8 SCIENTIFIC HIGHLIGHT OF THE MONTH: Evolutionary crystal structure prediction: overview of the USPEX method and some of its applications

Evolutionary crystal structure prediction: overview of the USPEX method and some of its applications

Artem R. Oganov[1,2*], Yanming Ma[1,3], Colin W. Glass[1], Mario Valle[4].

[1] *ETH Zurich, Switzerland*
[2] *Moscow State University Russia*
[3] *Jilin University, P. R. China*
[4] *Swiss National Supercomputing Centre (CSCS) Switzerland*
* Corresponding author. E-mail: a.oganov@mat.ethz.ch.

Prediction of the stable crystal structure on the basis of only the chemical composition is one of the central problems of condensed matter physics, which for a long time remained unsolved. The recently developed evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography) made an important progress in this field, enabling efficient and reliable prediction of structures with up to 30-40 atoms in the unit cell using *ab initio* methods (or up to 100-200 atoms/cell with classical forcefields). Here we review this methodology, illustrating its variation operators and visualising evolutionary runs using specifically formulated similarity matrices. We also show several recent applications – (1) prediction of new high-pressure phases of $CaCO_3$, (2) search for the structure of the polymeric phase of CO_2 ("phase V"), (3) search for new phases of transition metals, (4) high-pressure phases of oxygen, (4) new high-pressure phases of FeS and FeO, (5) exploration of possible stable compounds in the Xe-C system at high pressures, (6) investigation of previously proposed cluster-based insulators $Al_{13}K$ and $Al_{12}C$.

1 Introduction

The problem of finding the crystal structure stable at given pressure and temperature conditions, knowing just from chemical composition, occupies a central place in materials design. The ability to solve this problem would open new ways also for the understanding of the behaviour of materials at extreme conditions, e.g. those found in deep planetary interiors.

The simplest method is to explore relative energetics (i.e. compare the free energies) of a number of candidate structures (which can be known structures of analogous materials, or new structures guessed using chemical intuition), and this seems to be the most popular approach. Problems arise almost every time when a totally unexpected and hitherto unknown structure is stable. Instead of relying on real experiments and human intuition, the more advanced

approach of data mining [2] applies methods from machine learning to derive rules of stability of crystal structures from a large set of *ab initio* calculations. A number of simpler intuitive empirical schemes (e.g. structure diagrams, polyhedral clusters – see [3]) have appeared in literature, but unlike the data mining approach, their application usually requires a large experimental data set or good understanding of the compound at hand.

Reliable structure prediction can be performed without any prior knowledge or assumptions about the system. Simulated annealing [4]-[7], minima hopping [8] and metadynamics [9]-[11] have been used with some success. For very small systems, even relaxing randomly produced structures can deliver the stable structure [12]. In this article, we review the evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography) [13]-[15] and illustrate several of its applications. Section II presents basics of the method, section III 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 section IV. This review contains both new results and those published previously ([13][14][30][32][69][73]).

2 Evolutionary Algorithm USPEX

Several groups attempted the pioneering use of evolutionary algorithms to structure prediction - for crystals [19]-[22], colloids [23] and clusters [24]. The algorithm developed by Deaven and Ho [24] is perhaps especially interesting as some of its features (real-space representation of structures, local optimisation and spatial heredity) are similar to the USPEX method. Their algorithm has successfully reproduced the structure of the C_{60} buckminsterfullerene, but has never been extended to heteroatomic clusters, nor to periodic systems (i.e. crystals). The algorithm of Bush and Woodley [19]-[21] was originally developed for crystals and successfully produced a starting model for solving the structure of Li_3RuO_4 [19]. However, subsequent systematic tests [20][21] 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 a population, 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 [20][21]), which swaps

chemical identities in randomly chosen pairs of 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 neighbourhood 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 Figs. 1,2.

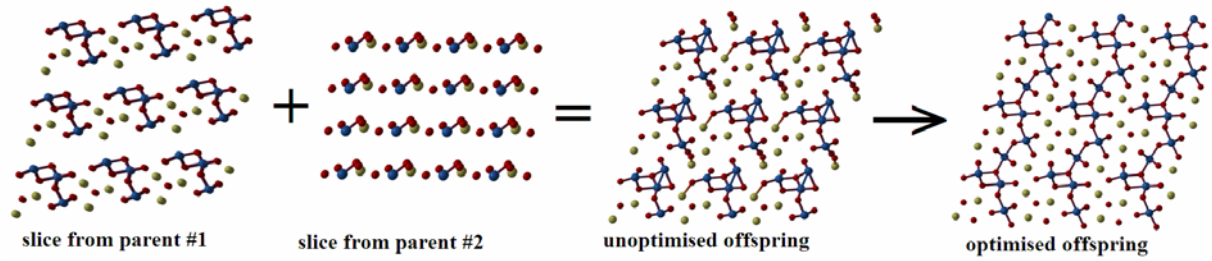


Fig. 1. Heredity operator: slices of two parent structures, and the offspring structure before and after local optimisation.

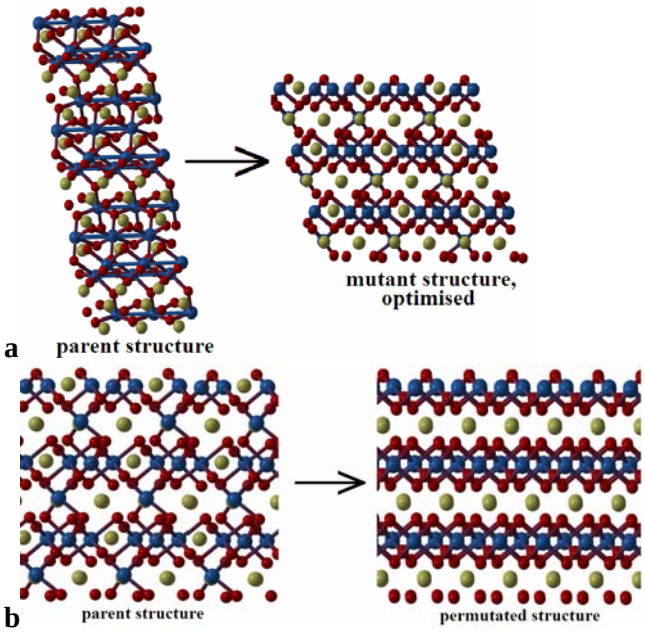


Fig. 2. Illustrations of lattice mutation and permutation operators.

Before new candidate structures are relaxed, they are tested against three constraints – first, all interatomic distances must be above the specified minimal values; second, cell angles must be between 60º and 120º; third, 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 [25], SIESTA [26], GULP [27]).

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. 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.

Recently [69] we improved the algorithm by more exhaustive removal of lattice redundancies. The constraint that the cell angles be between 60º and 120º (in simulations we use a slightly wider interval) removes only part of the redundancies and still allows inconvenient redundant cell shapes to be produced (e.g. cells with $\alpha=\beta=\gamma\sim120$º are practically flat). Certain advantages would be gained by transforming such cells to a cell shape with shorter cell vectors. Such a transformation can be done if there is at least one lattice vector whose projection onto any other cell vector (or onto the diagonal vector of the opposite face of the unit cell) is greater (by absolute value) than half the length of that vector, i.e. for vectors **a** and **b**, these criteria are:

$$\left|\frac{\mathbf{a}\bullet\mathbf{b}}{|\mathbf{b}|}\right| > \frac{|\mathbf{b}|}{2} \qquad (1a)$$

$$\left|\frac{\mathbf{a}\bullet\mathbf{b}}{|\mathbf{a}|}\right| > \frac{|\mathbf{a}|}{2} \qquad (1b)$$

E.g. for the criterion (1a) the new vector $\mathbf{a}^*$ equals:

$$\mathbf{a}^* = \mathbf{a} - ceil\left(\left|\frac{\mathbf{a}\bullet\mathbf{b}}{|\mathbf{b}|}\right|\right)(sign(\mathbf{a}\bullet\mathbf{b}))\mathbf{b} \quad , \qquad (2)$$

During this transformation atomic fractional coordinates are transformed so that the original and the transformed structures are identical (during the transformation Cartesian coordinates of the atoms remain invariant). For more details on the USPEX method, see [14][15].

An evolutionary algorithm, somewhat similar to USPEX, was proposed slightly later and independently from us in [71]; this method differs from USPEX in the absence of permutation (i.e. potential problems for binary and more complex compounds), different forms of heredity

and mutation, absence of cell rescaling. Much later, our USPEX method was reimplemented in paper [72], which contained a number of errors (see comment [73]), but for very simple test cases was also successful.

Why is the USPEX methodology successful? One of the reasons is that local optimization with high probability creates chemically reasonable local environments of the atoms. Evolutionary pressure, through selection, forces the population to improve from generation to generation. 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.

Another important reason is that the energy landscapes expected in chemical systems are likely to have an overall "funnel" shape (Fig. 3), 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. In some cases (where dramatically different atomic arrangements are energetically competitive – e.g. complex molecular crystals where different molecules can be assembled from the same atoms) the landscape is expected to contain multiple funnels, and such systems are then 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 overall landscape shape (Fig. 3) 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.

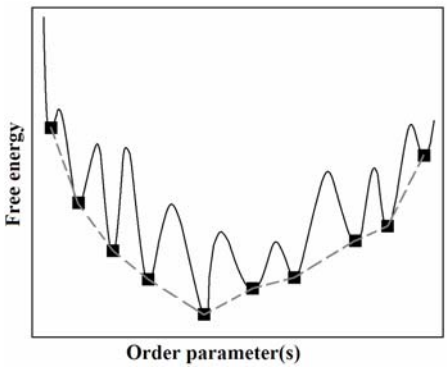


Fig. 3. Expected energy landscape in chemical systems. The original response surface is very "noisy" (i.e. contains very large energy variations, with high barriers). Local optimization reduces this surface to local minima points (black squares). The reduced response surface (dashed line) is well-behaved and has a simple overall shape. This is why the use of local optimization dramatically improves global optimization [15].

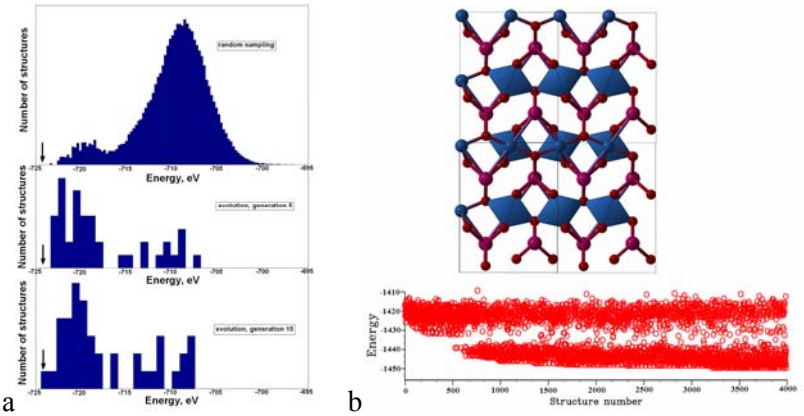
3 Tests of the Algorithm

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 [14][15][28][69][73]. 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 favourable fcc structure.

The largest test for a chemically complex system is the prediction of the structure of $MgSiO_3$ post-perovskite [33][34] using a relatively large 80-atom supercell (with fixed supercell parameters) and an empirical potential [34] describing interatomic interactions within a partially ionic model. Local optimization and energy calculations were done using the GULP code [27]. Previously [28] we have shown that already in a 40-atom supercell this test is unfeasible using the simple random sampling (with local optimization) [12]: the correct structure was not produced even after 1.2×10^5 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 (~3200) of local optimizations – see Fig. 4.



a b

Fig. 4. Evolutionary prediction of the structure of $MgSiO_3$ post-perovskite using the experimental cell parameters for a) 40-atom [28] and b) 80-atom [69] supercells. In both cases, each generation consisted of 41 structures. Panel (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 $1.2x10^5$ steps, whereas in the evolutionary simulation shown it was found within 15 generations (i.e. 600 local optimizations). Arrows mark the ground-state energy. Panel (b) shows the energies of structures along the evolutionary trajectory for the 80-atom run; the structure of post-perovskite was obtained within ~3200 local optimizations. One can see that the density of low-energy structures increases during the simulation.

Fig. 5 shows variable-cell *ab initio* results for $MgSiO_3$ at the pressure of 120 GPa. Several runs with somewhat different (but within a reasonable range) parameters have been 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 Fig. 5. The similarity matrix (Fig. 5b) shows a slow decrease of population diversity. This is what we usually find in simulations, except for very small or very special systems. The construction of the similarity matrix is outlined in the Appendix.

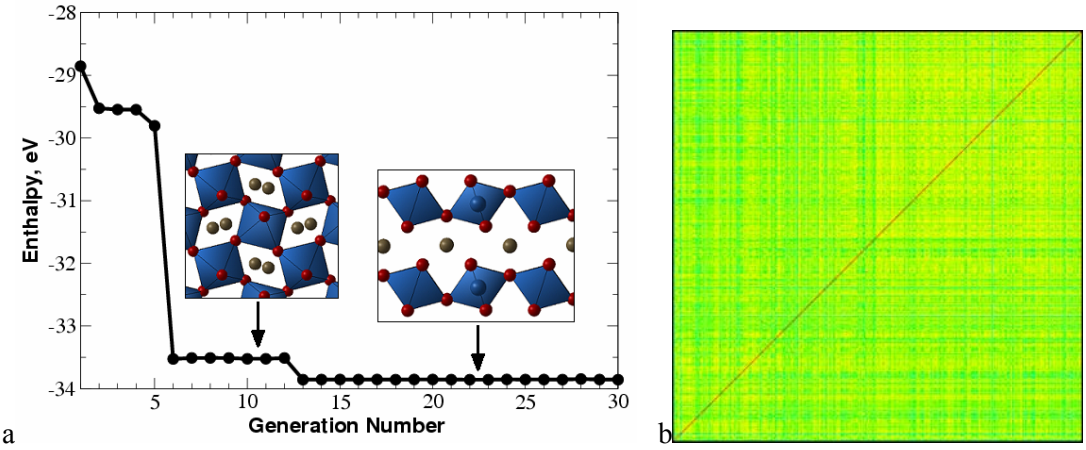


Fig. 5. Evolutionary structure search for $MgSiO_3$ at 120 GPa. a) evolution of the lowest enthalpy as a function of the generation (insets show the structures of perovskite and post-perovskite phases), from [14]. b) structure similarity matrix (dimensions 500x500).

An example of a very simple test, variable-cell *ab initio* structure search for GaAs with 8 atoms/cell [73], is given in Fig. 6. The ground-state structure for systems of such size can be found even by local optimisation of a reasonable number of randomly produced structures. The density of states of relaxed random structures (Fig. 6a), 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 ~0.2%, i.e. finding it with random search would on average take ~ 500 local optimizations. In evolutionary

simulations (Fig. 6b) it can be found within 3 generations, or just 30 structure relaxations. Similarity matrices for random (Fig. 6c) and evolutionary (Fig. 6c) searches clearly reveal a strong increase of structure similarity (i.e. decrease of diversity) 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.

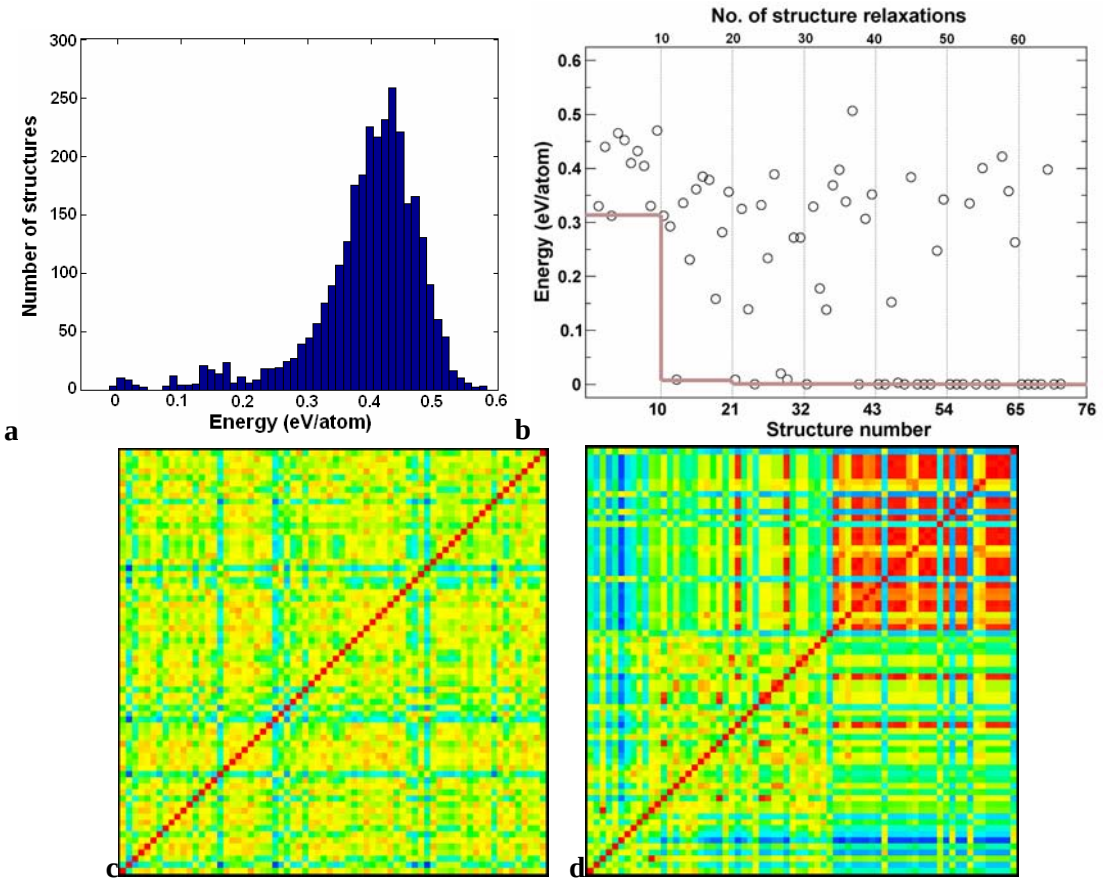
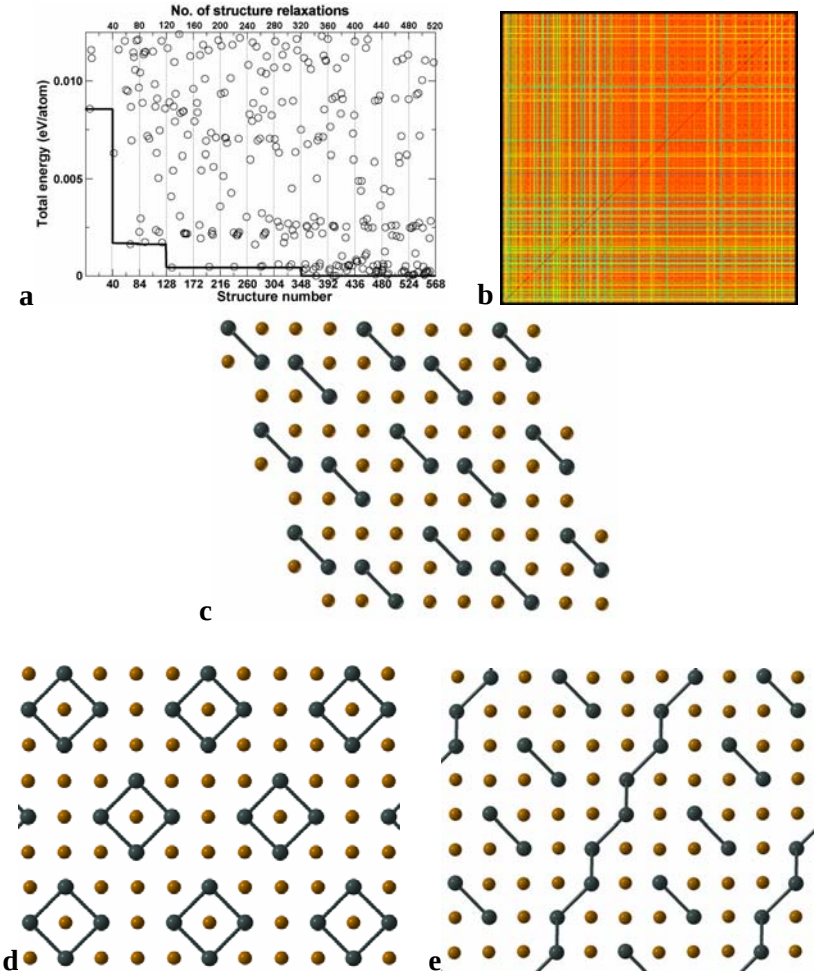


Fig. 6. Structure prediction for GaAs (8 atoms/cell): a) energy distribution for relaxed random structures, b) progress of an evolutionary simulation (thin vertical lines show generations of structures, and the grey line shows the lowest energy as a function of generation), c-d) similarity matrices (dimension 70x70) 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. Panels a,b are from [73]. Calculations are performed within the GGA [38].

Au_8Pd_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 [78][79][74][72][73]. 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.[74] found two energetically nearly degenerate structures (Fig. 7d,e), which seemed to be the only ground-state solutions. Evolutionary search done in Ref. [72] found another fcc-based structure (Fig. 7f), which is 2 meV/atom above the ground state (4

meV/atom from more precise calculations [73]). This failure has led the authors of Ref. [72] to incorrectly conclude that evolutionary structure prediction is "not able to mimic for metal alloys the resolution of cluster expansion". Errors that have led the authors of Ref. [72] to this conclusion have been analysed in [73]. Our evolutionary simulations [73], performed within the same approximations (the same LDA functional and ultrasoft pseudopotentials) found a new ground-state structure (Fig. 7c) that has been overlooked by the previous cluster-expansion study [74] and turned out to be ~0.1 meV/atom lower in energy than the previously known lowest-energy structures (Fig. 7d,e). Examination of all the produced structures shows that most of them are different ordering schemes of the fcc-structure and the energy differences are generally very small (Fig. 7a). The problem of structure prediction in this case reduces to that of finding the optimal ordering scheme within the basic fcc-structure; the ordering scheme is strongly coupled to the unit cell shape (especially for such small systems). Thus, maintaining a large diversity of cell shapes within a run is important and probably difficult to achieve with the small population size chosen in Ref. [72]. Similarity matrix (Fig. 7b) shows that even with these settings, the diversity in the system is rather limited and visibly decreases during the run. Reasonable choice of simulation parameters is essential for the success of any simulations; analyzing the evolutionary trajectories, the energies (e.g. Fig. 7a) and similarity matrices (Fig. 7b) for preliminary calculations can help in choosing the parameters.



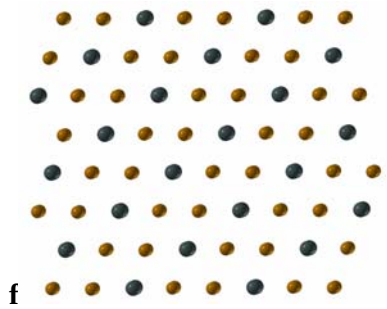


Fig. 7. Evolutionary structure search for Au_8Pd_4. a) evolution of the total energies (only the lowest-energy part is shown for clarity), b) structure similarity matrix (dimensions 873x873), c) the lowest-energy structure found in our evolutionary simulation, d-e) – the lowest-energy structures found by cluster expansion in Ref. [74] (structures No. 4905 and No. 4557, respectively, in the search catalogue of Ref. [74]), f) the lowest-energy structure found in Ref. [71]. All energies are given relative to the ground state.

4 Some Applications of the Method

In this section we will review some new insight that has been obtained using our method (see also [14]). All structure predictions described here were performed within the generalized gradient approximation (GGA [38]) and the PAW method [39][40], using VASP [25] 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 and may sometimes still be qualitatively correct. Most of the cases considered below (except ε-O_8, FeO and to some extent FeS) are well described by the GGA.

$CaCO_3$ polymorphs. High-pressure behaviour 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 carbon [43]. For $CaCO_3$, there is a well-known transition from calcite to aragonite at ~2 GPa, followed by a transition to a post-aragonite phase at ~40 GPa [45], the structure of which was solved [13] using USPEX, and the predicted structure matched the experimental X-ray diffraction pattern well. Furthermore, we have predicted [13] 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 [46] at pressures above 130 GPa. We note that both post-aragonite and the $C222_1$ structure (Fig. 8) 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 [48], which is one of the reasons for the enormous diversity of silicate structure types.

Fig 9 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 [48]. One can see only a shallow minimum at $\angle(\mathrm{Si}-\mathrm{O}-\mathrm{Si})$=135º, but a deep minimum at $\angle(C-\mathrm{O}-C)$=124º with steep energy variations for $H_6C_2O_6$ (Fig. 9). This suggests a much more limited structural variety of metacarbonates, compared to silicates. In both $CaCO_3$ and CO_2 the $\angle(C-\mathrm{O}-C)$ angles are close to 124º in a wide pressure range.

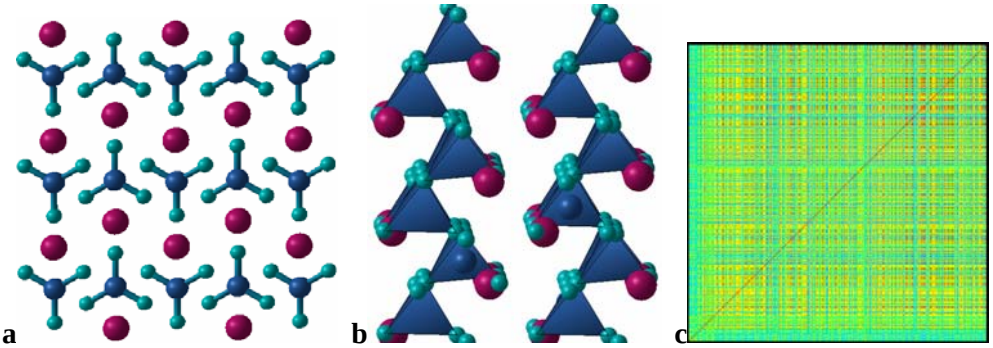


Fig. 8. $CaCO_3$ at high pressure. a) structure of post-aragonite phase, b) $C222_1$ phase, c) similarity matrix (dimensions 500x500) for $CaCO_3$ with 10 atoms/cell at 80 GPa, showing considerable diversity remaining in the population throughout the run.

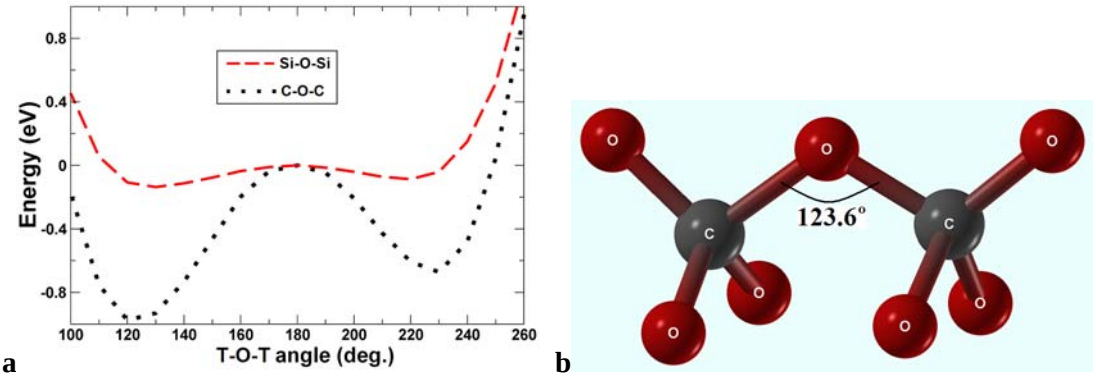
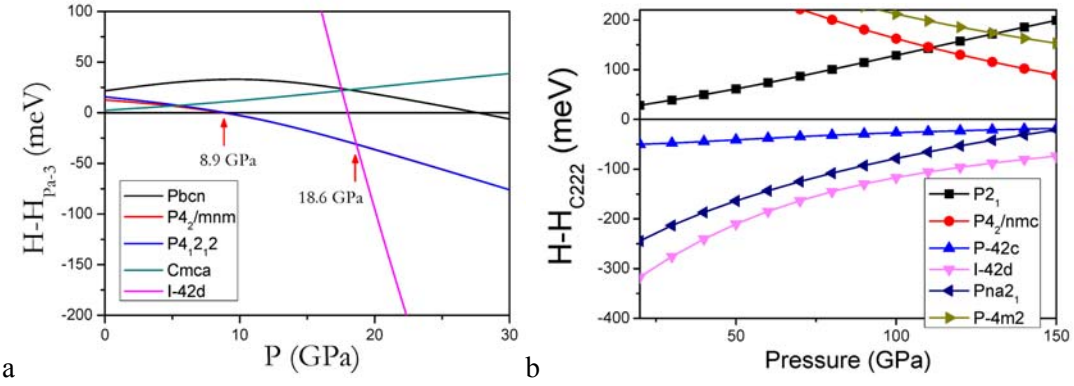


Fig. 9. Energy variation as a function of the T-O-T angle (red dashed line – T = Si, black dotted line – T = C). Calculations were performed on $H_6T_2O_7$ molecules; at each angle all T-O distances and O-T-O valence angles were optimised. Optimum angle C-O-C=124º, Si-O-Si=135º. These calculations were performed with SIESTA code [26] using the GGA functional [38], norm-conserving pseudopotentials and a double-ζ basis set with a single polarisation function for each atom.

Polymeric phase of CO_2 [50]. High-pressure behaviour of CO_2 is still controversial [51]. It is known that above ~20 GPa a non-molecular phase (called phase V [52]) with tetrahedrally coordinated carbon atoms becomes stable, but its structure is still under debate: in the first experimental study [52] a trydimite structure was proposed, but later theoretical works found it to be unstable (even not metastable) and much less favourable than the β-cristobalite structure [53][54]. At the same time, it was not possible to rule out that there may be even more stable

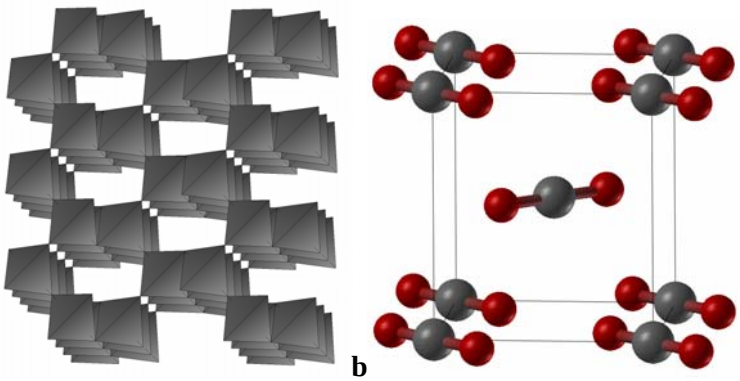
structures. Here, we have performed evolutionary structure searches at 50 GPa, 100 GPa and 150 GPa for system with 6, 9, 12, 18 and 24 atoms/cell. At all these pressures we confirmed stability of the β-cristobalite structure (Fig. 10), thus suggesting an experimental re-investigation of phase V of carbon dioxide. 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). Evolutionary simulations also uncovered several interesting metastable polymeric forms shown in Fig. 11. In agreement with our expectations, none of these structures are known for SiO_2. One can notice that for CO_2, the β-cristobalite structure is much lower in enthalpy compared to other structures – this is again in contrast with silica, where a large number of energetically similar structures are known. To a large extent this difference is explained by the different flexibilities of the intertetrahedral $\angle(\mathrm{T}-\mathrm{O}-\mathrm{T})$ angles (see above).

At lower pressures, between 8.9 GPa and 18.9 GPa, the $P\frac{4_2}{m}nm$ phase (see [68] for details) is stable, and at even lower pressures (0-8.9 GPa) the $Pa3$ structure is stable (Fig. 10). The $Pa3$-$P\frac{4_2}{m}nm$ transition pressure calculated here (8.9 GPa) is consistent with experiment and previous calculation [68].



a b

Fig. 10. Enthalpies of candidate forms of CO_2: a) in the low-pressure region, relative to the molecular *Pa*3 structure, b) in the high-pressure region, relative to the non-molecular *C*222 structure. From [50].



a **b**

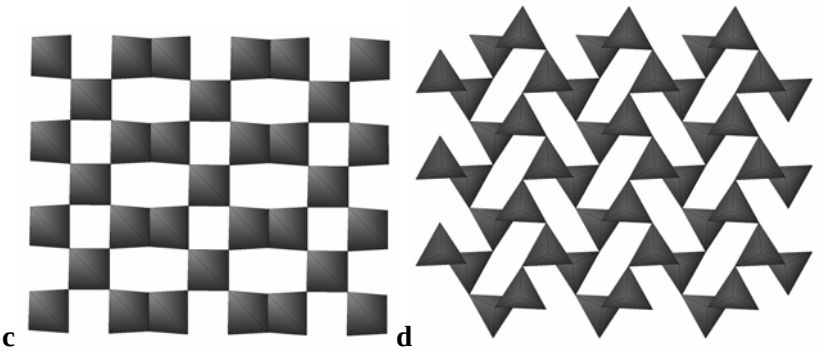


Fig. 11. CO_2 structures: a) β-cristobalite-type form of CO_2 at 100 GPa, suggested to be the structure of phase V and showing carbonate tetrahedra. Structural parameters at 100 GPa: space group $I\bar{4}2d$, a=b=3.2906 Å, c=6.0349 Å, C(0.5; 0; 0.25), O(0.2739; 0.25; 0.125), b) molecular $P\frac{4_2}{m}nm$ structure, stable at lower pressures than CO_2-V, c) polymeric $C222$ structure, d) metastable polymeric $Pna2_1$ structure. From [50].

Methane (CH_4) under pressure: formation of diamond or polymerization? Methane, one of the major components of giant planets Uranus and Neptune, is expected to become unstable at high pressure. However, neither the minimum pressure of its instability nor the product of its chemical transformation are well constrained. Some studies [80][81][82] suggested decomposition into diamond and hydrogen, others [83] proposed polymerization into heavier hydrocarbons. These problems still remain open. The simulations we have done up to now (at 100 GPa, 200 GPa, 300 GPa with 10, 15 and 20 atoms/cell in each case) are only preliminary and do not fully resolve them, but they already shed some light onto the nature of the chemical transformation.

Fig. 12a shows the most stable structure made of CH_4 molecules, which we found at 100 GPa and 200 GPa. At 200 GPa and 300 GPa we found more stable structures made of higher hydrocarbons (propane C_3H_8, butane C_4H_{10} and infinite-chain $(CH_2)_\infty$ polymer) mixed with H_2 molecules (Fig. 12b-d). To understand the chemical transformation of methane under pressure, we consider the following reactions at 300 GPa:

$$CH_4 = C + 2H_2 \quad , \qquad \text{(R1)}$$

i.e. decomposition of methane into diamond and hydrogen (structure of hydrogen was taken from our previous results [14]).

$$4CH_4 = C_4H_{10} + 3H_2 \quad , \qquad \text{(R2)}$$

i.e. decomposition into butane and hydrogen (the structure of butane was taken, as the first approximation, from the structure shown in Fig. 12d, removing the H_2 molecules)

$$2CH_4 = C_2H_4 + 2H_2 \quad , \qquad \text{(R3)}$$

i.e. decomposition into an infinite zigzag polymeric chain $(CH_2)_\infty$ structure (which here was taken, as the first approximation, from the structure shown in Fig. 12c, removing the H_2 molecules)

Additionally, we looked at the enthalpy of decomposition of butane into diamond and hydrogen (we find it to be strongly positive, i.e. butane is stable against this decomposition) and the enthalpy of removal of the H_2 molecules from the structure shown in Fig. 12d (we find it to be negative, i.e. phase separation is favourable; this also shows that the structure of butane we assumed here is energetically at least very reasonable).

At 300 GPa, we find reaction (R1) to be unfavourable (by 0.04 eV); since its products are 4% denser than reactants, it will become favourable at higher pressures. Reaction R2, however, is extremely favourable (by 0.72 eV) already at 300 GPa; its products are also significantly (2%) denser than the reactants and this reaction should be preferred to R1 up to very high pressures.

How far does polymerization go under pressure? Decomposition into diamond and hydrogen (R1) is unstable, and so is the formation of infinite polymers (R3, unfavourable by 0.29 eV). Therefore, we expect the formation of finite alkane polymers, such as butane, to be the most favourable process in a wide pressure range. It is possible that with pressure the length of the polymers would increase (a coexistence of different polymers is likely in the compressed fluid [83], and a gradual increase in fraction of heavier polymers is possible on increasing pressure). To resolve these issues and obtain precise polymerization pressure for methane, we need to perform additional evolutionary simulations for other compositions (C_3H_8, C_4H_{10}, etc.)

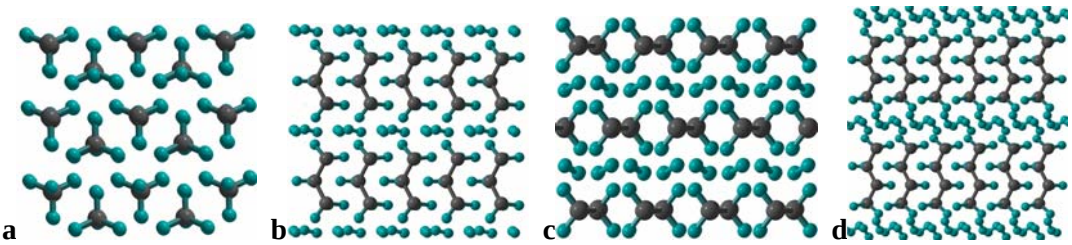


Fig. 12. Structures of CH_4: a) methane structure at 100 GPa, b) C_3H_8*$2H_2$ structure at 200, c) $(CH_2)_\infty$*H_2 structure at 300 GPa, d) C_4H_{10}*$3H_2$ structure at 300 GPa.

Exploration of transition metals at ultrahigh pressures. Many transition metals appear to have the same structure (typically, hcp, fcc or bcc) over a very wide pressure range. This is at first sight surprising, especially when contrasted with the behaviour of simple sp-metals, which often adopt d-character (i.e. become *de facto* transition metals) and a sequence of complex structures on compression. We have performed exploratory simulations of a number of transition metals that at ambient conditions are known in the hcp (Zn, Cd, Ru, Os, Tc, Re), bcc (Mo, W, Nb) and fcc (Cu) structures. For these metals, no phase transitions were experimentally detected up to 126 GPa (Zn), 174 GPa (Cd), 56 GPa (Ru), 75 GPa (Os), 10 GPa (Tc), 251 GPa (Re), 188 GPa (Cu), 272 GPa (Mo), 154 GPa (Nb), 364 GPa (W). Among

these metals, only for Os a transition into the hexagonal ω-phase structure (with 6 atoms/cell) was predicted to occur at 840 GPa [67].

The behaviour of each of these metals was explored at several pressures from 0 to 1000 GPa, using system sizes of 2 atoms/cell (Zn), 4 atoms/cell (Zn), 6 atoms/cell (Zn, Cd, Ru, Os, Tc, Re, Nb) and 8 atoms/cell (all metals). We have found that at all pressures up to 1000 GPa (at 0 K), all these metals remain in the same structure as at ambient conditions. This includes Os, for which our calculations find the hcp structure to be much more stable than ω-phase or any other structure at least up to 1000 GPa (Fig. 13). The difference with [67] is likely to be due to pseudopotential errors, but Fig. 13 shows that both PAW and pseudopotential calculations lead to the same conclusion when the pseudopotential is correctly constructed.

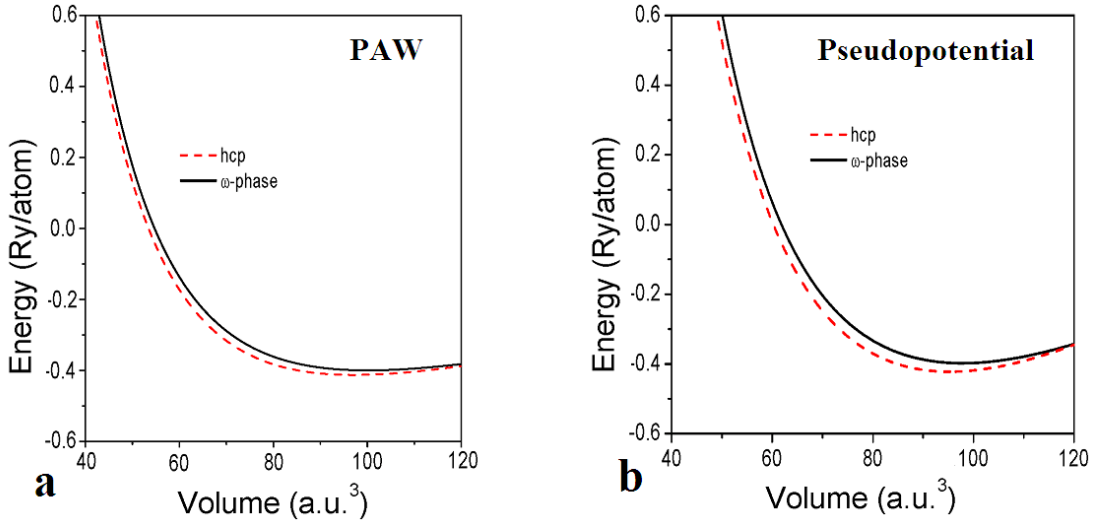


Fig. 13. Energy-volume curves for Os a) PAW results, b) pseudopotential calculations. PAW calculations were performed using the VASP code, pseudopotential results were obtained with the PWSCF code.

Testing the limits of DFT: new hypothetical phases of FeS and FeO. The Fe-S system is important for the understanding of the cores of terrestrial planets (Mercury, Venus, Earth, Mars). Numerous experimental studies of FeS uncovered six solid phases at pressures of up to 100 GPa and temperatures up to 3000 K ([32] and references therein) – all of which are based on the NiAs-type structure with various distortions and superstructures (one of the high-temperature phases has the undistorted hexagonal NiAs structure). Among these phases, two have the same MnP-type structure (*Pnma*-distortion of the NiAs-type structure) – but one is antiferromagnetic (stable between 3.5 GPa and 7 GPa [84]) and the other, recently discovered [32], is non-magnetic and stable above 40 GPa (i.e. at pressures of Martian core). The structure of the non-magnetic MnP-type phase (Fig. 14a) was established in the same study by both Rietveld analysis and evolutionary structure prediction [32].

Tremel et al. [98] have analyzed the nature of various distortions of the NiAs-type structure (assuming non-magnetic ground states), and explained them tracing these distortions in terms of the Jahn-Teller effect and electronic pseudogap formation. These authors found that the

MnP-type structure occurs when the number of valence electrons is 11-14 per formula. FeS, with 14 valence electrons, is therefore at the limit of stability of the MnP structure type and should be prone to phase transitions. The complicated phase diagram of FeS is consistent with this picture; magnetism (not considered in [98]), increases its complexity further. At very high pressures (above 130 GPa, according to GGA calculations) the MnP-type phase was predicted [32] to transform into a new phase. This structure of this seventh phase of FeS (Fig. 14b) belongs to a new structure type, but can be represented as a *Pmmn*-distortion of the NaCl-type structure. FeS is a difficult system for the GGA: while the sequence of phase transitions with pressure is well reproduced by this approximation, the transition pressures are not predicted accurately: most notably, the pressure range where the non-magnetic $P2_1/a$ phase is stable is predicted to be 0-90 GPa (in experiments it is much narrower, 7-40 GPa – see [32] and references therein). Therefore, there is a large uncertainty on the predicted stability range (> 130 GPa) of the *Pmmn* phase of FeS and an experimental determination of this pressure range is desirable, as this phase is potentially important for the cores of the Earth and Venus.

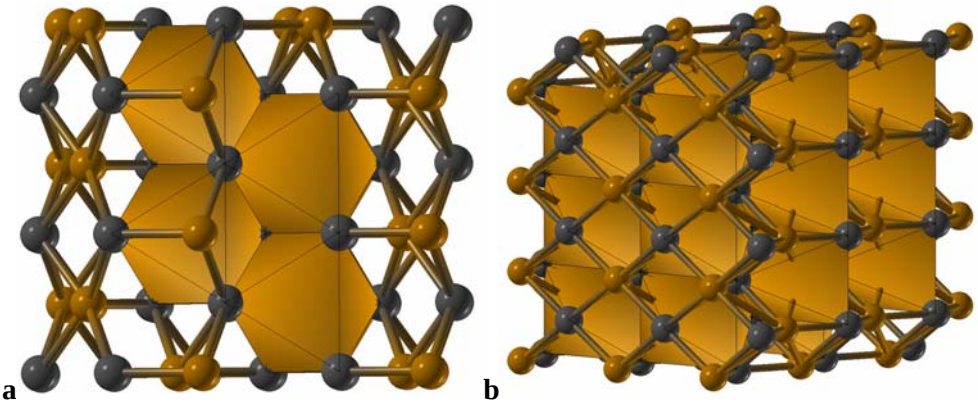


Fig. 14. Structures of high-pressure non-magnetic phases of FeS. a) MnP-type structure (FeS VI), space group *Pnma*; b) distorted NaCl-type structure (FeS VII), space group *Pmmn*. The yellow and black spheres denote iron and sulfur atoms, respectively. FeS6 octahedra are shown.

FeO has long been known to be a system where standard density-functional approximations, the LDA and GGA, fail not only quantitatively but even qualitatively. One of the worst failures is that these approximations predict the wrong ground state at 1 atm – the inverse NiAs-type (i-NiAs) antiferromagnetic phase [85][86] (Fig. 15b). The experimentally observed phase at these conditions has a rhombohedrally distorted antiferromagnetic NaCl-type structure (Fig. 15a), and for this phase LDA and GGA calculations incorrectly predict a metallic behaviour. High-pressure experiments produced strong evidence for the stability of the i-NiAs [88] or NiAs [87][89] structures above ~70-90 GPa.

We performed a series of evolutionary simulations at 100 GPa (with 4,6,8,12 and 16 atoms/cell, all non-magnetic) and found that, within the GGA, the most stable is the unexpected and hitherto unknown trigonal structure (space group $P3_221$, see Fig. 15c). Within the GGA, this structure is stable between 65 GPa and 190 GPa (Fig. 16), and above 190 GPa a

reentrant transition into the (non-magnetic) distorted NaCl-type phase is predicted. While these transition pressures are likely to be numerically inaccurate, there is a possibility that the sequence of phases is correct, because at high pressures, where the role of strong on-site Coulomb correlations diminishes and magnetism is suppressed, GGA calculations might produce at least qualitatively correct results (as for FeS). We will not be surprised, however, if the new structure is an artifact of the GGA. Detailed experiments at ultrahigh pressures are needed to address this and assess the accuracy of DFT calculations at ultrahigh pressures.

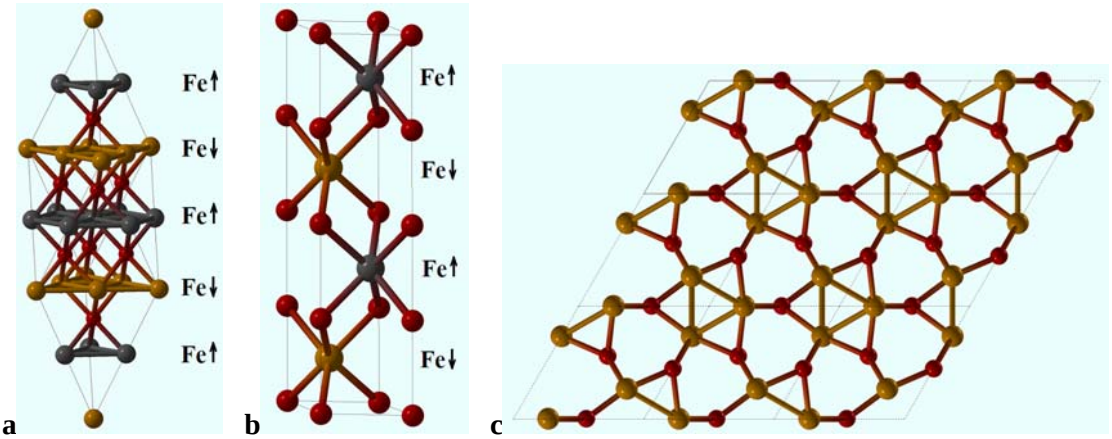


Fig. 15. Phases of FeO. Space group $P3_221$, structure parameters at 100 GPa: a=b=4.5365 Å, c=2.3648 Å, Fe(0; 0.2665; 0.3333), O(0.5564; 0; 0.16667). Distances: Fe-O = 1.77 (x2), 1.80 (x2), Fe-Fe= 2.24 (x2), 2.36 (x2), 2.62 (x2), O-O= 2.36 (x2), 2.44 (x2), 2.64 (x2). In panels a,b Fe atoms with opposite magnetic moments are shown by different colours; layers of atoms with the same magnetic polarization are marked.

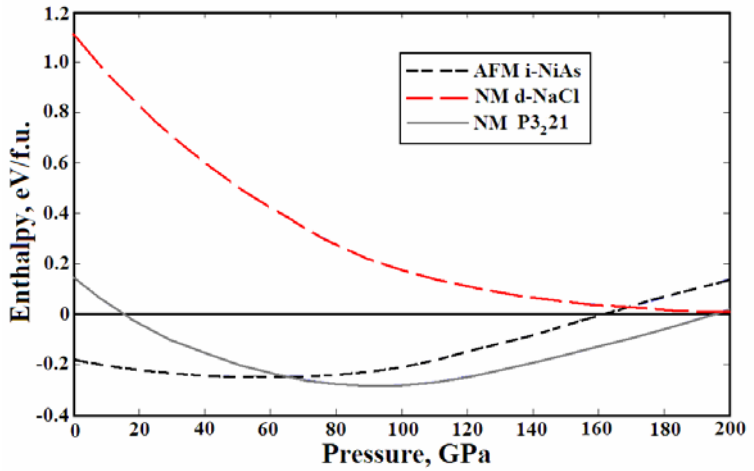


Fig. 16. Stability of FeO phases within the GGA. Enthalpies are relative to the non-magnetic rhombohedrally distorted NaCl-type structure. Spin transition of the NaCl-type is predicted to be gradual and to be completed at ~200 GPa. At pressures between 65 GPa and 190 GPa the new $P3_221$ possesses the lowest theoretical enthalpy. This is possibly a result of shortcomings of the GGA, as well as the prediction of the i-NiAs structure as the ground state at ambient conditions.

Semiconducting and metallic phases of solid oxygen: unusual molecular associations. The red ε-phase of oxygen, stable in the pressure range 8-96 GPa, was discovered in 1979 [55], but its structure was solved only in 2006 [57][58]. The metallic (superconducting at very low temperatures [60]) ζ-phase, stable above 96 GPa, was discovered in 1995 [56], but its structure still remains experimentally unsolved. Neutron diffraction showed [61] that already in the ε-phase (at 8 GPa) there is no long-range magnetic order. 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 [14][62], restricting ourselves only to non-magnetic solutions.

At 25 GPa we found two particularly interesting structures – one consisting of zigzag chains of O_2 molecules (*Cmcm* structure of Ref.[64][14], see Fig. 17b) and one with more complex chains of molecules (see Fig. 17c). These have strong similarities with the experimentally observed structure (Ref. [57][58], see Fig. 17a) 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 ~2.1-2.2 Å (the intermolecular distance is ~1.2 Å). The *Cmcm* structure, first suggested in [64], is the true GGA ground state, but it differs from experiment; as Fig. 18a shows, its enthalpy is ~10 meV/atom lower than for the experimentally found structure (Fig. 17a). Metastability of the experimentally studied structure cannot yet be ruled out, but, more likely, this discrepancy is due to deficiencies of the GGA. The $(O_2)_4$ clusters are held together by weak intermolecular covalent bonds: each O_2 molecule has two unpaired electrons occupying two molecular π*-orbitals, and sharing these electrons with neighbouring molecules creates two intermolecular bonds per molecule and a non-magnetic ground state [62][66]. 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 ζ-phase.

For the ζ-phase, evolutionary simulations at 130 GPa and 250 GPa uncovered two interesting structures with *C*2/*m* and *C*2/*c* space groups [62]. These have very similar enthalpies (Fig. 18a); the *C*2/*m* structure is slightly lower in enthalpy and matches experimental X-ray diffraction and Raman spectroscopy data very well, better than the *C*2/*c* structure [62]. 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 of the *C*2/*m* structure (Fig. 18b-d) 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 Γ-V direction, close to the Brillouin zone centre – such soft modes may result in small-amplitude long-wavelength modulations of the structure at very low temperatures.

The ε-ξ transition is isosymmetric, which implies that it is first-order at low temperatures but can become fully continuous above some critical temperature [65]. Given the small volume discontinuity upon transition and small hysteresis (one can obtain the *C*2/*m* structure of the ξ-phase by simple overcompression of the ε-O_8 structure, ~5 GPa above the thermodynamic transition pressure), one can expect this critical temperature to be rather low. We note that within the GGA the ε-ξ transition is predicted to occur at 45 GPa, 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 ζ-phase than for the semiconducting ε-O_8 phase, the enthalpy differences are expected to suffer from non-canceling errors, (ii) since the ε-ξ transition is not only structural, but also electronic (insulator-metal transition), one might expect metallization at lower pressures than in experiment.

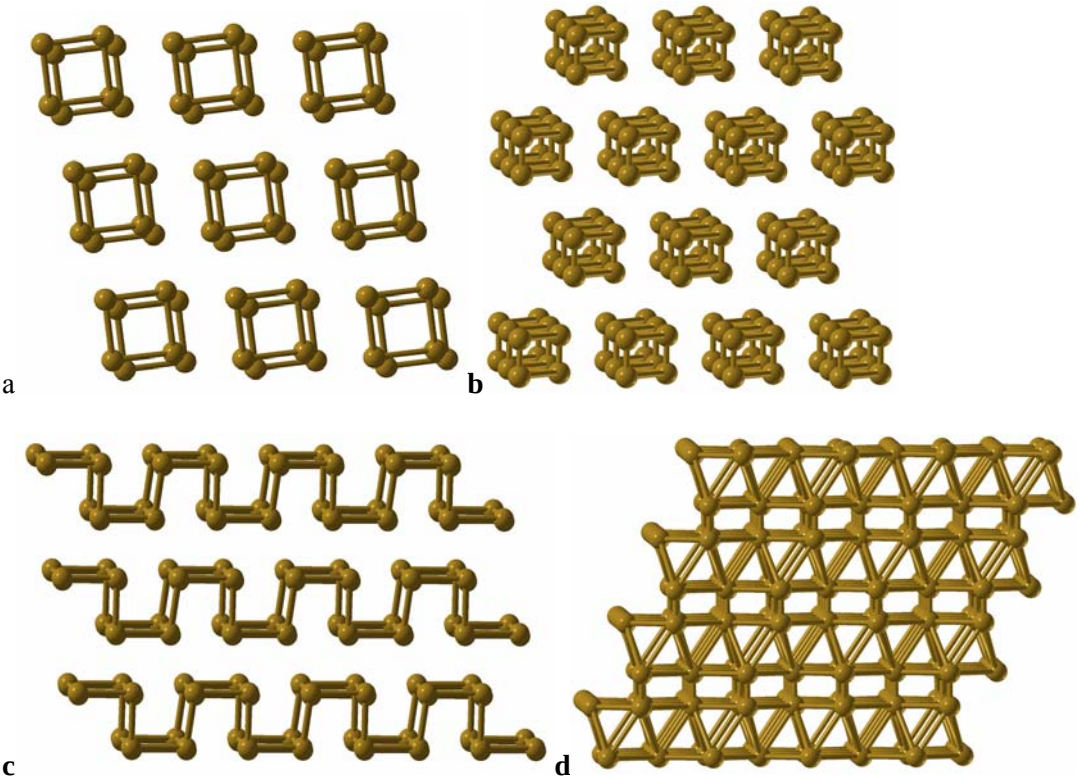


Fig. 17. High-pressure structures of oxygen: a) experimentally found ε-O_8 structure at 17.5 GPa [57], b) *Cmcm* chain structure [64][14], c) metastable chain structure at 25 GPa [14], d) *C*2/*m* structure of the ζ-phase at 130 GPa [62]. Contacts up to to 2.2 Å are shown as bonds.

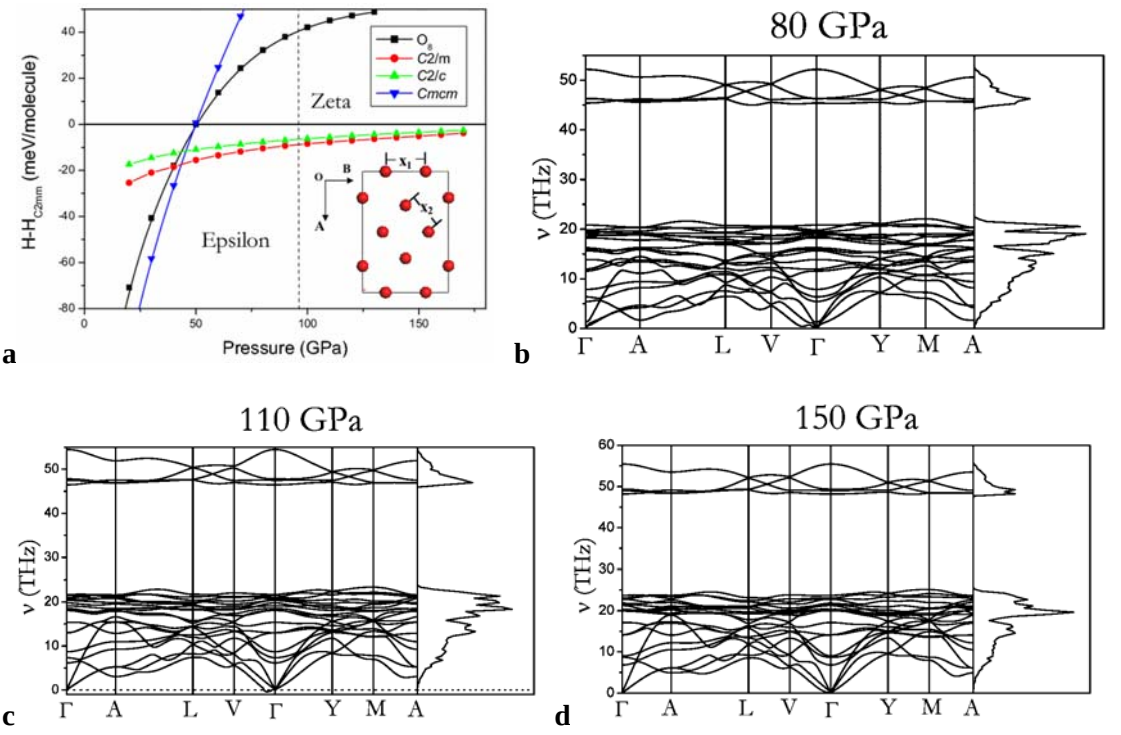


Fig. 18. High-pressure phases of oxygen: a) enthalpies (relative to the *C*2*mm* structure of Ref. [63]) of several possible structures as a function of pressure (from [62]) of several possible structures, b) phonon dispersion curves and densities of states of the *C*2/*m* ζ-phase at three pressures.

Reactivity of noble gases: are Xe-C compounds possible at high pressure? USPEX can be used also for studying multicomponent stoichiometric systems and their reactivity. Grochala [43] has suggested the possibility of stable Xe-C compounds at high pressure; indeed, carbon and xenon have similar valence orbital energies (*cf.* ionisation potentials of 12.13 eV and 11.26 eV for Xe and C, respectively) and one expects that pressure would make Xe more reactive [41]. We did simulations at 200 GPa, i.e. above the metallization pressure of Xe (132 GPa [77]), when its closed electronic shells are strongly perturbed. These calculations were done within the GGA [38] and on cells containing up to 14 atoms/cell. At this pressure all Xe carbides are extremely unstable (Fig. 19), but XeC_2 has a small negative volume of formation and might become stable at much higher pressures. All structures (Fig. 20) are made of close-packed Xe layers (i.e. fragments of the elemental Xe structure) and 3,4-connected carbon layers (intermediate between graphite and diamond), except the 3D-clathrate structure of XeC_8. The observed layering is, most likely, system size-dependent: increasing the number of atoms, we expect to see thicker elemental layers, to the point of complete phase separation. The presence of three-coordinate carbon and a clathrate structure at such high pressure can be explained by an effective negative pressure, which large Xe atoms exert on the much smaller C atoms. Some of the carbon layers are shown in Fig. 21; they are reminiscent of the low-pressure chemistry of carbon. Recently, we succeeded in predicting new stable oxides and silicates of xenon at much lower pressures (Jung *et al.*, in prep.). It seems that high electronegativity of the partner atom is more important for stability of noble gas compounds than similarity of orbital energies.

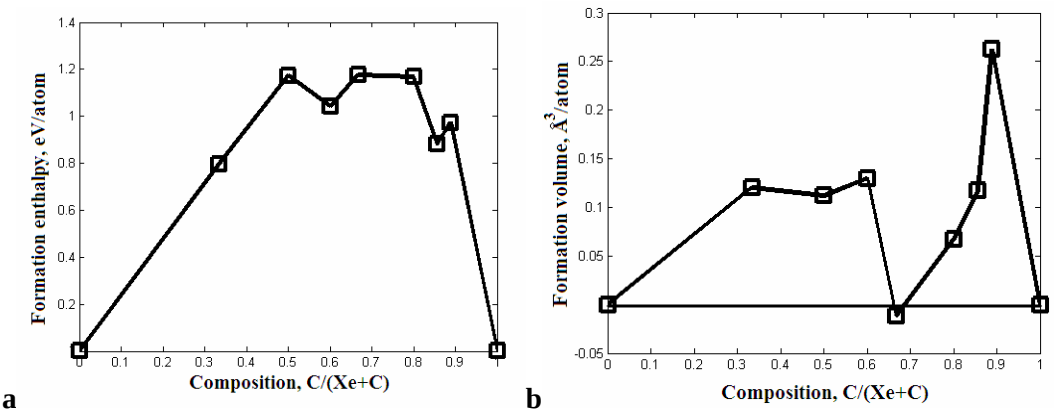


Fig. 19. Predicted enthalpy (a) and volume (b) of formation of Xe-C compounds at 200 GPa. The compounds shown are Xe (hcp), Xe_2C, XeC, Xe_2C_3, XeC_2, XeC_4, XeC_6, XeC_8, C(diamond).

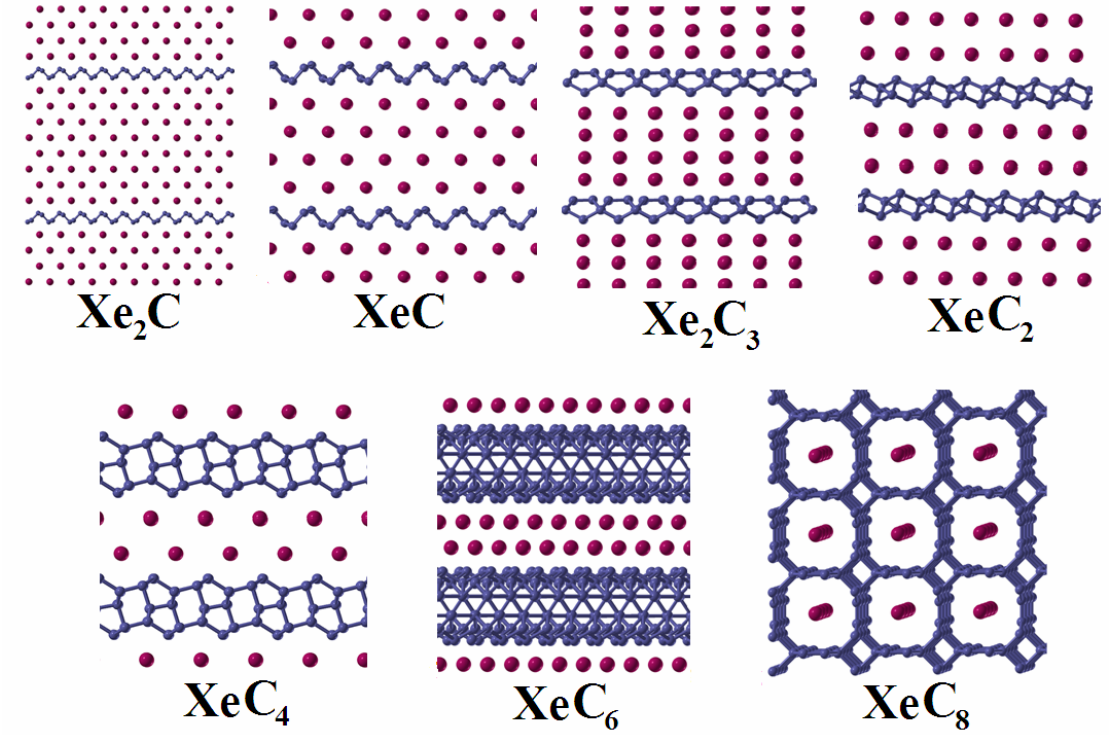


Fig. 20. Predicted structures of Xe_2C, XeC, Xe_2C_3, XeC_2, XeC_4, XeC_6, XeC_8 at 200 GPa.

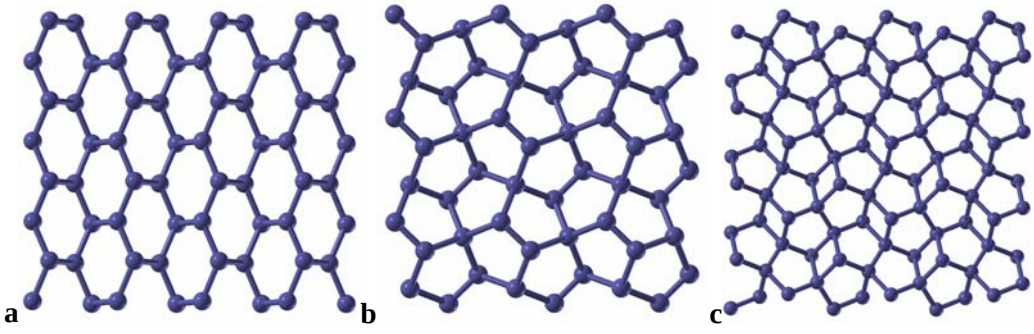


Fig. 21. Structures of carbon layers in a) Xe_2C and XeC, b) Xe_2C_3, c) XeC_2.

Another example of unmixing: Cu-C system. While many metals easily form carbides already at ambient pressure, no stable compounds are known in the Cu-C system. 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. 22). When the tendency to unmixing is not so large, simulations on small unit cells may find metastable "mixed" structures (see below, Fig. 23c). 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. 22).

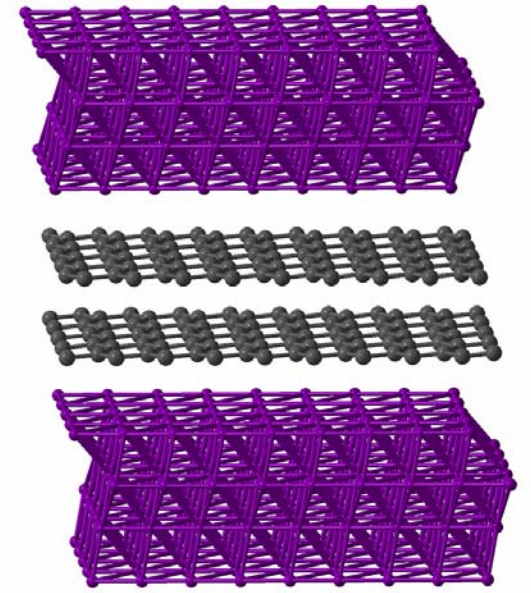


Fig. 22. Lowest-energy structure of Cu_2C with 12 atoms/cell at 1 atm.

Investigation of possible insulating phases of $Al_{12}C$ and $Al_{13}K$. Works [35]-[37] proposed a new class of solids, made of Al clusters – ionic insulator $(Al_{13})K$ with the CsCl-type structure and icosahedral [35] or cuboctahedral [36] Al_{13}^- ions, and molecular $Al_{12}C$ with icosahedral [37] $Al_{12}C$ clusters bound together by weak intermolecular interactions. Icosahedral nanoclusters are well known for aluminium, boron, and other elements, and according to the jellium model, clusters with 40 valence electrons should be particularly stable and $Al_{12}C$ and Al_{13}^- icosahedra groups may have closed-shell electronic configurations and produce insulating structures. Such materials, if they can be synthesized, would be technologically very interesting materials. Recently [69] we explored these systems (with the exception that for $Al_{12}C$ we study only the 13-atom system, rather than the previously [37] suggested 104-atom cell) using USPEX. In both cases the lowest-energy structures do not contain any well-defined clusters.

For $Al_{13}K$, when full relaxation is performed, the icosahedral structure (Fig. 23a) spontaneously transforms into the cuboctahedral one (Fig. 23b). The cuboctahedra are fragments of the fcc structure; however, these are not isolated in $Al_{13}K$, but form bonds with

each other (Al-Al distances within the cluster are 2.78 Å, and 2.72 Å between the clusters), which makes this structure metallic and indicates limitations of the jellium model. The structure found with USPEX (Fig. 23c) is 78 meV/atom lower in energy, has no clusters and shows a strong perturbation exerted by K atoms on the Al-sublattice. No stable compounds are known in the Al-K system – this is consistent with our result that even the USPEX-produced structure is 72 meV/atom higher in energy than the mechanical mixture of pure Al and K.

For $Al_{12}C$, the optimized icosahedral structure (Fig. 24a) also has bonds between the clusters: Al-Al distances are 2.67-2.72 Å within the cluster and 2.65 Å between the clusters. Thus, the expectation of a molecular structure made on the basis of the jellium model is again incorrect. This structure has a DFT band gap of 0.3 eV, but is not stable: with USPEX, we found a much more stable (by 557 meV/atom) structure shown in Fig. 24b. That structure is close-packed, similar to the fcc structure of pure Al, but with stacking faults at which C atoms occupy the octahedral voids in the aluminium close packing. Thus, C impurities in Al may be correlated with the presence of stacking faults. The USPEX-produced structure of $Al_{12}C$ is metallic and only 39 meV/atom less stable than the mechanical mixture of pure aluminium and graphite. There is a stable phase in the Al-C system, Al_4C_3.

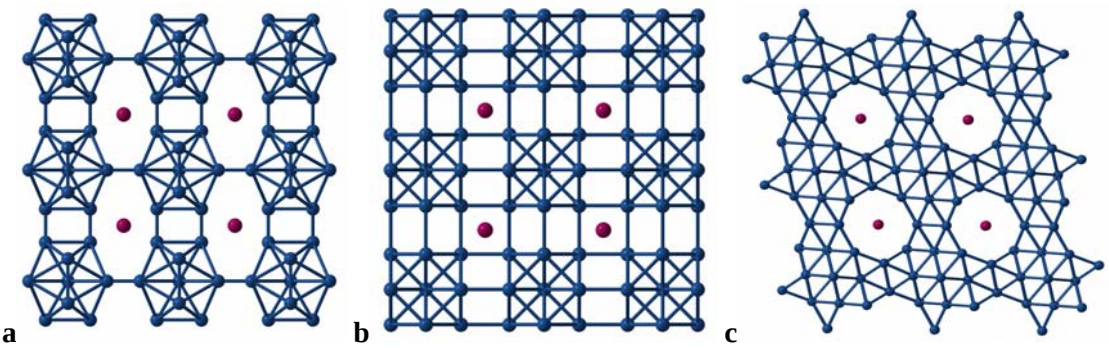


Fig. 23. Structures of $Al_{13}K$ with 14 atoms/cell: (a) with icosahedral Al_{13} clusters, (b) with cuboctahedral Al_{13} clusters, (c) structure produced by USPEX. From [69].

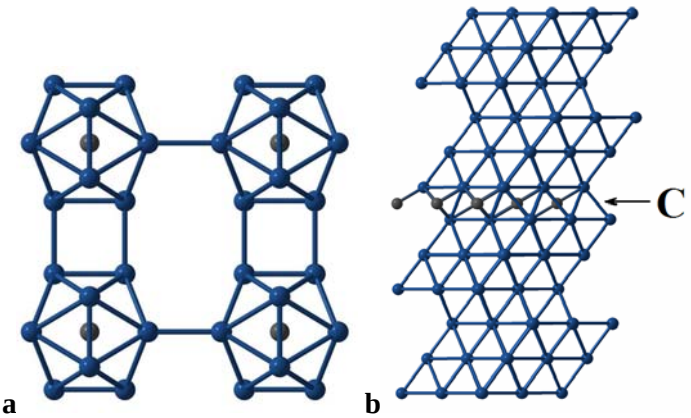


Fig. 24. Structures of $Al_{12}C$ with 13 atoms/cell: (a) with icosahedral clusters, (b) structure produces by USPEX. From [69].

5 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 impossible). We are suggesting USPEX as the method of choice for crystal structure prediction of systems with ~ 6 – 40 atoms/cell, where no information (or just the lattice parameters) is available. Above 40 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'. Fortunately, many (or most) systems of practical or fundamental interest fall in the size range accessible to USPEX. The extension of the method to molecular systems (i.e. handling whole molecules, rather than individual atoms) is already available. 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. Difficult cases are aperiodic and disordered systems (for which only the lowest-energy periodic approximants and ordered structures can be predicted at this moment). Additional work is required to extend it further to very large (> 200 atoms/cell) systems and to systems with variable composition. In the latter case, one would simultaneously predict all the stable stoichiometries in a given range of compositions, and the corresponding stable structures. A pioneering study [90] succeeded in predicting stable stoichiometries of alloys within a given structure type, but simultaneous prediction of structure and stoichiometry remains an unsolved problem. Currently, we are working on such extensions of the method.

Acknowledgements

Calculations were performed at the Joint Russian Supercomputer Centre (Russian Academy of Sciences, Moscow), ETH Zurich and CSCS (Manno). We thank R. Hoffmann and W. Grochala for discussions of Xe-C compounds, S. Ono for numerous discussions, D. Vicente (Barcelona Supercomputer Centre), T. Racic and G. Sigut (ETH Zurich) for computational assistance. This work is supported by the Swiss National Science Foundation (grant 200021–111847/1).

Appendix: How to compute the similarity matrix. Fingerprinting method.

Here we describe a method which one of us (M.V.) has developed to monitor population diversity in evolutionary simulations. To measure the similarity between the structures, we propose a clustering procedure based on some definition of the abstract distance between two structures. Any such method should be: 1) independent from translation and rotation of the structure; 2) independent from the choice of unit cell between equivalent unit cells; 3) robust

regarding small numerical errors; 4) independent from the ordering of cell axis and atoms in the cell; 5) independent from inversions and mirroring of the structures.

Among the number of papers (e.g. [91]-[94]) where structure descriptors and distance metrics were considered for crystals, we note the work of Chisholm and Motherwell [94], which employed interatomic distances and provided inspiration for our more general method. The work of Hundt et. al. [93] gives a comprehensive survey of existing methods with a focus on calculating some form of distance metric between structures.

Using this distance, we construct a similarity matrix where each element A_{ij} measures the similarity (in terms of the abstract distance) between the i-th and j-th structures. We prefer to rescale the abstract distances, so that for a maximally similar pair the similarity equals one, and for the most dissimilar pair it is zero.

Properties of similarity matrices:

1. The range of values is from 0 (blue) to 1 (red).
2. Similarity matrix is symmetric, with values of 1 on the main diagonal.
3. Blue lines indicate structures that have no significant similarity with any other structures explored.
4. Blocks of values close to the main diagonal show diversity of the population and its change along the simulation. Due to selection pressure, in evolutionary simulations diversity of the population decreases during the simulation – visually this can be detected as "reddening" of the similarity matrix towards its upper right corner.

To calculate the distance between the structures, a descriptor, called a fingerprint, is computed for each structure. The fingerprint can be defined in the real space (set of interatomic distances) or in the reciprocal space (structure factor), we prefer the more intuitive real-space constructions. One (out of an infinite number of possibilities) construction we use consists of components from all element types; each component contains the element type of the corresponding atom and an increasing ordered set of distances from it to the other atoms in an infinite crystal (Fig. 25). All components contain the same number of distances and it is sufficient to have distances up to one half of the longest unit cell side in the examined set of structures.

To cluster fingerprints, and therefore to cluster the associated structures, a distance between them should be defined. We start defining the distance between two fingerprint components. The simplest measure is the Euclidean distance between the N-dimensional points whose coordinates are the various component distances (*da* for first fingerprint and *db* for the second one).

$$D = \sqrt{\sum_i (da_i - db_i)^2}$$

To increase the discriminating power, one should account for the linear increase of errors with distance and the quadratic increase with distance of the number of distance values in the fingerprint, e.g.:

$$D = \sqrt{\sum_i ((da_i - db_i) / (da_i + db_i)^3)^2}$$

This algorithm has been implemented inside the molecular visualization environment STM3 [95][96] and uses its visualization facilities to help guide and control the algorithm behavior. The method uses a library for approximate nearest neighbor searching [97] to compute the number of distances needed for each structure atoms.

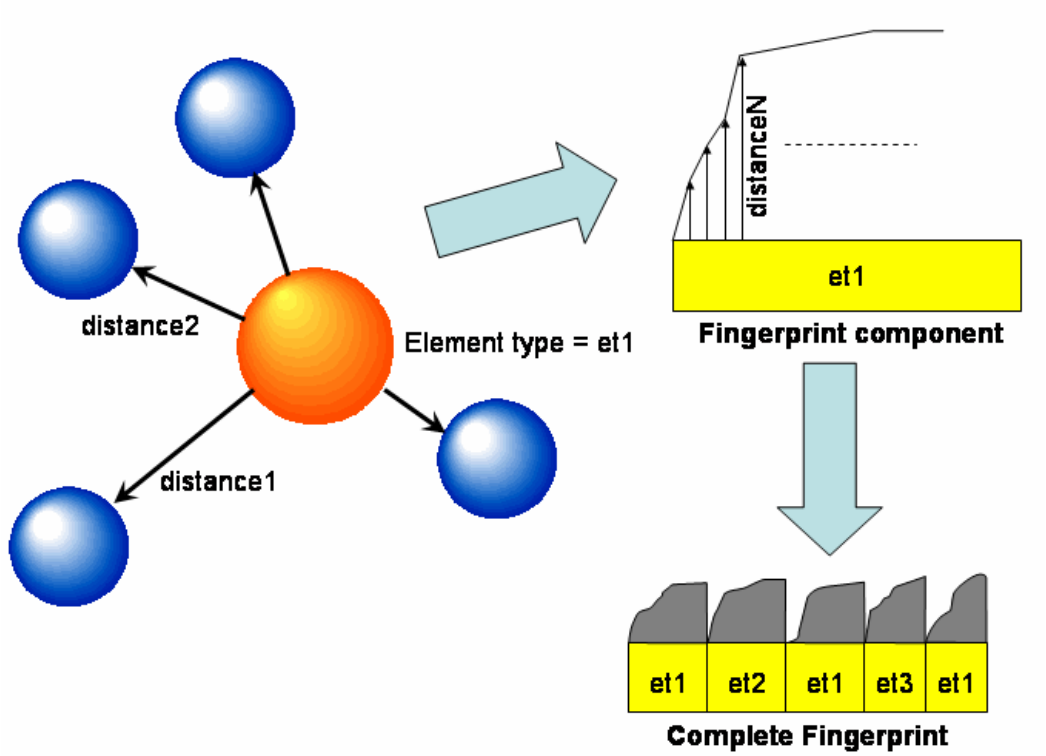


Fig. 25. Structure of a fingerprint.

References

[1] Maddox J. (1988). Crystals from first principles. *Nature* **335**, 201.

[2] 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.

[3] Urusov V.S., Dubrovinskaya N.A., Dubrovinsky L.S. (1990). Generation of likely crystal structures of minerals. Moscow State University Press, Moscow.

[4] Deem M. W., Newsam, J. M. (1989). Determination of 4-connected framework crystal structures by simulated annealing. *Nature* **342**, 260-262.

[5] Pannetier J., Bassasalsina J., Rodriguez-Carva jal J., and Caignaert V. (1990) Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature*, **346**, 343–345.

[6] Boisen M.B., Gibbs G.V., Bukowinski M.S.T. (1994). Framework silica structures generated using simulated annealing with a potential energy function based on an H6Si2O7 molecule. *Phys. Chem. Minerals* **21**, 269-284.

[7] Schön J.C. and 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.
[8] 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.
[9] Martonak R., Laio A., and Parrinello M. (2003). Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys. Rev. Lett*, **90**, 075503.
[10] Martonak R., Laio A., Bernasconi M., Ceriani C., Raiteri P., Zipoli F., and Parrinello M. (2005). Simulation of structural phase transitions by metadynamics. *Z. Krist.* **220**, 489–498.
[11] Martoňák R., Donadio D., Oganov A.R., Parrinello M. (2006). Crystal structure transformations in SiO_2 from classical and *ab initio* metadynamics. *Nature Materials* **5**, 623-626.
[12] Pickard C.J., Needs R.J. (2006). High-pressure phases of silane. *Phys. Rev. Lett.* **97**, art. 045504.
[13] Oganov A.R., Glass C.W., Ono S. (2006). High-pressure phases of $CaCO_3$: crystal structure prediction and experiment. *Earth Planet. Sci. Lett.* **241**, 95-103.
[14] Oganov A.R., Glass C.W. (2006). Crystal structure prediction using *ab initio* evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, art. 244704.
[15] Glass C.W., Oganov A.R., Hansen N. (2006). USPEX – evolutionary crystal structure prediction. *Comp. Phys. Comm.* **175**, 713-720.
[16] Raiteri P., Martonak R., Parrinello M. (2005). Exploring polymorphism: The case of benzene. *Angew. Chemie Intl. Ed.* **44**, 3769-3773.
[17] Oganov A.R., Martoňák R., Laio A., Raiteri P., Parrinello M. (2005). Anisotropy of Earth's D" layer and stacking faults in the $MgSiO_3$ post-perovskite phase. *Nature* **438**, 1142-1144.
[18] Martoňák R., Donadio D., Oganov A.R., Parrinello M. (2007). 4- to 6- coordinated silica: transformation pathways from metadynamics. *Phys. Rev.* **B76**, 014120.
[19] Bush T.S., Catlow C.R.A. & Battle P.D. (1995). Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.* **5**, 1269-1272.
[20] Woodley S.M. (2004). Prediction of crystal structures using evolutionary algorithms and related techniques. S*tructure and Bonding* **110**, 95-132.
[21] Woodley S.M., Battle P.D., Gale J.D., Catlow C.R.A. (1999). The prediction of inorganic crystal structures using a genetic algorithm and energy minimization. *Phys. Chem. Chem. Phys.* **1**, 2535-2542.
[22] Bazterra V.E., Ferraro M.B., Facelli J.C. (2002). Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *J. Chem. Phys.* **116**, 5984-5991.
[23] Gottwald D., Kahl G., Likos C.N. (2005). Predicting equilibrium structures in freezing processes. *J. Chem. Phys.* **122**, art. 204503.
[24] Deaven D.M., Ho K.M. (1995). Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **75**, 288-291.
[25] Kresse G. & Furthmüller J. (1996). Efficient iterative schemes for *ab initio* total-energy calculations using a plane wave basis set. *Phys. Rev.* **B54**, 11169-11186.
[26] Soler J.M., Artacho E., Gale J.D., 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.
[27] Gale J.D. (2005). GULP: Capabilities and prospects. *Z. Krist.* 220, 552-554.
[28] Martonak R., Oganov A.R., Glass C.W. (2007). Crystal structure prediction and simulations of structural transformations: metadynamics and evolutionary algorithms. *Phase Transitions* **80**, 277-298.
[29] Ono S., Kikegawa T., Ohishi Y. (2007). High-pressure transition of $CaCO_3$. *Am. Mineral.* **92**, 1246-1249.

[30] Ma Y.-M., Oganov A.R., Glass C.W. (2007). Structure of the metallic ζ-phase of oxygen and isosymmetric nature of the ε-ζ phase transition: *Ab initio* simulations. *Phys. Rev.* **B76**, art. 064101.
[31] Oganov A.R., Glass C.W., Ma Y.-Z., Ma Y.-M., Chen J. (2007). Ionic high-pressure form of elemental boron. *Nature*, under review.
[32] Ono S., Oganov A.R., Brodholt J.P., Vocadlo L., Wood I.G., Glass C.W., Côté A.S., Price G.D. (2007). High-pressure phase transformations of FeS: novel phases at conditions of planetary cores. *Earth Planet. Sci. Lett.*, under review.
[33] Oganov, A.R. & Ono, S. (2004). Theoretical and experimental evidence for a post-perovskite phase of MgSiO3 in Earth's D" layer. *Nature* **430**, 445-448.
[34] Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. (2004). Post-perovskite phase transition in MgSiO3. *Science* **307**, 855-858.
[35] Khanna S.N., Jena P. (1994). Designing ionic solids from metallic clusters. *Chem. Phys. Lett.* **219**, 479-483.
[36] Liu F., Mostoller M., Kaplan T., Khanna S.N., Jena P. (1996). Evidence for a new class of solids. First-principles study of $K(Al)_{13}$. *Chem. Phys. Lett.* **248**, 213-217.
[37] Gong X.G. (1997). Structure and stability of cluster-assembled solid $Al_{12}C(Si)$: a first-principles study. *Phys. Rev.* **B56**, 1091-1094.
[38] Perdew J.P., Burke K., Ernzerhof M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865-3868.
[39] Blöchl P.E. (1994). Projector augmented-wave method. *Phys. Rev.* **B50**, 17953-17979.
[40] Kresse G., Joubert D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev.* **B59**, 1758-1775.
[41] Sanloup C., Schmidt B.C., Perez E.M.C., Jambon A., Gregoryanz E., Mezouar M. (2005). Retention of xenon in quartz and Earth's missing xenon. *Science* **310**, 1174-1177.
[42] Hoffmann R., Hughbanks T., Kertesz M., Bird P.H. (1983). A hypothetical metallic allotrope of carbon. *J. Am. Chem. Soc.* **105**, 4831-4832.
[43] Grochala W. (2007). Atypical compounds of gases, which have been called 'noble'. *Chem. Soc. Rev.* **36**, 1632 -1655.
[44] Shcheka S.S., Wiedenbeck M., Frost D.J., Keppler H. (2006). Carbon solubility in mantle minerals. *Earth Planet. Sci. Lett.* **245**, 730-742.
[45] Ono S., Kikegawa T., Ohishi Y., Tsuchiya J. (2005). Post-aragonite phase transformation in $CaCO_3$ at 40 GPa, *Am. Mineral.* **90**, 667-671.
[46] Ono S., Kikegawa T., Ohishi Y. (2007). High-pressure phase transition of $CaCO_3$. Am. Mineral. **92**, 1246-1249.
[47] Arapan S., Souza de Almeida J., Ahuja R. (2007). Formation of sp^3 hybridized bonds and stability of $CaCO_3$ at very high pressure. *Phys. Rev. Lett.* **98**, art. 268501.
[48] Lasaga A.C., Gibbs G.V. (1987). Applications of quantum-mechanical potential surfaces to mineral physics calculations. *Phys. Chem. Minerals* **14**, 107-117.
[49] Santoro M., Gorelli F.A., Bini R., Ruocco G., Scandolo S., Crichton W.A. (2006). Amorphous silica-like carbon dioxide. *Nature* **441**, 857-860.
[50] Oganov A.R., Ma Y., Glass C.W., Garcia A. (2007). High-pressure structures of $MgCO_3$ and CO_2: an *ab initio* evolutionary study of carbon storage in the Earth's mantle. *In prep.*
[51] Bonev S.A., Gygi F., Ogitsu T., Galli G. (2003). High-pressure molecular phases of solid carbon dioxide. *Phys. Rev. Lett.* **91**, 065501.
[52] Yoo C.S., Cynn H., Gygi F., Galli G., Iota V., Nicol M., Carlson S., Hausermann D., Mailhiot C. (1999). Crystal structure of carbon dioxide at high pressure: "Superhard" polymeric carbon dioxide. *Phys. Rev. Lett.* **83**, 5527-5530.
[53] Holm B., Ahuja R., Belonoshko A., Johansson B. (2000). Theoretical investigation of high pressure phases of carbon dioxide. *Phys. Rev. Lett.* **85**, 1258-1261.
[54] Dong J.J., Tomfohr J.K., Sankey O.F., Leinenweber K., Somayazulu M., McMillan P.F. (2000). Investigation of hardness in tetrahedrally bonded nonmolecular CO_2 solids by density-functional theory. *Phys. Rev.* **B62**, 14685-14689.
[55] Nicol M., Hirsch K.R., Holzapfel W.B. (1979). Oxygen phase equilibria near 298 K. *Chem. Phys. Lett.* 68, 49-52.

[56] 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.
[57] Lundegaard L.F., Weck G., McMahon M.I., Desgreniers S., Loubeyre P. (2006). Observation of an O_8 molecular lattice in the epsilon phase of solid oxygen. *Nature* **443**, 201-204.
[58] 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.
[59] 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.
[60] Shimizu K., Suhara K., Ikumo M., Eremets M.I., Amaya K. (1998). Superconductivity in oxygen. *Nature* **393**, 767-769.
[61] Goncharenko I.N. (2005). Evidence for a magnetic collapse in the epsilon phase of solid oxygen. *Phys. Rev. Lett.* 94, art. 205701.
[62] Ma Y.-M., Oganov A.R., Glass C.W. (2007). Structure of metallic ζ-phase of oxygen. *Phys. Rev.***B**, in press.
[63] Serra S., Chiarotti G., Scandolo S., Tosatti E. (1998). Pressure-induced magnetic collapse and metallization of molecular oxygen: The ζ-O_2 phase. *Phys. Rev. Lett.* **80**, 5160-5163.
[64] Neaton J.B., Ashcroft N.W. (2002). Low-energy linear structures in dense oxygen: Implications for the epsilon phase. *Phys. Rev. Lett.* **88**, 205503.
[65] Christy A.G. (1995). Isosymmetric Structural Phase Transitions: Phenomenology and Examples. *Acta Cryst.* **B51**, 753-757.
[66] Steudel R., Wong M.W. (2007). Dark-red O_8 molecules in solid oxygen: rhomboid clusters, not S_8-like rings. *Angew. Chem. Int. Ed.* **46**, 1768-1771.
[67] Hebbache M., Zemzemi M. (2004). *Ab initio* study of high-pressure behavior of a low compressibility metal and a hard material: Osmium and diamond *Phys. Rev.* **B70**, 224107.
[68] Bonev S.A., Gygi F., Ogitsu T., Galli G. (2003). High-pressure molecular phases of solid carbon dioxide. Phys. Rev. Lett. 91, 065501.
[69] Oganov A.R., Glass C.W. (2007). Evolutionary crystal structure prediction as a tool in materials design. *J. Phys.: Cond. Matter*, in press.
[70] Chaplot S.L., Rao K.R. (2006). Crystal structure prediction – evolutionary or revolutionary crystallography? *Current Science* **91**, 1448-1450.
[71] Abraham N.L., Probert M.I.J. (2006). A periodic genetic algorithm with real-space representation for crystal structures and polymorph prediction. *Phys. Rev.* **B73**, art. 224104.
[72] Trimarchi G., Zunger A. (2007). Global space-group optimization problem: finding the stablest crystal structure without constraints. *Phys. Rev.* **B75**, art. 104113.
[73] Oganov A.R., Glass C.W. Comment on "Global space-group optimization problem: finding the stablest crystal structure without constraints", Phys. Rev. B., submitted.
[74] Barabash S.V., 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.* **B74**, art. 035108.
[75] Valle M. (2005). STM3: a chemistry visualization platform. *Z. Krist.* 220, 585-588.
[76] Martoňák R., Oganov A.R., Glass C.W. (2007). Crystal structure prediction and simulations of structural transformations: metadynamics and evolutionary algorithms. *Phase Transitions* **80**, 277-298.
[77] Goettel K.A., Eggert J.H., Silvera I.F., Moss W.C. (1989). Optical evidence for the metallization of xenon at 132(5) GPa. *Phys. Rev. Lett.* **62**, 665-668.
[78] 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.

[79] Sluiter M.H.F., Colinet C., Pasturel A. (2006). *Ab initio* calculation of phase stability in Au-Pd and Ag-Pt alloys. *Phys. Rev.* **B73**, 174204.
[80] Ross M., Ree F.H. (1980). Repulsive forces of simple molecules and mixtures at high density and temperature. *J. Chem. Phys.* **73**, 6146-6152.
[81] Benedetti L.R., Nguyen J.H., Caldwell W.A., Liu H., Kruger M., Jeanloz R. (1999). Dissociation of CH_4 at high pressures and temperatures: diamond formation in giant planet interiors? *Science* **286**, 100-102.
[82] Zerr A., Serghiou G., Boehler R., Ross M. (2006). Decomposition of alkanes at high pressures and temperatures. *High-Pres. Res.* **26**, 23-32.
[83] Ancilotto F., Chiarotti G.L., Scandolo S., & Tosatti E. (1997). Dissociation of methane into hydrocarbons at extreme (planetary) pressure and temperature. *Science* **275**, 1288-1290.
[84] King, H.E., Prewitt, C.T., 1982. High-pressure and high-temperature polymorphism of iron sulfide (FeS). *Acta Cryst.* **B38**, 1877-1887.
[85] Fang Z., Terakura K., Sawada H., Miyazaki T., Solovyev I. (1998). Inverse versus normal NiAs structures as high-pressure phases of FeO and MnO. *Phys. Rev. Lett.* **81**, 1027-1030.
[86] Mazin I.I., Fei Y.W., Downs R., Cohen R. (1998). Possible polytypism in FeO at high pressures. *Am. Mineral.* **83**, 451-457.
[87] Fei Y.W., Mao H.K. (1994). In situ determination of the NiAs phase of FeO at high pressure and temperature. *Science* **266**, 1678-1680.
[88] Kondo T., Ohtani E., Hirao N., Yagi T., Kikegawa T. (2004). Phase transitions of (Mg,Fe)O at megabar pressures. *Phys. Earth Planet. Int.* **143**, 201-213.
[89] Murakami M., Hirose K., Ono S., Tsuchiya T., Isshiki M., Watanuki T. (2004). High pressure and high temperature phase transitions of FeO. *Phys. Earth Planet. Int.* **146**, 273-282.
[90] Jóhannesson G.H., Bligaard T., Ruban A.V., Skriver H.L., Jacobsen K.W., and Nørskov J.K. (2002). Combined Electronic Structure and Evolutionary Search Approach to Materials Design. *Phys. Rev. Lett.* **88**, art. 255506.
[91] Dzyabchenko A.V. (1994). Method of crystal-structure similarity searching, *Acta Cryst.* **B50**, 414-425.
[92] Van Eijck B. P., Kroon J. (1997). Fast clustering of equivalent structures in crystal structure prediction, *J. Comput. Chem.* **18**, 1036-1042.
[93] Hundt R., Schön J.C. and Jansen M. (2006). CMPZ – an algorithm for the efficient comparison of periodic structures, *J. Appl. Cryst.* **39**, 6-16.
[94] Chisholm J.A., Motherwell S. (2005). COMPACK: a program for identifying crystal structure similarity using distances, *J. Appl. Cryst.* **38**, 228-231.
[95] Valle M. (2005). STM3: a chemistry visualization platform. *Z. Krist.* **220**, 585-588, 2005.
[96] Valle M. STM4, the STM3 successor: *http://www.cscs.ch/~mvalle/STM4* (accessed 2007-09-16).
[97] Mount D.M., Arya S. ANN: A Library for Approximate Nearest Neighbor Searching, *http://www.cs.umd.edu/~mount/ANN* (accessed 2007-06-01).
[98] Tremel W., Hoffmann R., Silvestre J. (1986). Transitions between NiAs and MnP type phases: an electronically driven distortion of triangular (3^6) nets. *J. Am. Chem. Soc.* **108**, 5174-5187.