

# Evolutionary Crystal Structure Prediction as a Method for the Discovery of Minerals and Materials

**Artem R. Oganov**

*Department of Geosciences, Department of Physics and Astronomy,  
and New York Center for Computational Sciences  
Stony Brook University  
Stony Brook, New York, 11794-2100, U.S.A.  
artem.oganov@sunysb.edu*

*Geology Department, Moscow State University  
119992 Moscow, Russia*

**Yanming Ma**

*National Lab of Superhard Materials  
Jilin University  
Changchun 130012, P. R. China*

**Andriy O. Lyakhov**

*Department of Geosciences  
Stony Brook University  
Stony Brook, New York, 11794-2100, U.S.A.*

**Mario Valle**

*Data Analysis and Visualization Group  
Swiss National Supercomputing Centre (CSCS)  
Cantonale Galleria 2  
6928 Manno, Switzerland*

**Carlo Gatti**

*CNR-ISTM, Istituto di Scienze e Tecnologie Molecolari  
via Golgi 19  
20133 Milano, Italy*

## ABSTRACT

Prediction of stable crystal structures at given pressure-temperature conditions, based only on the knowledge of the chemical composition, is a central problem of condensed matter physics. This extremely challenging problem is often termed “crystal structure prediction problem,” and recently developed evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography) made an important progress in solving it, enabling efficient and reliable prediction of structures with up to ~40 atoms in the unit cell using *ab initio* methods. Here we review this methodology, as well as recent progress in analyzing energy landscape of solids (which also helps to analyze results of USPEX runs). We show several recent applications – (1) prediction of new high-pressure phases of CaCO<sub>3</sub>, (2) search

for the structure of the polymeric phase of CO<sub>2</sub> (“phase V”), (3) high-pressure phases of oxygen, (4) exploration of possible stable compounds in the Xe-C system at high pressures, (5) exotic high-pressure phases of elements boron and sodium, as well as extension of the method to variable-composition systems.

## INTRODUCTION

Crystal structure prediction problem occupies a central place in materials design. Solving this problem would also open new ways for understanding the behavior of materials at extreme conditions, where experiments are difficult (in some cases, prohibitively difficult).

Often the approach has been to compare the free energies of a number of candidate structures (usually taken from analogous materials, or constructed by chemical intuition). Data mining (Curtarolo et al. 2003) is the pinnacle of this approach, as it very efficiently explores databases of known crystal structures and, using correlations between structures adopted by different compounds, indicates a list of likely candidate structures. Problems arise when a totally unexpected and hitherto unknown structure is actually stable (this often happens under pressure, or when the system does not have known good analogs). A number of simpler intuitive empirical schemes (e.g., structure diagrams, polyhedral clusters – see Urusov et al. 1990) have appeared in literature, but their application usually requires a large experimental data set or good understanding of the compound at hand.

Thanks to recent methodological developments, reliable structure prediction can be performed without any prior knowledge or assumptions about the system. Simulated annealing (Deem and Newsam 1989; Pannetier et al. 1990; Boisen et al. 1994; Schön and Jansen 1996), minima hopping (Gödecker 2004) and metadynamics (Martoňák et al. 2003, 2005, 2006) have been used with some success. For small systems, even relaxing randomly produced structures can deliver the stable structure (Pickard and Needs 2006). Here we review the evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography) (Oganov and Glass 2006; Oganov et al. 2006; Glass et al. Hansen 2006) and a small selection of the results it has provided so far. This review is an updated version of the previous account of the methodology (Oganov et al. 2007) and is based on the lectures delivered at the 2009 Erice School on high-pressure crystallography, and at the 2009 MSA Short Course on computational mineral physics. The “*Evolutionary Algorithm USPEX*” section presents basics of the method, the “*Tests of the Algorithm*” section shows several interesting test cases (mostly on systems with a known ground state), while a number of applications to systems where the stable structure is unknown are presented in the “*Some Applications of the Method*” section.

## EVOLUTIONARY ALGORITHM USPEX

Several groups attempted the pioneering use of evolutionary algorithms to structure prediction: for crystals (Bush et al. 1995; Woodley et al. 1999; Bazterra et al. 2002; Woodley 2004), colloids (Gottwald et al. 2005) and clusters (Deaven and Ho 1995). The algorithm developed by Deaven and Ho (1995) is perhaps especially interesting as some of its features (real-space representation of structures, local optimization and spatial heredity) are similar to the USPEX method. Their algorithm has successfully reproduced the structure of the C<sub>60</sub> buckminsterfullerene, but has never been extended to heteroatomic clusters, nor to periodic systems (i.e., crystals). The algorithm of Bush and Woodley (Bush et al. 1995; Woodley et al. 1999; Woodley 2004) was originally developed for crystals and successfully produced a starting model for solving the structure of Li<sub>3</sub>RuO<sub>4</sub> (Bush et al. 1995). However, subsequent systematic tests (Woodley 2004; Woodley et al. 1999) showed frequent failures even for rather simple systems containing ~10 atoms/cell. Other drawbacks are that this algorithm requires

experimental lattice parameters and simulations are very expensive, unless a cheap and crude heuristic expression is used for fitness. Unlike the Deaven-Ho algorithm and USPEX, in this method structures are represented by binary “0/1” strings, there is no local optimization and no spatial heredity.

In USPEX, structures are represented by fractional coordinates for the atoms and lattice vectors. USPEX operates with populations of structures; from them, parent structures are selected. The fitness of structures is the relevant thermodynamic potential derived from *ab initio* total energy calculations. The worst structures of a population are discarded; for the remaining structures the probability of being selected as parent is a function (e.g., linear) of its fitness rank. A new candidate structure is produced from parent structures using one of three operators: (i) heredity, which combines spatially coherent slabs (in terms of fractional coordinates) of two parent structures, while the lattice vectors matrices are weighted averages of the two parent lattice vectors matrices, (ii) permutation (as in Woodley et al. 1999 and Woodley 2004), which swaps chemical identities in randomly selected pairs of unlike atoms, (iii) lattice mutation, which distorts the cell shape by applying a random symmetric strain matrix. To avoid pathological lattices, all newly produced structures are rescaled to produce a predefined unit cell volume (a reasonable starting value should be supplied in the input, and then allowed to evolve during the run). Heredity enables very broad searches, while preserving already found local fragments of good structures, and introduces ideas of “two-phase” simulations. Permutation facilitates finding the optimal ordering of the atoms; in some situations (for systems with a large range in degree of chemical similarity between different atom types) it may be useful to swap only chemically more similar atoms (e.g., Al-Si in aluminosilicates). Lattice mutation enables better exploration of the neighborhood of parent structures, prevents premature convergence of the lattice, and essentially incorporates the ideas of metadynamics in our search. The action of these variation operators is illustrated in Figures 1 and 2.

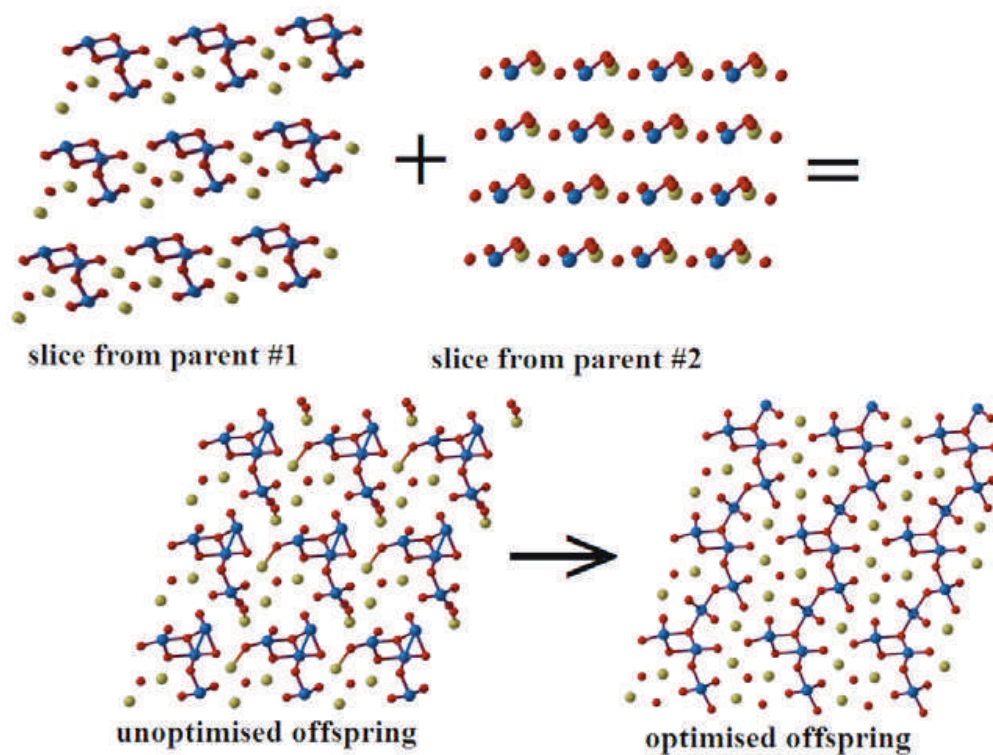
Before new candidate structures are relaxed, they are tested against three constraints:

- (1) all interatomic distances must be above the specified minimal values;
- (2) cell angles must be between 60° and 120°;
- (3) all cell lengths must be larger than a specified value (e.g., diameter of the largest atom).

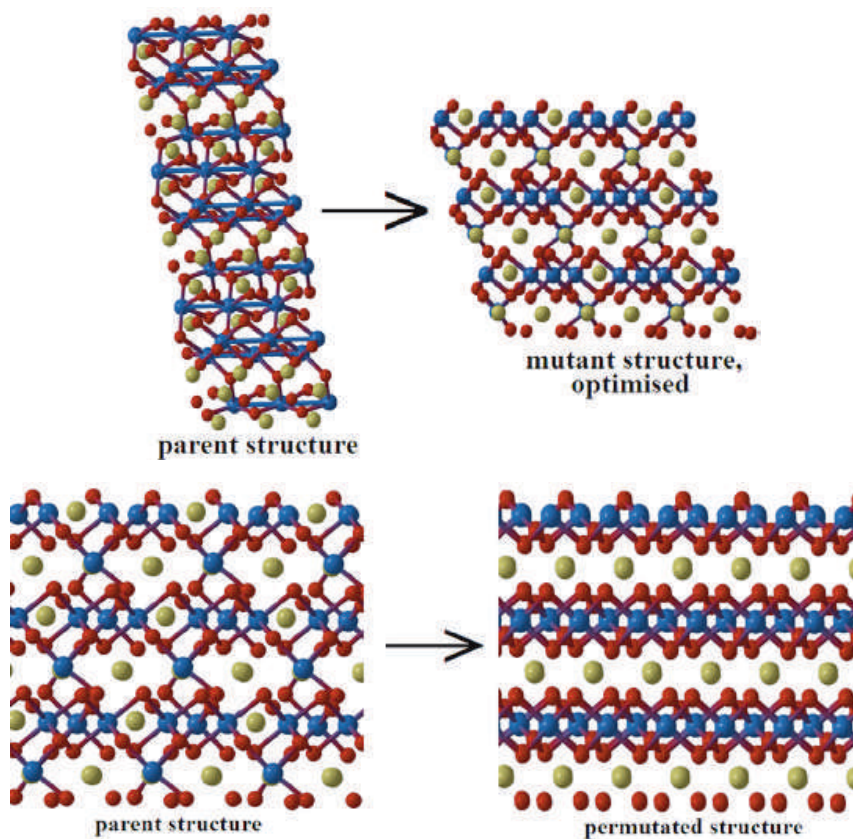
These constraints help to ensure stability of energy calculations and local optimization, and remove only redundant and infeasible regions of configuration space – thus the search is physically unconstrained. If in violation of these constraints, the candidate structure is discarded; otherwise, it is locally optimized (relaxed). Structure relaxations and energy calculations are done by external codes (currently, USPEX is interfaced with VASP (Kresse and Furthmüller 1996), SIESTA (Soler et al. 2002), GULP (Gale 2005)).

The relaxed structures are recorded and used for producing the next generation of structures. A new population of structures is made to contain one or more lowest-enthalpy structures from the previous population and the new structures produced using variation operators. Generation by generation, the above procedure is repeated in a loop.

The first generation usually consists of random structures, but it is possible to include user-specified structures. If lattice parameters are known, runs can be done in the fixed cell, but this is not required and in most cases simulations are done with variable cell shape. We have also improved the algorithm by more exhaustive removal of lattice redundancies (Oganov and Glass 2008). For more details on the USPEX method, see Oganov and Glass (2006) and Glass et al. (2006). A similar evolutionary algorithm was proposed slightly later and independently from us by Abraham and Probert (2006); this method differs from USPEX in the absence of permutation (with potential problems for binary and more complex compounds), different forms of heredity and mutation, and absence of cell rescaling. Recently, we also developed



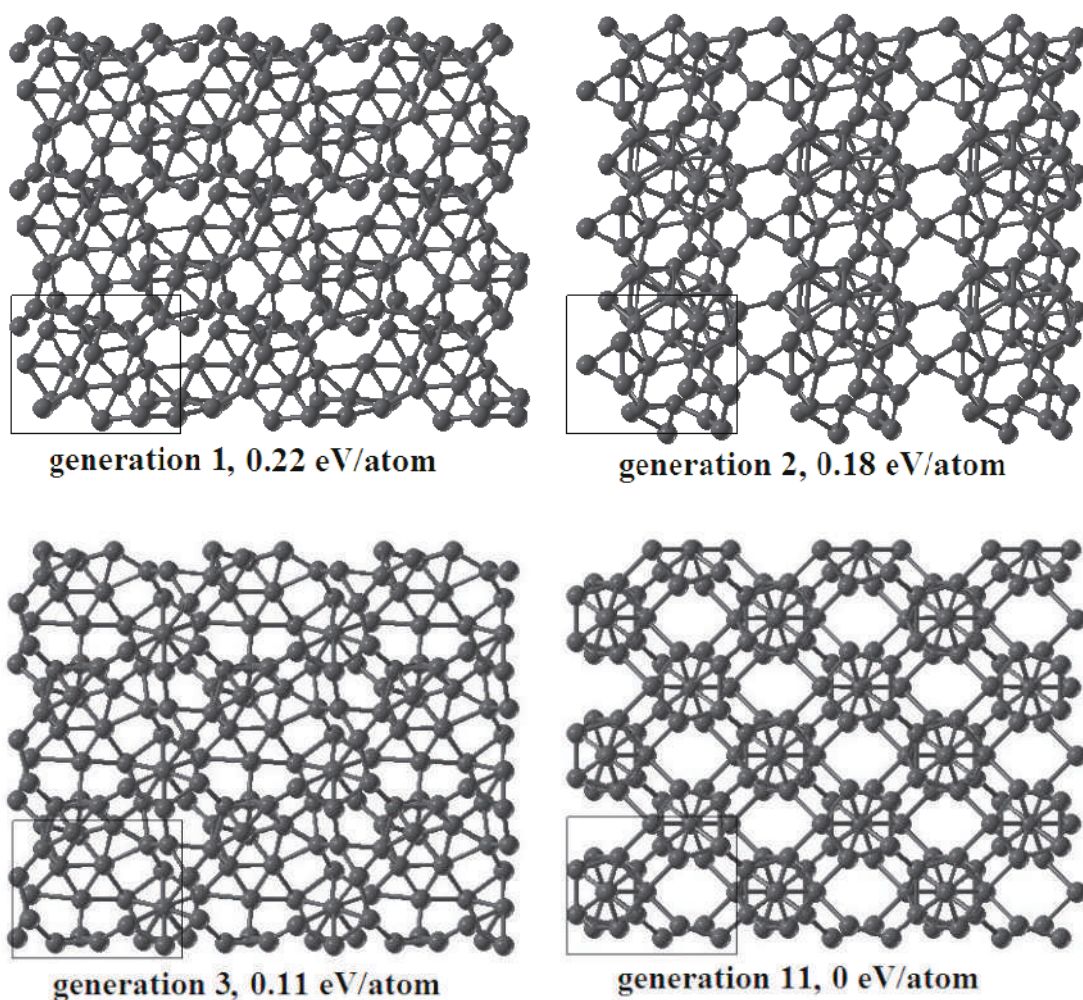
**Figure 1.** Heredity operator: slices of two parent structures, and the offspring structure before and after local optimization.



**Figure 2.** Illustrations of lattice mutation and permutation operators.

an approach, enabling deeper insight into the performance of structure prediction simulations (e.g., see below on similarity matrices) and into the energy landscape that is being sampled during the simulation (Valle and Oganov 2008; Oganov and Valle 2009).

Why is the USPEX methodology successful? One of the reasons is that local optimization creates chemically reasonable local environments for the atoms. Another reason is that evolutionary pressure (through selection) forces the population to improve from generation to generation. Yet another reason is the choice of variational operators. In heredity, local arrangements of atoms (spatially coherent pieces of structures) are partly preserved and combined. This respects the predominant short-ranged interactions in crystals and exploits information from the current population. For large systems it may be advantageous to combine slabs of several structures. On the other hand, for systems with very few atoms (or molecules) in the unit cell heredity becomes obsolete (in the limit of 1 atom/unit cell it is completely useless); these cases, however, are trivial for other variation operators and even for local optimization of random structures. As a general note, a successful evolutionary algorithm needs to maintain a balance between its “learning power” and maintaining diversity of the population. Figure 3 illustrates how, without any prior knowledge, a simulation of boron gradually “learned” about  $B_{12}$  icosahedra and arrived at the correct ground-state structure.

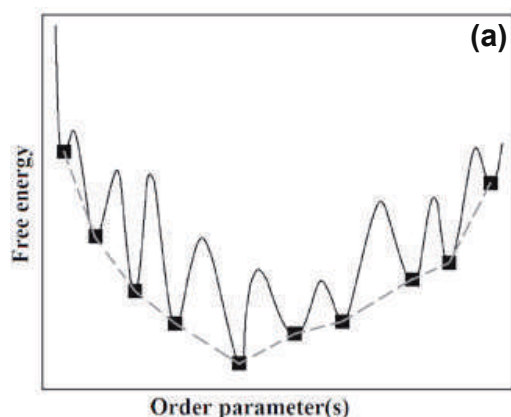


**Figure 3.** Illustration of an evolutionary search: 24 atoms of boron in a fixed cell. The best structure of the first random generation is 0.22 eV/atom above the ground state and is heavily disordered. In the second generation the best structure already contains an incomplete  $B_{12}$  icosahedron, the main building block of the ground-state structure. From Oganov et al. (2009).

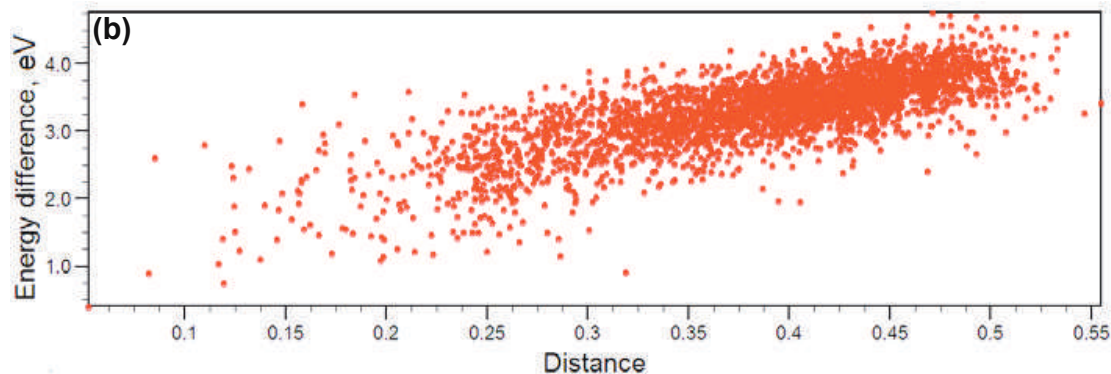
Another important reason is that the energy landscapes expected in chemical systems are likely to have an overall “funnel” shape (Fig. 4a), where lowest-energy structures are clustered in the same region of configuration (or order parameter) space. In such cases, evolutionary algorithms are particularly powerful: they “zoom in” on the most promising region of configuration space until the global minimum is found. This “zooming in” is enabled by selection of lower-energy structures as parents for the subsequent generation, and by the form of the variational operators.

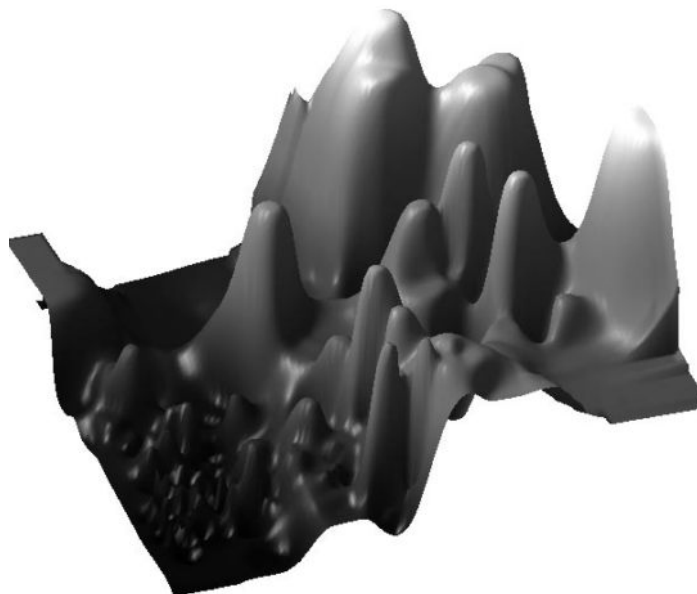
Actually, it is possible to test the assumption of an overall benign landscape shape using a recent approach (Oganov and Valle 2009) that enables mapping of energy landscapes. If the landscape has one funnel (like in Fig. 4a), there will be a direct correlation between the “distance” of all structures from the ground-state structure (this abstract “distance” measures the degree of structural dissimilarity) and the energy relative to the ground state – indeed, in many real systems (for example, GaAs with 8 atoms/cell – Fig. 4b) such a correlation is found. Even when more than one funnel is present, the number of funnels is usually small (up to three or four). Such situations arise when very different atomic arrangements are energetically competitive, and such systems are particularly challenging as the algorithm may tend to get stuck in one particular funnel. To avoid this, several tools can be used – including dense random or quasirandom sampling (to cover all funnels), tabu lists or special constraint techniques (to deal with each funnel, or a group of funnels, separately).

The energy-distance correlations (Fig. 4b) can be considered as 1D-projections of multidimensional energy landscapes. Projections can, actually, be performed on an arbitrary number of dimensions. Particular visual insight comes from 2D-projections that can be obtained by interpolating and smoothing the 2D-plots presented in Oganov and Valle (2009). One such depiction of a landscape (for  $\text{Au}_8\text{Pd}_4$  system) is given in Figure 5.



**Figure 4.** Energy landscapes in chemical systems. (a) A pedagogical cartoon. The original response surface is very “noisy” (i.e., contains very large energy variations, with high barriers). Local optimization reduces this surface to local minima points (black squares). The reduced response surface (dashed line) is well-behaved and has a simple overall shape. This is one of the reasons why the use of local optimization dramatically improves global optimization (Glass et al. 2006). From Oganov et al. (2007). (b) Energy-distance correlation for GaAs (8 atoms/cell). Each point is a locally optimized (i.e., relaxed) structure. The correlation proves that the energy landscape has a simple one-funneled topology. From Oganov and Valle (2009).





**Figure 5.** 2D-representation of the energy landscape of  $\text{Au}_8\text{Pd}_4$  system using method presented in Oganov and Valle (2009). The surface has the same meaning as the dashed line in Figure 4a – it is an interpolation between the points of local minima. Clearly, there is one energy funnel (dark region), which corresponds to different Au-Pd orderings of the underlying fcc-structure.

The overall landscape shape (Figs. 4, 5) implies that, *en route* to the global minimum some of the low-energy metastable minima can be discovered. This is important, as such phases are often interesting as well. Furthermore, metastable structures found during evolutionary simulations provide a deep insight into the structural chemistry of the studied compound. Thus, evolutionary simulations provide three major results – 1) the ground-state structure; 2) a set of low-energy metastable structures; 3) detailed information on the chemical regime of the compound.

### OF THE ALGORITHM

To measure the strengths and weaknesses of the algorithm, we consider several issues:

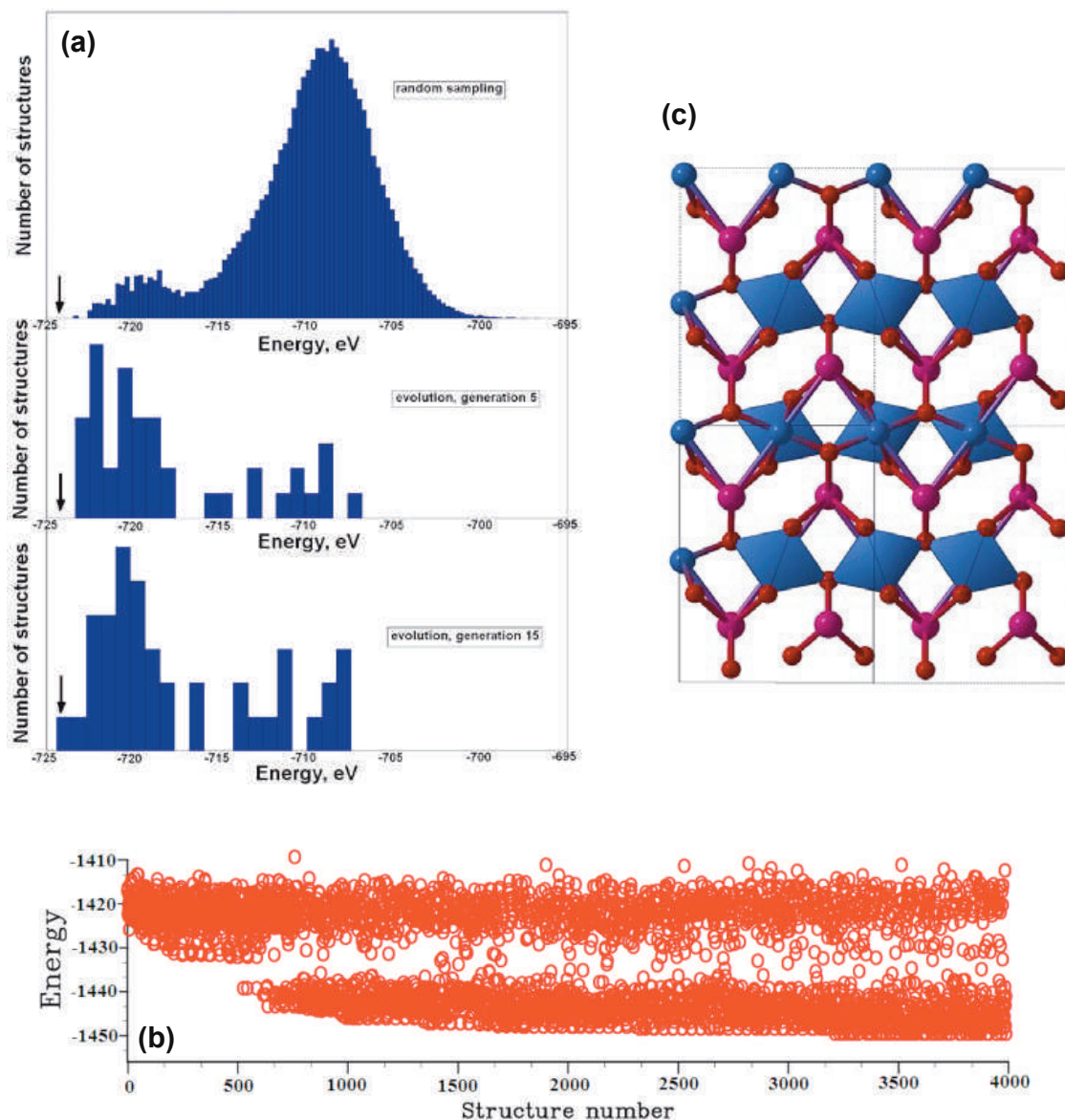
- (1) efficiency of finding the global minimum, in particular relative to a simple well-defined search method, the random sampling,
- (2) size of systems that can be studied in practice,
- (3) how fast the diversity decreases along the evolutionary trajectory.

A number of successful tests have been reported in Oganov and Glass (2006, 2008), Glass et al. (2006), Martoňák et al. (2007), and Oganov et al. (2007). The largest successful test is for a Lennard-Jones crystal with 128 atoms in the (super)cell with variable-cell structure search, which has correctly identified hcp structure as the ground state within 3 generations (each consisting of only 10 structures). For larger Lennard-Jones systems (256 and 512 atoms/cell) we found an energetically very slightly less favorable fcc structure.

The largest test for a chemically complex system is the prediction of the structure of  $\text{MgSiO}_3$  post-perovskite (Oganov and Ono 2004; Murakami et al. 2004) using a relatively large 80-atom supercell (with fixed supercell parameters) and an empirical potential (Murakami et al. 2004) describing interatomic interactions within a partially ionic model. Local optimization and energy calculations were done using the GULP code (Gale 2005). Previously in Martoňák et al. (2007), we have shown that already in a 40-atom supercell this test is unfeasible using

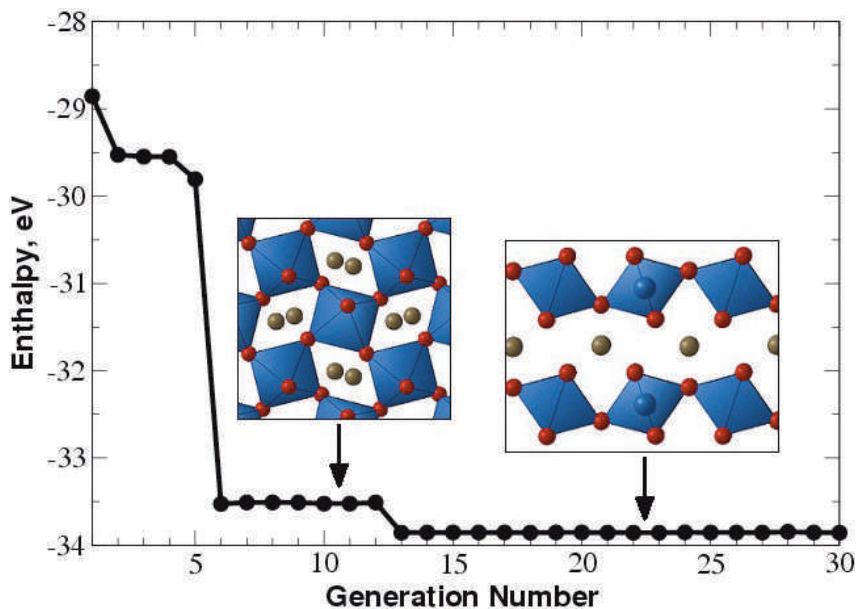
the simple random sampling (with local optimization) (Pickard and Needs 2006): the correct structure was not produced even after  $1.2 \times 10^5$  random attempts, but was found with 600-950 local optimizations of structures produced by USPEX. With 80 atoms/cell the problem becomes much more complicated (one expects an exponential increase of complexity with system size), but even in this case we correctly produced the post-perovskite structure in a reasonable number ( $\sim 3200$ ) of local optimizations – see Figure 6.

Figure 7 shows variable-cell *ab initio* results for  $\text{MgSiO}_3$  at the pressure of 120 GPa. Several runs with somewhat different parameters (but within a reasonable range) have been



**Figure 6.** Evolutionary prediction of the structure of  $\text{MgSiO}_3$  post-perovskite using the experimental cell parameters for a) 40-atom (Martoňák et al. 2007) and b) 80-atom supercells (Oganov and Glass 2008). In both cases, each generation consisted of 41 structures. (a) compares densities of states of optimized structures generated randomly (top) and in the evolutionary run. Random sampling did not find the correct structure within  $1.2 \times 10^5$  steps, whereas in the evolutionary simulation shown it was found within 15 generations (i.e., 600 local optimizations). Arrows mark the ground-state energy. (b) shows the energies of structures along the evolutionary trajectory for the 80-atom run; (c) shows the structure of post-perovskite was obtained within  $\sim 3200$  local optimizations. One can see that the density of low-energy structures increases during the simulation.



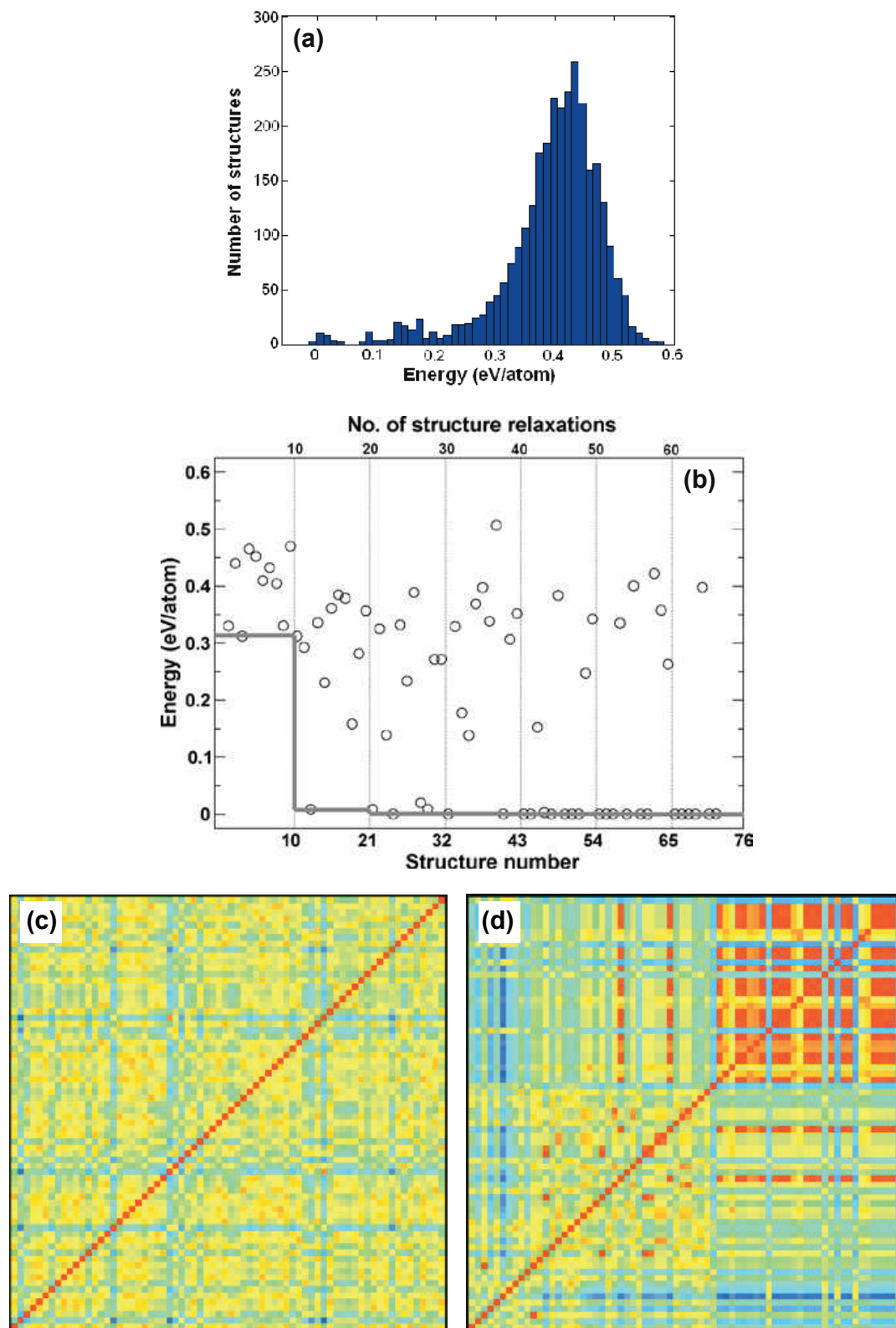


**Figure 7.** Evolutionary structure search for  $\text{MgSiO}_3$  at 120 GPa. Evolution of the lowest enthalpy is shown as a function of the generation (insets show the structures of perovskite and post-perovskite phases), from Oganov and Glass (2006).

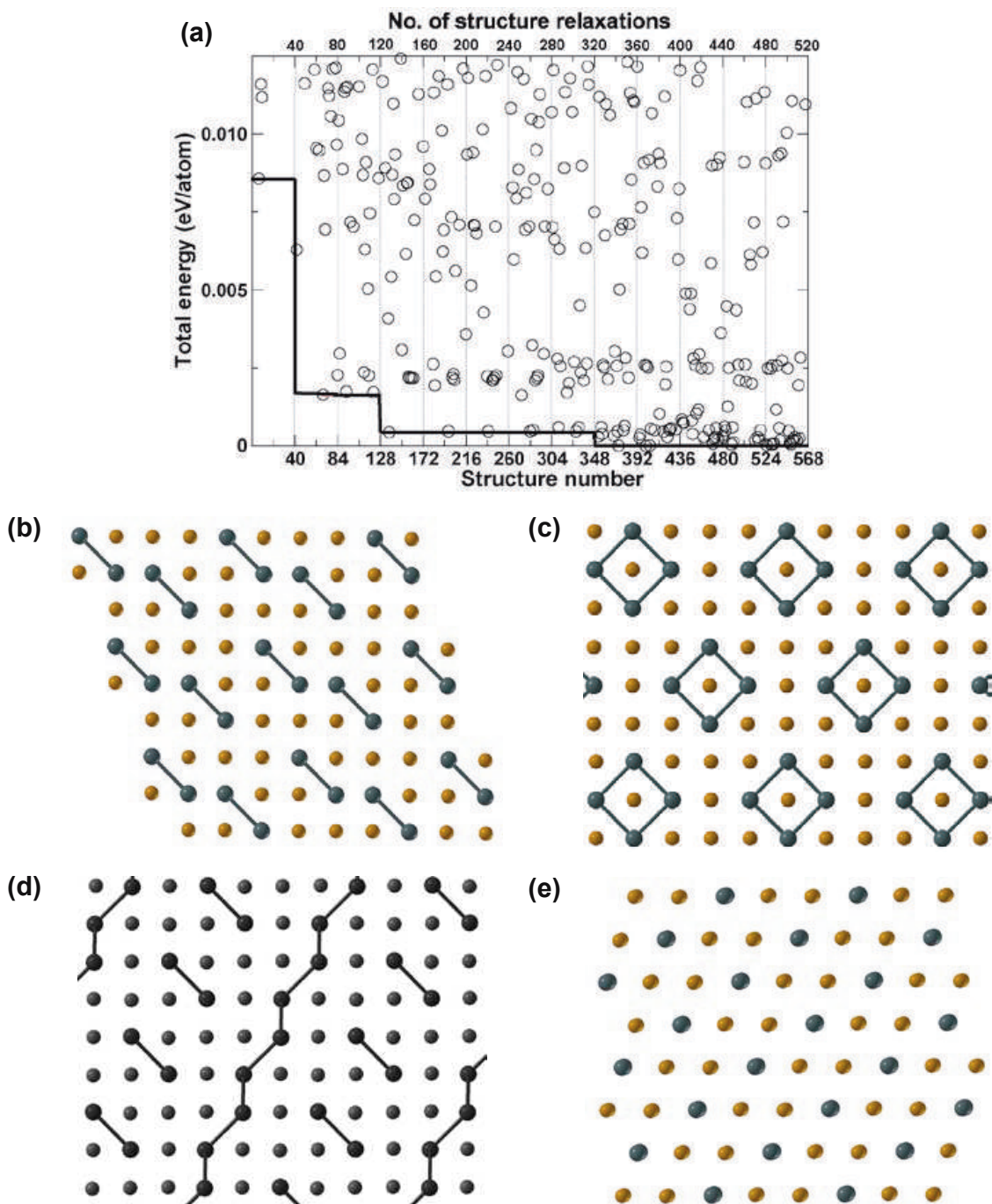
performed and all produced the correct ground-state structure of post-perovskite. The number of local optimizations performed before this structure was found ranged in different runs between 120 and 390; the longest run is shown in Figure 7.

An example of a very simple test, variable-cell *ab initio* structure search for GaAs with 8 atoms/cell (Oganov et al. 2007), is given in Figure 8. The ground-state structure for systems of such size can be found even by local optimization of a reasonable number of randomly produced structures. The density of states of relaxed random structures (Fig. 8a), obtained from 3000 random structures, has a characteristic multimodal shape, which seems to be a general feature of energy landscapes. The stable zincblende structure has the abundance of  $\sim 0.2\%$ , i.e., finding it with random search would on average take  $\sim 500$  local optimizations. In evolutionary simulations (Fig. 8b) it can be found within 3 generations, or just 30 structure relaxations. Similarity matrices for random (Fig. 8c) and evolutionary (Fig. 8d) searches clearly reveal a strong increase of structure similarity (i.e., decrease of diversity, which can be quantified using the approach of Valle and Oganov (2008) and Oganov and Valle (2009)) along the evolutionary run, after finding the global minimum. Even in this extreme case a significant number of dissimilar structures are produced long after the global minimum is found.

$\text{Au}_8\text{Pd}_4$  (12 atoms/cell) is an unusual system, where a number of different ordered decorations of the fcc structure have competitive energies. The ground state of this system is unknown, but was investigated in several computational studies (Curtarolo et al. 2005; Sluiter et al. 2006; Barabash et al. 2006; Oganov et al. 2007). Assuming that the ground-state structure should be an ordered variant of the cubic close-packed (“fcc”) structure and using the cluster expansion technique with parameters calibrated on a set of *ab initio* energies, Barabash et al. (2006) suggested that there are two energetically nearly degenerate structures (Fig. 9c,d). Our calculations found a new ground-state structure (Fig. 9b) that has been overlooked by the previous cluster-expansion study (Barabash et al. 2006) and turned out to be  $\sim 0.1$  meV/atom lower in energy than the previously known lowest-energy structures (Fig. 9c,d). Examination of all the produced structures shows that most of them are different ordering schemes of the fcc-structure and the energy differences are in most cases very small (Fig. 9a).

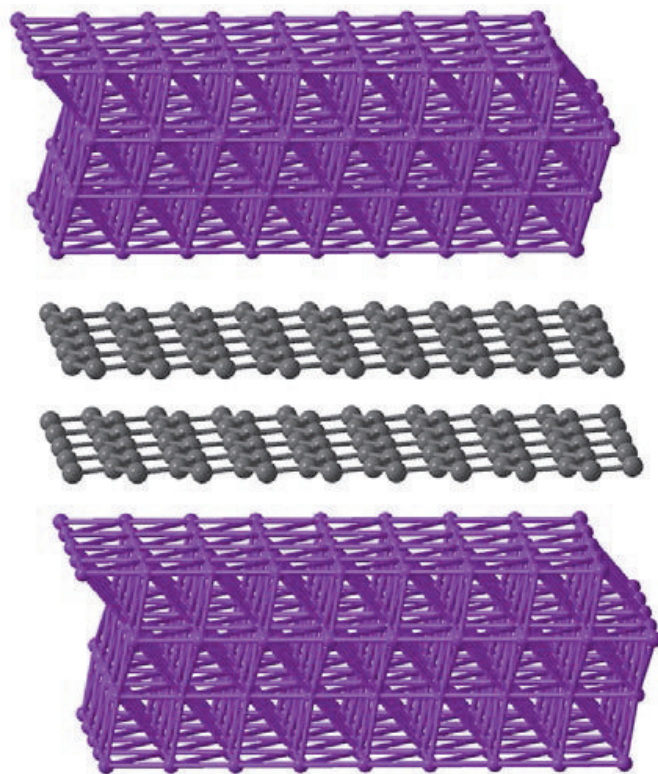


**Figure 8.** Structure prediction for GaAs (8 atoms/cell): a) energy distribution for relaxed random structures, b) progress of an evolutionary simulation (thin vertical lines show generations of structures, and the grey line shows the lowest energy as a function of generation), c-d) similarity matrices (dimension  $70 \times 70$ ) for the random and evolutionary searches, respectively. All energies are relative to the ground-state structure. The evolutionary simulation used a population of 10 structures. Calculations are performed within the GGA (Perdew et al. 1996). From Oganov et al. (2007). Color online.



**Figure 9.** Evolutionary structure search for  $\text{Au}_8\text{Pd}_4$ . a) evolution of the total energies (only the lowest-energy part is shown for clarity), b) the lowest-energy structure found in our evolutionary simulation, c-d) the lowest-energy structures found by cluster expansion in Barabash et al. (2006), e) a suboptimal structure (for discussion, see Oganov et al. 2007, and references therein). Energies are given relative to the ground state.

Periodic boundary conditions suppress decomposition, but when a compound is extremely unstable against decomposition, phase separation can be observed in USPEX simulations. Actually, this happens rather frequently in explorations of hypothetical compositions. A clear example is given by the Cu-C system, which does not have any stable compounds. The tendency to unmixing in this system is very strong and even simulations on small cells show clear separation into layers of fcc-structured Cu and layers of graphite (Fig. 10). When the tendency to unmixing is not so large, simulations on small unit cells may find metastable



**Figure 10.** Lowest-energy structure of  $\text{Cu}_2\text{C}$  with 12 atoms/cell at 1 atm.

“mixed” structures. Such structures have the lowest thermodynamic potential only at the given number of atoms in the unit cell; increasing the cell size would lead to phase separation. In the Cu-C system, phase separation is evident already at very small system sizes (Fig. 10).

### SOME APPLICATIONS OF THE METHOD

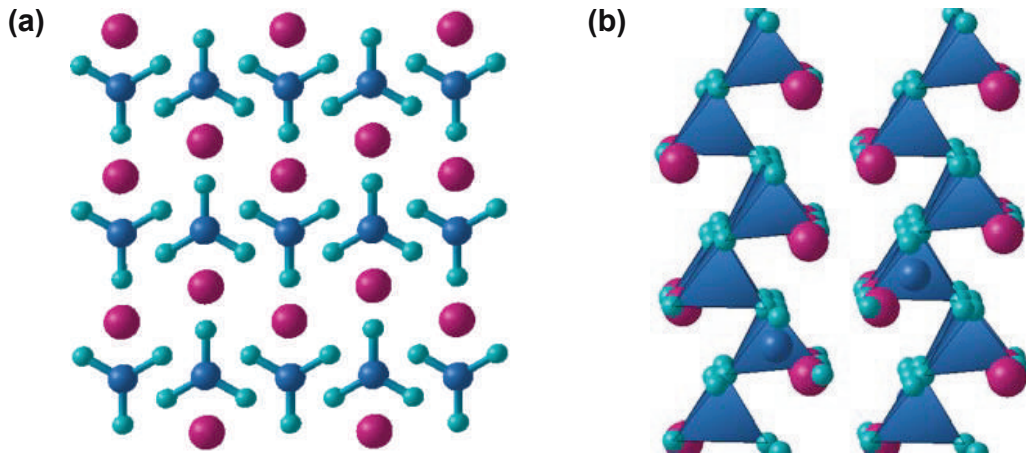
In this section we will review some new insight that has been obtained using our method (see also Oganov and Glass 2006). All structure predictions described here were performed within the generalized gradient approximation (GGA; Perdew et al. 1996) and the PAW method (Blöchl 1994; Kresse and Joubert 1999), using VASP code (Kresse and Furthmüller 1996) for local optimization and total energy calculations. The predicted structures correspond to the global minimum of the approximate free energy surface. For systems where the chosen level of approximation (GGA in cases considered below) is adequate, this corresponds to the experimentally observed structure. Where this is not the case, results of global optimization are invaluable for appraising the accuracy of the approximations.

#### **CaCO<sub>3</sub> polymorphs**

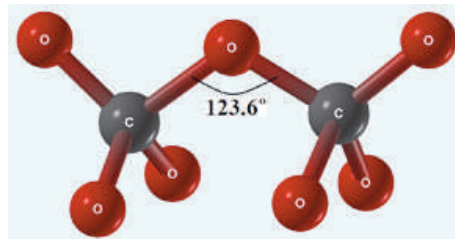
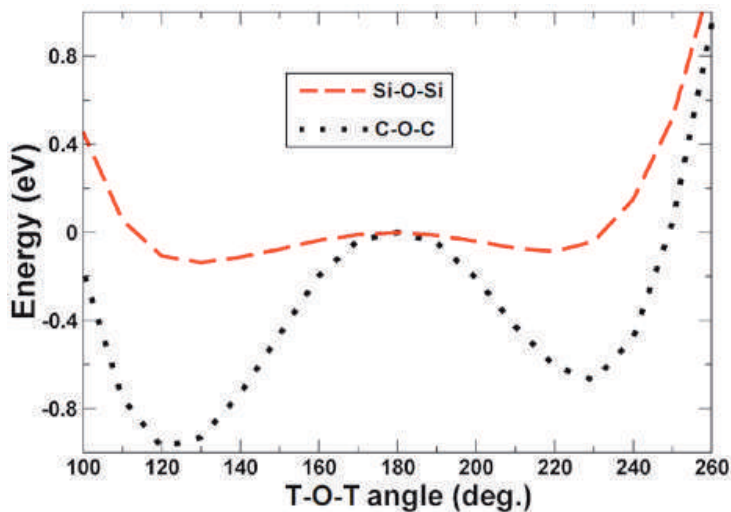
High-pressure behavior of carbonates is very important for the global geochemical carbon cycle, as high-pressure carbonates of Mg and Ca are expected to contain most of the Earth’s oxidized carbon (Shcheka et al. 2006). For  $\text{CaCO}_3$ , there is a well-known transition from calcite to aragonite at  $\sim 2$  GPa, followed by a transition to a post-aragonite phase at  $\sim 40$  GPa (Ono et al. 2005b), the structure of which was solved (Oganov et al. 2006) using USPEX, and the predicted structure matched the experimental X-ray diffraction pattern well. Furthermore, Oganov et al. (2006) have predicted that above 137 GPa a new phase, with space group  $C222_1$  and containing chains of carbonate tetrahedra, becomes stable. Recently this prediction was verified by experiments (Ono et al. 2007) at pressures above 130 GPa. We note that both post-

aragonite and the  $C222_1$  structure (Fig. 11) belong to new structure types and could not have been found by analogy with any known structures.

The presence of tetrahedral carbonate-ions at very high pressures invites an analogy with silicates, but the analogy is limited. In silicates, the intertetrahedral angle Si-O-Si is extremely flexible (Lasaga and Gibbs 1987), which is one of the reasons for the enormous diversity of silicate structure types. Figure 12 shows the variation of the energy as a function of the Si-O-Si angle in the model  $H_6Si_2O_7$  molecule – method borrowed from Lasaga and Gibbs



**Figure 11.**  $CaCO_3$  at high pressure. a) structure of post-aragonite phase, b)  $C222_1$  phase.



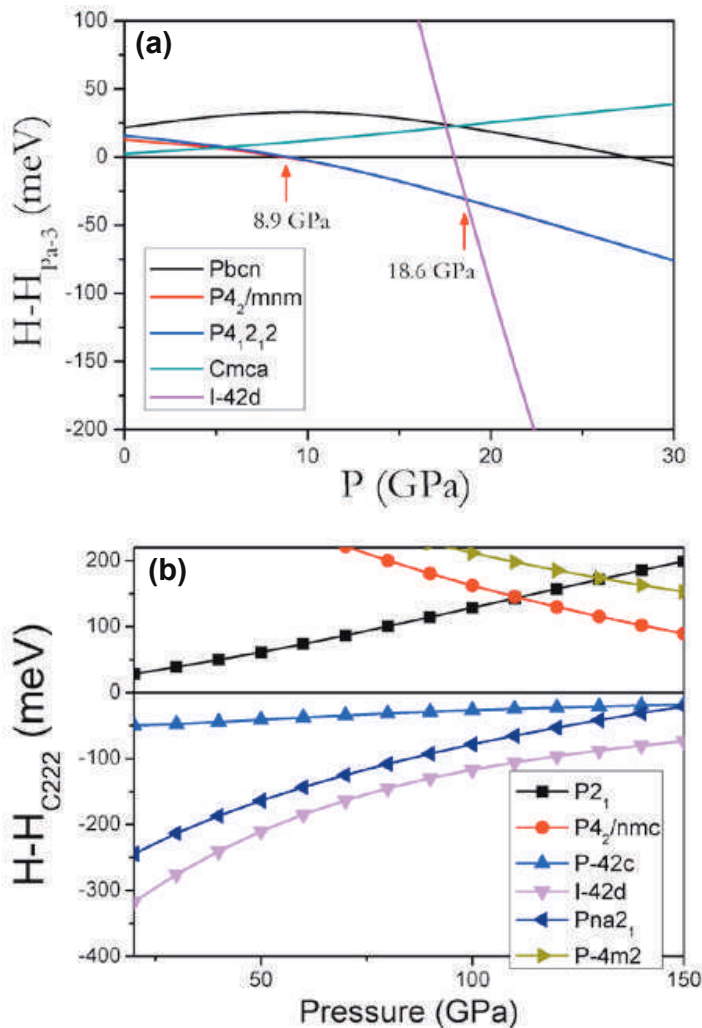
**Figure 12.** Energy variation as a function of the T-O-T angle (dashed line – T = Si, dotted line – T = C). Calculations were performed on  $H_6T_2O_7$  molecules; at each angle all T-O distances and O-T-O valence angles were optimized. Optimum angle C-O-C=124°, Si-O-Si=135°. These calculations were performed with SIESTA code (Soler et al. 2002) using the GGA functional (Perdew et al. 1996), norm-conserving pseudopotentials and a double- $\zeta$  basis set with a single polarization function for each atom.

(1987). One can see only a shallow minimum at  $\angle(\text{Si-O-Si}) = 135^\circ$ , but a deep minimum at  $\angle(\text{C-O-C}) = 124^\circ$  with steep energy variations for  $\text{H}_6\text{C}_2\text{O}_7$  (Fig. 12). This suggests a much more limited structural variety of metacarbonates, compared to silicates. In both  $\text{CaCO}_3$  and  $\text{CO}_2$  the  $\angle(\text{C-O-C})$  angles are close to  $124^\circ$  in a wide pressure range.

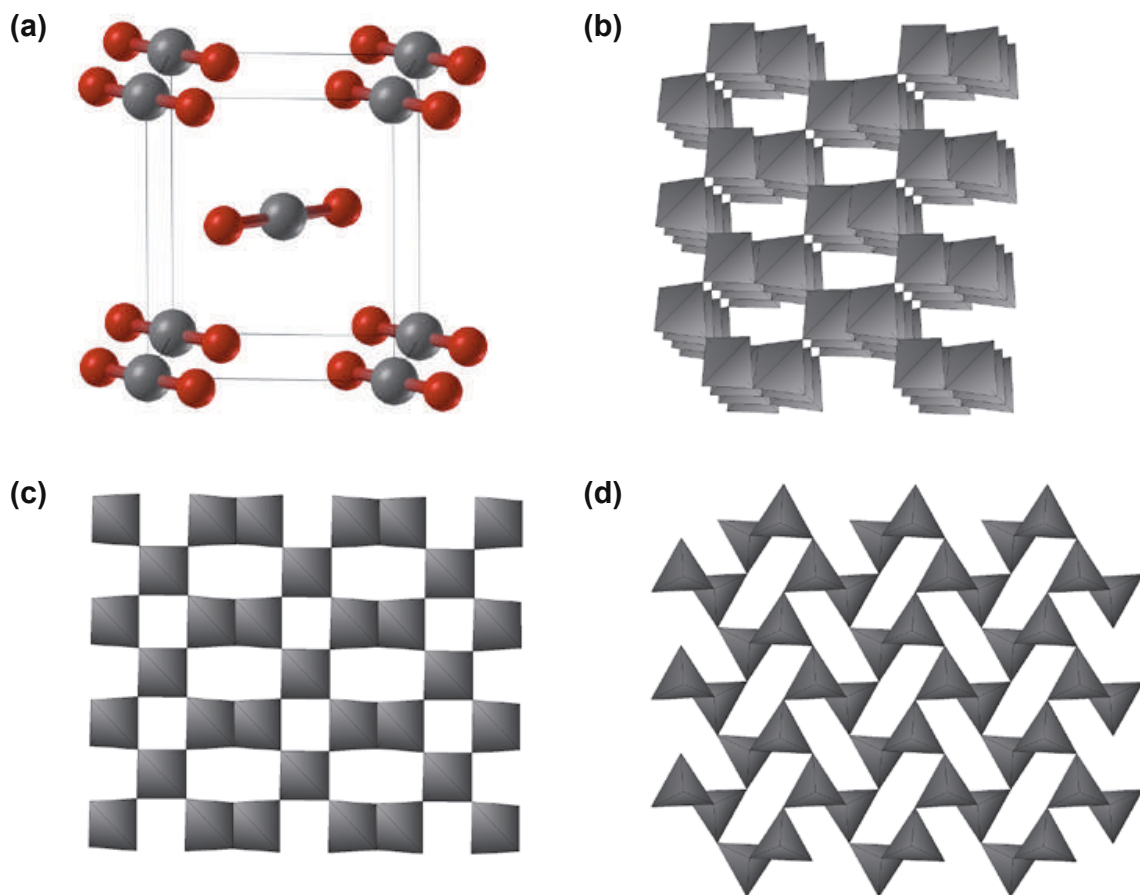
### Polymeric phase of $\text{CO}_2$

High-pressure behavior of  $\text{CO}_2$  is still controversial (Bonev et al. 2003). It is known that above  $\sim 20$  GPa a non-molecular phase (called phase V; Yoo et al. 1999) with tetrahedrally coordinated carbon atoms becomes stable, but its structure is still under debate: in the first experimental study (Yoo et al. 1999) a trydimite structure was proposed, but later theoretical works found it to be unstable (even not metastable) and much less favorable than the  $\beta$ -cristobalite structure (Dong et al. 2000; Holm et al. 2000). At the same time, it was not possible to rule out that there may be even more stable structures. We have performed evolutionary structure searches at 50 GPa, 100 GPa and 150 GPa for systems with 6, 9, 12, 18 and 24 atoms/cell (Oganov et al. 2007, 2008). At all these pressures we confirmed stability of the  $\beta$ -cristobalite structure (Figs. 13 and 14), thus suggesting an experimental re-investigation of phase V of carbon dioxide.  $\text{CO}_2$ -V is stable against decomposition into diamond and oxygen (the enthalpy of decomposition is very large and increases from 3.3 eV to 3.8 eV between 50 GPa and 200 GPa).

At lower pressures, between 8.9 GPa and 18.9 GPa, the  $P4_2/mnm$  phase (see Bonev et al. 2003 for details) is stable, and at even lower pressures (0-8.9 GPa) the  $Pa3$  structure is stable



**Figure 13.** Enthalpies of candidate forms of  $\text{CO}_2$ : a) in the low-pressure region, relative to the molecular  $Pa3$  structure, b) in the high-pressure region, relative to the non-molecular  $C222$  structure. From Oganov et al. (2008).



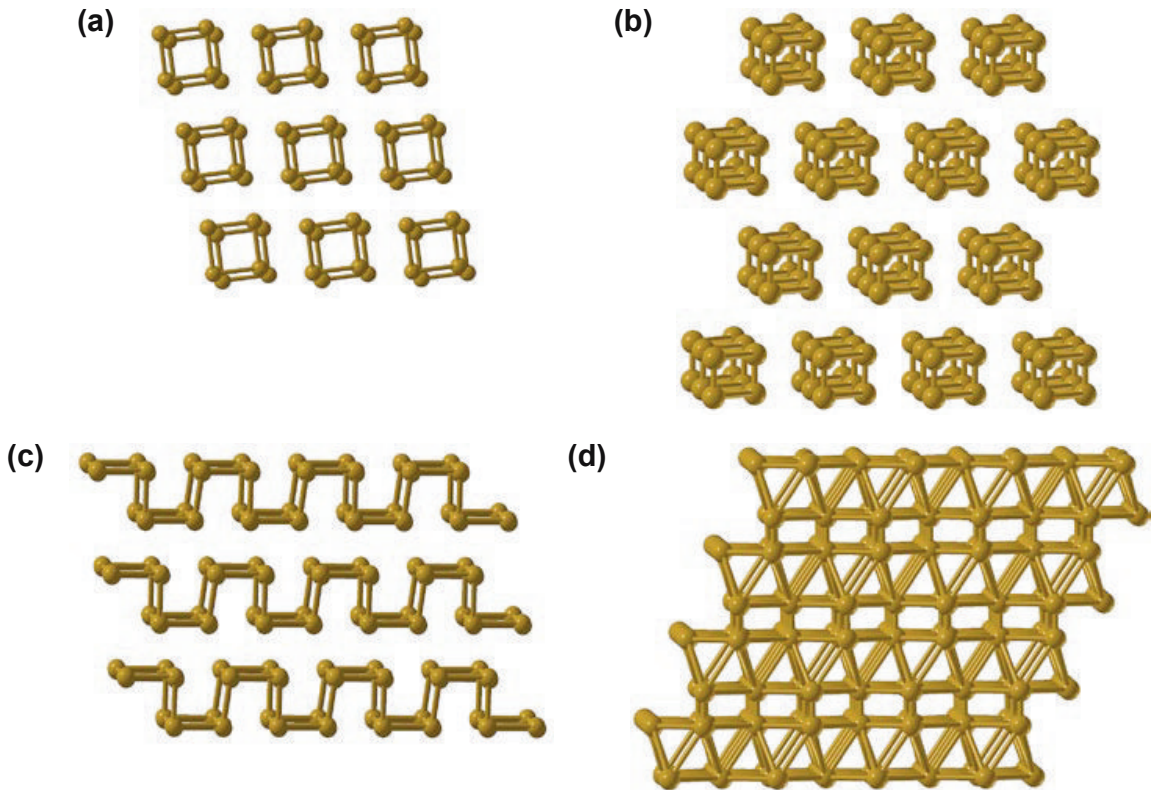
**Figure 14.** CO<sub>2</sub> structures: a) molecular  $P4_2/mnm$  structure, stable at lower pressures than CO<sub>2</sub>-V, b) polymeric  $\beta$ -cristobalite-type form of CO<sub>2</sub>, suggested to be the structure of phase V and showing carbonate tetrahedra. Structural parameters at 100 GPa: space group  $I42d$ ,  $a = b = 3.2906 \text{ \AA}$ ,  $c = 6.0349 \text{ \AA}$ , C(0.5; 0; 0.25), O(0.2739; 0.25; 0.125), c) polymeric  $C222$  structure, d) metastable polymeric  $Pna2_1$  structure. From Oganov et al. (2008).

(Fig. 13). The  $Pa3$ - $P4_2/mnm$  transition pressure calculated here (8.9 GPa) is consistent with experiment and previous calculation (Bonev et al. 2003).

### Semiconducting and metallic phases of solid oxygen: unusual molecular associations

The red  $\epsilon$ -phase of oxygen, stable in the pressure range 8–96 GPa, was discovered in 1979 (Nicol et al. 1979), but its structure was solved only in 2006 (Fujihisa et al. 2006; Lundegaard et al. 2006). The metallic (superconducting at very low temperatures; Shimizu et al. 1998)  $\zeta$ -phase, stable above 96 GPa, was discovered in 1995 (Akahama et al. 1995), and its structure remained controversial for a long time. Neutron diffraction showed that already in the  $\epsilon$ -phase (at 8 GPa) there is no long-range magnetic order and likely even no local moments (Goncharenko 2005). The disappearance of magnetism is a consequence of increasing overlap of molecular orbitals with increasing pressure. Ultimately, orbital overlap leads to metallization. To understand high-pressure chemistry of oxygen, we performed extensive structure searches at pressures between 25 GPa and 500 GPa, taking into account only non-magnetic solutions (Oganov and Glass 2006; Ma et al. 2007).

At 25 GPa, we found two particularly interesting structures – one consisting of zigzag chains of O<sub>2</sub> molecules ( $Cmcm$  structure of Neaton and Ashcroft 2002 and Oganov and Glass 2006; see Fig. 15b) and one with more complex chains of molecules (see Fig. 15c). These have strong similarities with the experimentally observed structure (Lundegaard et al. 2006; Fujihisa

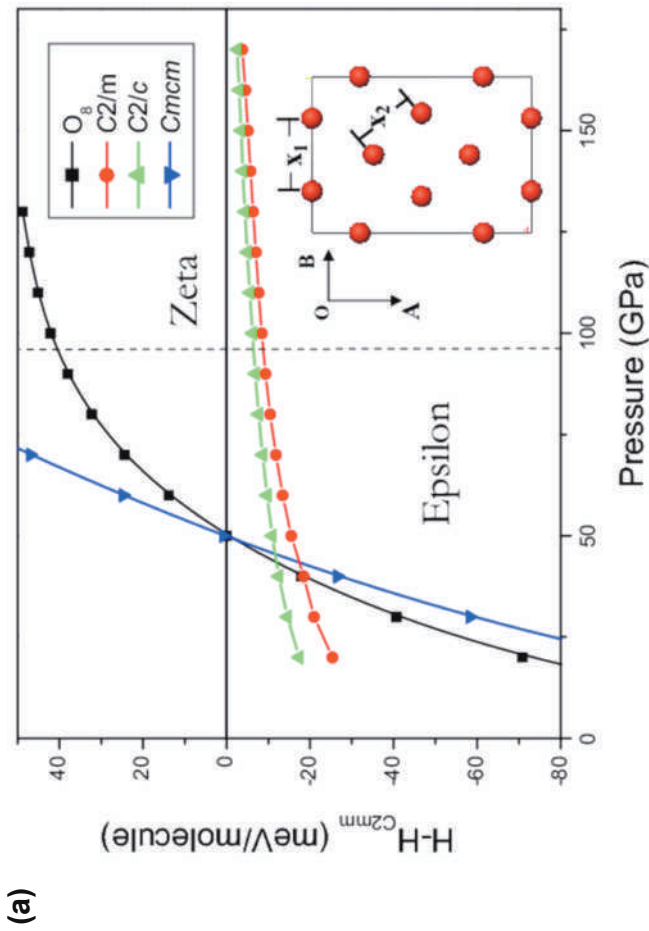
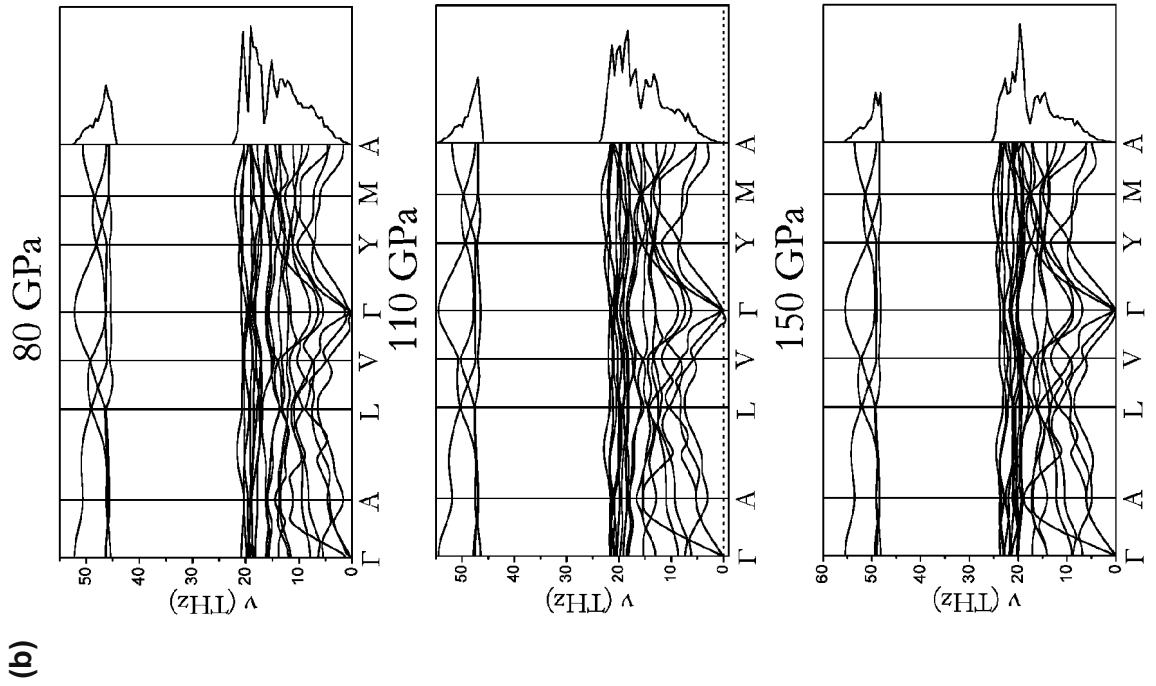


**Figure 15.** High-pressure structures of oxygen: a) experimentally found  $\epsilon$ - $O_8$  structure at 17.5 GPa (Lundegaard et al. 2006), b)  $Cmcm$  chain structure (Neaton and Ashcroft 2002; Oganov and Glass 2006), c) metastable  $P\bar{1}$  chain structure at 25 GPa (Oganov and Glass 2006), d)  $C2/m$  structure of the  $\zeta$ -phase at 130 GPa (Ma et al. 2007). Contacts up to 2.2 Å are shown as bonds. From Oganov et al. (2007).

et al. 2006; see Fig. 15a) consisting of  $O_8$  clusters: all of these structures are molecular, and in all of them each molecule is connected with two other molecules, at distances of  $\sim 2.1$ - $2.2$  Å (the intermolecular distance is  $\sim 1.2$  Å). The  $Cmcm$  structure, first suggested in Neaton and Ashcroft (2002), is the true GGA ground state, but it differs from experiment; as Figure 16a shows, its enthalpy is  $\sim 10$  meV/atom lower than for the experimentally found structure (Fig. 15a). Metastability of the experimentally studied structure cannot yet be ruled out, but it seems likely that this discrepancy is rather due to deficiencies of the GGA. Molecules in the  $(O_2)_4$  clusters interact by weak intermolecular covalent bonds: each  $O_2$  molecule has two unpaired electrons occupying two molecular  $\pi^*$ -orbitals, and sharing these electrons with neighboring molecules creates two intermolecular bonds per molecule and a non-magnetic ground state (Ma et al. 2007; Stuedel and Wong 2007). It is well known that DFT-GGA does not perform well for stretched covalent bonds, the root of the problem being in the locality of the exchange-correlation hole in DFT-GGA, whereas the true exchange-correlation hole in such cases is highly delocalized. At high pressure, intermolecular distances decrease, intermolecular bonds become more similar to normal covalent bonds and the true exchange-correlation hole becomes more localized. Therefore, we can apply the GGA with greater confidence for the prediction of the structure of the metallic  $\zeta$ -phase.

For the  $\zeta$ -phase, evolutionary simulations at 130 GPa and 250 GPa uncovered two interesting structures with  $C2/m$  and  $C2/c$  space groups (Ma et al. 2007). These have very similar enthalpies (Fig. 16a); the  $C2/m$  structure is slightly lower in enthalpy and matches experimental X-ray diffraction and Raman spectroscopy data very well, better than the  $C2/c$  structure (Ma et al. 2007). Both structures contain well-defined  $O_2$  molecules; our simulations show that oxygen remains a molecular solid at least up to 500 GPa. Phonon dispersion curves





**Figure 16.** High-pressure phases of oxygen: a) enthalpies (relative to the  $C2/m$  structure of Serra et al. 1998) of several possible structures as a function of pressure (from Ma et al. 2007), b) phonon dispersion curves and densities of states of the  $C2/m$   $\zeta$ -phase at three pressures. From Oganov et al. (2007).

of the  $C2/m$  structure (Fig. 16b) contain clearly separated molecular vibrons and show that the structure is dynamically stable, except at 110 GPa, where we see tiny imaginary frequencies in the  $\Gamma$ -V direction, close to the Brillouin zone center. Such soft modes may result in small-amplitude long-wavelength modulations of the structure at very low temperatures.

The  $\varepsilon$ - $\xi$  transition is isosymmetric, which implies that it is first-order at low temperatures but can become fully continuous above some critical temperature (Christy 1995). Given the small volume discontinuity upon transition and small hysteresis (one can obtain the  $C2/m$  structure of the  $\xi$ -phase by simple overcompression of the  $\varepsilon$ - $O_8$  structure,  $\sim 5$  GPa above the thermodynamic transition pressure), one can expect this critical temperature to be rather low. We note that within the GGA the  $\varepsilon$ - $\xi$  transition is predicted to occur at 45 GPa (Fig. 16a), much lower than the experimental transition pressure (96 GPa). This has two explanations – (i) as the GGA is expected to perform better for the metallic  $\xi$ -phase than for the semiconducting  $\varepsilon$ - $O_8$  phase, the enthalpy differences are expected to suffer from non-cancelling errors, (ii) since the  $\varepsilon$ - $\xi$  transition is not only structural, but also electronic (insulator-metal transition), one might expect metallization at lower pressures than in experiment. Typically, density functional calculations overstabilize metallic states relative to insulating ones, and this is exactly what happens in oxygen. The predicted  $C2/m$  structure of the  $\xi$ -phase was very recently confirmed by single-crystal experiments (Weck et al. 2009).

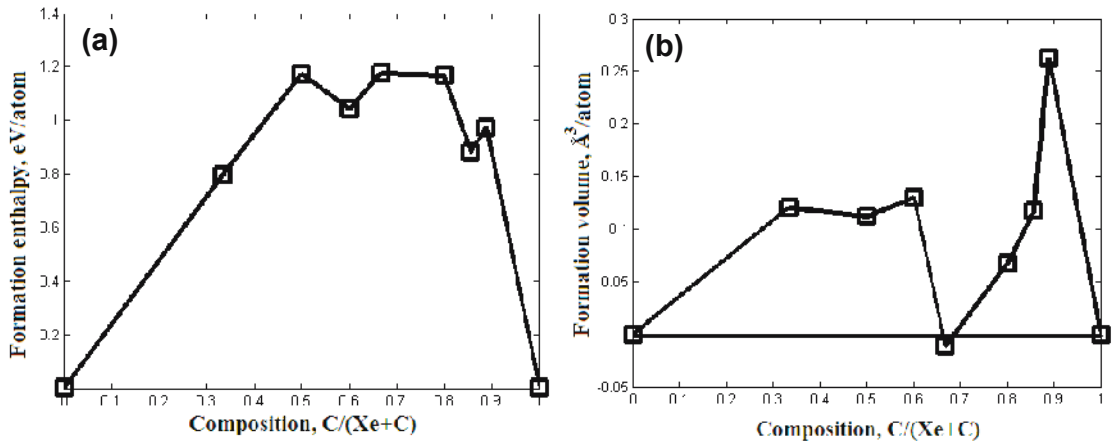
### **Reactivity of noble gases: are Xe-C compounds possible at high pressure?**

Inducing major changes in the electronic structure of atoms, high pressure may also change their reactivity. For instance, noble (i.e., largely unreactive) metal platinum under pressure easily forms carbide PtC (Oganov and Ono 2004; Ono et al. 2005a) and dinitride PtN<sub>2</sub> (Gregoryanz et al. 2004). One should not confuse chemical reactivity with propensity to phase transitions: so recently it was concluded that gold loses its “nobility” at 240 GPa, when it undergoes an fcc-hcp structural transition (Dubrovinsky et al. 2007). Structural transitions and reactivity are unrelated notions, however: in spite of becoming reactive, Pt does not change its fcc structure, and Cu (not a noble metal by any standards) is only known in one crystalline phase with the fcc structure.

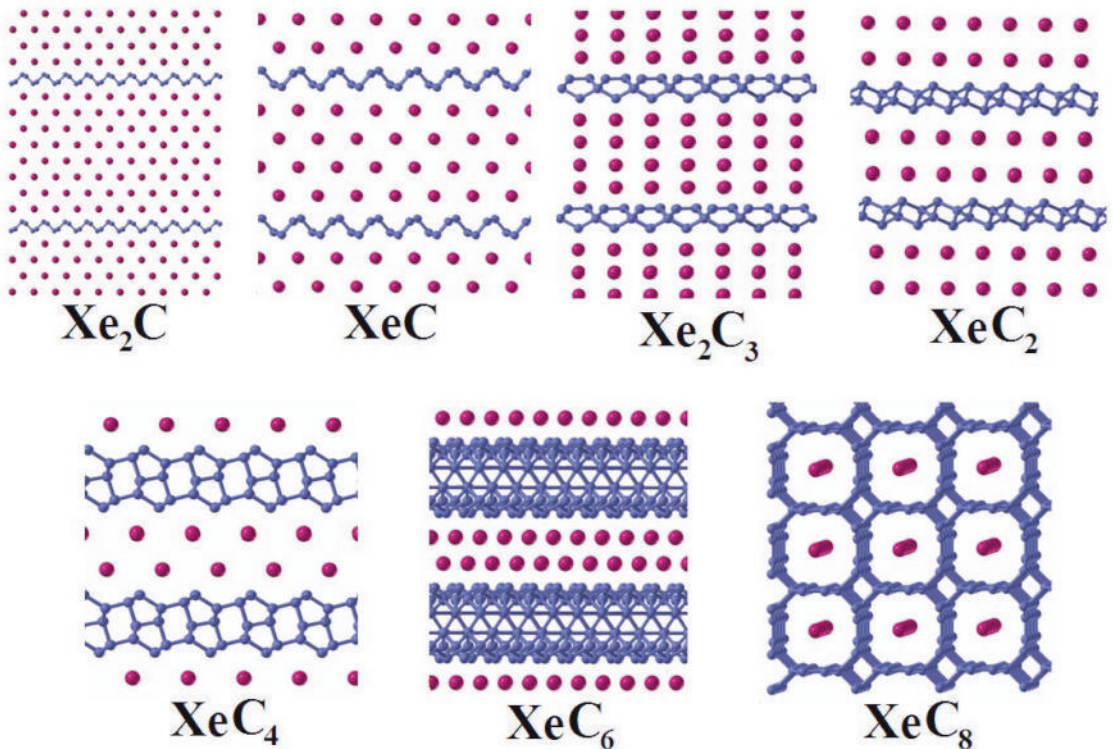
An interesting question is whether noble gases become reactive. Indeed, it was observed that a few percent Xe can be incorporated in quartz (SiO<sub>2</sub>) at elevated pressures and high temperatures (Sanloup et al. 2005). A possibility has been suggested by Grochala (2007) that stable Xe-C compounds may be stable at high pressure; indeed, carbon and xenon have similar valence orbital energies (cf. ionization potentials of 12.13 eV and 11.26 eV for Xe and C, respectively) and one expects that pressure would make Xe more reactive (Sanloup et al. 2005). We did simulations at 200 GPa, i.e., above the metallization pressure of Xe (132 GPa; Goettel et al. 1989), when its closed electronic shells are strongly perturbed. These calculations were done within the GGA (Perdew et al. 1996) and on cells containing up to 14 atoms/cell. At this pressure all Xe carbides are extremely unstable (Fig. 17) and their structures (Fig. 18) show clear separation into close-packed Xe layers (i.e., fragments of the elemental Xe structure) and 3,4-connected carbon layers (intermediate between graphite and diamond). The only exception is the 3D-clathrate structure of XeC<sub>8</sub>. The observed layering is consistent with the instability to decomposition. Although Xe carbides are unstable at 200 GPa, already at that pressure we observe considerable bonding Xe-C interactions and the effect of Xe on the carbon sublattice is far beyond simple mechanistic size factor – the carbon layers adopt unusual and very interesting configurations that may be prepared in the laboratory under certain conditions.

### **Boron: novel phase with a partially ionic character**

Boron is perhaps the most enigmatic element: at least 16 phases were reported in the literature, but most are believed or suspected to be compounds (rather than forms of the pure element), and until recently the phase diagram was unknown. A number of important results started with experimental findings of J. Chen and V. L. Solozhenko (both arrived independently

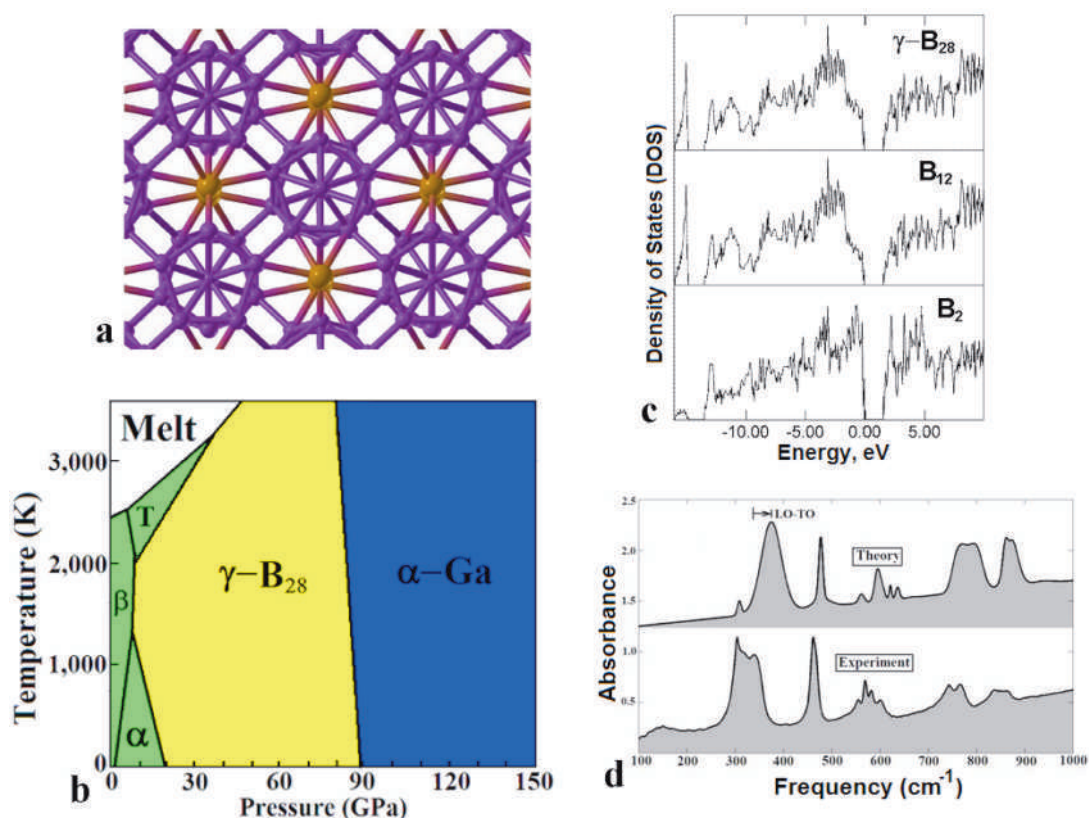


**Figure 17.** Predicted enthalpy (a) and volume (b) of formation of Xe-C compounds at 200 GPa. The compounds shown are Xe (hcp), Xe<sub>2</sub>C, XeC, Xe<sub>2</sub>C<sub>3</sub>, XeC<sub>2</sub>, XeC<sub>4</sub>, XeC<sub>6</sub>, XeC<sub>8</sub>, C(diamond). Note that XeC<sub>2</sub> has a small negative volume of formation and might become stable at much higher pressures. From Oganov et al. (2007).

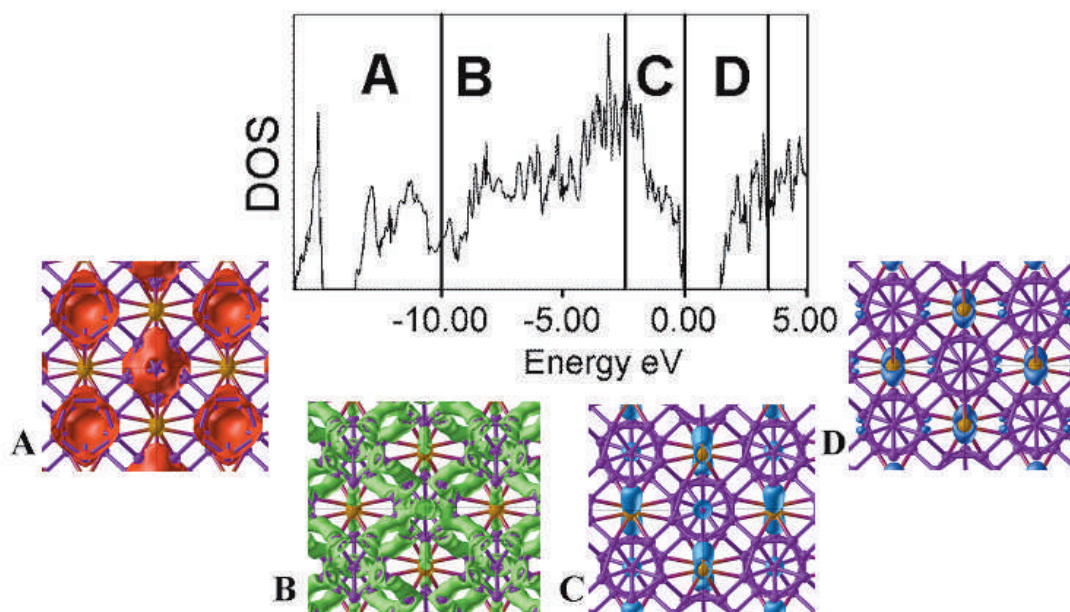


**Figure 18.** Predicted structures of Xe<sub>2</sub>C, XeC, Xe<sub>2</sub>C<sub>3</sub>, XeC<sub>2</sub>, XeC<sub>4</sub>, XeC<sub>6</sub>, XeC<sub>8</sub> at 200 GPa. From Oganov et al. (2007).

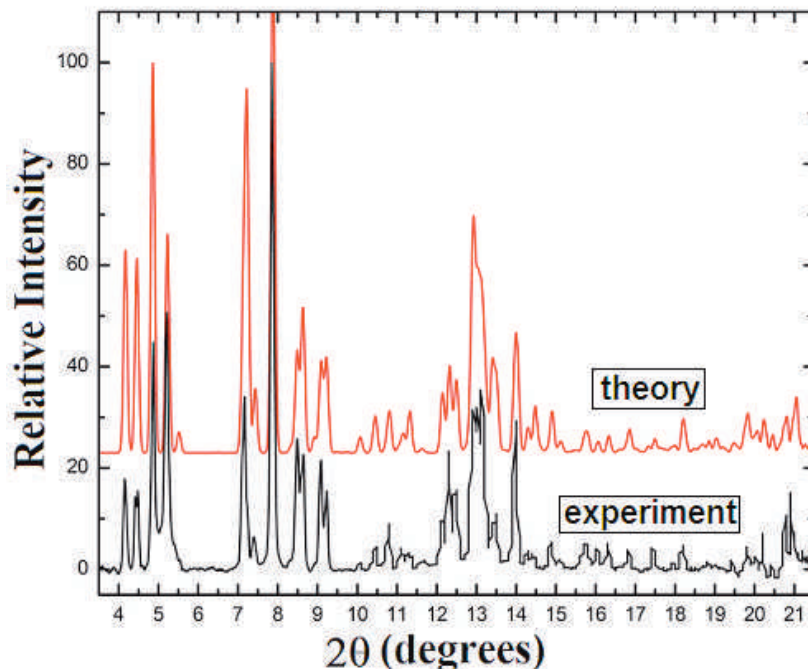
at the same conclusions in 2004) of a new phase at pressures above 10 GPa and temperatures of 1800-2400 K, though at first the structure of this new phase could not be determined from experimental data alone. We found the structure using USPEX and named this phase  $\gamma$ -B<sub>28</sub> (because it contains 28 atoms/cell). Its structure has space group *Pnmm* and is comprised of icosahedral B<sub>12</sub> clusters and B<sub>2</sub> pairs in a NaCl-type arrangement. This phase is stable between 19 and 89 GPa, and exhibits sizable charge transfer from B<sub>2</sub> pairs to B<sub>12</sub> clusters, quite unexpected for a pure element. Details are given in Oganov et al. (2009) and Figures 19 and 20. Figure 21 shows a comparison of theoretical and experimental X-ray powder diffraction profiles.



**Figure 19.** Boron: (a) structure of  $\gamma$ -B<sub>28</sub> (B<sub>12</sub> icosahedra and B<sub>2</sub> pairs are marked by different colors). (b) phase diagram of boron, showing a wide stability field of  $\gamma$ -B<sub>28</sub>. (c) electronic DOS and its projections onto B<sub>12</sub> and B<sub>2</sub> units (all DOSs are normalized per atom), (d) comparison of theoretical and experimental IR spectra. IR spectra indicate the presence of non-zero Born charges on atoms. From Oganov et al. 2009.



**Figure 20.**  $\gamma$ -B<sub>28</sub>: total electronic DOS and energy-decomposed electron densities. Lowest-energy valence electrons are dominated by the B<sub>12</sub> icosahedra, while top of the valence band and bottom of the conduction band (i.e., holes) are localized on the B<sub>2</sub> pairs. This is consistent with atom-projected DOSs (Fig. 19 c) and the idea of charge transfer B<sub>2</sub> → B<sub>12</sub>. Color online.



**Figure 21.** Comparison of theoretical and experimental X-ray powder diffraction profiles of  $\gamma$ -B<sub>28</sub>. X-ray wavelength  $\lambda = 0.31851$  Å. From Oganov et al. 2009.

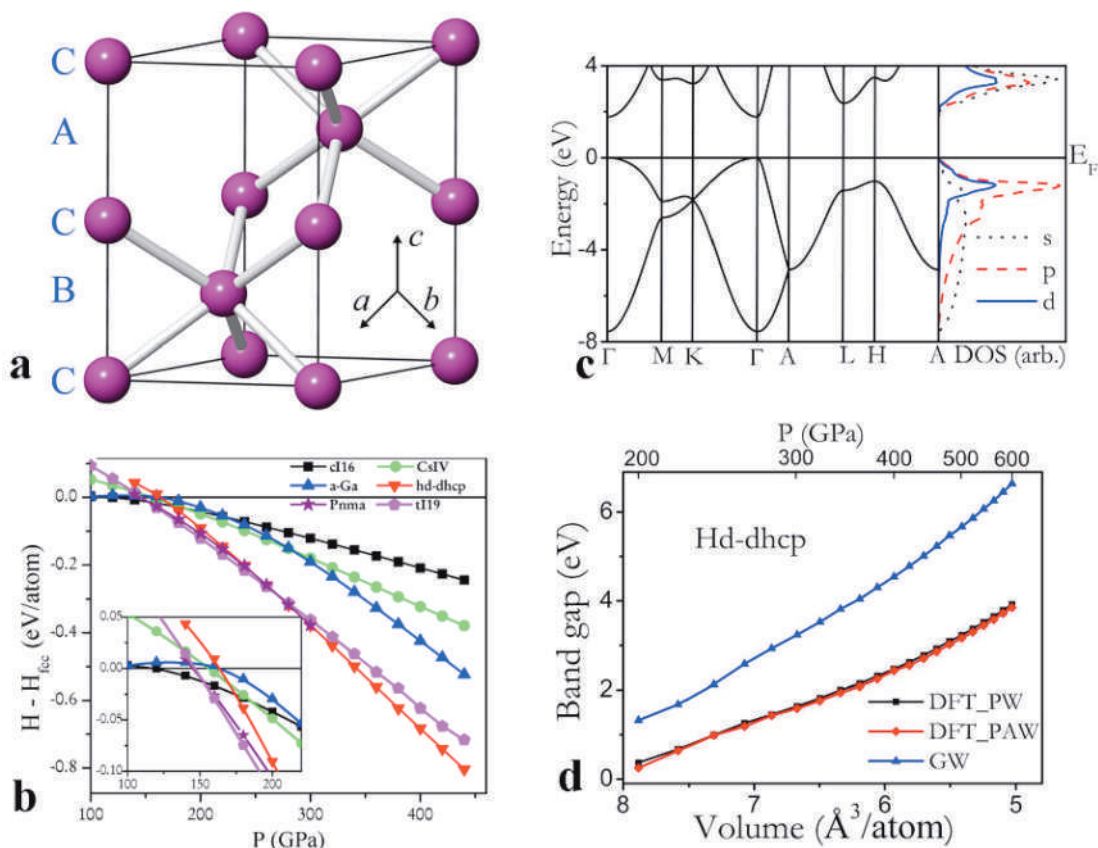
$\gamma$ -B<sub>28</sub> can be represented as a “boron boride” (B<sub>2</sub>)<sup>δ+</sup>(B<sub>12</sub>)<sup>δ-</sup>; although the exact value of the charge transfer  $\delta$  depends on the definition of an atomic charge, for all definitions that we used the qualitative picture is the same. Perhaps the most reliable definition of a charge, due to Bader (1990), gives  $\delta \sim 0.5$  (Oganov et al. 2009). Based on the similarity of synthesis conditions and many diffraction peaks, it seems likely that the same high-pressure boron phase may have been observed by Wentorf in 1965. However, material was generally not believed to be pure boron (due to the sensitivity of boron to impurities and lack of chemical analysis or structure determination in Wentorf 1965) and its diffraction pattern was deleted from Powder Diffraction File database.  $\gamma$ -B<sub>28</sub> is structurally related to several compounds – for instance, B<sub>6</sub>P (Amberger and Rauh 1974) or B<sub>13</sub>C<sub>2</sub> (Kwei and Morosin 1996), where the two sublattices are occupied by different chemical species (instead of interstitial B<sub>2</sub> pairs there are P atoms or C-B-C groups, respectively). Significant charge transfer can be found in other elemental solids, and observations of dielectric dispersion (Tsagareishvili et al. 2009), equivalent to LO-TO splitting, suggest it for  $\beta$ -B<sub>106</sub>. The nature of the effect is possibly similar to  $\gamma$ -B<sub>28</sub>. Detailed microscopic understanding of charge transfer in  $\beta$ -B<sub>106</sub> would require detailed knowledge of its structure, and reliable structural models of  $\beta$ -B<sub>106</sub> finally begin to emerge from computational studies (van Setten et al. 2007; Widom and Mikhalkovic 2008; Ogitsu et al. 2009). It is worth mentioning that  $\gamma$ -B<sub>28</sub> is a superhard phase, with a measured Vickers hardness of 50 GPa (Solozhenko et al. 2008), which puts it among half a dozen hardest materials known to date.

### **Sodium: a metal that goes transparent under pressure**

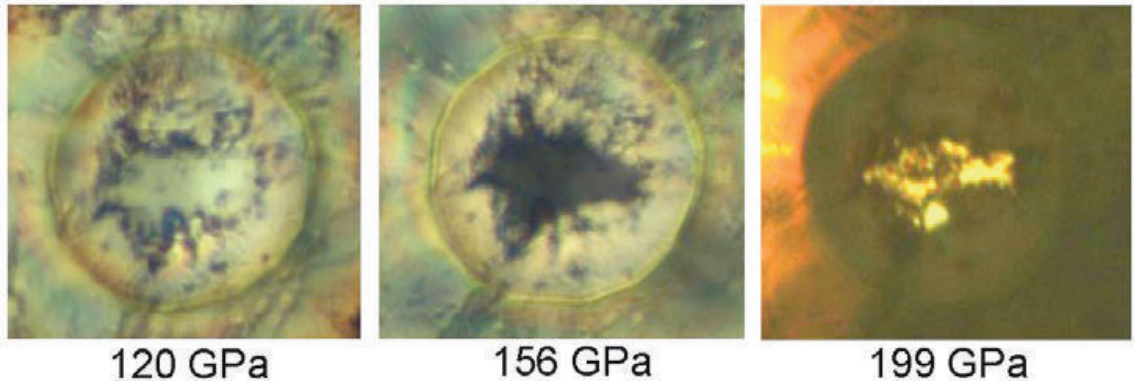
A sequence of recent discoveries demonstrated that sodium, a simple s-element at normal conditions, behaves in highly non-trivial ways under pressure. The discovery of an incommensurate host-guest structure (Hanfland et al. 2002), followed by finding of several complex structures (Gregoryanz et al. 2008) in the range of pressures corresponding to the minimum of the melting curve (Gregoryanz et al. 2005), and the very existence of that extremely deep minimum in the melting curve at about 110 GPa – all this evidence points to some unusual changes in the physics of sodium. Later it was shown also that the incommensurate

host-guest structure is a quasi-1D-metal (Lazicki et al. 2009), where conductivity is primarily due to chains formed by the guest sublattice. Yet another unusual phenomenon was predicted using USPEX and later (but within the same paper by Ma et al. 2009a) verified experimentally: on further compression sodium becomes a wide-gap insulator! This happens at  $\sim 190$  GPa, and Figure 22 shows the crystal structure of the insulating “hP4” phase, its enthalpy relative to other structures, and the electronic structure. The structure can be described as a double hexagonal close-packed (dhcp) structure, squeezed more than twice along the  $c$ -axis, as a result of which sodium atoms have 6-fold coordination. There are 2 inequivalent Na positions: Na1 and Na2, which have the octahedral and trigonal-prismatic coordination, and the hP4 structure can be described as the elemental analog of the NiAs structure type (the same way as diamond is the elemental analog of the zincblende structure type). Calculations suggest that sodium is no longer an s-element; instead, its outermost valence electron has significant s-, p- and d-characters (Fig. 22c). Strongly compressed sodium can be considered as a transition metal, because of its significant d-character.

The band gap is direct, and increases with pressure. At 200 GPa the bandgap calculated with the GW approximation (known to give rather accurate results) is 1.3 eV, and increases to 6.5 eV at 600 GPa. These predictions implied that above 200 GPa sodium will be red and transparent, and at  $\sim 300$  GPa it will become colorless and transparent (like wide-gap insulators). This has indeed been confirmed in experiments of M. I. Eremets (Ma et al. 2009a) as shown in Figure 23. The insulating behavior is explained by the extreme localization of the valence electrons in the interstices of the structure, i.e., the “empty” space (Fig. 24). These areas of localization are characterized by surprisingly high values of the electron localization function (nearly



**Figure 22.** Summary on the hP4 phase of sodium: a) its crystal structure, b) enthalpies of competing high-pressure phases (relative to the fcc structure), c) band structure, d) pressure dependence of the band gap, indicating rapid increase of the band gap on compression. From Ma et al. (2009a).



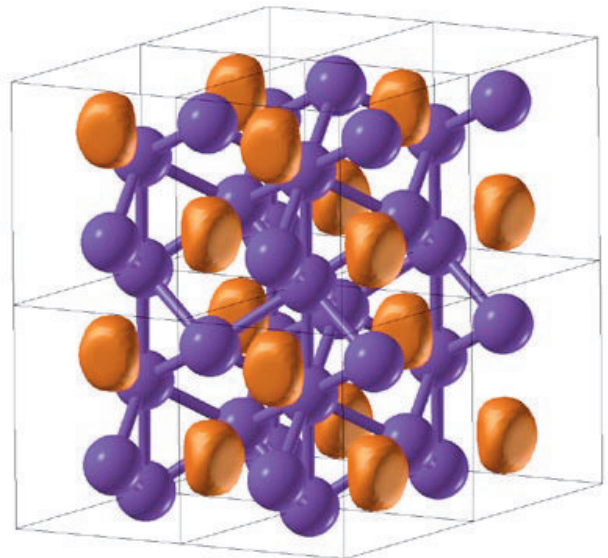
**Figure 23.** Photographs of sodium samples under pressure. At 120 GPa the sample is metallic and highly reflective, at 156 GPa the reflectivity is very low, and at 199 GPa the sample is transparent. From Ma et al. (2009a).

1.0) and maxima of the total electron density. The number of such maxima is half the number of sodium atoms, and therefore in a simple model we can consider Na atoms as completely ionized ( $\text{Na}^+$ ), and interstitial maxima as containing one entire electron pair. The hP4 structure can also be described as a  $\text{Ni}_2\text{In}$ -type structure, where Na atoms occupy positions of Ni atoms, and interstitial electron pairs in hP4-Na sit on the same positions as In atoms in  $\text{Ni}_2\text{In}$ . At first counter intuitively, the degree of localization of the interstitial electron pairs increases with pressure, explaining the increase of the band gap (Fig. 22d). hP4-Na can be described as an electride, i.e., an ionic “compound” formed by ionic cores and localized interstitial electron pairs. The very fact that sodium, one of the best and simplest metals, under pressure becomes a transparent insulator with localized valence electrons, is remarkable and forces one to reconsider classical ideas of chemistry.

Interstitial charge localization can be described in terms of (*s*)-*p*-*d* orbital hybridizations, and its origins are in the exclusionary effect of the ionic cores on valence electrons: valence electrons, feeling repulsion from the core electrons, are forced into the interstitial regions at pressures where atomic cores begin to overlap (Neaton and Ashcroft 1999).

## CONCLUSIONS

Evolutionary algorithms, based on physically motivated forms of variation operators and local optimization, are a powerful tool enabling reliable and efficient prediction of stable crystal structures. This method has a wide field of applications in computational materials design (where experiments are time-consuming and expensive) and in studies of matter at extreme conditions (where experiments are very difficult or sometimes beyond the limits of feasibility).



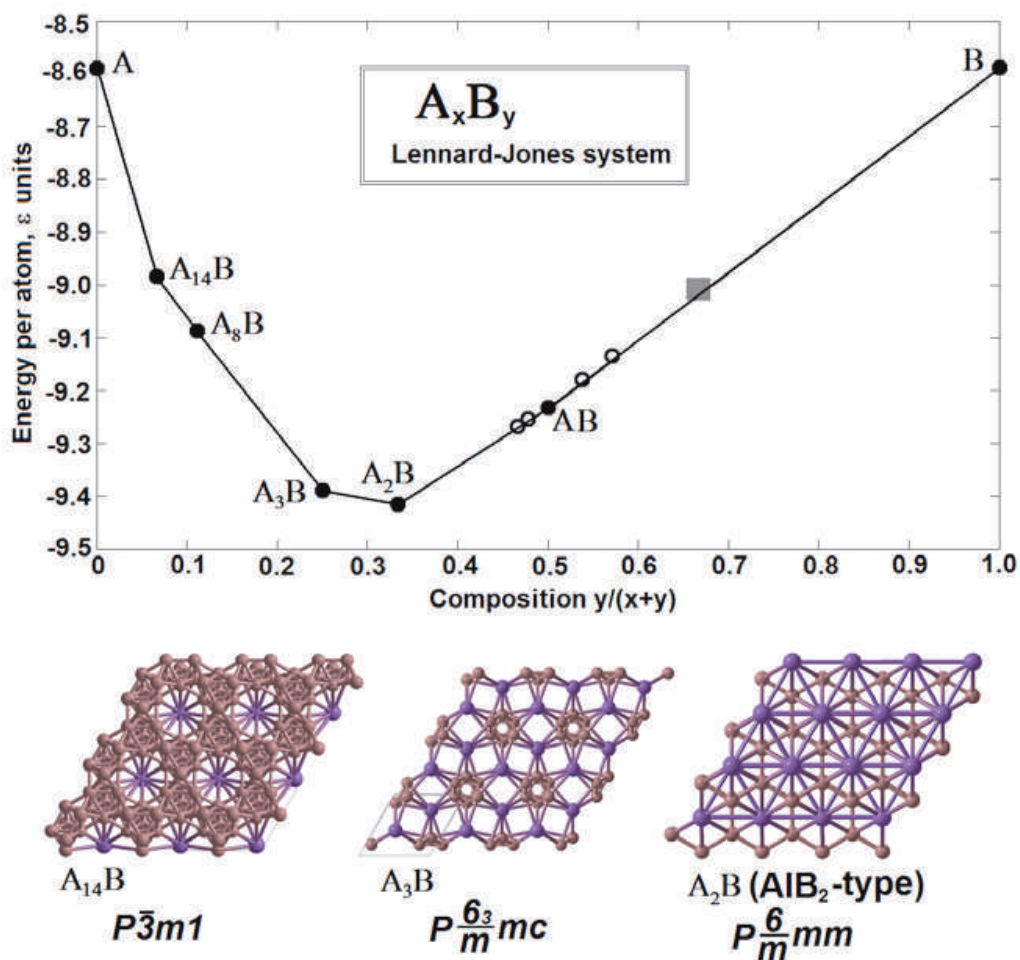
**Figure 24.** Crystal structure and electron localization function (isosurface contour 0.90) of the hP4 phase of sodium at 400 GPa. Interstitial electron localization is clearly seen.

One of the current limitations is the accuracy of today's *ab initio* simulations; this is particularly critical for strongly correlated and for van der Waals systems. Note, however, that the method itself does not make any assumptions about the way energies are calculated and can be used in conjunction with any method that is able to provide total energies. Most of practical calculations are done at  $T = 0$  K, but temperature can be included as long as the free energy can be calculated efficiently. Difficult cases are aperiodic and disordered systems (for which only the lowest-energy periodic approximants and ordered structures can be predicted at this moment).

We are suggesting USPEX as the method of choice for crystal structure prediction of systems with up to  $\sim 50$  atoms/cell, where no information (or just the lattice parameters) is available. Above 50-100 atoms/cell runs become expensive (although still feasible), eventually necessitating the use of other ideas within USPEX or another approach, due to the "curse of dimensionality." There is hope of enabling structure prediction for very large ( $> 200$  atoms/cell) systems. The extension of the method to molecular systems (i.e., handling whole molecules, rather than individual atoms) is already available. The first successful step has been made in adapting USPEX to clusters (Schönborn et al. 2009). Similar extensions, relatively straightforward, still need to be done for surfaces and interfaces. One major unsolved problem is the simultaneous prediction of all stable stoichiometries and structures (in a given range of compositions). A pioneering study by Jóhannesson et al. (2002) succeeded in predicting stable stoichiometries of alloys within a given structure type, while an approach for simultaneous prediction of structure and stoichiometry was proposed by Wang and Oganov (2008) and implemented by Trimarchi et al. (2009) and Lyakhov et al. (2010). Here, the basic ideas are: (1) to start with a population (sparsely) sampling the whole range of compositions of interest, (2) allow variation operators to change chemical composition (we lift chemistry-preserving constraints in the heredity operator and, in addition to the permutation operator, introduce a "chemical transmutation" operator), (3) evaluate the quality of each structure not by its (free) energy, but the (free) energy per atom minus the (free) energy of the most stable isochemical mixture of already sampled compounds. This means that this fitness function depends on history of the simulation. Such an approach works surprisingly well. While Trimarchi et al. (2009) introduced a constraint that in each simulation the total number of atoms in the unit cell is fixed, our method has no such constraint, and this proves beneficial and very convenient. An example of a (very difficult) system is given in Figure 25. Odd as it may seem, a binary Lennard-Jones system with 1:2 ratio of radii (see caption to Fig. 25 for details of the model) exhibits a large number of ground states — including the exotic  $A_{14}B$  compound and the well-known  $A_1B_2$ -type structure, and several marginally unstable compositions (such as  $A_8B_7$ ,  $A_{12}B_{11}$ ,  $A_6B_7$ ,  $A_3B_4$ ,  $AB_2$ ). The efficiency and reliability of the variable-composition algorithm is illustrated by the fact that a fixed-composition simulation at  $AB_2$  stoichiometry produced results (grey square in the upper graph of Fig. 25) perfectly consistent with the variable-composition runs.

USPEX has been applied to many important problems. Apart from the applications described above, several noteworthy results have been published by us recently. These include the high-pressure post-magnesite phases of  $MgCO_3$  (Oganov et al. 2008), polymeric phases of nitrogen (Ma et al. 2009b), superconducting phases of silane ( $SiH_4$ ) (Martinez-Canales et al. 2009) and germane ( $GeH_4$ ) (Gao et al. 2008), the latter predicted to have a remarkably high  $T_C = 64$  K (Gao et al. 2008). Its ability to predict not only the ground states, but also low-energy metastable structures has led to the finding of an interesting metastable structure of carbon (Oganov and Glass 2006), which has recently been shown by Li et al. (2009) to match the observed properties of the so-called "superhard graphite," a material scratching on diamond and formed by metastable room-temperature compression of graphite beyond 15 GPa (Mao et al. 2003). One expects many more applications to follow, both in high-pressure research and in materials design.





**Figure 25.** Variable-composition USPEX simulation of the  $A_x B_y$  binary Lennard-Jones system. In the upper panel: filled circles – stable compositions, open circles – marginally unstable compositions ( $A_8 B_7$ ,  $A_{12} B_{11}$ ,  $A_6 B_7$ ,  $A_3 B_4$ ). Gray square – fixed-composition result for  $AB_2$  stoichiometry, finding a marginally unstable composition in agreement with the variable-composition results. The lower panel shows some of the stable structures. While the ground state of the one-component Lennard-Jones crystal is hexagonal close packed (hcp) structure, ground states of the binary Lennard-Jones system are rather complex (e.g.,  $A_{14} B$ ). Note that the structure found for  $A_2 B$  belongs to the well-known  $AlB_2$  structure type. The potential is of the Lennard-Jones form for each atomic  $ij$ -pair:  $U_{ij} = \epsilon_{ij} [(R_{min,ij} / R)^{12} - 2(R_{min,ij} / R)^6]$ , where  $R_{min,ij}$  is the distance at which the potential reaches minimum, and  $\epsilon$  is the depth of the minimum). In these simulations we use additive atomic dimensions:  $R_{min,BB} = 1.5R_{min,AB} = 2R_{min,AA}$  and non-additive energies (to favor compound formation):  $\epsilon_{AB} = 1.25\epsilon_{AA} = 1.25\epsilon_{BB}$ .

## ACKNOWLEDGMENTS

ARO thanks R. Hoffmann, W. Grochala, R.J. Hemley and R.M. Hazen for exciting discussions. ARO also gratefully acknowledges financial support from the Research Foundation of Stony Brook University and from Intel Corporation. YM's work is supported by the China 973 Program under Grant No. 2005CB724400, the National Natural Science Foundation of China under grant No. 10874054, the NSAF of China under Grant No. 10676011, and the 2007 Cheung Kong Scholars Programme of China. We thank the Joint Supercomputer Center (Russian Academy of Sciences, Moscow) and Swiss National Supercomputing Centre (CSCS) (Manno) for providing supercomputer time. USPEX code is available on request from ARO.

## REFERENCES

- Abraham NL, Probert MIJ (2006) A periodic genetic algorithm with real-space representation for crystal structures and polymorph prediction. *Phys Rev B* 73, art. 224104
- Akahama Y, Kawamura H, Hausermann D, Hanfland M, Shimomura O (1995) New high-pressure structural transition of oxygen at 96 GPa associated with metallization in a molecular solid. *Phys Rev Lett* 74:4690-4693
- Amberger E, Rauh PA (1974) Struktur des borreichen Borphosphids. *Acta Crystallogr B* 30:2549-2553
- Bader (1990) *Atoms in Molecules. A Quantum Theory*. Oxford University Press: Oxford
- Barabash SV, Blum V, Muller S, Zunger A (2006) Prediction of unusual stable ordered structures of Au-Pd alloys via a first-principles cluster expansion. *Phys Rev B* 74, art. 035108
- Bazterra VE, Ferraro MB, Facelli JC (2002) Modified genetic algorithm to model crystal structures. I Benzene, naphthalene and anthracene. *J Chem Phys* 116:5984-5991
- Blöchl PE (1994) Projector augmented-wave method. *Phys Rev B* 50:17953-17979
- Boisen MB, Gibbs GV, Bukowinski MST (1994) Framework silica structures generated using simulated annealing with a potential energy function based on an  $H_6Si_2O_7$  molecule. *Phys Chem Mineral* 21:269-284
- Bonev SA, Gygi F, Ogitsu T, Galli G (2003) High-pressure molecular phases of solid carbon dioxide. *Phys Rev Lett* 91:065501
- Bush TS, Catlow CRA, Battle PD (1995) Evolutionary programming techniques for predicting inorganic crystal structures. *J Mater Chem* 5:1269-1272
- Christy AG (1995) Isosymmetric structural phase transitions: phenomenology and examples. *Acta Crystallogr B* 51:753-757
- Curtarolo S, Morgan D, Ceder G (2005) Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys. *CALPHAD: Comput Coupling of Phase Diagrams and Thermochem.* 29:163-211
- Curtarolo S, Morgan D, Persson K, Rodgers J, Ceder G (2003) Predicting crystal structures with data mining of quantum calculations. *Phys Rev Lett* 91, art. 135503
- Deaven DM, Ho KM (1995) Molecular geometry optimization with a genetic algorithm. *Phys Rev Lett* 75:288-291
- Deem MW, Newsam, JM (1989) Determination of 4-connected framework crystal structures by simulated annealing. *Nature* 342:260-262
- Dong JJ, Tomfohr JK, Sankey OF, Leinenweber K, Somayazulu M, McMillan PF (2000) Investigation of hardness in tetrahedrally bonded nonmolecular  $CO_2$  solids by density-functional theory. *Phys Rev B* 62:14685-14689
- Dubrovinsky L, Dubrovinskaia N, Crichton WA, Mikhailushkin AS, Simak SI, Abrikosov IA, de Almeida JS, Ahuja R, Luo W, Johansson B (2007) Noblest of all metals is structurally unstable at high pressure. *Phys Rev Lett* 98:045503
- Fujihisa H Akahama Y, Kawamura H, Ohishi Y, Shimomura O, Yamawaki H, Sakashita M, Gotoh Y, Takeya S, Honda K (2006)  $O_8$  cluster structure of the epsilon phase of solid oxygen. *Phys Rev Lett* 97, art. 085503
- Gale JD (2005) GULP: Capabilities and prospects. *Z Kristallogr* 220:552-554
- Gao G, Oganov AR, Bergara A, Martinez-Canalez M, Cui T, Iitaka T, Ma Y, Zou G (2008) Superconducting high pressure phase of germane. *Phys Rev Lett* 101:107002
- Glass CW, Oganov AR, Hansen N (2006) USPEX – evolutionary crystal structure prediction. *Comput Phys Commun* 175:713-720
- Gödecker S (2004) Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J Chem Phys* 120:9911–9917
- Goettel KA, Eggert JH, Silvera IF, Moss WC (1989) Optical evidence for the metallization of xenon at 132(5) GPa. *Phys Rev Lett* 62:665-668
- Goncharenko IN (2005) Evidence for a magnetic collapse in the epsilon phase of solid oxygen. *Phys Rev Lett* 94, art. 205701
- Gottwald D, Kahl G, Likos CN (2005) Predicting equilibrium structures in freezing processes. *J Chem Phys* 122, art. 204503
- Gregoryanz E, Degtyareva O, Somayazulu M, Hemley RJ, Mao HK (2005) Melting of dense sodium. *Phys Rev Lett* 94:185502.
- Gregoryanz E, Lundegaard LF, McMahon MI, Guillaume C, Nelmes RJ, Mezouar M (2008) Structural diversity of sodium. *Science* 320:1054–1057
- Gregoryanz E, Sanloup C, Somayazulu M, Badro J, Fiquet G, Mao HK, Hemley RJ (2004) Synthesis and characterization of a binary noble metal nitride. *Nat Mater* 3:294-297
- Grochala W (2007) Atypical compounds of gases, which have been called ‘noble’. *Chem Soc Rev* 36:1632-1655

- Hanfland M, Syassen K, Loa I, Christensen NE, Novikov DL (2002) Na at megabar pressures. Poster at 2002 High Pressure Gordon Conference, June 23-28, 2002. Academy Meriden, NH
- Holm B, Ahuja R, Belonoshko A, Johansson B (2000) Theoretical investigation of high pressure phases of carbon dioxide. *Phys Rev Lett* 85:1258-1261
- Jóhannesson GH, Bligaard T, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK (2002) Combined electronic structure and evolutionary search approach to materials design. *Phys Rev Lett* 88, art. 255506
- Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane wave basis set. *Phys Rev B* 54:11169-11186
- Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* 59:1758-1775
- Kwei GH, Morosin B (1996) Structures of the boron-rich boron carbides from neutron powder diffraction: implications for the nature of the inter-icosahedral chains. *J Phys Chem* 100:8031-8039
- Lasaga AC, Gibbs GV (1987) Applications of quantum-mechanical potential surfaces to mineral physics calculations. *Phys Chem Minerals* 14:107-117
- Lazicki A, Goncharov AF, Struzhkin VV, Cohen RE, Liu Z, Gregoryanz E, Guillaume C, Mao HK, Hemley RJ (2009) Anomalous optical and electronic properties of dense sodium. *Proc Natl Acad Sci* 106:6525-6528
- Li Q, Oganov AR, Wang H, Wang H, Xu Y, Cui T, Ma Y, Mao H-K, Zou G (2009) Superhard monoclinic polymorph of carbon. *Phys Rev Lett* 102:175506
- Lundegaard LF, Weck G, McMahon MI, Desgreniers S, Loubeyre P (2006) Observation of an O<sub>8</sub> molecular lattice in the epsilon phase of solid oxygen. *Nature* 443:201-204
- Lyakhov AO, Oganov AR, Wang Y, Ma Y (2010) Crystal structure prediction using evolutionary approach. *In: Crystal Structure Prediction*. Oganov AR (ed) Wiley-VCH, submitted
- Ma Y, Eremets MI, Oganov AR, Xie Y, Trojan I, Medvedev S, Lyakhov AO, Valle M, Prakapenka V (2009a) Transparent dense sodium. *Nature* 458:182-185
- Ma Y, Oganov AR, Xie Y, Li Z, Kotakoski J (2009b) Novel high pressure structures of polymeric nitrogen. *Phys Rev Lett* 102:065501
- Ma Y-M, Oganov AR, Glass CW (2007) Structure of the metallic  $\zeta$ -phase of oxygen and isosymmetric nature of the  $\epsilon$ - $\zeta$  phase transition: Ab initio simulations. *Phys Rev B* 76, art. 064101
- Mao WL, Mao HK, Eng PJ, Trainor TP, Newville M, Kao CC, Heinz DL, Shu J, Meng Y, Hemley RJ (2003) Bonding changes in compressed superhard graphite. *Science* 302:425-427
- Martinez-Canales M, Oganov AR, Lyakhov A, Ma Y, Bergara A (2009) Novel structures of silane under pressure. *Phys Rev Lett* 102:087005
- Martoňák R, Donadio D, Oganov AR, Parrinello M (2006) Crystal structure transformations in SiO<sub>2</sub> from classical and ab initio metadynamics. *Nat Mater* 5:623-626
- Martoňák R, Laio A, Bernasconi M, Ceriani C, Raiteri P, Zipoli F, Parrinello M (2005) Simulation of structural phase transitions by metadynamics. *Z Kristallogr* 220:489-498
- Martoňák R, Laio A, Parrinello M (2003) Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys Rev Lett* 90:075503
- Martoňák R, Oganov AR, Glass CW (2007) Crystal structure prediction and simulations of structural transformations: metadynamics and evolutionary algorithms. *Phase Transitions* 80:277-298
- Murakami M, Hirose K, Kawamura K, Sata N, Ohishi Y (2004) Post-perovskite phase transition in MgSiO<sub>3</sub>. *Science* 307:855-858
- Neaton JB, Ashcroft NW (1999) Pairing in dense lithium. *Nature* 400:141-144
- Neaton JB, Ashcroft NW (2002) Low-energy linear structures in dense oxygen: Implications for the epsilon phase. *Phys Rev Lett* 88:205503
- Nicol M, Hirsch KR, Holzapfel WB (1979) Oxygen phase equilibria near 298 K. *Chem Phys Lett* 68:49-52
- Oganov AR, Chen J, Gatti C, Ma Y-Z, Ma Y-M, Glass CW, Liu Z, Yu T, Kurakevych OO, Solozhenko VL (2009) Ionic high-pressure form of elemental boron. *Nature* 457:863-867
- Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J Chem Phys* 124, art. 244704
- Oganov AR, Glass CW (2008) Evolutionary crystal structure prediction as a tool in materials design. *J Phys Condens Matter* 20, art. 064210.
- Oganov AR, Glass CW, Ono S (2006) High-pressure phases of CaCO<sub>3</sub>: crystal structure prediction and experiment. *Earth Planet Sci Lett* 241:95-103
- Oganov AR, Ma Y, Glass CW, Valle M (2007) Evolutionary crystal structure prediction: overview of the USPEX method and some of its applications. *Psi-k Newsletter* 84:142-171
- Oganov AR, Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of MgSiO<sub>3</sub> in Earth's D'' layer. *Nature* 430:445-448
- Oganov AR, Ono S, Ma Y, Glass CW, Garcia A (2008) Novel high-pressure structures of MgCO<sub>3</sub>, CaCO<sub>3</sub> and CO<sub>2</sub> and their role in the Earth's lower mantle. *Earth Planet Sci Lett* 273:38-47

- Oganov AR, Valle M (2009) How to quantify energy landscapes of solids. *J Chem Phys* 130:104504
- Ogitsu T, Gygi F, Reed J, Motome Y, Schwegler E, Galli G (2009) Imperfect Crystal and Unusual Semiconductor: Boron, a Frustrated Element. *J Am Chem Soc* 131:1903-1909
- Ono S, Kikegawa T, Ohishi Y (2005a) A high-pressure and high-temperature synthesis of platinum carbide. *Solid State Comm* 133:55-59
- Ono S, Kikegawa T, Ohishi Y (2007) High-pressure phase transition of CaCO<sub>3</sub>. *Am Mineral* 92:1246-1249
- Ono S, Kikegawa T, Ohishi Y, Tsuchiya J (2005b) Post-aragonite phase transformation in CaCO<sub>3</sub> at 40 GPa. *Am Mineral* 90:667-671
- Pannetier J, Bassasalsina J, Rodriguez-Carvajal J, and Caignaert V (1990) Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* 346:343-345
- Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865-3868
- Pickard CJ, Needs RJ (2006) High-pressure phases of silane. *Phys Rev Lett* 97, art. 045504.
- Sanloup C, Schmidt BC, Perez EMC, Jambon A, Gregoryanz E, Mezouar M (2005) Retention of xenon in quartz and Earth's missing xenon. *Science* 310:1174-1177
- Schön JC, Jansen M (1996) First step towards planning of syntheses in solid-state chemistry: Determination of promising structure candidates by global optimization. *Angew Chem Int Ed* 35:1287-1304
- Schönborn S, Goedecker S, Roy S, Oganov AR (2009) The performance of minima hopping and evolutionary algorithms for cluster structure prediction. *J Chem Phys* 130:144108
- Serra S, Chiarotti G, Scandolo S, Tosatti E (1998) Pressure-induced magnetic collapse and metallization of molecular oxygen: The  $\zeta$ -O<sub>2</sub> phase. *Phys Rev Lett* 80:5160-5163
- Shcheka SS, Wiedenbeck M, Frost DJ, Keppeler H (2006) Carbon solubility in mantle minerals. *Earth Planet Sci Lett* 245:730-742
- Shimizu K, Suhara K, Ikumo M, Eremets MI, Amaya K (1998) Superconductivity in oxygen. *Nature* 393:767-769
- Sluiter MHF, Colinet C, Pasturel A (2006) Ab initio calculation of phase stability in Au-Pd and Ag-Pt alloys. *Phys Rev B* 73:174204
- Soler JM, Artacho E, Gale JD, Garcia A, Junquera J, Ordejon P, Sanchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. *J Phys Condens Matter* 14:2745-2779
- Solozhenko VL, Kurakevych OO, Oganov AR (2008) On the hardness of a new boron phase, orthorhombic  $\gamma$ -B<sub>28</sub>. *J Superhard Mater* 30:428-429
- Stuedel R, Wong MW (2007) Dark-red O<sub>8</sub> molecules in solid oxygen: rhomboid clusters, not S<sub>8</sub>-like rings. *Angew Chem Int Ed* 46:1768-1771
- Trimarchi G, Freeman AJ, Zunger A (2009). Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys Rev B* 80:092101
- Tsagareishvili OA, Chkhartishvili LS, Gabunia DL (2009) Apparent low-frequency charge capacitance of semiconducting boron. *Semiconductors* 43:14-20
- Urusov VS, Dubrovinskaya NA, Dubrovinsky LS (1990) Generation of likely crystal structures of minerals. Moscow State University Press, Moscow
- Valle M, Oganov AR (2008) Crystal structure classifier for an evolutionary algorithm structure predictor. *IEEE Symposium on Visual Analytics Science and Technology* (October 21 - 23, Columbus, Ohio, USA), p 11- 18
- van Setten MJ, Uijtewaal MA, de Wijs GA, de Groot RA (2007) Thermodynamic stability of boron: The role of defects and zero point motion. *J Am Chem Soc* 129:2458-2465
- Wang Y, Oganov AR (2008) Research on the evolutionary prediction of very complex crystal structures. *IEEE Computational Intelligence Society Walter Karplus. Summer Research Grant 2008 Final Report*. [http://www.ieee-cis.org/\\_files/EAC\\_Research\\_2008\\_Report\\_WangYanchao.pdf](http://www.ieee-cis.org/_files/EAC_Research_2008_Report_WangYanchao.pdf)
- Weck G, Desgreniers S, Loubeyre P, Mezouar M (2009) Single-crystal structural characterization of the metallic phase of oxygen. *Phys Rev Lett* 102:255503
- Wentorf RH (1965) Boron: another form. *Science* 147:49-50
- Widom M, Mikhalkovic M (2008) Symmetry-broken crystal structure of elemental boron at low temperature. *Phys Rev B* 77:064113
- Woodley SM (2004) Prediction of crystal structures using evolutionary algorithms and related techniques. *Struct Bond* 110:95-132
- Woodley SM, Battle PD, Gale JD, Catlow CRA (1999) The prediction of inorganic crystal structures using a genetic algorithm and energy minimization. *Phys Chem Chem Phys* 1:2535-2542
- Yoo CS, Cynn H, Gygi F, Galli G, Iota V, Nicol M, Carlson S, Hausermann D, Mailhot C (1999) Crystal structure of carbon dioxide at high pressure: "Superhard" polymeric carbon dioxide. *Phys Rev Lett* 83:5527-5530