

Cite this: DOI: 10.1039/c2ce06642d

www.rsc.org/crystengcomm

PAPER

# Evolutionary metadynamics: a novel method to predict crystal structures†

Qiang Zhu,<sup>\*a</sup> Artem R. Oganov<sup>ab</sup> and Andriy O. Lyakhov<sup>a</sup>

Received 7th December 2011, Accepted 8th March 2012

DOI: 10.1039/c2ce06642d

A novel method for crystal structure prediction, based on metadynamics and evolutionary algorithms, is presented here. This technique can be used to produce efficiently both the ground state and metastable states easily reachable from a reasonable initial structure. We use the cell shape as collective variable and evolutionary variation operators developed in the context of the USPEX method [Oganov, Glass, *J. Chem. Phys.*, 2006, **124**, 244704; Lyakhov *et al.*, *Comp. Phys. Comm.*, 2010, **181**, 1623; Oganov *et al.*, *Acc. Chem. Res.*, 2011, **44**, 227] to equilibrate the system as a function of the collective variables. We illustrate how this concept helps one to find stable and metastable states for Al<sub>2</sub>SiO<sub>5</sub>, SiO<sub>2</sub>, MgSiO<sub>3</sub>, and carbon. Apart from predicting crystal structures, the new method can also provide insight into the mechanisms of phase transitions.

## 1 Introduction

Crystal structure prediction (CSP) has long been a major fundamental challenge in physical sciences.<sup>1</sup> As the stable structure corresponds to the global minimum of the free energy surface (FES), several global optimization algorithms have been devised and applied.<sup>2–7</sup> Due to these efforts, in many cases it is now possible to predict the stable structure of a given inorganic compound at arbitrary external pressure. In addition, there is progress towards optimizing not only free energy, but also other properties (*e.g.*, hardness,<sup>8</sup> density,<sup>9</sup> *etc.*).

In general, there are two types of strategies for predicting crystal structures. One is to directly scan the whole energy landscape, and find the most stable crystal structures using random sampling<sup>7</sup> or generally more efficient evolutionary algorithms.<sup>6</sup> Alternatively, one could also use some known structures as the starting point, and predict the new structures by crossing the energy barriers, until the lowest energy structure is found.<sup>2–5</sup> The latter group of methods can be described as neighborhood search methods and some of them - in particular, metadynamics,<sup>3,10</sup> can be very efficient, but rely on availability of a good starting structure.

Metadynamics explores the properties of the multidimensional FES of complex many-body systems by means of a coarse-grained non-Markovian dynamics in the space defined by a few collective coordinates. By introducing a history-dependent potential term, it fills the minima in the FES and allows efficient exploration of the FES as a function of the collective

coordinates.<sup>10</sup> The technique is usually applied as an extension of the molecular dynamics (MD) simulation technique. Martoňák *et al.* used the edges of the simulation cell as collective variables for the study of pressure-induced structural transformations.<sup>3</sup> The method proved to be much more powerful in predicting crystal structure transformations,<sup>11–13</sup> compared with the normal MD approach. However, at each metastep it uses MD for equilibrating the system and MD is not always an efficient method for equilibration, which leads to trapping in metastable states and often amorphization rather than transition to a stable crystal structure. This motivated us to develop an alternative strategy.

In this paper, we present a method combining the features of both strategies for predicting crystal structures, basically a metadynamics-like method driven not by local MD sampling, but by efficient global optimization moves.<sup>6,14</sup> Following Martoňák *et al.*,<sup>3</sup> we adopt the cell edges as the collective variables, and equilibrate the system at each value of the collective variables using moves inspired by the evolutionary variation operators,<sup>14</sup> rather than previously used MD simulation. We find that this approach is very efficient for predicting stable and low-energy metastable states and avoids amorphization.

## 2 Methodology

MD method is widely used to study physical processes in liquids and solids,<sup>15</sup> but has difficulties in escaping from local energy minima and crossing high energy barriers.<sup>16</sup> Metadynamics<sup>10</sup> is an ingenious way to solve this problem and accelerate the activated processes involving barrier crossing, including first-order reconstructive transitions. In metadynamics, one has to specify collective variables that adequately describe the evolution of the system, and then perturb the FES defined in this reduced space of collective variables. Martoňák *et al.* used the cell box  $h$  ( $3 \times 3$  matrix) as a collective variable to distinguish different

<sup>a</sup>Department of Geosciences, Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA. E-mail: qiang.zhu@stonybrook.edu; Fax: +1631 632 8240; Tel: +1631 632 1449

<sup>b</sup>Geology Department, Moscow State University, Moscow, Russia

† Electronic Supplementary Information (ESI) available. See DOI: 10.1039/c2ce06642d

structures.<sup>3</sup> For a given system with volume  $V$  under external pressure  $P$ , the derivative of the free energy with respect to  $h$  is

$$-\frac{\partial G}{\partial h_{ij}} = V[h^{-1}(P-p)]_{ji} \quad (1)$$

where  $p$  is the internal pressure tensor calculated for each given geometry.

The collective variable evolves with a stepping parameter  $\delta h$

$$h(t+1) = h(t) + \delta h \frac{f(t)}{|f(t)|} \quad (2)$$

the driving force  $f = -\frac{\partial G}{\partial h}$  is from a history-dependent Gibbs potential  $G(t)$  where a Gaussian has been added to  $G^h$  at every point  $h(t')$  already visited in order to discourage it from being visited again,

$$G(t) = G^h + \sum_i W e^{-\frac{|h-h(t')|^2}{2\delta h^2}} \quad (3)$$

As the simulation proceeds, the history-dependent term fills the initial well of the FES, and concurrently the collective variables (the cell) undergo a sequence of changes, at each of which the atoms are re-equilibrated. At some critical cell distortion atoms rearrange dramatically, yielding a new crystal structure.

By adding a history-dependent Gaussian potential, metadynamics efficiently reaches and crosses the transition state, thus solving an intrinsic problem of MD simulations. However, it still relies on the ability of MD to equilibrate the system at each value of collective variables. Metadynamics encounters two general problems. First, MD is not a very efficient technique for equilibration involving crossing energy barriers, especially at low temperatures. Second, metadynamics reduces the full energy landscape to a low-dimensional projection, which is not always adequate.

For a crystal with  $N$  atoms in the unit cell, the number of degrees of freedom is  $3N + 3$ . Metadynamics assumes that the system can be described by 6 collective variables representing the box  $h$ . The remaining  $3N-3$  dimensions describe atomic positions, and we need to determine the global energy minimum with respect to these dimensions at given  $h$ . This can be done using MD (as in standard metadynamics), but as discussed above, this encounters certain problems. Here we do it using global optimization techniques based on evolutionary algorithms.<sup>6,14,17</sup> These algorithms use several variation operators - *i.e.* recipes for obtaining new structures from the previously sampled ones - for sampling the energy landscape. Here we use the softmutation operator,<sup>14</sup> which connects global optimization with lattice dynamics and group theory. Indeed, the  $3N-3$  variables describing atomic positions can be transformed into a set of normal modes that possess valuable properties. If a structure is not dynamically stable, a more stable structure is obtained by following the eigenvector of the softest vibrational mode. For structures without soft modes, there is a statistically valid Bell-Evans-Polanyi principle that states that low-energy structures are usually connected by low activation barriers.<sup>18</sup> Low barriers are, in turn, usually related to the direction of the lowest curvature of the FES - or eigenvectors of the softest vibrational mode. This is the principle behind the minima hopping method<sup>5</sup> and explains the efficiency of the softmutation operator

introduced in the evolutionary algorithm USPEX.<sup>14</sup> To calculate the vibrational modes, we construct the dynamical matrix from bond hardness coefficients:

$$D_{\alpha\beta}(a,b) = \sum_m \frac{\partial^2}{\partial \alpha_a^0 \partial \beta_b^m} \left( \frac{1}{2} \sum_{i,j,l,n} H_{ij}^{l,n} \left[ (x_i^l - x_j^n)^2 + (y_i^l - y_j^n)^2 + (z_i^l - z_j^n)^2 \right] \right) \quad (4)$$

Here  $\alpha, \beta$  denote coordinates  $(x,y,z)$ ;  $a,b,i,j$  describe the atom in the unit cell;  $l,m,n$  denote the unit cell number. Therefore  $x_i^l$  is the  $x$ -coordinate of atom  $i$  in the unit cell  $l$ .  $H_{ij}^{l,n}$  are the hardness coefficients for a bond between the atom  $i$  in the unit cell  $l$  and atom  $j$  in the unit cell  $n$ . The bond hardness coefficients are computed from bond distances, and tabulated covalent radii and electronegativities of the atoms.<sup>14,19</sup> Note that the dynamical matrix in Eq. (4) does not include atomic masses (to obtain pure low-curvature directions of the FES) and corresponds to zero wavevector (*i.e.*, as in original metadynamics (ref. 20), we keep the number of atoms in the supercell. Extensions allowing variable number of atoms in the supercell are straightforward).

To perform softmutation, we move the atoms along the eigenvector of the softest calculated mode. One structure can be softmutated many times using different non-degenerate modes and displacements. By properties of normal modes, the original and softmutated structures are linked by group-subgroup symmetry relations, but structure relaxation may result in a symmetry increase. In this case one observes a structural transition with a common subgroup. The magnitude of the displacement ( $d_{\max}$ ) along the mode eigenvector is an input parameter: with relatively small  $d_{\max}$  and displacements represented by a random linear mixture of all mode eigenvectors, we obtain a method similar to MD-metadynamics in its ability to cross energy barriers and equilibrate the system. With large  $d_{\max}$  along the softest mode eigenvectors, we obtain the softmutation operator,<sup>14</sup> capable of efficiently finding the global energy minimum. When needed, other evolutionary variation operators can be used - such as permutation (swaps of atomic identities) or heredity (combination of pieces of two parent structures).

Our algorithm as shown in Fig. 1 is in many ways similar to the original version by Martoňák *et al.*<sup>20</sup> We start from one known initial structure at a given external pressure  $P$ , and then compute its vibrational modes, which are used to produce typically 20–40

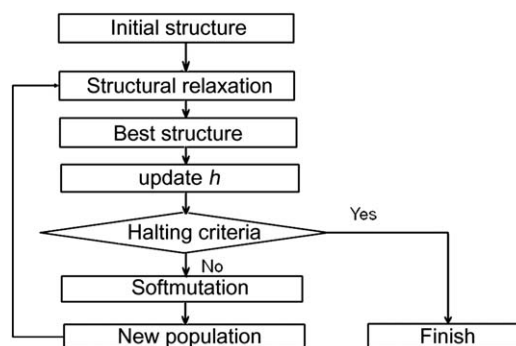


Fig. 1 Illustration of the evolutionary metadynamics algorithm.

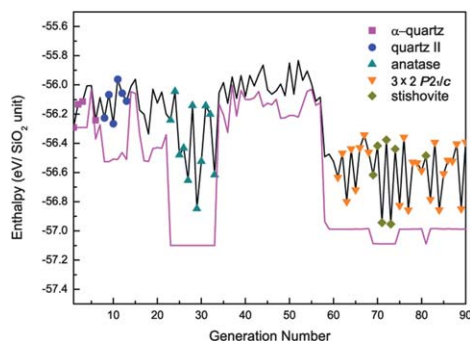
softmutated structures. These are relaxed at constant  $h$ , the lowest-enthalpy structure is selected and its pressure tensor  $p$  is computed. Its box  $h$  is then updated according to Eq. (2), and a new generation of softmutated structures are produced and relaxed in the fixed cell  $h$ . Repeated for a number of generations, this computational scheme leads to a series of structural transitions and is stopped when the maximum number of generations is reached. Due to the presence of a population of structures and a selection step, this algorithm is evolutionary, unlike original metadynamics.<sup>20</sup>

In addition to  $d_{\max}$ , there are two other important parameters - the Gaussian height ( $W$ ) and Gaussian width ( $\delta h$ ), for choosing which Martoňák *et al.*<sup>20</sup> proposed a recipe based on the curvature of the FES at the minimum. According to our experience, a good starting point is to set  $\delta h$  as one tenth of the minimum lattice vector, and  $W$  as around 8 000  $\delta h^2$  (with the units of kbar  $\text{\AA}^3$ ).

Note that Eq. (2) is not invariant with respect to the choice of the unit cell (or supercell) and works best for cells close to cubic shape. To remedy this, we developed a more sophisticated equation based on elasticity theory:

$$h_{im}(t+1) = h_{im}(t) + \frac{\delta h}{|f|V^{1/3}} S_{ijk} f_{ki} h_{jm}(t) \quad (5)$$

Here we use the elastic compliance tensor ( $S$ ) corresponding to an elastically isotropic medium with Poisson ratio 0.26, which corresponds to the border between brittle and ductile materials<sup>21</sup> and is a good average value to describe both metals and insulators. The choice of the Poisson ratio does not significantly affect the results; the main effect of Eq. (5) is to make the results independent of the simulation cell shape. The Young's modulus is chosen in such a way that for a cubic cell under uniform stress the original Eq. (2) is reproduced. We have implemented and tested the formalism based on Eq. (5) and found it to work extremely well, the efficiency improves compared to Eq. (2) (results not shown here).



**Fig. 2** Enthalpy evolution during the compression of a 72-atom supercell of  $\alpha$ -quartz ( $\text{SiO}_2$ ) at 10 GPa (black line: enthalpies for best structures as a function of  $h$ ; magenta line: enthalpies for best structures after full relaxation). In this calculation, we set Gaussian parameters  $W = 10\,000$  kbar  $\text{\AA}^3$ ,  $\delta h = 1.0$   $\text{\AA}$ , and  $d_{\max} = 3.0$   $\text{\AA}$ . Each generation contains 40 structures. Structures were relaxed using the GULP code<sup>22</sup> with the BKS potential.<sup>23</sup> We first observed the transition from  $\alpha$ -quartz (space group  $P3_121$ ) to quartz II (C2) in the 9th generation, and then quartz II amorphized until it transformed into the anatase structure ( $I4_1/amd$ ) in the 22th generation. Anatase amorphized again, and evolved into the  $3 \times 2 P2_1/c$  structure in the 58th generation, and then transformed into stishovite ( $P4_2/mmm$ ) at the 69th generation.

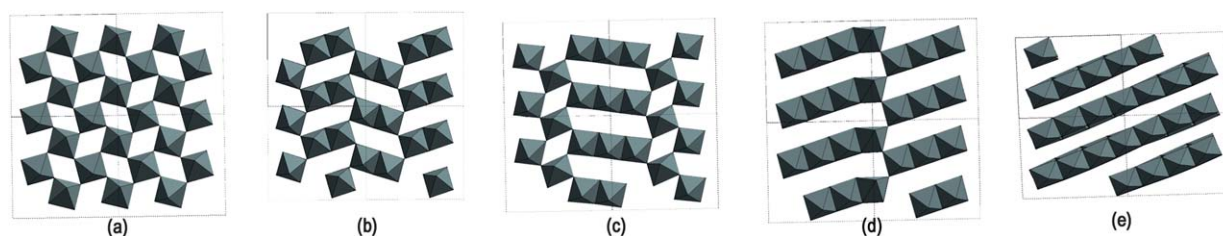
Below we discuss results obtained with Eq. (2) and Eq. (3), *i.e.* the original and simplest formulation of metadynamics<sup>20</sup> with isotropic Gaussians. In reality, the FES minima are anisotropic - the local FES curvature is lower for shear and higher for compressional distortions of the cell. It was found that the "isotropic" formalism based on Eq. (3) coupled with MD equilibration is incapable of predicting structural transformations of silica and gets stuck in amorphization.<sup>12</sup> This problem was remedied by the anisotropic extension of Eq. (2) and Eq. (3).<sup>12</sup> Here we find that this increase of complexity can be avoided and the same structural transformations are easily predicted with the isotropic formulation of metadynamics, if softmutation is used instead of MD for equilibration (see Fig. 2). We also checked that this approach also correctly reproduces the previous results on pressure-induced transitions in  $\text{MgSiO}_3$ <sup>11</sup> (see Fig. 3). After these successful tests, we applied this approach to two challenging and important problems, namely the pressure-induced transformations of elemental carbon and  $\text{Al}_2\text{SiO}_5$ . Modified formalism based on Eq. (5) gives very similar results, but is invariant to the choice of supercell and more efficient.

### 3 Results and discussion

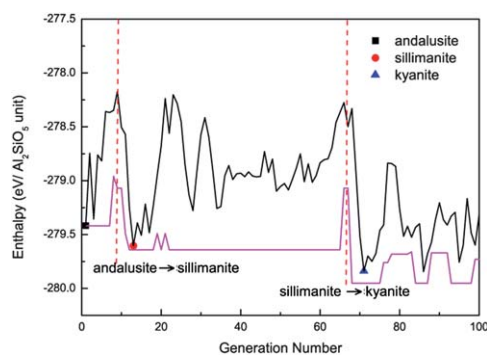
#### 3.1 Finding stable and metastable energy minima: example of $\text{Al}_2\text{SiO}_5$

The phase diagram of  $\text{Al}_2\text{SiO}_5$  is important for Earth sciences and has attracted great interest. Three known  $\text{Al}_2\text{SiO}_5$  polymorphs (kyanite, andalusite, sillimanite) are common minerals in the Earth's crust and upper mantle. In all these structures, Si atoms are tetrahedrally coordinated, while half of the Al atoms are octahedrally coordinated. The Al octahedra form the  $-\text{Al}-\text{Al}-$  chains, and the remaining Al and Si alternate in neighboring chains ( $-\text{Si}-\text{Al}-$ ). The coordination of Al in the  $-\text{Si}-\text{Al}-$  chains is either tetrahedral (sillimanite), fivefold (andalusite), or octahedral (kyanite).<sup>25</sup> It has to be noted that the prediction of these structures is an extremely challenging test for any global optimization method (These complex structures have low symmetry and relatively large primitive cells with 32 atoms. To illustrate the difficulty of finding the ground state, we generated 10,000 random structures and relaxed them at 10 GPa, and found that none of these structures correspond to the stable phase, kyanite). Energy barriers between these structures are very high and these phases can exist metastably, and even coexist, in nature for millions of years - making direct MD sampling (which covers timescales up to  $\mu\text{s}$ ) of these structural transitions clearly impossible. Impressively, an evolutionary metadynamics simulating starting from the low-pressure polymorph, andalusite, has successfully found the other two structures.

These simulations were carried out by using a classical potential<sup>26</sup> and the GULP code.<sup>22</sup> We performed the calculation with  $d_{\max} = 3.0$   $\text{\AA}$ ,  $W = 2\,500$  kbar  $\text{\AA}^3$  and  $\delta h = 0.4$   $\text{\AA}$ . Each generation contains 30 structures. Starting from andalusite (32 atoms per box), as shown in Fig. 4, in the 8th generation we observed breaking of an interchain Al-O bond in  $\text{AlO}_5$  polyhedra, and the structure transformed into sillimanite containing  $\text{AlO}_4$  tetrahedra. Sillimanite survived until the 68th generation, when the inter-chain Al-O bonds formed again, increasing coordination of Al from fourfold to sixfold and eventually



**Fig. 3** Phases observed in evolutionary metadynamics starting from a 160-atom supercell of perovskite ( $\text{MgSiO}_3$ ) at 250 GPa. (a) Perovskite (space group  $Pbnm$ ); (b)  $2 \times 2$  phase ( $Pbnm$ ); (c)  $3 \times 1$  phase ( $P2_1/m$ ); (d)  $4 \times 4$  phase ( $Pm$ ); (e) post-perovskite ( $Cmcm$ ). Only Si octahedra are shown (Mg atoms are omitted for clarity). In this calculation, we set Gaussian parameters  $W = 7000 \text{ kbar } \text{\AA}^3$ ,  $\delta h = 0.8 \text{ \AA}$ , and  $d_{\text{max}} = 3.0 \text{ \AA}$ . Each generation contains 40 structures. Structures were relaxed using the GULP code<sup>22</sup> with a partially ionic Buckingham potential.<sup>24</sup> Our calculation found the transition from perovskite to post-perovskite, and structures b, c, d were also observed in the simulation. Plane sliding with the formation of stacking faults is a possible pathway for this phase transition, in accordance with the previous metadynamics study.<sup>12</sup>



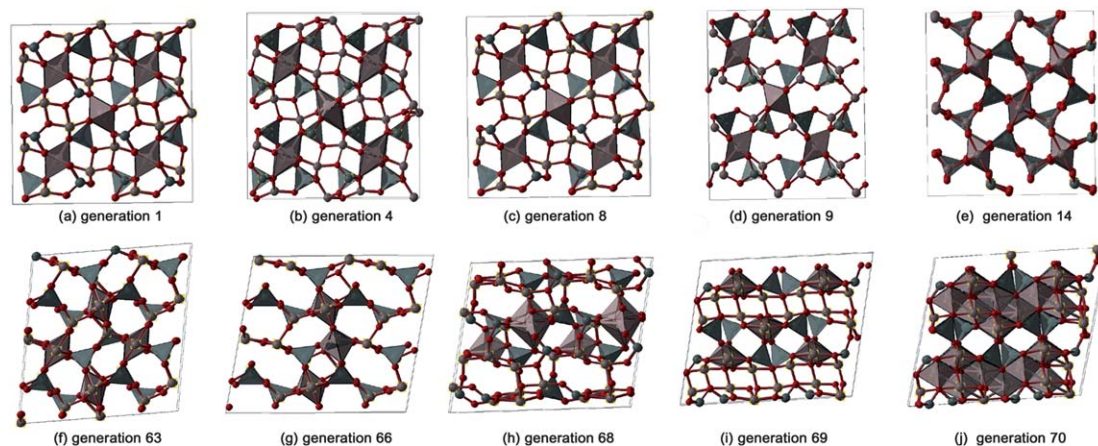
**Fig. 4** Enthalpy evolution during the compression on andalusite ( $\text{Al}_2\text{SiO}_5$ ) at 10 GPa (black line: enthalpies for best structures with constant  $h$ ; magenta line: enthalpies for best structures after full relaxation).

creating kyanite phase with all Al atoms in the  $\text{AlO}_6$  octahedra. The whole picture, as shown in Fig. 5, proves that our method was easily able to predict the transitions: andalusite  $\rightarrow$  sillimanite  $\rightarrow$  kyanite. As a bonus in addition to finding the global minimum structure, the simulation unravels a very non-trivial relationship between the structures (*e.g.* sillimanite–kyanite) and a crystallographic model of their transformations. For a reconstructive phase transition, one should expect a nucleation-and-

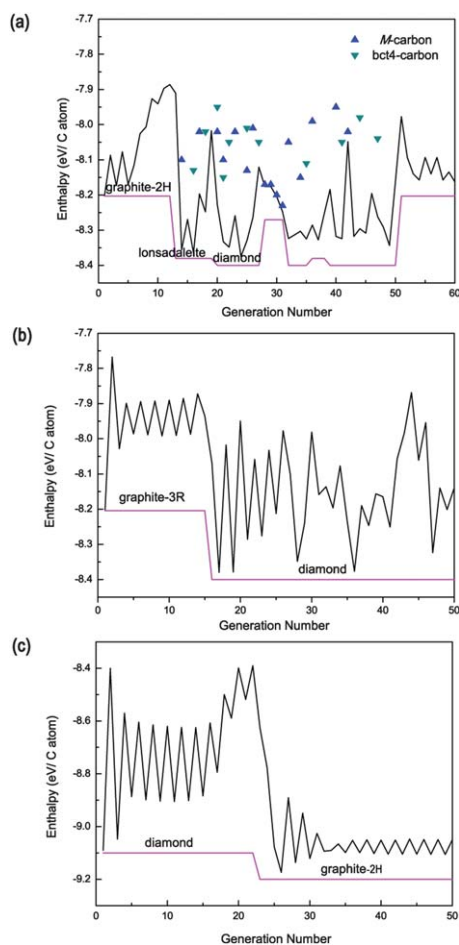
growth mechanism, rather than a concerted crystallographic mechanism. However, a concerted mechanism provides not only a useful simplified view of the real transition, but also input for mean-field theories of phase transformations, and for techniques (such as the Transition Path Sampling (TPS) (*e.g.* ref. 26)) that are capable of simulating nucleation and growth phenomena but require a reasonable initial mechanism.

### 3.2 Searching for low-energy metastable phases: elemental carbon

Carbon is unique in that it adopts a wide range of structures, from superhard insulating (diamond, lonsdaleite, and new allotrope<sup>28</sup>) to ultrasoft semi-metallic (graphite, fullerenes) and even superconducting (doped diamond<sup>29</sup> and alkali-doped fullerenes<sup>30</sup>). The number of possible metastable phases is infinite. By compressing graphite at room temperature to 15–20 GPa, a metastable transparent superhard phase was phase observed, but its structure remained uncertain.<sup>28</sup> Several structural models were proposed.<sup>31,32</sup> The correct model should have the lowest barrier of formation from graphite at 15–20 GPa. To determine the barrier, it is necessary to study the transition pathways from graphite to all candidate structures. The technique discussed here is capable of finding several (if not all) relevant candidates in one single simulation.



**Fig. 5** Phases observed during compressing andalusite ( $\text{Al}_2\text{SiO}_5$ ) at 10 GPa. (a) generation 1 (andalusite); (b) generation 4; (c) generation 8; (d) generation 9 (sillimanite); (e) generation 14; (f) generation 63; (g) generation 66; (h) generation 68; (i) generation 69; (j) generation 70 (kyanite).



**Fig. 6** Enthalpy evolution starting from (a) graphite-2H at 20 GPa; (b) graphite-3R at 20 GPa; (c) diamond at 0 GPa (black line: enthalpies for best structures with constant  $h$ ; magenta line: enthalpies for best structures after full relaxation).

We started the simulation at 20 GPa from the graphite-2H structure with 32 atoms per box and using  $d_{\max} = 2.5 \text{ \AA}$ ,  $W = 7000 \text{ kbar \AA}^3$  and  $\delta h = 0.7 \text{ \AA}$ . Each generation contained 25 structures. Structure relaxations were done using density functional theory (DFT) within the generalized gradient approximation (GGA)<sup>33</sup> together with the all-electron projector augmented wave (PAW)<sup>34</sup> method as implemented in the VASP code<sup>35</sup>. We used the plane wave kinetic energy cutoff of 520 eV and the Brillouin zone sampling resolution of  $2\pi \times 0.08 \text{ \AA}^{-1}$ , which showed excellent convergence of the energy differences,

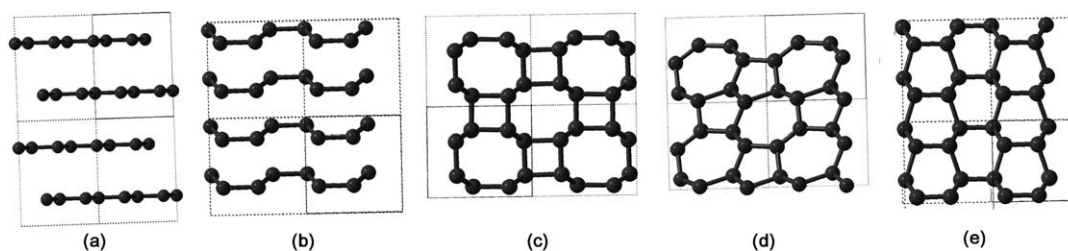
stress tensors and structural parameters. Fig. 6a shows the results. In the 13th generation, graphene layers buckled as shown in Fig. 7b, which led to the formation of 3D networks of  $sp^3$ -hybridized carbon atoms. We observed, in a single simulation, the formation of lonsdaleite with 6-membered rings in the 13th generation (Fig. 7e),  $M$ -carbon containing 5- and 7-membered rings in the 14th (Fig. 7d), and bct- $C_4$  structure with 4- and 8-membered rings (Fig. 7c) in the 16th generation. In the 20th generation, global minimum, diamond, was found. Diamond is dominant in most of the following generations, except some occurrences of lonsdaleite or  $M$ -carbon as the lowest-energy structure of a generation. At the 51th generation, the system reverted to graphite. We note that since at each value of  $h$  a number of structures are obtained by softmutation at each generation, or metastep (but only one structure per metastep was found using MD moves in the original metadynamics method<sup>3</sup>), the structural information is much richer in this version of metadynamics. Many candidate metastable structures can thus be produced.

Both bct- $C_4$  and  $M$ -carbon have been shown to match the experimental diffraction patterns,<sup>31,32</sup> and it was unclear which structure has the lower formation barrier. The best method to compute these barriers is TPS, but it requires an initial model of the pathway, which is then evolved using Monte Carlo sampling. Our results provide such an initial pathway both for the graphite to  $M$ -carbon and graphite to bct- $C_4$  transformations and structural relationships are very clear from Fig. 7. Our TPS simulations<sup>27</sup> suggest that  $M$ -carbon has the lowest barrier for the formation from graphite among all carbon phases, and is thus the likeliest structure for the phase observed in experiment. This example shows how the present metadynamics-based technique can be used in the search for metastable phases.

Starting the calculation at 20 GPa from another polytype, graphite-3R, we again easily found the diamond structure (ground state) and a number of low-energy metastable structures with  $sp^3$  hybridization. We also performed a search at  $P = 1 \text{ atm}$ , starting from the diamond structure. Graphite-2H (the ground state) and a number of mixed  $sp^2$ - $sp^3$  structures were easily found. The results of the simulations are shown in Fig. 6b and c.

## 4 Conclusions

We have developed a novel method for crystal structure prediction, based on the ideas of metadynamics and evolutionary global optimization. We use the cell shape (6 degrees of freedom) as the collective variables, sampling which we find and cross energy barriers in the directions of likely phase transitions. At



**Fig. 7** Structures found by metadynamics starting from graphite-2H at 20 GPa. (a) initial graphite-2H; (b) buckled graphite layer; (c) bct  $C_4$  with 4 + 8 membered rings; (d)  $M$ -carbon with 5 + 7 membered rings; (e) lonsdaleite.

each value of the collective variable  $h$ , we equilibrate the system using evolutionary variation operators (in the example given here - softmutation), which efficiently explores the remaining  $3N-3$  internal degrees of freedom. The success of tests on  $\text{SiO}_2$ ,  $\text{MgSiO}_3$ , carbon and  $\text{Al}_2\text{SiO}_5$  proves the power of this method. Going beyond crystal structure prediction, this method is also capable of producing transformation trajectories between phases, and is thus useful for understanding the transition mechanisms. The method marries ideas from the standard MD-metadynamics<sup>20</sup> and evolutionary algorithm USPEX.<sup>6</sup> Like standard metadynamics and unlike USPEX, present technique does require a reasonable starting structure and has the ability to find transition pathways (due to the use of finite displacements along the lowest-frequency mode eigenvectors). Yet, unlike MD-metadynamics, our method has a more efficient equilibration and at each metastep produces a set of structures (rather than a single structure), *i.e.* a richer chemical information. It does not get stuck in amorphization during the simulation - a common problem for MD-metadynamics. For large systems, it can in some cases be more efficient than USPEX, provided a good initial structure. Present technique has a very different philosophy from the USPEX method and in many ways is complementary to it.

## Acknowledgements

Calculations were performed on the cluster of Center for Functional Nanomaterials, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under contract No. DE-AC02-98CH10086, and on Skif-MSU supercomputer (Moscow State University, Russia) and at the Joint Supercomputer Center (Russian Academy of Sciences, Moscow, Russia). This work is funded by DARPA (grant N66001-10-1-4037) and National Science Foundation (grant EAR-1114313). We thank Prof. J. D. Gale for help with the GULP code, and Prof. R. Martonak for valuable discussions.

## References

- 1 J. Maddox, *Nature*, 1988, **335**, 201.
- 2 J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal and V. Caignaert, *Nature*, 1990, **346**, 343–345.
- 3 R. Martoňák, A. Laio and M. Parrinello, *Phys. Rev. Lett.*, 2003, **90**, 075503.
- 4 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
- 5 S. Goedecker, *J. Chem. Phys.*, 2004, **120**, 9911–9917.
- 6 A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
- 7 C. M. Freeman, J. M. Newsam, S. M. Levine and C. R. A. Catlow, *J. Mater. Chem.*, 1993, **3**, 531–535.
- 8 A. O. Lyakhov and A. R. Oganov, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 092103.
- 9 Q. Zhu, A. R. Oganov, M. A. Salvadó, P. Pertierra and A. O. Lyakhov, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 193410.
- 10 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.
- 11 A. R. Oganov, R. Martoňák, A. Laio, P. Raiteri and P. Michele, *Nature*, 2005, **438**, 1142–1144.
- 12 R. Martoňák, D. Donadio, A. R. Oganov and M. Parrinello, *Nat. Mater.*, 2006, **5**, 623–626.
- 13 P. Raiteri, R. Martoňák and M. Parrinello, *Angew. Chem., Int. Ed.*, 2005, **44**, 3769–3773.
- 14 A. O. Lyakhov, A. R. Oganov and M. Valle, *Comput. Phys. Commun.*, 2010, **181**, 1623–1632.
- 15 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.
- 16 A. F. Voter, *J. Chem. Phys.*, 1997, **106**, 4665–4677.
- 17 A. R. Oganov, A. O. Lyakhov and M. Valle, *Acc. Chem. Res.*, 2011, **44**, 227–237.
- 18 F. Jensen, *Computational Chemistry*, Wiley, New York, 1999.
- 19 K. Li, X. Wang, F. Zhang and D. Xue, *Phys. Rev. Lett.*, 2008, **100**, 235504.
- 20 R. Martoňák, A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, F. Zipoli and M. Parrinello, *Z. Kristallogr.*, 2005, **220**, 489–498.
- 21 S. F. Pugh, *Philos. Mag.*, 1954, **45**, 823–843.
- 22 J. Gale and A. Rohl, *Mol. Simul.*, 2003, **29**, 291.
- 23 B. W. H. van Beest, G. J. Kramer and R. A. van Santen, *Phys. Rev. Lett.*, 1990, **64**, 1955–1958.
- 24 A. R. Oganov, J. P. Brodholt and G. D. Price, *Phys. Earth Planet. Inter.*, 2000, **122**, 277–288.
- 25 A. R. Oganov, G. D. Price and J. P. Brodholt, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2001, **57**, 548–557.
- 26 V. Urusov, V. Oganov and N. Eremin, *Geochem. Int.*, 1998, **36**, 3897–414.
- 27 S. E. Boulfelfel, A. R. Oganov and S. Leoni, submitted.
- 28 W. L. Mao, H.-K. Mao, P. J. Eng, T. P. Trainor, M. Newville, C.-C. Kao, D. L. Heinz, J. Shu, Y. Meng and R. J. Hemley, *Science*, 2003, **302**, 425–427.
- 29 E. A. Ekimov, V. A. Sidorov, E. D. Bauer, N. N. Melnik, N. J. Curro, J. D. Thompson and S. M. Stishov, *Nature*, 2004, **428**, 542–545.
- 30 K. Tanigaki, T. W. Ebbesen, S. Saito, J. Mizuki, J. S. Tsai, Y. Kubo and S. Kuroshima, *Nature*, 1991, **352**, 222–223.
- 31 Q. Li, Y. Ma, A. R. Oganov, H. Wang, H. Wang, Y. Xu, T. Cui, H.-K. Mao and G. Zou, *Phys. Rev. Lett.*, 2009, **102**, 175506.
- 32 K. Umemoto, R. M. Wentzcovitch, S. Saito and T. Miyake, *Phys. Rev. Lett.*, 2010, **104**, 125504.
- 33 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 34 P. E. Blöchl, *Phys. Rev. B: Condens. Matter*, 1994, **50**, 17953–17979.
- 35 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter*, 1996, **54**, 11169–11186.