

Exploring Energy Landscapes

David J. Wales

¹Department of Chemistry, University of Cambridge, Cambridge, UK, CB2 1EW;
email: dw34@cam.ac.uk

xxxxxx 0000. 00:1–25
Copyright © 0000 by Annual Reviews.
All rights reserved

Keywords

energy landscapes, global optimisation, enhanced sampling, rare events, machine learning

Abstract

Recent advances in the potential energy landscapes approach are highlighted, including both theoretical and computational contributions. Treating the high dimensionality for molecular and condensed matter systems of contemporary interest is important for understanding how emergent properties are encoded in the landscape, and in calculating these properties while faithfully representing barriers between different morphologies. The pathways calculated in full dimensionality, which are used to construct kinetic transition networks, may prove useful in guiding such calculations. The energy landscape perspective has also produced new procedures for structure prediction and analysis of thermodynamic properties. Basin-hopping global optimisation, with alternative acceptance criteria and generalisations to multiple metric spaces, has been used to treat systems ranging from biomolecules to nanoalloy clusters and condensed matter. The Review also illustrates how all this methodology, developed in the context of chemical physics, can be transferred to landscapes defined by cost functions associated with machine learning.

Contents

1. INTRODUCTION	2
2. Visualisation of Landscapes and Pathways	3
3. Recent Developments in Basin-Hopping Global Optimisation	7
3.1. The Superposition Approach	7
3.2. Basin-Hopping with Alternative Thermodynamic Potentials	9
3.3. Generalised Basin-Hopping	10
4. Overcoming Broken Ergodicity Using Basin-Sampling	11
5. Insight from Occupation Probabilities	14
5.1. Assignment of Heat Capacity Features	14
5.2. Quantifying the Complexity of a Landscape	18
6. Energy Landscapes for Machine Learning	20
7. Conclusions	21

1. INTRODUCTION

Analysis of potential energy surfaces (PESs) for molecular and condensed matter systems can provide both fundamental insight into emergent observable properties, and computational tools that complement conventional simulation techniques. There are three basic components to this framework, focussing on structure prediction, sampling thermodynamic properties, and describing global dynamics. The unifying theme for calculations is the exploration of the PES using geometry optimisation to characterise local minima and the transition states that connect them via steepest-descent pathways. These stationary points are defined geometrically by vanishing forces with no imaginary normal mode frequencies for local minima, and one for a transition state (1).

The geometry optimisation techniques employed in these calculations are well established. Minima are obtained by local minimisation, and double-ended connections between specified pairs are obtained by first running a doubly-nudged (2) elastic band (3, 4) (DNEB) calculation. Transition state candidates are identified from images in the DNEB interpolation corresponding to local maxima, and these configurations are accurately refined to transition states using hybrid eigenvector-following (5–8); the connectivity is defined by approximate steepest-descent pathways. The resulting databases of local minima and transition states constitute kinetic transition networks (9–12) when supplemented with unimolecular rate theory (13). Discrete path sampling provides a range of schemes for refining such networks (14, 15); some recent applications to proteins and nucleic acids have recently been discussed elsewhere (16).

This Review starts by considering the inherent problems involved in visualising high-dimensional landscapes and pathways, especially the issues that arise in projecting the full space onto low-dimensional reaction coordinates and order parameters (§2). The superposition approach, where thermodynamic properties are calculated from a database of local minima (17–22), is then summarised in §3.1. Some recent developments in basin-hopping global optimisation are described in §3, highlighting schemes based on free energy and alternative thermodynamic potentials (§3.2). The generalised basin-hopping approach, where searches are extended over different metric spaces, is then outlined in §3.3. Superposition theory again plays a key role in §4 and §5, in the calculation and interpretation of

global thermodynamic properties. The basin-sampling approach (§4) couples results from basin-hopping global optimisation with parallel tempering (83–85), providing equilibrium thermodynamic properties for landscapes that feature broken ergodicity. In §5 the focus shifts to occupation probabilities for local minima. This perspective enables heat capacity features to be assigned as transitions between well-defined regions of the PES without the need for a structure-based order parameter (§5.1), and provides a quantitative measure of the complexity of the landscape in terms of a frustration index (§5.2). The ability to decompose thermodynamic features, without reference to atomic structure, allows us to analyse the corresponding quantities defined for abstract landscapes, where the stationary points correspond to an arbitrary function that supports multiple minima. The landscapes defined by cost functions, which play the role of potential energy in machine learning fits of training data, constitute a particularly important application. One such machine learning landscape is described in §6 for prediction of classifications corresponding to outcomes for molecular geometry optimisation. These results highlight the versatility of optimisation techniques from energy landscape theory: exploring, predicting and explaining the properties of atomistic, mesoscopic, and condensed matter systems can be extended to abstract functions by defining thermodynamic and dynamical analogues, potentially yielding new insight into emergent behaviour.

2. Visualisation of Landscapes and Pathways

The dimensionality of most molecular systems of interest poses immediate problems for visualisation. Theory suggests a power law growth for the number of minimum energy structures, $N_{\min}^{\text{st}}(N)$, as a function of the number of atoms, N (23, 24). If the system is large enough so that it can be divided into m equivalent subsystems of N atoms each, and the subsystems are independent, then

$$N_{\min}^{\text{st}}(mN) = N_{\min}^{\text{st}}(N)^m \quad \text{so that} \quad N_{\min}^{\text{st}}(N) = \exp(\xi N). \quad (1)$$

A similar argument can be made for the number of transition states, defined as stationary points with Hessian index one (1). If the rearrangement corresponding to a transition state is localised in one subsystem, then a transition state of the mN -atom system occurs when one of the subsystems is at a transition state and the rest are at a minimum, so

$$N_{\text{ts}}^{\text{st}}(mN) = m N_{\min}^{\text{st}}(N)^{m-1} N_{\text{ts}}^{\text{st}}(N) \quad \text{and} \quad N_{\text{ts}}^{\text{st}}(N) = N \exp(\xi N). \quad (2)$$

The ratio $N_{\text{ts}}^{\text{st}}/N_{\min}^{\text{st}}$ is therefore predicted to grow linearly with size, in reasonable agreement with numerical results for small atomic clusters (24), which suggest that ξ is of order one.

The scaling of connectivity with N means that plots of potential energy surfaces in three dimensions are usually only useful for low dimensional functions. For example, the surface in Figure 1 is well populated with local minima, but actually corresponds to a function of only two variables. Locating all the local minima, transition states, and pathways for such a landscape is straightforward.

Disconnectivity graphs provide an alternative approach (25, 26), providing a visualisation that respects the potential or free energy (27, 28) barriers between states. Quantitative results for thermodynamic and kinetic observables are obtained from the underlying kinetic transition network (9–12). However, inspection of the corresponding disconnectivity graph can immediately suggest likely properties, such as heat capacity features and separate relaxation time scales, as discussed below (19, 22, 29–37).

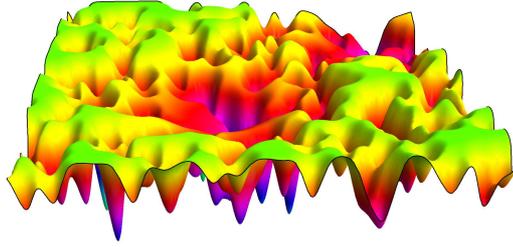


Figure 1

A three-dimensional landscape corresponding to a function of two variables. The direct connectivity between minima is limited by the low dimensionality.

Atomic clusters bound by simple potentials have provided a wealth of insight into such features. The pairwise additive Lennard-Jones (LJ) function (38) is often employed, where the potential energy for N atoms in an LJ_N cluster is

$$V = 4\epsilon \sum_{i < j} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right], \quad (3)$$

where r_{ij} is the distance between atoms i and j . ϵ and $2^{1/6}\sigma$ are the pair equilibrium well depth and separation, respectively, and are used to define reduced units with $\epsilon = \sigma = 1$, and a reduced temperature $k_B T/\epsilon$.

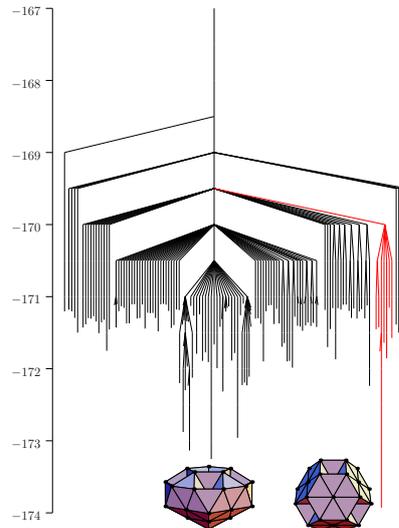


Figure 2

Disconnectivity graph for the LJ_{38} cluster. Structures corresponding to the lowest energy incomplete Mackay icosahedron and the global minimum truncated octahedron are indicated near the corresponding branches. Minima based on fcc packing are coloured red and the vertical axis is the potential energy in ϵ .

A disconnectivity graph for LJ₃₈ is presented in Figure 2. The graph is constructed using connectivity information from a database of local minima and transition states, where the minima are partitioned into disjoint sets (superbasins) at a regular series of energy thresholds. Minima are in the same superbasin if they can interconvert via a pathway where the highest transition state is below the threshold. Energy increases up the vertical axis, and the low energy terminus of each branch corresponds to the energy of a local minimum, with a horizontal displacement chosen to minimise crossings. The branches join at the lowest energy threshold where minima, or sets of minima, merge into the same superbasin.

Disconnectivity graphs have provided new insight into how observable properties are encoded in the landscape. Efficient self-organisation is associated with a single funnel structure, where relaxation to the global minimum can be achieved by downhill pathways associated with relatively low barriers. This is the structure we associate with ‘magic number’ clusters, crystals, and functional biomolecules. However, the landscape for LJ₃₈ in Figure 2 corresponds to a double funnel system (26, 39), with low-lying minima based upon an incomplete Mackay icosahedron (40) and fcc packing, separated by a relatively high barrier. Such multifunnel landscapes can exhibit broken ergodicity, and have been extensively investigated as benchmarks for global optimisation (41), thermodynamic sampling (22, 32, 33, 37, 42–45) and rare event dynamics (14, 15, 46).

The landscapes visualised for structural glass-formers correspond to multi-funnel systems, with many competing amorphous minima separated by high barriers (47–50). Here another level of organisation has been identified by separating different regions of the landscape in terms of cage-breaking processes required for atomic or molecular diffusion. In between the limits defined by efficient self-organisation and glass behaviour we have suggested that intrinsically multifunctional systems will be associated with multifunnel landscapes (51). For example, bistable molecules with the potential to act as switches are expected to have double funnel landscapes (52, 53), and the intrinsically disordered PUMA protein has a number of funnels corresponding to different secondary structure content, which might be modulated by ligand binding (51). A quantitative frustration index, described in §5.2, has been proposed to assess the complexity.

These examples illustrate how the ability to retain all the relevant degrees of freedom can provide detailed insight into the origin of functional behaviour. In particular, the double funnel character of the LJ₃₈ landscape (Figure 2) leads to a heat capacity feature corresponding to a low temperature solid-solid transition between the fcc and icosahedral morphologies, along with two distinct relaxation time scales (54). Projecting a potential or free energy landscape onto a few degrees of freedom can produce a misleading representation if states are lumped together when they are actually separated by significant barriers (10, 55–59). One particular path between the two lowest minima for LJ₃₈, consisting of ten local minima and the nine transition states that connect them, was chosen to examine the effects of projection in defining a reaction coordinate (59). The profile, defined geometrically in terms of stationary points and steepest-descent paths, is shown in Figure 3, where the potential energy is plotted as a function of the integrated path length.

Potentials of mean force were calculated using Monte Carlo sampling as a function of the integrated path length, s , and projection onto the bond order parameter Q_6 (60, 61). Q_6 provides a useful distinction between fcc and icosahedral packing, with smaller values for icosahedra (Figure 4). Configurations were saved along the pathways, and used to define a discretisation of the overall path (59). A Metropolis acceptance condition was employed with the additional constraint that configurations had to remain within a distance

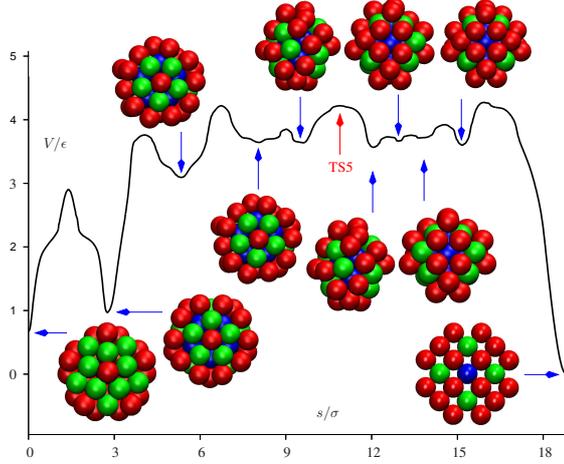


Figure 3

A nine-step pathway linking the two lowest minima of LJ_{38} , with potential energy V as a function of the integrated path length s . The ten local minima are illustrated with the most tightly bound atoms in blue, the least tightly bound in red, and the remainder in green.

d of at least one configuration, \mathbf{X}_i , on the path. Once probabilities for the bins have converged at a given temperature, T , we can compare the Landau free energies calculated as $F(s) = -k_B T \ln P(s)$ and $F(Q_6) = -k_B T \ln P(Q_6)$. A biasing potential was constructed to accelerate convergence, since the closest configuration on the saved path, n , is required, and its potential energy, V_n , is known. Accept/reject steps can then be based on the potential $V(\mathbf{X}) + W(\mathbf{X})$, where $W(\mathbf{X}) = -V_n(\mathbf{X})$, and we accumulate bin averages of $\exp[W(\mathbf{X})/k_B T]$ for each bin. This approach is similar in spirit to umbrella sampling, and enables the expectation value of an observable $O(T)$ to be calculated as (62)

$$\langle O(T) \rangle = \langle O(T) e^{W(\mathbf{X})/k_B T} \rangle_W / \langle e^{W(\mathbf{X})/k_B T} \rangle_W, \quad (4)$$

where the subscript W implies that the biased potential is sampled.

Parallel runs, labelled r , with overlapping bins were used, and the statistics were combined using weighted histogram analysis (63, 64) to obtain the best fit probability distribution $P(s)$ by minimising (65)

$$\chi_{1D}^2 = \sum_r \sum_s \mathcal{N}_{rs} [\ln P(s) - \ln P_{rs} Z_r]^2 \quad \text{where} \quad \mathcal{N}_{rs} = \sum_{j \in s} 1, \quad (5)$$

$$\text{and } P_{rs} = \frac{\mathcal{N}_{rs}}{\mathcal{N}_r} \quad (\text{unbiased}) \quad \text{or} \quad P_{rs} = \frac{\sum_{j \in s} \exp[W(\mathbf{X}_j)/k_B T]}{\sum_j \exp[W(\mathbf{X}_j)/k_B T]} \quad (\text{biased}),$$

where \mathcal{N}_r is the total number of MC steps for parallel run r , and the sums over j correspond to the configurations in the Markov chain for that run. Direct minimisation was employed using analytic derivatives with respect to $\ln P(s)$ and $\ln Z_r$.

For the Q_6 order parameter the statistics provide relative values within each block, and contributions to the same value of Q_6 can arise for different blocks. The required probability

distribution was obtained having solved for $P(s)$ with equation (5), using the average value for Q_6 in each bin; further details are provided in reference (59), including construction of free energy surfaces and convergence checks. Results for $F(s)$ as a function of temperature and constraint distance are shown in Figure 4. The corresponding profiles generally exhibit maxima and minima that correspond directly to the potential energy profile in Figure 3. In contrast, when we project onto Q_6 the minimum around $s = 2.8$ is missing, because this order parameter does not discriminate between the first two low-lying minima with icosahedral packing (Figure 4).

The problem with projecting the pathway onto Q_6 is clear from the two views of the free energy surface $F(s, Q_6)$ in Figure 5. The two minima corresponding to alternative surface packings are separated by a high barrier on the s axis, but overlap in terms of Q_6 . Such barriers, corresponding to orthogonal degrees of freedom, are commonly associated with ‘friction’ when attempts are made to recover the dynamics from a one-dimensional projection. Since the geometrical paths, including all degrees of freedom, are obtained as a byproduct of constructing a kinetic transition network, they could be used to guide reconstructions in lower dimensionality. This approach suggests a number of possible avenues for future research. For example, it should be possible to examine the effect of increasing temperature on the pathway entropy, where an increasing number of discrete paths make significant contributions to the interconversion rate.

3. Recent Developments in Basin-Hopping Global Optimisation

The fastest variant of basin-hopping for treating atomic clusters with multifunnel landscapes reported at the optbench benchmarking site (URL optbench.org) (66) corresponds to symmetrised steps (41). Further benchmarks have been reported and discussed recently (16) for GPU hardware (67) and local rigid body formulations (68). In the present Review the focus is on schemes that extend the sampling using different acceptance conditions. First we summarise the superposition approach, which plays a key role, both in the remainder of this section, and §4 and §5.

3.1. The Superposition Approach

In the superposition approach (17–22) an explicitly ergodic partition function is obtained by summing over contributions from the catchment basins of local minima. Here it is convenient to lump together local minima corresponding to permutation-inversion isomers of a given structure, α , since the corresponding statistical weight is determined purely by the order of the point group, o_α , and the composition, with N_s atoms of element s , etc.:

$$n_\alpha = \left(2 \prod_s N_s! \right) / o_\alpha \equiv P / o_\alpha. \quad (6)$$

The superposition partition function can now be written as a sum over structures:

$$Z(T) = \sum_{\alpha=1}^{N^{\text{st}}} \sum_{\zeta=1}^{n_\alpha} Z_\zeta(T) = \sum_{\alpha=1}^{N^{\text{st}}} n_\alpha Z_\alpha(T) = P \sum_{\alpha=1}^{N^{\text{st}}} Z_\alpha(T) / o_\alpha. \quad (7)$$

This formulation is explicitly ergodic because it reproduces the equilibrium occupation probability for each structure, irrespective of the barriers between the local minima. $Z_\alpha(T)$

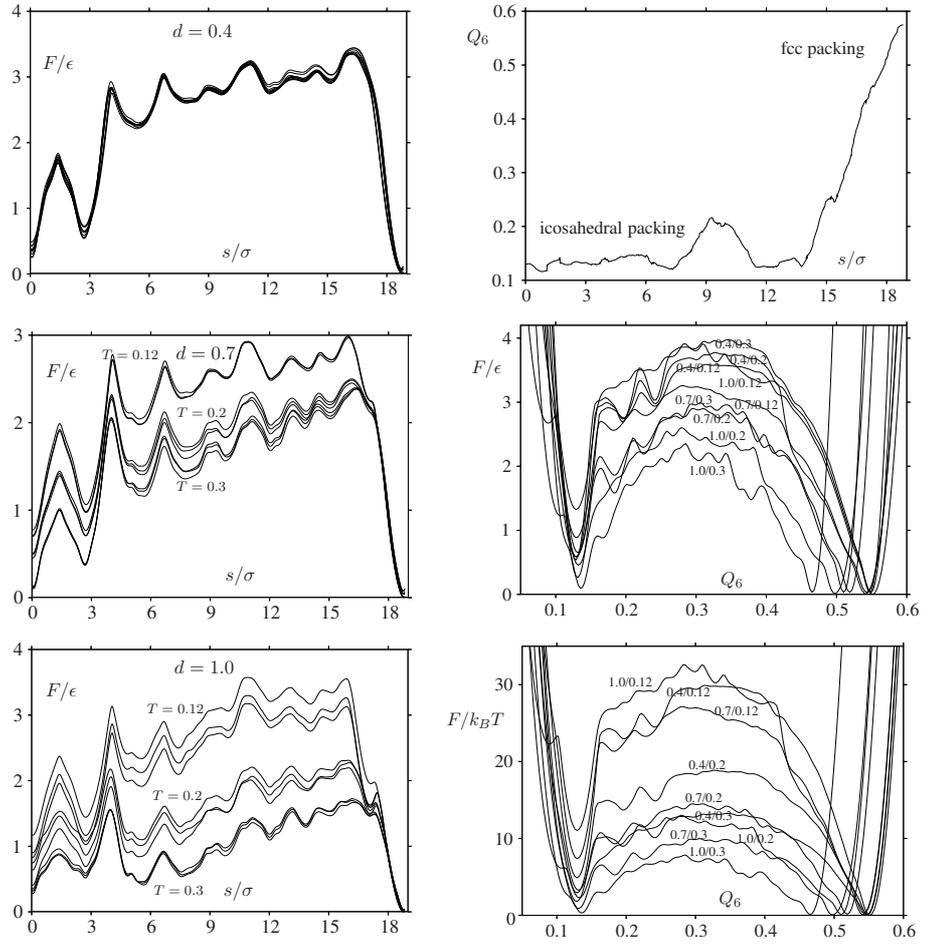


Figure 4

(Left) Free energy F as a function of the integrated path length s for the LJ₃₈ pathway in Figure 3. Results for $d = 0.4, 0.7$ and 1.0 are shown separately, each including temperatures of $k_B T/\epsilon = 0.12, 0.2$ and 0.3 , with three different starting points for each d, T combination to check convergence. For $d = 0.4$ the nine curves are practically coincident. (Right) Top: Q_6 as a function of the configurations characterising an interconversion pathway of LJ₃₈. Middle: free energy F as a function of Q_6 for the same path. The results correspond are for $d = 0.4, 0.7$ and 1.0 and $k_B T/\epsilon = 0.12, 0.2$ and 0.3 . Bottom: as for the middle plot but illustrating $F(Q_6)/k_B T$ for comparison with previous work (43).

is the partition function of structure α at temperature T , which is identical for the catchment basin associated with each of the n_α permutation-inversion isomers.

For molecules that are not translating or rotating it is the internal vibrational degrees of freedom that are of primary interest, and harmonic normal mode analysis gives

$$Z_\alpha(T) = \left(\frac{k_B T}{h \bar{\nu}_\alpha} \right)^\kappa e^{-V_\alpha/k_B T}, \quad (8)$$

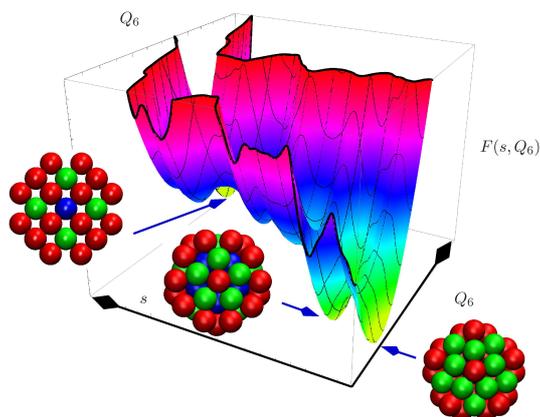


Figure 5

Free energy surface obtained from $-k_B T \ln P(s)P(Q_6|s)$ for constraint distance 0.4 and $k_B T/\epsilon = 0.12$ (59). The positions of three low-lying minima are marked.

where $\kappa = 3N - 6$ is the number of vibrational degrees of freedom, $N = \sum_s N_s$ is the total number of atoms, and $\bar{\nu}_\alpha$ is the geometric mean vibrational frequency for minimum α . The harmonic approximation neglects well anharmonicity, but corrections can be formulated in various ways (18, 69–72).

3.2. Basin-Hopping with Alternative Thermodynamic Potentials

Usually, basin-hopping steps between local minima are accepted or rejected based on a Metropolis-type condition using the potential energy difference $V_{\text{new}} - V_{\text{old}}$ between the new and old local minima with a fictitious temperature, T , chosen to permit a suitable fraction of uphill steps (29, 73, 74). Other choices are certainly possible, including downhill-only moves (75) and Tsallis rather than Boltzmann weights (76). In this Review, we focus on schemes that exploit the harmonic superposition approach described above, to provide alternative ways to survey the energy landscape. The first of these methods is free energy basin-hopping (FEBH) (77), where steps between local potential energy minima are accepted or rejected based on the change in local harmonic free energy,

$$F_{\text{new}}(T) - F_{\text{old}}(T) = V_{\text{new}} - V_{\text{old}} + k_B T \ln (o_{\text{new}} \bar{\nu}_{\text{new}}^\kappa / o_{\text{old}} \bar{\nu}_{\text{old}}^\kappa). \quad (9)$$

Initial benchmarks for peptides showed that the global free energy minimum was located faster using the FEBH scheme than by running conventional basin-hopping and calculating the local free energies for saved minima in a post-processing phase (77).

Steps that change the number or identities of atoms correspond to grand and semi-grand canonical basin-hopping (GCBH and SGCBH) (78). The accept/reject condition considered for GCBH employed the local grand potential:

$$\xi_\alpha^N = V_\alpha^N - \mu N - k_B T \ln \frac{n_\alpha^{(N)} \sqrt{8\pi} |\mathbf{I}_\alpha|^{1/2} (k_B T)^{3/2}}{\hbar^3 (\beta \hbar \bar{\nu}_\alpha^N)^{\kappa(N)}}, \quad (10)$$

which included the rigid rotor partition function, with inertia tensor \mathbf{I}_α . This factor was found to have little effect on the results. Blocks of conventional BH steps were employed

between changes of N , with an acceptance condition based on the lowest value of the grand potential in the new block of local minima compared to the lowest value for the previous block in the chain. The global minimum changes from dissociated states to clusters for larger values of the chemical potential, μ , and lower temperatures, as illustrated in Figure 6. GCBH runs provide a useful overview of low-lying minima as a function of cluster size, and the presence of a least probable size for a range of μ and T suggests an analogue of first order nucleation (78), which requires further investigation.

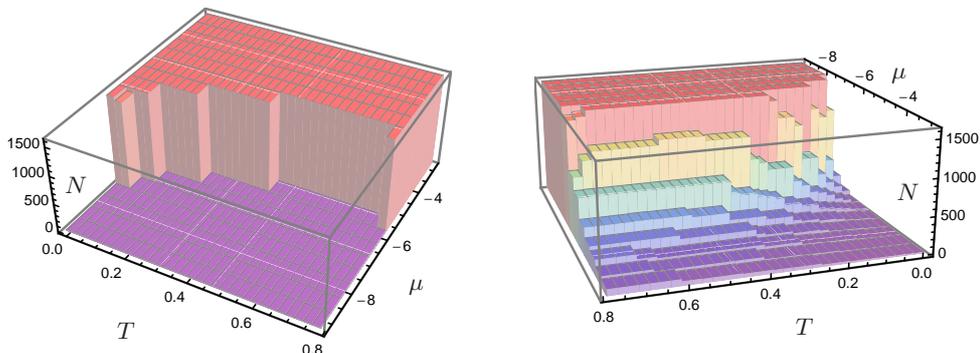


Figure 6

In GCBH runs for atoms interacting via the Lennard-Jones potential the most probable cluster size (left) corresponds to the smallest or largest size permitted. However, the least probable size can occur between the limits (right), analogous to a free energy barrier to nucleation (78).

For a binary semi-grand canonical potential we consider fixed $N = N_A + N_B$, variable N_A and N_B , and an acceptance criterion based on the potential

$$\xi_{\alpha}^{N_B} = V_{\alpha}^{N_B} - \Delta\mu N_B - k_B T \ln \frac{(k_B T)^{3/2} n_{\alpha}^{(N_A, N_B)} \sqrt{8\pi} |\mathbf{I}_{\alpha}|^{1/2}}{\hbar^3 (\beta \hbar \bar{V}_{\alpha}^N)^{\kappa(N)}}, \quad (11)$$

where $\Delta\mu = \mu_B - \mu_A$ is the chemical potential difference and $n_{\alpha}^{(N_A, N_B)} = 2N_A!N_B!/o_{\alpha}$. SGCBH runs for binary nanoalloy clusters highlighted the most favourable compositions as a function of $\Delta\mu$. Results are shown for $\text{Ag}_n\text{Pd}_{55-n}$ in Figure 7, which reveals distinct steps around $N_{\text{Ag}} = 12$ and $N_{\text{Ag}} = 42$, where high symmetry arrangements are possible. The steps are smoothed when wider sampling is used to calculate the probabilities (78).

3.3. Generalised Basin-Hopping

The generalised basin-hopping (GBH) approach was developed in the context of structure prediction for nanoalloy clusters, where the additional problem of chemical ordering (79) introduces an extra combinatorial dimension to the problem. Initial treatments focused on biminima, defined as configurations that are local minima for two different metrics, namely the usual continuous coordinate space, and the neighbourhood defined by swapping the chemical identity of a pair of atoms and reminimising (80, 81). These treatments are specific examples of a generalised formulation in terms of multiminima (82), defined as configurations that are local minima with respect to multiple metrics, which could be continuous, discrete, categorical, etc. For nanoalloy clusters, the minimal number of swaps for atomic species defines a local neighbourhood of isomers related by a single exchange in terms of the Hamming distance (82). Initial tests for nanoalloy clusters indicate that GBH

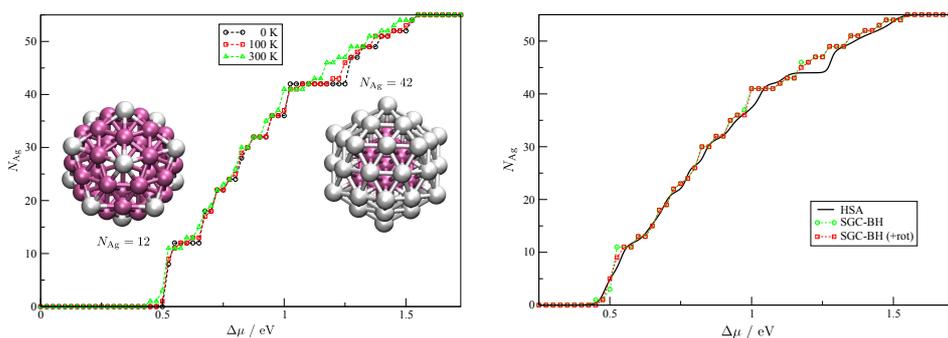


Figure 7

Left: the most favourable composition for icosahedral $\text{Ag}_n\text{Pd}_{55-n}$ as a function of $\Delta\mu$ exhibits steps at $N_{\text{Ag}} = 12$ and $N_{\text{Ag}} = 42$. Right: lumping probabilities for minima with the same composition and including non-icosahedral structures smooths the steps (78).

is most efficient for systems with lower lattice mismatch, where segregation is favoured over mixing, and fewer biminima exist. For $\text{Cu}_{13}\text{Ag}_n\text{Au}_{42-n}$ clusters, a scheme that scanned a subset of the local neighbourhood to check the biminimum property located the likely global minima orders of magnitude faster than in previous work (82).

The GBH procedure is very flexible, and many problems might be formulated in terms of this framework, beyond the chemical ordering challenge presented by homotop isomers in nanoalloys. For example, neighbourhoods can be defined in terms of point mutations for proteins and nucleic acids. New projects that exploit this capability are already in progress.

4. Overcoming Broken Ergodicity Using Basin-Sampling

Equilibrium sampling for systems exhibiting broken ergodicity is a challenging problem for computer simulation, which arises whenever we encounter alternative low-energy states separated by barriers that are high compared to the relevant thermal energy (19). Atomic clusters with double-funnel potential energy landscapes, provide well-defined benchmark examples, exhibiting solid-solid transitions at low temperatures (19, 22, 29–37). Obtaining converged densities of states and the corresponding thermodynamic properties can be difficult, even for enhanced sampling techniques, such as the parallel tempering (PT) Monte Carlo (83–85) multiple histogram (63, 64, 86, 87) approach.

Basin-sampling (BSPT) tackles the sampling problem by combining results from global optimisation and parallel tempering calculations (44). Broken ergodicity is treated using basin-hopping global optimisation §3 (29, 73, 74), while the configuration space corresponding to high temperature is sampled by parallel tempering. Minimisation at regular intervals is used to map configurations onto local minima and obtain a two-dimensional density of states. An anharmonic form is optimised using a multihistogram approach for potential energy bins corresponding to local minima, connecting the results obtained for low and high temperatures. This procedure provides accurate densities of states and thermodynamic properties for benchmark atomic clusters exhibiting broken ergodicity (22, 32, 33, 37, 42–45). It can also be used to calculate the potential energy density of local minima for permutation-inversion isomers and structures, providing new information about the complexity of the energy landscape as a function of system size.

Parallel tempering produces canonical probability distributions $P(V, T) \propto$

$\Omega_c(V)e^{-V/k_B T}$ at the set of temperatures simulated. Here $\Omega_c(V)$ is the configurational density of states; the momentum contribution is analytical and is added separately after the multihistogram analysis. For each replica r with temperature T_r we count the number of visits, \mathcal{N}_{ir} , to potential energy bins indexed as V_i^I , where the superscript I serves to distinguish instantaneous potential energies from quenched potential energies obtained after local minimisation, denoted by Q , and i is the bin label. The statistics for bin visits then provide an estimate for $P(V_i^I, T_r)$ as

$$P(V_i^I, T_r) = \mathcal{N}_{ir}/\mathcal{N}_r, \quad (12)$$

where $\mathcal{N}_r = \sum_i \mathcal{N}_{ir}$ is the total number of Monte Carlo steps for replica r .

To obtain optimal estimates for $\Omega_c(V_i^I)$ and $Z_c(T_r)$ (the configurational part of the canonical partition function) we minimise (44)

$$\chi_{1D}^2 = \sum_r \sum_i \mathcal{N}_{ir} \left[\ln \Omega_c(V_i^I) - \ln \left(\frac{\mathcal{N}_{ir} e^{V_i^I/k_B T_r}}{Z_c(T_r)} \right) \right]^2, \quad (13)$$

where the variables are $\Omega_c(V_i^I)$ and $Z_c(T_r)$. Once the best fit values $\Omega_c^*(V_i^I)$ and $Z_c^*(T_r)$ are known we can calculate $Z_c(T)$ for any temperature as $Z_c(T) = \sum_i \Omega_c^*(V_i^I) \exp(-V_i^I/k_B T)$.

At specified intervals the current configuration in each replica is quenched to a local minimum without changing the coordinates in the Markov chain. The number of visits to quench potential energy bin q from instantaneous potential energy bin i in replica r is \mathcal{N}_{iqr} : a two-dimensional histogram. The corresponding canonical probability distribution is

$$\bar{P}(V_i^I, V_q^Q, T_r) = \mathcal{N}_{iqr}/\mathcal{N}_r \propto \bar{\Omega}_c(V_i^I, V_q^Q) e^{-V_i^I/k_B T_r}. \quad (14)$$

We can now consider a two-dimensional best fit by minimising the statistic

$$\chi_{2D}^2 = \sum_r \sum_i \sum_q \mathcal{N}_{iqr} \left[\ln \bar{\Omega}_c(V_i^I, V_q^Q) - \ln \left(\frac{\mathcal{N}_{iqr} e^{V_i^I/k_B T_r}}{Z_c(T_r)} \right) \right]^2, \quad (15)$$

where the variables are the $\bar{\Omega}_c(V_i^I, V_q^Q)$ if we set $Z_c(T_r) = Z_c^*(T_r)$. The analytic density of states for a Morse potential suggests a model anharmonic function with two fitting parameters for each quench bin, A_q and B_q :

$$\ln \bar{\Omega}_c(V_i^I, V_q^Q) = (\kappa' + e^{A_q} V_{iq}) \ln V_{iq} + B_q, \quad (16)$$

where $\kappa' = \kappa/2 - 1$ and $V_{iq} = V_i^I - V_q^Q$.

In the limit $V_{iq} \rightarrow 0$ the dependence on V_{iq} reduces to the harmonic form: $\bar{\Omega}_c(V_i^I, V_q^Q) \propto V_{iq}^{\kappa/2-1}$. The contribution to $\bar{\Omega}_c(V_i^I, V_q^Q)$ from a given structure should also include the weight $n_\alpha = P/o_\alpha$, and the vibrational factor $1/\bar{v}_\alpha^\kappa$. Hence, if quench bin q contains a single minimum, α , we should recover a prefactor of $\ln n_\alpha/\bar{v}_\alpha^\kappa + c = -\ln o_\alpha \bar{v}_\alpha^\kappa + c'$, where the constants c and c' are independent of the structure considered and the energy bin. For each q fits were performed for different ranges of i to derive optimal parameters A_q^* and B_q^* .

For low-energy q bins all the minima will generally be sampled during a basin-hopping global optimisation run, and we can therefore calculate the corresponding prefactors directly from $\sum_{\alpha \in q} (o_\alpha \bar{v}_\alpha^\kappa)^{-1}$. Fitted B_q^* values are replaced up to a specified potential energy threshold, retaining previously fitted A_q^* parameters to account for anharmonicity (44). A

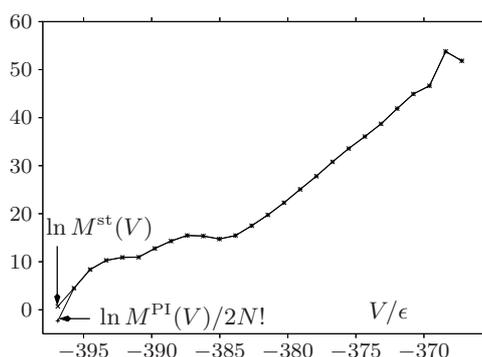


Figure 8

Calculated values for $\ln M^{\text{st}}(V)$ and $\ln M^{\text{st}}(V)/P$ as a function of potential energy for LJ₇₅ (44). The potential energy bin width was 1.18731ϵ .

hybrid configurational density of states was obtained by combining the results of the one- and two-dimensional fitting procedures (44).

LJ₃₁ and LJ₇₅ both exhibit broken ergodicity and a low temperature solid-solid phase transition due to competition between the global minimum and alternative structures. For LJ₃₁ accurate heat capacity curves were obtained using 10^6 equilibration steps (discarded), 5×10^6 PT, and 5×10^6 BSPT steps (abbreviated as 1/5/5 in terms of millions of steps) using 24 replicas spread geometrically through the temperature range 0.1 to 0.5 (Figure 9). The container radius was set to 2.5 reduced units and quenches in the BSPT phase were performed every 30 steps. This run required 21.8 minutes of wall clock time on Intel Xeon E5405 cpus (2.00 GHz). All the replicas were started from the global minimum, and only the lowest four minima were used to replace fitted B_q^* values. A BSPT simulation starting from random configurations takes more than twice as long because PT is not an efficient global optimisation procedure. For comparison, PT results for 24 replicas running at geometrically distributed temperatures in the range 0.0125 to 0.6 required about 110.5 hours of wall clock time on the same hardware for 10^{10} steps. Previous work suggests that this value should be sufficient for convergence (37).

Obtaining converged equilibrium thermodynamic properties for LJ₇₅ represents a more challenging test. Previous work indicates that even 10^{11} Monte Carlo steps are not enough to converge the heat capacity using adaptive exchange parallel tempering (22, 42), and 3×10^9 steps were needed when an auxiliary harmonic superposition reference was employed (37). Results for a BSPT run of 10/200/60 steps and 32 replicas exponentially spaced in the temperature range 0.15 to 0.375 and a container radius of 3.0 are shown in Figure 9. Quenches were performed every 30 steps in the BSPT phase, and the run required 22 hours of wall clock time. These results were obtained when the lowest 13 minima were used to replace the fitted B_q^* values (44).

The potential energy density of minima can also be obtained from the basin-sampling approach, providing an alternative to schemes that compute average catchment basin volumes using thermodynamic integration (88, 89). The mean vibrational contribution to the prefactor in $\bar{\Omega}_c(V_i^I, V_q^Q)$ is

$$\langle 1/\bar{\nu}^k \rangle_q = \sum_{\alpha \in q} (o_\alpha \bar{\nu}_\alpha^k)^{-1} / \sum_{\alpha \in q} o_\alpha^{-1}. \quad (17)$$

For a complete sample, dividing $\sum_{\alpha \in q} (o_\alpha \bar{v}_\alpha^\kappa)^{-1}$ by $\langle 1/\bar{v}^\kappa \rangle_q$ would give $\sum_{\alpha \in q} o_\alpha^{-1}$, which is the number of distinct permutation-inversion isomers in bin q divided by P . The fitted prefactors $\exp(B_q^*)$ should be proportional to $\sum_{\alpha \in q} (o_\alpha \bar{v}_\alpha^\kappa)^{-1}$ so that $\exp(B_q^*) \propto \langle 1/\bar{v}^\kappa \rangle_q \sum_{\alpha \in q} o_\alpha^{-1}$. Calculating $\langle 1/\bar{v}^\kappa \rangle_q$ from (17) using a sample of quench minima provides an estimate for $M^{\text{PI}}(V_q^Q)$, the number of permutation-inversion isomers in bin q , from $\exp(B_q^*) / \langle 1/\bar{v}^\kappa \rangle_q$, up to a constant factor.

Assuming we have a complete sample of minima in the lowest bin, denoted 0, the number of distinct permutation-inversion isomers in bin q is

$$\frac{M^{\text{PI}}(V_q^Q)}{M^{\text{PI}}(V_0^Q)} = \frac{e^{B_q^*} \langle 1/\bar{v}^\kappa \rangle_0}{e^{B_0^*} \langle 1/\bar{v}^\kappa \rangle_q} \quad \text{with} \quad M^{\text{PI}}(V_0^Q) = P \sum_{\alpha \in 0} o_\alpha^{-1},$$

so that

$$M^{\text{PI}}(V_q^Q) = P \frac{e^{B_q^*} \langle 1/\bar{v}^\kappa \rangle_0}{e^{B_0^*} \langle 1/\bar{v}^\kappa \rangle_q} \sum_{\alpha \in 0} o_\alpha^{-1}. \quad (18)$$

A similar analysis provides the number of structures (44):

$$M^{\text{st}}(V_q^Q) = \frac{e^{B_q^*} \langle 1/o \bar{v}^\kappa \rangle_0}{e^{B_0^*} \langle 1/o \bar{v}^\kappa \rangle_q} \sum_{\alpha \in 0} 1. \quad (19)$$

In the absence of point group symmetry the distributions $M^{\text{PI}}(V)$ and $M^{\text{st}}(V)$ are the same, aside from the factor of P .

The probability distributions calculated for LJ₇₅ are shown in Figure 8 (44). The presence of a shallow local minimum in the distributions around $V = -385\epsilon$ appears to be real. The more sparsely populated regions of potential energy may result from the superposition of distributions that correspond to different morphologies (44).

5. Insight from Occupation Probabilities

5.1. Assignment of Heat Capacity Features

The superposition approach and basin-sampling have recently been used to decompose the contributions of local minima to the heat capacity (90). The theory provides an assignment that does not require any structural information, or specification of an order parameter. Hence it provides an unbiased scheme, which can be used for the landscapes defined by any continuous function with multiple minima, such as neural network fits §6.

Two examples will be described here for LJ₃₁ and LJ₇₅. The low temperature heat capacity peaks in Figure 9 are reproduced by the harmonic superposition approach, while the melting transition requires basin-sampling (44). The analysis of C_V is most straightforward for the harmonic approximation, where the heat capacity can be written as

$$C_V = \left(\frac{\partial E}{\partial T} \right)_{N,V} = \kappa k_B - \frac{z_1(T)^2}{k_B T^2 z_0(T)^2} + \frac{z_2(T)}{k_B T^2 z_0(T)}, \quad (20)$$

$$\text{with} \quad z_r(T) = \sum_{\gamma} n_{\gamma} (V_{\gamma})^r \left(\frac{k_B T}{h \bar{v}_{\gamma}} \right)^{\kappa} e^{-V_{\gamma}/k_B T}. \quad (21)$$

The configurational part of C_V is obtained by subtracting $\kappa k_B/2$. The occupation probability for minimum α in the same approximation is

$$p_{\alpha}(T) = n_{\alpha} e^{-V_{\alpha}/k_B T} / \bar{v}_{\alpha}^{\kappa} \left/ \sum_{\gamma} n_{\gamma} e^{-V_{\gamma}/k_B T} / \bar{v}_{\gamma}^{\kappa} \right., \quad (22)$$

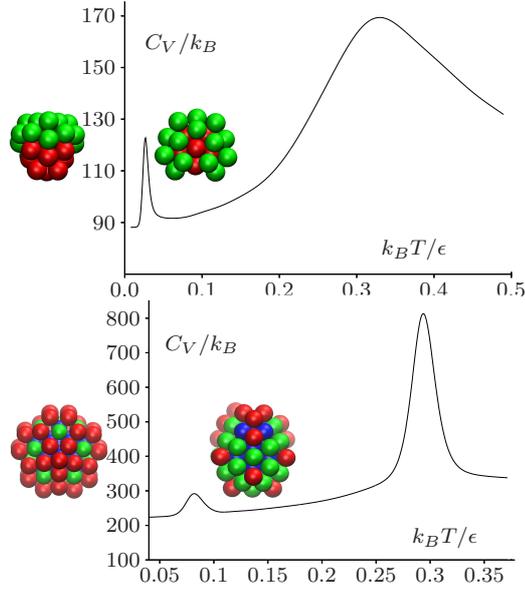


Figure 9

C_V calculated using basin-sampling (44) for LJ₃₁ (top) and LJ₇₅ (bottom).

$$\text{so that } z_r(T) = Z(T) \sum_{\gamma} p_{\gamma}(T) (V_{\gamma})^r = z_0(T) \sum_{\gamma} p_{\gamma}(T) (V_{\gamma})^r, \quad (23)$$

$$\text{and } \frac{\partial p_{\alpha}(T)}{\partial T} = -p_{\alpha}(T) \frac{\partial \ln Z(T)}{\partial T} + \frac{\kappa p_{\alpha}(T)}{T} + \frac{p_{\alpha}(T) V_{\alpha}}{k_B T^2} \equiv g_{\alpha}(T), \quad (24)$$

defining $g_{\alpha}(T)$, which can be written as

$$g_{\alpha}(T) = \frac{p_{\alpha}}{k_B T^2} \left(V_{\alpha} - \sum_{\gamma} p_{\gamma}(T) V_{\gamma} \right) \equiv \frac{p_{\alpha}}{k_B T^2} (V_{\alpha} - \langle V \rangle_{\min}). \quad (25)$$

Using equations (20) and (23) we have

$$C_V = \kappa k_B - \frac{\left(\sum_{\gamma} p_{\gamma}(T) V_{\gamma} \right)^2}{k_B T^2 \left(\sum_{\gamma} p_{\gamma}(T) \right)^2} + \frac{\sum_{\gamma} p_{\gamma}(T) (V_{\gamma})^2}{k_B T^2 \sum_{\gamma} p_{\gamma}(T)} = \kappa k_B + \frac{\langle V^2 \rangle_{\min} - \langle V \rangle_{\min}^2}{k_B T^2}. \quad (26)$$

$$\text{Since } \sum_{\gamma} g_{\gamma}(T) (V_{\gamma} - \langle V \rangle_{\min}) = \sum_{\gamma} \frac{p_{\gamma}}{k_B T^2} (V_{\gamma} - \langle V \rangle_{\min})^2 = \frac{\langle V^2 \rangle_{\min} - \langle V \rangle_{\min}^2}{k_B T^2},$$

$$\text{we have } C_V - \kappa k_B = \sum_{\gamma} g_{\gamma}(T) (V_{\gamma} - \langle V \rangle_{\min}) = k_B T^2 \sum_{\gamma} g_{\gamma}(T)^2 \frac{1}{p_{\gamma}(T)}. \quad (27)$$

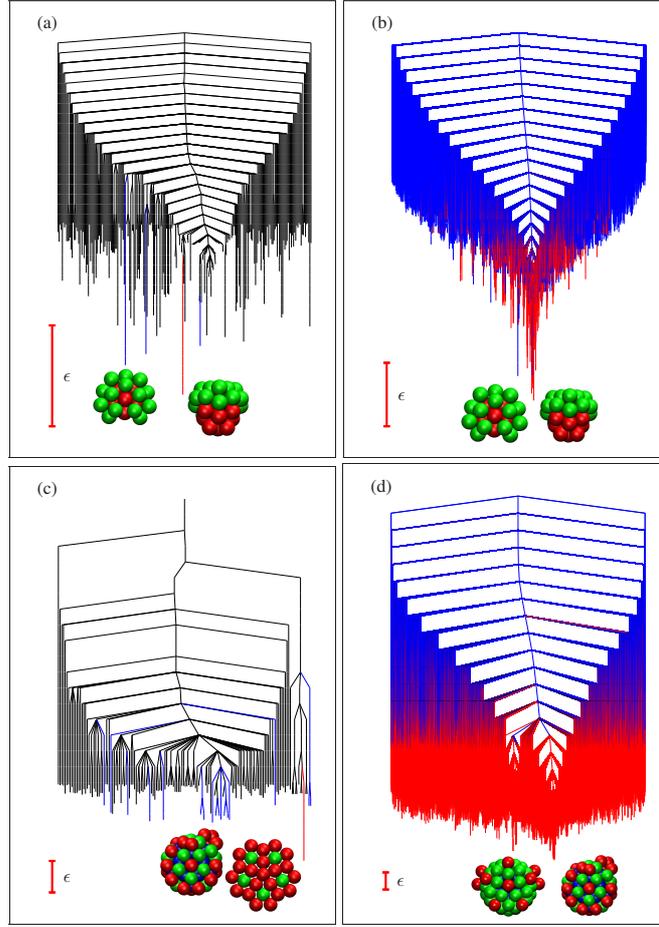


Figure 10

Disconnectivity graphs coloured according to the overall occupation probability gradients for the quench energy bin associated with each local minimum [equation (33)]. Minima are coloured for contributions to the heat capacity that account for 99% of the contributions from positive (blue) and negative (red) gradients (90). (a) LJ₃₁ at $k_B T/\epsilon = 0.0268$, (b) LJ₃₁ at $k_B T/\epsilon = 0.329$, (c) LJ₇₅ at $k_B T/\epsilon = 0.082$, (d) LJ₇₅ at $k_B T/\epsilon = 0.291$. The lowest two minima are illustrated for LJ₃₁ in panels (a) and (b). For LJ₇₅ the two lowest minima are illustrated in panel (c). The graphs in panels (b) and (d) exclude all minima that do not contribute to components of the heat capacity peak within the 99% threshold, and the global minimum does not appear for LJ₇₅ in (d). Instead the lowest minimum based on icosahedral packing is illustrated along with the second-lowest minimum, both of which have a negative occupation probability gradient.

Hence we can separate contributions to C_V from local minima with positive and negative occupation probability gradients:

$$C_V = \kappa k_B + \sum_{\gamma}^{g_{\gamma}(T) > 0} g_{\gamma}(T) (V_{\gamma} - \langle V \rangle_{\min}) + \sum_{\gamma}^{g_{\gamma}(T) < 0} g_{\gamma}(T) (V_{\gamma} - \langle V \rangle_{\min})$$

$$\equiv \kappa k_B + C^+(T) + C^-(T). \quad (28)$$

Ranking the minima according to the fraction of $C^\pm(T)$ they contribute provides an unbiased way to assign a heat capacity peak to features of the potential energy surface.

To describe the melting peaks in C_V accurately we need to account for anharmonicity. The occupation probabilities of the potential energy bins, $p_i(T)$, can be employed, using the weights obtained by basin-sampling (§4) for W_i^{1D} and W_{iq}^{2D} to write

$$p_i(T) = \frac{W_i^{1D} e^{-V_i^I/k_B T}}{\sum_j^{\text{I bins}} W_j^{1D} e^{-V_j^I/k_B T}} = \frac{e^{-V_i^I/k_B T} \sum_q^{\text{Q bins}} W_{iq}^{2D}}{\sum_j^{\text{I bins}} e^{-V_j^I/k_B T} \sum_r^{\text{Q bins}} W_{jr}^{2D}}. \quad (29)$$

The temperature derivative of $p_i(T)$ is

$$g_i(T) \equiv \frac{\partial p_i(T)}{\partial T} = \frac{p_i}{k_B T^2} (V_i^I - \langle V^I \rangle_{\text{I bins}}), \quad (30)$$

where $\langle V^I \rangle_{\text{I bins}} = \sum_i^{\text{I bins}} p_i(T) V_i^I$, and adding $\kappa k_B/2$ to the configurational heat capacity provides expressions for $C_V(T)$ (90) analogous to equation (27):

$$C_V = \frac{\kappa k_B}{2} + k_B - \frac{\Delta^2 e^{\Delta/k_B T}}{k_B T^2 (1 - e^{\Delta/k_B T})^2} + \sum_i^{\text{I bins}} g_i(T) (V_i^I - \langle V^I \rangle_{\text{I bins}}). \quad (31)$$

C_V can be written in terms of contributions from potential energy minima by reordering sums over i and q :

$$\begin{aligned} \sum_i^{\text{I bins}} g_i(T) (V_i^I - \langle V^I \rangle_{\text{I bins}}) &= \sum_i^{\text{I bins}} \frac{p_i}{k_B T^2} (V_i^I - \langle V^I \rangle_{\text{I bins}})^2 = \\ & \sum_q^{\text{Q bins}} \left[\sum_{g_i(T) > 0}^{\text{I bins}} \frac{(V_i^I - \langle V^I \rangle_{\text{I bins}})^2 W_{jq}^{2D} e^{-V_i^I/k_B T}}{k_B T^2 \sum_m^{\text{I bins}} W_m^{1D} e^{-V_m^I/k_B T}} + \sum_{g_i(T) < 0}^{\text{I bins}} \frac{(V_i^I - \langle V^I \rangle_{\text{I bins}})^2 W_{jq}^{2D} e^{-V_i^I/k_B T}}{k_B T^2 \sum_m^{\text{I bins}} W_m^{1D} e^{-V_m^I/k_B T}} \right] \\ &\equiv \sum_q^{\text{Q bins}} [c_q^+ + c_q^-]. \end{aligned} \quad (32)$$

The net contribution from quench bin q is assigned as positive or negative according to the largest of c_q^\pm (90).

Figure 10 illustrates disconnectivity graphs for LJ₃₁ and LJ₇₅ with branches coloured red and blue for the minima associated with 99% of the total contributions to C_V at the temperatures corresponding to the solid-solid and melting peaks. At low temperature, only the global minimum has a significant negative occupation probability temperature gradient in each case. However, for the melting peak the transition is widely delocalised over many local minima. These assignments agree with previous intuition, but require no knowledge of the underlying atomic structure.

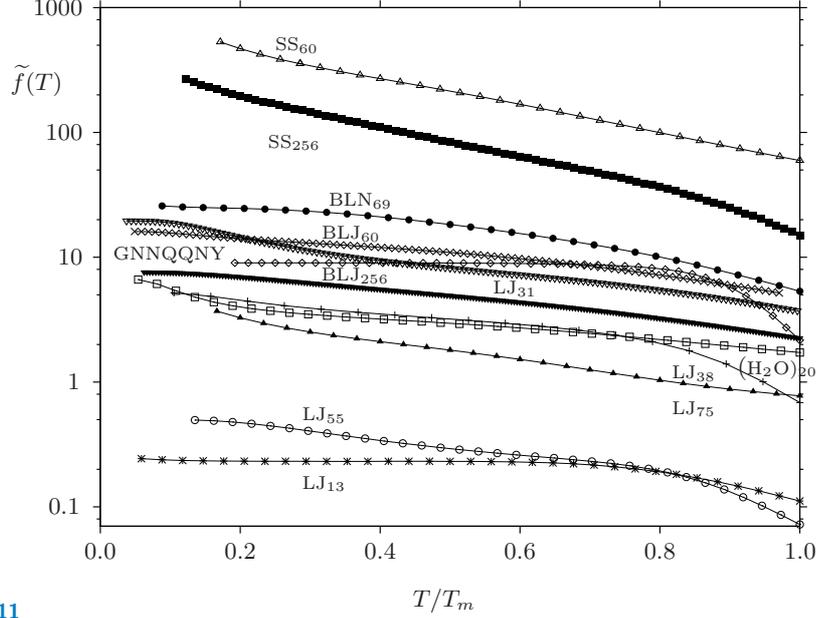


Figure 11

The frustration index calculated from renormalised probabilities, $\tilde{f}(T)$, for systems ranging from magic number clusters to jammed packings of soft spheres, as described in the text (50).

5.2. Quantifying the Complexity of a Landscape

Occupation probabilities of local minima have also been used to quantify the degree of frustration in an energy landscape by defining (50)

$$f(T) = \sum_{\alpha \neq \text{gmin}} p_{\alpha}^{\text{eq}}(T) \left(\frac{V_{\alpha}^{\dagger} - V_{\text{gmin}}}{V_{\alpha} - V_{\text{gmin}}} \right). \quad (33)$$

gmin labels the lowest minimum in the database, and V_{α}^{\dagger} is the potential energy of the highest transition state on the lowest energy path between minimum α and gmin . V_{gmin} corresponds to the global minimum for most of the databases considered, except for the condensed matter systems, where the crystalline region of configuration space was intentionally excluded. The formulation of $f(T)$ was based on schemes introduced to suggest pairs of minima for additional connection attempts in refining kinetic transition networks (91). Figure 11 illustrates results for $\tilde{f}(T)$ based on harmonic approximations to the renormalised occupation probabilities $\tilde{p}_{\alpha}^{\text{eq}} = p_{\alpha}^{\text{eq}} / (1 - p_{\text{gmin}}^{\text{eq}})$ obtained by removing the temperature dependence of the global minimum. A reduced temperature is defined using T_m at the melting peak in C_V obtained with the same database in the harmonic approximation (50). This representation enables us to compare ‘magic number’ clusters, which relax efficiently to a well defined global minimum (LJ₁₃ and LJ₅₅) on the same scale as structural glass formers (BLJ₆₀ and BLJ₂₅₆) and jammed packings of soft repulsive spheres (SS₆₀ and SS₂₅₆). In between these limits we see atomic clusters with double funnel landscapes (LJ₃₁, LJ₃₈ and LJ₇₅), a water cluster (H₂O)₂₀, the amyloidogenic GNNQQNY peptide dimer, and a coarse-grained model of a 69-residue β -barrel protein, BLN₆₉ (50).

$\tilde{f}(T)$ can provide a useful diagnostic of landscapes with qualitatively different organisa-

tion, in terms of competing low energy structures. The ability to quantify our intuition of frustration across such a wide range of systems in a single index may provide new insight into how observable properties are encoded in the energy landscape.

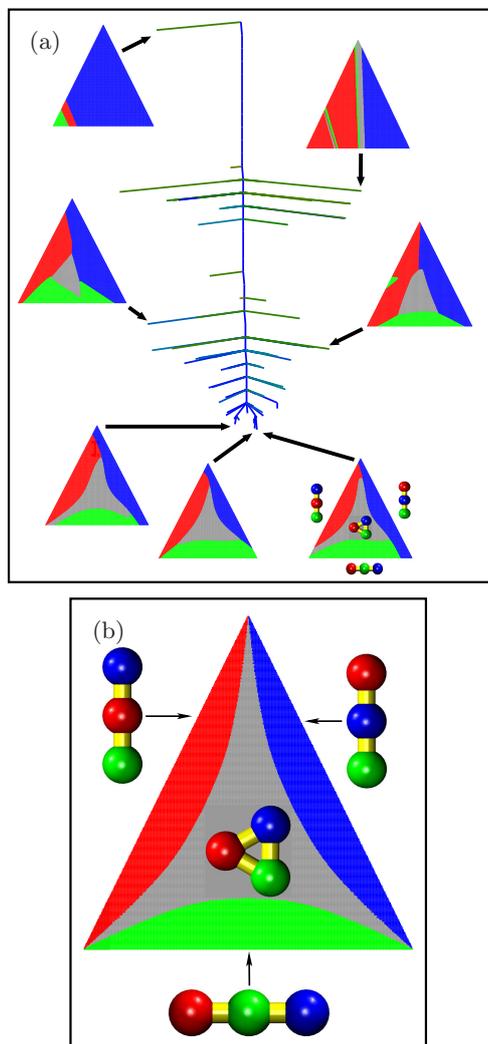


Figure 12

(a) Disconnectivity graph defined by a neural network fit to predict which of four local minima will be reached on geometry optimisation. (b) For a grid of starting configurations, the outcome of local minimisation is indicated by the gray, red, green and blue basins in the large triangle. The central gray region corresponds to the equilateral triangle, and the coloured regions correspond to the linear minima defined by the the central atom. For the triangles associated with seven selected minima in panel (a), the same configurations are used to colour the pixels on the grid according to the minimum with the highest predicted probability. The correspondence with the reference in (b) is very good for the low-lying minima, but degrades systematically for higher-lying solutions.

6. Energy Landscapes for Machine Learning

Part of the motivation for the theory developed in the previous section was provided by intriguing properties discovered for the landscapes defined by neural networks in the context of machine learning (ML) (92, 93). Here the network parameters take the place of atomic coordinates, and local minima of the ML landscape correspond to alternative fits to training data for a cost function that quantifies how well the network reproduces the known outcomes. The ML landscape can be explored using the same methods that have been developed for molecular potential energy surfaces, and analogues for thermodynamic properties and rate constants can be defined using second derivatives and harmonic densities of states. Understanding the organisation of ML landscapes from this viewpoint (92–94) may suggest ways to design better predictive tools.

The example considered here is a classification problem to predict which isomer will result from local minimisation, given only the coordinates of the starting configuration (92, 95). Here we must distinguish two landscapes: the first corresponds to the molecular potential energy surface, which supports four distinct minima for the system in question, and the second is the ML landscape of a neural network (NN) where the cost function corresponds to the accuracy of the predicted outcome. Reliable predictions would be useful for applications that involve systematic local minimisation to assign densities of states to particular isomers or regions of configuration space, as for basin-sampling (44), described in §4, and thermodynamic analysis of granular packings (88).

The triatomic cluster bound by pairwise Lennard-Jones (38) and three-body Axilrod–Teller interactions, corresponding to instantaneous induced dipole-induced dipole terms, exhibits one equilateral triangle minimum and three linear minima, which are distinguished by the identity of the middle atom (96–98). In this example, the aim is to predict which minimum will be obtained by geometry optimisation given three initial interparticle distances (92, 95). Neural networks were trained on a database of outcomes for random starting configurations, and the ML landscapes defined by this problem were explored using the geometry optimisation tools referred to in the Introduction §1.

Full details of the cost function and NN fitting can be found elsewhere (88). Here we focus on how the prediction quality can be visualised for one particular ML landscape. The disconnectivity graph in Figure 12 corresponds to a NN with an input layer of three nodes (for the three interparticle distances, r_{12} , r_{13} , and r_{23}), an output layer with four nodes (for the four local minima), and a hidden layer between them, of three nodes. A pictorial representation of how local minima on the ML landscape perform is provided by considering a two-dimensional triangular grid, where each pixel maps on to a starting configuration in the plane $r_{12} + r_{13} + r_{23} = 3r_e$, where r_e is the pair equilibrium bond length, $2^{1/6}\sigma$. In Figure 12b the pixels are coloured according to the outcome of local minimisation from the corresponding geometry. The red, green and blue regions correspond to basins of attraction for the three linear minima, and the gray region corresponds to runs that converge to the equilateral triangle. For selected minima in the disconnectivity graph in Figure 12a, the pixels are coloured according to the minimum with the highest predicted probability for the same grid of starting configurations. The low-lying minima on the ML landscape generally exhibit a close correspondence with the reference in Figure 12b; each of these solutions could be used with some confidence to predict the outcome of local minimisation given the starting geometry. However, the ML minima with larger residual cost functions show systematically poorer performance.

7. Conclusions

This Review has summarised some recent advances in the potential energy landscapes approach, and showed how this perspective can be extended to gain new insight into the solution space of local minima defined by abstract functions that arise in applications such as neural network fitting. Formulating observable properties in terms of local minima and the transition states that connect them on the underlying potential energy surface provides a powerful and flexible tool for developing both theory and computational methods. Calculating thermodynamic properties and kinetics within this framework, in combination with statistical mechanics and unimolecular rate theory, enables us to treat broken ergodicity and rare event dynamics from a viewpoint that is largely complementary to more conventional techniques. Visualising the corresponding landscapes and pathways can produce insight for designing self-assembly or multifunctional behaviour, along with the mechanisms corresponding to reaction coordinates that respect all the barriers in the full configurational space.

The principal themes developed here show how the superposition approach can be used to calculate accurate densities of states, and to analyse the resulting heat capacity features as emergent properties of the underlying landscape. Occupation probabilities and their temperature derivatives enable us to assign contributions to specific local minima or regions of configuration space, and can also be used to design a quantitative measure of complexity that places disparate systems on a common scale. Basin-hopping global optimisation can be adapted for searches based on local free energy, or grand potentials that enable the system size and composition to vary. Employing multiminima for an arbitrary combination of metric spaces enables us to define neighbourhoods, which can be searched efficiently to predict structure for alchemical or mutational transitions. The capability to transfer this insight to landscapes beyond the domain of atomistic systems suggests exciting possibilities for applications in different fields, such as machine learning.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am very grateful to all the group members and collaborators who have contributed to this research, and to the EPSRC and ERC for financial support.

LITERATURE CITED

1. Murrell JN, Laidler KJ. 1968. Symmetries of activated complexes. *Trans. Faraday. Soc.* 64:371–377
2. Trygubenko SA, Wales DJ. 2004. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* 120:2082–2094
3. Henkelman G, Jónsson H. 2000. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* 113:9978–9985
4. Henkelman G, Uberuaga BP, Jónsson H. 2000. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* 113:9901–9904
5. Munro LJ, Wales DJ. 1999. Defect migration in crystalline silicon. *Phys. Rev. B* 59:3969–3980

6. Henkelman G, Jónsson H. 1999. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* 111:7010–7022
7. Kumeda Y, Munro LJ, Wales DJ. 2001. Transition states and rearrangement mechanisms from hybrid eigenvector-following and density functional theory. application to $C_{10}H_{10}$ and defect migration in crystalline silicon. *Chem. Phys. Lett.* 341:185–194
8. Zeng Y, Xiao P, Henkelman G. 2014. Unification of algorithms for minimum mode optimization. *J. Chem. Phys.* 140:044115
9. Rao F, Caffisch A. 2004. The protein folding network. *J. Mol. Biol.* 342:299–306
10. Noé F, Fischer S. 2008. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Op. Struct. Biol.* 18:154–162
11. Prada-Gracia D, Gómez-Gardenes J, Echenique P, Fernando F. 2009. Exploring the free energy landscape: From dynamics to networks and back. *PLoS Comput. Biol.* 5:e1000415
12. Wales DJ. 2010. Energy landscapes: Some new horizons. *Curr. Op. Struct. Biol.* 20:3–10
13. Forst W. 1973. Theory of unimolecular reactions. New York: Academic Press
14. Wales DJ. 2002. Discrete path sampling. *Mol. Phys.* 100:3285–3305
15. Wales DJ. 2004. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* 102:891–908
16. Joseph JA, Röder K, Chakraborty D, Mantell RG, Wales DJ. 2017. Exploring biomolecular energy landscapes. *Chem. Commun.* in press
17. Stillinger FH, Weber TA. 1984. Packing structures and transitions in liquids and solids. *Science* 225:983–989
18. Wales DJ. 1993. Coexistence in small inert-gas clusters. *Mol. Phys.* 78:151–171
19. Wales DJ. 2003. Energy landscapes. Cambridge: Cambridge University Press
20. Stillinger FH. 1995. A topographic view of supercooled liquids and glass formation. *Science* 267:1935–1939
21. Strodel B, Wales DJ. 2008. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.* 466:105–115
22. Sharapov VA, Meluzzi D, Mandelshtam VA. 2007. Low-temperature structural transitions: Circumventing the broken-ergodicity problem. *Phys. Rev. Lett.* 98:105701
23. Stillinger FH, Weber TA. 1982. Hidden structure in liquids. *Phys. Rev. A* 25:978–989
24. Doye JPK, Wales DJ. 2002. Saddle points and dynamics of Lennard-Jones clusters, solids, and supercooled liquids. *J. Chem. Phys.* 116:3777–3788
25. Becker OM, Karplus M. 1997. The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys.* 106:1495–1517
26. Wales DJ, Miller MA, Walsh TR. 1998. Archetypal energy landscapes. *Nature* 394:758–760
27. Krivov SV, Karplus M. 2002. Free energy disconnectivity graphs: Application to peptide models. *J. Chem. Phys.* 117:10894–10903
28. Evans DA, Wales DJ. 2003. Free energy landscapes of model peptides and proteins. *J. Chem. Phys.* 118:3891–3897
29. Wales DJ, Doye JPK. 1997. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* 101:5111–5116
30. Doye JPK, Wales DJ, Miller MA. 1998. Thermodynamics and the global optimization of Lennard-Jones clusters. *J. Chem. Phys.* 109:8143–8153
31. Doye JPK, Miller MA, Wales DJ. 1999. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* 110:6896–6906
32. Calvo F, Neirotti JP, Freeman DL, Doll JD. 2000. Phase changes in 38 atom Lennard-Jones clusters. ii: A parallel tempering study of equilibrium and dynamic properties in the molecular dynamics and microcanonical ensembles. *J. Chem. Phys.* 112:10350–10357
33. Neirotti JP, Calvo F, Freeman DL, Doll JD. 2000. Phase changes in 38 atom Lennard-Jones clusters. i: A parallel tempering study in the canonical ensemble. *J. Chem. Phys.* 112:10340–10349

34. Frantsuzov PA, Mandelshtam VA. 2005. Size-temperature phase diagram for small Lennard-Jones clusters. *Phys. Rev. E* 72:037102
35. Predescu C, Frantsuzov PA, Mandelshtam VA. 2005. Thermodynamics and equilibrium structure of Ne₃₈ cluster: Quantum mechanics versus classical. *J. Chem. Phys.* 122:154305
36. Liu HB, Jordan KD. 2005. On the convergence of parallel tempering Monte Carlo simulations of LJ₃₈. *J. Phys. Chem. B* 109:5203–5207
37. Sharapov VA, Mandelshtam VA. 2007. Solid-solid structural transformations in Lennard-Jones clusters: Accurate simulations versus the harmonic superposition approximation. *J. Phys. Chem. A* 111:10284–10291
38. Jones JE, Ingham AE. 1925. On the calculation of certain crystal potential constants, and on the cubic crystal of least potential energy. *Proc. R. Soc. A* 107:636–653
39. Doye JPK, Miller MA, Wales DJ. 1999. Evolution of the potential energy surface with size for Lennard-Jones clusters. *J. Chem. Phys.* 111:8417–8428
40. Mackay AL. 1962. A dense non-crystallographic packing of equal spheres. *Acta Cryst.* 15:916–918
41. Oakley MT, Johnston RL, Wales DJ. 2013. Symmetrisation schemes for global optimisation of atomic clusters. *Phys. Chem. Chem. Phys.* 15:3965–3976
42. Mandelshtam VA, Frantsuzov PA, Calvo F. 2006. Structural transitions and melting in LJ_{74–78} Lennard-Jones clusters from adaptive exchange Monte Carlo simulations. *J. Phys. Chem. A* 110:5326–5332
43. Calvo F. 2010. Free-energy landscapes from adaptively biased methods: Application to quantum systems. *Phys. Rev. E* 82:046703
44. Wales DJ. 2013. Surveying a complex potential energy landscape: Overcoming broken ergodicity using basin-sampling. *Chem. Phys. Lett.* 584:1–9
45. Sehgal RM, Maroudas D, Ford DM. 2014. Phase behavior of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* 140:104312
46. Picciani M, Athenes M, Kurchan J, Tailleur J. 2011. Simulating structural transitions by direct transition current sampling: The example of LJ₃₈. *J. Chem. Phys.* 135:034108
47. de Souza VK, Wales DJ. 2008. Energy landscapes for diffusion: Analysis of cage-breaking processes. *J. Chem. Phys.* 129:164507
48. de Souza VK, Wales DJ. 2009. Connectivity in the potential energy landscape for binary Lennard-Jones systems. *J. Chem. Phys.* 130:194508
49. Niblett SP, De Souza VK, Stevenson JD, Wales DJ. 2016. Dynamics of a molecular glass former: Energy landscapes for diffusion in ortho-terphenyl. *J. Chem. Phys.* 145
50. de Souza VK, Stevenson JD, Niblett SP, Farrell JD, Wales DJ. 2017. Defining and quantifying frustration in the energy landscape: Applications to atomic and molecular clusters, biomolecules, jammed and glassy systems. *J. Chem. Phys.* 146:124103
51. Chebaro Y, Ballard AJ, Chakraborty D, Wales DJ. 2015. Intrinsically disordered energy landscapes. *Sci. Rep.* 5:10386
52. Chakrabarti D, Wales DJ. 2011. Coupled linear and rotary motion in supramolecular helix handedness inversion. *Soft Matter* 7:2325–2328
53. Röder K, Wales DJ. 2017. Transforming the energy landscape of a coiled-coil peptide via point mutations. *J. Chem. Theory Comput.* 13:1468–1477.
54. Doye JPK, Wales DJ. 1999. The dynamics of structural transitions in sodium chloride clusters. *J. Chem. Phys.* 111:11070–11079
55. Krivov SV, Karplus M. 2004. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Nat. Acad. Sci. USA* 101:14766–14770
56. Krivov SV, Karplus M. 2006. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B* 110:12689–12698
57. Muff S, Caffisch A. 2008. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β -sheet

- miniprotein. *Proteins: Struct., Func. and Bioinf.* 70:1185–1195
58. Krivov SV, Karplus M. 2008. Diffusive reaction dynamics on invariant free energy profiles. *Proc. Nat. Acad. Sci. USA* 105:13841–13846
 59. Wales DJ. 2015. Perspective: Insight into reaction coordinates and dynamics from the potential energy landscape. *J. Chem. Phys.* 142:130901
 60. Steinhardt PJ, Nelson DR, Ronchetti M. 1983. Bond-orientational order in liquids and glasses. *Phys. Rev. B* 28:784–805
 61. van Duijneveldt JS, Frenkel D. 1992. Computer simulation study of free energy barriers in crystal nucleation *J. Chem. Phys.* 96:4655–4668
 62. Frenkel D, Smit B. 2002. Understanding molecular simulation, second edition. London: Academic Press
 63. Ferrenberg AM, Swendsen RH. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638
 64. Ferrenberg AM, Swendsen RH. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198
 65. Weerasinghe S, Amar FG. 1993. Absolute classical densities of states for very anharmonic systems and applications to the evaporation of rare gas clusters. *J. Chem. Phys.* 98:4967–4983
 66. Chill ST, Stevenson J, Ruehle V, Shang C, Xiao P, et al. 2014. Benchmarks for characterization of minima, transition states, and pathways in atomic, molecular, and condensed matter systems. *J. Chem. Theor. Comput.* 10:5476–5482
 67. Mantell RG, Pitt CE, Wales DJ. 2016. Gpu-accelerated exploration of biomolecular energy landscapes. *Journal of Chemical Theory and Computation* 12:6182–6191
 68. Kusumaatmaja H, Whittleston CS, Wales DJ. 2012. A local rigid body framework for global optimization of biomolecules. *J. Chem. Theor. Comput.* 8:5159–5165
 69. Doye JPK, Wales DJ. 1995. Calculation of thermodynamic properties of small lennard-jones clusters incorporating anharmonicity. *J. Chem. Phys.* 102:9659–9672
 70. Doye JPK, Wales DJ. 1995. An order-parameter approach to coexistence in atomic clusters. *J. Chem. Phys.* 102:9673–9688
 71. Calvo F, Doye JPK, Wales DJ. 2001. Characterization of anharmonicities on complex potential energy surfaces: Perturbation theory and simulation. *J. Chem. Phys.* 115:9627–9636
 72. Georgescu I, Mandelshtam VA. 2012. Self-consistent phonons revisited. i. the role of thermal versus quantum fluctuations on structural transitions in large lennard-jones clusters. *J. Chem. Phys.* 137:144106
 73. Li Z, Scheraga HA. 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* 84:6611–6615
 74. Wales DJ, Scheraga HA. 1999. Global optimization of clusters, crystals and biomolecules. *Science* 285:1368–1372
 75. Leary RH, Doye JPK. 1999. New tetrahedral global minimum for the 98-atom lennard-jones cluster. *Phys. Rev. E* 60:R6320–R6322
 76. Shang C, Wales DJ. 2014. Communication: Optimal parameters for basin-hopping global optimization based on tsallis statistics. *J. Chem. Phys.* 141:071101
 77. Sutherland-Cash K, Wales D, Chakrabarti D. 2015. Free energy basin-hopping. *Chem. Phys. Lett.* 625:1–4
 78. Calvo F, Schebarchov D, Wales DJ. 2016. Grand and semigrand canonical basin-hopping. *J. Chem. Theory Comput.* 12:902–909
 79. Jellinek J, Krissinel E. 1996. Ni_nAl_m alloy clusters: analysis of structural forms and their energy ordering. *Chem. Phys. Lett.* 258:283–292
 80. Schebarchov D, Wales DJ. 2013. Communication: A new paradigm for structure prediction in multicomponent systems. *J. Chem. Phys.* 139:221101
 81. Schebarchov D, Wales DJ. 2014. Structure prediction for multicomponent materials using bi-minima. *Phys. Rev. Lett.* 113:156102

82. Schebarchov D, Wales DJ. 2015. Quasi-combinatorial energy landscapes for nanoalloy structure optimisation. *Phys. Chem. Chem. Phys.* 17:28331–28338
83. Swendsen RH, Wang JS. 1986. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* 57:2607–2609
84. Geyer G. 1991. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. EK Keramidas. Interface Foundation, Fairfax Station
85. Hukushima K, Nemoto K. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Japan* 65:1604–1608
86. Labastie P, Whetten RL. 1990. Statistical mechanics of the cluster solid-liquid transition. *Phys. Rev. Lett.* 65:1567–1570
87. Calvo F, Labastie P. 1995. Configurational density of states from molecular dynamics simulations. *Chem. Phys. Lett.* 247:395–400
88. Martiniani S, Schrenk KJ, Stevenson JD, Wales DJ, Frenkel D. 2016. Turning intractable counting into sampling: Computing the configurational entropy of three-dimensional jammed packings. *Phys. Rev. E* 93:012906
89. Martiniani S, Schrenk KJ, Stevenson JD, Wales DJ, Frenkel D. 2016. Structural analysis of high-dimensional basins of attraction. *Phys. Rev. E* 94:031301
90. Wales DJ. 2017. Decoding heat capacity features from the energy landscape. *Phys. Rev. E* 95:030105(R)
91. Strodel B, Whittleston CS, Wales DJ. 2007. Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. *J. Amer. Chem. Soc.* 129:16005–16014
92. Ballard AJ, Stevenson JD, Das R, Wales DJ. 2016. Energy landscapes for a machine learning application to series data. *J. Chem. Phys.* 144:124119
93. Ballard AJ, Das R, Martiniani S, Mehta D, Sagun L, et al. 2017. Energy landscapes for machine learning. *Phys. Chem. Chem. Phys.* 19:12585–12603
94. Pavlovskaja M, Tu K, Zhu SC. 2014. Mapping energy landscapes of non-convex learning problems. *arXiv:1410.0576 [stat.ML]*
95. Das R, Wales DJ. 2016. Energy landscapes for a machine-learning prediction of patient discharge. *Phys. Rev. E* 93:063310
96. Wales DJ. 1992. Basins of attraction for stationary-points on a potential-energy surface. *J. Chem. Soc. Faraday Trans.* 88:653–657
97. Wales DJ. 1993. Locating stationary-points for clusters in Cartesian coordinates. *J. Chem. Soc. Faraday Trans.* 89:1305–1313
98. Asenjo D, Stevenson JD, Wales DJ, Frenkel D. 2013. Visualizing basins of attraction for different minimization algorithms. *J. Phys. Chem. B* 117:12717–12723