Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 27 July 2009
Accepted 4 July 2010

Addressing chemical diversity by employing the energy landscape concept

Martin Jansen, Klaus Doll and J. Christian Schön*

Max Planck Institute for Solid State Research, Heisenbergstrasse 1, D-70569 Stuttgart, Germany. Correspondence e-mail: schoen@fkf.mpg.de

Exploring the structural diversity of a chemical system rests on three pillars. First, there is the global exploration of its energy landscape that allows one to predict which crystalline modifications can exist in a chemical system at a given temperature and pressure. Next, there is the development of new synthesis methods in solid-state chemistry, which require only very low activation energies such that even metastable modifications corresponding, for example, to minima on the landscape surrounded by low barriers can be realized. Finally, there is the theoretical design of optimal synthesis routes, again based on the study of the system's energy landscape. In this paper the energy landscape approach to the prediction of stable and metastable compounds as a function of temperature and pressure is presented, with a particular focus on possible phase transitions. Furthermore, several examples are presented, where such predicted compounds were subsequently successfully synthesized, often employing a newly developed synthesis method, low-temperature atom-beam deposition.

1. Introduction

Although every configuration of a chemical system of N atoms is mathematically represented by a point in a $3N$-dimensional Euclidean space, the corresponding atom arrangement necessarily resides in our familiar three-dimensional Euclidean space. This fact places severe restrictions on the mathematical transformations that can be applied to the system, *e.g.* limiting the number of possible space-group types to only 230. Combining these restrictions with some empirical knowledge about the size of the atoms, or their preferred bonding patterns, has already enabled us to rationalize the structures of known chemical compounds, and even to extrapolate to not yet realized configurations. Along such rather traditional approaches one matches, for example, three-dimensional space-group symmetries, and the Wyckoff multiplicities implied, with chemical composition, cell contents and preferred coordination numbers (Brown & Shannon, 1973; O'Keeffe & Hyde, 1984, 1985; Brown, 1992; Hansen, 1993) or with the properties of packing schemes of spheres (Müller, 1992, 1998). A quite independent set of procedures can be applied to those classes of chemical systems where the presence of certain uniform bonding geometries or rigid building blocks leads to well defined connectivities among the atoms and thus to topological restrictions on the feasible structures in these systems. The most stringent of such derivations has been the (correct) prediction of all possible alkanes and their isomers by Polya using graph theory (Polya, 1936). A similar graph-theoretical representation of the structure systematics of silicates has been suggested by Klee, Hahn and co-workers (Chung *et al.*, 1984; Klee *et al.*, 1997), and most recently it has been shown that the full structural diversity of, for example, zeolite structures can be accessed by applying graph theory (Treacy *et al.*, 1997; Winkler *et al.*, 1999; Strong *et al.*, 2004; Szu & Hartley, 1987) or tiling procedures in three dimensions (Heesch, 1934; Delgado-Friedrichs *et al.*, 1999, 2007; Foster *et al.*, 2004; Thomas & Klinowski, 2007).

All these approaches are based on using geometrical and/or topological constraints on feasible configurations to allow a complete mathematical enumeration. However, this leads to two intrinsic problems. For one, the procedure implies that there is a close correspondence between a structure obeying the geometrical restrictions and it being kinetically stable, chemically relevant and a reasonable modification. Since the number of mathematically permitted structures grows exponentially with any relaxation of the topological restrictions, it is clear that balancing the demands of computational resources, chemical relevance and completeness of the resulting set of structures is highly non-trivial. However, at the heart of this problem lies the use of geometrical and topological proxies instead of the true physical quantity determining the stability of chemical compounds, and thus their structures: the free energy of the system under investigation and its tendency to minimize, together with the energetic and entropic barriers surrounding the minimum configurations. In our approach to predicting chemical compounds capable of existence and their structures, we therefore focus on the energies associated with the configurations (Schön & Jansen, 1996, 2001, 2009; Jansen, 2002, 2008; Schön, Doll & Jansen, 2010). Quite generally, this representation of the multitude of

all known and still unknown chemical compounds on an energy landscape points the way to a deductive treatment of chemistry, quite in contrast to the inductive approach historically followed in this discipline (Jansen & Schön, 2006; Jansen, 2008).

A rather simple scenario results if one considers the hypothetic conditions of $T = 0\,\mathrm{K}$ and $p = 0\,\mathrm{GPa}$ (and suppresses the zero-point vibrations). Then each minimum of the continuous (hyper)surface of potential energy corresponds to a kinetically (infinitely) stable configuration, and *vice versa*. At finite temperature and pressure, *i.e.* at realistic thermodynamic conditions, each individual minimum becomes metastable, in principle. The relevant quantitity is now the locally ergodic region that can encompass one or many local minima, which corresponds to a macroscopic thermodynamic state. Depending on the thermodynamic boundary conditions applied, one of these regions corresponds to the thermodynamically stable state of the system under consideration, while the remaining regions represent metastable ones, exhibiting a wide spread of lifetimes.

Such physically realistic energy landscapes, exhibiting numerous locally ergodic regions, offer a firm foundation for dealing with virtually all aspects of chemistry on a rational basis. Since the sufficient and necessary precondition for any chemical configuration to exist is that it belongs to such a region, no matter whether one deals with a molecule, a liquid, a crystalline compound or an amorphous solid, all the diverse fields of preparative chemistry can be dealt with on a unified footing.

In our approach to the rational planning of solid-state and materials synthesis, we computationally search the respective energy landscapes for (meta)stable compounds and explore the barrier structures around the relevant local minima. For a given composition, the most stable predicted structure candidates and the lowest connecting barrier heights are displayed in a tree graph presentation. At a first glance, the enormous wealth and diversity of the candidates predicted appears to indicate an extreme (seemingly unrealistic) complexity of chemical matter. However, applying appropriate experimental tools has by now enabled us to realize parts of the tree graphs of the lithium halides and of the elusive sodium nitride, including almost all of its predicted polymorphs, many years after the predictions were published.

In this report we give a brief outline of our conception, including a summary of the basic energy landscape concepts and a short description of the algorithms we employ, and demonstrate the feasibility of our approach by presenting not only theoretical structure predictions but also examples of their experimental verification *via* successful syntheses of predicted compounds.

2. Energy landscape approach to the prediction of (meta)stable compounds and their phase diagrams

The starting point for any prediction of (meta)stable compounds and their phase diagrams without recourse to experimental information is the representation of the chemical system at hand as a collection of N atoms *via* a $3N$-dimensional position vector, $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ [plus a $3N$-dimensional momentum vector, $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$], and its (potential) energy $E(\mathbf{R})$. This energy hypersurface over the $3N$-dimensional space of all atom arrangements is commonly denoted as the energy landscape of the chemical system (Schön & Jansen, 1996, 2001; Jansen, 2002, 2008; Goldstein, 1969; Stillinger & Weber, 1982; Wales, 2003).

The crucial step in going from the classical mechanical description above to the thermodynamic one is the determination of the so-called locally ergodic regions on the energy landscape (Schön & Jansen, 2001; Schön *et al.*, 2003). For a given temperature T, a subset $\mathcal{R}$ of the configuration space is called locally ergodic on the observation timescale t_{obs} if the time $\tau_{\mathrm{eq}}(\mathcal{R}; T)$ it takes for the system to equilibrate within $\mathcal{R}$ is much shorter than t_{obs}, while the time $\tau_{\mathrm{esc}}(\mathcal{R}; T)$ it takes for the system to leave the region $\mathcal{R}$, the so-called escape time, is much larger than t_{obs},

$$\tau_{\mathrm{esc}}(\mathcal{R}; T) \gg t_{\mathrm{obs}} \gg \tau_{\mathrm{eq}}(\mathcal{R}; T). \tag{1}$$

If this holds true then the ergodic theorem tells us that we can replace the time averages of observables $O[\mathbf{R}(t), \mathbf{P}(t)]$ along a trajectory of length $t_{\mathrm{obs}} = t_2 - t_1$,

$$\langle O \rangle_{t_{\mathrm{obs}}} = \frac{1}{t_{\mathrm{obs}}} \int_{t_1}^{t_2} O[\mathbf{R}(t'), \mathbf{P}(t')]\, \mathrm{d}t', \tag{2}$$

inside the locally ergodic region $\mathcal{R}$ by the (Boltzmann) ensemble average of this observable,

$$\langle O \rangle_{\mathrm{ens}}(T) = \frac{\int O(\mathbf{P}, \mathbf{R}) \exp\left[-E(\mathbf{P}, \mathbf{R})/k_{\mathrm{B}}T\right] \mathrm{d}\mathbf{P}\, \mathrm{d}\mathbf{R}}{\int \exp\left[-E(\mathbf{P}, \mathbf{R})/k_{\mathrm{B}}T\right] \mathrm{d}\mathbf{P}\, \mathrm{d}\mathbf{R}}, \tag{3}$$

restricted to the region $\mathcal{R}$,

$$\left| \langle O \rangle_{t_{\mathrm{obs}}} - \langle O \rangle_{\mathrm{ens}}(T) \right| < a. \tag{4}$$

Of course, this 'equality' holds only within an accuracy a, since only local and not global ergodicity is asserted. In particular, we can compute for every locally ergodic region $\mathcal{R}_i$ the local free energy,

$$\begin{aligned} F(\mathcal{R}_i, T) &= -k_{\mathrm{B}}T \ln Z(\mathcal{R}_i, T) \\ &= -k_{\mathrm{B}}T \ln \sum_{j \in \mathcal{R}_i} \exp[-E(j)/k_{\mathrm{B}}T], \end{aligned} \tag{5}$$

and thus apply the usual laws of thermodynamics to the system as long as it remains within the region $\mathcal{R}_i$.

For any given observation timescale t_{obs}, the configuration space of the chemical system is split into a large number of disjoint locally ergodic regions, with the remainder of the configuration space consisting of transition regions connecting the locally ergodic regions. Each such region corresponds to a kinetically stable compound of the chemical system on the timescale of observation.

Quite generally, one notes that at low temperatures the escape times from the locally ergodic regions tend to follow Arrhenius' law and are therefore controlled by energetic

barriers on the energy landscape. Thus, at very low temperatures individual local minima[1] of the energy landscape are locally ergodic and their local free energies are determined by the energy of the minimum plus the contribution of the vibrations about these minima. Usually, the regions with the lowest free energy correspond to crystalline modifications of the system, while structures containing defects are also associated with local minima but with higher energies. At elevated temperatures and on sufficiently long timescales, locally ergodic regions will typically encompass many local minima. An important class is formed by locally ergodic regions that contain all the local minima associated with, for example, the rotation of complex anions in a solid or the oscillation of individual atoms in double-well potentials, or the many local minima representing possible atom arrangements belonging to solid-solution phases.

Constructing an equilibrium phase diagram from first principles thus involves as a first step the determination of the locally ergodic regions on the energy landscape of a chemical system as a function of temperature, pressure and composition, for observational timescales of interest. Next, the local free energies of the corresponding (meta)stable compounds and phases are computed and by minimization of these free energies the thermodynamically stable phase is obtained.

To achieve this goal, we need to identify and analyze the central quantities of interest that control the dynamics on energy landscapes of chemical systems. These include special points on the landscape (Quandt, 2008) such as local minima and saddle points (in particular those connecting pairs of minima), special regions such as locally ergodic regions (Schön & Jansen, 2001), transition regions (Schön *et al.*, 2001*a*), local densities of states, and the flow of probability on the landscape with the corresponding barrier landscape consisting of (generalized) barriers (Schön *et al.*, 2003) such as energetic, entropic and kinetic barriers (Hoffmann & Schön, 2005). Finally, visualization plays an important role in the analysis of the properties of energy landscapes by representing the high-dimensional complex multi-minima landscape in a simplified fashion, where, in particular, graph-based representations (Hoffmann & Sibani, 1988; Sibani *et al.*, 1993; Schön *et al.*, 1996; Becker & Karplus, 1997; Heuer, 1997; Wales *et al.*, 1998; Krivov & Karplus, 2002; Komatsuzaki *et al.*, 2005) and plots of, for example, the energy as a function of some characteristic order parameter or some reduced set of relevant coordinates (Gower, 1966; Abagyan & Argos, 1992; Amadei *et al.*, 1993; Troyer & Cohen, 1995; Becker, 1997; Das *et al.*, 2006) play an important role (for more information, see *e.g.* Schön & Jansen, 2001, 2009). All these quantities are of interest when trying to predict not-yet-synthesized solid compounds capable of existence, and the related phase diagrams. The number of methods that have been developed to determine these crucial features of energy landscapes is very large; we refer to the literature for reviews and more detail of the algorithms involved (Schön & Jansen, 2001, 2009).

[1] Both the perfect crystalline atom arrangement and defect configurations constitute such individual local minima. Note that amorphous 'phases' correspond to multi-minima regions that are not locally ergodic but consist of nested sets of marginally ergodic regions $\mathcal{R}_i, \ldots \subset \mathcal{R}_i \subset \mathcal{R}_{i+1} \subset \ldots$, with $\ldots \tau_{\rm esc}(\mathcal{R}_{i+1}) \simeq \tau_{\rm eq}(\mathcal{R}_{i+1}) \geq \tau_{\rm esc}(\mathcal{R}_i) \simeq \tau_{\rm eq}(\mathcal{R}_i) \ldots$.

3. Global landscape exploration techniques based on random walks

We have employed a variety of exploration techniques to identify local minima and locally ergodic regions in general, and to investigate the barrier structure of the landscape of chemical systems. All the algorithms used are based on random walks: stochastic simulated annealing and its variants, the threshold algorithm, the deluge algorithm, and the ergodicity search algorithm. In general, algorithms that employ random walkers to explore energy landscapes are based on the following four fundamental features: (i) a set of random walkers that can be (*a*) non-interacting, (*b*) interacting and/or (*c*) learning from each other; (ii) a configuration (or solution) space $\mathcal{S} = \{\mathbf{x}\}$ together with an energy (or cost) function $E(\mathbf{x})$ that can be unchanged or evolving as the algorithm proceeds; (iii) a moveclass (or neighborhood) $\mathcal{N}(\mathbf{x})$ which gives for each state $\mathbf{x}$ the neighboring states that can be accessed with a certain probability by the random walker if it is at state $\mathbf{x}$. This moveclass can remain unchanged or evolve as the algorithm proceeds; (iv) an acceptance criterion according to which the walker makes the move to the neighbor state selected. Again, this criterion can (and often does) vary during the run.

The prototype of such an algorithm is the so-called Monte Carlo Metropolis algorithm (Metropolis *et al.*, 1953) describing a single walker at a constant temperature T, which employs the Metropolis acceptance criterion. At the beginning, a starting point $\mathbf{x}_0$ for the walker is chosen, either at random or according to some deterministic scheme. The move from the ith to the $(i+1)$th position of the walker takes place as follows. From the neighborhood $\mathcal{N}(\mathbf{x}_i)$ of the current state $\mathbf{x}_i$ we select, at random with probability according to the moveclass, a target state $\mathbf{x}_{\rm target} \in \mathcal{N}(\mathbf{x}_i)$. Next, we compute the difference in energy between these two states, $E(\mathbf{x}_{\rm target}) - E(\mathbf{x}_i)$. If $E_{\rm target} \leq E_i$, the move is accepted. If $E_{\rm target} > E_i$, a random number $0 \leq r \leq 1$ is generated. If now

$$\exp\left[-(E_{\rm target} - E_i)/T\right] \geq r, \tag{6}$$

then the move is accepted, *i.e.* $\mathbf{x}_{i+1} = \mathbf{x}_{\rm target}$. Otherwise, the walker stays at $\mathbf{x}_i$, *i.e.* $\mathbf{x}_{i+1} = \mathbf{x}_i$. This procedure is repeated until the maximal number of steps $N_{\rm max}$ has been performed, and the full trajectory $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N_{\rm max}}\}$ has been obtained.

Under certain conditions the trajectory covers the space $\mathcal{S}$ ergodically, but for complex multi-minima systems $N_{\rm max}$ is usually much too short to yield global ergodicity, especially if T is varied or the landscape evolves during the run (Geman & Geman, 1984; Schön, 1997; Salamon *et al.*, 2002).

3.1. Global optimization techniques: simulated annealing and related algorithms

The most common generalization of the Metropolis algorithm consists of varying the temperature during the run. The so-called simulated-annealing algorithm (Kirkpatrick *et al.*,

1983; Czerny, 1985) proceeds by slowly lowering the temperature, thus moving the walker to (on average) states with lower and lower energy. The expectation is that if one proceeds slowly enough the walker will at the end of the run have reached the global minimum of the energy landscape. Another well known algorithm is the stochastic quench, where one performs a Metropolis random walk at zero temperature, *i.e.* only steps that lower the energy are being accepted. Especially if one performs such stochastic quenches from high-lying states one can gain additional information beyond that obtained by performing a gradient minimization, because by repeating the quench from the same starting point for different random-number sequences one can determine whether the starting point is associated with only one or several local minima, *i.e.* whether it 'belongs' to a transition region or a single basin (Schön *et al.*, 2001*a*). Such high-lying states can either be generated at random or represent periodic stopping points along the trajectories during a Monte Carlo simulated-annealing or threshold run.

Quite generally, a great advantage of the discrete steps in stochastic simulated annealing is the large freedom in choosing a moveclass most appropriate to the type of exploration being performed. Thus, one can replace the physically realistic moveclass of moving one or a few atoms by a small amount by a more optimization-effective moveclass that allows larger changes in the atom configuration during each move, in order to explore a larger part of the landscape during, for example, a global optimization run. In this case it is sometimes efficient to combine such large moves every time with a quench.[2] Besides the moveclass, there are a number of other features that can be adjusted to increase the efficiency of the algorithm (Salamon *et al.*, 2002; Delamarre & Virot, 1998). The temperature schedule $T(n)$, where n counts the number of moves along the trajectory, can be optimized; common schedules consist of an exponential or linear decrease of temperature with n. Also quite popular are schedules involving temperature cycling (Möbius *et al.*, 1997, 2004) where the temperature periodically increases and then decreases again, and adaptive schedules (Ruppeiner *et al.*, 1991; Salamon *et al.*, 1988) that take properties of the landscape explored up to now into account.

Finally, the acceptance criterion can be modified; the most popular alternatives accept a move according to a Fermi-function-like distribution [the so-called 'fast' annealing (Szu & Hartley, 1987)], the Tsallis distribution (Tsallis, 1988; Tsallis & Stariolo, 1996) or based on a temperature-dependent acceptance threshold (Dueck & Scheuer, 1990).

3.2. Lid-based algorithms for global optimization and studies of barrier structure

A rather different line of attack is taken by the lid-based methods for continuous energy landscapes. Characteristic is the presence of an energy lid, which denotes the maximal energy the walker is permitted to have, *i.e.* any move to a neighboring configuration with an energy exceeding this value is rejected. Typically, any move that leads to a state with energy below this lid is accepted, as if $T = \infty$. However, for special purposes we have also performed runs below the lid at constant temperature, usually in order to study the temperature dependence of the probability flow restricted to a pocket in the landscape.

One example of a lid-based global optimization technique is the deluge algorithm (Dueck, 1993), where the energy lid that must not be crossed during the random walk is slowly lowered from very high lid values, squeezing the walker into a low-energy minimum. Another approach employs the threshold algorithm (Schön *et al.*, 1996) that was originally developed as an implementation of the lid algorithm (Sibani *et al.*, 1993, 1999; Sibani & Schriver, 1994) for the study of the barrier structure of continuous landscapes.

Here, one starts with a set of local minima that had been identified, for example, by stochastic quenches or short simulated-annealing runs. Using each of these minima as starting points, for a sequence of fixed energy lids, the walker is allowed to move below the lid with every move accepted as in the deluge algorithm, and one checks periodically whether new local minima have been reached by performing several quench runs from stopping points along the trajectories below the current energy lid. In this fashion, one can gain an estimate of the barrier height between neighboring minima by registering for which lid value the barrier to a neighboring minimum can be crossed for the first time. Furthermore, the likelihood of reaching neighboring minima, or of returning to/staying inside the starting minimum basin gives us a measure of the probability flow as a function of energy slice, and yields a measure of the entropic barriers surrounding a minimum (Schön *et al.*, 1996; Wevers *et al.*, 1999). This procedure is repeated for all the original starting minima plus all the local minima that are being identified during the threshold run itself.

Besides the probability flow, the density of states is sampled as a function of energy and lid value. By performing not one but many quenches at each stopping point along the trajectories, one can determine the size of the transition regions by identifying the so-called characteristic regions (Schön *et al.*, 2001*a*) as a function of energy slice. These characteristic regions are defined *via* the probability of reaching a set of minima from a given stopping point along a trajectory when one performs many stochastic quenches from the same stopping point.

A great advantage of these lid-based algorithms when employed as global optimization tools is the fact that they explore the landscape in a way complementary to the standard simulated-annealing methods: the latter start from high energies and proceed downhill while staying approximately in local equilibrium, and thus they enter basins containing local minima based on the slope, *i.e.* the growth rate, of the local density of states. In contrast, the lid-based methods enter local minimum regions based on the size of the basin at a given energy lid. As a consequence, the distribution of minima found using these two methods is going

[2] This scheme is often called basin hopping (Wales & Doye, 1997; Iwamatsu & Okabe, 2004).

to be different, and minima overlooked using one class of approaches can be identified by the other type of search method.

3.3. Identifying locally ergodic regions at non-zero temperatures

At non-zero temperatures the locally ergodic regions can contain not only one but several or many local minima. In order to identify such regions, one proceeds either in a direct fashion *via* (long) constant-temperature Monte Carlo simulations, or indirectly by analyzing the set of minima found. The first approach is more likely to deliver regions encompassing few minima that are close neighbors on the landscape, while the second one is more suited when dealing with large multi-minima regions.

3.3.1. Ergodicity search algorithm. We have developed the so-called ergodicity search algorithm (Schön, Čančarevič *et al.*, 2008), where one registers the fluctuation of indicator variables, for instance the potential energy or the radial distribution function, within time windows during the long (Monte Carlo) simulations. If the average value of these variables jumps between two windows by more than the fluctuation, this suggests the existence of a new locally ergodic region. Next, swarms of short (Monte Carlo) simulations starting from points along the trajectory in the time window are employed to verify whether the system is in local equilibrium in this region. Finally, long simulations for a number of temperatures are used to measure the probability flow from the region and thus the escape time. Unsurprisingly, searching for locally ergodic regions in this fashion is quite expensive computationally.

3.3.2. Identifying multi-minima locally ergodic regions *via* structure families. Since straightforward Monte Carlo simulations that could directly yield locally ergodic regions encompassing many local minima are usually too time-consuming even with empirical potentials, one needs to proceed more indirectly when trying to identify such regions. In many chemical systems where such large regions are relevant, one can take advantage of the fact that these systems exhibit so-called controlled disorder, *i.e.* the minima show a high degree of structural similarity and belong to so-called structure families, where only some degrees of freedom are disordered while the remaining ones are essentially unchanged.

Thus, after the determination and local optimization of the minima, a structural analysis of the minimum configurations follows, to decide whether we are dealing with solid solutions or ordered crystalline modifications. The crucial issue is whether such structure families exist that fulfil three conditions (Schön & Jansen, 2005; Schön *et al.*, 2006): (i) for a given composition the energies of the members of the family should be very similar; (ii) there should exist no deeper-lying minima that do not belong to the family; and (iii) the same structure family should exist for all (or at least for a wide range of) compositions. If that is the case, the union of these local minima can be treated as a large locally ergodic region for each composition, and the free energy of this solid solution phase contains an entropy of mixing which favors the solution over ordered crystalline compounds which correspond to a single minimum basin on the energy landscape. In the case of intermetallic solutions, the minimum structures belonging to such a structure family can be identified by the fact that structurally they possess the same set of sublattices on which the atoms can be placed more or less randomly. In the case of ionic solid solutions, the same overall cation–anion superstructure is present for many compositions, and the different types of cations and/or anions are randomly distributed over the cation or anion positions in the superstructure, respectively.

4. Modular approach to exploring energy landscapes for rational synthesis planning

4.1. Modular approach

For the practical implementation of the procedure, outlined in §2, for the identification of locally ergodic regions and the computation of the phase diagram of a chemical system, we have developed a modular approach (Schön & Jansen, 1996). The first step is the determination of as many as possible of the relevant local minima on the energy landscape as a function of atom positions and cell parameters, using various global optimization techniques. This stage is repeated for many different pressure values and numbers of atoms in the variable simulation cell. In addition, locally ergodic regions at non-zero temperatures are identified using the ergodicity search algorithm.

Next, these configurations are analyzed with respect to their electronic, geometrical and topological properties, such as typical oxidation state of the ions, translational and rotational symmetries of the structure, or the occurrence of typical local coordination environments.[3] Furthermore, duplicates are eliminated from the set of candidates before the next steps are taken.

The third step consists of the determination of the energetic and entropic barriers that surround the local minima, in order to gain estimates on the kinetic stability of the structure candidates. As a side effect, additional structure candidates are identified and the local density of the minimum regions is sampled.

In a fourth step, the candidates are locally optimized on an *ab initio* level, in order to be able to rank them by energy. As part of this process, the $E(V)$ curves are computed for every hypothetical modification, and possible transitions between them as a function of pressure can be identified. Again, the structural features of the optimized structures are analyzed and the space group of each candidate is determined.

In a fifth step, the phonon spectrum is derived (in the harmonic approximation), and the free energy is computed for

[3] If appropriate, typical local coordinations can be used to define primary building units that are employed in a second round of global optimizations. Similarly, the global searches can be repeated using ions with fixed charge instead of atoms with variable charge.

each candidate. Furthermore, additional physical quantities such as the band gap can be computed at this stage.

A sixth step follows, where the set of local minima is analyzed with respect to the existence of structure families that might indicate the possibility of controlled disorder in the system and thus the existence of high-temperature phases with non-negligible configurational entropy contributions to the local free energy.

Finally, all the results are put together to derive the phase diagram of the chemical system, including metastable phases.[4]

4.2. Choice of energy function during the global exploration stage

The first step in trying to identify structure candidates consists of the choice of a cost function whose minima (or sets of minima) correspond to chemically and physically reasonable compounds. In principle, not only the atoms' positions and their electronic states but also the shape, size and content of the simulation cell can be varied as part of the global search. Thus, the general cost function is the thermodynamic potential of the grand isobaric ensemble (Schön & Jansen, 1996)

$$C = E + pV - \sum_{i=1,\,N_{\rm spec}} \mu_i N_i, \tag{7}$$

where $\sum_{i=1,\,N_{\rm spec}} N_i = N$, and $N_{\rm spec}$ is the number of different types of atoms involved. Ideally, the energy E is computed on an *ab initio* level; however, owing to the large amount of energy calculations required for the global searches, one commonly employs an empirical energy function. One potential difficulty encountered is choosing an appropriate value for the chemical potential μ_i of species i; typically, we use the standard enthalpy of formation per atom of the element. However, experience has shown that it tends to be more efficient to keep the composition and the number of atoms fixed during a single global optimization run, and to repeat the runs for different compositions and numbers of atoms.

4.2.1. Empirical potentials. When employing an empirical energy function, we usually choose a simple but robust potential that has proven to be quite efficient for ionic systems,

$$E = \sum_{i=1,\,N} E_{\rm ion}(i) + \sum_{i,\,j} V_{ij}, \tag{8}$$

where V_{ij} is a Coulomb-plus-Lennard-Jones potential depending on the charges q_i and the distance between atoms $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$,

$$V_{ij} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \varepsilon_{ij}\left\{\left[\frac{r_{\rm ion}(i)+r_{\rm ion}(j)}{r_{ij}}\right]^{12} - \left[\frac{r_{\rm ion}(i)+r_{\rm ion}(j)}{r_{ij}}\right]^{6}\right\}. \tag{9}$$

Note that the ionic radius depends on the current ionization state of the atom and thus its value is part of the optimization. The values of $r_{\rm ion}(i)$ are averages taken from databases with ionic structures, and are essentially the only experimental input into the potential.

The great advantage of this type of potential is that it is flexible, fast and unprejudiced with respect to the possible structures and local environments. Nevertheless, it has proven to be very useful to repeat the global searches on such empirical energy landscapes with slightly varied parameters for $r_{\rm ion}$ and ε_{ij}, in order to be able to access possible outliers among the candidates on the energy landscape. Of course, it is clear that this energy function will be less appropriate for intermetallic systems once one can no longer clearly assign the ionization states in the compound.[5]

An alternative is to use potentials that have been fitted to *e.g. ab initio* calculations. Of course, this also corresponds to a restriction of the available landscape, and their use can be problematic if the potentials are too much tied to the structures that have been used to fit the parameters. Thus, there might not be enough flexibility built into the system to exhibit the full spectrum of chemically reasonable modifications. On the other hand, the potential often describes the local environments of the atoms in most modifications quite well, and the global optimization more quickly reaches the deep-lying minima.

4.2.2. *Ab initio* energies. The alternative to the use of empirical potentials is to employ *ab initio* energy calculations at the global optimization stage. In order to be able to do this, it is necessary to speed up the calculation without losing the advantages of *ab initio* calculations.

For the *ab initio* energy calculations we have employed the code *CRYSTAL2006* (Dovesi *et al.*, 2006). This code uses a local Gaussian basis set where the basis functions are centered at the positions of the nuclei. The speed-up of the energy calculations can be achieved in several ways. Firstly, by reducing the number of matrix elements (integrals), which can be achieved by selecting less strict tolerances, which means that more integrals are discarded, or evaluated on a lower level of accuracy.

Another way to accelerate the calculations is by reducing the basis set, for example one may consider omitting polarization functions. Similarly, using less diffuse exponents helps both to enhance the numerical stability and to speed up the calculations. Other modifications that can speed up the computation are, for example, the use of fewer **k**-points or of less strict thresholds for the convergence of the self-consistent field cycles. Finally, the quality of the density-functional grid

[4] As described here, we have excluded the part of the energy landscape where the liquid state and the structurally amorphous states reside. To explore these regions, one needs to employ larger simulation cells than the ones typically used for the identification of crystalline structure candidates. Of course, unless these phases are included, the part of the phase diagram that can be computed does not include the temperature region where melting is expected to occur in the system.

[5] This breakdown of the applicability usually shows up in rather unusual features of the local minima found: for example, the atoms prefer to stay in the neutral state forming just various dense packings, or auto-ionization takes place, *i.e.* it appears to be energetically favorable for two metal atoms of the same type to form a M^+–M^- pair.

(*i.e.* the number of grid points) may be reduced, when the global search is performed on the density-functional level.

An additional complication one encounters during global explorations of energy landscapes of chemical systems is the necessity to converge the *ab initio* calculations for random atom arrangements, often far away from the experimental configuration, which is not a trivial task. Note that the main difficulty is the convergence of the atom configurations at the beginning of the simulated-annealing run, whereas the structures at later stages and after the simulated-annealing run has finished are easy to converge. The main reason is that the initial configurations resemble a gas, and thus the band structure is similar to that of localized electrons, with almost no dispersion, and with a much smaller gap than the final structure.[6]

Concerning the choice of the functionals, it turned out that, for insulators, Hartree–Fock calculations facilitate convergence at even the most unusual geometries. This is due to the large band gap typical for this method, whereas calculations with the local density approximation may exhibit severe convergence problems.

4.3. Global and local optimization

If one searches for crystalline modifications using one of the global optimization methods described above, one defines the system as a set of N atoms that are located inside a periodically repeated simulation cell, where both the atom positions and charges, and the cell parameters can be varied (Schön & Jansen, 1996). In the case of a cluster, one nearly always employs just a fixed large simulation cell, and varies only the atom positions. Concerning the moveclass, the ratio between atom movements and exchanges, and cell-parameter variations is typically about 4:1.

Two consecutive minimizations are performed after the simulated-annealing stage: a quench with the same moveclass and the same level of accuracy for the energy calculations as in simulated annealing, and subsequently a local optimization, usually with the *CRYSTAL2006* code at a high level of accuracy. Here, we use analytical gradients for the nuclear positions (Doll *et al.*, 2001; Doll, 2001) and the unit cell (Doll *et al.*, 2004, 2006), and the full geometry optimization as implemented in the present release (Civalleri *et al.*, 2001; Dovesi *et al.*, 2006), sometimes together with a heuristic algorithm (Čančarević *et al.*, 2004; Schön, Čančarević & Jansen, 2004) that employs an automated nested minimization scheme.

The optimization is usually performed both with the Hartree–Fock approximation and on the level of several density functionals. This comparison of the results for different *ab initio* methods is necessary in general, in order to be able to judge the robustness of the candidates found: since we predict new structures in chemical systems where no compounds have been synthesized so far, the modifications found on the basis of the global optimization cannot be compared with experiment!

4.4. Analysis tools: *SFND*, *RGS*, *CMPZ*

An important step in the modular approach is the analysis of the results of the global optimization runs. Since the searches are performed for variable simulation cells with space group $P1$, all the local minimum configurations will be recorded without any symmetry information. Furthermore, owing to numerical aspects of the algorithms, all these minima show slight minute deviations from the (highly) symmetric atom arrangements that correspond to the exact location of the minimum. Finally, the structures obtained are usually not given with the standard unit cell. To address this issue, we have developed two algorithms, *SFND* (Hundt *et al.*, 1999) and *RGS* (Hannemann *et al.*, 1998), which allow us to determine the symmetries a given periodic structure exhibits within a prescribed set of tolerances, to idealize the cell parameters and atom positions to be in agreement with the symmetries detected, and finally to deduce the correct space group and transform the structure to standard setting.

A second issue is the need to eliminate duplicates among many of the local minima produced by the optimization, to compare the structures found with already known structure types listed in the Inorganic Crystal Structure Database (ICSD), and finally to identify structure families by comparing, for example, the cation–anion superstructures in multi-cation/anion compounds. To deal with these tasks, we have developed the *CMPZ* algorithm (Hundt *et al.*, 2006), which allows us to compare two arbitrary periodic structures by generating a mapping of the two infinite periodic atom arrangements onto one another. As a criterion for similarity, we measure the deviations between the cell parameters of the appropriately transformed cells together with the deviations of the atom positions within these cells. We note that this is a geometric criterion for similarity, not a topological one. These algorithms have been implemented in the structure analysis program *KPLOT* (Hundt, 1979), which can be employed in an automated script-driven fashion.

5. Theoretical applications

The examples one finds in the literature under the heading of 'structure prediction' can be divided into three different classes: on the one extreme is the structure determination where structural information, typically a unit cell and its content, often together with a powder diffractogram, is known from experiment. On the other extreme is the unrestricted structure prediction, where only the stoichiometry but neither the unit cell nor the number of formula units is known. And if we only know the general chemical system, *i.e.* the composition is not fixed either, structure prediction becomes equivalent to the prediction of the phase diagram of the system.

Between these two extremes lies the case of restricted structure prediction, for example the prediction of structures

[6] An important advantage of a local basis set is that fewer integrals need to be computed if the atoms are far apart. In particular, during the initial steps of the search, where a large volume is enforced, the energy calculations are much faster than those at smaller volume.

in systems where certain structural elements or local environments of atoms are pre-defined or assumed at the outset, such as primary and secondary building units (Mellot-Draznieks *et al.*, 2002).[7] Here, one employs some general chemical constraints that are quite plausible in the system under investigation but are, strictly speaking, not admissible in a true structure prediction. Nevertheless, we count these restricted searches still as part of structure prediction, since these restrictions are made based on general chemical experience, in contrast to the prescription of the unit cell which implies that at least some powder diffraction data of the compound are available.

Finally, there is the theoretical treatment of syntheses and phase transitions. Here, we will present examples for the optimal control of phase transitions.

5.1. Structure prediction

5.1.1. Alkali metal halides. One of the first ionic systems whose energy landscape has been investigated (Schön & Jansen, 1995) in detail using simulated annealing without recourse to experimental data is NaCl. A large number of local minima were found on the empirical energy landscape, and the global minimum of the landscape corresponded to the experimentally observed rocksalt structure. The structures of most of the energetically low-lying minima could be identified with typical *AB*-structure types like NiAs, PtS, CsCl or sphalerite. However, one deep-lying local minimum, denoted $Na^{[5]}Cl^{[5]}$ (the so-called 5–5 structure type), exhibited a structure type previously unknown in ionic systems. Here, Na^+ and Cl^- were coordinated by Cl^- and Na^+, respectively, in a trigonally bipyramidal fashion, resulting in a topology that resembled that of hexagonal BN.

The energy barrier stabilizing this structure was only moderately high (∼0.01 eV per atom), suggesting that the structure might be difficult to synthesize with traditional solid-state synthesis methods in the NaCl system. Thus it came as a pleasant surprise when this new predicted structure type was found experimentally (Haas & Jansen, 1999) as the aristotype of Li_4SeO_5, where Li and Se occupy the Na positions and O the Cl positions in the $Na^{[5]}Cl^{[5]}$ structure, respectively. By now, this structure type has also been observed during the growth of ZnO films (Claeyssens *et al.*, 2005).

Analogous global optimizations have been performed since for all 20 alkali halides (Čančarević *et al.*, 2008) for a wide range of pressures. Similar to the case of NaCl, many possible modifications were found that included both well known *AB* structure types (rocksalt, NiAs, wurtzite, sphalerite, 5–5, CsCl *etc.*) and previously unknown structures.

[7] Primary building units are defined as groups of atoms that exist as independent groups within a solid or a molecule, and thus the composition of the system does not change unless one explicitly changes the number of atoms/building units present. In contrast, secondary building units are groups of atoms which are typically connected by *shared* atoms. Thus, for example, the stoichiometry of the final structure is not fixed from the outset but changes whenever two atoms belonging to two different secondary building units merge in order to establish a connection.

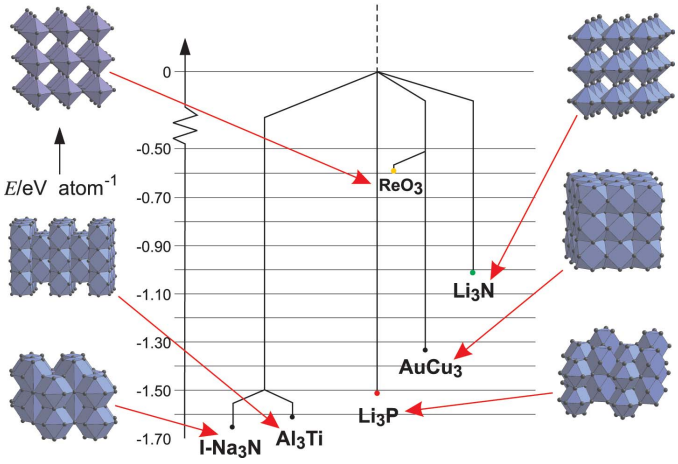

Figure 1
Excerpt of the tree graph of the energy landscape of Na_3N on an empirical energy level at standard pressure, depicting some of the most important local minima (Jansen & Schön, 1998; Schön *et al.*, 2000, 2001*b*; Fischer & Jansen, 2002*b*). I-Na_3N corresponds to a strongly distorted Li_3Bi structure type with 12(+2)-fold coordination of the nitrogen atoms by sodium atoms. Experimentally, the ReO_3 type (Fischer & Jansen, 2002*b*), the Li_3N, the Li_3P and the Li_3Bi structure types (Vajenine *et al.*, 2008, 2009) have all been synthesized, the latter three using high-pressure experiments starting from the ReO_3 type. Furthermore, at intermediary pressures another modification exhibiting the YF_3 type was observed (Vajenine *et al.*, 2008, 2009) that resembles several structures found as local minima on the enthalpy landscapes of the alkali nitrides.

Finally, simulated annealing was employed to find the minima of LiF on the *ab initio* energy landscape at standard pressure, where both the Hartree–Fock approximation and density functionals were used to compute the energy (Doll *et al.*, 2007). The same relevant minima as with the empirical potential were obtained, including the rocksalt, the wurtzite and sphalerite, the NiAs, and the 5–5 structure types. This study served both as a valuable validation of the many landscape explorations based on empirical potentials and as a proof-of-principle for the feasibility of global stochastic explorations on *ab initio* energy surfaces.

5.1.2. Sodium nitride Na_3N. Another deceptively simple chemical system where a successful synthesis followed the prediction is Na_3N.[8] In several studies (Jansen & Schön, 1998; Schön *et al.*, 2000, 2001*b*), the enthalpy landscapes of all alkali nitrides M_3N (M = Li, Na, K, Rb, Cs) were explored with simulated annealing and the threshold algorithm for a wide range of pressures. This resulted in a large number of structure candidates, including, for example, the Li_3N, the Li_3P, the Li_3Bi, the $AuCu_3$, the Al_3Ti, the ReO_3 and the UO_3 structure types, plus many previously unknown structure types. Fig. 1 shows a part of the tree graph for the empirical energy landscape of Na_3N containing some of the most important local minima. *Ab initio* calculations using the Hartree–Fock approximation suggested that for Na_3N the most likely

[8] The inability to synthesize any compound of this composition had for many decades been hailed as a blatant violation of the homologue rule, since Li_3N can be synthesized directly from the elements at ambient conditions (Brese & O'Keeffe, 1992).

candidate would be the Li_3P type, followed by the Li_3N and the ReO_3 type, with the Li_3Bi type expected at high pressures.

5.1.3. Alkali metal orthocarbonates $M_4(CO_4)$, with M = Li, Na, K, Rb, Cs. The existence of esters of the hypothetic acid $H_4(CO_4)$ suggests that its salts, the orthocarbonates, should also be accessible (Mellot-Draznieks *et al.*, 2002). A promising approach to synthesize, for example, the alkali metal carbonates would be to apply high hydrostatic pressures during syntheses to the phase equilibria $M_2O + M_2(CO_3) \rightleftharpoons M_4(CO_4)$ (M = alkali metal). Since the hypothetical orthocarbonate would compete with high-pressure phases of the corresponding regular carbonates plus oxides, it is necessary to study the parts of the enthalpy surfaces of the M/C/O system with composition $M_4(CO_4)$, M_2O and $M_2(CO_3)$ for many different pressures. In this way, one can theoretically establish the range of (thermodynamic) stability of the orthocarbonate phase *versus* the decomposition into the corresponding oxide and carbonate as a function of applied pressure.

To achieve this, the enthalpy landscapes of M_2O (Čančarević *et al.*, 2006*a*), $M_2(CO_3)$ (Čančarević *et al.*, 2006*b*) and $M_4(CO_4)$ (Mellot-Draznieks *et al.*, 2002; Čančarević *et al.*, 2007) were investigated for many different pressures using simulated annealing and an empirical Coulomb-plus-Lennard-Jones potential. In a first round of global optimizations, individual metal, carbon and oxygen atoms were used to describe atom configurations. After it turned out that the minimum configurations contained isolated trigonal CO_3 and tetrahedral CO_4 units, the latter at high pressures, a second round of global optimizations was performed, where CO_3 and CO_4 units were employed together with the metal atoms. All structure candidates in these systems were locally minimized on the *ab initio* level in the Hartree–Fock approximation. Next, for each pressure, the thermodynamically stable modification was determined (together with the transition pressures among the various modifications for each of the individual systems), and the enthalpy of $M_4(CO_4)$ was compared with that of $M_2O + M_2(CO_3)$ as a function of pressure (Čančarević *et al.*, 2007). It was found that for all alkali metals there should exist thermodynamically stable orthocarbonates at sufficiently high pressures, with the most easily accessible candidates being $K_4(CO_4)$ and $Rb_4(CO_4)$ where the phase equilibrium is expected to switch to the orthocarbonate from the oxide-plus-carbonate in the range 20–30 GPa.

5.1.4. Global exploration performed on the *ab initio* level: boron nitride and calcium carbide. As we have already described in §5.1.1, LiF was chosen as our first example (Doll *et al.*, 2007) to prove the feasibility of using *ab initio* calculations during the simulated-annealing stage. A significant extension was the application of simulated annealing to structure prediction for boron nitride (Doll *et al.*, 2008), where several kinds of, mostly covalent, contributions to the total energy are present. The global searches were performed on both the Hartree–Fock and the density-functional level, using four formula units. The starting configurations were random atom arrangements in a large unit cell resembling a gas phase. The searches with the Hartree–Fock approximation proved to be considerably faster (by a factor of about four) than the ones using density-functional theory (DFT)-based energy calculations, because the former could better handle the task of converging the electronic structure calculations at low density for random structures. After the low-lying minima were identified, local optimizations using standard *ab initio* tolerance parameters followed, where both Hartree–Fock and density-functional methods were employed.

The BN system is particularly fascinating as a test system because the experimentally observed modifications include both layered structures (hexagonal BN) and three-dimensional networks (wurtzite and sphalerite type). In the global optimizations, all experimentally observed structure types were indeed found. In addition, several new modifications were predicted such as layered structures but with a stacking order different from the experimentally observed structure h-BN. Two other very interesting new low-energy modifications consist of three-dimensional B–N networks, showing the β-BeO structure and the Al partial structure in $SrAl_2$.

The mixed covalent–ionic system CaC_2 currently under investigation represents another challenge, since it would be very difficult to model a system that tends to form complex anions of varying size [preferably $(C_2)^{2-}$ ions] with empirical potentials while still retaining the freedom to break up such C_2^{2-} units. Again, up to four formula units per simulation cell were employed, and the optimizations followed the same procedure and approximations as in the case of boron nitride and lithium fluoride. The calculations indicate so far that, besides the experimental structures, a new high-pressure modification should exist (Kulkarni *et al.*, 2010).

5.1.5. Free-energy landscape as a function of temperature and pressure for the example of SrO. The free-energy landscape of SrO was constructed by combining runs with the ergodicity search algorithm and the threshold algorithm for a global exploration of the energy landscape (Schön, Čančarevič *et al.*, 2008), where an empirical Coulomb-plus-Lennard-Jones potential served as an energy function. After a preliminary global optimization of the landscape using simulated annealing, the local minima identified during the global optimization were used as starting points for a large number of threshold runs at several different pressures (Schön, 2004). This yielded an overview over both the low-lying local minima on the enthalpy landscapes and the barriers separating the different modifications. Next, the ergodicity search algorithm (ESA) was applied at standard pressure, and for a large number of different temperatures, in order to identify possible high-temperature phases. The potential energy and the radial distribution function served as indicator variables. All the structure candidates found with ESA turned out to be associated with individual local minima that had already been detected during the threshold run phase. Finally, the appearance of the melt phase was observed by checking the stability of the underlying crystalline lattice of the rock-salt-type modification (the thermodynamically stable solid modification of SrO at standard pressure, both according to the experiment and the calculations) during very long Monte

Carlo simulations for large simulation cells as a function of temperature.

In the fourth step, the free energies of the structure candidates found were computed in the quasi-harmonic approximation on the empirical potential level, and also on the DFT-B3LYP level. Combining these free energies as a function of temperature with the energy barriers computed *via* the threshold algorithm resulted in the free-energy landscape shown in Fig. 2.

5.1.6. Phase diagrams of the quasi-binary mixed alkali halides. As discussed in §3.3, the crucial issue for the existence of solid-solution phases is whether so-called structure families exist among the minima observed for many different compositions which have essentially the same energy for a given composition (Schön & Jansen, 2005; Schön *et al.*, 2006). Once this has been established, one can compute the Gibbs free energy *via* the convex hull method (Voronin, 2003) using a combination of the ideal entropy of mixing and a Redlich–Kister polynomial *ansatz* (Redlich & Kister, 1948) for the enthalpy of mixing.

This approach has been applied to about 20 different quasi-binary alkali halide systems (Schön *et al.*, 2006, 2007; Pentin *et al.*, 2007, 2008; Schön, Pentin & Jansen, 2008), where the focus was on the low-temperature region of the phase diagram since the solid–liquid region was already known experimentally. The global landscape exploration for many different compositions was performed using simulated annealing and a Coulomb-plus-Lennard-Jones potential for the energy evaluation, followed by local optimizations using both Hartree–Fock and DFT methods. In all cases the calculations correctly predicted whether a solid solution or ordered crystalline modifications were thermodynamically stable, and for those systems where the miscibility gap had been measured the computed binodal was in good quantitative agreement with the experimental data (Sangster & Pelton, 1987) (the error in the critical temperatures of the computed miscibility gaps was estimated to be about ±100 K), as is shown in Fig. 3 for the system NaBr–LiBr (Schön *et al.*, 2006; Doornhof *et al.*, 1984). For those systems where crystalline modifications were predicted to be thermodynamically stable, these agreed with those already known from experiment, and several additional stable and metastable compounds could be predicted (Pentin *et al.*, 2007).

Such global landscape explorations as a function of composition had also been performed in the 1990s for quasi-binary systems such as Ca_2Si–$CaBr_2$ (Putz *et al.*, 1999*a*) and MgF_2–MgO (Putz *et al.*, 1998), where solid solutions are not expected to occur. In these two systems a number of (meta)-stable phases were predicted, with a very promising one found

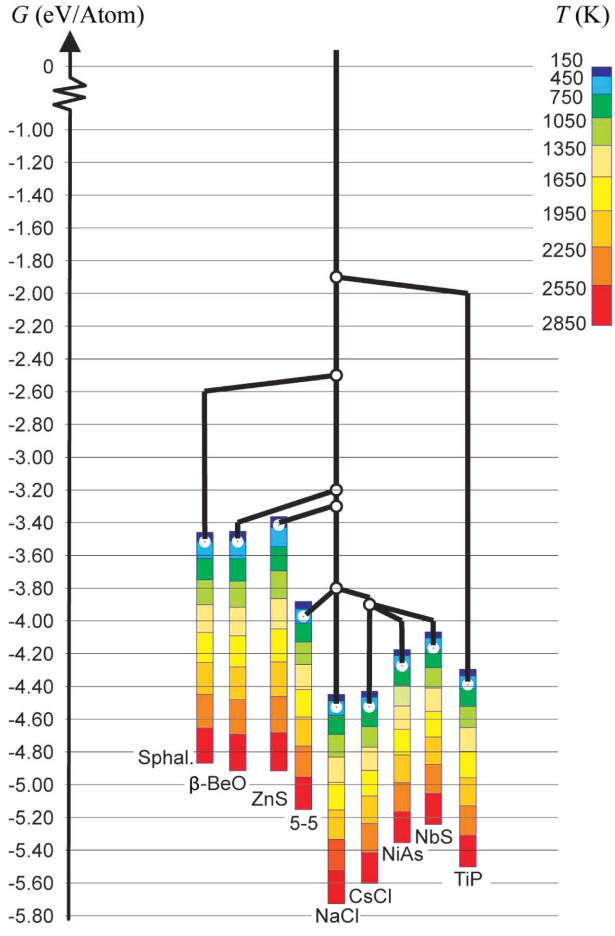

Figure 2
Free-enthalpy landscape of SrO at p = 0 GPa for eight different temperatures (T = 150, ..., 2850 K) (Schön, Čančarevič *et al.*, 2008) computed using global landscape explorations followed by free-energy calculations in the quasi-harmonic approximation on the empirical potential and *ab initio* level. The energetic contributions to the barriers stabilizing locally ergodic regions exhibiting different structure types are given by the energy difference between the minima (black circles) and transition regions (white circles). Entropic barrier contributions (for a typical example see, for example, Wevers *et al.*, 1999) are not shown to avoid overloading the figure.

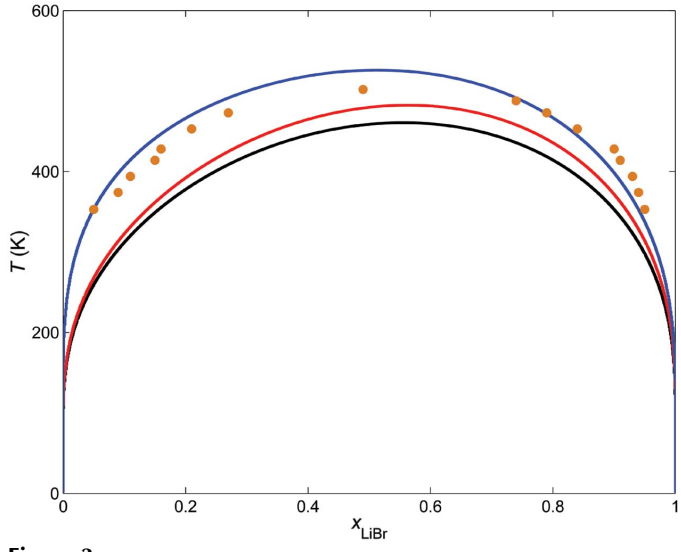

Figure 3
Low-temperature region of the phase diagram for the system NaBr–LiBr showing the miscibility gap in the system (Schön *et al.*, 2006). The gap was computed using global landscape explorations followed by the determination of free enthalpies employing both Hartree–Fock (black curve) and DFT–B3LYP (red curve) calculations. The blue curve is a fit to experimental data (Sangster & Pelton, 1987); the yellow dots are experimental data points (Doornhof *et al.*, 1984).

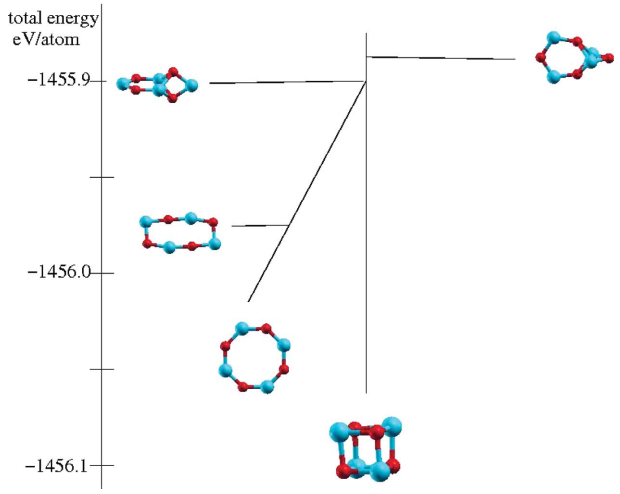

Figure 4
Tree graph for the *ab initio* energy landscape of the $(LiF)_4$ cluster generated using the threshold algorithm.

at the 1:1 composition of the $CaSi_2$–$CaBr_2$ system, which was computed to be thermodynamically stable on the Hartree–Fock level.

5.1.7. LiF clusters. When trying to understand the existence of the many feasible modifications of crystalline solids, and trying to develop possible synthesis routes, it is very useful to study the energy landscape of small- and medium-size clusters with the same chemical composition. Thus we have performed extensive global explorations of the *ab initio* landscape of clusters in the LiF system up to the size $(LiF)_8$ (Doll *et al.*, 2010).

Going beyond the search for local minima, a major challenge is the calculation of the barriers on the landscape on the *ab initio* level with the threshold algorithm. We have studied the landscape of the $(LiF)_4$ cluster using the threshold algorithm and we have obtained the tree graph shown in Fig. 4.

5.2. Structure prediction employing structural restrictions: complex ions as primary building units

One of the most important types of structural restrictions on the allowed configuration space during structure predictions is the use of primary building units, *i.e.* groups of atoms that exist as individual groups of atoms in the final structure (*e.g.* forming a complex ion or a whole molecule). One drawback is that one cannot *a priori* correctly assign the charge distribution of the building unit when computing the total energy of the test configurations during the random walk. Unless one can perform the calculation with *ab initio* energies or there are very strong arguments favoring a particular charge distribution, experience has shown that one should repeat the global searches for different charge distributions in order to ensure that one does not overlook important structure candidates.

As an example, we present a study of the energy landscape of KNO_2 (Schön & Jansen, 2001). Here, a NO_2^- building unit was employed during simulated-annealing runs, with geometrical data taken from compounds listed in the ICSD. The charge distribution of the building unit was varied from $q(N) = +3$ and $q(O) = -2$ to $q(N) = -1$ and $q(O) = 0$. It was found that the most prominent structure candidates exhibit a distorted rocksalt structure if one considers only the centers of mass of the NO_2 groups and the potassium atoms. All these minima taken together constitute a structure family which at elevated temperatures forms the basis of a large locally ergodic region that corresponds to a high-temperature phase of the system. Such an 'average' rocksalt structure is also observed experimentally at room temperature for KNO_2, where one assumes either a positional or rotational disorder of the NO_2 groups to be present (Solbakk & Stromme, 1969; Onoda-Yamamuro *et al.*, 1998). Note that one will never observe such thermal disorder directly in the global optimizations, since the cost function always refers to $T = 0$. Thus, the structures one finds are always low-temperature structures, which in this case correspond to some low-symmetry modification of the highly symmetric high-temperature structure. To identify the high-temperature phase directly, one needs to perform long-time molecular dynamics or Monte Carlo simulations, as was done by Duan *et al.* (2001), or employ the ergodicity search algorithm. However, it is possible to study the activation barriers of the rotation of the NO_2 units around various axes of the unit, and one finds that two barriers of about 10 K and about 100 K, respectively, appear to dominate the dynamics (Schön & Jansen, 2005; Schön, Salamon & Jansen, 2010). This suggests that at least one intermediary structure with limited rotational freedom of the NO_2 units should exist between the ordered global minimum and the freely rotating high-temperature structure. In the experiment, both a low-temperature structure (in space group $P2_1/c$) corresponding to one of the local minima found on the energy landscape and a structure with the NO_2 units rotating along the threefold axis of the crystal (space group $R\bar{3}m$) at intermediary temperatures have been found. Our results suggest that there might be a second, not yet observed, intermediary phase where the NO_2 units rotate along a twofold axis of the structure.

5.3. Structure determination

While the true structure prediction described above is a fascinating area of fundamental research in crystallography, chemistry and materials science, a more restricted somewhat modified application of the same methodology looks ready to become an invaluable tool in applied crystallography and solid-state chemistry: structure determination from limited experimental information using energy landscapes, *i.e.* the generation of structure candidates that can be refined using standard methods.

A number of methods have been developed, such as the reverse Monte Carlo method (Kaplow *et al.*, 1968; McGreevy, 1997; LeBail, 2000; Mellergård & McGreevy, 1999) or structure determination using experimental cell information (Freeman *et al.*, 1993; Catlow *et al.*, 1994; Pannetier *et al.*, 1990; Belashchenko, 1994). We have introduced a Pareto-optimization approach, where the cost function modifies the

energy landscape through an explicit incorporation of experimental data by adding the difference R_B (sometimes denoted the R value) between the measured Bragg intensities and those calculated for the current atomic configuration to the potential energy (Putz *et al.*, 1999*b*). Here, the structure is optimized both with respect to the energy and the diffractogram,

$$E = \lambda E_{\rm pot} + (1-\lambda)R_B \qquad (0 \le \lambda \le 1). \tag{10}$$

Besides prescribing the cell parameters, one can include various additional constraints, *e.g.* keep the positions of some of the atoms fixed. In analogy to the restricted structure prediction, it is also possible to employ rigid or flexible building units, in particular for dealing with complex ions and molecules.

This approach has been tested successfully for a large number of ionic, quasi-ionic and metallic systems (Putz *et al.*, 1999*b*; Coelho, 2000; Lanning *et al.*, 2000), where simulated annealing (Putz *et al.*, 1999*b*; Coelho, 2000) and genetic algorithms (Lanning *et al.*, 2000) were used as the global optimization tool. Typically, simple two-body potentials with Coulomb and Lennard-Jones terms served as energy functions; such simple potentials were sufficient because the combination of experimental input and theoretical energy function delivered a high synergy by eliminating many unrealistic local minima on the energy landscape. One up-to-date implementation, the program *ENDEAVOUR* (Putz, 2000), has already been very successful in 'real life' applications, generating convincing structure candidates for such different systems as K_2CN_2 (Becker & Jansen, 2000), sulfur (Crichton *et al.*, 2001), Na_3PSO_3 (Pompetzki & Jansen, 2002), Ag_2NiO_2 (Schreyer & Jansen, 2002), Ag_2PdO_2 (Schreyer & Jansen, 2001), $GaAsO_4$ (Santamaría-Pérez *et al.*, 2006), ammonium metatungstate (Christian & Whittingham, 2008), the zeolite-like structure $Na_{1-x}Ge_{3+z}$ (Beekman *et al.*, 2007), Tl_2CS_3 (Beck & Benz, 2009) and BiB_3O_6 (Dinnebier *et al.*, 2009).

5.4. Optimal control of phase transitions

In practical studies of phase transitions, one needs to be aware that there is only a finite time available to perform the actual transition from phase A to phase B. As a consequence, one has to perform excess work or generate excess entropy compared with the infinite-time case where the system can proceed along a quasi-equilibrium curve in state space. Thus, one has to employ methods of finite-time thermodynamics and optimal control to optimize the transition process (Schön, 2009).

This applies not only to experiment but also to the issue of simulating a phase transition and measuring, for example, the difference in free energies between the two phases with as high an accuracy as possible for a finite amount of computer time (function evaluations). In this case, too, the problem can be cast as a finite-time thermodynamics problem, and an optimality criterion be derived that allows the minimization of the difference between the computed free-energy difference and the true value.

5.4.1. Optimal control of the gas-to-liquid phase transition. As the most simple example of an analysis of a phase transition that is to take place in finite time, we considered the liquification of one mole of the gaseous phase of a molecular compound (N_2) at constant temperature and standard pressure (Santoro *et al.*, 2007). As temperature, we chose the boiling temperature at standard pressure, $T = T^{\rm boil}$ at the pressure $p_g = p_g^{\infty}(T^{\rm boil}) = 1$ atm. Thus, the chemical potentials of the liquid and the gaseous phase are equal, and, if the process could proceed infinitely slowly, no excess work would be required. The goal is to perform this liquification within a finite time τ and to adjust the pressure of the system in such a way that the total amount of external work needed is minimized. The excess work is indicated by the shaded area in Fig. 5, and its rate of change is given by

$$\frac{{\rm d}W_{\rm exc}}{{\rm d}t} = \left(p_g - p_g^{\infty}\right)\left[(n_{\rm total} - n_l)\frac{k_B T}{p_g^2}\frac{{\rm d}p_g}{{\rm d}t} + \left(\frac{k_B T}{p_g} - v_l\right)\frac{{\rm d}n_l}{{\rm d}t}\right], \tag{11}$$

where v_l is the amount of volume per particle in the liquid state, and $n_{\rm total}$ is the total number of particles in the system.

We describe the system by its internal pressure $p_g(t)$ and the amount of liquid phase $n_l(t)$, which have admissible ranges of values $[p_g^{\infty}, \infty]$ and $[0, n_{\rm total}]$, respectively. The control variable is the externally applied pressure $p_a(t)$, with admissible range $[p_g^{\infty}, \infty]$, and we assume that the internal pressure reacts instantaneously to variations of the applied pressure, ${\rm d}p_a/{\rm d}t = {\rm d}p_g/{\rm d}t$. Regarding the cluster size distribution, we focus for simplicity only on the clusters of critical size by making the assumption that, once this size has been reached,

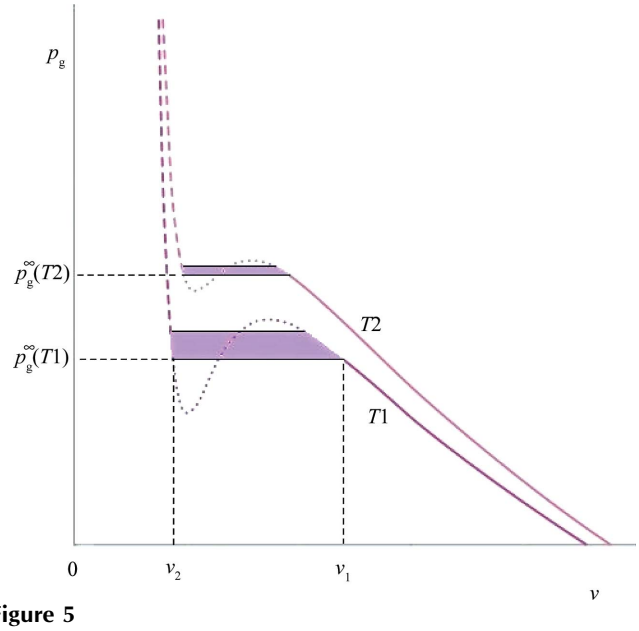

Figure 5
Sketch of the excess work (shaded region) needed to perform a first-order phase transition within a finite time by applying a pressure in excess of the equilibrium pressure $p_g^{\infty}(T)$ at a fixed temperature $T = T^{\rm boil}$.

Table 1
Initial pressure $p_a^*(t_0)$ and total excess work W_{exc} as a function of total available time τ.

τ (s)	$p_a^*(t_0)$ (Pa)	W_{exc} (J)
1	259730	5.430
10	251695	5.407
10^2	244640	5.380
10^3	238365	5.362
10^4	232750	5.346
10^5	227685	5.333

on average the clusters will continue to grow. Furthermore, we employ classical nucleation theory (Zettlemoyer, 1969) to describe the steady-state rate at which clusters are generated, J_{ss}^c. This leads to the following evolution equation for the amount of liquid phase,

$$\frac{dn_l}{dt} = n_l^c J_{ss}^c + \frac{(4\pi)^{1/3}(3v_l)^{2/3}}{(2\pi m k_B T)^{1/2}} n_l^{2/3}(p_g - p_g^\infty)\left(\int_0^t J_{ss}^c \, dt'\right)^{1/3}, \tag{12}$$

where the first term is the cluster generation rate (n_l^c is the amount of molecules inside a critical cluster), and the second one is the average cluster growth rate.

Applying standard calculus of variation methods to the optimization problem, we derive a set of coupled differential equations, which need to be solved numerically. As a result, we find for the optimal trajectory $p_a(t) = p_g(t)$ that one should rapidly increase the pressure from the standard equilibrium pressure $p_g^\infty(T^{boil})$ to an optimal value $p_a^*(t_0)$ ($t_0 = 0$ in the limit of infinitely fast pressure increases) in order to start a burst of nucleation, followed by a fast decrease of $p_g(t)$ to slightly above p_g^∞ for the remainder of the allowed time τ. Fig. 6 shows these optimal curves for six different values of τ (1 s, ..., 10^5 s) plotted *versus* the scaled time $t' = t/\tau$. Table 1 gives the values of $p_a^*(t_0)$ and the excess work as a function of τ. As expected, the amount of excess work required decreases with increasing available time τ; this decrease can be approximated by a power law.

5.4.2. Optimality criterion for free-energy difference calculation. One of the most popular ways to compute free-energy differences between two systems A and B is the adiabatic procedure proposed by Watanabe & Reinhardt (1990). Here one proceeds by changing system A into system B along some route in parameter space. After each step along the trajectory, one computes the work performed, and then lets the system relax to equilibrium. After having reached system B, one follows the trajectory in parameter space in the reverse direction. Summing up these pieces of work along each of the two paths yields an upper and a lower bound on the free-energy difference, respectively. Ideally, one would proceed in a quasi-equilibrium fashion by taking very small steps along the trajectory and fully relaxing the system after each step. However, the computational effort to do so is very large, and one needs to efficiently allocate the limited computer time available.

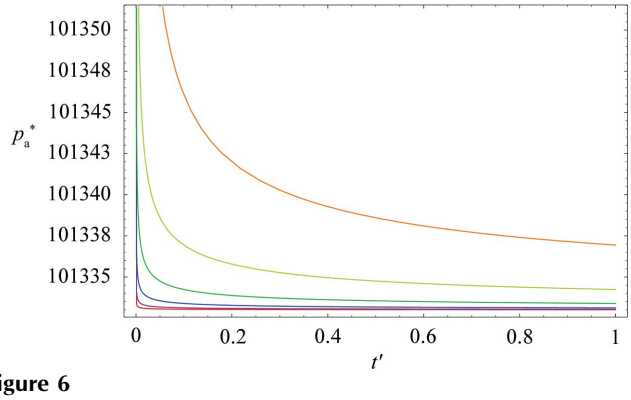

Figure 6
Optimal applied pressure $p_a^*(t)$ *versus* $t' = t/\tau$ in units of Pa for nitrogen N_2. $\tau = 1$, orange; $\tau = 10$, light green; $\tau = 10^2$, green; $\tau = 10^3$, blue; $\tau = 10^4$, magenta; $\tau = 10^5$, red.

Thus, it is clearly of interest to determine an optimal path in the sense that one wants to (*a*) find the optimal trajectory in parameter space between system A and B, and (*b*) determine the optimal step sizes along this trajectory and the length of time spent on relaxing the system at each parameter value, such that the difference between the free-energy change ΔF and the work W along the path, $I_{path} = W - \Delta F$, is minimal. Trying to solve this optimal control problem turns out to be equivalent to a finite-time thermodynamics problem in probability space, and one can derive a lower bound on the loss of availability or the generation of excess entropy in the phase transition by minimizing the thermodynamic path length along the trajectory in parameter space (Schön, 1996). Taking only the lowest terms into account, we find that along the optimal path the quantity I_{path} is bounded from below by the thermodynamic path length,

$$\begin{aligned} I_{path} &\simeq k_B T \sum_{i=1}^{N} (\boldsymbol{\pi}_i - \boldsymbol{\pi}_{i-1})(1/\boldsymbol{\pi}_i)(\boldsymbol{\pi}_i - \boldsymbol{\pi}_{i-1}) \\ &= \sum_{i=1}^{N} (\Delta L_i)^2 \;\geq\; L^2/N. \end{aligned} \tag{13}$$

Here, $\boldsymbol{\pi}_i$ is the equilibrium probability distribution over all states of the system at step i along the path between A and B, N is the number of steps along the path, and ΔL_i is the length of the piece $(\boldsymbol{\pi}_i - \boldsymbol{\pi}_{i-1})$ measured in the metric on probability space $k_B T(1/\boldsymbol{\pi}_i)$ at step i. Note that the equality holds only if all the individual lengths ΔL_i are equal. Thus, the optimization problem is equivalent to determining the path between the two phases A and B that is minimal measured in the metric in probability space, and then to allocate the available N steps in such a fashion that each individual piece has the same length along the path. Applying this prescription to the switching of a classical paramagnet with the help of a magnetic field shows that the thermodynamically optimal route is clearly preferable to typical heuristic ways of resource allocation.

6. Experimental verification

The second step of planning chemical syntheses consists of rationally developing a viable path to the desired configura-

tion, predicted to be either kinetically or thermodynamically stable. This is a task of intriguing complexity, which includes monitoring the structural and compositional evolution of the system under consideration as a function of time. The reactions involved need to proceed spontaneously, and the system thus follows a descending trajectory on the hyperspace of free enthalpy. In many instances, such pathways would be a spin-off of the determination of the free enthalpy landscapes, addressed above. However, upon approaching the synthesis target many pathways leading to different modifications compete, and the final outcome is the result of a bifurcation in the cluster population in sub- and super-critical nuclei. This final step in the synthesis of a specific solid is determined by the kind of nuclei that first reach critical size and start growing. Therefore, special measures need to be taken to direct the system into the minimum region corresponding to the desired configuration. To exert an influence on this decisive final step would require at least some control of the population dynamics of transient states occurring in the pre-organization stage during which the stable (supercritical) nuclei develop. Regrettably, neither the theoretical treatment nor the experimental control of such a process has yet reached a satisfactory level. In our attempts to experimentally realize metastable predicted compounds we have therefore restricted ourselves to creating synthesis conditions that would be particularly suited to preserve metastable configurations, *i.e.* employing only low thermal activation.

For this purpose, we have constructed a dedicated experimental setup (see Fig. 7) based on a high-vacuum chamber. This allows us to evaporate the elements constituting the desired compound as free atoms and to deposit them on a cooled substrate (at liquid-nitrogen or at liquid-helium temperature) in a random spatial distribution. The structural evolution of this solid reaction mixture, which is very much reminiscent of the starting configurations for the global computational exploration of the respective energy landscape, is monitored as a function of time and temperature. Amazingly, such mixtures undergo all solid-state reactions at temperatures far below room temperature, yielding well crystallized products (Fischer & Jansen, 2002*a*).

Synthesizing crystalline solids at these low temperatures, *i.e.* at such extremely mild thermal conditions, is unprecedented, and we have, indeed, been able to prepare several metastable compounds and metastable modifications along this novel approach. Some characteristics of our low-temperature atomic beam deposition (LT-ABD) have already become apparent over the past decade. Obviously, when crystalline nuclei form inside the amorphous deposit, a shrinkage of volume occurs, thus generating effective negative pressures on the surfaces of these nuclei (Fischer, Čančarević *et al.*, 2004). Consequently, the first structures evolving are the metastable low-density ones, quite in line with our intention.

Among the most impressive results is the synthesis of the elusive sodium nitride Na_3N in an energetically high-lying structure, the ReO_3 type (Fischer & Jansen, 2002*b*). Even more importantly, almost the full set of the most stable predicted, and published five years in advance, polymorphs of Na_3N (see Fig. 1 and §5.1.2) have recently been realized (Vajenine *et al.*, 2008, 2009). This impressively documents the strength of our new rational way for solid-state synthesis, and this result has attracted a lot of interest, which in part arose because of generations of solid-state chemists having tried in vain to synthesize Na_3N in the past.

As we have already discussed in §5.1.1, for each of the 20 alkali metal halides numerous polymorphs have been predicted (Schön & Jansen, 1995; Čančarević *et al.*, 2008). Out of these, the energetically still low-lying low-density wurtzite-type configurations have been realized for LiI (Fischer, Müller & Jansen, 2004; Čančarević *et al.*, 2005), LiBr (Liebold-Ribeiro *et al.*, 2008) and LiCl (Bach *et al.*, 2009), using our LT-ABD technique. Fig. 8 shows the typical structural evolution from

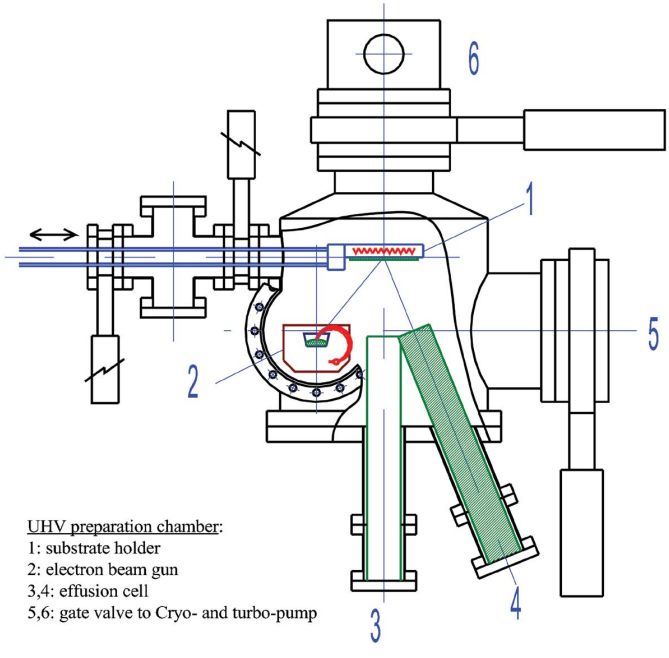

Figure 7
Experimental setup of the low-temperature atom-beam deposition method (Fischer & Jansen, 2002*a*).

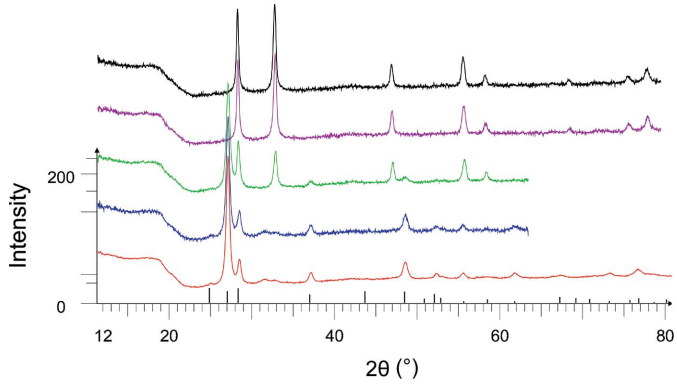

Figure 8
X-ray powder diffractograms of LiBr deposited *via* the LT-ABD method, as a function of temperature (bottom to top: 223, 243, 263, 283, 298 K) (Liebold-Ribeiro *et al.*, 2008). The line diagram indicates the peaks corresponding to the wurtzite modification of LiBr.

the amorphous phase (not depicted here) to the metastable wurtzite and finally to the thermodynamically stable rocksalt structure for LiBr. It is noteworthy how well developed the powder diffractograms are in spite of the extremely low crystallization temperatures.

7. Conclusion

In the preceding sections, we have reviewed the energy landscape approach to the rational planning of solid-state syntheses. One pillar is the realization that all feasible (meta)stable modifications of a chemical system are represented by the locally ergodic regions of the energy landscape of the system. We have presented our modular approach to the identification of such regions, together with the algorithms and analysis tools we have employed for this purpose for a variety of systems ranging from ionic compounds over systems where covalent or mixed covalent–ionic bonding is expected, to materials that exhibit solid-solution phases. While the first studies focused on standard pressure and zero temperature modifications, the thermodynamic space that can now be explored includes high pressures, elevated temperatures and variations of composition.

Parallel to this theoretical work, explorations of the landscape *via* the development of new synthesis methods have taken place that aim at the experimental verification of the predictions. One class comprises high-pressure experiments, where, for example, in the case of the alkali metal sulfides Li_2S (Grzechnik *et al.*, 2000), Na_2S (Vegas *et al.*, 2001) and K_2S (Vegas *et al.*, 2002) the independent predictions of several high-pressure phases (Schön *et al.*, 2001*b*; Schön, Čančarević & Jansen, 2004) could be confirmed. Similarly, the X-ray powder diffractogram for the high-pressure phase of Li_2O (Kunc *et al.*, 2005) is in good agreement with the predicted high-pressure phases (Čančarević *et al.*, 2006*b*); note that in both the alkali sulfides and oxides the predicted high-pressure phases of the remaining compounds Rb_2S, Cs_2S, Na_2O, K_2O, Rb_2O and Cs_2O still await their experimental verification.

A second class of experiments using the LT-ABD method described in the preceding section has led to the confirmation of the existence of competing metastable modifications of the alkali halides in the wurtzite type for LiBr and LiCl. It appears to be only a matter of time until, for example, the predicted 5–5 structure type (Schön & Jansen, 1995; Schön, 2004; Čančarević *et al.*, 2008) that has already been realized as the aristotype of Li_4SeO_5 (Haas & Jansen, 1999), and also been seen in thin films of ZnO (Claeyssens *et al.*, 2005), will also have become accessible in the alkali halides or perhaps the earth alkali metal oxides. And, as mentioned above, a combination of the LT-ABD method and high-pressure syntheses has resulted in the successful synthesis of four previously predicted modifications of the elusive sodium nitride (Fischer & Jansen, 2002*b*; Vajenine *et al.*, 2008, 2009).

Standing firmly on this solid theoretical and experimental basis, the time appears to have come to address the third pillar of the rational planning of solid-state synthesis: the modeling and optimization of chemical syntheses. For certain types of syntheses, such as growth of crystals from a melt, or the generation of new phases *via* phase transitions upon changes in temperature and/or pressure, these tasks can be achieved by analyzing pathways on the energy landscape of the chemical system alone. In contrast, many typical syntheses involve additional chemical species, solvents and/or catalysts whose influence must be taken into account during the modeling process. But even in the case of pure phase transitions, the fact that many of these transformations are of first order leads to technical problems in atomistic modeling, owing to the large size of the simulations that have to be able to describe nuclei of critical sizes containing hundreds or even thousands of atoms.

Thus, it will be necessary to combine models on many time- and length scales to reach an approximately analytical description, which then can be analyzed and employed as input to an optimal control approach aimed at achieving a specific synthesis outcome.[9]

The widespread availability of fast computers and clusters thereof has led in recent years to a rapid increase in the number of research groups involved both in the prediction of new compounds and in the atomistic modeling of phase transitions and other synthesis routes, in particular the nucleation and growth of crystals from solution (Catlow *et al.*, 2007). While the success rate of the predictions has steadily increased with time (and available computer time), not everyone seems to be conscious of the fact that at each given temperature and pressure there are many possible metastable modifications capable of existence. Thus it is crucial not only to search for the thermodynamic minimum configuration but also to identify the competing metastable ones, and to estimate their kinetic stability. Once this has been achieved, we look again towards the second pillar of our approach, the experimental verification, fully realizing that the challenge to the experimentalist to develop new low-activation-energy, and thus kinetically controllable, synthesis routes capable of producing metastable compounds, is at least as great as the task for the theorist to predict new compounds and suggest routes to their synthesis.

The work was funded by the BMBF project 03C0352 and the Multiscale-Materials-Modeling initiative of the Max Planck Society.

References

Abagyan, R. & Argos, P. (1992). *J. Mol. Biol.* **225**, 519–532.
Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. (1993). *Protein Struct. Funct. Genet.* **17**, 412–425.
Bach, A., Fischer, D. & Jansen, M. (2009). *Z. Anorg. Allg. Chem.*, doi: 10.1002/zaac.200900357.
Beck, J. & Benz, S. (2009). *Z. Anorg. Allg. Chem.* **635**, doi:10.1002/zaac.200801408.
Becker, M. & Jansen, M. (2000). *Solid State Sci.* **2**, 711–715.
Becker, O. M. (1997). *Protein Struct. Funct. Genet.* **27**, 213–226.
Becker, O. M. & Karplus, M. (1997). *J. Chem. Phys.* **106**, 1495–1517.

[9] An example of such a stepping-stone description is the modeling of the sol–gel synthesis of the amorphous ceramic a-$Si_3B_3N_7$ (Schön, Hannemann & Jansen, 2004; Hannemann *et al.*, 2005).

Beekman, M., Kaduk, J. A., Huang, Q., Wong-Ng, W., Yang, Z., Wang, D. & Nolas, G. S. (2007). *Chem. Commun.* pp. 837–839.
Belashchenko, D. K. (1994). *Inorg. Mater.* **30**, 966–976.
Brese, N. E. & O'Keeffe, M. (1992). *Structure and Bonding*, p. 307. Heidelberg: Springer.
Brown, I. D. (1992). *Acta Cryst.* B**48**, 553–572.
Brown, I. D. & Shannon, R. D. (1973). *Acta Cryst.* A**29**, 266–282.
Čančarević, Ž., Schön, J. C., Fischer, D. & Jansen, M. (2005). *Mater. Sci. Forum*, **494**, 61–66.
Čančarević, Ž., Schön, J. C. & Jansen, M. (2004). *Mater. Sci. Forum*, **453**, 71–76.
Čančarević, Ž., Schön, J. C. & Jansen, M. (2006*a*). *Z. Anorg. Allg. Chem.* **632**, 1437–1448.
Čančarević, Ž., Schön, J. C. & Jansen, M. (2006*b*). *Phys. Rev. B*, **73**, 224114.
Čančarević, Ž., Schön, J. C. & Jansen, M. (2007). *Chem. Eur. J.* **13**, 7330–7348.
Čančarević, Ž., Schön, J. C. & Jansen, M. (2008). *Chem. Asian J.* **3**, 561–572.
Catlow, C. R. A., Bell, R. G. & Gale, J. D. (1994). *J. Mater. Chem.* **4**, 781–792.
Catlow, C. R. A., DeLeeuw, N. H., Anwar, J., Davey, R. J., Roberts, K. J. & Unwin, P. R. (2007). Editors. *Faraday Discussions 136: Crystal Growth and Nucleation.* London: Royal Society of Chemistry.
Christian, J. B. & Whittingham, M. S. (2008). *J. Solid State Chem.* **181**, 1782–1791.
Chung, S. J., Hahn, T. & Klee, W. E. (1984). *Acta Cryst.* A**40**, C212.
Civalleri, B., D'Arco, P., Orlando, R., Saunders, V. R. & Dovesi, R. (2001). *Chem. Phys. Lett.* **348**, 131–138.
Claeyssens, F., Freeman, C. L., Allan, N. L., Sun, Y., Ashfold, M. N. R. & Harding, J. H. (2005). *J. Mater. Chem.* **15**, 139–148.
Coelho, A. A. (2000). *J. Appl. Cryst.* **33**, 899–908.
Crichton, W. A., Vaughan, G. B. M. & Mezouar, M. (2001). *Z. Kristallogr.* **216**, 417–419.
Czerny, V. (1985). *J. Optim. Theor. Appl.* **45**, 41–51.
Das, P., Moll, M., Stamati, H., Kavraki, L. E. & Clementi, C. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 9885–9890.
Delamarre, D. & Virot, B. (1998). *RAIRO Rech. Oper.* **32**, 43–73.
Delgado-Friedrichs, O., Dress, A. W. M., Huson, D. H., Klinowski, J. & Mackay, A. L. (1999). *Nature (London)*, **400**, 644–647.
Delgado-Friedrichs, O., O'Keefe, M. & Yaghi, O. M. (2007). *Phys. Chem. Chem. Phys.* **9**, 1035–1043.
Dinnebier, R. E., Hinrichsen, B., Lennie, A. & Jansen, M. (2009). *Acta Cryst.* B**65**, 1–10.
Doll, K. (2001). *Comput. Phys. Commun.* **137**, 74–88.
Doll, K., Dovesi, R. & Orlando, R. (2004). *Theor. Chem. Acc.* **112**, 394–402.
Doll, K., Dovesi, R. & Orlando, R. (2006). *Theor. Chem. Acc.* **115**, 354–360.
Doll, K., Saunders, V. R. & Harrison, N. M. (2001). *Int. J. Quant. Chem.* **82**, 1–13.
Doll, K., Schön, J. C. & Jansen, M. (2007). *Phys. Chem. Chem. Phys.* **9**, 6128–6133.
Doll, K., Schön, J. C. & Jansen, M. (2008). *Phys. Rev. B*, **78**, 144110.
Doll, K., Schön, J. C. & Jansen, M. (2010). *J. Chem. Phys.* **133**, 024107.
Doornhof, D., Wijk, H. V. & Hoonk, H. (1984). *Thermochim. Acta*, **76**, 171–178.
Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, P. & Llunell, M. (2006). *CRYSTAL2006.* University of Torino, Italy.
Duan, C. G., Mei, W. N., Smith, R. W., Liu, J., Ossowski, M. M. & Hardy, J. R. (2001). *Phys. Rev. B*, **63**, 144105.
Dueck, G. (1993). *J. Comput. Phys.* **104**, 86–92.
Dueck, G. & Scheuer, T. (1990). *J. Comput. Phys.* **90**, 161–175.
Fischer, D., Čančarević, Ž., Schön, J. C. & Jansen, M. (2004). *Z. Anorg. Allg. Chem.* **630**, 156–160.
Fischer, D. & Jansen, M. (2002*a*). *J. Am. Chem. Soc.* **124**, 3488–3489.
Fischer, D. & Jansen, M. (2002*b*). *Angew. Chem. Int. Ed.* **41**, 1755–1756.
Fischer, D., Müller, A. & Jansen, M. (2004). *Z. Anorg. Allg. Chem.* **630**, 2697–2700.
Foster, M. D., Simperler, A., Bell, R. G., Delgado-Friedrichs, O., Paz, F. A. & Klinowski, J. (2004). *Nat. Mater.* **3**, 234–238.
Freeman, C. M., Newsam, J. M., Levine, S. M. & Catlow, C. R. A. (1993). *J. Mater. Chem.* **3**, 531–535.
Geman, S. & Geman, D. (1984). *IEEE Trans. Pattern Anal.* **6**, 721–741.
Goldstein, M. (1969). *J. Chem. Phys.* **51**, 3728–3739.
Gower, J. C. (1966). *Biometrika*, **53**, 325–338.
Grzechnik, A., Vegas, A., Syassen, K., Loa, I., Hanfland, M. & Jansen, M. (2000). *J. Solid State Chem.* **154**, 603–611.
Haas, H. & Jansen, M. (1999). *Angew. Chem. Int. Ed. Engl.* **38**, 1910–1911.
Hannemann, A., Hundt, R., Schön, J. C. & Jansen, M. (1998). *J. Appl. Cryst.* **31**, 922–928.
Hannemann, A., Schön, J. C. & Jansen, M. (2005). *J. Mater. Chem.* **15**, 1167–1178.
Hansen, S. (1993). *J. Solid State Chem.* **105**, 247–254.
Heesch, H. (1934). *Nachr. Ges. Wiss. Göttingen.* pp. 35–42.
Heuer, A. (1997). *Phys. Rev. Lett.* **78**, 4051–4054.
Hoffmann, K. & Schön, J. C. (2005). *Found. Phys. Lett.* **18**, 171–182.
Hoffmann, K. H. & Sibani, P. (1988). *Phys. Rev. A*, **38**, 4261–4270.
Hundt, R. (1979). *KPLOT: A Program for Plotting and Investigation of Crystal Structures* Version 9. University of Bonn, Germany.
Hundt, R., Schön, J. C., Hannemann, A. & Jansen, M. (1999). *J. Appl. Cryst.* **32**, 413–416.
Hundt, R., Schön, J. C. & Jansen, M. (2006). *J. Appl. Cryst.* **39**, 6–16.
Iwamatsu, M. & Okabe, Y. (2004). *Chem. Phys. Lett.* **399**, 396–400.
Jansen, M. (2002). *Angew. Chem. Int. Ed.* **41**, 3747–3766.
Jansen, M. (2008). *Turning Points in Solid-State, Materials and Surface Science*, edited by K. M. Harris & P. Edwards, p. 22. Cambridge: RSC Publishing.
Jansen, M. & Schön, J. C. (1998). *Z. Anorg. Allg. Chem.* **624**, 533–540.
Jansen, M. & Schön, J. C. (2006). *Angew. Chem. Int. Ed.* **45**, 3406–3412.
Kaplow, R., Rowe, T. A. & Averbach, B. L. (1968). *Phys. Rev.* **168**, 1068–1079.
Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
Klee, W. E., Bader, M. & Thimm, G. (1997). *Z. Kristallogr.* **212**, 553–558.
Komatsuzaki, T., Hoshino, K., Matsunaga, Y., Rylance, G. J., Johnston, R. L. & Wales, D. J. (2005). *J. Chem. Phys.* **122**, 084714.
Krivov, S. V. & Karplus, M. (2002). *J. Chem. Phys.* **117**, 10894–10903.
Kulkarni, A., Doll, K., Schön, J. C. & Jansen, M. (2010). *J. Phys. Chem. B*. Submitted.
Kunc, K., Loa, I., Grzechnik, A. & Syassen, K. (2005). *Phys. Status Solidi. B*, **242**, 1857–1863.
Lanning, O. J., Habershon, S., Harris, K. D. M., Johnston, R. L., Kariuki, B. M., Tedesco, E. & Turner, G. W. (2000). *Chem. Phys. Lett.* **317**, 296–303.
LeBail, A. (2000). *Proc. EPDIC-7.* Preprint.
Liebold-Ribeiro, Y., Fischer, D. & Jansen, M. (2008). *Angew. Chem. Int. Ed.* **47**, 4428–4431.
McGreevy, R. L. (1997). *Computer Modelling in Inorganic Crystallography*, edited by C. R. A. Catlow. pp. 151–184. San Diego: Academic Press.
Mellergård, A. & McGreevy, R. L. (1999). *Acta Cryst.* A**55**, 783–789.
Mellot-Draznieks, C., Girard, S., Ferey, G., Schön, J. C., Čančarević, Ž. & Jansen, M. (2002). *Chem. Eur. J.* **8**, 4102–4113.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.

Möbius, A., Hoffmann, K. & Schön, J. C. (2004). *Complexity, Metastability and Nonextensivity*, edited by C. Beck, G. Benedek, A. Rapisarda & C. Tsallis. pp. 215–219. Singapore: World Scientific.
Möbius, A., Neklioudov, A., Diaz-Sanchez, A., Hoffmann, K. H., Fachat, A. & Schreiber, M. (1997). *Phys. Rev. Lett.* **79**, 4297–4301.
Müller, U. (1992). *Acta Cryst.* B**48**, 172–178.
Müller, U. (1998). *Z. Anorg. Allg. Chem.* **624**, 529–532.
O'Keeffe, M. & Hyde, B. G. (1984). *Nature (London)*, **309**, 411–414.
O'Keeffe, M. & Hyde, B. G. (1985). *Struct. Bond.* **61**, 77.
Onoda-Yamamuro, N., Honda, H., Ikeda, R., Yamamuro, O., Matsuo, T., Oikawa, K., Kamiyama, T. & Izumi, F. (1998). *J. Phys. Condens. Matter*, **10**, 3341–3351.
Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J. & Caignaert, V. (1990). *Nature (London)*, **346**, 343–345.
Pentin, I. V., Schön, J. C. & Jansen, M. (2007). *J. Chem. Phys.* **126**, 124508.
Pentin, I. V., Schön, J. C. & Jansen, M. (2008). *Solid State Sci.* **10**, 804–813.
Polya, G. (1936). *Z. Kristallogr.* **93**, 415.
Pompetzki, M. & Jansen, M. (2002). *Z. Anorg. Allg. Chem.* **628**, 641–646.
Putz, H. (2000). *Endeavour 1.0.* Crystal Impact GbR, Bonn, Germany.
Putz, H., Schön, J. C. & Jansen, M. (1998). *Comput. Mater. Sci.* **11**, 309–322.
Putz, H., Schön, J. C. & Jansen, M. (1999*a*). *Z. Anorg. Allg. Chem.* **625**, 1624–1630.
Putz, H., Schön, J. C. & Jansen, M. (1999*b*). *J. Appl. Cryst.* **32**, 864–870.
Quandt, A. (2008). *Lect. Notes Phys.* **739**, 437–469.
Redlich, O. & Kister, A. T. (1948). *Ind. Eng. Chem.* **40**, 345.
Ruppeiner, G., Pedersen, J. M. & Salamon, P. (1991). *J. Phys. I*, **1**, 455–470.
Salamon, P., Nulton, J., Robinson, J., Pedersen, J. M., Ruppeiner, G. & Liao, L. (1988). *Comput. Phys. Commun.* **49**, 423–428.
Salamon, P., Sibani, P. & Frost, R. (2002). *Facts, Conjectures, and Improvements for Simulated Annealing.* Philadelphia: SIAM Monographs.
Sangster, J. & Pelton, A. (1987). *J. Phys. Chem. Ref. Data*, **16**, 509.
Santamaría-Pérez, D., Haines, J., Amador, U., Morán, E. & Vegas, A. (2006). *Acta Cryst.* B**62**, 1019–1024.
Santoro, M., Schön, J. C. & Jansen, M. (2007). *Phys. Rev. E*, **76**, 061120.
Schön, J. C. (1996). *J. Chem. Phys.* **105**, 10072–10083.
Schön, J. C. (1997). *J. Phys. A*, **30**, 2367–2389.
Schön, J. C. (2004). *Z. Anorg. Allg. Chem.* **630**, 2354–2366.
Schön, J. C. (2009). *Z. Anorg. Allg. Chem.* **635**, 1794–1806.
Schön, J. C., Čančarevič, Ž. & Jansen, M. (2004). *J. Chem. Phys.* **121**, 2289–2304.
Schön, J. C., Čančarevič, Ž. P., Hannemann, A. & Jansen, M. (2008). *J. Chem. Phys.* **128**, 194712.
Schön, J. C., Doll, K. & Jansen, M. (2010). *Phys Status Solidi B*, **247**, 23–39.
Schön, J. C., Hannemann, A. & Jansen, M. (2004). *J. Phys. Chem. B*, **108**, 2210–2217.
Schön, J. C. & Jansen, M. (1995). *Comput. Mater. Sci.* **4**, 43–58.
Schön, J. C. & Jansen, M. (1996). *Angew. Chem. Int. Ed. Engl.* **35**, 1286–1304.
Schön, J. C. & Jansen, M. (2001). *Z. Kristallogr.* **216**, 307–325.
Schön, J. C. & Jansen, M. (2005). *Proceedings of the Materials Research Society Symposium*, Vol. 848, *Solid State Chemistry of Inorganic Materials V*, edited by J. Li, N. E. Brese, M. G. Kanatzidis & M. Jansen, pp. 333–344. Warrendale: MRS.
Schön, J. C. & Jansen, M. (2009). *Int. J. Mater. Res.* **100**, 135–152.
Schön, J. C., Pentin, I. V. & Jansen, M. (2006). *Phys. Chem. Chem. Phys.* **8**, 1778–1784.
Schön, J. C., Pentin, I. V. & Jansen, M. (2007). *J. Phys. Chem. B*, **111**, 3943.
Schön, J. C., Pentin, I. V. & Jansen, M. (2008). *Solid State Sci.* **10**, 455–460.
Schön, J. C., Putz, H. & Jansen, M. (1996). *J. Phys. Condens. Matter*, **8**, 143–156.
Schön, J. C., Salamon, P. & Jansen, M. (2010). In preparation.
Schön, J. C., Wevers, M. A. C. & Jansen, M. (2000). *Solid State Sci.* **2**, 449–456.
Schön, J. C., Wevers, M. A. C. & Jansen, M. (2001*a*). *J. Phys. A*, **34**, 4041–4052.
Schön, J. C., Wevers, M. A. C. & Jansen, M. (2001*b*). *J. Mater. Chem.* **11**, 69–77.
Schön, J. C., Wevers, M. A. C. & Jansen, M. (2003). *J. Phys. Condens. Matter*, **15**, 5479–5486.
Schreyer, M. & Jansen, M. (2001). *Solid State Sci.* **3**, 25–30.
Schreyer, M. & Jansen, M. (2002). *Angew. Chem. Int. Ed.* **41**, 643.
Sibani, P., van der Pas, R. & Schön, J. C. (1999). *Comput. Phys. Commun.* **116**, 17–27.
Sibani, P., Schön, J. C., Salamon, P. & Andersson, J. O. (1993). *Europhys. Lett.* **22**, 479–485.
Sibani, P. & Schriver, P. (1994). *Phys. Rev. B*, **49**, 6667–6671.
Solbakk, J. K. & Stromme, K. O. (1969). *Acta Chem. Scand.* **23**, 300.
Stillinger, F. & Weber, T. A. (1982). *Phys. Rev. A*, **25**, 978–989.
Strong, R. T., Pickard, C. J., Milman, V., Thimm, G. & Winkler, B. (2004). *Phys. Rev. B*, **70**, 045101.
Szu, H. & Hartley, R. (1987). *Phys. Lett. A*, **122**, 157.
Thomas, J. M. & Klinowski, J. (2007). *Angew. Chem. Int. Ed.* **46**, 7160.
Treacy, M. M. J., Randall, K. H., Rao, S., Perry, J. A. & Chadi, D. J. (1997). *Z. Kristallogr.* **212**, 768–791.
Troyer, J. M. & Cohen, F. E. (1995). *Protein Struct. Funct. Genet.* **23**, 97–110.
Tsallis, C. (1988). *J. Stat. Phys.* **52**, 479–487.
Tsallis, C. & Stariolo, S. A. (1996). *Physica A*, **233**, 395–406.
Vajenine, G. V., Wang, X., Efthimiopoulus, I., Karmakar, S., Syassen, K. & Hanfland, M. (2008). *Z. Anorg. Allg. Chem.* **634**, 2015.
Vajenine, G. V., Wang, X., Efthimiopoulus, I., Karmakar, S., Syassen, K. & Hanfland, M. (2009). *Phys. Rev. B*, **79**, 224107.
Vegas, A., Grzechnik, A., Hanfland, M., Mühle, C. & Jansen, M. (2002). *Solid State Sci.* **4**, 1077–1081.
Vegas, A., Grzechnik, A., Syassen, K., Loa, I., Hanfland, M. & Jansen, M. (2001). *Acta Cryst.* B**57**, 151–156.
Voronin, G. (2003). *Russ. J. Phys. Chem.* **77**, 1874.
Wales, D. J. (2003). *Energy Landscapes with Applications to Clusters, Biomolecules and Glasses.* Cambridge University Press.
Wales, D. J. & Doye, J. P. K. (1997). *J. Phys. Chem.* **101**, 5111–5116.
Wales, D. J., Miller, M. A. & Walsh, T. R. (1998). *Nature (London)*, **394**, 758–760.
Watanabe, M. & Reinhardt, W. P. (1990). *Phys. Rev. Lett.* **65**, 3301–3304.
Wevers, M. A. C., Schön, J. C. & Jansen, M. (1999). *J. Phys. Condens. Matter*, **11**, 6487–6499.
Winkler, B., Pickard, C. J., Milman, V., Klee, W. E. & Thimm, G. (1999). *Chem. Phys. Lett.* **312**, 536–541.
Zettlemoyer, A. C. (1969). *Nucleation.* New York: Dekker.