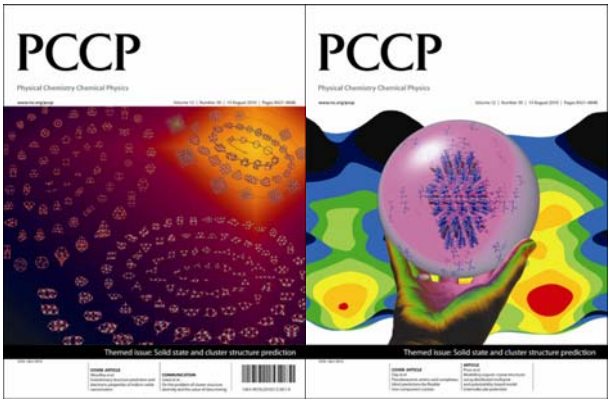
This paper is published as part of a *PCCP* themed issue on solid state and cluster structure prediction

Guest Editors: Richard Catlow and Scott M. Woodley



Editorial

Solid state and cluster prediction
Scott M. Woodley and Richard Catlow, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp90058c

Communication

On the problem of cluster structure diversity and the value of data mining
Alexey A. Sokol, C. Richard A. Catlow, Martina Miskufova, Stephen A. Shevlin, Abdullah A. Al-Sunaidi, Aron Walsh and Scott M. Woodley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00068j

Papers

Evolutionary structure prediction and electronic properties of indium oxide nanoclusters
Aron Walsh and Scott M. Woodley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00056f

Exploration of multiple energy landscapes for zirconia nanoclusters
Scott M. Woodley, Said Hamad and C. Richard A. Catlow, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00057d

Pseudoracemic amino acid complexes: blind predictions for flexible two-component crystals
Carl Henrik Görbitz, Bjørn Dalhus and Graeme M. Day, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004055j

Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials
Sarah L. Price, Maurice Leslie, Gareth W. A. Welch, Matthew Habgood, Louise S. Price, Panagiotis G. Karamertzanis and Graeme M. Day, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004164e

Ab initio prediction of low-temperature phase diagrams in the Al–Ga–In–As system, MAs–M´As (M, M´= Al, Ga or In) and AlAs–GaAs–InAs, *via* the global study of energy landscapes
Ilya V. Pentin, J. Christian Schön and Martin Jansen, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004040c

Importance of London dispersion effects for the packing of molecular crystals: a case study for intramolecular stacking in a bis-thiophene derivative
Jonas Moellmann and Stefan Grimme, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003432k

An extensive theoretical survey of low-density allotropy in silicon
Martijn A. Zwijnenburg, Kim E. Jelfs and Stefan T. Bromley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004375c

Predicting transition pressures for obtaining nanoporous semiconductor polymorphs: oxides and chalcogenides of Zn, Cd and Mg
Winyoo Sangthong, Jumras Limtrakul, Francesc Illas and Stefan T. Bromley, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c0cp00002g

Databases of virtual inorganic crystal structures and their applications
Armel Le Bail, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003907c

Flexibility of ideal zeolite frameworks
V. Kapko, C. Dawson, M. M. J. Treacy and M. F. Thorpe, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003977b

Constant pressure molecular dynamics simulations for ellipsoidal, cylindrical and cuboidal nano-objects based on inertia tensor information
Clive Bealing, Giorgia Fugallo, Roman Martoňák and Carla Molteni, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004053c

Appearance of bulk-like motifs in Si, Ge, and Al clusters
Wen-Cai Lu, C. Z. Wang, Li-Zhen Zhao, Wei Zhang, Wei Qin and K. M. Ho, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004059b

Small germanium sulfide clusters: mass spectrometry and density functional calculations
Joseph J. BelBruno and Andrei Burnin, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003704d

Prediction of the structures of free and oxide-supported nanoparticles by means of atomistic approaches: the benchmark case of nickel clusters
Giulia Rossi, Luca Anghinolfi, Riccardo Ferrando, Florin Nita, Giovanni Barcaro and Alessandro Fortunelli, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003949g

Crystal structure prediction and isostructurality of three small organic halogen compounds
Aldi Asmadi, John Kendrick and Frank J. J. Leusen, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c003971c

Aspects of crystal structure prediction: some successes and some difficulties
Michael O'Keeffe, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004039h

Nanopolycrystalline materials; a general atomistic model for simulation
Dean C. Sayle, Benoît C. Mangili, David W. Price and Thi X. Sayle, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/b918990d

Isomorphism between ice and silica
Gareth A. Tribello, Ben Slater, Martijn A. Zwijnenburg and Robert G. Bell, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/b916367k

Investigation of the structures and chemical ordering of small Pd–Au clusters as a function of composition and potential parameterisation
Ramli Ismail and Roy L. Johnston, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004044d

Predicting crystal structures *ab initio*: group 14 nitrides and phosphides
Judy N. Hart, Neil L. Allan and Frederik Claeyssens, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004151c

Zeolitic polyoxometalates metal organic frameworks (Z-POMOF) with imidazole ligands and ε-Keggin ions as building blocks; computational evaluation of hypothetical polymorphs and a synthesis approach
L. Marleny Rodriguez Albelo, A. Rabdel Ruiz-Salvador, Dewi L. Lewis, Ariel Gómez, Pierre Mialane, Jérome Marrot, Anne Dolbecq, Alvaro Sampieri and Caroline Mellot-Draznieks, *Phys. Chem. Chem. Phys.*, 2010
DOI: 10.1039/c004234j

On the problem of cluster structure diversity and the value of data mining

Alexey A. Sokol,*[a] C. Richard A. Catlow,[a] Martina Miskufova,[a] Stephen A. Shevlin,[a] Abdullah A. Al-Sunaidi,[b] Aron Walsh[a] and Scott M. Woodley[a]

Received 30th March 2010, Accepted 19th June 2010
First published as an Advance Article on the web 8th July 2010
DOI: 10.1039/c0cp00068j

Data mining, involving cross examination of cluster structure pools collected for ZnO, GaN, LiF and AgI, has been applied to predict plausible cluster structures of related binary materials. We consider the energy landscapes of $(MX)_{12}$ clusters for materials that possess tetrahedral bulk phases, wurtzite or sphalerite, including LiF, BeO, BN, AlN, SiC, CuF, ZnO, GaN, GeC and AgI. The energy is evaluated using the hybrid PBEsol0 density functional for structures optimised at the PBEsol level. We report a novel encapsulated iodide structure for AgI and a series of new CuF structures, where significant differences are found between the results for the two functionals.

Structure determination is, perhaps, the most challenging contemporary problem in the field of nanoscience, where computational techniques are often the only available tool. At the nanoscale even simple binary systems that are well characterised in the bulk state exhibit surprising diversity. To understand this behaviour we apply a combination of global search techniques with density functional theory (DFT) calculations, using a state-of-the-art treatment of exchange and correlation, to elucidate the energy landscapes underlying this phenomenon. Our methodology combines evolutionary algorithms[1–5] with data mining from the structures found[6–8] rather than from the literature.[9–14] We calibrate this approach on a range of main group and filled *d*-shell transition metal binary heteropolar compounds LiF, BeO, BN, AlN, SiC, CuF, ZnO, GaN, and GeC, and contrast the cluster structures with AgI, which is isostructural in bulk but has substantially different electronic structure of the anion.

We have previously applied an evolutionary algorithm (EA) to investigate the energy landscape of $(ZnO)_n$, n = 1–32, clusters using interatomic potentials.[15] A large set of viable structures has thus been generated and proved to be robust in further assessment by DFT calculations.[16] Hereafter we refer to this set of structures as a *pool*, and it would typically contain stationary points of interest within the configurational space. In nature, a wide range of compounds of 1 : 1 stoichiometry adopt bulk structures similar to those of zinc oxide, with tetrahedral coordination of both ions arranged in a wurtzite or zinc blende (sphalerite) structure. A higher coordination of constituent ions is also observed in these compounds under pressure or in related compounds, and it is commonly believed that the structural preference is controlled by the ratio of the constituents' ionic radii. Based on a simple ionic model, where ions are considered to be incompressible nonpolarisable spheres, the coordination of the smaller cations relative to the larger anion, and thus the structural stability, is limited by ionic radius ratio rules.[17,18] For example, the rock-salt structure with octahedral coordination should be stable for r_M/r_X of 0.414–0.717, as demonstrated by MgO (0.46), LiF (0.44) and AgCl (0.70). The cubic sphalerite (or hexagonal wurtzite) structures with tetrahedral coordination should be favoured for a lower radius ratio, while higher coordination structures such as the eightfold coordinated CsCl structure should be favoured by a higher radius ratio. Due to the simple nature of the theory, there are many known exceptions to the rule. Cuprous fluoride, CuF, was reported in the sphalerite structure (a = 4.264 Å)[19] in 1933; however, there has been an absence of more recent reports. Whereas cuprous oxide is characterised by linear Cu coordination (O–Cu–O dumbbells), Cu(I) adopts tetrahedral coordination in chalcogenide compounds such as the chalcopyrite and kesterite family of materials.[20,21] Furthermore, the heavier cuprous halides (CuCl, CuBr and CuI) adopt the sphalerite structure. The difficulty in synthesising CuF is likely to be related to the competitive formation of cupric fluoride (CuF_2) and complete dissociation into metallic Cu.

One problem of interest in cluster chemistry is the structural diversity, as for each cluster size a large number of candidate structures are identified by computational means. Which of these structures are realised in nature or can be targeted in synthesis? As the structure of clusters and nanoparticles of these compounds remains largely undetermined by experiment, a computational approach could be a useful means of assessment. Based on the similarity of the bulk structures of "tetrahedral" compounds, in this work we hypothesise that the energy landscapes of their cluster structures are also similar, *i.e.* showing comparable features. Pursuing this idea, we data mined the pool of structures obtained for zinc oxide clusters of one particular size n = 12, in which a sodalite cage has previously been identified as a particularly stable structural unit for ZnO and a number of related compounds of 1 : 1 stoichiometry by computational means[7,22,23] and, which for BN, has been observed experimentally.[24–27]

Of the twenty highest ranked (lowest energy) structures of ZnO, we have selected nine structures that are not only of the

[a] *Department of Chemistry, University College London, 3rd Floor, Kathleen Lonsdale Building, Gower Street, London WC1E 6BT, United Kingdom. E-mail: a.sokol@ucl.ac.uk*
[b] *King Fahd University of Petroleum and Minerals, Department of Physics, Dhahran 31261, Saudi Arabia*

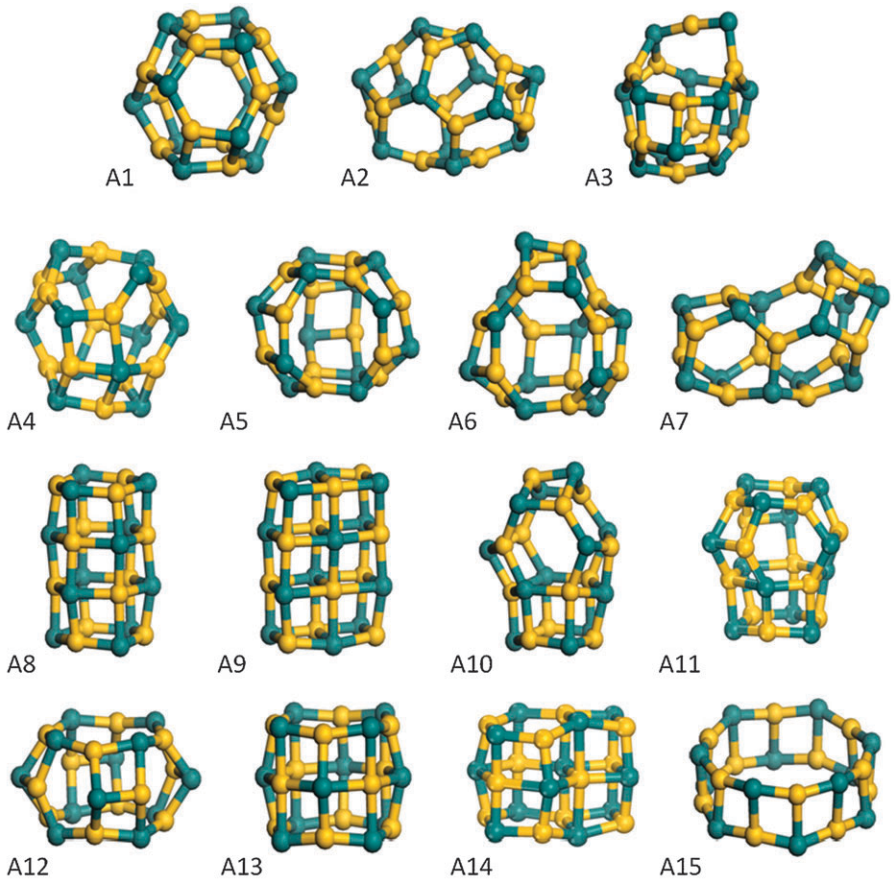


Fig. 1 Optimised $(MX)_{12}$ silicon carbide structures derived from data mining. Yellow is reserved for cations and teal for anions.

lowest energy but also possess clearly defined three-dimensional morphology as discussed below (structures A1–A3, A5–A7, A10 and A13 shown in Fig. 1). To explore the reliability of data mining using only one parent compound, we performed three complementary EA searches of energy landscapes[15,28] employing interatomic potentials: (i) for GaN,[29] which is a close match of ZnO regarding both the structure and physical properties, (ii) LiF,[30] which is a borderline case between tetrahedral and octahedral compounds, with the rock-salt bulk structure under ambient conditions,[31,32] but the sodalite cage structure adopted in the ground state as an $n = 12$ cluster,[33] and finally, (iii) for AgI,[34] which exhibits markedly different behaviour from the main group compounds. Search (i) has resulted only in one additional structure (A15), which lies too high in energy on the rigid ion energy landscape of ZnO for the structure to remain in the final pool. Search (ii) generated three new structures (A4, A8 and A11), which are characterised by multiple tetragonal faces. Search (iii) produced one new structure A4, which proved to be one of the most stable local minima for a great number of compounds, but has not previously been identified using rigid ion potentials. Further, we have constructed two bulk like structures, one, A9, closely related to A8, and previously investigated for MgO by Roberts and Johnston,[4] using a partial charge model, and another (A16), which is a two-dimensional cluster cut from the bulk of the tetrahedral compounds (a hexagonal sheet).

Details of recent applications of the Evolutionary Algorithm to zirconia[8] and india (indium oxide)[11] clusters can be found in this PCCP themed issue on structure prediction.

Total energy electronic structure calculations were performed using an all-electron DFT method with a local atomic numerical orbital basis set, as implemented in the FHI-AIMS code.[35,36] A "Tier-2" basis was employed for each element with scalar-relativistic effects treated at the ZORA level.[37] Geometry optimization was performed using a Quasi-Newton (BFGS) algorithm and the PBEsol exchange–correlation functional.[38] The equilibrium structures were then subject to a final total energy evaluation using the PBEsol0 hybrid functional, where 25% of the semi-local PBEsol exchange functional is replaced by non-local Hartree–Fock exchange.[39] This systematic approach allows for the utilisation of a modern DFT functional with only moderate computational expense.

We should compare our results for the energetics of different $(SiC)_{12}$ isomers as calculated using the method described above to previous work, which employed the PBE exchange–correlation functional and a plane wave basis set.[40] We find that the energetic ordering of isomers is identical in the two methodologies. The first isomer structure not investigated in our previous work is the A11 structure, which has the fourth lowest total energy. The PBE calculations underestimate total energy differences with respect to the

PBEsol0 calculations, with the difference in relative energies ranging from 0.37 eV for the A6 isomer to 0.61 eV for the A12 isomer.

The sodalite (β-)cage (A1) proved to be the ground state configuration for the majority of compounds studied, and for these materials we observe remarkable similarity in the order of stability of lowest energy local minima, while for the less stable structures the energy landscapes become much more compound specific in giving preference to differing structural motifs (see Table 1). In particular, those energetically favoured are closed, one-layer structures with three- and four-coordinated ions forming patchworks of edge-sharing tetra-, hexa- and

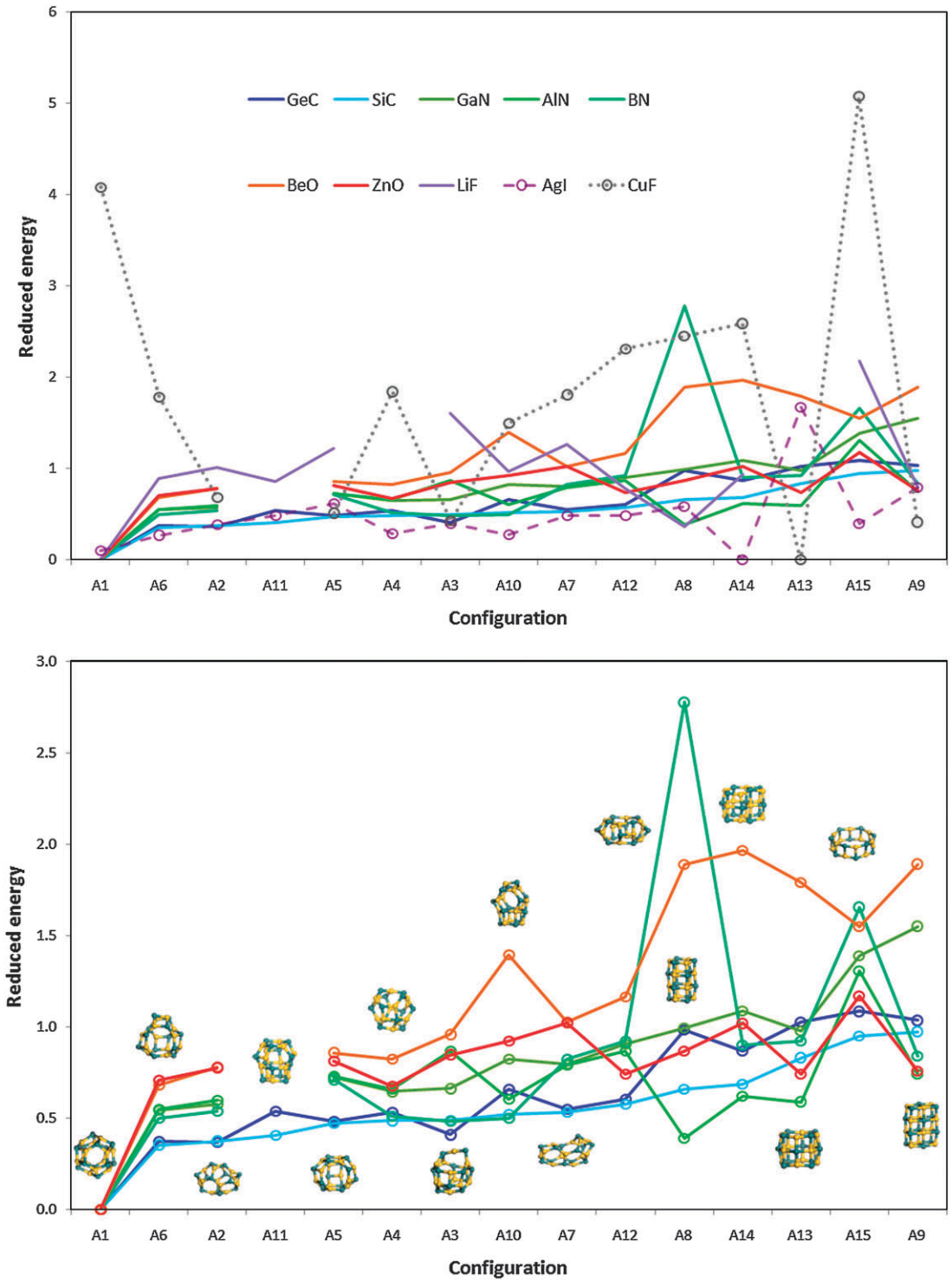


Fig. 2 Relative stability graphs for $(MX)_{12}$ cluster configurations with respect to the global minimum for each compound (energy in eV). The upper panel presents a summary of results while the lower concentrates on the main group and filled *d*-shell transition metal tetrahedral compounds. The gaps in the graphs correspond to the cases where a cluster configuration transformed on relaxation as indicated in Table 1.

Table 1 Relative energies of $(MX)_{12}$ cluster configurations with respect to the global minimum for each compound, shown in bold. Change in configuration on relaxation is indicated by the label of the final optimised configuration highlighted in italics. Cluster configurations for CuF (C configurations) and AgI (B configurations) are ordered according to the starting configurations (A). All energies are in eV per cluster

	LiF	BeO	BN	AlN	SiC	CuF	ZnO	GaN	GeC	AgI
A1	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	2.15	**0.00**	**0.00**	**0.00**	0.10
A2	0.57	2.05	3.26	3.01	3.39	0.36	1.69	2.83	3.18	0.38
A3	0.91	2.52	2.93	4.38	4.42	0.23	1.84	3.25	3.53	0.40
A4	*A1*	2.17	3.11	3.33	4.39	0.97	1.47	3.17	4.61	0.29
A5	0.69	2.26	4.32	3.67	4.25	0.27	1.77	3.56	4.18	0.62
A6	0.51	1.79	3.03	2.77	3.20	0.94	1.54	2.68	3.24	0.27
A7	0.72	2.71	4.99	4.00	4.81	0.96	2.22	3.92	4.74	0.49
A8	0.21	4.98	16.85	1.97	5.93	1.29	1.88	4.87	8.46	0.58
A9	0.45	4.98	5.10	3.75	8.79	0.22	1.65	7.62	8.93	0.79
A10	0.55	3.67	3.03	3.05	4.71	0.79	2.01	4.06	5.66	0.27
A11	0.49	*A1*	*A1*	*A1*	3.68	1.29	*A1*	*A1*	4.64	0.49
A12	0.44	3.07	5.59	4.38	5.20	1.22	1.61	4.45	5.20	0.49
A13	*A12*	4.72	5.59	2.98	7.50	**0.00**	1.61	4.81	8.84	1.67
A14	0.53	5.18	5.47	3.13	6.19	1.37	2.22	5.34	7.48	**0.00**
A15	1.23	4.08	10.04	6.59	8.55	2.68	2.55	6.82	9.35	0.40
A16	3.24	4.73	4.78	12.67	13.09	0.11	3.17	6.92	8.87	1.87

octagonal rings. The patchworks can be seen to resolve in bubble (A1 and A2), tubular (drum—A8, A13, A15) or related defective structures, characterised by "handles", *i.e.* links formed by two-coordinated ions (A3). (The complete EA pool of ZnO structures includes a very large number of such defective configurations, which are of relatively low energy but not considered here. This abundance hinders the search of the higher energy part of the energy landscape for new structural motifs.) We recognise bubbles including octagonal rings (A4–A7) and tetragonal rock-salt-related patchworks based on the A9 structure in two subgroups, those constructed from four (A8–A11) and three layers (A12–A14).

The ionic nature of the materials helps in furthering our analysis. In the graphs of relative stability shown in Fig. 2 we scale the raw energies summarised in Table 1, by the Coulombic energies per bond, q^2/r (where q is the formal ionic charge, r is an interionic distance determined here from the relaxed sodalite cage), which highlights the close similarities within the groups of oxides, nitrides and carbides, arranged thus by charge. Perhaps, even closer convergence between compounds could be achieved by considering the ionic polarisability, with higher charged anions being more efficient at screening the Coulombic interactions. However, this observation can be taken only so far as, for example, the fourth-ranked structure for carbides proves to be unstable for other compounds with its structure transforming on relaxation into the sodalite cage. At the same time the LiF and AgI compounds readily adopt this structure (as rank 5 and 9, respectively).

Considering Fig. 2, it is tempting to make statements about the influence of structural motifs on the cluster stability. For example, the calculated energetics of bulk-like configuration A9 implies that higher coordination is less beneficial for GaN and even less so for BeO, possibly due to the small size of the Be ions. Moreover, comparing bubble configurations

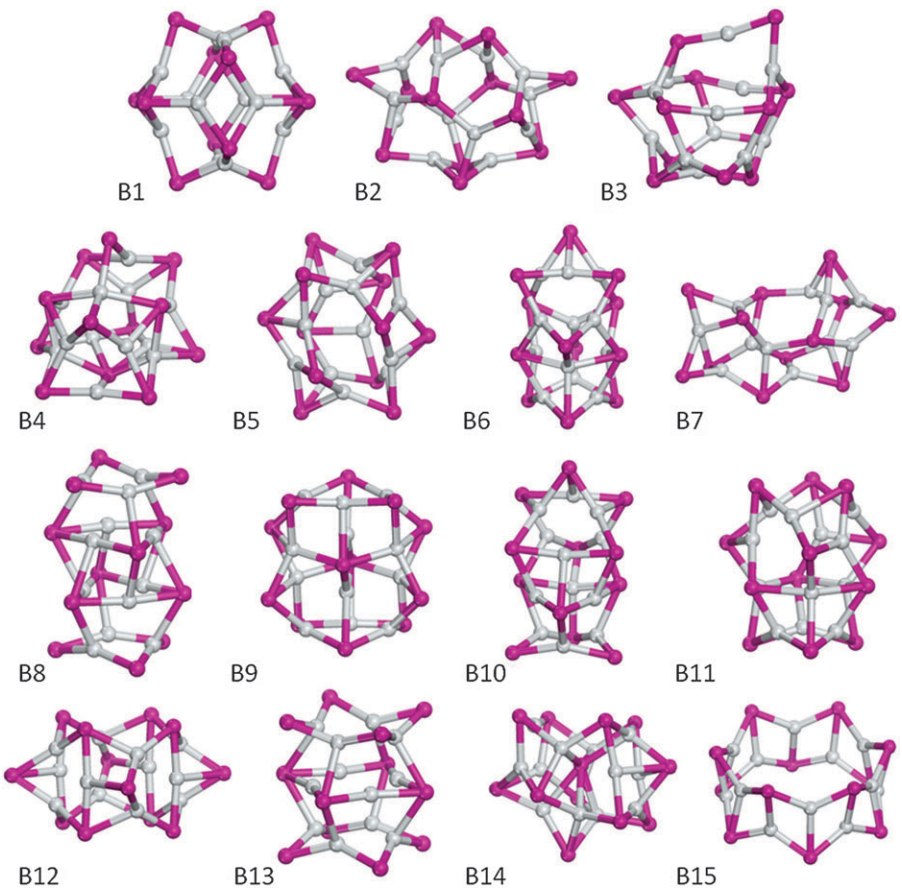


Fig. 3 $(AgI)_{12}$ cluster configurations obtained by data mining structures refined for SiC. Purple is reserved for iodine and silver for silver.

A8, A11 and A5, we observe that the tetragonal patchwork on the A8 cluster opens up in one case to create three hexagons and in another three octagons, thus reducing the average coordination number of constituent ions. In both cases, for GaN and to a greater extent for BeO, the cluster with the lower average coordination number is more stable, which corroborates our observation for A9. However, to solidify our conclusions, these trends need to be explored in further detail, using larger pools of specially designed structures of different sizes, *i.e.* using an alternative (to data mining) approach to cluster generation, which is outside of the scope of the current communication.

The underlying energy landscape is maintained even by such a disparate compound as AgI. However, the landscape is now severely distorted (see Fig. 3), which we ascribe to the exceptionally large size and polarisability of the iodide ion.

The high polarisability of AgI is also probably the reason for the stabilisation of a new cation-rich cage structure encapsulating an anion, which proved to be the global minimum although only 0.1 eV lower than the sodalite cage. This type of structure was first proposed by Kasuya *et al.*[41] to rationalise the appearance of $(MX)_{13}$ and $(MX)_{34}$ magic number clusters in mass spectra of laser ablated CdS, CdSe, CdTe, ZnS and ZnSe. Later studies have confirmed these results for cadmium, but not zinc.[28,42,43] Perhaps the more polarisable Cd is the source of extra stabilisation of the encapsulated anion configurations: polarisability of Ag(I) cations could also be the contributing factor. We note that the encapsulated anion structure reverts to one of the cage configurations for all other compounds. Another feature, which should be ascribed to the anion's large size, is the breaking of a large number of AgI bonds as a large proportion of the structures are stabilised by forming handles instead of the three-coordinated ions. This effect is not just a result of particular criteria used to define bonds for visualisation, but we typically find the corresponding interatomic distances increased by more than 0.5 Å.

CuF is a notable exception to the previous observations, as it exhibits a tendency for copper and fluorine segregation with a copper dominated core decorated by fluoride species on the surface (see Fig. 4). The copper core is reminiscent of the copper clusters that we previously modelled on ZnO surfaces.[44] Remarkably, even among the non-segregated structures of CuF, the sodalite cage is least stable with the emergence of a novel planar structure C16 (second lowest local minimum), which is in contrast to all other compounds considered (see Fig. 5). The new global minimum structure C13 features three stoichiometric CuF layers only weakly bound to each other. Within each layer, we observe a perfectly square Cu cluster inscribed within the F square rotated by 45°, which achieves linear coordination for each Cu(I) ion.

To assess the stability of the local minimum structures C1, C13 and C16, molecular dynamics simulations were performed

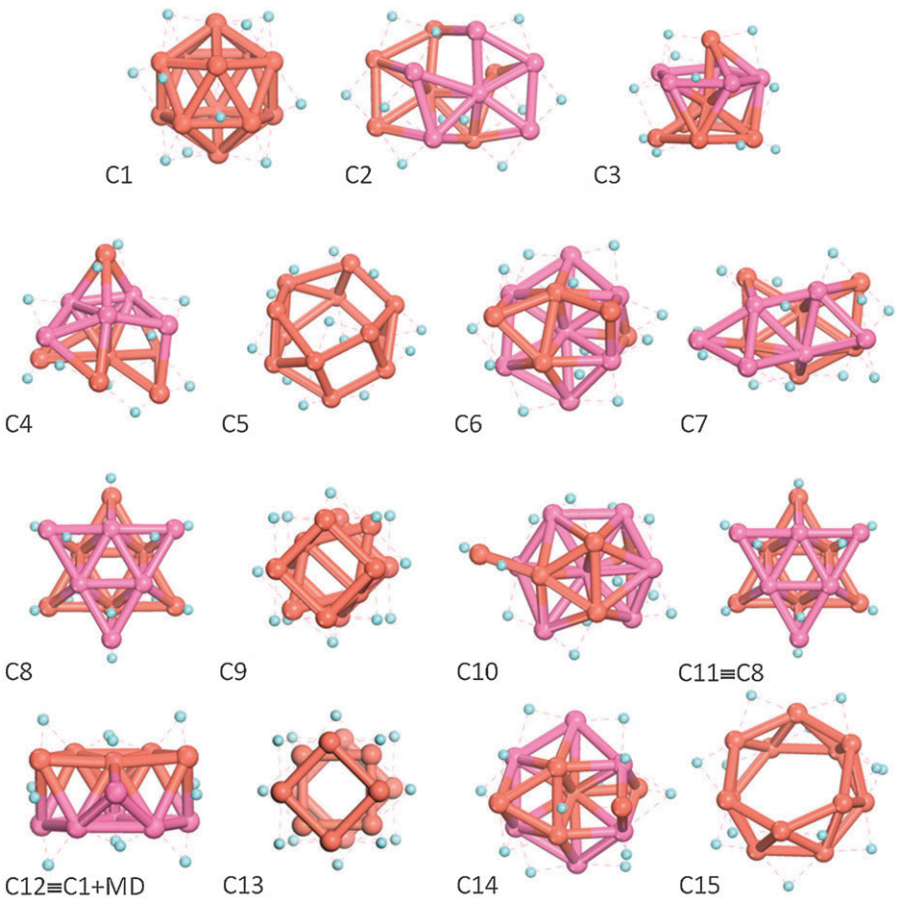


Fig. 4 $(CuF)_{12}$ cluster configurations obtained by data mining. Cyan is reserved for fluorine and copper for copper. To distinguish the layered structure of Cu clusters we also highlight alternative layers by interchanging copper and pink colours.

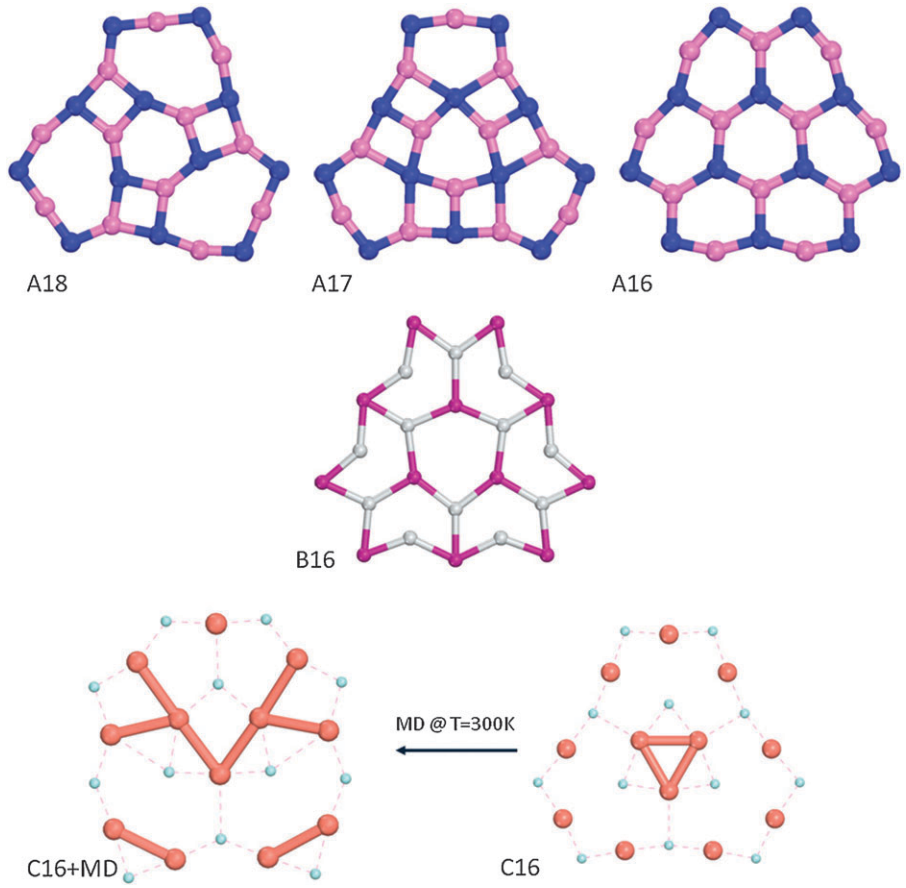


Fig. 5 Planar cluster configurations obtained as a cut from a hexagonal sheet and relaxed for each material of interest. The CuF cluster configuration C16 + MD is obtained by quenching after a molecular dynamics run at room temperature on the PBE energy landscape.

at the PBEsol level: 2 ps (NVT) simulations were run at 300 K with a Nosé–Hoover thermostat and using a time step of 1 fs. The final structures were quenched to 0 K (geometry optimised).

On quenching after molecular dynamics runs, structure C13, transforms into configuration C14, which lies 0.25 eV below in energy at the PBEsol level of theory. However, the preferred energy measure using the PBEsol0 functional reverts the order putting the initial configuration at 2.59 eV lower in energy, thus defining the global minimum on the CuF energy landscape. The disagreement between the two functionals is unexpected, and to clarify the matter we considered the difference between the PBEsol and PBEsol0 landscapes for all materials. Notably, they proved to be quite similar for all materials but CuF. We have plotted the difference in Fig. 6. It appears that there is a direct correlation between the energy gaps in one-electron states (difference between highest occupied and lowest unoccupied orbital energies). As the energy gaps show similar behaviour for both functionals we chose to show the trend as an average value of the two. For each cluster that exhibits a well ordered cation–anion separation which keeps copper ions apart we observe a sharp increase in the energy gap and stabilisation with respect to the alternative structures. This behaviour can be rationalised by considering self-interaction effects in the filled *d*-shells of the copper ions. Whenever copper ions are sufficiently close to each other electron delocalisation can be expected to occur as in resonance bonding in copper metal, which should result in a delocalisation of such electrons and a reduction of the corresponding self-interaction. The hybrid functional, PBEsol0, by making use of the Hartree–Fock-like exchange partially cancels the self-interaction and thus favours more localised states. The PBEsol functional, therefore, favours metallic bonding in CuF clusters, whereas the PBEsol0 favours ionic bonding. Hence, for those structures where the largest one-electron energy gap is observed and there is a large disagreement regarding the relative stability predictions between GGA and hybrid functionals, we expect that the hybrid functional is closer to that which may be observed in experiment.

To check that the metal segregation within the CuF clusters does not cause problems with our choice of the local basis set, we have performed plane-wave calculation using the VASP code as described above for the SiC clusters. We find that with few exceptions delineated below the cluster energies are in good correspondence between both the codes, basis sets and two particular functionals, PBE and PBEsol. The exceptions are the planar structures C16 and C16 + MD (quenched), for which the relative cluster energies are significantly lower for the PBE exchange and correlation functional than PBEsol. Curiously, the energy landscape for PBE shown in Fig. 6 is similar to that of PBEsol0, but the difference with PBEsol is not as pronounced. This observation is in line with the more empirical nature of the PBE functional, which is biased towards molecules.[38]

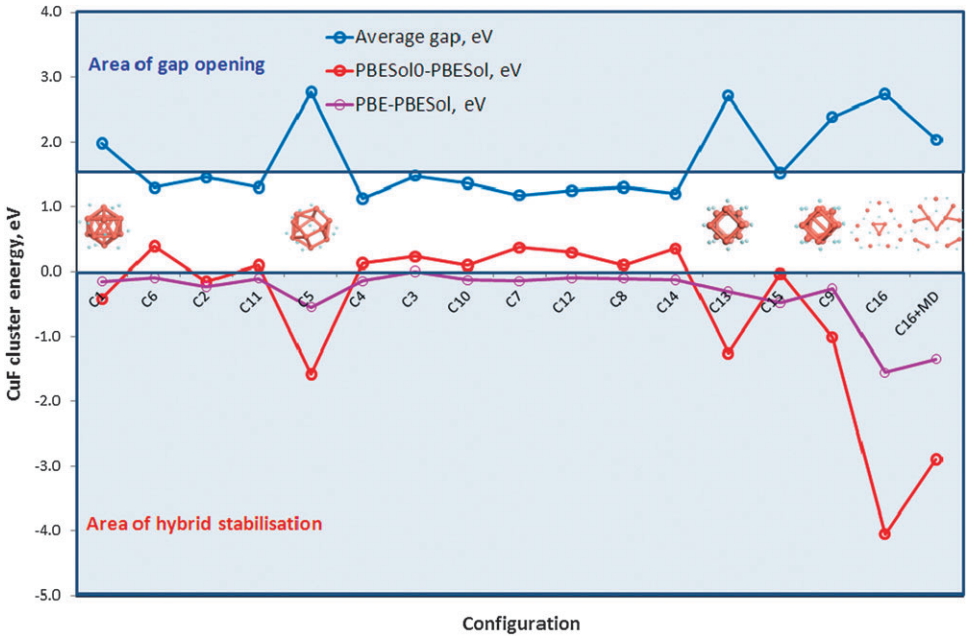


Fig. 6 Energy stabilisation diagram for ordered $(CuF)_{12}$ cluster configurations correlated with one-electron energy gaps (the gap calculated as an arithmetic mean value between the two functionals). The energy differences (PBESol0-PBESol and PBE-PBESol) refer to the relative energy differences between the two respective functionals for each cluster.

The planar structures shown in Fig. 5 are of highest energy for all compounds but CuF and could not be generated by the global search techniques including EA, unless targeted specifically. They are still of interest as they would form the basis of larger clusters structurally related to the bulk phases. The planar structures exhibit remarkable diversity in their coordination environments, where interestingly AlN is capable of supporting three patterns of connectivity, while SiC only adopts configuration A16, ZnO A17, and BN A18. It is the latter configuration that is the global minimum among the planar structures of AlN (stabilised by 1.73 eV from A16). Both AgI and CuF also adopt the A16 structure. The exceptional stability of the CuF C16 planar structure (only at 0.11 eV above the ground state) can be understood considering the strong preference of Cu(I) ions for linear coordination, which is satisfied for the maximum number of Cu ions within the C16 configuration, and is necessarily lower in three-dimensional structures.

In summary, we have demonstrated the power of the data mining technique applied to the structure prediction of $(MX)_{12}$ clusters. We have shown that the cluster structure even of such disparate compounds as ZnO, AgI and CuF can be ascertained using sufficiently representative pools of structures generated beforehand with global search techniques. Combining structure pools of sufficiently different compounds that span the configurational space has proved to be essential to the success of a data mining technique.

We would like to thank Christian Schön, Claudio Silva and Z. Xiao Guo for useful discussions. We gratefully acknowledge funding by EPSRC (grants GR/T11364/01, GR/S26965/01, EP/E040071/1, GR/S52636/01, EP/E046193/1, EP/D504872 and EP/F067496) and our membership of the Sustainable Hydrogen Energy and HPC Materials Chemistry Consortia. A.W. would like to acknowledge funding from a Marie-Curie Intra-European Fellowship from the European Union under the Seventh Framework Programme.

References

1 B. Hartke, in *Applications of Evolutionary Computation in Chemistry*, R. L. Johnston, Springer-Verlag Berlin, Berlin, 2004, vol. 110, pp. 33–53.
2 S. M. Woodley, in *Applications of Evolutionary Computation in Chemistry*, R. L. Johnston, Springer-Verlag Berlin, Berlin, 2004, vol. 110, pp. 95–132.
3 S. M. Woodley and R. Catlow, *Nat. Mater.*, 2008, **7**, 937–946.
4 C. Roberts and R. L. Johnston, *Phys. Chem. Chem. Phys.*, 2001, **3**, 5024–5034.
5 D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.*, 1995, **75**, 288–291.
6 M. E. Zandler, E. C. Behrman, M. B. Arrasmith, J. R. Myers and T. V. Smith, *THEOCHEM*, 1996, **362**, 215–224.
7 S. A. Shevlin, Z. X. Guo, H. J. J. van Dam, P. Sherwood, C. R. A. Catlow, A. A. Sokol and S. M. Woodley, *Phys. Chem. Chem. Phys.*, 2008, **10**, 1944–1959.
8 S. M. Woodley, S. Hamad and C. R. A. Catlow, *Phys. Chem. Chem. Phys.*, 2010, DOI: 10.1039/c0cp00057d.
9 K. Rajan, in *Data Mining in Crystallography*, D. W. M. Hofmann and L. N. Kuleshova, Springer-Verlag Berlin, Berlin, 2010, vol. 134, pp. 59–87.
10 S. Curtarolo, D. Morgan, K. Persson, J. Rodgers and G. Ceder, *Phys. Rev. Lett.*, 2003, **91**, 135503.
11 A. Walsh and S. M. Woodley, *Phys. Chem. Chem. Phys.*, 2010, DOI: 10.1039/c0cp00056f.
12 H. Zenasni, H. Aourag, S. R. Broderick and K. Rajan, *Phys. Status Solidi B*, 2010, **247**, 115–121.
13 A. N. Kolmogorov and S. Curtarolo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**.
14 D. W. M. Hofmann, in *Data Mining in Crystallography*, D. W. M. Hofmann and L. N. Kuleshova, Springer-Verlag Berlin, Berlin, 2010, vol. 134, pp. 89–134.
15 A. A. Al-Sunaidi, A. A. Sokol, C. R. A. Catlow and S. M. Woodley, *J. Phys. Chem. C*, 2008, **112**, 18860–18875.
16 A. A. Sokol, C. R. A. Catlow, A. A. Al-Sunaidi and S. M. Woodley.
17 J. D. Dunitz and L. E. Orgel, *Adv. Inorg. Chem. Radiochem.*, 1960, **2**, 1.

18 L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, 1939.
19 F. Ebert and H. Woitinek, *Z. Anorg. Allg. Chem.*, 1933, **210**, 269–272.
20 A. F. Wells, *Structural Inorganic Chemistry*, Oxford University Press, Oxford, 1984.
21 S. Chen, X. G. Gong, A. Walsh and S.-H. Wei, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **79**, 165211.
22 E. C. Behrman, R. K. Foehrweiser, J. R. Myers, B. R. French and M. E. Zandler, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1994, **49**, R1543–R1546.
23 S. M. Woodley, M. B. Watkins, A. A. Sokol, S. A. Shevlin and C. R. A. Catlow, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3176–3185.
24 D. Golberg, Y. Bando, O. Stephan and K. Kurashima, *Appl. Phys. Lett.*, 1998, **73**, 2441–2443.
25 T. Oku, A. Nishiwaki, I. Narita and M. Gonda, *Chem. Phys. Lett.*, 2003, **380**, 620–623.
26 T. Oku, T. Hirano, M. Kuno, T. Kusunose, K. Niihara and K. Suganuma, *Mater. Sci. Eng., B*, 2000, **74**, 206–217.
27 F. Jensen, *Chem. Phys. Lett.*, 1993, **209**, 417–422.
28 S. M. Woodley, A. A. Sokol and C. R. A. Catlow, *Z. Anorg. Allg. Chem.*, 2004, **630**, 2343–2353.
29 C. R. A. Catlow, Z. X. Guo, M. Miskufova, S. A. Shevlin, A. G. H. Smith, A. A. Sokol, A. Walsh, D. J. Wilson and S. M. Woodley, *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.*, 2010, **368**, 3379–3456.
30 M. J. L. Sangster and R. M. Atwood, *J. Phys. C: Solid State Phys.*, 1978, **11**, 1541–1555.
31 K. Doll, J. C. Schon and M. Jansen, *Phys. Chem. Chem. Phys.*, 2007, **9**, 6128–6133.
32 J. C. Schon and M. Jansen, *Comput. Mater. Sci.*, 1995, **4**, 43–58.
33 J. Carrasco, F. Illas and S. T. Bromley, *Phys. Rev. Lett.*, 2007, **99**.
34 A. Wootton and P. Harrowell, *J. Phys. Chem. B*, 2004, **108**, 8412–8418.
35 V. Havu, V. Blum, P. Havu and M. Scheffler, *J. Comput. Phys.*, 2009, **228**, 8367–8379.
36 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175–2196.
37 E. Vanlenthe, E. J. Baerends and J. G. Snijders, *J. Chem. Phys.*, 1994, **101**, 9783–9792.
38 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 136406–136404.
39 J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
40 M. B. Watkins, S. A. Shevlin, A. A. Sokol, B. Slater, C. R. A. Catlow and S. M. Woodley, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3186–3200.
41 A. Kasuya, R. Sivamohan, Y. A. Barnakov, I. M. Dmitruk, T. Nirasawa, V. R. Romanyuk, V. Kumar, S. V. Mamykin, K. Tohji, B. Jeyadevan, K. Shinoda, T. Kudo, O. Terasaki, Z. Liu, R. V. Belosludov, V. Sundararajan and Y. Kawazoe, *Nat. Mater.*, 2004, **3**, 99–102.
42 C. R. A. Catlow, S. T. Bromley, S. Hamad, M. Mora-Fonz, A. A. Sokol and S. M. Woodley, *Phys. Chem. Chem. Phys.*, 2010, **12**, 786–811.
43 A. Burnin and J. J. BelBruno, *Chem. Phys. Lett.*, 2002, **362**, 341–348.
44 S. A. French, A. A. Sokol, C. R. A. Catlow and P. Sherwood, *J. Phys. Chem. C*, 2008, **112**, 7420–7430.