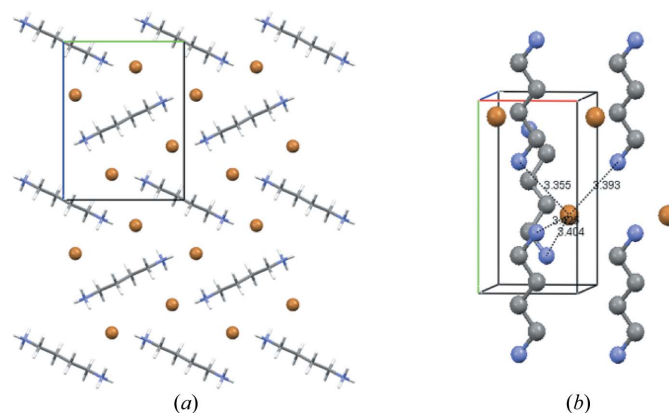


**Figure 10**
Butane-1,4-diammonium dibromide polymorph found by USPEX (space group: $P2_1/c$, $Z = 2$). (*a*) Representation of the network, with a view from the *a* axis; (*b*) $Br^-$ coordination environment, with a view from the *b* axis. For clarity, H atoms are not shown in (*b*). The $C_4H_{14}N_2^{2+}$ molecular ion has six flexible angles, and the unit cell of the stable polymorph contains 44 atoms.

rhombic $CO_2$-(III) (Olinger, 1982; Tajima *et al.*, 1997; Holm *et al.*, 2000). According to the theoretical calculation, $CO_2$-(III) is metastable, while $CO_2$-(II) with the $P42/mnm$ symmetry is believed to be thermodynamically stable (Bonev *et al.*, 2003). It is known that above 20 GPa a non-molecular phase [called phase (V)] with tetrahedrally coordinated C atoms becomes stable (Santoro & Gorelli, 2006).

In the previous prediction (Oganov *et al.*, 2008), unconstrained USPEX calculations succeeded in finding the correct $CO_2$ structures in a wide pressure range. By applying molecular constraint we have found the $P42/mnm$ phase (Fig. 7) quicker, just in two generations ($\sim$ 80 structural relaxations). The $P42/mnm$ phase remains the most stable structure composed of discrete $CO_2$ molecules at least up to 80 GPa. Both experiment (Yoo *et al.*, 1999) and theory (Bonev *et al.*, 2003; Oganov & Glass, 2008) show that $CO_2$ polymerizes above 20 GPa, while the molecular form ($P42mnm$ phase) exists as a metastable form at low temperatures and higher pressures. This example shows how imposing constraints gives the most stable molecular form, while unconstrained search finds the global minimum (which for $CO_2$ is non-molecular above 20 GPa). Both cases correspond to situations that are experimentally achievable, and thus important.

## 3.5. Benzene

Benzene is the simplest aromatic compound and it has a purely planar molecule, the packing of which is stabilized by $\pi$–$\pi$ interactions. The crystal structure of benzene is one of the most basic and most actively investigated structures. The first proposed phase diagram was very complex and contained six solid phases (Thiery & Leger, 1982). However, recent experimental studies simplified it (Ciabini *et al.*, 2005, 2007). At normal conditions, benzene crystallizes in the orthorhombic phase (I) ($Pbca$). A monoclinic phase (II) ($P2_1/c$), with two molecules per unit cell, was identified above 1.75 GPa. Phase (II) is stable up to the onset of chemical reactions (at 41 GPa and 298 K).
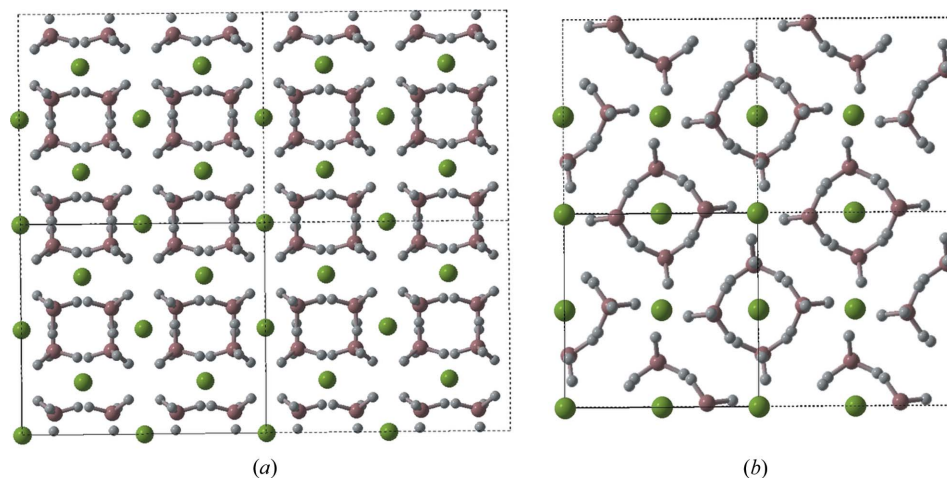
In our simulation we started with the same empirical potential (Spoel *et al.*, 1996) as used in a recent metadynamics study (Raiteri *et al.*, 2005), and we reproduced the multiple phases of benzene found there and corresponding to the old phase diagram. This potential was calibrated at normal conditions and may fail at high pressure. Its predicted many stable phases at different pressures (this is consistent with the old phase diagram, but most of these phases should be metastable according to the new experiment). To remedy this, we repeated our structure prediction runs at the level of DFT + D. We performed the calculation at 0, 5, 10 and 25 GPa with $Z = 4$. In our simulation, the experimentally observed orthorombic phase ($Pbca$) was identified as the most stable phase at 0 GPa, and then it transforms to the $P4_32_12$ phase at 4 GPa. We also found the monoclinic phase ($P2_1/c$; Fig. 8) as the ground state above 7 GPa. Our DFT + D results give fewer stable phases, in agreement with the new phase diagram – the only difference is in the $P4_32_12$ phase. This phase is experimentally known, and according to the latest experimental results (Ciabini *et al.*, 2005, 2007) is metastable. A previous DFT calculation (Wen *et al.*, 2011) suggested this phase to be stable at pressures of 4–7 GPa, which is consistent with our results. The thermodynamic stability of the $P4_32_12$ phase needs to be revisited experimentally.

## 3.6. Glycine

Glycine, with the formula $NH_2CH_2COOH$, is the smallest of 20 aminoacids commonly found in proteins. Aminoacids are important in nutrition and widely used in the pharmaceutical industry.

The polymorphism of glycine was intensely studied (Chisholm *et al.*, 2005; Hamad *et al.*, 2008; Dawson *et al.*, 2005; Moggach *et al.*, 2008; Boldyrea *et al.*, 2003; He *et al.*, 2006; Pervolich *et al.*, 2001). Glycine is known to crystallize in four polymorphs with space groups $P2_1/c$, $P2_1$, $P3_2$ and $P2_1/c$, which are labeled $\alpha$, $\beta$, $\gamma$ and $\sigma$, respectively (Chisholm *et al.*, 2005). The $\alpha$, $\beta$ and $\gamma$ phases are found at ambient pressure, with $\alpha$ and $\beta$ phases being metastable with respect to the $\gamma$ phase. $\sigma$-Glycine has recently been found to form under pressure (Dawson *et al.*, 2005). In the gas phase glycine is in a nonionic form, while in all four of the crystal structures glycine is zwitterionic (as shown in Fig. 9a). In this form an $-NH_3^+$ group on one ion electrostatically interacts with a $-COO^-$ group on a neighboring ion. Although zwitterionization causes an increase in energy with respect to the gas-phase molecule, it is thought that the zwitterionic crystals are stabilized by an increase in the number of hydrogen bonds that can be



**Figure 11**
$Mg(BH_4)_2$ polymorphs found by USPEX. (*a*) $F222$ phase; (*b*) $I4_122$ phase.

formed in comparison to the number that would be formed in the nonionic case.

Since the glycine zwitterion only has the point symmetry $C_1$ (*i.e.* no symmetry), structure prediction of glycine is more challenging compared with benzene. We performed variable cell prediction at 1 GPa with 2–4 molecules per cell. Without any experimental information, we found $\beta$-glycine (Fig. 9c) as the metastable structure with $Z = 2$; and $\gamma$-glycine (Fig. 9d) as the best structure with $Z = 3$. We also found $\alpha$-glycine as a metastable form in the calculation with $Z = 4$ (Fig. 9b) at 2.0 GPa. This shows the power of our evolutionary search method. However, GGA + D results show that $\alpha$-glycine possesses the lowest enthalpy, while the $\gamma$ and $\beta$ phases are 20 and 30 meV mol$^{-1}$ higher, respectively. Yet the experimental results demonstrated the relative thermodynamic stability to be $\gamma > \alpha > \beta$. This shows the need for better ways of computing intermolecular interaction energies.

### 3.7. Butane-1,4-diammonium dibromide

The molecules we discussed so far are rigid or nearly rigid. Is it possible to use this approach to study the packing of flexible molecules? To investigate this, we applied it to the prediction of the crystal structure of butane-1,4-diammonium dibromide, in which Br$^-$ and $C_4H_{14}N_2^{2+}$ can be described as two molecular units that form the structure.

By using the experimental cell parameters (van Blerk & Kruger, 2007) we indeed observed numerous structures with different conformations of the $C_4H_{14}N_2^{2+}$ molecular ion. USPEX firstly found the energetically favorable conformation and then identified the ground-state structure at the 12th generation ($\sim$ 500 structural relaxations): $P2_1/c$ butane-1,4-diammonium dibromide. In this structure, as shown in Fig. 10, the organic hydrocarbon chains are found to pack in a herringbone-type stacking with hydrogen bonds to Br$^-$. Each Br$^-$ anion is surrounded by four $-NH_3^+$ groups. During the process of rotational mutation, both the orientation of the whole molecular group and its flexible torsional angles are allowed to change. A large fraction of rotation ($\sim$ 40%) of the molecules is found to greatly speed up the prediction. This success confirmed that our constrained evolutionary algorithm can be straightforwardly adapted to deal with flexible molecules.

### 3.8. Inorganic crystals

Apart from molecular crystals, this new approach is also applicable to inorganic crystals with complex ions or clusters. Below are a few illustrations.

**3.8.1. Complex ionic solids: example of hydrogen storage materials.** Reversible hydrogen storage materials recently attracted great interest (Schlapbach & Züttel, 2001). Two groups of complex metal hydrides: alumohydrides containing AlH$_4$ groups and borohydrides with BH$_4$ groups have recently been under intensive study (Her *et al.*, 2007; Ozolins *et al.*, 2008; Dai *et al.*, 2008; Zhou *et al.*, 2009). Numerous dehydriding and rehydriding processes have been predicted theoretically and tested experimentally. In a good candidate material, dehydridation should happen at acceptably low temperatures. Structure prediction for such systems can guide the experimentalists to synthesize the desired compounds in the laboratory.

The crystal structure of Mg(BH$_4$)$_2$ has been extensively investigated. It was experimentally solved and found to be extremely complex (330 atoms per unit cell for the low-temperature phase with $P6_1$ symmetry; Her *et al.*, 2007). Recent theoretical work then predicted a new body-centered tetragonal phase (with $I\bar{4}m2$ symmetry), which has slightly lower energy than the $P6_1$ phase; it was found using the prototype electrostatic ground-state approach (PEGS; Ozolins *et al.*, 2008). Later, based on the prototype structure of Zr(BH$_4$)$_4$, another orthorhombic phase with $F222$ symmetry was found to have even lower energy than all the previously proposed structures (Zhou *et al.*, 2009).

In general, the previous theoretical discoveries of novel Mg(BH$_4$)$_2$ phases were conducted either by *ad hoc* extensive searching or by chemical intuition. However, our evolutionary algorithm does not rely on any prior knowledge except chemical composition, and could be particularly useful for predicting stable crystal structures for these complex metal hydride systems. If we consider the BH$_4^-$ ion as a molecular group, the search space would be dramatically reduced. Within 10 generations ($\sim$ 400 structure relaxations), USPEX found the $F222$ phase (Fig. 11a) as the most stable structure at ambient pressure. In addition, $I\bar{4}m2$ (Fig. 11b) was also found by USPEX in the same calculation, with enthalpy less than 1.2 meV per atom above that of the $F222$ phase. Compared with the previous work, our method is clearly more universal and robust, and enables efficient structure prediction for complex molecular systems, both organic and inorganic.

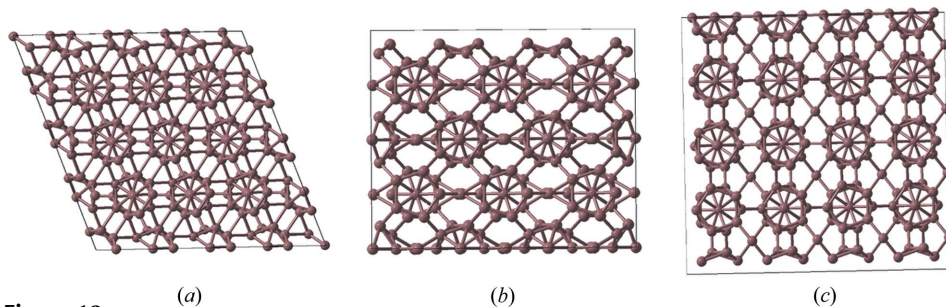**3.8.2. Cluster-based crystals: example of elemental boron.**



**Figure 12**
Crystal structures of boron. (a) $\alpha$-B$_{12}$; (b) $\gamma$-B$_{28}$; (c) novel metastable B$_{52}$ phase, space group $Pnn2$, $a = 8.868$, $b = 8.777$, $c = 5.000$ Å. B1(0.5777,0.7728,0), B2(0.9187,0.7321,0.3222), B3(0.7464,0.7488,0.4978), B4(0.5902,0.6690,0.3087), B5(0.6243,0.8678,0.3055), B6(0.8209,0.9090,0.3202), B7(0.8698,0.6288,0.0229), B8(0.7835,0.5798,0.3273), B9(0.6684,0.5795,0.0182), B10(0.7190,0.9189,0.0056), B11(0.8991,0.8309, 0.0115), B12(0.7461,0.7461,0.8302), B13(0,0,0.7529), B14(0,0.5,0.9161).

Boron, located in a unique position of the periodic table, is an element of chemical complexity due to the subtle balance between localized and delocalized electronic states. All known structures of boron contain icosahedral $B_{12}$ clusters. A recent experiment (Oganov *et al.*, 2009) found a new phase of pure boron ($\gamma$-$B_{28}$) at pressures above 10–12 GPa, and its structure was solved using USPEX with fixed experimental cell dimensions (Oganov *et al.*, 2009). Surprisingly, $\gamma$-$B_{28}$ showed different chemistry compared with all the other elemental boron allotropes. In the $\gamma$-$B_{28}$ structure (Fig. 12*b*), the centers of the $B_{12}$ icosahedra form a distorted cubic close packing as in $\alpha$-$B_{12}$ (Fig. 12*a*), but all octahedral voids are occupied by $B_2$ pairs. The $\gamma$-$B_{28}$ structure resembles an NaCl-type structure, with the $B_{12}$ icosahedra and $B_2$ pairs as 'anions' and 'cations'. Finding this structure without fixing cell parameters was reported to be exceedingly difficult (Ji *et al.*, 2010), but the latest methodological developments enable this (Lyakhov *et al.*, unpublished). However, the problem can be made very simple if we recall that all boron phases contain $B_{12}$ icosahedra. Here we treated $B_{12}$ icosahedral and $B_2$ pairs as separate rigid units, and performed structure prediction runs at different numbers of $B_{12}$ and $B_2$ units (2:1, 1:1, 2:2, 2:4 *etc.*) at ambient conditions. We could easily find $\gamma$-$B_{28}$ within 2–3 generations or $\sim 100$ structural relaxations. Meanwhile, we observed a set of low-energy and chemically interesting structures with different proportions of $B_{12}$ and $B_2$. For instance, the novel metallic phase $B_{52}$ with the $Pnn2$ symmetry (Fig. 12*c*) was calculated to be only 12 meV per atom higher than $\gamma$-$B_{28}$ at atmospheric pressure. Its energy is lower in energy than those of the experimentally observed phases (such as the T-50 phase; Hoard *et al.*, 1958) and this example shows that our method can be used for even non-molecular and inorganic solids that contain clusters or complex ions.

## 4. Discussion and conclusions

In the original version of USPEX (Oganov & Glass, 2006) the stable crystal structure was assembled from individual atoms, which was also shown to work well for atomic crystals and for simple molecular systems (carbon dioxide, water, urea). However, it is clear that for molecular crystals improvements of the efficiency can be made if the structure is assembled from whole molecules rather than individual atoms. This is confirmed by the present study. Our *constrained global optimization* method allows the stable crystal structure of a given molecular compound to be found, and provides a set of low-energy metastable structures at a highly affordable cost.

The reasons why evolutionary algorithms succeed in crystal structure prediction have been discussed before (Oganov *et al.*, 2011). As mentioned in §2, in addition to these, the *constrained global optimization* fixes the molecular connectivity and brings the need for new variation operators (rotational mutation and molecular softmutation), developed and described here.

For efficient and reliable polymorph prediction, the population of structures should be sufficiently diverse. A major difficulty in the prediction of molecular crystals is the large number of plausible candidate structures that can have very close energies (Neumann & Perrin, 2005). Given the complexity of their energy landscape, the high diversity of the population of the structures is mandatory for successful prediction of molecular crystal structures. The initial population is particularly important and it is usually a good idea to add a number of random symmetrized structures in each generation, to keep sampling of the landscape diverse.

The presented algorithm provides not only the theoretical ground state, but also a number of low-energy metastable structures. With the inclusion of zero-point energy and entropic contributions, such structures may become stable. Even if this does not happen, low-energy metastable structures have a relatively high chance of being synthesized under special conditions.

While DFT + D is today's state of the art and its accuracy is often sufficient, for some systems (glycine) DFT + D is too crude and more reliable approaches for computing the energy are needed. Under high pressure many of the difficulties disappear, because the vdW interactions (poorly accounted for by today's *ab initio* methods) become relatively less important.

Clearly, the quality of the global minimum found by USPEX depends on the accuracy of the theory used for energy calculations and structure relaxation. Current levels of theory can be roughly divided into empirical, semiempirical and *ab initio* approaches. Accurate empirical force fields are appropriate for CSP, but reliable parameterizations are hard to generate for most molecules. In contrast to empirical force fields, *ab initio* calculations provide a more accurate and rigorous description without parameterization, but the calculations are much more time-consuming. In our prediction, we adopt the DFT + D level of theory, which combines 'the best of both worlds', *i.e.* an accurate representation of intermolecular repulsions, hydrogen bonding, electrostatic interactions and vdW dispersions. DFT + D proved to be reliable for most systems, but its results are not fully satisfactory for glycine. This shows that further improvements in theoretical calculations of intermolecular interactions energies are needed. In parallel with the improvement of methods for energy ranking, there is a need for efficient and reliable algorithms for global optimization of the theoretical energy landscape, and the present work is an important development in this direction. In the present paper we describe the most important ingredients of this method, and demonstrate how it enables affordable structure prediction for many complex organic and inorganic systems at the *ab initio* level.

In summary, we have presented a new efficient and reliable approach for global energy optimization for molecular crystal structure prediction. It is based on the evolutionary algorithm USPEX extended to molecular crystals by additional variation operators and constraints of partially or completely fixed molecules. The high efficiency of this method enables fully quantum-mechanical structure predictions to be performed at an affordable computational cost. Using this method, we succeeded in finding the stable structures for systems with various rigid molecular shapes (tetrahedral, linear, bent, planar and complex molecules), and different bonding situa-

tions (vdW bonding, ionic, covalent, metallic, weak and strong hydrogen bonding, $\pi$–$\pi$ stacking *etc.*). We showed that even large systems can be efficiently dealt with by this approach, even at the *ab initio* level of theory. This new approach also has wide applicability to inorganic crystals containing clusters and complex ions.

## References

Baburin, I. A., Leoni, S. & Seifert, G. (2008). *J. Phys. Chem. B*, **112**, 9437–9443.

Bardwell, D. A. *et al.* (2011). *Acta Cryst.* B**67**, 535–551.

Baur, W. H. & Kassner, D. (1992). *Acta Cryst.* B**48**, 356–369.

Bini, R. & Pratesi, G. (1997). *Phys. Rev. B*, **55**, 14800.

Blerk, C. van & Kruger, G. J. (2007). *Acta Cryst.* E**63**, o342–o344.

Blochl, P. E. (1994). *Phys. Rev. B*, **50**, 17953–17979.

Boldyreva, V., Drebushchak, T. N. & Shutova, E. S. (2003). *Z. Kristallogr.* **218**, 366–376.

Bonev, S. A., Gygi, F., Ogitsu, T. & Galli, G. (2003). *Phys. Rev. Lett.* **91**, 065501.

Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.

Chisholm, J. A., Motherwell, S., Tulip, P. R., Parsons, S. & Clark, S. J. (2005). *Cryst. Growth Des.* **5**, 1437–1442.

Ciabini, L., Gorelli, F. A., Santoro, M., Bini, R., Schettino, V. & Mezouar, M. (2005). *Phys. Rev. B*, **72**, 094108.

Ciabini, L., Santoro, M., Gorelli, F. A., Bini, R., Schettino, V. & Raugei, S. (2007). *Nat. Mater.* **6**, 39–43.

Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. (2003). *Phys. Rev. Lett.* **91**, 135503.

Dai, B., Sholl, D. S. & Johnson, J. K. (2008). *J. Phys. Chem. C*, **112**, 4391–4395.

Datchi, F., Ninet, S., Gauthier, M., Saitta, A. M., Canny, B. & Decremps, F. (2006). *Phys. Rev. B*, **73**, 17411.

Dawson, A., Allan, D. R., Belmonte, S. A., Clark, S. J., David, W. I. F., McGregor, P. A., Parsons, S., Pulham, C. R. & Sawyer, L. (2005). *Cryst. Growth Des.* **5**, 1415–1427.

Day, G. M. *et al.* (2005). *Acta Cryst.* B**61**, 511–527.

Day, G. M. (2011). *Cryst. Rev.* **17**, 3–52.

Day, G. M. *et al.* (2009). *Acta Cryst.* B**65**, 107–125.

Fortes, A. D., Brodholt, J. P., Wood, I. G., Vočadlo, L. & Jenkins, H. D. B. (2001). *J. Chem. Phys.* **115**, 7006.

Freeman, C. M., Newsam, J. M., Levine, S. M. & Catlow, C. R. A. (1993). *J. Mater. Chem.* **3**, 531–535.

Fukazawa, H., Ikeda, S. & Mae, S. (1998). *Chem. Phys. Lett.* **282**, 215–218.

Gale, J. D. & Rohl, A. L. (2003). *Mol. Simul.* **29**, 291–341.

Gao, G., Oganov, A. R., Ma, Y., Wang, H., Li, P., Li, Y., Iitaka, T. & Zou, G. (2010). *J. Chem. Phys.* **133**, 144508.

Gavezzotti, A. (1994). *Acc. Chem. Res.* **27**, 309–314.

Glass, C. W., Oganov, A. R. & Hansen, N. (2006). *Comput. Phys. Commun.* **175**, 713–720.

Goedecker, S. (2004). *J. Chem. Phys.* **120**, 9911–9917.

Grimme, S. (2006). *J. Comput. Chem.* **27**, 1787–1799.

Hamad, S., Hughes, C. E., Catlow, C. R. & Harris, K. D. (2008). *J. Phys. Chem. B*, **112**, 7280–7288.

He, G. W., Bhamidi, V., Wilson, S. R., Tan, R. B. H., Kenis, P. J. A. & Zukoski, C. F. (2006). *Cryst. Growth Des.* **6**, 1746–1749.

Her, J.-H., Stephens, P. W., Gao, Y., Soloveichik, G. L., Rijssenbeek, J., Andrus, M. & Zhao, J.-C. (2007). *Acta Cryst.* B**63**, 561–568.

Hoard, J. L., Hughes, R. E. & Sands, D. E. (1958). *J. Am. Chem. Soc.* **80**, 4507–4515.

Hoft, R., Gale, J. D. & Ford, M. J. (2006). *Mol. Simul.* **32**, 595–600.

Holm, B., Ahuja, R, Belonoshko, A. & Johansson, B. (2000). *Phys. Rev. Lett.* **85**, 1258.

Hubbard, W. B., Nellis, W. J., Mitchell, A. C., Holmes, N. C., Limaye, S. S. & McCandless, P. C. (1991). *Science*, **253**, 648–651.

Ji, M., Wang, C. Z. & Ho, K. M. (2010). *Phys. Chem. Chem. Phys.* **12**, 11617–11623.

Kim, S., Orendt, A. M., Ferraro, M. B. & Facelli, J. C. (2009). *J. Comput. Chem.* **30**, 1973–1985.

Kresse, G. & Furthmuller, J. (1996). *Phys. Rev. B*, **54**, 11169–11186.

Leadbetter, A. J., Ward, R. C., Clark, J. W., Tucker, P. A., Matsuo, T. & Suga, H. (1985). *J. Chem. Phys.* **82**, 424–428.

Li, Y., Lu, D., Nguyen, H. V. & Galli, G. (2010). *J. Phys. Chem. A*, **114**, 1944–1952.

Lommerse, J. P. M. *et al.* (2000). *Acta Cryst.* B**56**, 697–714.

Loveday, J. S., Nelmes, R. J., Marshall, W. G., Besson, J. M., Klotz, S. & Hamel, G. (1996). *Phys. Rev. Lett.* **76**, 74–77.

Lyakhov, A. L., Oganov, A. R. & Valle, M. (2010). *J. Comput. Phys. Commun.* **181**, 1623–1632.

Maddox, J. (1988). *Nature*, **335**, 201.

Martonák, R., Laio, A. & Parrinello, M. (2003). *Phys. Rev. Lett.* **90**, 075503.

Maynard-Casely, H. E., Bull, C. L., Guthrie, M., Loa, I., McMahon, M. I., Gregoryanz, E., Nelmes, R. J. & Loveday, J. S. (2010). *J. Chem. Phys.* **133**, 064504.

Moggach, S. A., Parsons, S. & Wood, P. A. (2008). *Cryst. Rev.* **14**, 143–184.

Motherwell, W. D. S. *et al.* (2002). *Acta Cryst.* B**58**, 647–661.

Murray, B. J., Knopf, D. A. & Bertram, A. K. (2005). *Nature*, **434**, 202–205.

Nakahata, I., Matsui, N., Akahama, Y. & Kawamura, H. (1999). *Chem. Phys. Lett.* **302**, 359–362.

Neumann, M. A. & Perrin, M. A. (2005). *J. Phys. Chem. B*, **109**, 15531–15541.

Oganov, A. R., Ono, S., Ma, Y., Glass, C. W. & Garcia, A. (2008). *Earth Planet. Sci. Lett.* **273**, 38–47.

Oganov, A. R. *et al.* (2009). *Nature*, **457**, 863–867.

Oganov, A. R. & Glass, C. W. (2006). *J. Chem. Phys.* **124**, 244704.

Oganov, A. R. & Glass, C. W. (2008). *J. Phys. Cond. Matter*, **20**, 064210.

Oganov, A. R., Lyakhov, A. O. & Valle, M. (2011). *Acc. Chem. Res.* **44**, 227–237.

Oganov, A. R. & Valle, M. (2009). *J. Chem. Phys.* **130**, 104504.

Olinger, B. (1982). *J. Chem. Phys.* **77**, 6255–6258.

Ozolins, V., Majzoub, E. H. & Wolverton, C. (2008). *Phys. Rev. Lett.* **100**, 135501.

Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J. & Calgnaert, V. (1990). *Nature*, **346**, 343–345.

Perdew, J. P., Burke, K. & Ernzerhof, M. (1996). *Phys. Rev. Lett.* **77**, 3865–3868.

Pervolich, G. L., Hansen, L. K. & Bauer-Brandl, A. (2001). *J. Therm. Anal. Calorim.* **66**, 699–715.

Pickard, C. J. & Needs, R. J. (2008). *Nat. Mater.* **7**, 775–779.

Price, S. L. (2004). *Adv. Drug Deliv. Rev.* **56**, 301–319.

# feature articles

Price, S. L., Leslie, M., Welch, G. W., Habgood, M., Price, L. S., Karamertzanis, P. G. & Day, G. M. (2010). *Phys. Chem. Chem. Phys.* **12**, 8478–8490.

Raiteri, P., Martonak, R. & Parrinello, M. (2005). *Angew. Chem. Int. Ed.* **44**, 3769–3773.

Santoro, M. & Gorelli, F. A. (2006). *Chem. Soc. Rev.* **35**, 918–931.

Schlapbach, L. & Züttel, A. (2001). *Nature*, **414**, 353–358.

Schon, J. C. & Jansen, M. (1996). *Angew. Chem. Int. Ed. Engl.* **35**, 1286–1304.

Soler, J. M., Artacho, E., Gale, J. D., Garcia, A., Junquera, J., Ordejón, P. & Sánchez-Portal, D. (2002). *J. Phys. Condens. Matter*, **84**, 2745.

Spoel, D. van der, van Buuren, A. R., Tieleman, D. P. & Berendsen, H. J. (1996). *J. Biomol. NMR*, **8**, 229–238.

Sun, L., Yi, W., Wang, L., Shu, J., Sinogeikin, S., Meng, Y., Shen, G., Bai, L., Li, Y., Liu, J. & Mao, H. (2009). *Chem. Phys. Lett.* **473**, 72–74.

Tajima, N., Tsuzuki, S., Tanabe, K., Aoki, K. & Hirano, T. (1997). *Electron. J. Theor. Chem.* **2**, 139–148.

Thiery, M. M. & Leger, J. M. (1982). *J. Chem. Phys.* **89**, 4255–4271.

Umemoto, K., Wentzcovitch, R. M., Saito, S. & Miyake, T. (2010). *Phys. Rev. Lett.* **104**, 125504.

Wales, D. J. & Doye, J. P. K. (1997). *J. Phys. Chem. A*, **101**, 5111–5116.

Wen, X. D., Hoffmann, R. & Ashcroft, N. W. (2011). *J. Am. Chem. Soc.* **133**, 9023–9035.

Yang, J., Meng, S., Xu, L. F. & Wang, E. G. (2005). *Phys. Rev. Lett.* **92**, 146102.

Yoo, C. S., Cynn, H., Gygi, F., Galli, G., Lota, V., Nicol, M., Carlson, S., Hausermann, D. & Mailhiot, C. (1999). *Phys. Rev. Lett.* **83**, 5527–5530.

Zhou, X. F., Qian, Q. R., Zhou, J., Xu, B., Tian, Y. & Wang, H. T. (2009). *Phys. Rev. B*, **79**, 212102.

Zhou, X. F., Qian, G. R., Dong, X., Zhang, L. X., Tian, Y. J. & Wang, H. T. (2010). *Phys. Rev. B*, **82**, 134126.