Structure prediction in solid state chemistry as an approach to rational synthesis planning*

M. Jansen, J. C. Schön
Max-Planck-Institut für Festkörperforschung
Heisenbergstr. 1, D-70569 Stuttgart

July 22, 2011

Abstract

Traditionally, solid state chemistry has followed the inductive paradigm, where experimental synthesis and observations provide information about the possible compounds in a chemical system and phenomenological and semi-phenomenological models are employed to rationalize a compound's existence (or "non-existence"). Over the past twenty years, a new methodology has been developing, the aim of which is the prediction of chemical compounds without any recourse to experimental information, followed by their synthesis. The founding stone of this deductive approach to the rational planning of solid state syntheses is the global study of the energy landscape of the chemical system under consideration. In this contribution, we present an introduction to the concept of energy landscapes in the context of structure predictions, and its implications for synthesis planning. The latter step gives access to calculate phase diagrams without resorting to any prior experimental information. Particularly noteworthy, the approach developed allows the derivation of extended phase diagrams that include metastable compounds in a systematic fashion.

Key words: structure prediction, energy landscape, rational synthesis planning, phase diagrams

*To appear as a chapter in: Comprehensive Inorganic Chemistry II

1 Introduction

Stable configurations of atoms that can exist for a finite lifetime constitute the foundation of the whole field of chemistry. All properties of a chemical compound defined by such an equilibrium structure are fully determined by the basic physical laws of nature: a stable chemical compound for given boundary conditions displays a distinct composition, topology, electronic structure and a specific set of physical properties. Therefore, stable atomic configurations serve as the basis and starting point for any action taken in chemistry, be it for the purpose of fundamental research or for the development of applications. The most prominent task of chemistry, virtually defining its identity, is to identify and realize stable compounds. From the beginnings of chemistry as a science, and even in its alchemistic predecessor[1], the focus has been the experimental synthesis of new compounds, which then was rationalized afterwards. This inductive approach to chemistry, relying upon more or less heuristic concepts, reasoning by analogy, and chemical intuition to select promising synthesis targets has e.g. led to great success stories in the field of organic and molecular chemistry, exemplified by the retrosynthesis approach of Corey[2, 3]. The general applicability of this methodology rests on two pillars: the ability of the molecular chemist to predict with high reliability, which molecule is capable of existence, i.e. whether a kinetically stable modification exists and how it looks like, and furthermore the great sophistication of molecular synthesis methods that allow the design of a controlled sequence of chemical reactions leading to the proposed compound in a particular modification. In the field of solid state chemistry, however, analogous heuristic concepts such as radius ratio rules[4], specifying preferred coordination polyhedra, have a much more limited range of applicability for the prediction of whether a particular compound is capable of existence or what its preferred structural modification will be.

Thus, it has been recognized that a more systematic and fundamental approach is needed to solve the problem of predicting the structures of crystalline solids. Addressing this fundamental objective has resulted in a fully rational, and thus deductive[5, 6] approach to planning chemical syntheses, where the goal is to first predict possible new compounds as synthesis targets without any experimental input, and subsequently to synthesize them in a controlled manner. This highly versatile approach is based on the energy landscape of the chemical system[7]. There exist an essentially infinite number of spatial arrangements (or configurations) of the atoms that belong

to a given chemical system, and the (meta)stable compounds feasible in the system correspond to those specific subsets of configurations that are stable for given thermodynamic boundary conditions; in particular, at $T = 0$ K, they are local minima of the energy.

Mathematically, every arrangement of the N atoms belonging to the chemical system under consideration is represented by a point on the $3N$-dimensional hypersurface of the potential energy, and all the dynamics of the system is represented by trajectories on this surface. Thus, the knowledge of the properties of the energy landscape is essential, but also sufficient, for the identification of all compounds in the system that are thermodynamically or kinetically stable at the given boundary conditions. While some of these landscape concepts might appear to be rather abstract, they have greatly increased our understanding of what kind of an entity a chemical compound really is, and their application to structure prediction and rational synthesis planning has already led to several exciting success stories. Among the most impressive ones are the successful prediction[8, 9, 10] and subsequent synthesis[11, 12, 13] of the elusive sodium nitride Na_3N in four different predicted modifications at standard and elevated pressures, and similarly the prediction[14] and subsequent synthesis of new modifications of lithium bromide[15] and chloride[16] in the wurtzite structure type.

In this contribution, a detailed introduction to the energy landscape concept for solid state chemistry is given, followed by its application to the prediction of solids, including both individual modifications and whole phase diagrams. Those readers mostly interested in gaining a general overview over the many chemical systems where successful predictions of solid compounds and phase diagrams have been performed might want to start with the example sections (5 - 6) and return to the methodological part (2 - 4) for deeper insight and conceptual background as needed.

The fundamental approach presented here is equally applicable for the prediction of large molecules, clusters, molecular crystals and extended solids, and furthermore relevant for the understanding of amorphous materials. However, due to spatial limitations, the focus is on extended crystalline solids, and the reader is referred to reviews in the literature[17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30] for more information on structure prediction for the other types of systems.

2 The energy landscape concept: configuration space, barriers, lifetimes, local ergodicity

The starting point of the energy landscape approach is the analysis of the chemical system on the level of individual atoms by adopting the fundamental physics point of view, where a chemical system is represented classically as a collection of N atoms. At each moment in time the system can be characterized by two $3N$-dimensional vectors, $\vec{R} = (\vec{r}_1, \ldots, \vec{r}_N)$ and $\vec{P} = (\vec{p}_1, \ldots, \vec{p}_N) = (m_1\vec{v}_1, \ldots, m_N\vec{v}_N)$, giving the positions and momenta of all atoms as function of time. As long as the Born–Oppenheimer approximation holds and zero point vibrations can be neglected, such a classical picture is also obtained in a quantum mechanical treatment after integrating out the electronic degrees of freedom[31]. Since the kinetic energy of the atoms is just a quadratic function of the momenta, the dynamics of the system (via Newton's equations) is given once the Born-Oppenheimer surface over the $3N$-dimensional space of all atom arrangements, the so-called configuration space of the system, is known. This energy hypersurface is commonly denoted as the energy landscape of the chemical system[32, 33, 34, 7, 19, 5, 21, 35, 36, 37, 38]. The first application of energy landscape concepts to chemical systems has been in the area of glasses[32, 33], in order to understand the non-equilibrium features of these systems. On the other hand, the landscape of crystalline compounds is being investigated with the goal of predicting new compounds and developing routes to their synthesis[7, 5], while clusters[34, 21] and large molecules such as proteins and polymers[17, 19, 21] are studied both as model systems for complex landscapes and in order to understand the properties of finite chemical systems such as the existence of magic number clusters or protein folding.

If one tries to analyze the dynamics on this energy landscape, the simplest case is obtained by considering the chemical system at a temperature of zero Kelvin. As a consequence, the kinetic energy gained due to the "downhill" acceleration from an arbitrary starting point on the potential energy landscape is rapidly dissipated away. Thus, no matter which atom configuration one starts with, the trajectory quickly reaches the nearest local minimum of the potential energy, and remains at this minimum configuration for the remaining duration of the experiment or simulation. Thus, the kinetically stable structures of a chemical system at zero Kelvin are precisely all the minimum configurations of the energy landscape. The system is essentially

static, since the lifetimes of these configurations are infinite, at least on the classical level - quantum mechanically, tunneling processes can take place, of course.

The situation is considerably more complex for non-zero temperatures, since now the time evolution of the chemical system corresponds to a complicated trajectory in configuration space. In particular, none of the configurations observed during an experiment or a molecular dynamics simulation of a chemical system at a given temperature perfectly agrees with the ideal crystal structure, no matter how accurate an energy function one employs. Lattice vibrations and other thermally induced excitations are always present in the crystal, and thus what one calls a solid compound is not associated with a single atom configuration but with a large number of such arrangements, whose thermal (Boltzmann) average corresponds to the actually observed crystal structure. In particular, the average of the simulated trajectory over the whole simulation (i.e. observation) time, the so-called time average, is equal to the average over the thermally activated configurations around the crystal structure, the so-called ensemble average. Such an equality of time and ensemble average means that the system fulfills the ergodic condition or ergodic hypothesis[39], which is a fundamental concept in the field of statistical thermodynamics and constitutes a crucial step in going from the classical mechanical description given above to the thermodynamic one. In particular, every chemical compound that is thermodynamically stable, must necessarily obey this ergodicity condition. This also applies to all metastable compounds, with the obvious restriction that the ergodic condition can only hold approximately because of the finite lifetime of a metastable compound. Thus, one can conclude that identifying all metastable compounds of a chemical system corresponds to the determination of all so-called locally ergodic regions on the energy landscape [19].

In the next step, these qualitative considerations need to be translated into a constructive mathematical definition of local ergodicity[19, 40]: For a given temperature T, a subset $\mathcal{R}$ of the configuration space is called locally ergodic on the observation time scale 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 or lifetime, 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 ensures that one can replace the time averages of

observables $O(\vec{R}(t), \vec{P}(t))$ along a trajectory of length $t_{\rm obs} = t_2 - t_1$

$$< O >_{t_{\rm obs}} = \frac{1}{t_{\rm obs}} \int_{t_1}^{t_2} O(\vec{R}(t'), \vec{P}(t'))\mathrm{d}t' \tag{2}$$

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

$$< O >_{\rm ens} (T) = \frac{\int O(\vec{P}, \vec{R}) \exp(-E(\vec{P}, \vec{R})/k_{\rm B}T)\mathrm{d}\vec{P}\mathrm{d}\vec{R}}{\int \exp(-E(\vec{P}, \vec{R})/k_{\rm B}T)\mathrm{d}\vec{P}\mathrm{d}\vec{R}} \tag{3}$$

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

$$| < O >_{t_{\rm obs}} (\mathcal{R}) - < O >_{\rm ens} (\mathcal{R}; T)| < a. \tag{4}$$

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

$$F(\mathcal{R}_i; T) = -k_{\rm B}T \ln Z(\mathcal{R}_i; T) = -k_{\rm B}T \ln \sum_{j \in \mathcal{R}_i} \exp(-E(j)/k_{\rm B}T), \tag{5}$$

and thus apply the usual laws of thermodynamics to the system as long as it remains within the region $\mathcal{R}_i$. Figure 1 depicts a sketch of a simplified two-dimensional landscape where the locally ergodic regions are indicated. Local ergodicity of a region $\mathcal{R}$ on some time scale $t_{\rm obs}$ implies that all the minima $\{M_i\}$ contained in the region are equilibrated among each other on this time scale, $\tau_{\rm esc}(\mathcal{R}) \gg t_{\rm obs} \gg \tau_{\rm eq}(M_i, M_j)$. However, the reverse is not true: for $t_{\rm obs} > \tau_{\rm esc}(\mathcal{R}) \gg \tau_{\rm eq}(M_i, M_j)$, the minima remain equilibrated among each other, of course, but there is a net flow of probability out of $\mathcal{R}$ and thus $\mathcal{R}$ is no longer locally ergodic on the time scale $t_{\rm obs}$ [41, 19, 40, 36].

As a result, for any given observation time scale $t_{\rm obs}$, 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 ergodic region corresponds to a kinetically stable compound of the chemical system on the time scale of observation. It is important to note that the set of all locally ergodic regions on an energy landscape depends not only on the observation time chosen, but also on temperature since the escape time in particular tends to vary strongly with temperature. As long as the probability of being found in one of these locally ergodic regions is much larger than

the probability of being found in a transition region,

$$\frac{\sum_i p(\mathcal{R}_i)}{1-\sum_i p(\mathcal{R}_i)} \gg 1 \tag{6}$$

where

$$p(\mathcal{R}_i) = \sum_{j\in\mathcal{R}_i} p(j) = \sum_{j\in\mathcal{R}_i} \frac{\exp(-E(j)/k_{\mathrm{B}}T)}{Z_{\mathrm{total}}(T)} = \frac{\exp(-F(\mathcal{R}_i;T)/k_{\mathrm{B}}T)}{\exp(-F_{\mathrm{total}}(T)/k_{\mathrm{B}}T)}, \tag{7}$$

the system can be considered to be in (meta)stable thermodynamic equilibrium on the time scale t_{obs} at temperature T. Here, $Z_{\mathrm{total}}(T)$ is the sum over states over the whole configuration space. Note that if the system has been given an essentially infinite time $\tau_{\mathrm{eq}}^{\mathrm{global}}(T) \gg \tau_{\mathrm{esc}}(\mathcal{R}_i;T)$ to equilibrate before we perform our measurement on the time scale $t_{\mathrm{obs}}(\ll \tau_{esc}(\mathcal{R}_i;T))$, the system can be treated as globally ergodic, and the likelihood of finding the system at the time of the measurement in a particular locally ergodic region $\mathcal{R}_i$ is given by $p(\mathcal{R}_i)$. As a consequence, the locally ergodic region with the lowest free energy $F(\mathcal{R}_i;T)$ has the highest probability of being occupied during the measurement, and the compound corresponding to this region is customarily designated to be the thermodynamically stable phase.

Quite generally, at low temperatures, the escape times from the locally ergodic regions are mostly controlled by energetic barriers on the energy landscape and tend to follow an Arrhenius' law. Thus, at very low temperatures individual local minima 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, plus possible electronic and magnetic contributions to the free energy. Usually, the regions with the lowest free energy correspond to crystalline modifications of the system ($\mathcal{R}_{\mathrm{cryst}}$), while structures containing structural or chemical ordering defects ($\mathcal{R}_{\mathrm{defect}}$) are also associated with local minima but with slightly higher energies. At elevated temperatures and/or on sufficiently long time scales, locally ergodic regions will typically encompass many local minima. The most common case is that the region consists of a large basin containing both the perfect crystalline minimum and the minima corresponding to equilibrium defects of this structure, and the free energy includes a contribution due to the various equilibrium defects. Finally, there are parts of the landscape that describe the glassy state. These regions are at best marginally ergodic, i.e. the escape time from the region $\mathcal{R}_{\mathrm{marginal}}$ is essentially the same or somewhat shorter than the equilibration time, $\tau_{esc}(\mathcal{R}_{\mathrm{marginal}}) \approx \tau_{eq}(\mathcal{R}_{\mathrm{marginal}})$. As a consequence, aging processes take place[42] and the marginally ergodic regions grow in size,

until the system has finally, perhaps after a very long time, reached a kinetically stable crystalline state. If these marginally ergodic regions together with the transition regions dominate the landscape, equation 6 is violated, and one would consider the system a glass former, where it is rather difficult to synthesize one of the possible crystalline modifications.

3 Computational tools for the exploration of chemical energy landscapes: search strategies, cost functions

3.1 General aspects of search strategies

In principle, the determination of locally ergodic regions consists of three steps: the generation of a candidate for such a region (which at low temperatures corresponds to identifying an individual candidate structure), the verification that the candidate is locally equilibrated on the time scale of observation, and the verification that the candidate is kinetically stable on the time scale of observation [43]. Nearly all the work performed so far in the field of structure prediction has focussed on the first step. In contrast, the verification that these candidates are equilibrated (step 2) and kinetically stable (step 3) on the relevant observational time scales are in most studies reduced to only checking whether the candidate corresponds to a local minimum of the energy. As a consequence, nearly all predictions found in the literature are only strictly valid for very low temperatures, $T \approx 0$ K.

Clearly, the approach to structure prediction via globally exploring the energy landscape of the chemical system is extremely powerful, since it does not require any assumptions or experimental information about the system under investigation. The only input is the energy function of the chemical system, which is given by the basic laws of physics. Once the chemical system has been defined and its energy function has been provided, the global search will deliver all (meta)stable structures and phases of the system, in principle.

However, the great complexity of the energy landscapes of bulk chemical systems and the high computational cost of evaluating the energy of every configuration encountered on ab initio level limit the usefulness of the most straightforward implementation of such a global search via a systematic scanning of configuration space. Furthermore, one should keep in mind that the search aims at identifying all local minima that are surrounded by sufficiently high barriers, since

every one corresponds to a possible metastable compound. Instead, the high computational effort required has led to the development of a modular multi-stage exploration strategy[7], where the global exploration of the chemical system employs simplified energy functions. The identified structure candidates capable of existence are subsequently refined during a local optimization stage using highly accurate empirical potentials or ab initio energy calculations. Depending on the complexity of the system (formation of complex ions etc.), a second global search stage is sometimes inserted, where groups of atoms (e.g. coordination polyhedra[44, 45, 46] or secondary building units in zeolites[47, 46]) are kept fixed during the second global optimization. Nevertheless, the sheer computer power required makes the systematic determination of all relevant local minima on an energy landscape of a solid without any direct or indirect input from experiment still essentially impossible for systems with more than about 40 atoms / simulation cell, irrespective of the global search algorithm employed.

In contrast, it is relatively straightforward to construct individual structure candidates containing dozens or even hundreds of atoms in a unit cell by copying and modifying known structures of chemically (more-or-less) related compounds from various databases[48] or in generating periodic bond-networks for a given set of allowed bond-connectivities[49]. The main problem when relying on a database lies in the inability to identify new and unusual but nevertheless energetically very low-lying minimum structures that do not exist in databases or do not obey the topological rules, but which one often finds during unbiased global optimizations on energy landscapes. Of course, with the topological network approach, one could in principle construct all possible atom-networks by including all conceivable types of bond-connectivities into the topological network generation; in the extreme limit, this would correspond to systematically scanning the configuration space more or less densely[50]. But then the number of hypothetical candidates would grow exponentially or possibly factorially fast with the number of atoms/cell [51], and the local minimizations required for these network-based structure candidates would also overwhelm the computational resources.

Quite generally, the methods discussed in the literature for generating structure candidates fall into three general classes: 1) direct determination of local minima and structure families of such minima on the energy landscape (see e.g. [7, 5, 43, 52, 35, 53, 36]), 2) chemically inspired [54, 55, 56, 57, 58] and/or systematic construction [59, 60, 50] of hypothetical candidates

including database-driven searches [48, 61], topological bond-network methods employing graph theory and mathematical tiling theory[51, 62, 63, 49, 64, 65, 66, 67] and so-called 'bottom-up' approaches[68, 69, 70, 71] where clusters are first optimized and then assembled to crystal structures, and, finally, 3) dynamical simulations that try to reproduce or imitate the chemical and physical processes leading to new compounds, such as pressure or temperature induced solid–solid phase transitions [72, 73, 74, 75], the sol-gel process [76, 77, 78], synthesis via deposition from the gas phase[79], or crystallization from solution [80] or the melt [81].

3.2 Algorithms for global landscape explorations

When predicting the existence and structure of stable compounds in a chemical system, the central quantities of interest on the system's energy landscape are the local minima and the locally ergodic regions[19] in general. In order to estimate the kinetic stability of these compounds, the barrier structure of the landscape consisting of (generalized) barriers[40] such as energetic, entropic and kinetic barriers[82] needs to be analyzed. To do so, one must identify the saddle points and transition regions[83] that connect these minima, and measure the flow of probability on the landscape. Finally, the local densities of states have to be determined, in order to compute the local free energies of the compounds. For the analysis of the results of these investigations, visualization plays an important role, since the depiction of the high-dimensional complex multi-minima landscape in a simplified fashion yields insights into the properties of, and the dynamics on, such hypersurfaces. The most important tools are graph-based representations, such as tree graphs[84, 41, 85, 86, 87, 88, 89, 90], networks[91, 92, 93, 94, 95, 96, 97, 98, 99, 100], or transition probability matrices[41, 86, 101, 102] and transition maps[103], besides a range of other methods for the reduction of the number of relevant coordinates[104, 105, 106, 107, 108, 83, 109, 110]. However, due to space constraints, landscape representations are not discussed in more detail and the reader is referred to the literature.

In this subsection, an overview is provided over the most popular algorithms that have been employed for the global exploration of energy landscapes. Of course, all these algorithms have been implemented in their own special fashion by the research groups using them. Most of the corresponding computer programs are specialized to one method only, although a few contain several algorithms as sub-modules, such as the G42-code[7] that implements about ten

different global and local exploration algorithms in various degrees of sophistication, such as simulated annealing[111], thermal cycling[112], basin hopping[113], threshold explorations[85], ergodicity search[114], parallel tempering[115], multi-walker annealing, evolutionary search[116], or prescribed path explorations[110], or the GULP-program[117, 118] that incorporates both force-based methods and genetic algorithms[119]. Furthermore, several of the codes can connect to external programs that provide e.g. ab initio or empirical potential energy calculations or sophisticated local minimization techniques. The reader is referred to the individual publications for more details and contact information regarding the authors of the various programs.

3.2.1 Identification of local minima

A large number of methods to identify local minima on the energy landscape have been developed for the solution of discrete and continuous optimization problems in mathematics, physics, chemistry, biology, technology and economics. These have been adapted and applied to the goal of identifying structure candidates in chemical systems as a tool for rational planning of syntheses[7, 5]. The two crucial issues are the efficiency of the search, and the effect of simplifications and modifications of the energy landscape with the goal of speeding up the search procedure. One should note that in the context of synthesis planning, one is not only interested in the global minimum – all minima with low energies and sufficiently high barriers surrounding them are of importance since they represent possible metastable compounds.

Quite generally, the global optimization procedures can be divided into two main groups, which overlap to a certain degree: Stochastic approaches and deterministic methods. Fundamental to the stochastic methods is the concept of a random 'walker' (or a set of such walkers) on the landscape, whose trajectory on the landscape describes the time evolution of the chemical system. For example, in a Monte Carlo simulation, the trajectory consists of a stochastic sequence of jumps of the walker from a given configuration with energy E_i to neighbor configurations with energies E_{i+1}, which are accepted if $\Delta E = E_{i+1} - E_i < 0$ or if $\exp(-\Delta E/T) > r$ for a random number $0 \leq r < 1$. In principle, these jumps can be generated by some arbitrary, non-physical rule, allowing very large steps on the landscape. In contrast, molecular dynamics simulations reproduce the actual physical trajectory of the system in a deterministic fashion, although the coupling to an external heat bath introduces a certain degree of stochasticity. The

classic deterministic global optimization methods, on the other hand, employ systematic scans of the landscape based on various heuristic or exhaustive rules. They are most useful, if the landscape can be divided into a hierarchy of regions each containing a number of local minima, thus enabling a divide-and-conquer approach.

Simulated annealing The most straightforward stochastic optimization approach consists in performing very long Monte-Carlo or molecular dynamics simulations at constant temperature, where, by definition, every possible minimum region is visited according to its Boltzmann probability. However, much computing time is essentially wasted by exploring high-energy regions of the landscape. If enough computer power is available or the landscape contains only few important basins and a simple barrier structure, such a constant temperature simulation can yield satisfactory results. However, the study of landscapes belonging to complex chemical systems such as large molecules, e.g. clusters or proteins, glassy systems, or non-trivial crystalline solids, shows that more efficient algorithms are required.

This has led to, among other methods (see below), the development[111, 120, 121, 122] and the subsequent application to crystal structure prediction[7] of the so-called simulated annealing algorithm, where during the MC/MD simulations the temperature is decreased, in order to focus the exploration on regions where deep-lying minima are expected to be found, since it can be proved that a logarithmically slow schedule $T(t) \propto A/(1 + \ln t)$ guarantees that one will find the global minimum[123]. One great advantage of employing MC based global optimizations is that one can replace the physically realistic moveclass of moving one or a few atoms by a small amount, by more optimization effective moveclasses that allow larger changes in the atom configuration during each move. One should be aware of the inherent danger, however, that an unphysical moveclass can lead to the inadvertent elimination of physically relevant local minima. But if one is reasonably careful when designing the moveclass, this drawback is outweighed by the higher overall efficiency of the algorithm. In this case, it is sometimes efficient, to combine such large moves every time with a quench; this scheme is often called basin hopping[113, 124, 125].

Besides the moveclass, there are a number of other features of simulated annealing that can be adjusted to increase the efficiency of the algorithm[126, 121]. The temperature schedule $T(n)$, where n counts the number of moves along the trajectory, can be optimized; common sched-

ules are exponential or linear decrease with n, schedules involving temperature cycling[112, 127] where the temperature periodically increases and then decreases again, and adaptive schedules[128, 129] that take properties of the landscape explored up to now into account. Multi-walker implementations have also been used[130, 126], such as the Demon-algorithm[131], methods that generate an averaged landscape[132, 133, 134, 135] such as conformation-family Monte Carlo[136], the particle swarm method[137], superposition state molecular dynamics[138] and SWARM molecular dynamics[139], or multi-overlap dynamics[140, 141, 142], parallel tempering[115] and J-walking[143] where different walkers run at different temperatures and periodically switch positions (or temperatures), in order to overcome barriers more efficiently. Finally, the acceptance criterion can be modified; the most popular ones accept a move according to the classical Metropolis distribution[144], the Tsallis distribution[145, 146], or based on a temperature dependent acceptance-threshold[147]. For more details we refer to the literature[121, 122].

Taboo searches and lid-based optimizations A different class of global optimization methods are the so-called taboo searches[148, 149, 150], where one performs MC/MD simulations and already visited regions are forbidden, either by rigid exclusion constraints or via penalty terms added to the energy function[151]. A crucial issue here is the length of the memory chain, i.e. the size of the list of configurations already visited that can be kept in memory. Combining a taboo search with quenches and large moves like in basin hopping schemes can alleviate the memory problem to a certain extent[152]. One of the more recent incarnations of a taboo search is the metadynamics approach[153, 154]. Furthermore, there exist a number of methods designed to achieve a more efficient barrier crossing, which are based on the modification and/or simplification of the energy landscape[151, 155, 156, 157, 158, 159, 160, 161, 162, 163], e.g. by locally elevating visited areas[151] (a precursor of metadynamics), by lowering barriers relative to the local minima[159], stochastic tunneling[164, 160, 165, 166], dynamic-lattice searching[167], or by modifying the potential between the atoms[157]. Typically, the landscape modifications during such accelerated (molecular or Monte Carlo) dynamics runs are adaptive, i.e., they vary with the progress of the simulation and depend on the information already gained about the system, in this way being related to the taboo-searches.

Finally, there exist several lid-based methods for the stochastic global exploration of con-

tinuous energy landscapes. Examples are the deluge algorithm[168], where an energy lid that must not be crossed during the random walk (like a Monte Carlo simulation where every move is accepted, i.e. where $T = \infty$) is slowly lowered from very high lid values, and the threshold algorithm[85], where the walker is allowed to move below an ascending sequence of energy lids, and one checks periodically, whether new local minima have been reached by performing quench runs from stopping points along the trajectories.

Genetic and evolutionary optimization Genetic or evolutionary algorithms[169, 170, 119, 171, 172, 173] have become very popular in crystal structure prediction over the past decade[174, 118, 175, 176, 177]. They essentially produce a stochastic evolution of an interacting ensemble of walkers which exchange information about the regions of the landscape explored by generating new configurations via a mixture (crossover move) of two or more walkers from one or more previous ensemble generations[170, 178], in addition to simple modifications of individual walkers (mutation move). These "mixtures" can consist of joining pieces of the (two or more) parent configurations, averaging the parent configurations, or combinations thereof. Based on the energies of the members of the extended ensemble, the size of the ensemble is reduced by elimination of the "weaker" members according to some selection scheme. One feature of this reduction process is that one often inserts a preparatory step before the actual selection, where one assigns a fitness value to the configurations based on some rule. One might use the ratio of the energy to the average energy of the population[119], the rank of the configuration by energy[179], or generate the fitness from the rank assigned via some sampling method such as a tournament[180], where the members of the enlarged ensemble are compared pairwise. Furthermore, one often employs hybrid moves, i.e. one adds a local minimization after the cross-over has taken place but before a fitness is assigned[171, 181].

As in the case of simulated annealing, many possibilities for the optimization of the algorithm have been explored. Besides the obvious importance of designing efficient mutation and cross-over moves that are adapted to the energy landscape of chemical systems, two further aspects need to be addressed when optimizing the moveclass: the ratio of mutation to cross-over moves, and the fraction of completely new randomly generated configurations that are added to the ensemble, in order to keep the diversity of the ensemble sufficiently high such that the

algorithm does not end up concentrating all walkers into the same basin on the landscape. Regarding the optimization of the selection procedure for reducing the enlarged parent+offspring ensemble to the original size, three variants appear to be most commonly employed: "Quench" selection, where only the best configurations are kept, "Monte Carlo" selection, where the best configurations are kept but according to their energies weighted by the appropriate Boltzmann factor, or a related (weighted) "roulette-wheel" selection, where the new ensemble is generated by a stochastic sampling with (or without) replacement proportional to their energy or fitness, or by a so-called stochastic universal sampling[182]. Furthermore, one often keeps some of the newly generated configurations in the ensemble, irrespective of their energies, or, alternatively, retains some of the parent configurations in spite of their not so favorable energies, in order to maintain a high diversity of the ensemble[183]. This can effectively lead to multi-generational selection strategies, which are common in some classes of evolutionary optimizations[170, 178].

A feature of special relevance for structure prediction is the representation of the members of the ensemble for the purpose of cross-over moves, i.e. either as a genotype[119, 184] (discrete encoding of an atom configuration) or as a phenotype[116, 185, 186] (the actual atom configuration). In particular during phenotype-based optimizations, hybrid moves are commonly introduced[116, 187], similar to the basin hoping procedure. One problem with evolutionary algorithms is the limited physical realism of the moveclass which makes it difficult to estimate the stability of the minima and can lead to the elimination of a number of potentially relevant local minima. In this context, one should also mention the extremal optimization method[188, 166], which is another ensemble-based exploration approach that produces a very wide spread in the energy distribution of the local minima that can be potentially quite useful when dealing with large energy barriers on the landscape.

Deterministic and multiple-quench optimizations The most basic deterministic global optimization method are exhaustive searches, where the whole energy landscape (or pockets thereof) are explored in such a way that one is guaranteed to find the global minimum. These methods are most applicable if the energy landscape is discrete or can be easily discretized via fast local minimizations. Branch-and-bound methods[189, 92] can be quite effective, and similarly the lid algorithm[41, 190, 191] which is based on a complete enumeration of all states

reachable from a starting point without crossing a sequence of prescribed energy lids. However, the system size usually becomes a serious problem as the number of atoms/cell increases.

Combining an exhaustive, heuristic or stochastic search for promising regions on the landscape with deterministic local minimizations leads to the many multiple-quench approaches found in the literature. Basically, the local minimization procedure is repeated for a very large number of starting points that are generated either by systematically[50] or randomly scanning the configuration space[192, 193] or by chemically inspired choices[56], e.g. via network model generation[194, 195, 196] or the selection of structures from databases[48, 61, 197]. Similarly, one can generate such starting points from long stochastic simulations where periodically quenches are performed along the trajectory of the system[198, 199].

3.2.2 Identification of locally ergodic regions at elevated temperatures

The most straightforward approach applicable for finding locally ergodic regions at low and moderate temperatures consists in finding the local minima of the energy landscape as discussed above and checking that they are surrounded by sufficiently high energy barriers[85, 103]. At higher temperatures, one can use long Monte Carlo or molecular dynamics simulations and attempt to visually identify stable structures about which the system oscillates, even if the structure does not correspond to a single, or even any, local minimum of the energy landscape[72, 200, 75]. A more systematic approach is the ergodicity search algorithm[114], where one registers the fluctuation of indicator variables, e.g. the potential energy or the radial distribution function, within time windows during long simulations at constant pressure and temperature. 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, and one or several configurations within the time window are saved. At the next stage, swarms of short simulations starting from these points along the trajectory in the time window are employed to verify, whether the system is in local equilibrium in this region, and 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. Thus, this procedure will be most useful for finding locally ergodic regions that either contain no clearly delineated local minimum and are stabilized only by entropic barriers[40], or consist of very few

minima separated only by a small energy barrier.

Clearly, the restricted focus on individual local minima is not sufficient if one wants to determine locally ergodic regions that encompass many local minima, which are often important at high temperatures, without recourse to experimental data. Two such types of locally ergodic regions correspond to phases that exhibit local or global controlled disorder[37], respectively, resulting in a configurational entropy contribution to the free energy of the region. Controlled disorder is often associated with locally ergodic regions containing many local minima that can equilibrate to a certain degree on the relevant time scales, while global controlled disorder is found in solid solutions or compounds with partially occupied atom positions. Identifying locally ergodic regions containing many local minima requires the analysis and structural comparison of all the local minima found during the global optimization. If a subset of these minimum configurations can be assigned to a particular structure family, and the energies of these minima are approximately the same[43, 201], a candidate for a partly disordered phase at elevated temperatures has been found.

3.2.3 Analysis of the barrier structure: saddle points, transition paths and probability flow

Studying saddle points, and transition paths in general, yields information about the barriers that surround local minima and locally ergodic regions, and thus allows to estimate the lifetimes of the hypothetical metastable compounds. Furthermore, the knowledge of the transition path between two locally ergodic regions representing two phases[202, 203] is also a crucial prerequisite when studying the mechanisms of phase transitions or of chemical reactions.

In general, the methods to find saddle points are considerably more involved, computationally more expensive and often less robust than the methods employed to determine local minima. One popular approach to identify saddle points proceeds by first finding the minima of the $|\nabla E|^2 (\geq 0)$ surface and then checking, which among these points are actual saddle points of the original energy surface[204]. Alternatives are the slowest slides procedure and other closely related methods[205, 206, 34, 207], and eigenvector following[208, 209].

If two minima in close proximity are known, one can use elastic band methods and related procedures[210, 211, 212, 110] to identify a locally optimal path connecting these minima. In

general, however, both the individual phases and the transition regions between them tend to contain many minima; thus, alternative methods are employed, in particular the so-called transition path sampling[213, 214, 215, 216], where one performs essentially a Monte Carlo walk on the space of transition paths, and related procedures[217, 212, 218, 219, 220]. Based on information gained during such a walk, one can compute the escape times via transition state theory[221, 222, 223, 224].

The metadynamics scheme[153] mentioned earlier and related methods[225, 226] provide another alternative, where only a starting minimum/phase is provided, and one expects the system to find the new phase 'on its own', by slowly building up a penalty type potential inside the starting region which makes this region energetically unfavorable and forces the walker to leave the region. This taboo-like search works in principle at any temperature, and one can measure the local free energy in the process, too[227, 228], but the metadynamics requires a good choice of an order parameter, which allows to distinguish between the original phase and the new one(s), which is a non-trivial issue. The same holds true for alternative methods such as some type of coarse molecular dynamics[229, 230, 231, 232, 233, 234] or various steered dynamics procedures[235, 236, 237, 238, 239, 240, 241].

An important measure of the stability of locally ergodic regions is the size of the probability flows among the locally ergodic regions on the landscape. For simple known transition paths, standard transition state theory yields the likelihood of moving between the various minima from the knowledge of the minima, and of the energy and width of the saddles[21]. However, in general, one does not deal with simple two-minima-plus-saddle systems, and one needs to use an alternative such as the threshold algorithm mentioned above, where the probability flow is measured not as function of temperature but instead as function of energy slice for each energy lid[85, 103]. No target is given, the density of states is sampled, and all minima that have been identified serve as starting points for new threshold runs. By performing many quenches along the way, one can identify all local minima on the landscape, gain estimates on the energetic and entropic barriers and determine the size of the transition regions by identifying the so-called characteristic regions[83] as function of energy slice.

3.3 Approximate description of the energy landscape and cost functions

Since a typical set of global optimization runs for the determination of the local minima involves millions or even billions of energy evaluations, a modular multi-scale approach has become standard[7, 45, 242, 175, 36]. A global search on an empirical or simplified ab initio energy or cost function landscape generates structure candidates, which are subsequently locally optimized on the full quantum mechanical level using e.g. the Hartree-Fock approximation or density functional theory[243, 244, 245]. Here, different levels of approximation and subsequent refinements for the description of the energy landscape are applied, both with regard to the choice of energy function (e.g. ab initio energies[176, 246, 247, 248], empirical atomic interaction potentials, or atom-group based cost functions[47, 46]), and concerning the quantities that describe the atom configuration and are varied during the global optimization (such as single atoms, groups of atoms[46], or nodes in bond networks[65]).

In addition to the level of approximation in the energy calculations, one needs to be aware that, in principle, not only the atoms' positions and their electronic states, but also the parameters and content of the simulation cell can be varied as part of the global search. Thus, the general cost function is actually not just the energy but the thermodynamic potential of the grand isobaric ensemble at pressure p [7]

$$C = E + pV - \sum_{i=1,N_{spec}} \mu_i N_i, \tag{8}$$

where $\sum_{i=1,N_{spec}} N_i = N$, and N_{spec} is the number of different types of atoms involved. One potential difficulty one encounters is choosing an appropriate value for the chemical potential of species i; typically, one uses the standard enthalpy of formation / atom of the element (as function of temperature). However, experience has shown that it often 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.

3.3.1 Ab initio energy landscapes

While already in the 1990s the global exploration of the ab initio energy landscape of few-atom clusters had become feasible[249, 250, 251], computers are only now reaching the speed and ubiquity that allow the global optimization on the ab initio level for extended solids as well.

Clearly, there are several reasons why it is desirable to employ ab initio energy calculations already during the global exploration stage[247, 248, 252]: Ab initio energies are expected to be more accurate as they mainly depend on the functional but are otherwise free of empirical parameters. In addition, they are more generally applicable than empirical energy functions, e.g. empirical potentials work well for ionic systems, but less so for covalent or metallic ones. In particular, ab initio energies are required if the type of bonds that will be formed is not known in advance. In this case, choosing a certain model potential to describe the interaction would introduce a dangerous bias in the calculations.

The major disadvantage of using ab initio energy functions is the computational cost, which can easily slow the calculations by a factor of 10^3 to 10^5 compared to calculations based on empirical potentials[247, 248, 252]. One should recall that all calculations have to be performed in space group $P1$, because all possible atom arrangements must be accessible during the global exploration, and thus no restrictions on the symmetry of the configurations are allowed.

One way to achieve the crucial speed-up is to use slightly less accurate but much faster ways to compute the ab initio energies, where once again, as with empirical potentials, one exploits the fact that it is often sufficient to know the structure candidates only approximately at the end of the global optimization stage. Only the final local optimizations that yield the correct ranking and the accurate cell parameters and atom positions of the minimum structures need to be performed with good parameters. Possible ways to speed up the ab initio energy calculations are to reduce the number of matrix elements or integrals that must be computed, to use a smaller basis and fewer $\vec{k}$ points, or to reduce the number of points of the density functional grid, depending on what type of ab initio calculation one performs[247].

Here one should note that the necessity to converge the ab initio calculations for random atom arrangements poses an additional complication during global explorations of energy landscapes of chemical systems: the Hartree-Fock or Kohn-Sham equations are often difficult to converge for configurations which are far away from realistic solid state structures. The main reason is that the initial configurations resemble a gas, and thus the band structure is similar to that of localized electrons, with nearly no dispersion, and with a much smaller gap than the one of the final structure. This influences the choice of the ab initio method one would use for the global search: it appears that Hartree-Fock calculations facilitate convergence at even the most unusual

geometries, due to the relatively large band gaps typically produced by this method, whereas calculations with the local density approximation may exhibit severe convergence problems[247, 248, 252]. Therefore, especially for non-metallic systems, the Hartree-Fock based energy function is often the best choice. For more details, we refer to the literature[247, 248, 252, 37]

3.3.2 Empirical potential landscapes

While the use of ab initio energy landscapes during the global exploration is clearly preferable, in principle, it is often not feasible to do so, especially if a system with many atoms in the simulation cell needs to be studied. Thus, one will take the risk of missing some potentially important structure candidates as long as one has a good chance of finding most of the relevant modifications. Quite generally, there exist a large number of empirical potentials of ever increasing complexity that have been employed to model solid compounds, in particular ionic ones. These range from two-body potentials like the Born-Mayer-potential[253], the Buckingham-potential[254], simple robust Coulomb-plus-Lennard-Jones potentials[255] or a refined version thereof with environment dependent radii[243, 256], over dipole[257] and quadrupole[258] shell models, to various kinds of breathing potentials[259, 260, 261, 262, 263, 264, 265, 266] of increasing levels of refinement.

While the more refined empirical potentials typically allow the computation of at least some properties of a given material with relatively high accuracy, it is not clear, whether their use during the global optimization is worth the cost as long as one is only interested in identifying possible structure candidates in a chemical system. For one, such potentials tend to be much more complicated and computationally expensive than the simple potentials commonly employed. Secondly, the number of parameters involved typically increases with the complexity of the potential, which makes it very difficult to construct such a potential without detailed a priori experimental knowledge of the system. However, this contradicts the premise of the whole enterprise of structure prediction, i.e., to identify the stable and metastable modifications of a chemical system without any prior information except the identity of the participating atoms. Of course, one can attempt to fit the parameters of the empirical potential to ab initio calculations in the system, but this typically involves a large amount of effort. Finally, it has often been observed that many of these potentials are not globally applicable. They strongly favor the

structure(s) to which they have been fitted, and even successfully reproduce their properties. But at the same time, these potentials often weaken or even eliminate the minima representing important alternative modifications on the landscape of the chemical system.

In contrast, the parameters of the simple two-body potentials are often transferable and can be derived from average ionic properties such as the radius and the hardness of the ion, but they lack the precision of the refined potentials, of course. Thus, in order to overcome the limitations of the simple potentials as much as possible while still retaining their advantages, one should repeat the global explorations for slightly varied values of the parameters characterizing the potential.[255, 7] The experience over the past two decades has shown, that the most important structure candidates usually are quite robust, i.e., they are found as local minima on many of the slightly modified energy landscapes. Furthermore, this also alleviates the concern that for many elements the parameters that enter the potential such as the effective ionic radius of an atom in a compound vary e.g. as function of the number of atoms in the first coordination sphere around the atom.

One interesting alternative approach that has the potential for generating robust approximate energy functions has recently been developed[267, 268], where one employs a neural network to optimize the implicit parameters in a generic empirical neural-network based energy function. For the training of the network, a test set of ab initio energies of (randomly chosen) atom configurations is used. While being a very promising approach, it still requires a very substantial computational effort during the learning phase of the neural network, and the transferability of the parameters to related but different chemical systems is rather limited.

3.3.3 Cost function landscapes

A second alternative to ab initio energy functions are so-called (non-physical) cost functions that contain, or wholly consist of, terms that reflect special chemical knowledge about the system under investigation. These additional terms incorporate empirical features, such as the validity of the bond-valence rules[269], or the existence of geometric and/or topological requirements of the structural elements in the chemical system. Similarly, already available structural information can be included into the cost function, e.g. by adding a penalty term to the energy, which measures the deviation of the proposed atom configuration from the data

available from experiment. Depending on the degree to which one "trusts" the experiment or the energy calculation, one can weigh the contributions of the energy E and the penalty terms P.

$$C = \lambda E + (1 - \lambda)P, \tag{9}$$

where $0 \leq \lambda \leq 1$. Alternatively, one can directly implement the desired constraints by restricting the configuration space the walker is allowed to explore, by e.g. fixing the shape and volume of the simulations cell or introducing rigid (or flexible) building units that consist of several atoms that are connected according to a predefined topology, such as primary and secondary building units [46].

3.4 Structure analysis tools: SFND, RGS, CMPZ

An important step during a structure prediction study for extended solids 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, due to numerical aspects of the algorithms, all these minima show slight minute deviations from the more symmetric atom arrangements that correspond to the exact location of the minimum. Finally, the structures obtained are usually not given with the unit cells according to crystallographic convention. To address this issue, two algorithms, SFND [270] and RGS [271] have been developed, which allow the determination of 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 the many local minima produced by the optimization, to compare the structures found with already known structure types listed in the ICSD, and finally to identify structure families by comparing e.g. the cation-anion superstructures in multi-cation/anion compounds. To deal with these tasks, the CMPZ algorithm [272] has been developed, which allows the comparison of two arbitrary periodic structures, by generating a mapping of the two infinite periodic atom arrangements onto one another. As criterion for similarity, one measures the deviations between the cell parameters of the appropriately transformed cells together with the deviations of the atom positions within these cells. One

should note that this is a geometric criterion for similarity, not a symmetry-based or topological one. All three algorithms mentioned have been implemented in the structure analysis program KPLOT [273] available at[274].

4 Moving to thermodynamic space: densities of states, free enthalpy, phase diagrams from first principles

When moving from the atomistic energy landscape to thermodynamic space, the next step after the identification of possible (meta)stable phases is the computation of the local free energies for the (meta)stable compounds that are kinetically stable for a prescribed set of thermodynamic boundary conditions (pressure p, temperature T, phase composition x_i, ...), for the given observation time t_{obs}. Usually, this requires the determination of the density of states, using one of the methods described below. From this information, one can derive an extension of the standard equilibrium phase diagrams, which adds the observation time t_{obs} as an additional coordinate. In such a diagram, one marks for each of the possible phases $\mathcal{P}_i$ those regions in thermodynamic space where $\mathcal{P}_i$ is kinetically stable on the timescale t_{obs}. The standard equilibrium phase diagram, where at each point in thermodynamic space only one phase (or weighted combination of several phases) is plotted, is then equivalent to the $(t_{obs} = \infty)$-slice of the extended $(p, T, x_i, \ldots; t_{obs})$-diagram[275]. Here, one should note that once the phases that can exist have been established and their free energies have been computed, standard procedures can be used for the depiction of the phase boundaries of each (meta)stable phase for the individual t_{obs}-slices of the extended phase diagram. For this purpose, a number of computer programs[276, 277, 278, 279] have been developed over the past four decades within the CALPHAD effort [280, 281]. We refer to the extensive literature for more details on the various interpolation and construction methods involved.

4.1 Measurement of local and global densities of states

The most straightforward method to measure local densities of states is to sample them via long unbiased random walks[282, 283], and then to normalize them to e.g. the vibrational density of states around local minima. The classical histogram methods belong to this group[284, 285]. The

most obvious problem are the long simulation times; other issues are to separate the contributions of many different minima. As a consequence, re-weighting methods have been developed[286, 287], where one performs the simulations e.g. for many different temperatures and re-scales the sampled densities of states with the acceptance probabilities such that the simulation effectively corresponds to diffusion on a 'flat' landscape. Many different schemes have been proposed to achieve this end: weighting of histograms on the fly, e.g. WHAM[288, 289], re-analysis of data of constant temperature runs taken e.g. from expanded ensembles[290], parallel tempering or multi-canonical simulations[291, 286, 292, 293, 294, 142, 295, 296, 297, 298], modification of the landscape such that it effectively becomes flat while keeping track of the measured density of states[299] as in metadynamics simulations, global or local transformation of the landscape using e.g. hyperbolic functions to make the landscape look 'flat' and re-weight the sampled DOS afterwards[286]. Important issues are the statistics, of course, the homogeneous sampling of different metabasins containing many local minima, and the overlap of distributions taken at different temperatures.

Similar concerns arise when mapping the DOS measured within overlapping energy slices, with subsequent slice-matching as one does when using e.g. the threshold algorithm[85]. The normalization in particular is usually performed with respect to the phonon spectrum of individual minima. Finally, for low temperatures, where most of the sampled DOS during a realistic physical trajectory inside a locally ergodic region comes from within the harmonic regions around the local minima and saddle points, one can use analytical methods to compute the density of states from the phonon spectrum of the Hessian of the minimum under consideration[300]. Even anharmonic corrections[301, 302] and magnetic contributions[303, 304, 305] can be included to some degree. Alternatively, one can compute the phonon spectrum from simulations e.g. via the Fourier transform of the velocity autocorrelation function[306].

4.2 Computation of free enthalpy

Thermodynamically, the free enthalpy $G(p, T, x_i, \vec{B}, \ldots)$ of a phase is given by the ground state free energy plus the free energy contributions due to the various excitations that are present in thermodynamic equilibrium at given temperature, pressure, magnetic field, etc. As long as the various excitations with respect to the energy minimum can be treated as independent, the sum

over states is given as a product of separate partition functions. Then the free enthalpy can be written as

$$G = G_0 + G_{vib} + G_{electr} + G_{defects} + G_{magn} + \ldots, \tag{10}$$

and standard formulas can be employed to evaluate each contribution separately. Once this is no longer possible, e.g. because of high temperatures or due to a strong coupling between vibrational and electronic or magnetic excitations, more complex methods must be employed. Here, two different types of calculations need to be distinguished: Energy landscape based methods, and procedures that can deal with phases encompassing so many local minima on the energy landscape that special tools based on statistical mechanical models are necessary to evaluate the free energy[307].

The most straightforward approach to compute the free energy of a system is via the (local) density of states of a locally ergodic region on the energy landscape. As far as the accuracy of such a procedure is concerned, the main issue is the quality of the density of states used as input for the calculation. An alternative procedure is the computation of the free energy of a system A via the difference between two free energies belonging to systems A and B, where the free energy of system B usually serves as a reference energy that can be computed analytically (e.g. the free energy of a system of Einstein oscillators) or is known from very detailed computer simulations. Three slightly different approaches have been developed for this purpose, called thermodynamic integration[308], thermodynamic perturbation[309] and computational alchemy[310]. One feature common to all of them is that one chooses some path between the two systems, along which one smoothly transforms one of the two systems into the other and back, and thus derives bounds on the free energy difference between A and B. While the choice of path and the allocation of time along the path is very important for the efficiency of the algorithm[311], the path as such does not necessarily have to be dynamically meaningful. Similarly, one can also directly compute the free energy along a chosen path, e.g. in order parameter space or along a reaction coordinate, via integration over the remaining coordinates[312, 313], or via the average force calculated along the path[314, 315, 316]. Ideally, one performs such computations along a true simulation trajectory using umbrella sampling[317]. Finally, grand canonical[318], semi-grand canonical[319, 320] and extended Gibbs ensemble[321] computations where atoms are essentially moved between two copies of the system[322], one describing phase A and the

other phase B, (in equilibrium) have become quite popular for computing phase equilibria and differences in free energies. Nevertheless, sampling error inaccuracies are bedeviling essentially all numerical estimates of free energies derived from dynamical simulations[323].

In contrast, computing the free energies of complex phases such as alloys or solid solutions requires more problem-specific methodologies. Such methods have their origin in simple ideal and regular solution models and have been continuously refined[324, 325, 326, 327, 328, 329]. Nowadays, the cluster expansion[326, 330, 331, 332, 329] and cluster variation[333, 334, 335] methods are among the most popular approaches, together with the explicit computation of the free energy via summing over the local free energies of a large number of atom configurations that are supposed to constitute a representative sample of the set of alloy configurations[336]. A precondition of this approach is that the basic structure (reference lattice) of the alloy is essentially independent of the distribution of the atoms over the structure or sublattices thereof (except possibly for small relaxations away from the reference positions), and thus an approximate energy function can be constructed. Since this function can be evaluated very quickly even for large numbers of atoms/cell, one can now calculate free energies via statistical averages over many thousands or millions of configurations to reasonably high accuracy within relatively short times[337, 338].

4.3 Phase diagrams from first principles incorporating metastable phases

In principle, the determination of the locally ergodic regions, their corresponding phases and the computation of their free energies as function of the thermodynamic parameters $(T, p, x_i, \ldots)$, provides all the information necessary for the construction of the equilibrium phase diagram of a chemical system. For every given value of $(T, p, x_i, \ldots)$, one can compare the free energies of all phases in the system and select the phase with the lowest free energy at this point, i.e. one constructs the convex hull of the free energy surfaces of the phases. Projecting this convex hull onto the space of thermodynamic parameters gives the classical equilibrium phase diagram of the system.

However, trying to incorporate all metastable modifications and phases into the traditional thermodynamic description of a chemical system, a task following directly from the goal of comprehensive unbiased structure prediction, requires the development of an appropriate gener-

alization of the classical phase diagram,which should contain the traditional equilibrium phase diagram as a limiting case. Seen from the energy landscape point of view, the natural choice for an additional coordinate in such an extended phase diagram is the observation time. For a given observation time t_{obs}, we can determine all phases that are kinetically stable for prescribed thermodynamic boundary conditions $(p, T, \ldots)$, i.e.: at this point in e.g. the (p, T)-plane, $n(p, T; t_{obs})$ phases $\mathcal{P}_i$ are capable of existence. Since for each of these phases $\tau_{esc}(\mathcal{P}_i) \gg t_{obs}$, it does *not* matter, whether $F(\mathcal{P}_i)$ is the lowest free energy or not: If the system had been prepared at time $t = 0$ in the locally ergodic region $\mathcal{R}_i$ that is associated with phase $\mathcal{P}_i$, then when the observation time $t = t_{obs}$ has been reached, we still reside inside $\mathcal{R}_i$, and thus the potential existence of other kinetically stable phases $\mathcal{P}_j$ is irrelevant, even if $F(\mathcal{P}_j) < F(\mathcal{P}_i)$.

Of course, for each value $t_{obs}^{(0)}$ of the observation time, the corresponding $(t_{obs} = t_{obs}^{(0)})$-slice through the extended phase diagram is going to look slightly different. For very short times, $t_{obs} \to 0$, only the individual local minima on the energy landscape can be locally ergodic, i.e. at each point in thermodynamic space, all the local minima on the energy landscape individually belong on the list of kinetically stable phases. In contrast, in the opposite limit, $t_{obs} \to \infty$, only the thermodynamically stable phase survives, and the $(t_{obs} = \infty)$-slice of the extended phase diagram is equivalent to the classical equilibrium phase diagram. One should note that this strictly applies only for macroscopic systems where the thermodynamic limit, $N, V \to \infty$ with $N/V =$ constant, exists. For finite-size systems such as clusters or proteins, even metastable regions can be observed with with a finite probability for infinite observation times.

An important point is the correct interpretation of the co-existence of these many phases at a particular point $(p, T, \ldots)$ in thermodynamic space for a given observation time t_{obs}. In equilibrium phase diagrams, Gibbs's rule holds, and the existence of two phases at the same point implies that their free energies are equal. This is typically not the case for the many metastable phases that can co-exist on the finite timescale t_{obs} although they usually have different free energies. This is a major consequence of the concept of local ergodicity: Once the system is prepared in a locally ergodic region of the landscape, then for observational timescales larger than the equilibration time of this region but smaller than the escape time from the region, we can treat the system with our traditional concepts of equilibrium thermodynamics restricted to the subset of phases that can dominate for various thermodynamic boundary conditions within

this locally ergodic region. In particular, Gibbs's rule holds for all phases present within such a region, and we can derive e.g. miscibility gaps or other solid-solid co-existence curves for this subset of phases.

Finally, one notes that systems such as glasses that have not reached at least local equilibrium after a relaxation over a time period t_{obs} cannot be represented in the extended phase diagram. The reason is that the extended phase diagram incorporates all phases that are both kinetically stable and locally equilibrated on some timescale t_{obs}, but excludes all thermodynamic non-equilibrium states. Of course, predicting the complete extended phase diagram of a chemical system is usually extremely expensive computationally, since all the locally ergodic regions on the landscape for all values of the thermodynamic boundary conditions must be determined. Nevertheless, the construction of this generalization of the classical phase diagram is the crowning step of the unbiased structure prediction in a chemical system.

5 Examples

The examples presented in this section are divided into two categories: 1) examples illustrating various general aspects of structure prediction, and 2) a short overview together with a selection of studies drawn from the field of structure prediction, plus some examples from the closely related field of structure determination of existing compounds. However, due to the great increase in the number of publications dealing with structure prediction and structure determination over the past couple of years, only a limited selection can be presented.

5.1 Illustrative methodological examples

5.1.1 Modular approach: study of the alkali halides

To illustrate the general modular approach, we turn to one of the first groups of ionic systems whose energy landscape has been investigated in detail using simulated annealing without recourse to experimental data, the alkali halides[255, 14]. A very detailed study was performed for NaCl[255], where a large number of local minima was found on the empirical energy landscape (Coulomb+Lennard-Jones potential) by global optimization using simulated annealing. The global minimum of the landscape corresponded to the experimentally observed rock salt

structure. The structures of most of the energetically low-lying minima could be identified with typical AB-structure types like NiAs, PtS, CsCl, wurtzite, or sphalerite. However, one deep-lying local minimum, denoted $Na^{[5]}Cl^{[5]}$ (the so-called 5-5-structure type), exhibits a previously unknown structure type. Here, Na^+ and Cl^- are coordinated by Cl^- and Na^+, respectively, in a trigonally bipyramidal fashion, resulting in a topology that resembles the one of hexagonal BN which displays a 3+2 coordination.

Threshold investigations showed that the energy barrier stabilizing this structure was only moderately high (≈ 0.01 eV/atom), suggesting that the structure might be difficult to synthesize with traditional solid state synthesis methods in the NaCl-system. However, several years later, this new predicted structure type was found experimentally [339] 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 found as a minimum on the energy landscapes of many other AB-systems[340, 14, 110], and has furthermore been observed during the growth of ZnO-films [341].

Analogous global optimizations have been performed since for all twenty alkali halides [14] for a wide range of pressures, where all the energies were refined through local minimizations of the structure candidates on ab initio level. Similar to the case of NaCl, many possible modifications were found that included both well-known AB-structure types (rock salt, NiAs, wurtzite, sphalerite, 5-5, CsCl, etc.), and previously unknown structures.

Finally, the ab initio energy landscape of LiF has been explored at standard pressure using simulated annealing, where both the Hartree-Fock approximation and density functionals were used to compute the energy [247]. The relevant minima found agreed with those determined on the empirical potential landscape, including the rocksalt-, the wurtzite- and sphalerite-, the NiAs-, and the 5-5-structure type. This study provided a valuable validation of the many landscape explorations based on empirical potentials and also served as a proof-of-principle for the feasibility of global stochastic explorations on ab initio energy surfaces.

5.1.2 Tree graph landscape representation: the landscapes of MgF_2 and CaF_2

An important step when studying the energy landscape of chemical systems is the construction of a simplified representation of the landscape, typically in the form of a tree graph, which depicts

the relative energies of the minima, together with the energetic and entropic barriers separating them. An early example of this kind for solid compounds is the study of the empirical energy landscapes of MgF_2[103, 83] and CaF_2[103, 40] (Coulomb+Lennard-Jones potential). Besides the global minimum exhibiting the rutile or fluorite structure, respectively, in agreement with experiment, many other minima were found showing e.g. the anatase type, the CdI_2-type, a half-occupied rocksalt structure-type, and structures built up from MgF_7 and CaF_7 monocapped prisms, just to name a few. Figure 2 shows the corresponding tree graph for MgF_2, where the energies at the points where branches connect indicate the energetic barriers, while the grey/black bars indicate the height of the entropic barriers that additionally stabilize the locally ergodic regions.

5.1.3 Free energy tree graph representation: the landscape of SrO

An alternative to the extended phase diagram discussed above is the so-called free energy landscape, typically represented in the shape of a tree graph for every observation time. This representation is particularly useful, if one deals only with ordered crystalline phases, and wants to depict both the free energy and the barriers separating the metastable modifications as function of temperature for a fixed composition and pressure. An example is the free energy landscape of SrO that was constructed by combining runs with the ergodicity search algorithm and the threshold algorithm for a global exploration of the energy landscape [114], 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 [340]. 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[114] (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 rocksalt-type modification (the thermodynamically stable solid modification of SrO at standard pressure, both according to the experiment and the calculations) during very long MC-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 Figure 3, which is valid for moderately long timescales, at least for low and intermediary temperatures.

5.1.4 Multinary phase diagrams: the quasi-ternary semiconductor (Al,In,Ga)-Sb, and similar systems

The ab initio prediction of multinary phase diagrams is nicely demonstrated at the example of the quasi-ternary semiconductor system (Al,In,Ga)-Sb.[342] By globally exploring the empirical energy landscape of all three quasi-binary systems ((Al,Ga)-Sb, (Al,In)-Sb and (Ga,In)-Sb) and the full ternary system for many different compositions, followed by a structure family analysis of the set of candidates obtained and a comparison of their energies on ab initio level, it was found that no ordered crystalline phase should be thermodynamically stable. The thermodynamically stable phase is a solid solution and exhibits the sphalerite structure type, in agreement with experimental observations. Fitting the excess enthalpy as function of compositions x_{AlSb}, x_{GaSb} and x_{InSb} to a Redlich-Kister polynomial[343], and adding the configurational entropy of an ideal solution, yields the free enthalpy $G(x_{\mathrm{AlSb}}, x_{\mathrm{GaSb}}, x_{\mathrm{InSb}}, T)$ as function of composition and temperature. The convex hull of G determines the miscibility dome of the system, which for (Al,Ga,In)-Sb has its maximum inside the quasi-ternary region (c.f. figure 4). So far, experimental thermodynamic data are only available for the high-temperature solidus-liquidus region, and thus a quantitative comparison with experiment has not yet been possible.

The same holds true for the quasi-ternary system (Al,Ga,In)-As[344], whose low-temperature phase diagrams was predicted in an analogous fashion. Furthermore, for more than twenty mixed quasi-binary alkali halides[201, 345, 346, 347, 348], the low temperature phase diagram was determined based on simulated annealing optimizations of an empirical landscape followed by the ranking of the crystalline modifications and the computation of miscibility gaps where

applicable, on ab initio level. The results were in good agreement with experimental data where available, and included the prediction of new ordered crystalline modifications in the RbX-LiX and CsX-LiX (X = halogen) systems.

In this context we note that special attention needs to be paid to the analysis of the many structure candidates obtained during the global searches for different compositions, since one has to be able to distinguish between ordered crystalline phases and solid solution phases, without experimental pre-information. The crucial issue is whether so-called structure families exist among the minima observed for different compositions which have essentially the same energy (for a given composition)[43, 201]. If that is the case, the union of these local minima can be treated as a large locally ergodic region, and the free energy of this solid solution phase contains an entropy of mixing which favors the solid solution over ordered crystalline compounds which correspond to a single minimum basin on the energy landscape. Finally, one can add the free energy contribution of the thermal excitations for the various phases, which become relevant at elevated temperatures[201, 305].

5.1.5 Structure prediction in multinary systems: study of the alkali metal orthocarbonates $M_4(CO_4)$, with M = Li, Na, K, Rb, Cs

When predicting the structure(s) of ternary or other multinary systems, one encounters the problem that for the given composition at which one performs the global search, alternative phases, such as appropriately weighted mixtures of stable compounds of the subsystems that together have the same overall composition as one of the multinary compounds, can exist but which cannot be accessed during a global optimization with a typical (small) number of atoms per simulation cell (< 40). As a consequence, one needs to also perform global searches for all alternative binary, ternary, etc. compounds that might be present in the system at some pressure or temperature.

An example system, where such an extended search has been performed are the still elusive alkali metal orthocarbonates. A promising approach to synthesize these compounds 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). But since the hypothetical orthocarbonate would compete with the high-pressure phases of the corresponding regular carbonates plus oxides, it is necessary for the

study to include, at least, the enthalpy surfaces with composition $M_4(CO_4)$, M_2O and $M_2(CO_3)$ for many different pressures, in order to establish the range of (thermodynamic) stability of the orthocarbonate phase vs. the decomposition into the corresponding oxide and carbonate as a function of applied pressure.

To achieve this, the enthalpy landscapes of M_2O [349], $M_2(CO_3)$ [245] and $M_4(CO_4)$ [46, 350] were investigated for many different pressures using simulated annealing and an empirical Coulomb+Lennard-Jones potential. In a first round of global optimizations, individual metal-, carbon-, and oxygen-atoms were used to describe atom configurations. Since many of 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 fixed CO_3- and CO_4-units were employed together with the metal atoms. All structure candidates found were locally minimized on 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 the one of M_2O + $M_2(CO_3)$ as function of pressure [350]. 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 of 20 - 30 GPa.

5.2 Structure prediction and structure determination using global landscape exploration: Overview and presentation of selected examples

5.2.1 Elemental modifications

With the exception of the noble gases, no simple empirical potentials are available for elemental solids, which are at the same time robust and sufficiently precise to allow the identification of all (meta)stable phases in the system. Thus besides early studies on the noble gases[351] and their mixtures[352], predictions of elemental solids using global optimization techniques have only been performed rather recently; earlier comparisons of $E(V)$-curves for a few typical structure types such as fcc-, hcp- or bcc-arrangements of metal atoms[353] do not qualify as structure prediction in the sense discussed in this contribution. Of particular interest have been

predictions of high-pressure phases, often spurred by some new high-pressure experiments and geophysical questions.

First attempts of predicting elemental carbon structures relied on the generation of periodic graphs consisting of sp^2 [194] and sp^3 [195] hybridized carbon atoms, where these network configurations were subsequently minimized on DFT-level. In this way, a number of metastable networks were generated that may be of relevance in understanding amorphous carbon or nanotube fragments. Similarly, two-dimensional necklace-like carbon structures were obtained using basin hopping simulated annealing with a tight-binding energy function[354]. More recently, the multiple-quench approach using a DFT-based energy function was employed for the determination of high-pressure phases of silicon, carbon, and hydrogen, where a gradient based relaxation was employed[246]. For silicon, the four lowest-pressure modifications were correctly obtained. Furthermore, in recent studies, the large number of possible clathrate-like structures for elemental silicon have been explored[355]. Both a multiple-quench[246] and an evolutionary algorithm[356] were used for silane to predict new high-pressure phases.

Similarly, an evolutionary algorithm was employed with an ab initio energy function for sulphur[176], silicon[357], and carbon[176, 358], to predict stable and metastable phases of these elements at ambient or at (ultra)high pressure. Furthermore, dense phases of lithium[359] and iron at terapascal pressures[360] have been studied, and structure prediction methods have been applied to boron[361] to achieve the structure solution of a previously unsolved high pressure modification of boron[362].

5.2.2 Ionic compounds

In nearly all of the studies of the energy landscape of ionic compounds, empirical potentials of some kind have been employed, since these have proven to be quite useful when modeling the properties of such systems. All types of prediction methodologies have been applied to this class of compounds, from simulated annealing and genetic/evolutionary algorithms to database searches and network generation. The systems investigated span a wide range of anion-cation combinations, preferably of main group elements.

A large body of work deals with oxides, e.g. the alkali[245] and alkaline earth oxides[340], V_2O_5[363], B_2O_3[364], SnO_2[7] and SnO[19], NiO[7], SiO_2[7, 365], $CaTiO_3$[7], $SrTiO_3$[7], $SrTi_2O_5$[7],

$MgSiO_3$[176] and TiO_2[366], which have been investigated to find possible phases at standard and high pressures. Similar studies were performed for a number of nitrides, such as the alkali nitrides[8, 9, 10, 367], Cu_3N[368], $Si_3B_3N_7$ (via simulated annealing[7] and network construction[56]) and other compositions in the Si-B-N-system[369] (via simulated annealing and the threshold algorithm), tantalum nitride (using ab initio energies, together with database information[370] or multiple quenches[371]) and lanthanum pernitride[372] (via minimization of structures taken from a database of AB_2 compounds). Analogous work was performed for the sulfides, e.g. the alkali sulfides[244, 10] and lead sulfide[110] employing an ab initio energy landscape in the latter case, and the carbonates of the alkali metals[349], magnesium and calcium[373, 176, 374]. The landscape of AlF_3 has been studied via the construction of polyhedra-networks[65], and furthermore the chlorides and fluorides of magnesium and calcium[375], $SrCl_2$[7], and all the binary alkali halides[255, 14] have been investigated with simulated annealing, yielding promising structure candidates. In addition, the ground state structures of SiC and GaAs have been found in global landscape studies[357], and the empirical energy landscape of the ternary system $Mg(BH_4)_2$[376] has been explored using simulated annealing.

In addition, a number of studies have been performed, where many different compositions in a binary or quasi-binary system have been investigated. Systems that have been studied in this way are e.g. the FeB-system using ab initio energies with an evolutionary algorithm[177], or the quasi-binary Ca_2Si-$CaBr_2$[256] and MgO-MgF_2[243] systems employing simulated annealing and the threshold algorithm for an environment-dependent potential. In the latter case, the energy barriers surrounding the minima were also determined, and the structure candidates refined on ab initio level. Such a modular approach combining simulated annealing and the threshold algorithm on an empirical landscape with a subsequent ab initio refinement, was also employed for the exploration of the landscape of the mixed lanthanum halides[377], showing that in all these systems thermodynamically stable ternary compounds should exist. Furthermore, the quasi-binary (Li,Na)-nitrides[9] were investigated using simulated annealing, with the main result being that the compositions Li:Na = 2 : 1 and 1 : 2 appear to be the most promising candidates for ternary crystalline compounds in this system.

As more detailed examples, the global investigations of the landscapes of $CaCO_3$ and of

B_2O_3[364] are discussed. Using an evolutionary algorithm, the enthalpy landscape of $CaCO_3$ at high-pressures was investigated[373] and several promising structures were found. One of them allowed the solution of the structure of the already known first post-aragonite phase of $CaCO_3$ (at about 40 GPa)[378], while another one corresponded to the structure of a second high-pressure phase[379], which was found in parallel experimental work at about 130 GPa. This second structure contains chains of CO_4 tetrahedra (c.f. figure 5), which is reasonable, since the presence of CO_4-tetrahedra at such high pressures has also been predicted for e.g. the alkaline orthocarbonates[350]. In the $CaCO_3$ study, particular attention was paid to maintaining the diversity of the population of walkers, which is an important concern in the use of genetic/evolutionary algorithms. Interestingly, a surprisingly large number of the structure candidates found (compared to what is often produced via e.g. simulated annealing or just random search) appears to exhibit a layer-like structure; a possible reason might be the frequent cross-over move that merges two halves of two configurations by cutting their unit cells along some plane and thus favors the stacking of partially optimized structure segments such as layers.

The study of B_2O_3[364] combines elements from database searches and network construction to generate structure candidates for crystalline and amorphous phases. Starting point were the known crystalline compounds (B_2O_3-I and B_2O_3-II) and several multinary borate structures containing e.g. water H_2O or a modifier oxide such as Cs_2O. After removing all Cs- and H-atoms, respectively, together with the appropriate number of oxygen atoms, the new network structures were locally optimized using ab initio methods and potentials fitted to ab initio calculations. In addition, new networks were generated by replacing e.g. a B_3O_6-unit by a BO_3-unit. The optimized structures exhibited the same local environments that are found in known crystalline and amorphous boron oxide compounds. For some of the new hypothetical structures, the lattice energies were lower than the one computed for B_2O_3-I, suggesting that some of the predicted structures should be capable of existence at standard conditions.

5.2.3 Intermetallic solids

Depending on the amount of charge transfer and polarization effects between the different metals involved, a successful global search for intermetallic compounds capable of existence can either be performed using modified ionic or embedded atom potentials or has to take place on an ab initio

energy surface. An example of the latter type is the alloy Au_8Pd_4. Several studies have been performed on this system, starting with a systematic scan of the possible Au-Pd-arrangements on an underlying fcc-lattice where the energy was computed using a cluster expansion fitted to ab initio calculations[380], followed by two evolutionary optimization studies[357, 381], where the global search on the ab initio landscape succeeded in finding the underlying fcc lattice, and several energetically close alloy configurations corresponding to different Au-Pd arrangements on this lattice. In contrast, the embedded atom potentials such as the Gupta potential have so far predominantly been used to study intermetallic clusters[28].

Partly due to the many similarities among the known intermetallic compounds, and partly due to the fact that a great majority of these compounds can be described via random or ordered atom arrangements on an underlying lattice (or group of sub-lattices), database methods[48] and exhaustive sublattice-occupation[338] procedures have been employed for predicting intermetallic compounds and their structures. One such example is the high-throughput analysis of eighty binary intermetallic alloys[382] that can be built up from the 4d transition metals, plus several systems containing Al, Au, Mg, Pt, Sc, Na, Ti, and Tc. Employing database information about typical binary intermetallic structures, a large number of prototypes was compiled by either taking an actual existing structure-type or generating a new one as a fcc-, hcp- or bcc based superstructure. For each of the chemical systems, these prototype structures were locally optimized for eighteen different compositions. The resulting structure candidates were analyzed, ranked by energy, and, in a final step, the convex hull as function of composition was constructed. The results are mostly in agreement with experimental data, but in a number of cases, discrepancies appear, which might be due to experimental or calculational errors. Furthermore, a substantial number of candidates for new structures are proposed that might in the future be accessible as stable or metastable phases.

5.2.4 Covalent solids

In a good first approximation, and with plausible results, compounds developing polar covalent bonds can often be successfully described by applying ionic, empirical potentials. This is especially true, as long as the local coordination polyhedra of the anions that surround the individual cations are highly symmetric and the overall charge distribution in the solid is approximately

isotropic. In such a case, covalent and ionic models in many cases produce the same set of low-energy structure candidates, just with some different energy rankings in the higher-lying minima and possibly different energy barriers between the minima. However, this does not apply to systems where important local minima exhibit different kinds of anisotropic bonding arrangements, e.g. sp^2- and sp^3-hybridized carbon atoms, and one then needs to perform global optimizations on the ab initio energy landscape.

For example, in boron nitride several kinds of, mostly covalent, contributions to the total energy are present, and the global searches need to be performed on ab initio level. The BN-system is particularly interesting as a test system, because the experimentally observed modifications include both layered structures (hexagonal BN) and three-dimensional networks (wurtzite- and sphalerite-type). In global optimizations employing both Hartree-Fock and density functionals[248], 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. The strength of the general landscape approach has been impressively demonstrated by the discovery of two remarkable new framework structures with low-energies exhibiting the β-BeO structure, and the Al-partial structure in $SrAl_2$, respectively.

Another recent study has employed a combination of data mining, network generation and local optimization with ab initio energy calculations (DFT), in order to predict crystal structures of group 14 nitrides and phosphides[383]. An important step was the generation of new candidates by judiciously substituting atoms of different types in known basic networks, resulting in a multitude of many interesting structures. This procedure is somewhat similar to one of the approaches taken to find crystalline candidates in the $Si_3B_3N_7$-system[56].

5.2.5 Compounds exhibiting two or more types of bonding

When trying to deal with systems where groups of atoms form complex ions that are usually kept together via covalent bonds, two options are available: One can introduce rigid (or flexible) building units with a pre-defined charge distribution (based on experiment or ab initio calculations) together with individual atoms, or perform the whole search on an ab initio landscape, where these complex ions are expected to form as part of the optimization process.

As an example for the use of rigid building units serves the system KNO_2[19], where 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 rock salt 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" rock salt 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 [384, 385].

Using the prescribed path algorithm to study the activation barriers of the rotation of the NO_2-units around various axes of the unit, one finds that two barriers of about 10 K and about 100 K, respectively, appear to dominate the dynamics [43, 386]. 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 three-fold axis of the crystal (space group $R\bar{3}m$) at intermediate temperatures have been found, while the barrier analysis suggests that there might exist a second, not yet observed, intermediary phase where the NO_2-units rotate along a twofold axis of the structure.

The alternative approach of directly searching on the ab initio energy landscape was employed for the mixed covalent-ionic system CaC_2[387], with simulated annealing as the global optimization method. In all of the low-energy structures found, the carbon atoms had paired up in C_2^{2-}-units. At standard pressure, the C_2-units were located in the centers of approximate Ca-octahedra, which were edge-connected analogous to the arrangement of the Cl-octahedra around the Na atoms in the rock salt structure type. Many very similar structure candidates differing in the orientation of the C_2-units inside the distorted octahedra were observed. They could be assigned to one large structure family also containing nearly all known experimental modifications, which suggests that actually many more modifications might be accessible to experiment. In addition, a representative of a family of high-pressure modifications analogous to

the CsCl-type of structure was found (the C_2-units were located in the centers of distorted Ca_8 cubes), which should be capable of existence at high pressures.

5.2.6 Molecular crystals

General aspects An important class of crystalline compounds for which building units play a central role in predicting their structures are the molecular crystals[388, 389, 390, 391, 392, 393, 394]. In these crystals, the molecules can essentially always be treated as indestructible units; one often even fixes the conformation of the molecules during the global search.

There are a number of critical issues that workers in this area have to contend with: 1) The number of local minima that have very similar energies is overwhelmingly large, and thus it is not easy to select the "relevant" structure candidates from among them. 2) Related to this problem is the difficulty to properly compute the total energy of the structures. One typically uses simple empirical potentials, to describe the inter-molecular interactions during the global optimization stage. But since high-quality ab-initio calculations of crystals containing both inter- and intra-molecular interactions are still rather challenging, one usually employs refined empirical potentials also for the refinement optimizations. Such potentials are either fitted to experiment or to results from ab-initio calculations on individual molecules. 3) In principle, the molecules are flexible and not rigid, and treating them as inflexible during the global search can easily lead to important candidates being overlooked. 4) Due to the similarity in the ground state energies of many structure candidates, the effect of the thermodynamic conditions (pressure, temperature) at which the crystal is synthesized can lead to a change in the ranking according to the free energies. 5) Finally, it is quite likely that with the system being able to "choose" from among many structures with nearly the same energies, the kinetic aspects of the synthesis process will end up controlling the structure of the molecular crystal found in the laboratory; this is especially likely, since the likelihood of formation of critical nuclei of the various phases is not necessarily correlated with the thermodynamic stability of the corresponding macroscopic crystal.

Seen from the perspective of extended solids, it comes as a surprise that a small set of space groups (18) suffices to describe over 90 % of all molecular crystals found so far[395]. Furthermore, over 90 % of these crystals contain only one molecule in the asymmetric unit[396].

This statistical observation is often exploited during the theoretical structure determination and/or prediction of a molecular crystal by restricting the allowed structure candidates to those which exhibit one of these space groups and contain only one or two molecules in the asymmetric unit[397]. As a consequence, in most studies, the random walk based global minimizations tend to be combined with massive exhaustive searches, and the choice of simulated annealing, genetic algorithms and other stochastic walker based algorithms for the global search has not been very critical to the success of the prediction. In spite of the large body of work in this sub-field of structure prediction, only two examples are presented in some detail to illustrate the issues facing the investigator; for further cases, we refer to one of the many useful review articles that have appeared in recent years[390, 391, 19, 398, 52, 30, 53], and to the results of the relatively recent third blind test[393].

Examples The so-called conformation-family Monte Carlo algorithm was used to predict the crystal structure of nine small organic molecules such as benzene, pyrimidine, dimethoxymethane or formamidoxime, using several empirical potentials during the global search, and several refined potentials during the local optimizations[136]. Both the location, orientation, and conformation of the molecule, and the shape and size of the unit cell were modified during the global search. No symmetry constraints were applied. For the rigid molecules, their shape was taken from crystal structures where these molecules were present, if such structures were available. In the four studies involving rigid molecules, the energy of the lowest minimum was lower than the energy of the experimental structure, indicating that the global optimization technique was working quite well although the energy functions employed were not yet accurate enough. In nearly all cases, the experimental structure was found as a local minimum, but often not as the global minimum. When searching for crystal structures for the five flexible molecules, the success rate was not so high: in most instances, the experimental structure was not found during the global search, and the energy of the lowest minimum found was higher than the one belonging to the experimental structure in several cases.

It is interesting to note that in the majority of cases, the global search found local minima whose energies were lower than those of the experimentally observed crystal structures; sometimes the experimental modifications were not even among the top ten minima by energy. This

appears to be a general problem for molecular crystal structure prediction, in contrast to the prediction of extended crystalline solids, where the experimental structure is usually among the four or five lowest minima in energy, and often corresponds to the global minimum on ab initio level, at least for simple binary and ternary systems.

The general strategy pursued in the investigation of the energy landscape of phenobarbital[399], a medium-sized molecule with two flexible torsion angles, provides another good example of the currently popular approach to structure prediction of molecular crystals: In a first step, the molecule is analyzed on refined potential and/or ab initio (here: DFT) level, and possible conformations of the molecule that are expected to be the most relevant ones are determined. Then, a divide-and-conquer approach is used: Global optimizations using simulated annealing are performed for several of the space groups most common in molecular crystals, with very few molecules in the asymmetric unit, and the runs are repeated for each of the different molecular conformations the molecule can exhibit. During a given run, the conformation is kept fixed, but the position and orientation of the molecule in the cell, and the cell parameters are allowed to vary. The minima found are locally optimized using refined potentials, where both the positions and the conformations of the molecules are allowed to change.

Similar to the experience in the preceding study, the experimentally observed structure "III" ranked only 15th by energy among the local minima found, indicating that even the highly refined potentials fitted to ab initio data or combinations of ab initio and potential energies were not yet sufficiently accurate. Neither of the other two confirmed experimental structures "I" and "II" was found during the global optimization, and their energies proved to be considerably lower than the energies of the lowest minima found during the global search. It is encouraging, however, that the search produced many energetically competitive structures, which can serve as possible candidates for a fourth known modification of phenobarbital whose structure has not yet been determined. Nevertheless, even this rather extensive careful study (over 620,000 crystal structures were locally minimized) encountered serious problems when dealing with a flexible molecule with non-negligible hydrogen bonding.

5.2.7 Zeolites

An important class of extended solid compounds for which structures have been predicted are the zeolites and zeolite-analogues [62, 400, 401, 47, 46, 402, 52, 53, 403]. However, when using unrestricted global optimization techniques such as simulated annealing, zeolite framework prediction encounters serious problems. The reason is that one cannot expect the simulated system to produce the zeolite-framework "on its own" within a reasonable time when starting from individual atoms, because of the competition from the high-density phases. On the other hand, structure determination of zeolites with restricted energy landscapes has already been successfully performed[404]. Addressing this issue has led to the development of the AASBU-procedure[47], and, alternatively, the use of restrictions on the overall cell volume of the allowed configurations.

The "automated assembly of secondary building blocks" (AASBU) proceeds by taking structural elements like coordination polyhedra and joining them at their corners, edges and faces, in order to generate new structures. As starting configurations, a random arrangement of a fixed number of rigid SBUs is chosen. Periodic boundary conditions are employed, and commonly a space group is prescribed. Simulated annealing is used for the global optimization, where the SBUs are allowed to rotate. Furthermore, the cell parameters can be changed, and the distances among the SBUs adjusted.

In order to generate zeolite-like structures, ML_4-tetrahedra were picked as SBUs in the first study[47], in agreement with experimental data, where the interactions were chosen to favor corner-sharing networks. One or two SBU per asymmetric unit, and a selected set of space groups were employed. The authors found the expected structure types, e.g. the GME-, FAU-, RHO-, and LTL-frameworks (for the notation of zeolite classifications, c.f. the database of the international zeolite association[405]). For LTL-like frameworks, two new candidate structures were found, and an energy minimization with an empirical potential showed that their lattice energies were only slightly higher than that of LTL itself. These results and subsequent successes[52] are quite encouraging. An unknown quantity is the effect of fixing the space group during the optimization - it is not clear, whether the above frameworks could be generated in $P1$, too.

A related way to generate zeolite-like frameworks is their construction from bubble clusters[70,

406, 407]. A starting point is the investigation of the global landscape of clusters of various sizes for a given chemical system. The most stable compact and bubble clusters serve as building blocks analogous to the SBUs to construct either densely packed structures or porous frameworks, respectively. For example, Woodley et al.[406] employed an evolutionary algorithm to identify promising ZnO clusters such as a sodalite-cage analogue which is the basic SBU e.g. for the LTA or the FAU zeolite, and subsequently constructed a variety of such networks according to the usual connection rules for zeolite (see figure 6). Besides generating many candidates for single-SBU based frameworks and testing their viability, the authors investigated possible frameworks consisting of two different SBUs, and frameworks where the cages are not merged but connected by linkers such as a ZnO-atom pair, leading to structures analogous to those known from mesoporous compounds[408].

Another approach to zeolite prediction consists in varying the positions of individual atoms and the cell parameters while ensuring that the overall density of the configuration remains within a given interval during the simulated annealing[43]. After local optimization (quench) on empirical potential level, the resulting structures (for a generic SiO_2-zeolite) were quite reasonable, but still exhibited many defects such as under-coordination or dangling bonds, underlining the large degree of freedom the system has in forming low-energy porous structures.

Finally, a second alternative method consists in introducing stationary or mobile exclusion zones together with individual atoms or larger building units. Such work has been mostly performed with genetic algorithms[400, 409, 410, 411], although simulated annealing has also been employed, e.g. for the study of possible zeolite-type structures in the SiO_2-analogue BeF_2 using spherical mobile zones of varying diameter[412]. When using such mobile exclusion zones, they tend to cluster during the global exploration stage resulting in structures that contain isolated columnar pores or consist of alternating sheets of essentially bulk material and layers of exclusion zones, because for fixed large cell volume such isolated slabs of bulk material are energetically competitive with three-dimensional porous structures, especially for very large numbers of atoms / simulation cell.

In a detailed study[411], the generation of zeolitic SiO_2-networks via hybrid genetic and evolutionary algorithms with and without the use of exclusion zones was investigated, for fixed overall density and various fixed shapes of the simulation cells. An ionic empirical energy func-

tion was employed, where the exclusion zones were represented by hard boundaries in order to accelerate the cost function evaluation. For the cases studied, the hybrid evolutionary algorithm was to be preferred to the hybrid genetic algorithm with about twice the success rate of finding one of the expected zeolites; in particular, it was able to generate small-pore frameworks such as JBW or BIK even without needing exclusion zones. However, for lower-density frameworks with larger pores, exclusion zones were always necessary for finding the zeolite frameworks such as SOD and CHA on the cost function landscape.

5.2.8 Selected examples of structure determination

Clearly, the more restrictions are enforced during the global search, the less one should call the investigation an unbiased structure prediction. Instead, one should divide the examples one finds in the literature under the heading of "structure prediction" into three different classes: On the one extreme is the unrestricted structure prediction, where only the stoichiometry but neither the unit cell nor the number of formula units is known, while on the other 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. Between these two extremes lies the case of restricted structure prediction, e.g. the prediction of structures in systems where certain structural elements or local environments of atoms are pre-defined or assumed at the outset based on chemical pre-knowledge, such as primary and secondary building units [46], or where the atoms are assumed to reside on prescribed sublattices usually known from experiment[413, 320, 414, 337, 415, 416, 417, 418, 419, 420].

Such structure determination from limited experimental information is rapidly becoming a highly valuable tool in the arsenal of applied crystallography and solid state chemistry. A number of methods have been developed that use global optimization techniques to generate structure candidates for newly synthesized compounds, which can serve as starting points of standard structure refinement techniques. The most common approaches are probably the so-called Reverse-Monte-Carlo method where the cost function equals the difference between measured and computed diffractogram[421, 422, 423, 424], and the use of experimental cell information to restrict the configuration space that the walker can explore while still employing a simple potential or cost function for computing the energy[269, 366, 425, 426, 118]. As global

optimization methods, both simulated annealing and genetic algorithms have been employed for systems such as NbF_4[269] and Li_3RuO_4[174], respectively.

Incorporating the experimental information directly into the energy landscape via penalty-type terms leads to 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 the ones calculated for the current atomic configuration, to the potential energy[427]. Here, the structure is optimized both with respect to the energy and the diffractogram,

$$E = \lambda E_{pot} + (1 - \lambda) R_B, (0 \leq \lambda \leq 1). \tag{11}$$

Besides prescribing the cell parameters, one can include various additional constraints, e.g. keep the positions of some of the atoms fixed or enforce certain symmetry requirements. In analogy to the restricted structure prediction, it is also possible to employ rigid or flexible building units, in particular if the existence of complex ions and molecules in the compound of interest has been established.

This approach has been tested successfully for a large number of ionic, quasi-ionic and metallic systems [427, 428, 429], where simulated annealing [427, 428] and genetic algorithms [429] 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 using simulated annealing, the program ENDEAVOUR [430], has already been very successful in "real-life" applications, generating convincing structure candidates for such different systems as K_2CN_2 [431], sulphur [432], Na_3PSO_3 [433], Ag_2NiO_2 [434], Ag_2PdO_2 [435], $GaAsO_4$ [436], ammonium metatungstate [437], the zeolite-like structure $Na_{1-x}Ge_{3+z}$ [438], Tl_2CS_3 [439], and BiB_3O_6 [440].

6 Feasibility and experimental verification

While there exist by now a large number of predicted structures in many types of chemical system, the instances of a successful synthesis after an unbiased structure prediction had taken

place are quite rare. One of the reasons is that for many of the elementary, binary and ternary compounds, i.e. precisely the kind of systems where the theorist stands a good chance of systematically exploring the energy landscape without guidance from experiment, the experimental chemist has usually already attempted to synthesize potential compounds, either successfully (and thus no prediction of the thermodynamically stable phase is possible any more) or unsuccessfully (which usually implies that the experiment is probably very challenging and requires a serious effort on the side of the experimentalist). Thus predictions can often only hope to be validated by either the synthesis of a new metastable modification or by the predicted structure being in a range of thermodynamic conditions which are unusual and/or difficult to reach with standard synthesis methods. The latter case especially refers to the high-pressure region, but also thermodynamically stable phases at very low temperatures that are difficult to reach by e.g. quenching from the melt in quasi-binary or -ternary systems fall into this category.

Nevertheless, there have been a number of noteworthy successes in recent years described below. This is partly due to the development of new synthesis techniques such as the low-temperature atom beam deposition method (LT-ABD)[441]. In this technique, atoms are deposited atomically dispersed on a very cold substrate (at liquid nitrogen or liquid helium temperatures) forming an amorphous deposit that is subsequently slowly heated, resulting in the crystallization of (in particular) low-density modification at very low temperatures. The other recent development favoring successful predictions in the future is the more frequent use of high-pressure methods in synthetic chemistry[442, 443, 444].

6.1 Prediction and synthesis of sodium nitride Na_3N

Among the most impressive results of a successful synthesis of a predicted compound is the synthesis of the elusive sodium nitride Na_3N in an energetically high-lying structure, the ReO_3-type using the LT-ABD-method[11], several years after a large number of metastable phases had been predicted to exist in this deceptively simple chemical system. One should note that 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 [445]. In several studies [8, 9, 10], the enthalpy landscapes of all alkali nitrides M_3N (M = Li, Na, K, Rb, Cs) had been explored with simulated annealing and the

threshold algorithm for a wide range of pressures. This resulted in a large number of structure candidates, including e.g. the Li_3N-, Li_3P-, Li_3Bi-, $AuCu_3$-, Al_3Ti-, ReO_3-, and UO_3-structure types, plus many previously unknown structures. Figure 7 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 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.

Even more impressively, almost the full set of the most stable predicted polymorphs of Na_3N (in addition to the ReO_3-type also the Li_3N-, the Li_3P-, and the Li_3Bi-type) have recently been realized in the correct sequence of appearance as function of pressure, using high-pressure experiments[12, 13]. One should note that in contrast to the usual situation prevailing when predictions of high-pressure phases are attempted, not even the standard pressure modification of Na_3N existed at the time when the predictions were made.

6.2 Prediction and synthesis of metastable lithium halides LiX (X = I, Br, Cl)

Another beautiful success in synthesizing a new modification after the prediction had occurred, was the synthesis of metastable modifications in the lithium halides. As mentioned earlier, for each of the twenty alkali metal halides numerous polymorphs had been predicted [255, 14]. In the case of the lithium halides, the structure that is most competitive energetically with the experimentally observed modification (rock salt) exhibits the low-density wurtzite type structure, and thus appears to be the most promising candidate for a metastable modification in the lithium halides.

Using the LT-ABD-technique, this predicted modification has subsequently been realized for LiI[446, 447], LiBr[15] and LiCl[16]. Figure 8 shows the typical structural evolution from the amorphous phase 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. Furthermore, this result underlines the fact that in many if not all simple supposedly well-known compounds additional, as yet unknown, metastable modifications exist that are accessible to experimental realization.

6.3 Prediction and validation of high-pressure alkali metal sulfides

As a final example, we consider the alkali metal sulfides, a case of successful predictions of new high-pressure modifications for a system where the standard pressure modification had been known and studied for a long time before the predictions and the high-pressure experiments were performed. At the beginning of the past decade, parallel work took place in predicting the possible high-pressure phases of Li_2S[10], Na_2S[10], K_2S[244], Rb_2S[244] and Cs_2S[244], and performing the corresponding high-pressure experiments on Li_2S[448], Na_2S[449], K_2S[450].

The high-pressure modifications found in the experiment agreed with those determined from the global optimization runs, serving as a satisfying validation of the suitability of global exploration techniques for structure predictions of high-pressure phases. However, only recently has the experimental work continued[451], confirming the original prediction[244] of the existence of a Ni_2In-type modification in the Rb_2S-system at high pressures.

7 Outlook

7.1 Future developments of exploration algorithms

"Beware the claims of the producer!"[452] is a word of caution one should heed quite generally when reviewing the performance of global optimization techniques, such as those employed for the purpose of structure prediction. Between the moveclass of the random walker, the temperature schedule, the size of the ensemble of walkers, the penalty terms included, and the degree to which chemical information about the system and already available landscape information is incorporated, there exists a large amount of freedom to tune and optimize any particular algorithm under consideration, and thus the comparability of the algorithms is often highly problematic.

Considering the general task of finding most of the relevant deep-lying local minima of a crystalline solid, or quite generally a chemical system, using one of the methods presented, what is the current state of the field? Of course, the size of the system that can be handled depends on the computer power available and the amount of already available information about the system and its landscape. Assuming that the researcher can perform several hundred or perhaps several thousand single walker optimization runs within a reasonable amount of time,

for binary and ternary systems, then e.g. standard "plain vanilla" simulated annealing can deal with about 20 - 30 atoms per simulation cell when using empirical potentials, and about 8 - 12 atoms per cell using ab initio energy calculations. More refined, but often computationally more expensive and/or algorithmically more complex, variants such as thermal cycling, basin hopping, multi-walker annealing or genetic/evolutionary algorithms are on average more successful than standard simulated annealing for up to 40 - 50 atoms per cell (for empirical potentials), but beyond this number, none of the methods appears to be truly reliable. This applies even more strongly if one attempts to deal with large quaternary or even more diverse systems, whose landscapes are dominated by minima corresponding to essentially amorphous configurations: unless the energy function possesses structure directing features, e.g. the energy hypersurface contains large funnels guiding the search towards the most important local minima or large essentially invariant building units control the possible structures, one has to ask oneself "Am I feeling lucky today?" when exploring such systems.

However, the task of optimizing the optimization algorithm will surely remain high on the priority list of future goals in structure prediction. A corollary of this is that newer and bigger computers are not the magic bullet, since the difficulty of solving typical structure prediction problems grows exponentially with system size. But experience shows that there often appear to be approximately hierarchical elements in those periodic structures that contain many atoms per primitive unit cell, and this must surely be reflected in the properties of the corresponding energy landscape. Once it is possible to use e.g. short preliminary exploration runs to identify such general features of the landscape of crystalline solids that can influence the progress of the walker[7, 453, 454, 103, 121], one should be able to greatly improve the performance of various search algorithms by e.g. employing adaptive moveclasses or reducing the size of the landscape by introducing fixed building units or restricting the search to sublattices.

Hierarchical or divide-and-conquer approaches that work by restricting the allowed configuration space during a part of the optimization stage often appear to be able to deal with large simulation cells or molecules. However, one always runs a considerable risk of overlooking important structure candidates, for example, when basing one's decision to employ certain building units only on e.g. database information about related chemical systems. This is especially true when one predominantly relies on chemical intuition instead of mathematical information about

the shape of the energy landscape. Furthermore, finding the local minima is not everything that one cares about when studying an energy landscape, even if one only wants to determine structure candidates. Identifying complex locally ergodic regions that are important at elevated temperatures, and estimating the kinetic stability of the candidates requires landscape information beyond the local minima, such as (generalized) barriers and local densities of states. While already several algorithms are available for this purpose, most are still rather clumsy and inefficient. Optimizing these exploration tools will clearly be a major enterprise in the future.

Closely connected to this issue is the fundamental question of the determination of 'phase diagrams' for systems that are not close to the thermodynamic limit, e.g. finite size systems such as clusters[455, 456] or systems that have not yet reached global equilibrium, as one daily encounters in solid state chemistry. One of the major issues in this context is to what extent the traditional concept of a phase diagram that presupposes the existence of the thermodynamic limit and global equilibrium can be extended in a systematic fashion to deal with such types of systems. As discussed in this work, in principle, such questions can be resolved within the context of locally ergodic regions and the proper observational time scales on which the movement of the system on the energy landscape takes place. This leads naturally to the concept of an extended phase diagram with the observational timescale as an additional coordinate, that encompasses the standard equilibrium phase diagram in the limit of infinite observation time.

Alternatively, one might want to incorporate the history of the phases in the experimentally determined phase diagram. Here, a quite interesting approach is the combination of thermodynamics and kinetic modeling visualized in the well-known time-temperature-transformation (TTT) diagrams[457, 458, 459, 460, 461], where the development of a phase is depicted as function of time and temperature. Clearly, the fact that metastable or marginally stable compounds such as glasses are of great importance in applications makes it imperative, to address this question in a systematic fashion.

7.2 Rational development of synthesis routes

The successful validation of the existence of the predicted compounds requires also great efforts on the part of the experimental solid state chemist, for the second step of planning chemical syntheses consists of rationally developing a viable path to the desired configuration, 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 hypersurface of the 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[5].

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.

One attempt aimed at improving this situation is the development of new synthesis routes such as the low-temperature atom beam deposition method mentioned above. Particularly noteworthy from the synthetic point of view is the fact that the structural evolution of the random solid reaction mixture, which is very much reminiscent of the starting configurations for the global computational exploration of the respective energy landscape, undergoes unprecedented all-solid state reactions at temperatures far below room-temperature, yielding well-crystallized products [441] of metastable phases. Since a shrinkage of volume occurs when crystalline nuclei form inside the amorphous deposit, effective negative pressures on the surfaces of these nuclei are generated[367], and thus the first structures that evolve are the metastable, low-density ones.

Working in tandem with experiment, at least conceptually, the time also appears to have come for theory, to address the third pillar of the rational planning of solid state synthesis: the modeling and optimization of chemical synthesis routes. 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[462]. 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, due 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[463, 464]. First examples of such a stepping-stone description are the modeling of the sol-gel synthesis of the amorphous ceramic a-$Si_3B_3N_7$ [76, 77], and the deposition and growth of Xe on a cold sapphire substrate[79].

The wide-spread 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 [465, 462]. 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.

However, perhaps the most fundamental change over the past two decades has been the development of a new paradigm for materials science and solid state chemistry[5, 6, 35]: the switch from the traditional inductive approach based on explorative synthesis followed by phenomenological modeling and a posteriori interpretation and interpolation of the experimental data to a deductive one of predicting new compounds, determining the phase diagrams where they occur, and designing routes for their synthesis from first principles. This rational synthesis planning[466, 7, 5, 35] is finally coming into its own as the success of the combination of initial prediction of a new compound and subsequent synthesis, via newly developed synthesis methods[11, 441], clearly demonstrates[15, 467]. This transformation of solid state chemistry and materials science from an inductive to a deductive science is a monumental step, and while

currently the focus is still on the development of more efficient tools to study energy surfaces and to compute phase diagrams without experimental input, perhaps an even greater challenge is faced by the experimentalist: the need for new and more refined synthesis techniques that will provide physical access to the plethora of new compounds and modifications waiting on the energy landscapes of chemical systems.

Acknowledgments

The authors would like to thank I. Pentin for assistance with preparing the figures, and S. Woodley for providing figure 6.

References

[1] L. Principe, editor. *Chymists and Chymistry: Studies in the History of Alchemy and Early Modern Chemistry*. Science History Publications, Sagamore Beach, 2007.

[2] E. J. Corey. *Pure Appl. Chem.*, 14:19–37, 1967.

[3] E. J. Corey. *Angew. Chem. Int. Ed. Eng.*, 30:455–465, 1991.

[4] V. M. Goldschmidt. *Skrift. Nor. Vidensk. Akad. Kl. 1: Mat. Naturvidensk.*, page 1, 1926.

[5] M. Jansen. *Angew. Chem. Int. Ed.*, 41:3747–3766, 2002.

[6] M. Jansen and J. C. Schön. *Angew. Chem. Int. Ed.*, 45:3406–3412, 2006.

[7] J. C. Schön and M. Jansen. *Angew. Chem. Int. Ed. Eng.*, 35:1286–1304, 1996.

[8] M. Jansen and J. C. Schön. *Z. Anorg. Allg. Chem.*, 624:533–540, 1998.

[9] J. C. Schön, M. A. C. Wevers, and M. Jansen. *Solid State Sci.*, 2:449–456, 2000.

[10] J. C. Schön, M. A. C. Wevers, and M. Jansen. *J. Mater. Chem.*, 11:69–77, 2001.

[11] D. Fischer and M. Jansen. *Angew. Chem. Int. Ed.*, 41:1755–1756, 2002.

[12] G. V. Vajenine, X. Wang, I. Efthimiopoulus, S. Karmakar, K. Syassen, and M. Hanfland. *Z. Anorg. Allg. Chem.*, 634:2015, 2008.

[13] G. V. Vajenine, X. Wang, I. Efthimiopoulus, S. Karmakar, K. Syassen, and M. Hanfland. *Phys. Rev. B*, 79:224107, 2009.

[14] Ž. Čančarević, J. C. Schön, and M. Jansen. *Chem. Asian. J.*, 3:561–572, 2008.

[15] Y. Liebold-Ribeiro, D. Fischer, and M. Jansen. *Angew. Chem. Int. Ed.*, 47:4428–4431, 2008.

[16] A. Bach, D. Fischer, and M. Jansen. *Z. Anorg. Allg. Chem.*, 635:2406–2409, 2009.

[17] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. *Prot. Science*, 4:561–602, 1995.

[18] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, and T. R. Walsh. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics, volume 115*, pages 1–111. Wiley, New York, 2000.

[19] J. C. Schön and M. Jansen. *Z. Krist.*, 216:307–325;361–383, 2001.

[20] F. Rao and A. Caflisch. *J. Mol. Biol.*, 342:299–306, 2004.

[21] D. J. Wales. *Energy Landscapes with Applications to Clusters, Biomolecules and Glasses.* Cambridge Univ. Press, 2004.

[22] S. A. Adcock and J. A. McCammon. *Chem. Rev.*, 106:1589–1615, 2006.

[23] M. C. Prentiss, C. Zong, C. Hardin, M. P. Eastwood, and P. G. Wolynes. *J. Chem. Theo. Comp.*, 2:705–716, 2006.

[24] N. G.. Phillips, C. W. S. Conover, and L. A. Bloomfield. *J. Chem. Phys.*, 94:4980–4987, 1991.

[25] S. K. Lai, P. J. Hsu, K. L. Wu, W. K. Liu, and M. Iwamatsu. *J. Chem. Phys.*, 117:10715–10725, 2002.

[26] R. L. Johnston and C. Roberts. In H. M. Cartwright and L. M. Sztendera, editors, *Soft computing approaches in chemistry*, pages 161–204. Springer-Verlag, Berlin, 2003.

[27] B. Hartke. *Struct. Bonding*, 110:33–53, 2004.

[28] R. Ferrando, J. Jellinek, and R. L. Johnston. *Chem. Rev.*, 108:845–910, 2008.

[29] A. A. Sokol, C. R. A. Catlow, M. Miskufova, S. A. Shevlin, A. A. Al-Sunaidi, A. Walsh, and S. M. Woodley. *Phys. Chem. Chem. Phys.*, 12:8423–8435, 2010.

[30] S. L. Price. *Phys. Chem. Chem. Phys.*, 10:1996–2009, 2008.

[31] J. Callaway. *Quantum Theory of the Solid State.* Academic Press, New York, 1974.

[32] M. Goldstein. *J. Chem. Phys.*, 51:3728–3739, 1969.

[33] F.H. Stillinger and T. A. Weber. *Phys. Rev. A*, 25:978–989, 1982.

[34] R. S. Berry. *Chem. Rev.*, 93:2379–2394, 1993.

[35] M. Jansen. In K. M. Harris and P. Edwards, editors, *Turning points in Solid-State, Materials and Surface Science*, page 22. RSC Publishing, Cambridge, UK, 2008.

[36] J. C. Schön and M. Jansen. *Int. J. Mat. Res.*, 100:135–152, 2009.

[37] J. C. Schön, K. Doll, and M. Jansen. *Phys Stat. Sol. (b)*, 247:23–39, 2010.

[38] M. Jansen, K. Doll, and J. C. Schön. *Acta. Cryst. A*, 66:518–534, 2010.

[39] L. D. Landau and E. M. Lifshitz. *Statistical Physics, Part 1, 3rd Ed.* Pergamon, New York, 1980.

[40] J. C. Schön, M. A. C. Wevers, and M. Jansen. *J. Phys.: Cond. Matter*, 15:5479–5486, 2003.

[41] P. Sibani, J. C. Schön, P. Salamon, and J.-O. Andersson. *Europhys. Lett.*, 22:479–485, 1993.

[42] A. Hannemann, J. C. Schön, M. Jansen, and P. Sibani. *J. Phys. Chem. B*, 109:11770–11776, 2005.

[43] J. C. Schön and M. Jansen. In J Li, N. E. Brese, M. G. Kanatzidis, and M. Jansen, editors, *Mat. Res. Soc. Symp. Proc. Vol 848: Solid State Chemistry of Inorganic Materials V*, page 333. MRS, Warrendale, 2005.

[44] J. C. Schön and M. Jansen. *Acta Cryst A (Suppl.)*, 55, 1999.

[45] J. C. Schön and M. Jansen. *Z. Krist.*, 216:307–325, 2001.

[46] C. Mellot-Draznieks, S. Girard, G. Ferey, J. C. Schön, Ž. Čančarević, and M. Jansen. *Chem. Eur. J.*, 8:4102–4113, 2002.

[47] C. Mellot-Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman, and G. Ferey. *Angew. Chem. Int. Ed. Eng.*, 39:2270–2275, 2000.

[48] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder. *Phys. Rev. Lett.*, 91:135503, 2003.

[49] O. Delgado-Friedrichs, A. W. M. Dress, D. H. Huson, J. Klinowski, and A. L. Mackay. *Nature*, 400:644–647, 1999.

[50] B. P. van Eijck, W. T. M. Mooij, and J. Kroon. *Acta Cryst. B*, 51:99–103, 1995.

[51] W. E. Klee, M. Bader, and G. Thimm. *Z. Krist.*, 212:553–558, 1997.

[52] C. Mellot-Draznieks. *J. Mater. Chem.*, 17:4348–4358, 2007.

[53] S. M. Woodley and C. R. A. Catlow. *Nature Mater.*, 7:937, 2008.

[54] A Y. Liu and M. L. Cohen. *Phys. Rev. B*, 41:10727–10734, 1990.

[55] D. M. Giaquinta and H. C. zur Loye. *Chem. Mater.*, 6:365–372, 1994.

[56] P. Kroll and R. Hoffmann. *Angew. Chem. Int. Ed. Eng.*, 37:2527–2530, 1998.

[57] L. Stixrude and D. R. Peacor. *Nature*, 420:165–168, 2002.

[58] M.-W. Lumey and R. Dronskowski. *Z. Anorg. Allg. Chem.*, 631:887–893, 2005.

[59] U. Müller. *Acta Cryst. B*, 42:557–564, 1986.

[60] U. Müller. *Acta Cryst. B*, 48:172–178, 1992.

[61] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder. *Nat. Mater.*, 5:641, 2006.

[62] M. M. J. Treacy, K. H. Randall, S. Rao, J. A. Perry, and D. J. Chadi. *Z. Krist.*, 212:768–791, 1997.

[63] J. Klinowski. *Curr. Opin. Solid State Mater. Sci.*, 3:79–85, 1998.

[64] M. D. Foster, A. Simperler, R. G. Bell, O. Delgado-Friedrichs, F. A. Paz, and J. Klinowski. *Nature Mater.*, 3:234–238, 2004.

[65] A. Le Bail and F. Calvayrac. *J. Solid State Chem.*, 179:3159–3166, 2006.

[66] O. Delgado-Friedrichs, M. O'Keefe, and O. M. Yaghi. *Phys. Chem. Chem. Phys.*, 9:1035–1043, 2007.

[67] V. A. Blatov and D. M. Proserpio. In A. R. Oganov, editor, *Modern methods of crystal structure prediction*, chapter 1, pages 1–28. Wiley VCh, Weinheim, 2011.

[68] S. N. Khanna and P. Jena. *Phys. Rev. Lett.*, 69:1664–1667, 1992.

[69] J. C. Wojdel, M. A. Zwijnenburg, and S. T. Bromley. *Chem. Mater.*, 18:1464–1469, 2006.

[70] J. Carrasco, F. Illas, and S. T. Bromley. *Phys. Rev. Lett.*, 9:235502, 2007.

[71] W. Sangthong, J. Limtrakul, F. Illas, and S. T. Bromley. *J. Mater. Chem.*, 18:5871–5879, 2008.

[72] C. G. Duan, W. N. Mei, R. W. Smith, J. Liu, M. M. Ossowski, and J. R. Hardy. *Phys. Rev. B*, 63:144105, 2001.

[73] R. Martonak, A. Laio, and M. Parrinello. *Phys. Rev. Lett.*, 90:075503, 2003.

[74] D. Zahn and S. Leoni. *Phys. Rev. Lett.*, 92:250201, 2004.

[75] C. Asker, A. B. Belonoshko, A. S. Mikhaylushkin, and I. A. Abrikosov. *Phys. Rev. B*, 77:220102, 2008.

[76] J. C. Schön, A. Hannemann, and M. Jansen. *J. Phys. Chem. B*, 108:2210–2217, 2004.

[77] A. Hannemann, J. C. Schön, and M. Jansen. *J. Mater. Chem.*, 15:1167–1178, 2005.

[78] M. Jansen, J. C. Schön, and L. van Wüllen. *Angew. Chem. Int. Ed.*, 45(26):4244–4263, 2006.

[79] N. Toto, J. C. Schön, and M. Jansen. *Phys. Rev. B*, 82:115401, 2010.

[80] D. Zahn. *Phys. Rev. Lett.*, 92:040801, 2004.

[81] I. M. Svishchev and P. G. Kusalik. *Phys. Rev. Lett.*, 75:3289–3292, 1995.

[82] K.H. Hoffmann and J. C. Schön. *Found. Phys. Lett.*, 18:171–182, 2005.

[83] J. C. Schön, M.A.C. Wevers, and M. Jansen. *J. Phys. A: Math. Gen.*, 34:4041–4052, 2001.

[84] K. H. Hoffmann and P. Sibani. *Phys. Rev. A*, 38:4261–4270, 1988.

[85] J. C. Schön, H. Putz, and M. Jansen. *J. Phys.: Cond. Matt.*, 8:143–156, 1996.

[86] O. M. Becker and M. Karplus. *J. Chem. Phys.*, 106:1495–1517, 1997.

[87] A. Heuer. *Phys. Rev. Lett.*, 78:4051–4054, 1997.

[88] D. J. Wales, M. A. Miller, and T. R. Walsh. *Nature*, 394:758–760, 1998.

[89] S. V. Krivov and M. Karplus. *J. Chem. Phys.*, 117:10894–10903, 2002.

[90] T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G. J. Rylance, R. L. Johnston, and D. J. Wales. *J. Chem. Phys.*, 122:084714, 2005.

[91] A. E. Garcia, R. Blumenfeld, G. Hummer, and J. A. Krumhansl. *Physica D*, 107:225–239, 1997.

[92] T. Klotz, S. Schubert, and K. H. Hoffmann. *J. Phys.: Cond. Matt.*, 10:6127–6134, 1998.

[93] P. Garstecki, T. X. Hoang, and M. Cieplak. *Phys. Rev. E*, 60:3219, 1999.

[94] J. P. K. Doye and C. P. Massen. *J. Chem. Phys.*, 122:084105, 2005.

[95] G. Cox, R. S. Berry, and R. L. Johnston. *J. Phys. Chem. A*, 110:11543–11550, 2006.

[96] F. Noe, D. Krachtus, J. C. Smith, and S. Fischer. *J. Chem. Theo. Comp.*, 2:840–857, 2006.

[97] F. Noe, I. Horenko, C. Schütte, and J. C. Smith. *J. Chem. Phys.*, 126:155102, 2007.

[98] D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao. *Proc. Nat. Acad. Sci.*, 104:1817–1822, 2007.

[99] D. Gfeller, P. De Los Rios, D. M. de Lachapelle, G. Caldarelli, and F. Rao. *Phys. Rev. E*, 76:026113, 2007.

[100] F. Noe and S. Fischer. *Curr. Opin. Struct. Biol.*, 18:154–162, 2008.

[101] K. D. Ball and R. S. Berry. *J. Chem. Phys.*, 111:2060–2070, 1999.

[102] M. A. Miller and D. J. Wales. *J. Chem. Phys.*, 111:6610–6616, 1999.

[103] M. A. C. Wevers, J. C. Schön, and M. Jansen. *J. Phys.: Cond. Matt.*, 11:6487–6499, 1999.

[104] J. C. Gower. *Biometrika*, 53:325–338, 1966.

[105] R. Abagyan and P. Argos. *J. Mol. Biol.*, 225:519–532, 1992.

[106] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. *Proteins: Struct. Funct. Gen.*, 17:412–425, 1993.

[107] J. M. Troyer and F. E. Cohen. *Proteins: Struct. Funct. Gen.*, 23:97–110, 1995.

[108] O. M. Becker. *Proteins: Struct. Funct. Gen.*, 27:213–226, 1997.

[109] P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi. *Proc. Nat. Acad. Sci.*, 103:9885–9890, 2006.

[110] D. Zagorac, J. C. Schön, K. Doll, and M Jansen. *Acta Physica Polonica A*, in press, 2011.

[111] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. *Science*, 220:671–680, 1983.

[112] A. Möbius, A. Neklioudov, A. Diaz-Sanchez, K. H. Hoffmann, A. Fachat, and M. Schreiber. *Phys. Rev. Lett.*, 79:4297–4301, 1997.

[113] D. J. Wales and J. P. K. Doye. *J. Phys. Chem.*, 101:5111–5116, 1997.

[114] J. C. Schön, Ž. P. Čančarević, A. Hannemann, and M. Jansen. *J. Chem. Phys.*, 128:194712, 2008.

[115] B. A. Berg and T. Neuhaus. *Phys. Rev. Lett.*, 68:9–12, 1992.

[116] D. M. Deaven and K. M. Ho. *Phys. Rev. Lett.*, 75:288–291, 1995.

[117] J. D. Gale. *J. Chem. Soc. Farad. Trans.*, 93:629–637, 1997.

[118] S. M. Woodley, P. D. Battle, J. D. Gale, and C. R. A. Catlow. *Phys. Chem. Chem. Phys.*, 1:2535–2542, 1999.

[119] J. H. Holland. *Adaptation in Natural and Artificial Systems.* Univ. Mich. Press, Ann Arbor, 1975.

[120] V. Czerny. *J. Optim. Theo. Appl.*, 45:41–51, 1985.

[121] P. Salamon, P. Sibani, and R. Frost. *Facts, Conjectures, and Improvements for Simulated Annealing.* SIAM Monographs, Philadelphia, 2002.

[122] J. C. Schön and M. Jansen. In A. R. Oganov, editor, *Modern methods of crystal structure prediction*, chapter 4, pages 67–106. Wiley VCh, Weinheim, 2011.

[123] S. Geman and D. Geman. *IEEE T. Pattern Anal*, 6(6):721–741, 1984.

[124] M. Iwamatsu and Y. Okabe. *Chem. Phys. Lett.*, 399:396–400, 2004.

[125] D. J. Wales. In A. R. Oganov, editor, *Modern methods of crystal structure prediction*, chapter 2, pages 29–54. Wiley VCh, Weinheim, 2011.

[126] D. Delamarre and B. Virot. *RAIRO - Rech. Oper. Oper. Res.*, 32:43–73, 1998.

[127] A. Möbius, K.H. Hoffmann, and J. C. Schön. In C. Beck, G. Benedek, A. Rapisarda, and C. Tsallis, editors, *Complexity, Metastability and Nonextensivity*, pages 215–219. International School of Solid State Physics, World Scientific Singapore, 2004.

[128] G. Ruppeiner, J. M. Pedersen, and P. Salamon. *J. Phys. I*, 1:455–470, 1991.

[129] P. Salamon, J. Nulton, J. Robinson, J. M. Pedersen, G. Ruppeiner, and L. Liao. *Comp. Phys. Comm.*, 49:423–428, 1988.

[130] J. A. Chandy, S. Kim, B. Ramkumar, S. Parkes, and P. Banerjee. *IEEE Trans. Comp. Aided Des. ICS*, 16:398–410, 1997.

[131] T. Zimmermann and P. Salamon. *Int. J. Comp. Math.*, 42:21–32, 1992.

[132] J. Ma, D. Hsu, and J. E. Straub. *J. Chem. Phys.*, 99:4024–4035, 1993.

[133] A. Roitberg and R. Elber. *J. Chem. Phys.*, 95:9277–9287, 1991.

[134] J. E. Straub and M. Karplus. *J. Chem. Phys.*, 94:6737–6739, 1990.

[135] J. G. Kim, Y. Fukunishi, A. Kidera, and H. Nakamura. *Chem. Phys. Lett.*, 392:34–39, 2004.

[136] J. Pillardy, Y. A. Arnautova, C. Czaplewski, K. D. Gibson, and H. A. Scheraga. *Proc. Nat. Acad. Sci.*, 98:12351–12356, 2001.

[137] Y. H. Shi and R. Eberhart. *CEC 1998: Proceedings IEEE Congr. on evolutionary computation*, page 69, 1998.

[138] A. Venkatnathan and G. A. Voth. *J. Chem. Theo. Comp.*, 1:36–40, 2005.

[139] T. Huber and W. F. van Gunsteren. *J. Phys. Chem.*, 102:5937–5943, 1998.

[140] B. A. Berg, H. Nogushi, and Y. Okamoto. *Phys. Rev. E*, 68:036126, 2003.

[141] S. G. Itoh and Y. Okamoto. *Chem. Phys. Lett.*, 400:308–313, 2004.

[142] J. Kim and T. Keyes. *J. Chem. Phys.*, 121:4237–4245, 2004.

[143] D. D. Frantz, D. L. Freeman, and J. D. Doll. *J. Chem. Phys.*, 93:2769–2784, 1990.

[144] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. *J. Chem. Phys.*, 21:1087–1092, 1953.

[145] C. Tsallis. *J. Stat. Phys.*, 52:479–487, 1988.

[146] C. Tsallis and S. A. Stariolo. *Physica A*, 233:395–406, 1996.

[147] G. Dueck and T. Scheuer. *J. Comp. Phys.*, 90:161–175, 1990.

[148] F. Glover. *Interfaces*, 20:74–94, 1990.

[149] D. Cvijovic and J. Klinowski. *Science*, 267:664–666, 1995.

[150] M. Ji and J. Klinowski. *Proc. Roy. Soc. A*, 462:3613–3627, 2006.

[151] T. Huber, A. Torda, and W. F. van Gunsteren. *J. Comput. Aided Mol. Des.*, 8:695–708, 1994.

[152] S. Goedecker. *J. Chem. Phys.*, 120:9911–9917, 2004.

[153] A. Laio and M. Parrinello. *Proc. Nat. Acad. Sci.*, 99:12562–12566, 2002.

[154] R. Martonak. In A. R. Oganov, editor, *Modern methods of crystal structure prediction*, chapter 5, pages 107–130. Wiley VCh, Weinheim, 2011.

[155] A. F. Voter. *J. Chem. Phys.*, 106:4665–4677, 1997.

[156] A. F. Voter. *Phys. Rev. Lett.*, 78:3908–3911, 1997.

[157] R. J. Wawak, J. Pillardy, A. Liwo, K. D. Gibson, and H. A. Scheraga. *J. Phys. Chem. A*, 102:2904–2918, 1998.

[158] Y. Zhang, D. Kihara, and J. Skolnick. *Proteins: Struct. Funct. Gen.*, 48:192–201, 2002.

[159] Z. Zhu, M. E. Tuckerman, S. O. Samuelson, and G. J. Martyana. *Phys. Rev. Lett.*, 88:100201, 2002.

[160] H. Merlitz and W. Wenzel. *Chem. Phys. Lett.*, 362:271–277, 2002.

[161] D. Hamelberg, J. Morgan, and J. A. McCommon. *J. Chem. Phys.*, 120:11919–11929, 2004.

[162] D. Hamelberg, T. Shen, and J. A. McCommon. *J. Chem. Phys.*, 122:241103, 2005.

[163] W. Zhang and Y. Duan. *Protein: Eng. Design Struct.*, 19:55–65, 2006.

[164] W. Wenzel and K. Hamacher. *Phys. Rev. Lett.*, 82:3003–3007, 1999.

[165] K. Hamacher. *Europhys. Lett.*, 74:944–950, 2006.

[166] K. Hamacher. *Physica A*, 378:307–314, 2007.

[167] L. Cheng, W. Cai, and X. Shao. *ChemPhysChem*, 6:261–266, 2005.

[168] G. Dueck. *J. Comp. Phys.*, 104:86–92, 1993.

[169] L. J. Fogel, A. J. Owens, and M. J. Walsh. *Artificial Intelligence Through Simulated Evolution*. Wiley, New York, 1966.

[170] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, Stuttgart, 1973.

[171] L. D. Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, 1991.

[172] D. Whitley. *Statist. Comput.*, 4:65–85, 1994.

[173] D. A. Coley. *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific, Singapore, 1999.

[174] T. S. Bush, C. R. A. Catlow, and P. D. Battle. *J. Mater. Chem.*, 5:1269, 1995.

[175] S. M. Woodley. *Struct. Bonding*, 110:95, 2004.

[176] A. R. Oganov and C. W. Glass. *J. Chem. Phys.*, 124:244704, 2006.

[177] A. Kolmogorov, S. Shah, E. R. Margine, A. F. Bialon, T. Hammerschmidt, and R. Drautz. *Phys. Rev. Lett.*, 105:217003, 2010.

[178] H. P. Schwefel. *Numerical Optimization of Computer Models*. Wiley, New York, 1981.

[179] J. Baker. In J. Grefenstette, editor, *Proceedings of the international conference on genetic algorithms and their applications*. Erlbaum, Hillsdale, 1985.

[180] D. E. Goldberg and K. Deb. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 66–93. Morgan Kaufmann, San Mateo, CA, 1991.

[181] H. Mühlenbein. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature 2*. North-Holland, Amsterdam, 1992.

[182] J. Baker. In J. Grefenstette, editor, *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference*. Erlbaum, Hillsdale, 1987.

[183] D. Whitley and J. Kauth. Genitor: a different genetic algorithm. In *Proceedings of the Rocky Mountains Conference on Artificial Intelligence*, pages 118–130, Denver, 1988.

[184] Y. Xiao and D. E. Williams. *Chem. Phys. Lett.*, 215:17–24, 1993.

[185] Y. Zeiri. *Phys. Rev. E*, 51:R2769–R2772, 1995.

[186] V. E. Bazterra, M. B. Ferraro, and J. C. Facelli. *J. Chem. Phys.*, 116:5984–5991, 2002.

[187] J. A. Niesse and H. R. Mayne. *J. Comp. Chem.*, 18:1233–1244, 1997.

[188] S. Boettcher and A. G. Percus. *Phys. Rev. Lett.*, 86:5211–5214, 2001.

[189] T. Klotz and S. Kobe. *Acta Phys. Slov.*, 44:347–356, 1994.

[190] P. Sibani and P. Schriver. *Phys. Rev. B*, 49:6667–6671, 1994.

[191] P. Sibani, R. v. d. Pas, and J. C. Schön. *Comp. Phys. Comm.*, 116:17–27, 1999.

[192] W. W. Tipton and R. G. Hennig. In A. R. Oganov, editor, *Modern methods of crystal structure prediction*, chapter 3, pages 55–66. Wiley VCh, Weinheim, 2011.

[193] C. J. Pickard and R. J. Needs. *J. Phys. Cond. Matter*, 23:053201, 2011.

[194] B. Winkler, C. J. Pickard, V. Milman, and G. Thimm. *Chem. Phys. Lett.*, 337:36–42, 2001.

[195] R. T. Strong, C. J. Pickard, V. Milman, G. Thimm, and B. Winkler. *Phys. Rev. B*, 70:045101, 2004.

[196] A. Le Bail. *J. Appl. Cryst.*, 38:389–395, 2005.

[197] G. Ghosh S. Delsante, G. Borzone, M. Asta, and R. Ferro. *Acta Mater.*, 54:4977–4997, 2006.

[198] Z. Li and H. A. Scheraga. *Proc. Nat. Acad. Sci.*, 84:6611–6615, 1987.

[199] V. Buch, R. Martonak, and M. Parrinello. *J. Chem. Phys.*, 124:204705, 2006.

[200] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope. *J. Chem. Phys.*, 126:155101, 2007.

[201] J. C. Schön, I. V. Pentin, and M. Jansen. *Phys. Chem. Chem. Phys.*, 8:1778–1784, 2006.

[202] H. B. Schlegel. *J. Comp. Chem.*, 24:1514–1527, 2003.

[203] E. E. Santiso and K. E. Gubbins. *Mol. Sim.*, 30:699–748, 2004.

[204] L. Angelani, R. Di Leonardo, G. Ruocco, A. Scala, and F. Sciortino. *J. Chem. Phys.*, 116:10297–10306, 2002.

[205] A. Banerjee, N. Adams, J. Simmons, and R. Shepard. *J. Phys. Chem.*, 89:52–57, 1985.

[206] R. S. Berry, H. L. Davis, and T. L. Beck. *Chem. Phys. Lett.*, 147:13–17, 1988.

[207] I. V. Ionova and E. A. Carter. *J. Chem. Phys.*, 98:6377–6386, 1993.

[208] D. J. Wales. *J. Chem. Soc. Farad. Trans.*, 89:1305–1313, 1993.

[209] J. C. Mauro, R. J. Loucks, and J. Balakrishnan. *J. Phys. Chem. A*, 109:9578–9583, 2005.

[210] G. Mills and H. Jonsson. *Phys. Rev. Lett.*, 72:1124–1127, 1994.

[211] H. Tanaka. *J. Chem. Phys.*, 113:11202–11211, 2000.

[212] E. Weinan, R. Weiqing, and E. Vanden-Eijnden. *Phys. Rev. B*, 66:052301, 2002.

[213] P. G. Bolhuis, C. Dellago, and D. Chandler. *Faraday Discuss.*, 110:421–436, 1998.

[214] H. Grubmüller. *Phys. Rev. E*, 52:2893–2906, 1995.

[215] C. Dellago, P. Bolhuis, F. S. Csajka, and D. Chandler. *J. Chem. Phys.*, 108:1964, 1998.

[216] C. Dellago, P. Bolhuis, and P. L. Geissler. In M. Ferrario, G. Ciccotti, and K. Binder, editors, *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, page 124. Springer, New York, 2006.

[217] L. R. Pratt. *J. Chem. Phys.*, 85:5045–5048, 1986.

[218] B. Peters, W. Z. Liang, A. T. Bell, and A. Chakraborty. *J. Chem. Phys.*, 118:9533–9541, 2003.

[219] L. Y.. Chen and P. L. Nash. *J. Chem. Phys.*, 119:12749–12752, 2003.

[220] R. J. Dimelow, R. A. Bryce, A. J. Masters, I. H. Hillier, and N. A. Burton. *J. Chem. Phys.*, 124:114113, 2006.

[221] H. A. Kramers. *Physica*, VII:284–304, 1940.

[222] D. G. Truhlar and B. C. Garrett. *Acc. Chem. Res.*, 13:440–448, 1980.

[223] P. Hänggi, P. Talkner, and M. Borkovec. *Rev. Mod. Phys.*, 62:251–341, 1990.

[224] E. Vanden-Eijnden and F. A. Tal. *J. Chem. Phys.*, 123:184103, 2005.

[225] J. Henin and C. Chipot. *J. Chem. Phys.*, 121:2904–2914, 2004.

[226] L. Maragliano and E. Vanden-Eijnden. *Chem. Phys. Lett.*, 426:168–175, 2006.

[227] B. Ensing, M. de Vivo, Z. Liu, P. Moore, and M. L. Klein. *Acc. Chem. Res.*, 39:73–81, 2006.

[228] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello. *J. Phys. Chem. B*, 110:3533–3539, 2006.

[229] G. Hummer and I. G. Kevrekidis. *J. Chem. Phys.*, 118:10762–10773, 2003.

[230] J. He, Z. Zhang, Y. Shi, and H. Liu. *J. Chem. Phys.*, 119:4005–4017, 2003.

[231] B. Alakent, M. C. Camurdan, and P. Doruker. *J. Chem. Phys.*, 123:144910, 2005.

[232] B. Alakent, M. C. Camurdan, and P. Doruker. *J. Chem. Phys.*, 123:144911, 2005.

[233] M. A. Amat, I. G. Kevrekidis, and D. Maroudas. *Phys. Rev. B*, 74:132201, 2006.

[234] M. A. Amat, I. G. Kevrekidis, and D. Maroudas. *Appl. Phys. Lett.*, 90:171910, 2007.

[235] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer. *Mol. Sim.*, 10:291–308, 1993.

[236] J. Schlitter, W. Swegat, and T. Mülders. *J. Mol. Model.*, 7:171–177, 2001.

[237] S. A. Corcelli, J. A. Rahman, and J. C. Tully. *J. Chem. Phys.*, 118:1085–1088, 2003.

[238] K. W. Borrelli, A. Vitalis, R. Alcantara, and V. Guallar. *J. Chem. Theo. Comp.*, 1:1304–1311, 2005.

[239] S. Yang, J. N. Onuchic, and H. Levine. *J. Chem. Phys.*, 125:054910, 2006.

[240] J. Hu, A. Ma, and A. R. Dinner. *J. Chem. Phys.*, 125:114101, 2006.

[241] C. Burisch, P. R. L. Markwick, N. L. Doltsinis, and J. Schlitter. *J. Chem. Theo. Comp.*, 4:164–172, 2008.

[242] J. C. Schön and M. Jansen. *Z. Krist.*, 216:361–383, 2001.

[243] H. Putz, J. C. Schön, and M. Jansen. *Comp. Mater. Sci.*, 11:309–322, 1998.

[244] J. C. Schön, Ž. Čančarević, and M. Jansen. *J. Chem. Phys.*, 121:2289–2304, 2004.

[245] Ž. Čančarević, J. C. Schön, and M. Jansen. *Phys. Rev. B*, 73:224114, 2006.

[246] C. J. Pickard and R. J. Needs. *Phys. Rev. Lett.*, 97:045504, 2006.

[247] K. Doll, J. C. Schön, and M. Jansen. *Phys. Chem. Chem. Phys.*, 9:6128, 2007.

[248] K. Doll, J. C. Schön, and M. Jansen. *Phys. Rev. B*, 78:144110, 2008.

[249] D. Hohl, R. Jones, R. Car, and M. Parrinello. *Chem. Phys. Lett.*, 139:540–545, 1987.

[250] R. O. Jones and G. Seifert. *J. Chem. Phys.*, 96:7564–7572, 1992.

[251] B. Hartke. *Theor. Chem. Acc.*, 99:241–247, 1998.

[252] K. Doll, J. C. Schön, and M. Jansen. *J. Phys.: Conf. Ser.*, 117:012014, 2008.

[253] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. Oxford University Press, London, 1954.

[254] R. A. Buckingham. *Proc. Roy. Soc. London A*, 168:264, 1938.

[255] J. C. Schön and M. Jansen. *Comp. Mater. Sci.*, 4:43–58, 1995.

[256] H. Putz, J. C. Schön, and M. Jansen. *Z. Anorg. Allg. Chem.*, 625:1624–1630, 1999.

[257] B. G. Dick and A. W. Overhauser. *Phys. Rev.*, 112:90, 1958.

[258] K. Fischer, H. Blitz, R. Habernkorn, and W. Weber. *Phys. Status Solidi B*, 54:285, 1972.

[259] R. E. Cohen. *Geophys. Res. Lett.*, 14:37–40, 1987.

[260] V. Nusslein and U. Schröder. *Phys. Status Solidi B*, 21:309, 1967.

[261] R. G. Gordon and Y. S. Kim. *J. Chem. Phys.*, 56:3122, 1972.

[262] L. L. Boyer, M. J. Mehl, J. L. Feldman, J. R. Hardy, J. W. Flocken, and C. Y. Fong. *Phys. Rev. Lett.*, 54:1940–1943, 1985.

[263] G. H. Wolf and M. S. T. Bukowinski. *Phys. Chem. Min.*, 15:209–220, 1988.

[264] O. V. Ivanov and E. G. Maksimov. *Phys. Rev. Lett.*, 69:108–111, 1992.

[265] R. E. Cohen. In P.J. Heaney, C.T. Prewitt, and G.V. Gibbs, editors, *MSA Reviews in Mineralogy; Silica: Physical Behavior, Geochemistry, and Materials Applications*, volume 29, pages 369–402. Mineralogical Society of America, Washington, D.C., 1994.

[266] B. B. Karki, L. Stixrude, and R. M. Wentzcovitch. *Rev. Geophys.*, 39:507–534, 2001.

[267] J. Behler, R. Martonak, D. Donadio, and M. Parrinello. *Phys. Rev. Lett.*, 100:185501, 2008.

[268] H. Eshet, R. Z. Khaliullin, T. D. Kuhne, J. Behler, and M. Parrinello. *Phys. Rev. B*, 81:184107, 2010.

[269] J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, and V. Caignaert. *Nature*, 346:343–345, 1990.

[270] R. Hundt, J. C. Schön, A. Hannemann, and M. Jansen. *J. Appl. Cryst.*, 32:413–416, 1999.

[271] A. Hannemann, R. Hundt, J. C. Schön, and M. Jansen. *J. Appl. Cryst.*, 31:922–928, 1998.

[272] R. Hundt, J. C. Schön, and M. Jansen. *J. Appl. Cryst.*, 39:6–16, 2006.

[273] R. Hundt. KPLOT: A Program for Plotting and Investigation of Crystal Structures, University of Bonn, Germany, Version 9, 2007, 1979.

[274] http://www.crystalimpact.com/download/kplot.htm.

[275] M. Jansen, I. V. Pentin, and J. C. Schön. *submitted to Nature Materials*, 2011.

[276] http://www.thermocalc.com/.

[277] http://www.factsage.com.

[278] http://www.mtdata.software.com/.

[279] http://www.computherm.com/pandat.html/.

[280] N. Saunders and A. P. Miodownik. *CALPHAD (Calculation of Phase Diagrams): A comprehensive guide*. Elsevier, New York, 1998.

[281] Z.-K. Liu, S. Hansen, J. Murray, P. Spencer, and J. Saal. *Comp. Coupl. Phase Diagr. Thermochem.*, 32:9–31, 2008.

[282] T. V. Bogdan, D. J. Wales, and F. Calvo. *J. Chem. Phys.*, 124:044102, 2006.

[283] J. C. Mauro and R. J. Loucks. *Phys. Rev. B*, 76:174202, 2007.

[284] R. H. Swendsen and A. M. Ferrenberg. *Phys. Rev. Lett.*, 61:2635, 1989.

[285] R. H. Swendsen and A. M. Ferrenberg. *Phys. Rev. Lett.*, 63:1195, 1993.

[286] F. Wang and D. P. Landau. *Phys. Rev. Lett.*, 86:2050, 2001.

[287] F. Wang and D. P. Landau. *Phys. Rev. E*, 64:056101, 2001.

[288] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. *J. Comp. Chem.*, 13:1011, 1992.

[289] B. Roux. *Comp. Phys. Comm.*, 91:275–282, 1995.

[290] A. P. Lyubartsev. *J. Chem. Phys.*, 96:1776–1783, 1992.

[291] G. R. Smith and A. D. Bruce. *Phys. Rev. E*, 53:6530–6543, 1996.

[292] S. B. Opps and J. Schofield. *Phys. Rev. E*, 63:056701, 2001.

[293] N. Kamiya and J. Higo. *J. Comp. Chem.*, 22:1098–1106, 2001.

[294] F. Yasar, H. Arkin, T. Celik, B. A. Berg, and H. Meirovitch. *J. Comp. Chem.*, 23:1127–1134, 2002.

[295] R. Jono, K. Shimizu, and T. Terada. *Chem. Phys. Lett.*, 432:306–312, 2006.

[296] J. Kim, J. E. Straub, and T. Keyes. *Phys. Rev. Lett.*, 97:050601, 2006.

[297] J. Kim, J. E. Straub, and T. Keyes. *Phys. Rev. E*, 76:011913, 2007.

[298] P. Liu and G. A. Voth. *J. Chem. Phys.*, 126:045106, 2007.

[299] S. G. Itoh and Y. Okamoto. *J. Chem. Phys.*, 124:104103, 2006.

[300] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Harcourt Brace, New York, 1976.

[301] G. Kern, G. Kresse, and J. Hafner. *Phys. Rev. B*, 59:8551–8559, 1999.

[302] B. Grabowski, T. Hickel, and J. Neugebauer. *Phys. Rev. B*, 76:024309, 2007.

[303] R. Drautz, A. Diaz-Ortiz, M. Fähnle, and H. Dosch. *Phys. Rev. Lett.*, 93:067202, 2004.

[304] P. G. Gonzales-Ormeno, H. M. Petrilli, and C. G. Schön. *Scripta Mater.*, 54:1271–1276, 2006.

[305] F. Körmann, A. Dick, B. Grabowski, B. Hallstedt, T. Hickel, and J. Neugebauer. *Phys. Rev. B*, 78:033102, 2008.

[306] D. Beeman and R. Alben. *Advances in Phys.*, 26(3):339–361, 1977.

[307] G. J. Ackland. *J. Phys.: Cond. Matter*, 14:2975–3000, 2002.

[308] J. G. Kirkwood. *J. Chem. Phys.*, 3:300–313, 1935.

[309] R. Zwanzig. *J. Chem. Phys.*, 22:1420–1426, 1954.

[310] M. Watanabe and W. P. Reinhardt. *Phys. Rev. Lett.*, 65:3301–3304, 1990.

[311] J. C. Schön. *J. Chem. Phys.*, 105:10072–10083, 1996.

[312] L. Rosso, P. Minary, Z. Zhu, and M. E. Tuckerman. *J. Chem. Phys.*, 116:4389–4402, 2002.

[313] D. J. Wales and T. V. Bogdan. *J. Phys. Chem. B*, 110:20765–20776, 2006.

[314] E. Darve and A. Pohorille. *J. Chem. Phys.*, 115:9169–9183, 2001.

[315] X. Wu and B. R. Brooks. *Chem. Phys. Lett.*, 381:512–518, 2003.

[316] L. Maragliano and E. Vanden-Eijnden. *J. Chem. Phys.*, 128:184110, 2008.

[317] G. M. Torrie and J. P. Valleau. *Chem. Phys. Lett.*, 28:578, 1977.

[318] M. Mezei. *Mol. Sim.*, 9:257–267, 1992.

[319] D. A. Kofke and E. D. Glandt. *Mol. Phys.*, 64:1105–1131, 1988.

[320] N. L. Allan, G. D. Barrera, M. Y. Lavrentiev, I. T. Todorov, and J. A. Purton. *J. Mater. Chem.*, 11:63–68, 2001.

[321] M. Strnad and I. Nezbeda. *Mol Sim.*, 22:183–198, 1999.

[322] O. Delgado-Buscalioni and P. V. Coveney. *J. Chem. Phys.*, 119:978–990, 2003.

[323] D. M. Zuckerman and T. B. Woolf. *J. Stat. Phys.*, 114:1303–1323, 2004.

[324] N. S. Golosov and A. M. Tolstik. *J. Phys. Chem. Solids*, 36:899–902, 1975.

[325] F. Ducastelle and F. Gautier. *J. Phys. F: Metal Phys.*, 6:2039–2061, 1976.

[326] J. M. Sanchez and D. De Fontaine. In *Structure and Bonding in Crystals*, page 117. Academic, New York, 1981.

[327] D. De Fontaine. In F. Seitz, D. Turnbull, and H. Ehrenreich, editors, *Solid State Physics Vol. 47*, pages 33–180. Academic Press, New York, 1994.

[328] N. Saunders. *Phil. Trans. Roy. Soc. Lond. A*, 351:543–561, 1995.

[329] A. V. Ruban and I. A. Abrikosov. *Rep. Prog. Phys.*, 71:046501, 2008.

[330] J. W. D. Connolly and A. R. Williams. *Phys. Rev. B*, 27:5169–5172, 1983.

[331] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger. *Phys. Rev. B*, 46:12587–12605, 1992.

[332] C. Wolverton and A. Zunger. *Phys. Rev. B*, 50:10548–10560, 1994.

[333] A. Gonis, X.-G. Zhang, A. J. Freeman, P. Turchi, G. M. Stocks, and D. M. Nicholson. *Phys. Rev. B*, 36:4630–4646, 1987.

[334] P. E. A. Turchi, M. Sluiter, and G. M. Stocks. *J. Phase Equil.*, 13:391–399, 1992.

[335] D. E. Nanu, Y. Deng, and A. J. Böttger. *Phys. Rev. B*, 74:014113, 2006.

[336] R. Grau-Crespo, S. Hamad, C. R. A. Catlow, and N. H. DeLeeuw. *J. Phys.: Cond. Matter*, 19:256201, 2007.

[337] M. Sanati, L. G. Wang, and A. Zunger. *Phys. Rev. Lett.*, 90:045502, 2003.

[338] V. Blum and A. Zunger. *Phys. Rev. B*, 69:020103, 2004.

[339] H. Haas and M. Jansen. *Angew. Chem. Int. Ed. Eng.*, 38:1910–1911, 1999.

[340] J. C. Schön. *Z. Anorg. Allg. Chem.*, 630:2354–2366, 2004.

[341] F. Claeyssens, C. L. Freeman, N. L. Allan, Y. Sun, M. N. R. Ashfold, and J. H. Harding. *J. Mater. Chem.*, 15:139–148, 2005.

[342] I. V. Pentin, J. C. Schön, and M. Jansen. *Phys. Rev. B*, 82:144102, 2010.

[343] O. Redlich and A. T. Kister. *Ind. Eng. Chem.*, 40:345, 1948.

[344] I. V. Pentin, J. C. Schön, and M. Jansen. *Phys. Chem. Chem. Phys.*, 12:8491–8499, 2010.

[345] I. V. Pentin, J. C. Schön, and M. Jansen. *J. Chem. Phys.*, 126:124508, 2007.

[346] J. C. Schön, I. V. Pentin, and M. Jansen. *J. Phys. Chem. B*, 111:3943, 2007.

[347] I. V. Pentin, J. C. Schön, and M. Jansen. *Solid State Sci.*, 10:804–813, 2008.

[348] J. C. Schön, I. V. Pentin, and M. Jansen. *Solid State Sci.*, 10:455–460, 2008.

[349] Ž. Čančarević, J. C. Schön, and M. Jansen. *Z. Anorg. Allg. Chem.*, 632:1437–1448, 2006.

[350] Ž. Čančarević, J. C. Schön, and M. Jansen. *Chemistry Europ. J.*, 13:7330–7348, 2007.

[351] J. C. Schön and M. Jansen. *Ber. Bunsenges.*, 98:1541–1544, 1994.

[352] H. Putz, J. C. Schön, and M. Jansen. *Ber. Bunsenges.*, 99:1148–1153, 1995.

[353] M. M. Dacorogna and M. L. Cohen. *Phys. Rev. B*, 34:4996–5002, 1986.

[354] J. C. Schön and M. Jansen. *unpubl.*

[355] M. A. Zwijnenburg, K. E. Jelfs, and S. T. Bromley. *Phys. Chem. Chem. Phys.*, 12:8505–8512, 2010.

[356] M. Martinez-Canales, A. R. Oganov, Y. Ma, Y. Yan, A. O. Lyakhov, and A. Bergara. *Phys. Rev. Lett.*, 102:087005, 2009.

[357] G. Trimarchi and A. Zunger. *Phys. Rev. B*, 75:104113, 2007.

[358] N. L. Abraham and M. I. J. Probert. *Phys. Rev. B*, 73:224104, 2006.

[359] C. J. Pickard and R. J. Needs. *Phys. Rev. Lett.*, 102:146401, 2009.

[360] C. J. Pickard and R. J. Needs. *J. Phys. Cond. Matter*, 21:452205, 2009.

[361] A. R. Oganov and V. L. Solozhenko. Boron: a hunt for superhard polymorphs. *J. Superhard Mater.*, 31:285–291, 2009.

[362] R. H. Wentorf, Jr. *Science*, 147:49–50, 1965.

[363] P. Balog, D. Orosel, Ž. Čančarević, J. C. Schön, and M. Jansen. *J. Alloys Comp.*, 429:87–98, 2007.

[364] A. Takada, C. R. A. Catlow, and G. D. Price. *J. Phys.: Cond. Mat.*, 7:8659–8692, 1995.

[365] R. Martonak, D. Donadio, A. R. Oganov, and M. Parrinello. *Nature Materials*, 5:623, 2006.

[366] C. M. Freeman, J. M. Newsam, S. M. Levine, and C. R. A. Catlow. *J. Mater. Chem.*, 3:531–535, 1993.

[367] D. Fischer, Ž. Čančarević, J. C. Schön, and M. Jansen. *Z. Anorg. Allg. Chem.*, 630:156–160, 2004.

[368] Ž. Čančarević, J. C. Schön, and M. Jansen. *Z. Anorg. Allg. Chem.*, 631:1167–1171, 2005.

[369] J. C. Schön, A. Hannemann, G. Sethi, I. V. Pentin, and M. Jansen. *submitted to Serb. J. Ceram. Res.*, 2011.

[370] P. Kroll, T. Schröter, and M. Peters. *Angew. Chem. Int. Ed.*, 44:4249–4254, 2005.

[371] C. Jiang, Z. Lin, and Y. Zhao. *Phys. Rev. Lett.*, 103:185501, 2009.

[372] M. Wessel and R. Dronskowski. *J. Am. Chem. Soc.*, 132:2421, 2009.

[373] A. R. Oganov, C. W. Glass, and S. Ono. *Earth Planet. Sci. Lett.*, 241:95–103, 2006.

[374] A. R. Oganov, S. Ono, C. W. Glass, and A. Garcia. *Earth Planet. Sci. Lett.*, 273:38–47, 2008.

[375] M. A. C. Wevers, J. C. Schön, and M. Jansen. *J. Solid State Chem.*, 136:223–246, 1998.

[376] V. Ozolins, E. H. Majzoub, and C. Wolverton. *Phys. Rev. Lett.*, 100:135501, 2008.

[377] I. V. Pentin, J. C. Schön, and M. Jansen. *Z. Anorg. Allg. Chem.*, 636:1703–1709, 2010.

[378] S. Ono, T. Kikegawa, Y. Ohishi, and J. Tsuchiya. *Am. Miner.*, 90:667, 2005.

[379] S. Ono, T. Kikegawa, and Y. Ohishi. *Am. Miner.*, 92:1246, 2007.

[380] S. V. Barabash, V. Blum, S. Muller, and A. Zunger. *Phys. Rev. B*, 74:035108, 2006.

[381] A. R. Oganov, Y. Ma, C. W. Glass, and M. Valle. *Psi-k Newsletter*, 84:142–171, 2007.

[382] S. Curtarolo, D. Morgan, and G. Ceder. *Calphad*, 29:163, 2005.

[383] J. N. Hart, N. L. Allan, and F. Claeyssens. *Phys. Chem. Chem. Phys.*, 12:8620, 2010.

[384] J. K. Solbakk and K. O Stromme. *Acta Chem. Scand.*, 23:300, 1969.

[385] N. Onoda-Yamamuro, H. Honda, R. Ikeda, O. Yamamuro, T. Mtsuo, K. Oikawa, T. Kamiyama, and F. Izumi. *J. Phys. Cond. Matter*, 10:3341–3351, 1998.

[386] J. C. Schön, P. Salamon, and M. Jansen. *in prep.*, 2011.

[387] A. Kulkarni, K. Doll, J. C. Schön, and M. Jansen. *J. Phys. Chem. B*, 114:15573–15581, 2010.

[388] A. Gavezzotti. *J. Amer. Chem. Soc.*, 113:4622–4629, 1991.

[389] J. R. Holden, Z. Du, and H. L. Ammon. *J. Comp. Chem.*, 14:422–437, 1993.

[390] R. J. Gdanitz. In A. Gavezzotti, editor, *Theoretical Aspects and Computer Modeling*, pages 185–201. Wiley, New York, 1997.

[391] P. Verwer and F. J. J. Leusen. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews of Computational Chemistry*, volume 12, pages 327–365. Wiley-VCH, New York, 1998.

[392] B. P. van Eijck and J. Kroon. *J. Comp. Chem.*, 20:799–812, 1999.

[393] G. M. Day et al. *Acta Cryst. B*, 61:511–527, 2005.

[394] P. G. Karamertzanis and C. C. Pantelides. *Mol. Phys.*, 105:273–291, 2007.

[395] W. H. Baur and D. Kassner. *Acta Cryst. B*, 48:356–369, 1992.

[396] N. Padmaja, S. Ramakumar, and M. A. Wiswamitra. *Acta Cryst. A*, 46:725–730, 1990.

[397] G. Filippini and A. Gavezzotti. *Mol. Cryst. Liq. Cryst.*, 219:37–41, 1992.

[398] S. L. Price. *Adv. Drug Deliv. Rev.*, 56:301–319, 2004.

[399] G. M. Day, W. D. S. Motherwell, and W. Jones. *Phys. Chem. Chem. Phys.*, 9:1693–1704, 2007.

[400] S. M. Woodley, C. R. A. Catlow, P. D. Battle, and J. D. Gale. *Chem. Comm.*, 2004:22–23, 2004.

[401] N. Engel. *Acta Cryst. B*, 47:849–858, 1991.

[402] M. A. Zwijnenburg, S. T. Bromley, M. D. Foster, R. G. Bell, O. Delgado-Friedrichs, J. C. Jansen, and T. Maschmeyer. *Chem. Mater.*, 16:3809–3820, 2004.

[403] L. M. R. Albelo, A. R. Ruiz-Salvador, D. W. Lewis, A. Gomez, P. Mialane, J. Marrot, A. Dolbecq, A. Sampieri, and C. Mellot-Draznieks. *Phys. Chem. Chem. Phys.*, 12:8632, 2010.

[404] M. W. Deem and J. M. Newsam. *Nature*, 342:260–262, 1989.

[405] http://www.iza-structure.org/database.

[406] S. M. Woodley, M. B. Watkins, A. A. Sokol, S. A. Shevlin, and C. R. A. Catlow. *Phys. Chem. Chem. Phys.*, 11:3176, 2009.

[407] M. B. Watkins, S. A. Shevlin, A. A. Sokol, B. Slater, C. R. A. Catlow, and S. M. Woodley. *Phys. Chem. Chem. Phys.*, 11:3186, 2009.

[408] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keefe, and O. M. Yaghi. *Acc. Chem. Res.*, 38:176–182, 2005.

[409] S. M. Woodley, P. D. Battle, J. D. Gale, and C. R. A. Catlow. *Phys. Chem. Chem. Phys.*, 6:1815–1822, 2004.

[410] S. M. Woodley. *Phys. Chem. Chem. Phys.*, 6:1823–1829, 2004.

[411] S. M. Woodley. *Phys. Chem. Chem. Phys.*, 9:1070–1077, 2007.

[412] J. C. Schön and M. Jansen. unpublished.

[413] M. Laradji, D. P. Landau, and B. Dünweg. *Phys. Rev. B*, 51:4894–4902, 1995.

[414] R. Hirschl, J. Hafner, and Y. Jeanvoine. *J. Phys.: Cond. Matter*, 13:3545–3572, 2001.

[415] C. Wolverton, V. Ozolins, and M. Asta. *Phys. Rev. B*, 69:144109, 2004.

[416] S. Curtarolo, A. N. Kolmogorov, and F. H. Cocks. *Comp. Coupl. Phase Diagr. Thermochem.*, 29:155–161, 2005.

[417] D. Fuks, S. Dorfman, S. Piskunov, and E. A. Kotomin. *Phys. Rev. B*, 71:014111, 2005.

[418] N. L. Allan, G. D. Barrera, M. Y. Lavrentiev, C. L. Freeman, I. T. Tordov, and J. A. Purton. *Comp. Mater. Sci.*, 36:42–48, 2006.

[419] J. A. Purton, N. L. Allan, M. Yu. Lavrentiev, I. T. Todorov, and C. L. Freeman. *Chem. Geol.*, 225:176–188, 2006.

[420] S. Bärthlein, G. L. W. Hart, A. Zunger, and S. Müller. *J. Phys.: Cond. Matter*, 19:032201, 2007.

[421] R. Kaplow, T. A. Rowe, and B. L. Averbach. *Phys. Rev.*, 168:1068–1079, 1968.

[422] R. L. McGreevy. In C. R. A. Catlow, editor, *Computer Modelling in Inorganic Crystallography*, pages 151–184. Acad. Press, San Diego, 1997.

[423] A. LeBail. In *Proc. EPDIC-7*. preprint, 2000.

[424] A. Mellergard and R. L. McGreevy. *Acta Cryst. A*, 55:783–789, 1999.

[425] C. R. A. Catlow, R. G. Bell, and J. D. Gale. *J. Mater. Chem.*, 4:781–792, 1994.

[426] D. K. Belashchenko. *Inorg. Mater. (Engl. Trans.)*, 30:966–976, 1994.

[427] H. Putz, J. C. Schön, and M. Jansen. *J. Appl. Cryst.*, 32:864–870, 1999.

[428] A. A. Coelho. *J. Appl. Cryst.*, 33:899–908, 2000.

[429] O. J. Lanning, S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, and G. W. Turner. *Chem. Phys. Lett.*, 317:296–303, 2000.

[430] H. Putz. *Endeavour 1.0*. Crystal Impact GbR, Bonn, 2000.

[431] M. Becker and M. Jansen. *Sol. State Sci.*, 2:711–715, 2000.

[432] W. A. Crichton, G. B. M. Vaughan, and M. Mezouar. *Z. Krist.*, 216:417–419, 2001.

[433] M. Pompetzki and M. Jansen. *Z. Anorg. Allg. Chem.*, 628:641–646, 2002.

[434] M. Schreyer and M. Jansen. *Angew. Chem. Int. Ed.*, 41:643, 2002.

[435] M. Schreyer and M. Jansen. *Sol. State Sci.*, 3:25–30, 2001.

[436] D. Santamaria-Perez, J. Haines, U. Amador, E. Moran, and A. Vegas. *Acta Cryst. B*, 62:1019–1024, 2006.

[437] J. B. Christian and M. S. Whittingham. *J. Solid State Chem.*, 181:1782–1791, 2008.

[438] M. Beekman, J. A. Kaduk, Q. Huang, W. Wong-Ng, Z. Yang, D. Wang, and G. S. Nolas. *Chem. Comm.*, 2007:837–839, 2007.

[439] J. Beck and S. Benz. *Z. Anorg. Allg. Chem.*, 635:962–965, 2009.

[440] R. E. Dinnebier, B. Hinrichsen, A. Lennie, and M. Jansen. *Acta Cryst. B*, 65:1–10, 2009.

[441] D. Fischer and M. Jansen. *J. Am. Chem. Soc.*, 124:3488, 2002.

[442] J. V. Badding. *Ann. Rev. Mater. Sci.*, 28:631–658, 1998.

[443] P. F. McMillan. *Chem. Commun.*, 8:919–923, 2003.

[444] H. Huppertz. *Z. Krist.*, 219:330–338, 2004.

[445] N. E. Brese and M. O'Keeffe. In *Struct. Bonding*, page 307. Springer, Heidelberg, 1992.

[446] D. Fischer, A. Müller, and M. Jansen. *Z. Anorg. Allg. Chem.*, 630:2697–2700, 2004.

[447] Ž. Čančarević, J. C. Schön, D. Fischer, and M. Jansen. *Progress in Materials Science and Processes (Mat. Sci. Forum)*, 494:61–66, 2005.

[448] A. Grzechnik, A. Vegas, K. Syassen, I. Loa, M. Hanfland, and M. Jansen. *J. Solid State Chem.*, 154:603, 2000.

[449] A. Vegas, A. Grzechnik, K. Syassen, I. Loa, M. Hanfland, and M. Jansen. *Acta Cryst. B*, 57:151–156, 2001.

[450] A. Vegas, A. Grzechnik, M. Hanfland, C. Mühle, and M. Jansen. *Solid State Sci.*, 4:1077–1081, 2002.

[451] D. Santamaria-Perez, A. Vegas, C. Mühle, and M. Jansen. *Acta Cryst. B*, 67:109–115, 2011.

[452] P. Salamon. Priv. comm., 1988.

[453] J. C. Schön. *J. Phys. A: Math. Gen.*, 30:2367–2389, 1997.

[454] J. C. Schön and Paolo Sibani. *J. Phys. A: Math. Gen.*, 31:8165–8178, 1998.

[455] H.-P. Cheng, X. Li, R. L. Whetten, and R. S. Berry. *Phys. Rev. A*, 46:791–800, 1992.

[456] Z. H. Li, A. W. Jasper, and D. G. Truhlar. *J. Amer. Chem. Soc.*, 129:14899–14910, 2007.

[457] J. S. Kirkaldy and E. A. Baganis. *Met. Trans.*, 9:495, 1978.

[458] K. Hasiguchi, J. S. Kirkaldy, T. Fukuzumi, and V. Pavaskar. *CALPHAD*, 8:173, 1984.

[459] M. Enomoto and H. I. Aaronson. *CALPHAD*, 9:43, 1985.

[460] A. A. B. Sugden and H. K. D. B. Bhadeshia. *Mat. Sci. Tch.*, 5:977, 1989.

[461] L. Ge, X. Hui, E. R. Wang, G. L. Chen, R. Arroyave, and Z. K. Liu. *Intermet.*, 16:27–33, 2008.

[462] J. Anwar and D. Zahn. *Angew. Chem. Int. Ed.*, 50:1996–2013, 2011.

[463] M. Santoro, J. C. Schön, and M. Jansen. *Phys. Rev. E*, 76:061120, 2007.

[464] J. C. Schön. *Z. Anorg. Allg. Chem.*, 635:1794–1806, 2009.

[465] C. R. A. Catlow, N. H. DeLeeuw, J. Anwar, R. J. Davey, K. J. Roberts, and P. R. Unwin, editors. *Faraday Discussions 136: Crystal Growth and Nucleation.* Roy. Soc. Chem., London, 2007.

[466] M. Jansen. *Abh. Rhein. Westf. Akad. Wiss.*, N420:7–35, 1996.

[467] D. C. Johnson. *Nature*, 454:174–175, 2008.

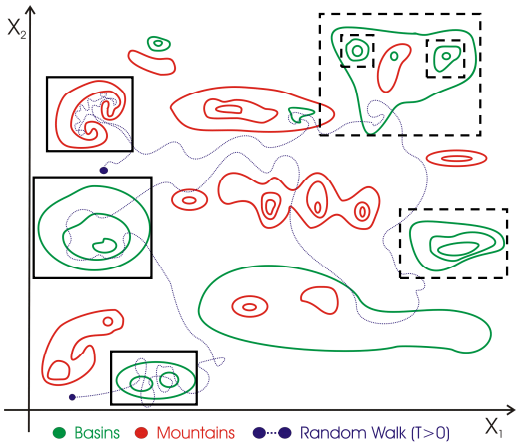

Figure 1: Sketch of an energy landscape consisting of minima regions (green level lines) and mountain regions (red level lines). Blue dotted line indicates one of the many possible trajectories, i.e. random walks, of the system. Regions contained in black rectangles are locally ergodic regions on the time scale of observation encountered during the simulation; dashed black rectangles are locally ergodic regions that have not been visited during this particular simulation run. Note that locally ergodic regions do not have to contain a local minimum - entropic barriers can be sufficient to establish local ergodicity[40].

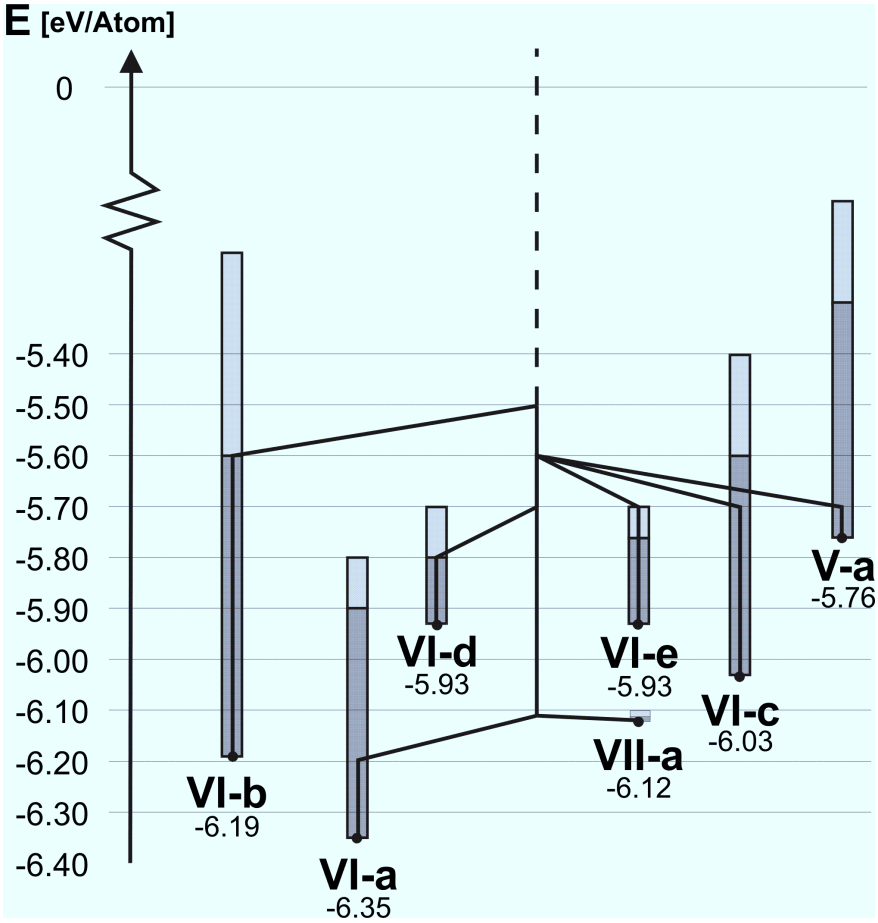

Figure 2: Excerpt of the tree graph of the MgF_2 system[103] including entropic barriers, indicated by the grey/black bars. Black/grey bar: Probability to leave the minimum during the simulation time at the given energy level was below 1% and 20%, respectively. Notation for minima: VI-a: rutile, VI-b: anatase, VI-c: half-filled rock salt structure, VII-a: structure consisting of monocapped prisms, VI-d: CaI_2 structure, VI-e: irregular structure containing F_6-prisms around Mg-atoms, V-a: structure containing F_5-trigonal bipyramids and square pyramids around Mg-atoms. At standard pressure, the only known modification of MgF_2 with an ordered structure exhibits the rutile structure-type.

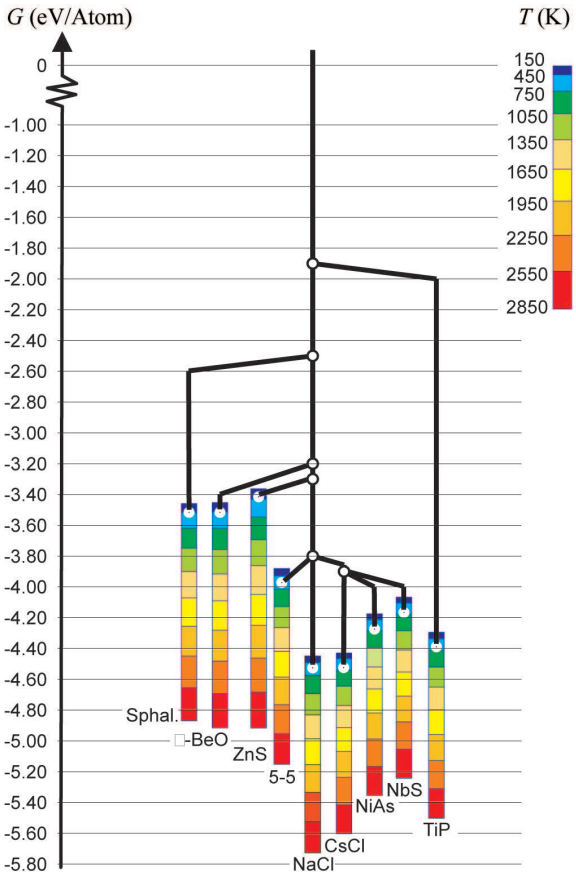

Figure 3: Free enthalpy landscape of SrO at p = 0 GPa for eight different temperatures (T = 150 K,..., 2850 K)[114] computed using global landscape explorations followed by free energy calculations in the quasi-harmonic approximation on empirical potential and ab initio level. The energetic contributions 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 e.g. [103]) are not shown to avoid overloading the figure.

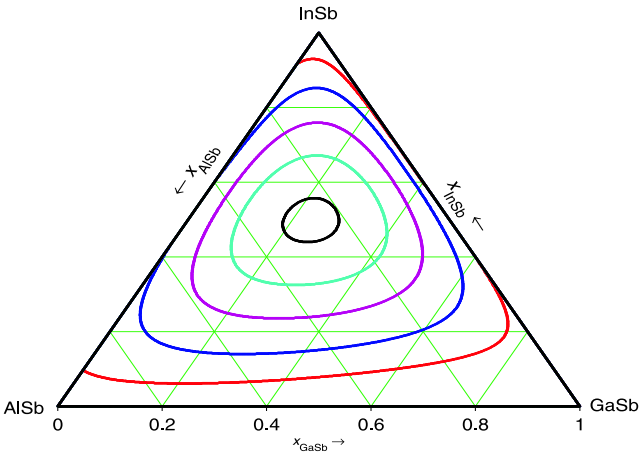

Figure 4: Five isothermal projections of the miscibility dome in the phase diagram for the AlSb-GaSb-InSb system at [250 350 450 520 570] K, based on the HF-calculations[342]. Red curve - 250 K, blue curve - 350 K, magenta curve - 450, cyan curve - 520 K and black curve - 570 K.

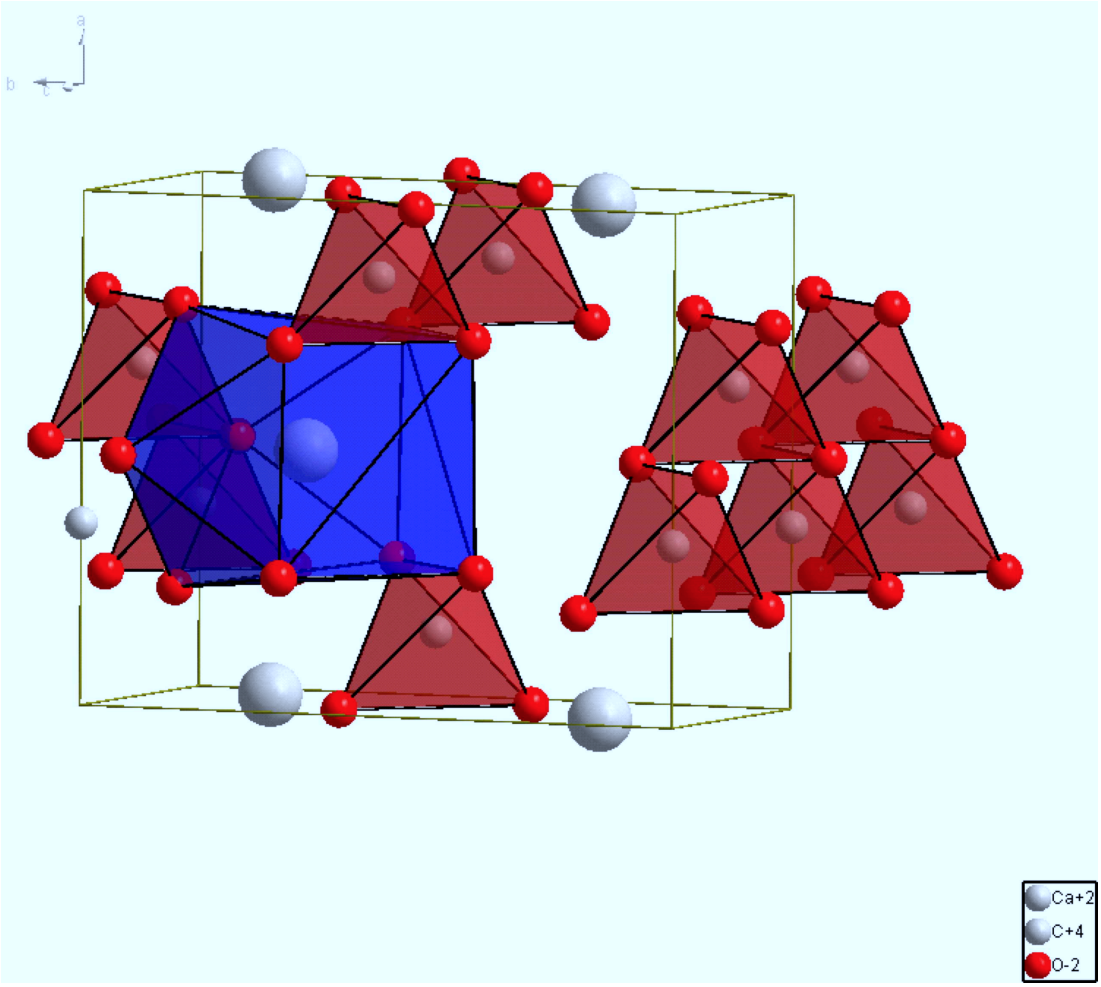

Figure 5: Predicted high-pressure structure of $CaCO_3$ exhibiting a corner-connected chain of CO_4 tetrahedra[373]. The calcium atoms are ten-fold coordinated by oxygen.

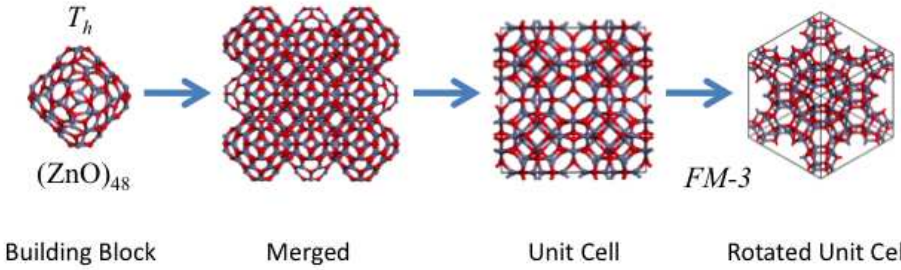

Figure 6: Predicted $(ZnO)_{48}$ bubble cluster, and subsequently predicted mesoporous structure (called 'framework F48', with space group $Fm\bar{3}$) generated by using the bubble cluster as a secondary building unit.[406] F48 is shown in three different perspectives.

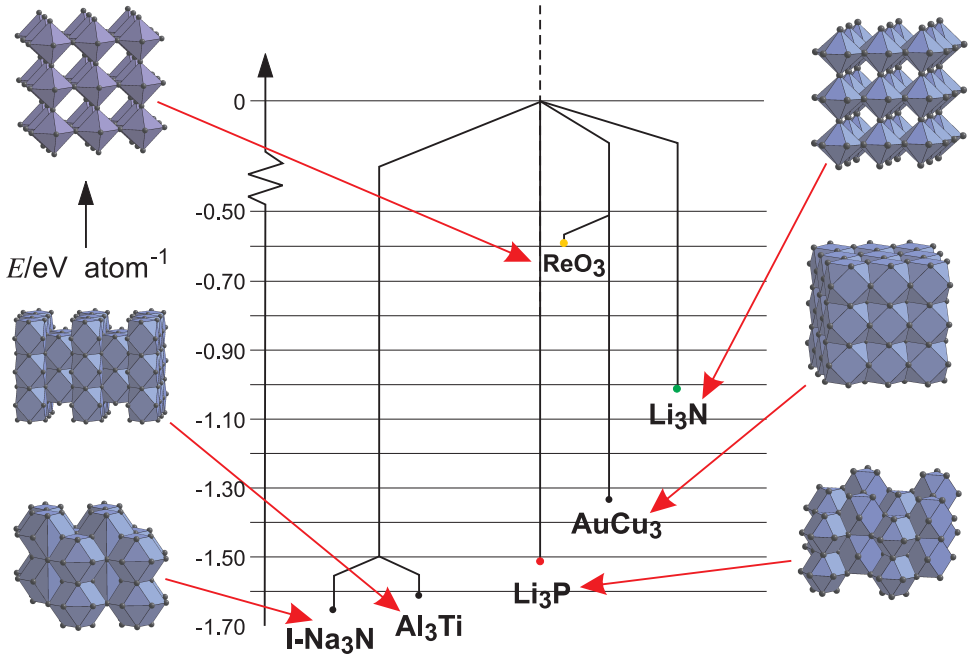

Figure 7: Excerpt of the tree graph of the energy landscape of Na_3N on empirical energy level at standard pressure, depicting some of the most important local minima [8, 9, 10, 11]. I-Na_3N corresponds to a strongly distorted Li_3Bi-structure type with 12(+2)-fold coordination of the nitrogen atoms by sodium atoms. Experimentally, the ReO_3-type [11], the Li_3N-, the Li_3P- and the Li_3Bi-structure types [12, 13] 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 [12, 13] that resembles several structures found as local minima on the enthalpy landscapes of the alkali nitrides.

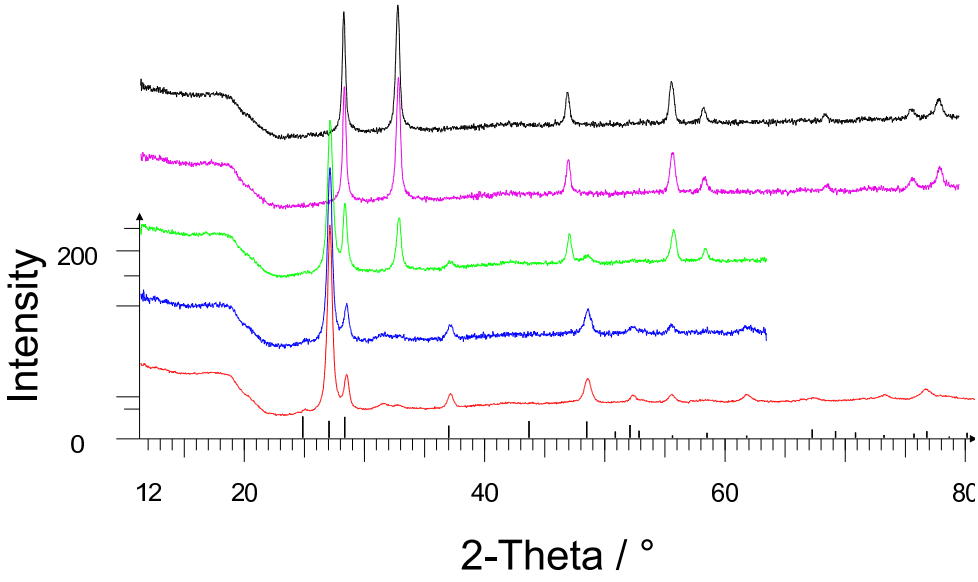

Figure 8: X-ray powder diffractograms of LiBr deposited via the LT-ABD method, as function of temperature (bottom to top: 223 K, 243 K, 263 K, 283 K, 298 K) [15]. The line diagram indicates the peaks corresponding to the wurtzite modification of LiBr.