MARCH 2011
VOLUME 44
NUMBER 3
pubs.acs.org/accounts

ACCOUNTS
of chemical research

ACS Publications
MOST TRUSTED. MOST CITED. MOST READ.
www.acs.org

ACCOUNTS
of chemical research

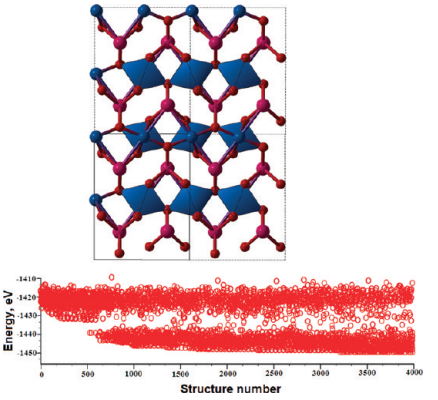
How Evolutionary Crystal Structure Prediction Works—and Why

ARTEM R. OGANOV,*,†,‡ ANDRIY O. LYAKHOV,† AND MARIO VALLE§

†*Department of Geosciences and Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-2100, United States,* ‡*Geology Department, Moscow State University, 119992 Moscow, Russia, and* §*Data Analysis and Visualization Group, Swiss National Supercomputing Centre (CSCS), via Cantonale, Galleria 2, 6928 Manno, Switzerland*

RECEIVED ON OCTOBER 2, 2010

CONSPECTUS



Once the crystal structure of a chemical substance is known, many properties can be predicted reliably and routinely. Therefore if researchers could predict the crystal structure of a material before it is synthesized, they could significantly accelerate the discovery of new materials. In addition, the ability to predict crystal structures at arbitrary conditions of pressure and temperature is invaluable for the study of matter at extreme conditions, where experiments are difficult.

Crystal structure prediction (CSP), the problem of finding the most stable arrangement of atoms given only the chemical composition, has long remained a major unsolved scientific problem. Two problems are entangled here: search, the efficient exploration of the multidimensional energy landscape, and ranking, the correct calculation of relative energies. For organic crystals, which contain a few molecules in the unit cell, search can be quite simple as long as a researcher does not need to include many possible isomers or conformations of the molecules; therefore ranking becomes the main challenge. For inorganic crystals, quantum mechanical methods often provide correct relative energies, making search the most critical problem. Recent developments provide useful practical methods for solving the search problem to a considerable extent. One can use simulated annealing, metadynamics, random sampling, basin hopping, minima hopping, and data mining. Genetic algorithms have been applied to crystals since 1995, but with limited success, which necessitated the development of a very different evolutionary algorithm. This Account reviews CSP using one of the major techniques, the hybrid evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography).

Using recent developments in the theory of energy landscapes, we unravel the reasons evolutionary techniques work for CSP and point out their limitations. We demonstrate that the energy landscapes of chemical systems have an overall shape and explore their intrinsic dimensionalities. Because of the inverse relationships between order and energy and between the dimensionality and diversity of an ensemble of crystal structures, the chances that a random search will find the ground state decrease exponentially with increasing system size. A well-designed evolutionary algorithm allows for much greater computational efficiency.

We illustrate the power of evolutionary CSP through applications that examine matter at high pressure, where new, unexpected phenomena take place. Evolutionary CSP has allowed researchers to make unexpected discoveries such as a transparent phase of sodium, a partially ionic form of boron, complex superconducting forms of calcium, a novel superhard allotrope of carbon, polymeric modifications of nitrogen, and a new class of compounds, perhydrides. These methods have also led to the discovery of novel hydride superconductors including the "impossible" LiH_n ($n = 2, 6, 8$) compounds, and $CaLi_2$. We discuss extensions of the method to molecular crystals, systems of variable composition, and the targeted optimization of specific physical properties.

Published on the Web 03/01/2011 www.pubs.acs.org/accounts
10.1021/ar1001318

1. Combinatorial Complexity of the Problem

Following a simple combinatorial argument,[1] the number of possible distinct structures can be evaluated as

$$C = \binom{V/\delta^3}{N} \prod_i \binom{N}{n_i} \quad (1)$$

where N is the total number of atoms in the unit cell of volume V, δ is a relevant discretization parameter (for instance, 1 Å) and n_i is the number of atoms of ith type in the unit cell. Already for small systems ($N \approx 10-20$), C is astronomically large (roughly 10^N if one uses $\delta = 1$ Å and typical atomic volume of 10 Å^3).

It is useful to consider the dimensionality of the energy landscape:

$$d = 3N + 3 \quad (2)$$

where $3N - 3$ degrees of freedom are the atomic positions, and the remaining six dimensions are lattice parameters. For example, a system with 20 atoms/cell poses a 63-dimensional problem! We can rewrite eq 1 as $C \sim \exp(\alpha d)$, where α is some system-specific constant. With such high-dimensional problems, simple exhaustive search strategies are clearly unfeasible.

The global optimization problem can be greatly simplified if combined with relaxation (local optimization). During relaxation, certain correlations between atomic positions set in: interatomic distances adjust to reasonable values and unfavorable interactions are avoided to some extent. The intrinsic dimensionality of this reduced energy landscape consisting only of local minima (Figure 1) is now

$$d^* = 3N + 3 - \kappa \quad (3)$$

where κ is the (noninteger) number of correlated dimensions. d^* depends both on system size and on chemistry. We found[2] $d^* = 10.9$ ($d = 39$) for Au_8Pd_4, $d^* = 11.6$ ($d = 99$) for $Mg_{16}O_{16}$, and $d^* = 32.5$ ($d = 39$) for $Mg_4N_4H_4$. The number of local minima is then

$$C^* \sim \exp(\beta d^*) \quad (4)$$

with $\beta < \alpha$, $d^* < d$, and $C^* \ll C$, implying that efficient search must include local optimization. Even simple random sampling, when combined with relaxation,[3] can deliver correct solutions for systems with $N < 8-10$. With USPEX, the limit is much higher, but the exponential increase of C^* with system size means that CSP is an NP-hard problem and for sufficiently large sizes CSP will always be intractable. In most cases, we are interested in systems with $N < 20-200$, and systems with $N < 100$ are tractable, while the range $100 < N < 200$ may become accessible in the foreseeable future.

2. How the Method Works

Evolutionary algorithms work best when the energy (or, more generally, fitness) landscape has an overall shape, as in Figure 1. Analysis[2] suggests such overall shape in the energy landscapes of chemical systems and implies that evolutionary algorithms are highly appropriate for CSP. Such overall structure is also expected for landscapes of many physical properties. In evolutionary simulations, a population of structures evolves, gradually "zooming in" on the most promising regions of the landscape and leading to further reduction of d^*.

The evolutionary algorithm USPEX (Universal Structure Predictor: Evolutionary Xtallography[1,4,5]), unlike a previous

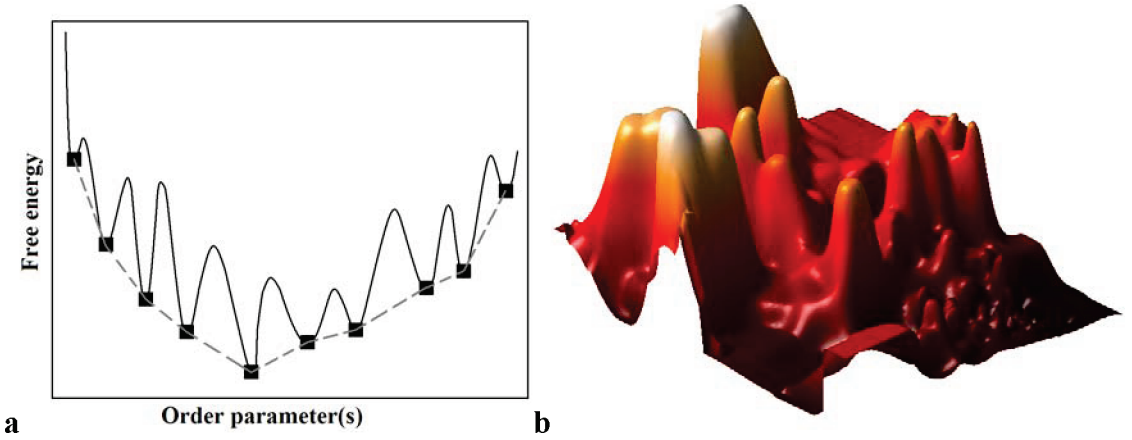


FIGURE 1. Energy landscape:[2] (a) 1D scheme showing the full landscape (solid line) and reduced landscape (dashed line joining local minima); (b) 2D projection of the reduced landscape of Au_8Pd_4, showing clustering of low-energy structures in one region.

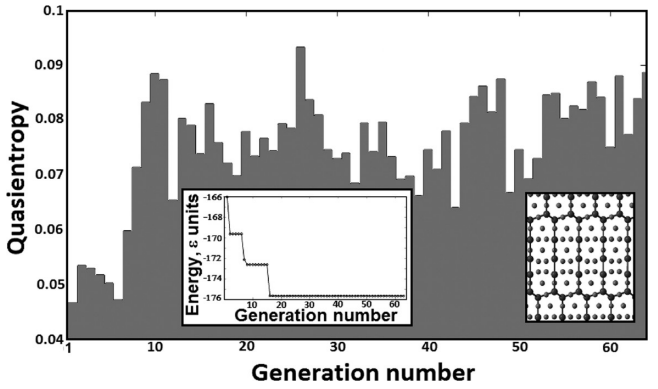


FIGURE 2. Evolutionary simulation of the binary Lennard-Jones A_5B_{16} (the potential model used here is known to yield low-energy quasicrystalline structures[9]). The insets show the lowest energy as a function of generation number, and the lowest-energy structure.

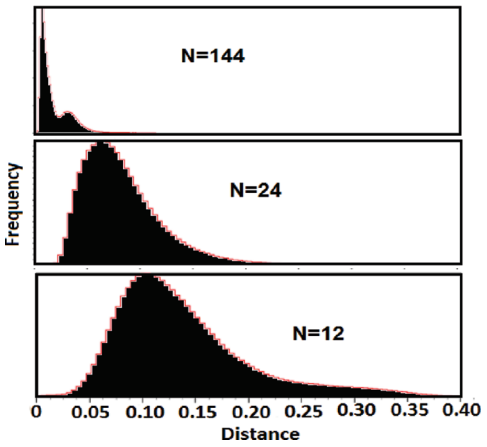


FIGURE 3. Distribution of distances between randomly sampled local minima in a binary Lennard-Jones system AB_2.

genetic algorithm for crystals[6] but similarly to an algorithm for clusters,[7] includes local optimization and treats structural variables as physical numbers, instead of nonintuitive binary "0/1" strings (the latter is the defining difference between "genetic" and more general "evolutionary" algorithms; the former use binary strings). Other important considerations are as follows:

(1) The algorithm incorporates "learning from history" (i.e., offspring structures bear resemblance to the more successful of the previously sampled structures), which is done through selection of the low-energy structures to become parents of the new generation, survival of the fittest structures, and variation operators (i.e., recipes for producing child structures from parents). Acting upon low-energy structures, variation operators lead, with high probability, to yet other low-energy structures. Four variation operators are used in our method:[1,4,5]
   (i) heredity (creating child structures from planar slabs cut from two parent structures[8])
   (ii) lattice mutation (large random deformation applied to the unit cell shape)
   (iii) permutation (swaps of chemical identity in pairs of chemically different atoms)
   (iv) special coordinate mutations (displacements of the atoms, but not in a fully random way, see below). For molecular crystals, where the structure is assembled from entire molecules (of a particular isomer), rigid or flexible, the above variation operators act on molecular centers, and additional variation operators must act on orientation and conformation of the molecules.

(2) The population should remain diverse, allowing very different solutions to be produced throughout the simulation. Diversity can be measured by the collective quasientropy, S_{coll}:

$$S_{coll} = -\langle (1 - D_{ij}) \ln(1 - D_{ij}) \rangle \quad (5)$$

where D_{ij} are abstract cosine distances between all pairs of structures (these distances measure structural dissimilarity and can only take values between 0 and 1[2]). Figure 2 shows that in a good simulation quasientropy retains large values and can exceed quasientropy of the first random generation, that is, evolutionary search not only is more efficient in finding low-energy structures but also can have more structural diversity than random search, thus depriving the latter of any potential advantages.

Initialization of the first generation can be random for small systems ($N < 20$). For large systems, most of the structures produced by random sampling will be very similar (Figure 3), disordered and with high energies.[2] It will be hard to produce good structures from such a population. There is an inverse relationship between the intrinsic dimensionality and the mean $\bar{\mu}$ of the distance distribution,

$$\bar{\mu} \approx (d^*)^{-m} \quad (6a)$$

and variance of this distribution,

$$\sigma \approx (d^*)^{-n} \quad (6b)$$

where positive *m and n* depend on the distance measure used (cosine vs Euclidean distances).

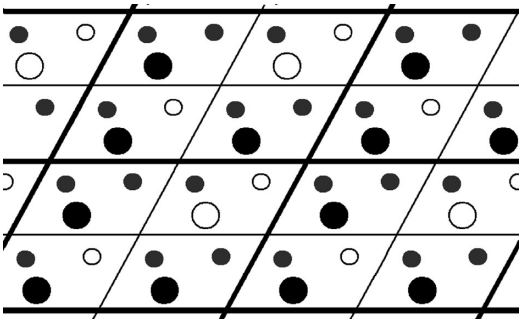


FIGURE 4. Pseudo-subcells for composition A_3B_6 (atoms A, large black circles; atoms B, small filled circles; vacancies, empty circles). The true cell (thick lines) is split into four pseudo-subcells (thin lines).

To obtain a diverse population, one should reduce the number of degrees of freedom in the first generation by (i) assembling initial structures from ready-made building blocks (molecules, coordination polyhedra, and low-energy seed structures) or (ii) generating the initial population using symmetry and/or pseudosymmetry. Since variation operators break symmetry, structures with different symmetries will have a chance to emerge. Consider splitting the unit cell into S subcells. When all n_i/S are integers, splitting is done into identical subcells, introducing additional translational symmetry. When n_i/S is a noninteger, random vacancies are created to maintain the correct number of atoms (Figure 4), introducing pseudosymmetry and leading to nontrivial ordered structures that are difficult to create otherwise. Such structures are well-known in nonstoichiometric compounds and even in the elements, for example, complex high-pressure phases Cs-III and Rb-III can be represented as supercells of the body-centered cubic structure with additional atoms (which can be thought of as additional partially occupied sublattices).

After a generation is completed, locally optimized structures are compared using their fingerprints,[10] and all nonidentical structures are ranked in order of their free energies.

The probability P of selecting a structure to be a parent is determined by its fitness rank i, e.g. in a linear scheme:

$$P(i) = P_1 - (i-1)\frac{P_1}{c}, \quad \sum_{i=1}^{c} P(i) = 1 \qquad (7)$$

where c is a selection cutoff. This scheme is superior to Boltzmann-type selection, because it is insensitive to peaks and gaps in energy distributions and does not require an additional parameter ("temperature") needed for defining Boltzmann probabilities; a quadratic analogue of (7) often works even better.

Niching (i.e., removal of identical structures using fingerprints[5,11]) allows a large number of lowest-energy structures to be carried over into the next generation, increasing the learning power, retaining diversity, and enabling a more thorough exploration of low-energy metastable structures.

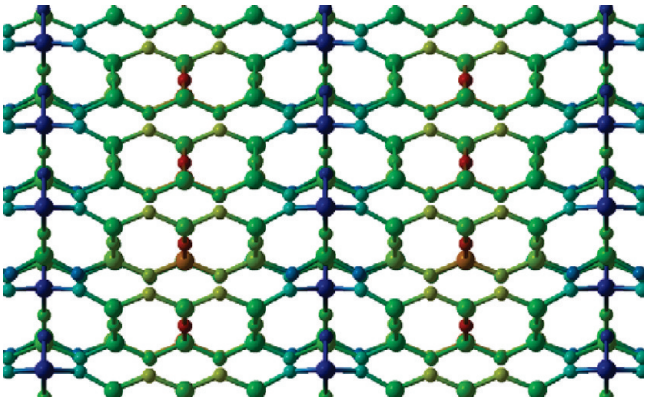


FIGURE 5. Illustration of the concept of local order for defective SiO_2. Low-order atoms are blue; high-order atoms are red. Low-order regions correspond to the planar defect.

The current algorithm is efficient for systems with <300 degrees of freedom and can be enhanced.[5] Directed moves that have a higher likelihood of leading to lower-energy structures or new promising areas of the energy landscape will be essential. For instance, moving the atoms along the eigenvectors of lowest-frequency phonon modes typically leads to low-energy structures. For this, one has to construct and solve the dynamical matrix, which is computationally extremely demanding at the *ab initio* level. We have efficiently solved this problem[5] by constructing an approximate dynamical matrix using the bond hardnesses computed from bond lengths and atomic covalent radii and electronegativities.

Given the usually good correlation between the energy and degree of order,[2] one could preferentially use the more ordered pieces of parent structures in the heredity operator. Figure 5 shows local order in a hypothetical defective structure of SiO_2; clearly, defective regions correspond to low-order atoms. Giving low-order atoms larger displacements while preserving positions of high-order atoms leads to a very effective coordinate mutation operator;[5] note that "blind" indiscriminate displacement of all atoms is much more likely to destroy than to create good structural motifs.[4]

Figure 6 shows different types of optimization. These examples show that with very minor adaptations, this method is powerful for solving a wide range of problems.

The limits of applicability of this approach are not fully known. Using a partly successful reimplementation of the method,[1,4] Trimarchi and Zunger[14] failed to predict

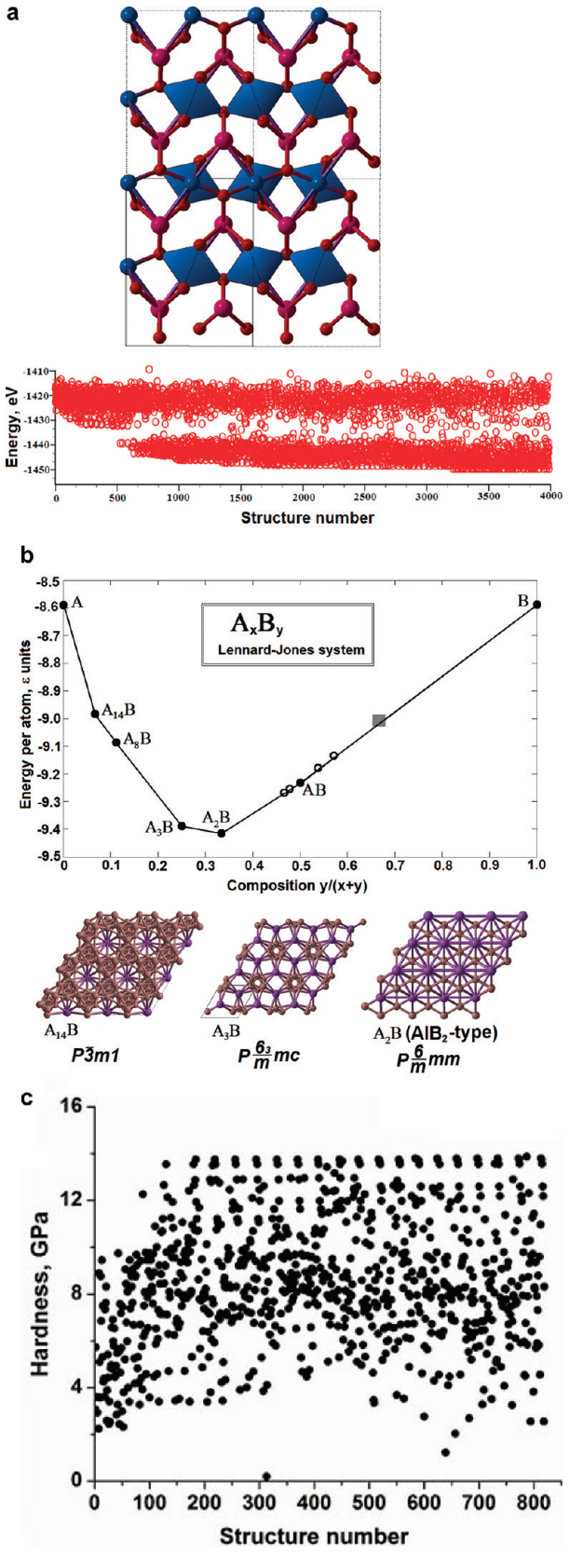


FIGURE 6. Predictions of (a) stable crystal structure of $MgSiO_3$ with 80 atoms in the unit cell,[12] (b) stable compounds and their structures in a binary Lennard-Jones system,[12] and (c) the hardest structure of TiO_2.[13]

fcc-ordered structures of the Au_8Pd_4 alloy that would be competitive with structures predicted by cluster expansion; however, this failure was probably due to inadequate Brillouin zone sampling. With this in mind, we found[12] a $P2_1/m$ ordering, more stable than any structure found to be stable by cluster expansion. Thus, atomic ordering can be efficiently predicted. What are the real limitations then? First, like for any NP-hard problem, dimensionality is a great restriction. The current version of the method was found to work very well for dimensionalities below 100−200. Recently, we could determine, at DFT level of theory, a very complex high-pressure structure of methane with 105 atoms/cell (Q. Zhu, unpublished), but to enable this, we had to operate with molecular, rather than atomic, entities, which considerably lowers the dimensionality of the problem.

Topology of the landscape is an important factor – single-funnel landscapes (as in Figure 1) are much more amenable for evolutionary algorithms than multifunnel or, even worse, featureless landscapes. Energy landscapes usually have an overall shape with a small number of funnels,[2] but landscapes of some properties may be more erratic.

Below we consider some this method's recent applications to high-pressure chemistry; most of these findings could not be expected by chemical intuition and required a powerful CSP method, nonempirical, unbiased and capable of arriving at completely unexpected solutions.

3. Applications

3.1. The High-Pressure Chemistry of "Inert" Atomic Cores and of the Electron Gas. Highly compressible atoms of alkali and alkali earth metals enter a chemically interesting regime at strong compression when their cores begin to overlap:[15] valence electrons get increasingly "trapped" in the interstitial space, and valence bandwidth decreases on compression. Vacant in the free atoms, p- and d-orbitals become dominant at strong compression, eventually making K, Rb, Cs, Ca, Sr, and Ba d-metals, while Li becomes a predominantly p-element, even adopting the diamond structure above 483 GPa.[16] The most interesting picture occurs for Na: in this element at megabar pressures, valence s-, p-, and d-orbitals are populated nearly equally.[17]

Sodium has a deep minimum on the melting curve at ~118 GPa and 300 K, below which extremely complex (and not resolved) crystal structures were observed.[18] We predicted[17] that above 250 GPa an *insulating* hP4 structure becomes stable, and this prediction was experimentally

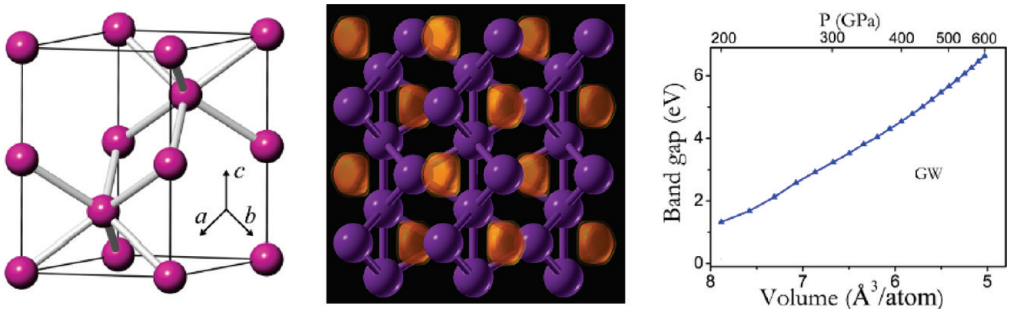


FIGURE 7. hP4 phase of sodium: crystal structure, electron localization function, and band gap.[17]

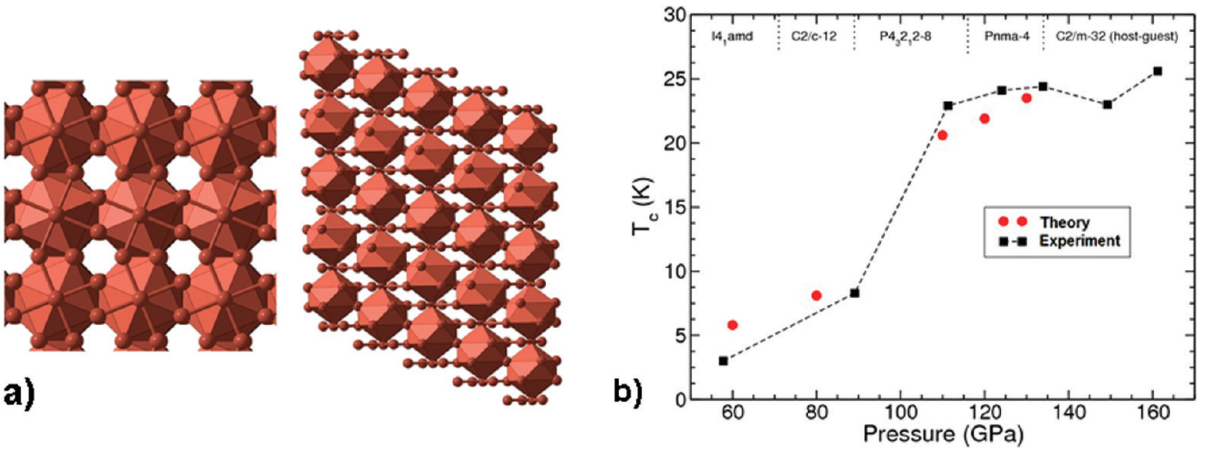


FIGURE 8. Calcium under pressure: (a) host−guest structure; (b) superconducting T_C as a function of pressure (experiment,[27] theory[26]). Dense Ca has the highest T_C among all elements.

verified,[17] although at lower pressures (>200 GPa).[19] The band gap, estimated using the state-of-the-art GW calculations,[17] turned out to be remarkably wide, from ~2 eV at 200 GPa to over 5 eV at 500 GPa. This implies that the hP4 phase should be optically transparent already at 200 GPa (this was experimentally confirmed[17]) and even colorless above ~320 GPa. The band gap is due to strong interstitial localization of electron pairs. The hP4 structure (Figure 7) can be described in several equivalent ways as

(i) NiAs-type structure where both sites are occupied by Na atoms.

(ii) Ni-sublattice of the Ni_2In structure, In sublattice being occupied by the interstitial electron pairs. The Ni_2In structure is known to be remarkably dense, considered to be the highest-pressure phase for AB_2 compounds[20] until a post-Ni_2In transition was discovered.[21]

(iii) Double hexagonal close packed structure. The stacking of close-packed layers of Na atoms is CACBCACB... (underlined layers contain interstitial electron pairs, Figure 7) and is squeezed by a factor of >2 along the *c*-axis, while the interstitial electron pairs form a nearly ideal hcp ABAB-stacking ($c/a \approx 1.3-1.6$).

The hP4 structure of Na minimizes core−valence overlap and maximizes packing efficiency of the interstitial electron pairs. It can be called an "electride",[22] that is, a "compound" made of ionic cores and localized interstitial electron pairs; electride forms have also been predicted in Li under pressure,[16] and Li was experimentally shown to become a semiconductor under pressure.[23] There are surprisingly faint hints of electride behavior in K in a narrow pressure range,[24] while Rb and Cs do not show electride behavior at all. This is due to the presence of shallow d-orbitals in heavy alkali and alkali earth elements; when d-states get populated under pressure, the atom becomes more compact (compactification due to s → d transfer was observed in experiments on collisions of Ca atoms;[25] since the Ca atom has no d-electrons in the core, the 3d-orbital, when occupied, is close to the nucleus). In sodium, d-levels are very high in energy and cannot be dominant below very high pressures; after the s → p transition, the 3p-valence electron still experiences strong repulsion from the core due to orthogonality to 2p-orbitals of the core, resulting in an increased interstitial localization on compression.

According to our calculations,[26] Ca is an electride in its low-pressure fcc (where it is a semiconductor in a narrow

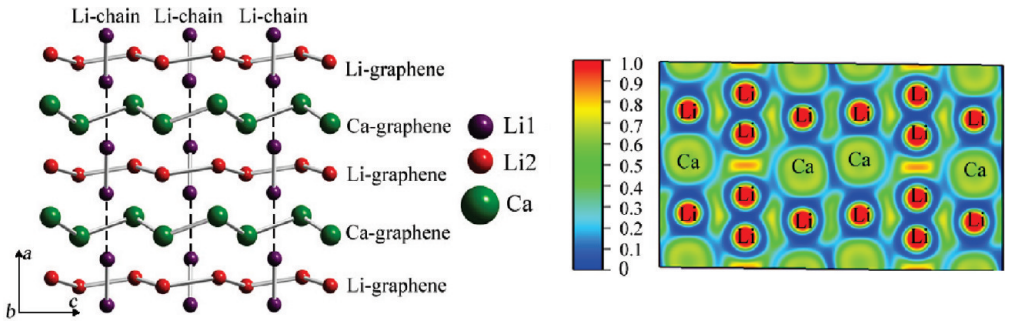


FIGURE 9. Structure and electron localization function of the $P2_1/c$ phase of $CaLi_2$, stable at 54−105 GPa.[28]

pressure range), bcc, and to some extent the β-tin forms, but at pressures above 71 GPa, interstitial electron localization disappears. Ca undergoes a large number of phase transitions, leading to an extremely stable (from 134 to 564 GPa) incommensurate host−guest structure (Figure 8) and, eventually, an hcp phase above 564 GPa. Calculations successfully reproduced its superconducting properties[27] across these transitions (Figure 8). The disappearance of interstitial electron localization can be viewed as "squeezing" the valence electrons into the vacant 3d-orbital in the core region, but it is unclear what makes the host−guest structure in Ca (or the Rb-IV structure in Na) so remarkably stable. Electride behavior was also seen[28] in the predicted high-pressure structures of $CaLi_2$ (Figure 9). For this very interesting material, initially variants of the low-pressure Laves phase were assumed,[29] but subsequent experiments[30] observed a very different diffraction pattern and superconductivity at pressures above 23 GPa. It was proposed[31] that $CaLi_2$ decomposes under pressure. The current picture[28] is that this compound is unstable to decomposition at 20−35 GPa, and at >105 GPa, while at intermediate pressures, novel polymorphs of $CaLi_2$ are stable: the $C2/c$ phase (related to the $ThSi_2$ structure and containing Li_2 pairs) is stable at 35−54 GPa, and the $P2_1/c$ structure (containing alternating graphene-like layers of Ca and Li, pierced by 1D-chains of Li_2 pairs, Figure 9) is stable at 54−105 GPa. The simulated diffraction patterns and T_C of these structures match the experimental data.[30] The $P2_1/c$ structure has a unique combination of 2D (graphene sheets) and 1D (chains of paired Li atoms) features; the pairing in the chain of Li atoms is superficially similar to the textbook case of pairing of hydrogen atoms in a 1D chain but is related not to the formation of covalently bonded molecules but to the exclusionary effect of atomic cores.[15] Figure 9 shows electron depletion within Li_2 pairs and charge accumulation between them.

Clearly, traditional rules of chemistry (including the Periodic Law) are violated under pressure, when the PV term in the free energy exceeds the energy of the bonds and, by the virial theorem, the kinetic energy T of the electrons outweighs the potential energy U:

$$\langle T \rangle = -\frac{1}{2}\langle U \rangle + \frac{3}{2}PV \qquad (8)$$

This would normally imply electronic delocalization (tendency toward the free electron gas) at high pressure, but this tendency can be thwarted by the exclusionary effect of atomic cores, as we have seen for sodium. Thus, ultrahigh-pressure chemistry may be less about atomic and molecular orbitals but more about the interplay of the tendency toward the free electron gas and complicating effects of atomic cores. One expects stability of "weird" stoichiometries that do not obey standard rules of chemistry (e.g., the octet rule) but rather Hume−Rothery rule of physics, which optimizes the interaction between the Fermi surface and the Brillouin zone. Which regime (physical or chemical) is taken, depends on how localized the electrons are. It has been predicted[32] that in the Li-H system above 100 GPa there will be stable compounds LiH_8, LiH_6, and LiH_2, which are as abnormal to the chemist as $NaCl_8$, $NaCl_6$, or $NaCl_2$!

3.2. Superconductors. Superconducting hydrogen-rich hydrides may fulfill the elusive dream of the physicists to create solid metallic hydrogen: precompressed by chemical bonding, hydrides may attain the metallic state at lower pressures than pure hydrogen.[33] Following an intuitive prediction[34] of high-pressure structures of silane SiH_4 (with a remarkably high T_C = 166 K at 202 GPa), Pickard and Needs[35] using random sampling found three high-pressure structures that are more stable than any of the previous proposals.[34] Soon it was found[36−39] that two out of their three predicted ground-state structures are metastable; many similar failings of random sampling are documented, for example, for stannane SnH_4 (see below) and for nitrogen

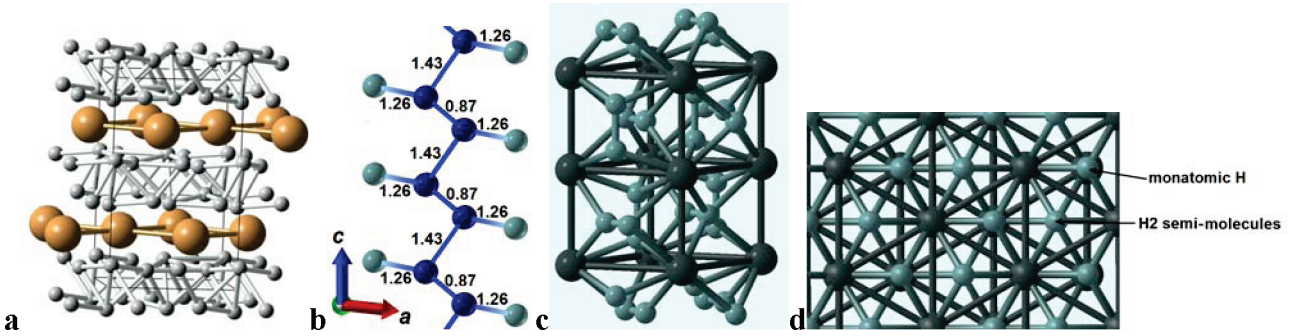


FIGURE 10. Crystal structures of (a) the *Pbcn* phase of silane, (b) a chain of hydrogen atoms in the *C*2/*c* phase of germane, and (c) *Cmcm* and (d) $P6_3/mmc$ phases of stannane. For GeH_4, the two types of hydrogen atoms are shown by different colors.

(where random sampling produced a dynamically unstable $P\bar{4}2_1m$ structure instead of *Iba*2 stable at 188−320 GPa[40]). Metallization of silane is predicted[39] to occur at 220 GPa, with the formation of a layered *Pbcn* structure (Figure 10a) with theoretical[39] $T_C = 16.5$ K.

Interesting structures were predicted for germane[41] GeH_4 and stannane[42] SnH_4. The *C*2/*c* structure of GeH_4, the first one to emerge stable against decomposition, becomes stable at 196 GPa and was predicted[41] to have a remarkably high $T_C = 64$ K at 200 GPa. Stannane has a lower T_C but should[42] become stable against decomposition at much lower pressures (96 GPa) and adopts a *Cmcm* structure at 96−163 GPa and a $P6_3/mmc$ structure above 163 GPa (both phases are superconductors;[41] *Cmcm* has $T_C = 37$ K at 100 GPa, and $P6_3/mmc$ has $T_C = 49$ K at 200 GPa). These are more stable than structures predicted with simulated annealing[43] or random sampling.[44] These superconducting MH_4 phases have M−M and H−H bonds. *Pbcn*-SiH_4 has a layered structure (Figure 10a); the *C*2/*c* phase of GeH_4 has a complex structure with Peierls-distorted H-chains (Figure 10b). The elegant structures of SnH_4 are based on a simple hexagonal packing of Sn atoms in the *Cmcm* phase (with $c/a = 0.84$ at 120 GPa) and hcp packing in the $P6_3/mmc$ phase ($c/a = 1.84$ at 200 GPa). Interestingly, GeH_4 and SnH_4 (and the predicted[32] LiH_2, LiH_6, and LiH_8) contain semimolecular H_2 units (Figure 10b−d) with H−H distances of 0.81 Å (0.87 Å in GeH_4), clearly longer than 0.74 Å in the isolated H_2 molecule. We call such compounds *perhydrides*.[42] At atmospheric pressure, organometallic perhydrides are known,[45] though bonding patterns are quite different. Perhydrides become ubiquitous only under pressure. *Cmcm*-SnH_4 can be represented as $Sn(H_2)_2$, that is, all hydrogens form H_2 units and tin atoms satisfy their valence mostly through Sn−Sn bonds. $P6_3/mmc$-SnH_4 can be represented as $SnH_2(H_2)$, that is, half of the hydrogens are in the monatomic form; because of similar Sn−Sn bonding, we conclude that tin's valence must increase by two; we assign Sn(II) valence state in *Cmcm*-SnH_4 and Sn(IV) in $P6_3/mmc$-SnH_4. Perhydride structures were erroneously taken[44] as a sign of instability to decomposition, while they are in fact stable against decomposition[41,32] and more stable than the *Cccm*-SnH_4 structure predicted[44] by random sampling. Ge, Sn, and Li perhydrides are metallic due to *partial occupancy* of the σ_u^*-bands (made predominantly of H_2 antibonding orbitals) by electrons donated by the metal atom.[32]

3.3. Hunt for Superhard Materials: Boron. Boron is arguably the most mysterious element, the phase diagram of which was clarified only recently,[46] when a new allotropic structure was determined by evolutionary structure prediction. Its experimental[47] Vickers hardness is 50 GPa, making it the hardest boron allotrope and one of the hardest known materials. The calculated stability field of this phase, along with other experimental and theoretical data, allowed us[46] to propose the first phase diagram of boron (Figure 11a). The new phase, called γ-B_{28}, has a peculiar structure made of B_2 pairs and B_{12} icosahedra, whose centers of mass occupy the same positions as ions in the NaCl-type structure. It can be represented as $(B_2)^{\delta+}(B_{12})^{\delta-}$ with charge transfer $\delta \approx 0.5$ electrons from Bader analysis, and 90% of δ is due to the chemical interaction between the atoms rather than geometric asymmetries (Table 1). γ-B_{28} is structurally related to the well-known allotrope α-B_{12}; the main difference is the insertion of B_2 pairs between the icosahedra (Figure 11), which leads to charge separation, affecting physical properties[46] and clearly visible in energy-decomposed electron density distributions. Lowest-energy valence electrons are concentrated on the B_{12} icosahedra, whereas the top of the valence and bottom of the conduction bands correspond to the B_2 pairs. It also has similarities with the structures of boron carbide ("B_4C") and phosphide ($B_{12}P_2$) and can also be viewed as an intermediate structure between the low-pressure semiconducting α-B_{12} (composed entirely of B_{12} icosahedra) and the ultrahigh-pressure metallic α-Ga-type

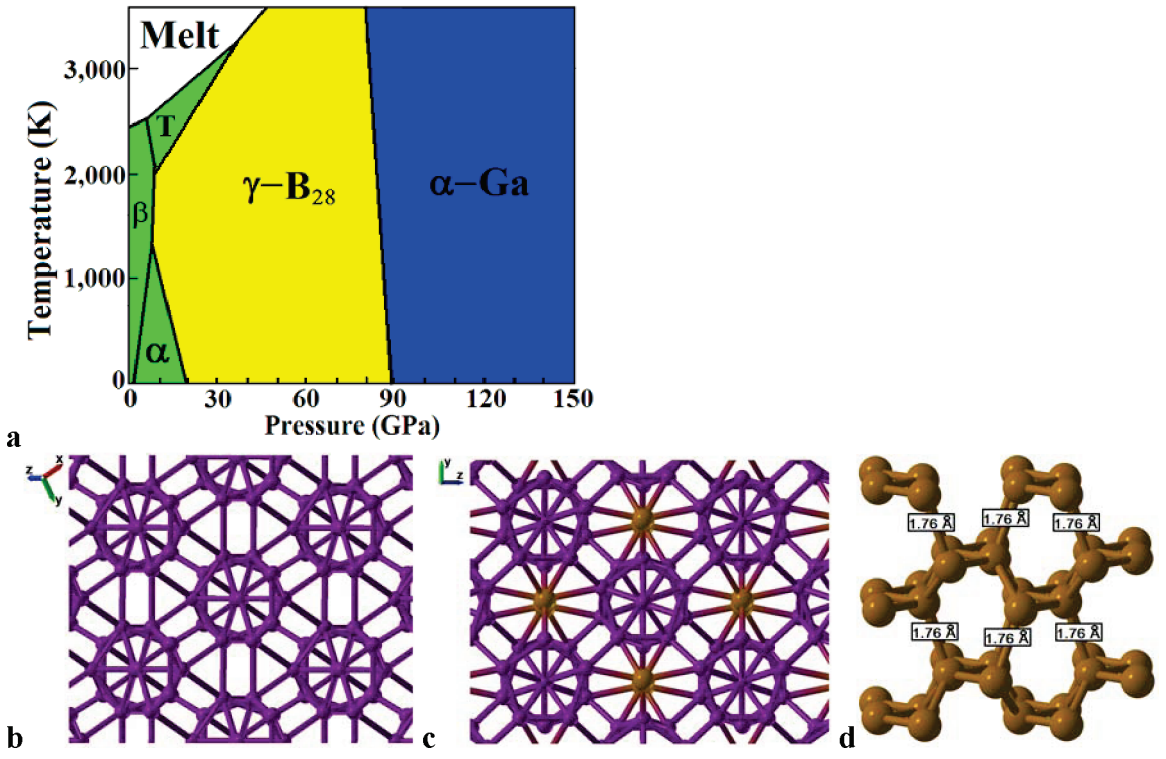


FIGURE 11. Boron:[46] (a) its schematic phase diagram and structures of (b) α-B_{12}, (c) γ-B_{28}, an (d) the α-Ga-type phase of boron.

TABLE 1. Theoretical Structure of γ-B_{28} and Its Atomic Charges[46a]

				Bader charge	
Wyckoff position	*x*	*Y*	*z*	noninteracting	interacting
	space group *Pnnm*; *a* = 5.043 (5.054) Å, *b* = 5.612 (5.620) Å, *c* = 6.921 (6.987) Å				
B1 (4g)	0.1702	0.5206	0	+0.025	+0.2418
B2 (8h)	0.1606	0.2810	0.3743	−0.0153	−0.1680
B3 (8h)	0.3472	0.0924	0.2093	+0.0035	+0.0029
B4 (4g)	0.3520	0.2711	0	−0.0003	+0.0636
B5 (4g)	0.1644	0.0080	0	−0.0011	+0.0255

[a]Experimental unit cell parameters are in parentheses.

phase (consisting entirely of B_2 pairs[48]), see Figure 11. Its stability field is larger than the fields of all other experimentally known phases combined, making it a crucial, and until recently missing, piece in the puzzle of boron's phase diagram.

4. Conclusions

This brief discourse focused on the evolutionary approach to CSP. Some of the other approaches are described elsewhere (ref 3, 6, 50, 52−56) and reviewed in ref 49. An interesting feature of evolutionary algorithms is that they can easily incorporate ideas and features of other methods. It may be promising to incorporate ideas from the field of data mining, which by itself is a successful method for CSP.[50]

We illustrated this methodology by applications to high-pressure chemistry, where many new and exciting phenomena have been discovered. But equally well, it can be applied to the search for new materials at ambient conditions or to the exploration of possible chemical regimes of different compounds. While our focus has been on the most stable structures, the method is also capable of finding a large number of low-energy metastable structures.[1,51]

We should also note that it is possible to extend and apply the evolutionary approach to other related problems, for instance, global optimization of properties other than energy[13] and prediction of low-dimensional structures (surfaces, interfaces, polymers, clusters). Great wealth of new science is to be discovered in those areas.

ARO thanks Intel Corporation, Research Foundation of Stony Brook University, and DARPA (Grant 54751) for funding, his former student C.W. Glass and postdoctoral associate Y. M. Ma

for collaboration, and K. Syassen and L. Lazareva for discussions on alkali metals.

BIOGRAPHICAL INFORMATION

Artem R. Oganov received his Diploma in Crystallography from Moscow State University (Russia), Ph.D. in Crystallography from University College London (U.K.), and Habilitation in Materials Science from ETH Zurich (Switzerland). He is a Full Professor at Stony Brook University. His research interests include development of methods to predict structure and properties of materials, mechanisms of structural transformations, and high-pressure crystallography and planetary sciences. He is an author of 90 scientific publications.

Andriy O. Lyakhov obtained his Diploma in Theoretical Physics from Chernivtsy University (Ukraine), and Ph.D. in Theoretical Physics from the University of Basel (Switzerland). He is currently a Research Fellow at Stony Brook University. His research interests include global optimization methods and crystal structure prediction. He has authored 20 publications in fields of quantum computing and crystal structure prediction.

Mario Valle got an Electronic Engineering Degree from Università di Roma "La Sapienza" (Italy) and worked for DEC and AVS companies. He works at the Data Analysis and Visualization Group of the Swiss National Supercomputing Centre (CSCS). He is interested in data analysis and visualization techniques in very different fields, from crystallography to bioinformatics.

REFERENCES

1 Oganov, A. R.; Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **2006**, *124*, No. 244704.

2 Oganov, A. R.; Valle, M. How to quantify energy landscapes of solids. *J. Chem. Phys.* **2009**, *130*, No. 104504.

3 Freeman, C. M.; Newsam, J. M.; Levine, S. M.; Catlow, C. R. A. Inorganic crystal structure prediction using simplified potentials and experimental unit cells — application to the polymorphs of titanium dioxide. *J. Mater. Chem.* **1993**, *3*, 531–535.

4 Glass, C. W.; Oganov, A. R.; Hansen, N. USPEX — evolutionary crystal structure prediction. *Comput. Phys. Commun.* **2006**, *175*, 713–720.

5 Lyakhov, A. O.; Oganov, A. R.; Valle, M. How to predict large and complex crystal structures. *Comput. Phys. Commun.* **2010**, *181*, 1623–1632.

6 Bush, T. S.; Catlow, C. R. A.; Battle, P. D. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.* **1995**, *5*, 1269–1272.

7 Deaven, D. M.; Ho, K. M. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **1995**, *75*, 288–291.

8 Such cuts were proposed in the context of cluster structure search,[13] where the cut plane went exactly through the center of the cluster. For crystal structures, the "center" of periodic structure is ill-defined, and such 50:50 cuts introduce a bias. Therefore, our cuts are randomly positioned within the unit cell and have random thickness (with contributions of each parent within 25—75% limits).

9 Roth, J.; Henley, C. L. A new binary decagonal Frank-Kasper quasicrystal phase. *Philos. Mag.* **1997**, *A75*, 861–887.

10 Fingerprint function is related to the pair distribution function.[2] Two structures are considered similar if the cosine distance between their fingerprints is <0.01. One alternative is to introduce a penalty function depending on the similarity of structure factors, but this distorts the fitness landscape: Abraham, N. L.; Probert, M. I. J. Improved real-space genetic algorithm for crystal structure and polymorph prediction. *Phys. Rev.* **2008**, *B77*, No. 134117. Another approach considered two structures identical if their radial distribution functions and space groups are exactly the same, but some identities will be missed when numerical errors are present: Yao, Y.; Tse, J. S.; Tanaka, K. Metastable high-pressure single-bonded phases of nitrogen predicted via genetic algorithm. *Phys. Rev.* **2008**, *B77*, No. 052103.

11 Niching has also been successfully done in the context of cluster structure prediction, but without fingerprinting: Hartke, B. Global cluster geometry optimization by a phenotype algorithm with niches: Location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.* **1999**, *20*, 1752–1759.

12 Oganov, A. R.; Ma, Y.; Lyakhov, A. O.; Valle, M.; Gatti, C. Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Rev. Mineral. Geochem.* **2010**, *71*, 271–298.

13 Oganov, A. R.; Lyakhov, A. O. Towards the theory of hardness of materials. *J. Superhard Mater.* **2010**, *32*, 143–147.

14 Trimarchi, G.; Zunger, A. Global space-group optimization problem: Finding the stablest crystal structure without constraints. *Phys. Rev. B75*, No. 104113, **2007**, Erratum: *Phys. Rev. B. 82*, 219903, **2010**.

15 Neaton, J. B.; Ashcroft, N. W. Pairing in dense lithium. *Nature* **1999**, *400*, 141–144.

16 Pickard, C. J.; Needs, R. J. Dense low-coordination phases of lithium. *Phys. Rev. Lett.* **2009**, *102*, No. 146401.

17 Ma, Y.; Eremets, M. I.; Oganov, A. R.; Xie, Y.; Trojan, I.; Medvedev, S.; Lyakhov, A. O.; Valle, M.; Prakapenka, V. Transparent dense sodium. *Nature* **2009**, *458*, 182–185.

18 Gregoryanz, E.; Lundegaard, L. F.; McMahon, M. I.; Guillaume, C.; Nelmes, R. J.; Mezouar, M. Structural diversity of sodium. *Science* **2008**, *320*, 1054–1057.

19 This is most likely due to shortcomings of the approximate density functionals, which overstabilize metallic phase; hence, the insulating hP4 phase was artificially pushed to higher pressures than in experiment. The whole sequence is bcc → (65 GPa) → fcc → (103 GPa) → cI16 → (118 GPa) → oP8 → (118 GPa) → tI19 → (200 GPa) → hP4, the most noteworthy phases being tI19 (incommensurate host—guest tetragonal structure with 19.3 atoms in the unit cell) and insulating hP4. The tI19 structure was shown to be a quasi-1D metal with conductivity along the chains of guest atoms: Lazicki, A.; Goncharov, A. F.; Struzhkin, V. V.; Cohen, R. E.; Liu, Z.; Gregoryanz, E.; Guillaume, C.; Mao, H. K.; Hemley, R. J. Anomalous optical and electronic properties of dense sodium. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6525–6528.

20 Leger, J. M.; Haines, J.; Atouf, A.; Schulte, O.; Hull, S. High-pressure X-ray diffraction and neutron diffraction studies of BaF_2 — an example of a coordination number of 11 in AX_2 compounds. *Phys. Rev.* **1995**, *B52*, 13247–13256.

21 Kinoshita, K.; Nishimura, M.; Akahama, Y.; Kawamura, H. Pressure-induced phase transition of BaH_2: post-Ni_2In phase. *Solid State Commun.* **2007**, *141*, 69–72.

22 Electrides known at ambient pressure usually have unpaired interstitial electons: Dye, J. L. Electrides: from 1D Heisenberg chains to 2D pseudo-metals. *Inorg. Chem.* **1997**, *36*, 3816–3826. In high-pressure electrides, they are paired. The situation is analogous to the high spin—low spin transition seen under pressure in transition metal compounds. The electrons are unpaired at low pressure due to the Hundt rules, but pairing wins at high pressure because it enables greater density.

23 Matsuoka, T.; Sjimizu, K. Direct observation of a pressure-induced metal-to-semiconductor transition in lithium. *Nature* **2009**, *458*, 186–189.

24 Marques, M.; Ackland, G. J.; Lundegaard, L. F.; Stinton, G.; Nelmes, R. J.; McMahon, M. I.; Contreras-Garcia, J. Potassium under pressure: A pseudobinary ionic compound. *Phys. Rev. Lett.* **2009**, *103*, No. 115501.

25 Khan, M. A.; Connerade, J.-P.; Rafique, M. Orbital contraction in Ca and its effect on collisions. *J. Phys. B: At. Mol. Opt. Phys.* **1994**, *27*, L563–L569.

26 Oganov, A. R.; Ma, Y. M.; Xu, Y.; Errea, I.; Bergara, A.; Lyakhov, A. O. Exotic behavior and crystal structures of calcium under pressure. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7646–7651.

27 Yabuuchi, T.; Matsuoka, T.; Nakamoto, Y.; Shimizu, K. Superconductivity of Ca exceeding 25 K at megabar pressures. *J. Phys. Soc. Jpn.* **2006**, *75*, No. 083703.

28 Xie, Y.; Oganov, A. R.; Ma, Y. Novel structures and high pressure superconductivity of $CaLi_2$. *Phys. Rev. Lett.* **2010**, *104*, No. 177005.

29 Feng, J.; Ashcroft, N. W.; Hoffmann, R. Theoretical indications of singular structural and electronic features of Laves-phase $CaLi_2$ under pressure. *Phys. Rev. Lett.* **2007**, *98*, No. 247002.

30 Debessai, M.; Matsuoka, T.; Hamlin, J. J..; et al. Superconductivity under high pressure in the binary compound $CaLi_2$. *Phys. Rev.* **2008**, *B78*, No. 214517.

31 Tse, J. S.; Klug, D. D.; Desgreniers, S.; Smith, J. S.; Dutrisac, R. Instability of $CaLi_2$ at high pressures: theoretical prediction and experimental results. *EPL* **2009**, *86*, 56001.

32 Zurek, E.; Hoffmann, R.; Ashcroft, N. W.; Oganov, A. R.; Lyakhov, A. O. A little bit of lithium does a lot for hydrogen. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17640–17643.

33 Ashcroft, N. W. Hydrogen dominant metallic alloys: High temperature superconductors? *Phys. Rev. Lett.* **2004**, *92*, No. 187002.

34 Feng, J; Grochala, W; Jaron, T; Hoffmann, R; Bergara, A; Ashcroft, N. W. Structures and potential superconductivity in SiH_4 at high pressure: En route to "metallic hydrogen". *Phys. Rev. Lett.* **2006**, *96*, No. 017006.

35 Pickard, C. J.; Needs, R. J. High-pressure phases of silane. *Phys. Rev. Lett.* **2006**, *97*, No. 045504.

36 Degtyareva, O.; Martinez-Canales, M.; Bergara, A.; Chen, X. J.; Song, Y.; Struzhkin, V. V.; Mao, H. K.; Hemley, R. J. Crystal structure of SiH_4 at high pressure. *Phys. Rev.* **2007**, *B76*, No. 064123.

37 Chen, X. J.; Wang, J. L.; Struzhkin, V. V.; Mao, H. K.; Hemley, R. J.; Lin, H. Q. Superconducting behavior in compressed solid SiH_4 with a layered structure. *Phys. Rev. Lett.* **2008**, *101*, No. 077002.

38 Kim, D. Y.; Scheicher, R. H.; Lebegue, S.; Prasongkit, J.; Arnaud, B.; Alouani, M.; Ahuja, R. Crystal structure of the pressure-induced metallic phase of SiH_4 from ab initio theory. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16454–16459.

39 Martinez-Canales, M.; Oganov, A. R.; Lyakhov, A.; Ma, Y.; Bergara, A. Novel structures of silane under pressure. *Phys. Rev. Lett.* **2009**, *102*, No. 087005.

40 Ma, Y.; Oganov, A. R.; Xie, Y.; Li, Z.; Kotakoski, J. Novel high pressure structures of polymeric nitrogen. *Phys. Rev. Lett.* **2009**, *102*, No. 065501.

41 Gao, G.; Oganov, A. R.; Bergara, A.; Martinez-Canalez, M.; Cui, T.; Iitaka, T.; Ma, Y.; Zou, G. Superconducting high pressure phase of germane. *Phys. Rev. Lett.* **2008**, *101*, No. 107002.

42 Gao, G.; Oganov, A. R.; Li, Z.; Li, P.; Cui, T.; Bergara, A.; Ma, Y.; Iitaka, T.; Zou, G. Crystal structures and superconductivity of stannane under high pressure. *Proc. Natl. Acad. Sci. U. S.A.* **2010**, *107*, 1317–1320.

43 Tse, J. S.; Yao, Y.; Tanaka, K. Novel superconductivity in metallic SnH_4 under high pressure. *Phys. Rev. Lett.* **2007**, *98*, No. 117004.

44 Pickard, C. J.; Needs, R. J. Structures at high pressure from random searching. *Phys. Status Solidi* **2008**, *B246*, 536–540.

45 Kubas, G. J.; Ryan, R. R.; Swanson, B. I.; Vergamini, P. J.; Wasserman, H. J. Characterization of the first examples of isolable molecular hydrogen complexes, $M(CO)_3(PR_3)_2(H_2)$ (M = Mo, W; R = Cy, i-Pr). Evidence for a side-on bonded H_2 ligand. *J. Am. Chem. Soc.* **1984**, *106*, 452–454.

46 Oganov, A. R.; Chen, J.; Gatti, C.; Ma, Y.-Z.; Ma, Y.-M.; Glass, C. W.; Liu, Z.; Yu, T.; Kurakevych, O. O.; Solozhenko, V. L. Ionic high-pressure form of elemental boron. *Nature* **2009**, *457*, 863–867.

47 Solozhenko, V. L.; Kurakevych, O. O.; Oganov, A. R. On the hardness of a new boron phase, orthorhombic γ-B_{28}. *J. Superhard Mater.* **2008**, *30*, 428–429.

48 The α-Ga structure type can be considered identical to that of I_2 and black phosphorus. All of these contain dimeric units, but in metallic α-Ga-type materials "dimerization" is weaker: the computed average intra- and interdimeric bond lengths in this phase of boron are 1.76 and 1.86 Å, respectively (in α-Ga, the corresponding theoretical values are 2.55 and 2.77 Å). Metallicity of the α-Ga-type phase of boron implies that B_2 dimers can easily give away their electrons; here they lose the electrons to the conduction band, whereas in γ-B_{28}, they donate electrons to the B_{12} icosahedra.

49 Oganov, A. R., Ed. *Modern Methods of Crystal Structure Prediction*; Wiley-VCH: 2010; ISBN: 978-3-527-40939-6.

50 Fischer, C. C.; Tibbets, K. J.; Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nature Materials* **2006**, *5*, 641–646.

51 Li, Q.; Ma, Y.; Oganov, A. R.; Wang, H. B.; Wang, H.; Xu, Y.; Cui, T.; Mao, H.-K.; Zou, G. Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.* **2009**, *102*, No. 175506.

52 Pannetier, J.; Bassasalsina, J.; Rodriguez-Carvajal, J.; Caignaert, V. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* **1990**, *346*, 343–345.

53 Schon, J. C.; Jansen, M. First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimisation. *Angew. Chem.-Int. Ed.* **1996**, *35*, 1287–1304.

54 Martoňák, R.; Laio, A.; Parrinello, M. Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys. Rev. Lett.* **2003**, *90*, 075503.

55 Wales, D. J.; Doye, J. P. K Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem.* **1997**, *A101*, 5111–5116.

56 Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.