International Reviews in Physical Chemistry,
Vol. 25, Nos. 1–2, January–June 2006, 237–282

Taylor & Francis
Taylor & Francis Group

Energy landscapes: calculating pathways and rates

DAVID J. WALES*

Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK

(*Received 27 January 2006*)

The stationary points of a potential energy surface provide a convenient framework for coarse-graining calculations of thermodynamics and kinetics. Thermodynamic properties can be extracted from a database of local minima using the superposition approach, where the total partition function is written as a sum over the contributions from each minimum. To analyse kinetics, we must also consider the transition states that link individual local minima, and evaluate rate constants for the corresponding elementary rearrangements. For small molecules the assignment of separate thermodynamic quantities, such as free energies, to individual isomers, and the notion of isomerisation rates between these structures, is usually straightforward. However, for larger systems the experimental states of interest generally correspond to sets of local minima with some common feature, such as a particular structural motif. This review focuses upon the discrete path sampling approach to obtaining phenomenological two-state rate constants between ensembles of local minima that are distinguished by suitable order parameters. Examples are discussed for atomic and molecular clusters, and for two peptides.

Contents

PAGE

1. Introduction 238

2. Discrete path sampling 240
2.1. Rate constant formulations 240
2.2. Building stationary point databases 246
2.3. Calculation of rate constants 248

3. Characterising stationary points and pathways 252

4. Disconnectivity graphs 255

5. Results 259
5.1. Lennard-Jones clusters 259
5.2. Water clusters 266
5.3. Peptides 270

References 277

*Email: dw34@cam.ac.uk

International Reviews in Physical Chemistry
ISSN 0144–235X print/ISSN 1366–591X online
http://www.tandf.co.uk/journals
DOI: 10.1080/01442350600676921

1. Introduction

Obtaining thermodynamic or kinetic properties for complex molecules or models of bulk matter using computer simulation is often a difficult task. The problems associated with length and time scales that are hard to access computationally are further exacerbated if there are large potential energy barriers (compared to the thermal energy k_BT) between important regions of configuration space. Here k_B is Boltzmann's constant and T is the temperature. Transitions between such regions then correspond to 'rare events', and developing alternative approaches to treat the kinetics is an active area of current research [1–12]. A number of these methods involve some sort of additional coarse-graining of the phase space, such as the interface formulation [4] of transition path sampling [5, 6], milestoning [9], Markovian state models [8], master equation approaches [10, 13], and discretised reaction paths [11]. The main focus of the present review is on the discrete path sampling approach [2, 7, 14], where coarse-graining is achieved using stationary points of the underlying potential energy surface (PES).

For small molecules it is usually straightforward to consider separate partition functions or densities of states for distinct isomers. These isomers correspond to local minima on the PES, and we can associate these states with distinct free energies so long as they are separated by barriers that are large compared to k_BT. Isomerisations, or transitions between local minima, each correspond to distinct rate constants. The same framework may be applied to larger systems, where the number of local minima usually grows exponentially with the number of atoms [15, 16]. For such systems the experimental states of interest often correspond to sets of local minima distinguished by an order parameter, such as the number of native hydrogen bonds for a protein. To calculate thermodynamic properties for such a state we add the partition functions for the component minima to obtain a partition function for the set of local minima. In the same way, the complete partition function for the whole system can be written as a sum over the partition functions of all the local minima. This summation is the basis of the *superposition approach* to thermodynamics [14, 15, 17–22], and is formally exact.

The appropriate superposition expressions for the microcanonical and canonical ensembles are

$$\Omega(E) = \sum_{\alpha} \Omega_{\alpha}(E) \quad \text{and} \quad Z(T) = \sum_{\alpha} Z_{\alpha}(T), \tag{1}$$

where $\Omega_{\alpha}(E)$ and $Z_{\alpha}(T)$ are the microcanonical density of states and canonical partition function for local minimum α, respectively. The superposition decomposition is rigorous if the total configuration space can be partitioned into contributions associated with all the local minima. This breakdown can be achieved by defining the 'catchment basin' [23, 24], or 'basin of attraction', for each local minimum as the region of configuration space for which steepest-descent paths converge to that minimum [14]. Since steepest-descent paths are defined by a first-order differential equation, they are determined uniquely by any point lying on the path that has a non-zero gradient. It is assumed that the boundaries between catchment basins, which

correspond to transition surfaces between minima, make a negligible contribution to the thermodynamics.

The superposition approach has several attractive features. In particular, the contributions to any thermodynamic property calculated from the resulting partition function can be broken down into terms corresponding to different local minima, or regions of configuration space. Furthermore, the total partition function written in the superposition form is naturally ergodic, providing access to the global thermodynamics of systems involving two or more states separated by large potential energy barriers. The local minima involved are also sometimes referred to as 'inherent structures', following Stillinger and Weber, who extended the superposition approach to bulk systems [15]. Although harmonic vibrational densities of states are normally used when implementing the superposition method, anharmonic [25–30] and quantum [29, 31] corrections have both been considered.

Transforming the PES into the basins of attraction of local minima also provides the basis for the basin-hopping global optimisation algorithm [14, 32, 33]. In this approach moves are proposed in configuration space followed by minimisation of the potential energy. Moves between local minima are then accepted or rejected using a Metropolis criterion involving the potential energy difference and a fictitious temperature, which is an adjustable parameter. We have recently adapted this method to obtain total densities of states and partition functions in the 'basin-sampling' (BS) approach [34]. Here we first perform a Wang-Landau-type [35–37] sampling to obtain the potential energy density of local minima, where each proposed move includes a minimisation step, as in basin-hopping. Average anharmonic vibrational partition functions are then calculated using a representative sample of local minima in each potential energy bin using a distance constraint, or 'tether'. This part of the calculation is similar to the confinement approach [38], but unlike that method it does not require additional minimisations to identify the instantaneous catchment basin. The first basin-sampling results for atomic clusters suggest that the method is almost as accurate as parallel tempering [39–43], and can be much faster for quasi-ergodic systems [44–49]. This technique therefore represents a systematic way to assure proper sampling of local minima for superposition calculations in larger systems.

Local minima on the PES are stationary points (where the gradient of the potential energy vanishes) and any infinitesimal displacement of the internal coordinates increases the energy. To discuss kinetics we must also consider transition states, which are defined here geometrically [50] as stationary points with a single imaginary normal mode frequency. Positive and negative displacements along this particular normal mode lower the energy, and usually correspond to the steepest-descent paths that lead to two local minima. These minima are then said to be 'adjacent' and connected by the transition state in question. Infinitesimal displacements along the other vibrational normal modes raise the energy, just as for a local minimum. Kunz and Berry [51] employed connected stationary point databases to calculate global dynamics using a master equation approach [52, 53]. A number of studies followed this work, including applications to clusters [26–28, 54–56], peptides [57, 58] and model proteins [59]. In these studies, the stationary point databases were obtained by starting from the global potential energy minimum and performing a fixed number of transition state searches for a specified number of low-lying structures.

For a sufficiently long search, this approach would eventually include all the higher-energy minima that might be important in mediating the kinetics between the particular regions of configuration space. However, in the spirit of the dynamics-based transition path sampling [5, 6] it should be more efficient to develop a sampling scheme for local minima and transition states that is specifically geared to converging a phenomenological two-state rate constant. The discrete path sampling approach [2, 7, 14] was developed for this purpose, and has now been applied to a variety of atomic and molecular clusters, as well as several peptides and a small protein [2, 7, 60–63].

2. Discrete path sampling

Many different methods have been suggested for studying 'rare events' via computer simulation [1, 5, 6, 64–83]. The present review focuses on the discrete path sampling (DPS) approach, which is based on stationary points of the PES, as described in the Introduction. Here a discrete path is defined as a connected sequence of local minima together with the intervening transition states, which links particular endpoint structures [2]. The number of steps is defined as the number of transition states in the sequence, which is one less than the total number of minima. In the simplest case, two minima may be linked by a single transition state. The link is defined by the two steepest-descent paths that lead downhill in energy from the transition state following infinitesimal displacements parallel and antiparallel to the Hessian eigenvector corresponding to the unique negative eigenvalue. Discrete paths between endpoints that are well separated in configuration space are likely to involve many intervening local minima and transition states.

Suppose that we wish to calculate the phenomenological two-state rate constant between states A and B. We first need some way to define these states unambiguously, which may entail the use of an order parameter of some sort. In particular, we must be able to determine whether a given local minimum belongs to A or B, or to neither, in which case it will be placed in the set of intervening minima, I. The most straightforward case involves fixed A and B sets, which are predefined. However, it is also possible, in principle, to change the A and B sets during or after the construction of the stationary point database.

2.1. *Rate constant formulations*

For single-exponential, two-state kinetics to apply we must have local equilibrium within each of the A and B regions on the time scale corresponding to $A \leftrightarrow B$ transitions. These transitions must therefore correspond to the slowest relaxation time scale in the system [73, 84]. In this case the occupation probability of minima $a \in A$ and $b \in B$ at time t, $p_a(t)$ and $p_b(t)$, can be written as

$$p_a(t) = \frac{p_a^{eq} p_A(t)}{p_A^{eq}} \qquad \text{and} \qquad p_b(t) = \frac{p_b^{eq} p_B(t)}{p_B^{eq}}, \tag{2}$$

where the superscript 'eq' stands for 'equilibrium'. Assuming that the dynamics between adjacent local minima are Markovian, so that the system loses its memory of the previous minimum before it makes another transition, the overall kinetics can be represented by the master equation [52, 53]

$$\frac{dp_\alpha(t)}{dt} = \sum_{\beta\neq\alpha}[k_{\alpha\beta}p_\beta(t) - k_{\beta\alpha}p_\alpha(t)], \tag{3}$$

where $k_{\beta\alpha}$ is the rate constant from minimum α to minimum β and the sums are over all adjacent minima.

First suppose that all minima belong to either the A or the B set. Then assuming local equilibrium within the A and B regions we can write the master equation as

$$\frac{dp_\mathrm{A}(t)}{dt} = k_\mathrm{AB}p_\mathrm{B}(t) - k_\mathrm{BA}p_\mathrm{A}(t) \quad \text{and} \quad \frac{dp_\mathrm{B}(t)}{dt} = k_\mathrm{BA}p_\mathrm{A}(t) - k_\mathrm{AB}p_\mathrm{B}(t), \tag{4}$$

where $p_\mathrm{A}(t) = \sum_{\mathrm{a}\in\mathrm{A}} p_\mathrm{a}(t)$, $p_\mathrm{B}(t) = \sum_{\mathrm{b}\in\mathrm{B}} p_\mathrm{b}(t)$,

$$k_\mathrm{AB} = \frac{1}{p_\mathrm{B}^\mathrm{eq}}\sum_{\mathrm{a}\in\mathrm{A}}\sum_{\mathrm{b}\in\mathrm{B}} k_\mathrm{ab}p_\mathrm{b}^\mathrm{eq} \quad \text{and} \quad k_\mathrm{BA} = \frac{1}{p_\mathrm{A}^\mathrm{eq}}\sum_{\mathrm{a}\in\mathrm{A}}\sum_{\mathrm{b}\in\mathrm{B}} k_\mathrm{ba}p_\mathrm{a}^\mathrm{eq}. \tag{5}$$

This result is instructive, because it reveals that the phenomenological rate constants can be evaluated in terms of minimum-to-minimum rates between A and B minima on the boundary between the two regions, weighted by the conditional equilibrium probability $p_\mathrm{a}^\mathrm{eq}/p_\mathrm{A}^\mathrm{eq}$ or $p_\mathrm{b}^\mathrm{eq}/p_\mathrm{B}^\mathrm{eq}$. This separation into a probability, which can be calculated from thermodynamics, and a purely dynamical factor is a well-known result that is used in other approaches to rare events [64, 84–86].

More generally, it is often the case that some minima will belong to the intervening set, I, although this does not preclude the existence of direct $\mathrm{A}\leftrightarrow\mathrm{B}$ connections via single transition states as well. Applying the steady-state approximation for each minimum $\mathrm{i}\in\mathrm{I}$ then gives

$$\begin{aligned} \frac{dp_\mathrm{i}(t)}{dt} &= \sum_\alpha k_{\mathrm{i}\alpha}\,p_\alpha(t) - p_\mathrm{i}(t)\sum_\beta k_{\beta\mathrm{i}} \approx 0, \\ \text{so that} \quad p_\mathrm{i}(t) &= \frac{\sum_\alpha k_{\mathrm{i}\alpha}\,p_\alpha(t)}{\sum_\beta k_{\beta\mathrm{i}}}. \end{aligned} \tag{6}$$

We can now systematically replace every occurrence of $p_\mathrm{i}(t)$ in the master equation using equation (6), and each substitution introduces a linear combination of probabilities $p_\mathrm{j}(t)$ ($\mathrm{j}\in\mathrm{I}$), $p_\mathrm{a}(t)$ and $p_\mathrm{b}(t)$, along with extra factors of k_ji, k_ai and k_bi, respectively. Hence $dp_\mathrm{A}(t)/dt$ and $dp_\mathrm{B}(t)/dt$ can finally be written as sums over all

possible paths that start and finish on the boundaries of the A and B regions:

$$\frac{dp_{\mathrm{A}}(t)}{dt} = \sum_{\mathrm{a}\leftarrow\mathrm{a}'} \frac{k_{\mathrm{ai}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{a}'}\, p_{a'}(t)}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}} + \sum_{\mathrm{a}\leftarrow\mathrm{b}} \frac{k_{\mathrm{ai}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{b}}\, p_{\mathrm{b}}(t)}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}} - \sum_{\mathrm{a}}\sum_{\mathrm{i}} k_{\mathrm{ia}}\, p_{\mathrm{a}}(t),$$

$$\frac{dp_{\mathrm{B}}(t)}{dt} = \sum_{\mathrm{b}\leftarrow\mathrm{b}'} \frac{k_{\mathrm{bi}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{b}'}\, p_{\mathrm{b}'}(t)}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}} + \sum_{\mathrm{b}\leftarrow\mathrm{a}} \frac{k_{\mathrm{bi}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{a}}\, p_{\mathrm{a}}(t)}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}} - \sum_{\mathrm{i}}\sum_{\mathrm{b}} k_{\mathrm{ib}}\, p_{\mathrm{b}}(t), \tag{7}$$

where, for example, a $\leftarrow$ b indicates that the sum is over all the possible paths that begin from a minimum $\mathrm{b}\in\mathrm{B}$ and end in a minimum $\mathrm{a}\in\mathrm{A}$, passing only through minima $\mathrm{i}\in\mathrm{I}$. We can immediately identify the k_{AB} term in the expression for $dp_{\mathrm{A}}(t)/dt$, and the k_{BA} term in $dp_{\mathrm{B}}(t)/dt$, so that

$$k^{\mathrm{SS}}_{\mathrm{AB}} = \frac{1}{p^{\mathrm{eq}}_{\mathrm{B}}} \sum_{\mathrm{a}\leftarrow\mathrm{b}} \frac{k_{\mathrm{ai}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{b}}\, p^{\mathrm{eq}}_{\mathrm{b}}}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}}$$

$$\text{and} \qquad k^{\mathrm{SS}}_{\mathrm{BA}} = \frac{1}{p^{\mathrm{eq}}_{\mathrm{A}}} \sum_{\mathrm{b}\leftarrow\mathrm{a}} \frac{k_{\mathrm{bi}_1}\, k_{\mathrm{i}_1\mathrm{i}_2}\cdots k_{\mathrm{i}_n\mathrm{a}}\, p^{\mathrm{eq}}_{\mathrm{a}}}{\sum_{\alpha_1} k_{\alpha_1\mathrm{i}_1} \sum_{\alpha_2} k_{\alpha_2\mathrm{i}_2}\cdots\sum_{\alpha_n} k_{\alpha_n\mathrm{i}_n}}, \tag{8}$$

where we have also used the local equilibrium condition from equation (2). The SS superscript is used to emphasise that the steady-state condition has been applied to intervening minima, and distinguish these rate constants from those obtained using different approximations below. Note that $dp_{\mathrm{A}}(t)/dt = -dp_{\mathrm{B}}(t)/dt$ holds when we use $k^{\mathrm{SS}}_{\mathrm{AB}}$ and $k^{\mathrm{SS}}_{\mathrm{BA}}$ in equation (4), because $dp_{\mathrm{I}}(t)/dt = 0$ by assumption. Detailed balance applies to any particular $a \leftrightarrow b$ discrete path, and to the overall rates, so that $k_{\mathrm{AB}}/k_{\mathrm{BA}} = p^{\mathrm{eq}}_{\mathrm{A}}/p^{\mathrm{eq}}_{\mathrm{B}}$.

Terms of the form $k_{\mathrm{ji}}/\sum_\alpha k_{\alpha\mathrm{i}}$ can be replaced by the branching probability P_{ji} in equation (8). If the escape routes from a given minimum i are regarded as competing Poisson processes, then escape via any route follows Poisson statistics with rate $k = \sum_\alpha k_{\alpha\mathrm{i}}$ [87]. The waiting time for a transition to occur to any adjacent minimum is exponentially distributed as $k\exp(-kt)$ [88], and the mean waiting time for escape is $\tau_{\mathrm{i}} = 1/\sum_\alpha k_{\alpha\mathrm{i}}$. Hence $P_{\mathrm{ji}} = k_{\mathrm{ji}}\tau_{\mathrm{i}}$ and

$$\begin{aligned} k^{\mathrm{SS}}_{\mathrm{AB}} &= \frac{1}{p^{\mathrm{eq}}_{\mathrm{B}}} \sum_{\mathrm{a}\leftarrow\mathrm{b}} P_{\mathrm{ai}_1}\, P_{\mathrm{i}_1\mathrm{i}_2}\cdots P_{\mathrm{i}_{n-1}\mathrm{i}_n}\, k_{\mathrm{i}_n\mathrm{b}}\, p^{\mathrm{eq}}_{\mathrm{b}} \\ &= \frac{1}{p^{\mathrm{eq}}_{\mathrm{B}}} \sum_{\mathrm{a}\leftarrow\mathrm{b}} P_{\mathrm{ai}_1}\, P_{\mathrm{i}_1\mathrm{i}_2}\cdots P_{\mathrm{i}_{n-1}\mathrm{i}_n}\, P_{\mathrm{i}_n\mathrm{b}}\, \tau^{-1}_{\mathrm{b}}\, p^{\mathrm{eq}}_{\mathrm{b}}, \end{aligned} \tag{9}$$

and similarly for $k_{\mathrm{BA}}^{\mathrm{SS}}$. The sums in equation (9) are again over all paths that start from a B minimum and end at an A minimum after passing through only I minima, which may be revisited. The sum of the branching probability product over all such paths starting from $\mathrm{b} \in \mathrm{B}$ is the probability that the system will visit an A minimum before it returns to the B region, which we write as $C_{\mathrm{b}}^{\mathrm{A}}$. Similarly, $C_{\mathrm{b}}^{\mathrm{B}} = 1 - C_{\mathrm{b}}^{\mathrm{A}}$ is the probability that a random walk starting from b will return to any minimum in the B region before it encounters an A minimum. $C_{\mathrm{b}}^{\mathrm{A}}$ is referred to as a committor probability, defined as

$$C_{\mathrm{b}}^{\mathrm{A}} = \sum_{\mathrm{A} \leftarrow \mathrm{b}} P_{\mathrm{a}\mathrm{i}_1} P_{\mathrm{i}_1\mathrm{i}_2} \cdots P_{\mathrm{i}_{n-1}\mathrm{i}_n} P_{\mathrm{i}_n\mathrm{b}}. \tag{10}$$

Hence we can write the rate constants that involve the steady-state approximation as

$$k_{\mathrm{AB}}^{\mathrm{SS}} = \frac{1}{p_{\mathrm{B}}^{\mathrm{eq}}} \sum_{\mathrm{b} \in \mathrm{B}} \frac{C_{\mathrm{b}}^{\mathrm{A}} p_{\mathrm{b}}^{\mathrm{eq}}}{\tau_{\mathrm{b}}} \qquad \text{and} \qquad k_{\mathrm{BA}}^{\mathrm{SS}} = \frac{1}{p_{\mathrm{A}}^{\mathrm{eq}}} \sum_{\mathrm{a} \in \mathrm{A}} \frac{C_{\mathrm{a}}^{\mathrm{B}} p_{\mathrm{a}}^{\mathrm{eq}}}{\tau_{\mathrm{a}}}. \tag{11}$$

The parameter $P_{\alpha}^{\mathrm{fold}}$, defined as the probability that a protein will fold before unfolding, starting from some initial condition α [8, 89, 90], is a more specific example of a committor probability.

In deriving the above expressions for $k_{\mathrm{AB}}^{\mathrm{SS}}$ and $k_{\mathrm{BA}}^{\mathrm{SS}}$ we considered a Markov process in which the state space consisted of all the local minima on the PES. To remove the steady-state approximation, we now describe the kinetics in terms of a Markov process in the restricted space consisting only of the A and B minima. Writing a master equation in terms of effective rate constants K_{ab} and K_{ba} for transitions between A and B minima we obtain

$$\begin{aligned} k_{\mathrm{AB}}^{\mathrm{NSS}} &= \frac{1}{p_{\mathrm{B}}^{\mathrm{eq}}} \sum_{\mathrm{b} \in \mathrm{B}} \sum_{\mathrm{a} \in \mathrm{A}} K_{\mathrm{ab}}\, p_{\mathrm{b}}^{\mathrm{eq}} = \frac{1}{p_{\mathrm{B}}^{\mathrm{eq}}} \sum_{\mathrm{b} \in \mathrm{B}} K_{\mathrm{Ab}}\, p_{\mathrm{b}}^{\mathrm{eq}} \\ \text{and} \qquad k_{\mathrm{BA}}^{\mathrm{NSS}} &= \frac{1}{p_{\mathrm{A}}^{\mathrm{eq}}} \sum_{\mathrm{a} \in \mathrm{A}} \sum_{\mathrm{b} \in \mathrm{B}} K_{\mathrm{ba}}\, p_{\mathrm{a}}^{\mathrm{eq}} = \frac{1}{p_{\mathrm{A}}^{\mathrm{eq}}} \sum_{\mathrm{a} \in \mathrm{A}} K_{\mathrm{Ba}}\, p_{\mathrm{a}}^{\mathrm{eq}}, \end{aligned} \tag{12}$$

where the superscript 'NSS' stands for 'non-steady-state'. If we treat transitions from minimum b to the A and B regions as independent Poisson processes then the mean waiting time between transitions is $t_{\mathrm{b}} = 1/(K_{\mathrm{Ab}} + K_{\mathrm{Bb}})$, with an analogous expression for A minima. Here K_{Bb} corresponds to the effective rate constant for a trajectory to return to any member of the B set starting from b. If we now identify $K_{\mathrm{Ab}}/(K_{\mathrm{Ab}} + K_{\mathrm{Bb}}) = K_{\mathrm{Ab}} t_{\mathrm{b}}$ with the committor probability $C_{\mathrm{b}}^{\mathrm{A}}$ then we obtain

$$k_{\mathrm{AB}}^{\mathrm{NSS}} = \frac{1}{p_{\mathrm{B}}^{\mathrm{eq}}} \sum_{\mathrm{b} \in \mathrm{B}} \frac{C_{\mathrm{b}}^{\mathrm{A}} p_{\mathrm{b}}^{eq}}{t_{\mathrm{b}}}, \qquad \text{and} \qquad k_{\mathrm{BA}}^{\mathrm{NSS}} = \frac{1}{p_{\mathrm{A}}^{\mathrm{eq}}} \sum_{\mathrm{a} \in \mathrm{A}} \frac{C_{\mathrm{a}}^{\mathrm{B}} p_{\mathrm{a}}{}^{\mathrm{eq}}}{t_{\mathrm{a}}}. \tag{13}$$

The difference between these rate constants and the steady-state analogues lies in the average waiting times between events. In equation (13) t_{b} is the average waiting time

for transitions between minimum b and any minimum in the A or B sets. In contrast, τ_b is the waiting time for a transition from minimum b to any of its adjacent minima, so that $\tau_b \leq t_b$. In the steady-state limit the waiting times in all the intervening minima should be negligible, and in this case $t_b \rightarrow \tau_b$ and the rate constants obtained from equations (13) and (11) should coincide. If these results do not agree, it is likely that the steady-state approximation does not hold, and one or more kinetic intermediates are present.

To evaluate the waiting times t_a and t_b we can use the kinetic Monte Carlo (KMC) approach [91–95]. The KMC simulations generate a memoryless (Markovian) random walk between the local minima, which is governed by the branching probabilities $P_{\alpha\beta}$. An efficient algorithm to propagate such trajectories is the '*n*-fold way' of Bortz, Kalos and Lebowitz [91], where the ratios between transition probabilities of different events are preserved, but rejections are eliminated. Methods to accelerate such calculations also exist, such as 'Monte Carlo with absorbing Markov chains' [96] and alternative sampling schemes [97, 98]. In the *n*-fold scheme, we choose the next step from minimum i according to the probabilities $P_{\alpha i}$. The simulation time can simply be incremented by τ_i if only the mean first passage time (MFPT) between the endpoints is of interest.

The KMC runs required to calculate t_a and t_b are generally short, because they terminate as soon as the trajectory encounters any A or B minimum. For example, if the potential energy barrier between the A and B regions is large compared to k_BT, trajectories started from a B minimum will probably return to a B minimum in a few steps, and similarly for trajectories started from an A minimum. The averages over KMC runs used to calculate t_a and t_b should converge rapidly in this case, and need not include any events where the trajectory actually crosses the potential energy barrier. Such events will be rare, and their contribution to the average will not be very different from the trajectories that return to the starting region, because waiting times for minima in the intervening barrier region should be relatively small. As noted above, a significant difference between t_α and τ_α for $\alpha \in$ A or B indicates a breakdown in the steady-state approximation and in the two-state description.

An alternative formulation for the two-state rate constants is obtained by initiating KMC trajectories from minima in the starting region, and following them until they reach any minimum in the other region. This approach provides an MFPT from one region to the other when averaged over a suitable number of KMC runs. The corresponding rate constants will be denoted by a superscript KMC, and are calculated as

$$k_{AB}^{KMC} = \frac{1}{p_B^{eq}} \sum_{b \in B} \frac{p_b^{eq}}{\mathcal{T}_{Ab}} \qquad \text{and} \qquad k_{BA}^{KMC} = \frac{1}{p_A^{eq}} \sum_{a \in A} \frac{p_a^{eq}}{\mathcal{T}_{Ba}}, \tag{14}$$

where $\mathcal{T}_{Ba}$ is the MFPT for a KMC trajectory initiated at minimum a to reach any B minimum, etc. In contrast to t_a and t_b the MFPT's $\mathcal{T}_{Ba}$ and $\mathcal{T}_{Ab}$ include an arbitrary number of revisits to any minima in the starting region. Hence $\mathcal{T}_{Ba} \geq t_a \geq \tau_a$. In particular, for a barrier that is large compared to k_BT, the KMC runs used to calculate MFPT's will include many revisits to the starting region [97].

The committor probabilities required for the 'SS' and 'NSS' rate constants can be calculated using a first-step analysis [99], for example,

$$C_\alpha^{\mathrm{A}} = \sum_\beta C_\beta^{\mathrm{A}} P_{\beta\alpha}. \tag{15}$$

An analogous expression is used to calculate P_α^{fold} in reference [8]. Our implementation of this method uses the successive overrelaxation technique with an extrapolation factor of 1.999 [100]. We have also found it advantageous to store the branching probabilities $P_{\beta\alpha}$ using the compressed row storage scheme [101] for sparse matrices.

To derive a formal connection between the above KMC rate constants and $k_{\mathrm{AB}}^{\mathrm{NSS}}$ and $k_{\mathrm{BA}}^{\mathrm{NSS}}$ we again consider a Markov process in the state space defined by the local minima in regions A and B. We can then write a first-step relation [99] between the MFPT's as

$$\mathcal{T}_{\mathrm{Ab}_1} = t_{\mathrm{b}_1} + \sum_{\mathrm{b}_2} C_{\mathrm{b}_1}^{\mathrm{b}_2} \mathcal{T}_{\mathrm{Ab}_2}, \tag{16}$$

where t_{b_1} is the expected waiting time for a transition from b_1 to any other A or B minimum, and $C_{\mathrm{b}_1}^{\mathrm{b}_2}$ is the corresponding committor probability that the transition is to minimum b_2 within this state space. Adding and subtracting $\mathcal{T}_{\mathrm{Ab}_1}$ from $\mathcal{T}_{\mathrm{Ab}_2}$ in the sum and rearranging then gives

$$C_{\mathrm{b}_1}^{\mathrm{A}} \mathcal{T}_{\mathrm{Ab}_1} = t_{\mathrm{b}_1} + \sum_{\mathrm{b}_2} C_{\mathrm{b}_1}^{\mathrm{b}_2} \left(\mathcal{T}_{\mathrm{Ab}_2} - \mathcal{T}_{\mathrm{Ab}_1}\right) \equiv t_{\mathrm{b}_1} + Y_{\mathrm{b}_1}, \tag{17}$$

which defines the quantity Y_{b_1}. Now we can write

$$\begin{aligned}
p_{\mathrm{B}}^{\mathrm{eq}} k_{\mathrm{AB}}^{\mathrm{KMC}} &= \sum_{\mathrm{b}_1} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}}}{\mathcal{T}_{\mathrm{Ab}_1}} = \sum_{\mathrm{b}_1} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{A}}}{t_{\mathrm{b}_1} + Y_{\mathrm{b}_1}}, \\
&= \sum_{\mathrm{b}_1} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{A}}}{t_{\mathrm{b}_1}} - \sum_{\mathrm{b}_1} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{A}} Y_{\mathrm{b}_1}}{t_{\mathrm{b}_1}\left(t_{\mathrm{b}_1} + Y_{\mathrm{b}_1}\right)} \\
&= p_{\mathrm{B}}^{\mathrm{eq}} k_{\mathrm{AB}}^{\mathrm{NSS}} - \sum_{\mathrm{b}_1} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}}}{t_{\mathrm{b}_1} \mathcal{T}_{\mathrm{Ab}_1}} \sum_{\mathrm{b}_2} C_{\mathrm{b}_1}^{\mathrm{b}_2} \left(\mathcal{T}_{\mathrm{Ab}_2} - \mathcal{T}_{\mathrm{Ab}_1}\right) \\
&= p_{\mathrm{B}}^{\mathrm{eq}} k_{\mathrm{AB}}^{\mathrm{NSS}} - \sum_{\mathrm{b}_1 < \mathrm{b}_2} \left(\mathcal{T}_{\mathrm{Ab}_2} - \mathcal{T}_{\mathrm{Ab}_1}\right) \left[\frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{b}_2}}{t_{\mathrm{b}_1} \mathcal{T}_{\mathrm{Ab}_1}} - \frac{p_{\mathrm{b}_2}^{\mathrm{eq}} C_{\mathrm{b}_2}^{\mathrm{b}_1}}{t_{\mathrm{b}_2} \mathcal{T}_{\mathrm{Ab}_2}}\right] \\
&= p_{\mathrm{B}}^{\mathrm{eq}} k_{\mathrm{AB}}^{\mathrm{NSS}} - \sum_{\mathrm{b}_1 < \mathrm{b}_2} \left(\mathcal{T}_{\mathrm{Ab}_2} - \mathcal{T}_{\mathrm{Ab}_1}\right) \left[\frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{b}_2}}{t_{\mathrm{b}_1} \mathcal{T}_{\mathrm{Ab}_1}} - \frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{b}_2}}{t_{\mathrm{b}_1} \mathcal{T}_{\mathrm{Ab}_2}}\right] \\
&= p_{\mathrm{B}}^{\mathrm{eq}} k_{\mathrm{AB}}^{\mathrm{NSS}} - \sum_{\mathrm{b}_1 < \mathrm{b}_2} \frac{\left(\mathcal{T}_{\mathrm{Ab}_2} - \mathcal{T}_{\mathrm{Ab}_1}\right)^2}{\mathcal{T}_{\mathrm{Ab}_1} \mathcal{T}_{\mathrm{Ab}_2}} \frac{p_{\mathrm{b}_1}^{\mathrm{eq}} C_{\mathrm{b}_1}^{\mathrm{b}_2}}{t_{\mathrm{b}_1}}.
\end{aligned} \tag{18}$$

In the penultimate line we have used the detailed balance condition $K_{b_1 b_2} p^{eq}_{b_2} = K_{b_2 b_1} p^{eq}_{b_1}$, which combined with the identity $C^{b_2}_{b_1} = K_{b_2 b_1} t_{b_1}$ gives $p^{eq}_{b_2} C^{b_1}_{b_2} / t_{b_2} = p^{eq}_{b_1} C^{b_2}_{b_1} / t_{b_1}$. Here, as in equation (12), we use a capital K to denote effective rate constants within the restricted state space of A and B minima; the corresponding minima need not be connected by a single transition state. If the B minima are in rapid equilibrium relative to the time scale for transitions to the A region, then any $\mathcal{T}_{Ab}$ will be large compared to any t_b. In fact, to a first approximation the MFPT's from alternative B minima will only differ by terms of order t_b. In this limit the last term of equation (18) vanishes. Hence

$$\lim_{B\ eq} k^{KMC}_{AB} \rightarrow k^{NSS}_{AB} \quad \text{and} \quad \lim_{\tau_i \ll \tau_b} k^{NSS}_{AB} \rightarrow k^{SS}_{AB}, \tag{19}$$

and similarly for the A to B rate constants. We have therefore established a clear hierarchy of approximations: the KMC rate constants are equal to the 'NSS' values when local equilibrium is established in the A and B regions, and the 'NSS' rate constants are equal to the steady-state values when the waiting times for minima in the I region can be neglected. It is particularly noteworthy that the local equilibrium assumption for the A and B regions enables us to eliminate revisits to A and B minima.

2.2. *Building stationary point databases*

In the preceding section we have presented three different formulations for phenomenological two-state rate constants, all of which assume that a connected database of local minima and transition states is available. We now consider how this database might be constructed to provide a useful description of kinetics over a specified temperature range. The scheme outlined here was employed to produce all the results discussed in subsequent sections. However, alternative methods to build databases are still under consideration, and future DPS studies may employ a different approach.

Each DPS study described below started with the definition of local minima in fixed A and B sets, and the construction of an initial discrete path linking any A minimum to any B minimum. This first path may have to span a large distance in configuration space, which can itself present a difficult problem. However, the connection algorithm presented in section 3, which is based on analysis of the shortest path in a network with missing connections, has enabled this part of the calculation to be automated. The geometry optimisation techniques involved in characterising local minima, transition states, and the steepest-descent paths that connect them, are also described in section 3.

For the results reported in section 5 the stationary point databases were produced by considering a current discrete path and systematically perturbing it. The current path at any point in the calculation is usually the path with the largest contribution to k^{SS}_{AB} (or k^{SS}_{BA}, depending on which rate constant is required), for which a specified maximum number of perturbations has yet to be completed. For the A$\leftarrow$B rate this contribution is

$$\frac{1}{p^{eq}_B} P_{a i_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n} P_{i_n b} \tau_b^{-1} p^{eq}_b. \tag{20}$$

Note that there is a conditional probability term in this contribution, i.e. p_b^{eq}/p_B^{eq}, and so the sampling is not based purely on how fast the path is, but also on how likely the starting minimum is to be occupied in equilibrium. Hence, criticisms [102–104] of schemes based only on the fastest paths [105–107] should not apply.

The perturbations to the current path consist of choosing a minimum at random and replacing it with a minimum that does not appear on the path in question. The replacement minimum must be connected to the original minimum by a single transition state, so that the perturbation has a chance of generating a new path that is kinetically relevant. Suitable minima may already exist in the database, and if not, a candidate is generated using a single-ended transition state search (section 3). A connection is then sought between the replacement minimum and the two minima one or more steps away from the original minimum on either side of it on the current path. The two minima chosen can also be randomised subject to maximum and minimum separations along the path. These two connections are obtained using the double-ended network-based shortest-path algorithm described in section 3. If this process succeeds then any new minima and transition states are added to the database, and the resulting discrete path is compared with previous paths that have been perturbed, and checked for shortcuts via known transition states. If the contribution to the appropriate steady-state rate constant calculated from equation (20) is larger than that of the original path then the new path becomes the current path. Otherwise it is recorded, and the original path is restored for a further perturbation. If the maximum number of perturbations has been reached for the current path then we move instead to the recorded path with the largest contribution to the rate for which the maximum number of perturbations has not been exceeded. The sampling terminates when a specified number of paths with the largest contributions to the rate constant have all been perturbed the maximum number of times. These limits can be increased for a subsequent run to check the convergence of the overall rates. For long initial paths it can also be beneficial to run a number of direct double-ended searches between minima in the path to seek shortcuts, and there is a provision for this in our `PATHSAMPLE` program. `PATHSAMPLE` basically acts as a driver and bookkeeping routine for our public domain code `OPTIM`, which performs all the geometry optimisation and pathway searches [108].

The main results of the stationary point sampling phase of the DPS calculation are the recorded discrete paths obtained via perturbations, and the database of local minima and transition states. An infinite number of distinct discrete paths connecting the A and B regions can be constructed from the database, so long as there is at least one path containing two or more intervening minima. For such paths we should include arbitrary recrossings between intervening minima, since these all contribute to the sums in equation (11). For simple cases, all the important contributions to the rate constants may be contained in the recorded discrete paths that were generated explicitly during the sampling phase. However, it is possible to include all the pathways through the database using the techniques described in section 2.3. It is also possible to identify the path with the largest contribution in equation (11) during the sampling phase using Dijkstra's algorithm [109, 110]. Some of the later DPS results described below employed this approach to make sure that the most relevant paths were considered during the perturbation phase.

2.3. *Calculation of rate constants*

We now address the extraction of phenomenological two-state rate constants from the stationary point database. For every formulation considered here, we require the conditional occupation probabilities p_a^{eq}/p_A^{eq} and p_b^{eq}/p_B^{eq}, along with the rate constants for transitions between adjacent minima, $k_{\alpha\beta}$. For nontranslating, nonrotating systems the simplest approximation to these quantities is to use harmonic vibrational densities of states. In this case the microcanonical density of states and the canonical partition function associated with minimum α with potential energy V_α are

$$\Omega_\alpha(E) = \frac{n_\alpha (E - V_\alpha)^{\kappa-1}}{\Gamma(\kappa)(h\bar{\nu}_\alpha)^\kappa} \quad \text{and} \quad Z_\alpha(T) = \frac{n_\alpha e^{-V_\alpha/k_B T}}{(\beta h\bar{\nu}_\alpha)^\kappa}, \tag{21}$$

where $\beta = 1/k_B T$, $n_\alpha = 2N_A!N_B!N_C!\cdots/o_\alpha$, o_α is the order of the point group for the system in question with chemical formula $A_{N_A}B_{N_B}C_{N_C}\cdots$, and n_α is the number of distinct permutation-inversion isomers of the minimum in question [14]. $\bar{\nu}_\alpha$ is the geometric mean vibrational frequency of minimum α, and κ is the number of vibrational degrees of freedom. We can then write the canonical conditional occupation probability of minimum a′ within the A set as $Z_{a'}(T)/Z_A(T)$, etc., where $Z_A(T) = \sum_{a\in A} Z_a(T)$. To consider rates between particular permutation-inversion isomers of the same minimum we need to break this degeneracy factor down appropriately. This breakdown can be achieved automatically by distinguishing atoms of the same element using fictitious masses when assigning the point group.

Unimolecular rate theory [111–116] can be used to calculate the elementary rate constants corresponding to the pathways mediated by a single transition state. For example, the transition state theory expression for the canonical rate constant from minimum α to minimum β via a given transition state is [111–119]

$$k_{\beta\alpha}(T) = \frac{k_B T}{h}\frac{Z_{\beta\alpha}(T)}{Z_\alpha(T)} e^{-(V_{\beta\alpha}-V_\alpha)/k_B T}, \tag{22}$$

where $Z_{\beta\alpha}$ omits the unique mode with the imaginary frequency for the transition state linking minima α and β, which has potential energy $V_{\beta\alpha}$. $k_{\beta\alpha}(T)$ corresponds to the rate constant for a particular minimum-to-minimum transition. However, we often wish to group all the distinct permutation-inversion isomers of each minimum and transition state together. In the harmonic approximation the resulting microcanonical and canonical rate constants are

$$k_{\beta\alpha}(E) = \sigma \frac{(\bar{\nu}_\alpha)^\kappa}{(\bar{\nu}_{\beta\alpha})^{\kappa-1}} \left(\frac{E - V_{\beta\alpha}}{E - V_\alpha}\right)^{\kappa-1} \tag{23}$$

and

$$k_{\beta\alpha}(T) = \sigma \frac{(\bar{\nu}_\alpha)^\kappa}{(\bar{\nu}_{\beta\alpha})^{\kappa-1}} e^{-(V_{\beta\alpha}-V_\alpha)/k_B T}, \tag{24}$$

where the symmetry number $\sigma = o_a/o_{\beta\alpha}$ for a non-degenerate rearrangement and $\sigma = 2o_a/o_{\beta\alpha}$ for a degenerate rearrangement between permutation-inversion isomers of the same structure. $\bar{\nu}_{\beta\alpha}$ is the geometric mean vibrational frequency for the transition state excluding the unique imaginary frequency.

One advantage of the DPS approach is that the partition functions and minimum-to-minimum rate constants could be recalculated using increasingly sophisticated approaches once the stationary point database has been constructed. Consistent levels of theory must be used, so that detailed balance remains satisfied for each transition state. Another advantage is that the rates may be calculated over a range of total energy or temperature from the same database. The database construction is based upon contributions to the phenomenological two-state rate constant calculated at some fixed temperature. The contributions of different stationary points to the overall rate are temperature dependent, but it would only be necessary to build a fresh database for conditions where new stationary points start to make a significant contribution.

Several distinct approaches have previously been used to extract two-state rate constants from a DPS stationary point database [2, 12, 14]. For simple cases, where only a few discrete paths are significant, we can obtain $k_{\mathrm{AB}}^{\mathrm{SS}}$ and $k_{\mathrm{BA}}^{\mathrm{SS}}$ by summing these contributions, which should all have been encountered during the pathway perturbation process. More work is required to sum all the possible contributions from the entire stationary point database. Not only must all the possible paths between A and B be enumerated along with their contributions to equation (11), we must also include all possible recrossings between minima from the I set. These recrossings must be distinguished from dynamical corrections for recrossing of the transition surface, which are not included in transition state theory, for example. Such corrections would enter into the minimum-to-minimum rate constants $k_{\alpha\beta}$, and could be calculated by evaluating the reactive flux [73, 84–86].

All the above contributions in equation (11) can be summed using a matrix multiplication (MM) technique [7]. We first define a weighted adjacency matrix **M** with non-zero elements

$$\begin{aligned} M_{\mathrm{ab}} &= k_{\mathrm{ab}} p_{\mathrm{b}}^{\mathrm{eq}} / p_{\mathrm{B}}^{\mathrm{eq}}, & & M_{\mathrm{ab}} = k_{\mathrm{ab}} p_{\mathrm{b}}^{\mathrm{eq}} / p_{\mathrm{B}}^{\mathrm{eq}}, \\ M_{\mathrm{ai}} &= k_{\mathrm{ai}}, & & M_{\mathrm{ai}} = k_{\mathrm{ai}} / \textstyle\sum_{\gamma} k_{\gamma \mathrm{i}}, \\ M_{\mathrm{ij}} &= k_{\mathrm{ij}} \Big/ \textstyle\sum_{\gamma} k_{\gamma \mathrm{j}}, & \text{or} \quad & M_{\mathrm{ij}} = k_{\mathrm{ij}} \Big/ \textstyle\sum_{\gamma} k_{\gamma \mathrm{j}}, \\ M_{\mathrm{ib}} &= k_{\mathrm{ib}} p_{\mathrm{b}}^{\mathrm{eq}} \Big/ p_{\mathrm{B}}^{\mathrm{eq}} \textstyle\sum_{\gamma} k_{\gamma \mathrm{i}}, & & M_{\mathrm{ib}} = k_{\mathrm{ib}} p_{\mathrm{b}}^{\mathrm{eq}} \Big/ p_{\mathrm{B}}^{\mathrm{eq}}, \end{aligned} \tag{25}$$

where $\mathrm{i}, \mathrm{j} \in \mathrm{I}$. Both the above formulations should converge to the same result for $k_{\mathrm{AB}}^{\mathrm{SS}}$, and an analogous matrix can be defined to calculate $k_{\mathrm{BA}}^{\mathrm{SS}}$. The contribution to k_{AB} from paths with n steps (transition states) linking minima a and b is $[\mathbf{M}^n]_{\mathrm{ab}}$. Hence we simply sum matrix elements of increasing powers of **M** until a convergence condition is satisfied. For large databases a sparse representation is essential for **M**.

The MM approach converges fastest when only a small number of discrete paths make a significant contribution to $k_{\mathrm{AB}}^{\mathrm{SS}}$ and $k_{\mathrm{BA}}^{\mathrm{SS}}$. In particular, at low temperatures only paths corresponding to the lowest overall potential energy barrier are likely to contribute. This is the regime where the master equation and KMC approaches become less efficient, and the MM method was therefore used to obtain low temperature rates in the early DPS studies [7]. However, when numerous long paths

make an important contribution to the rate, the MM calculation can be slow to converge, and numerical precision may be lost.

The master equation [52, 53] and KMC methods [91, 94, 120–122] provide two alternative approaches to analysing the global kinetics of a stationary point database. In both cases the steady-state approximation and assumption of local equilibrium in the A and B regions are relaxed, providing a useful comparison with k_{AB}^{SS} and k_{BA}^{SS}.

The master equation method [52, 53] involves diagonalisation of a matrix whose dimension equals the number of local minima, N_{min}. The non-zero eigenvalue with the smallest magnitude, λ, corresponds to the slowest relaxation process, which decays as $\exp(-|\lambda|t)$, where $\lambda < 0$. If λ is well separated from the eigenvalue that is next-smallest in magnitude then the sum of the two-state rate constants can be calculated as $k_{AB} + k_{BA} = |\lambda|$. This condition, combined with detailed balance, can be used to extract the individual rate constants. Unfortunately, the diagonalisation process scales with the number of minima as $\mathcal{O}(N_{min}^3)$, and problems of numerical precision may arise at low temperature, even when projection methods are applied [123]. Some of these difficulties can be overcome by grouping and pruning the database [14, 59, 60, 62, 63], which reduces the dimension of the problem. Sets of minima can be grouped together according to various dynamic or thermodynamic criteria to simplify the kinetic analysis; expressions for the corresponding group free energies and rates are given in section 4. In contrast, pruning the database simply refers to the removal of minima that are deemed to be unimportant. For example, recursively removing minima with only one connection usually has little effect on the calculated rate constants.

The KMC runs do not suffer from the same numerical precision issues as master equation calculations. However, the required computer time may increase exponentially as the temperature falls and $k_B T$ becomes small compared to the intervening potential energy barrier between the A and B regions. The problem in this case is that the KMC trajectory revisits minima in the starting region an inordinate number of times before it manages to cross the barrier. In these circumstances the committor probability formulation of k_{AB}^{NSS} and k_{BA}^{NSS} can provide a viable alternative, especially if the minima are regrouped according to an energy threshold [12].

The most powerful approach for kinetic analysis of stationary point databases now appears to be the graph transformation (GT) method [12]. The theory extends previous results for 'leapfrog' moves to second neighbours in KMC calculations [7] and the Monte Carlo with absorbing Markov chains approach [96]. The method can also be viewed as a development of the *n*-fold way [91] where we exclude not only the transitions from the current state to itself, but also those involving adjacent minima. Starting from the complete database, we progressively remove local minima from the I region but renormalise the branching probabilities and waiting times to leave the MFPT between the A and B regions unchanged.

Consider the removal of minimum $i \in I$ for specificity, and a KMC trajectory that arrives at minimum β adjacent to i. If i is removed then the trajectory must continue by stepping to any of the minima adjacent to either β or i. We denote these minima by the set Γ. We must also define new branching probabilities, $P'_{\gamma\beta}$ for all $\gamma \in \Gamma$ and a new waiting time for escape from β, τ'_β, so that the MFTP between A and B is conserved. The new branching probabilities must therefore

account for all possible recrossings involving minimum β before escape to Γ. Hence we obtain [12]

$$P'_{\gamma\beta} = (P_{\gamma i}P_{i\beta} + P_{\gamma\beta})\sum_{m=0}^{\infty}(P_{\beta i}P_{i\beta})^m = (P_{\gamma i}P_{i\beta} + P_{\gamma\beta})/(1 - P_{\beta i}P_{i\beta}). \tag{26}$$

This renormalisation can be applied for any β. Although it changes the branching probabilities out of A and B minima, these are only used when calculating the rate constants for transitions out of the corresponding region. It is easy to show that the new branching probabilities are normalised, so that $\sum_{\gamma\in\Gamma} P'_{\gamma\beta} = 1$.

To calculate the corresponding waiting time in β we replace each branching probability $P_{\alpha\beta}$ by $\widetilde{P}_{\alpha\beta} = P_{\alpha\beta}\exp(\zeta\tau_\beta)$ and use the result that

$$\begin{aligned}\left[\frac{d}{d\zeta}\widetilde{P}_{\alpha_1\alpha_2}\widetilde{P}_{\alpha_2\alpha_3}\widetilde{P}_{\alpha_3\alpha_4}\cdots\widetilde{P}_{\alpha_{n-1}\alpha_n}\right]_{\zeta=0} \\ = P_{\alpha_1\alpha_2}P_{\alpha_2\alpha_3}P_{\alpha_3\alpha_4}\cdots P_{\alpha_{n-1}\alpha_n}(\tau_{\alpha_2} + \tau_{\alpha_3} + \cdots + \tau_{\alpha_n}),\end{aligned} \tag{27}$$

which gives the waiting time multiplied by the corresponding probability factor for any path. Hence we obtain [12]

$$\tau'_\beta = \left[\frac{d}{d\zeta}\sum_{\gamma\in\Gamma}\frac{P_{\gamma i}P_{i\beta}e^{\zeta(\tau_i+\tau_\beta)} + P_{\gamma\beta}e^{\zeta\tau_\beta}}{1 - P_{\beta i}P_{i\beta}e^{\zeta(\tau_i+\tau_\beta)}}\right]_{\zeta=0} = \frac{\tau_\beta + P_{i\beta}\tau_i}{1 - P_{\beta i}P_{i\beta}}. \tag{28}$$

Because the renormalised branching probabilities include the effect of all possible paths involving minimum i the probability associated with any given a $\leftrightarrow$ b path is preserved. Any discrete path that jumps from β to γ acquires a time increment equal to the average value, τ'_β, for escape to minima in Γ, so that the MFPT's corresponding to particular a $\leftrightarrow$ b paths are not conserved. However, it is possible to prove that the average MFPT for paths between the A and B regions does not change under this transformation [12]. If we remove all the I minima, and denote the final renormalised branching probabilities and waiting times by a single prime, then rate constants can be calculated as

$$k_{AB}^{GT} = \frac{1}{p_B^{eq}}\sum_{b\in B}\frac{p_b^{eq}}{\tau'_b}\sum_{a\in A}P'_{ab} \qquad \text{and} \qquad k_{BA}^{GT} = \frac{1}{p_A^{eq}}\sum_{a\in A}\frac{p_a^{eq}}{\tau'_a}\sum_{b\in B}P'_{ba}. \tag{29}$$

These rate constants are not exactly the same as those obtained from the committor probability or KMC formulations discussed above. In particular, τ'_b corresponds to the waiting time for a transition to any of the A or B minima aside from b itself. However, it is also possible to adapt the GT approach to calculate precisely the rate constants corresponding to the KMC approach or k_{AB}^{NSS} and k_{BA}^{NSS} [124].

The GT approach avoids the steady-state approximation for the I minima and the local equilibrium assumption for regions A and B. Most importantly, the computational

expense of the transformation is independent of temperature, and so GT becomes the method of choice at low temperatures, where master equation, committor probability, and KMC calculations all become unfeasible.

If the A and B sets are fixed and include all the minima of interest in these regions then the relative equilibrium occupation probabilities can also be calculated within the superposition approximation. In this case, detailed balance requires $k_{AB}/k_{BA} = p_A^{eq}/p_B^{eq}$. Hence it is only necessary to obtain the rate constant corresponding to one direction from a kinetic analysis. The reverse rate constant can then be obtained from the above equation. In more general cases one or both of the two states might correspond to a high entropy region of configuration space, and we must then consider how many minima we need in the A and B sets. Suppose that we require the rate constant k_{AB} from B to A, and that the A state is a liquid-like phase, or the denatured state of a protein. If a two-state description is applicable then local equilibration in the two states must be fast compared to the A$\leftarrow$B rate, and if we calculate rates for different B minima, they should only differ by a small term corresponding to the local equilibration time in B. Hence we can proceed by calculating rate constants from a chosen set of B minima, and calculate an average by weighting the terms k_{Ab} according to p_b^{eq}/p_B^{eq}. In practice, the operational test of convergence for an individual k_{Ab} contribution would be to sample more discrete paths until the result satisfies some predefined condition. Similarly, to test the convergence of k_{AB} we can include more members of the B set in the average until it appears to be stable. To converge each k_{Ab} contribution we need to include sufficient discrete paths and A minima to converge the average value.

The DPS sampling described above, which aims to produce a kinetically relevant database of stationary points, should be clearly distinguished from the schemes that involve explicit dynamics via transition path sampling [5, 6]. In the latter approach, it is possible to accept and reject proposed moves to new dynamical paths using a Metropolis criterion. Importance sampling of this kind cannot be used in DPS calculations because analytical results are not available for the probability of locating different transition states using the methods described in section 3. It is also important to emphasise that the advantage of DPS in focusing on stationary points relies upon the Markov assumption for the minimum-to-minimum dynamics, and upon the statistical rate theory and models for the vibrational density of states. It is these approximations that uncouple the calculations from the short time scale corresponding to vibrations, and enable us to use geometry optimisation techniques instead of explicit dynamics to overcome large potential energy barriers.

3. Characterising stationary points and pathways

To exploit the coarse-grained approach described above, we clearly need efficient methods to locate the required stationary points on the PES and link them together. All the DPS studies discussed in the following sections employed Nocedal's L-BFGS algorithm [125, 126] to approximate the two unique downhill steepest-descent paths that usually connect a transition state to two local minima. These paths are constructed by minimisation of the potential energy following infinitesimal positive and negative

displacements along the Hessian eigenvector corresponding to the unique negative Hessian eigenvalue at the transition state. Energy minimisation is generally much faster than methods that follow the steepest-descent paths accurately. However, some of the approximate paths can actually lead to different local minima, because they deviate from the true steepest-descent trajectories. In particular, the larger steps taken during L-BFGS minimisation may cause the calculation to skip over a low barrier separating the correct minimum from an adjacent one. Since the number of such events is small, and the extra step that is missed out corresponds to a low barrier, the effect on the calculated rate constants is not usually significant.

Locating transition states is generally harder than characterising local minima, because the stationary point is balanced on a knife-edge in one degree of freedom. The DPS results described below employed two types of transition state search, namely single-ended, where only a starting configuration is specified, and double-ended, where we try to link two particular endpoint minima. In fact, all transition state candidates from double-ended searches are refined using hybrid eigenvector-following [127, 128], which is also used in the single-ended searches. The eigenvector and smallest nonzero Hessian eigenvalue that define the uphill search direction were either obtained iteratively from the analytic Hessian, avoiding full diagonalisation, or from a variational approach where the Hessian is not required [129]. In the latter method we define the Rayleigh-Ritz ratio as

$$\lambda(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{H} \mathbf{x}}{\mathbf{x}^2}, \tag{30}$$

where $\mathbf{x}$ is the displacement from the current configuration, superscript ^{T} denotes the transpose and $\mathbf{H}$ is the Hessian matrix. We then minimise $\lambda(\mathbf{x})$ after rewriting the ratio as [69, 127]

$$\lambda(\mathbf{x}) \approx \frac{V(\mathbf{X}+\xi\mathbf{x}) + V(\mathbf{X}-\xi\mathbf{x}) - 2V(\mathbf{X})}{(\xi\mathbf{x})^2}, \tag{31}$$

where $V(\mathbf{X})$ is the potential energy at configuration $\mathbf{X}$ and $\xi \ll 1$. The gradient is then

$$\frac{\partial\lambda(\mathbf{x})}{\partial\mathbf{x}} = \frac{\nabla V(\mathbf{X}+\xi\mathbf{x}) - \nabla V(\mathbf{X}-\xi\mathbf{x})}{\xi\mathbf{x}^2} - \frac{2\lambda(\mathbf{x})\mathbf{x}}{\mathbf{x}^2}. \tag{32}$$

To prevent the minimisation from arriving at a solution corresponding to overall rotation or translation, these directions are projected out using the known analytical forms for the corresponding eigenvectors [14, 130, 131]. An uphill step is then taken along the chosen eigenvector, followed by L-BFGS minimisation in the tangent space [127, 128]. The minimisation consists of a small, fixed number of steps, so that the uphill direction does not change too much, thus avoiding collapse to a minimum.

The magnitude of the uphill step is given by

$$x = \frac{\pm 2g}{|\varepsilon^2|(1+\sqrt{1+4g^2/\varepsilon^4})}, \tag{33}$$

where g is the component of the gradient along the eigenvector in question, and ε^2 is the corresponding Hessian eigenvalue [132–134]. We also adopt a trust radius scheme [134–137] for the maximum step size for this direction. The gradient at the current point, n, and the previous point, $n-1$, are used to estimate the Hessian eigenvalue as

$$\varepsilon^2_{\text{est}} = \frac{g_n - g_{n-1}}{x_{n-1}}, \tag{34}$$

where x_{n-1} is the step length for cycle $n-1$, etc. The trust ratio is defined as

$$r = \left| \frac{\varepsilon^2_{\text{est}} - \varepsilon^2_n}{\varepsilon^2_n} \right|, \tag{35}$$

and the maximum step is changed depending upon whether r is greater or less than a specified trust radius [134–136].

For the peptides in section 5.3 all the geometry optimisations were performed using internal coordinates according to the linear scaling scheme of Németh *et al.* [138, 139]. For calculations in Cartesian coordinates with explicit Hessians the eigenvalues corresponding to overall translation and rotation were shifted to the top of the spectrum [14]. Double-ended searches between specified local minima were performed using the doubly nudged elastic band (DNEB) approach [140], which builds upon previous strategies involving a chain of states [65, 141–146]. The chain consists of images, which correspond to a discretised representation of the geometry along a path through configuration space that connects the two endpoints. Each image experiences a force derived from the gradient of the potential energy and a contribution corresponding to harmonic springs that connect adjacent minima. In particular, the DNEB method retains a portion of the spring gradient perpendicular to the path. With this modification L-BFGS minimisation can be applied to the images in the band until they are reasonably stable. Images that lie above their neighbours are then selected as starting points for hybrid eigenvector-following calculations with tight convergence conditions.

For local minima that are widely separated in configuration space, the shortest discrete path with the fewest steps may involve dozens, or even hundreds, of intervening transition states and the corresponding local minima. Furthermore, we cannot hope to find all of these transition states in one DNEB calculation. Hence a strategy is required for guiding successive DNEB searches by judicious choice of the next pair of local minima to connect. This procedure has been automated using Dijkstra's shortest path algorithm [109, 110] by defining a suitable weight (or cost) function between every pair of known minima. If we know a transition state that connects the pair, then the weight is zero, otherwise a weight is calculated as a monotonically increasing function of the minimum Euclidean distance between the points [147]. Dijkstra's algorithm is then used to find the discrete path with the minimum total weight, and double-ended transition state searches are initiated for every missing connection in this path. The Dijkstra connection procedure enables an initial discrete path to be located between members of the A and B states in a single call to `OPTIM` [108].

4. Disconnectivity graphs

Disconnectivity graphs [14, 65, 148–150] have played a key role in recent efforts to explain how diverse processes such as protein folding, crystallisation, self-assembly, and the appearance of 'magic number' clusters occur via directed searches of the PES. In each case the probability of finding the structure in question via a random search is inconsistent with the experimental time scales observed, as Levinthal pointed out in the field of protein folding [151]. However, if the PES is organised in the right way, the system is effectively guided to the structure in question. This organisation can be visualised using disconnectivity graphs.

At a given total potential energy, V, a connected database of local minima can be partitioned into disjoint sets, or 'superbasins' [148]. The members of each set can be interconverted via one or more transition states without exceeding the threshold V, while the lowest transition state on a discrete path between members of different sets lies above this energy. The superbasin analysis is usually performed at a regularly spaced set of potential energies, $V_1 < V_2 < V_3 < \cdots$. In the corresponding disconnectivity graph the potential energy increases on the vertical axis, while the horizontal axis is usually arbitrary, although it can be used to reflect properties of the minima [152]. As V increases more and more local minima become accessible. We draw a vertical line upwards from the energy corresponding to each local minimum to connect it with the appropriate superbasin node at the next superbasin energy. Lines are joined together at superbasin analysis energy V_n if the corresponding minima lie in the same superbasin at that energy. A minimum lies in its own superbasin until V_n exceeds the potential energy of the lowest transition state that connects it to a different minimum.

The line that terminates at the bottom of the graph corresponds to the global potential energy minimum of the system. As the potential energy increases more local minima become accessible, and their branches merge together when they can interconvert via a discrete path where the highest transition state lies below the superbasin analysis energy. If this threshold is raised high enough then eventually all the branches will merge into one at the top of the graph, because all the local minima can interconvert. The energy interval at which the superbasin analysis is carried out affects the form of the graph. If it is too large then interesting structure will be washed out; if it is too small then the graph may contain too much detail to provide any insight into topological features of the landscape.

Figure 1 illustrates three archetypal potential energy landscapes [149]. Since each graph splits into two parts if an edge is cut, they are technically classified as tree

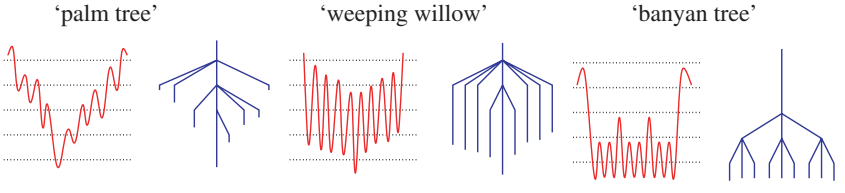

Figure 1. One-dimensional potential energy functions (left) and the corresponding disconnectivity graphs (right). The dotted lines indicate the energies at which a superbasin analysis was performed.

graphs [155]. The 'palm tree' pattern corresponds to a surface with a well-defined global minimum and relatively small downhill barriers. This is the motif that we associate with good 'structure seekers', where efficient relaxation to the global minimum is expected over a wide range of temperature. It provides a direct visualisation of the 'folding funnel', first envisaged for proteins as a set of kinetically convergent pathways [156]. In contrast, for the 'willow tree' pattern the downhill barriers are large compared to the potential energy difference between successive minima. This motif is realised in the PES of C_{60} [149, 157]. Efficient relaxation to the global minimum is still possible if the system has enough thermal energy to overcome the barriers, since there is a potential energy gradient to 'guide' the system downhill [158–160]. The 'banyan tree' motif is qualitatively different from the others, since whole sets of minima can be disconnected by cutting certain edges. This hierarchical structure results from barrier heights on two or more different energy scales.

The calculations of global thermodynamics and kinetics do not depend upon the disconnectivity graph approach. However, it may be possible to make predictions from the form of a graph without further calculations. In particular, it is the palm tree graph that may help to unify the nonrandom searches that result in protein folding, crystallisation, 'magic number' clusters in mass spectra, and self-assembly. By choosing an order parameter based upon structural similarity to the global minimum we can produce a free energy surface with 'funneling' properties [156, 161–172]. We can also identify a landscape supporting a single potential energy funnel with the principle of 'minimal frustration' [161, 162, 166, 168]. Here, frustration would correspond to alternative low-lying minima separated from the global minimum by large barriers, i.e. to separate potential energy funnels. The single potential energy funnel should also produce a large value for the ratio T_f/T_g, where T_f is the 'folding' temperature, below which the potential energy and free energy global minima coincide, and T_g is the 'glass' temperature, at which relaxation slows down below some given time scale [173, 174]. For a large value of T_f/T_g the PES supports a global free energy minimum that is kinetically accessible over a wide range of temperature [14, 161, 175, 176].

Two particular examples of the palm tree motif are shown in figures 2 and 3. The first graph is for a bulk model of silicon that employs the Stillinger-Weber potential [153] and uses periodic boundary conditions for a supercell containing 216 atoms. Here we associate nucleation and growth of the crystal with the palm tree structure. The second example illustrates self-assembly of an icosahedral shell from twelve rigid pentamers [150]. Each pentamer has five attractive sites around the equator, and a single repulsive site at the apex. This simple intermolecular potential was designed to study the minimal conditions required for a hollow icosahedral shell to become a kinetically accessible global minimum. Disconnectivity graph calculations revealed that palm tree landscapes occur for pyramids that are not too flat, and not too spiky [150]. This model provides a crude representation of a simple virus capsid. It is interesting to note that Crick and Watson first suggested that such capsids might be composed of repeated subunits on the basis of the limited genetic information that could be contained in the shell [177]. The disconnectivity graph analysis further suggests that capsid formation may also exploit the properties of a funnelled potential energy landscape to achieve self-assembly [150].

Surfaces with two well-defined potential energy palm tree structures correspond to systems with competing morphologies, which can also be considered in terms of 'frustration'. Examples will be illustrated for atomic clusters in section 5.1. Such landscapes often lead to a separation of time scales for relaxation, as for previous lattice models [178–180]. The first fast component corresponds to direct relaxation to the global

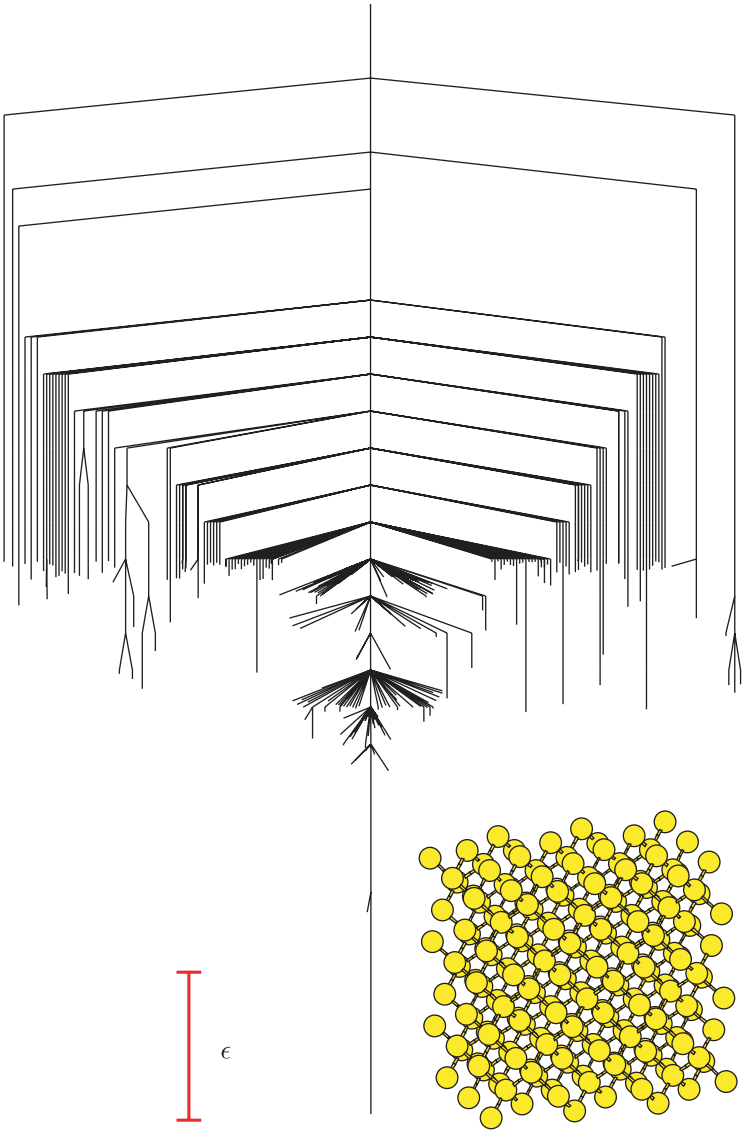

Figure 2. Disconnectivity graph containing the lowest 850 minima in the neighbourhood of the Stillinger-Weber silicon [153] crystal [154], which is illustrated at the bottom. The supercell contains 216 atoms, and ϵ is the pair well depth.

free energy minimum, whilst the second slow component corresponds to trajectories that are first trapped in the competing structure, and subsequently escape. In some cases one region of the PES is favoured by potential energy, and the other by entropy. As the temperature increases from absolute zero, the global free energy minimum will change, and the system exhibits a first-order-like phase transition, where the singularity in the heat capacity is rounded in finite systems [14, 181, 182]. In particular, the change in the global free energy minimum from one morphology to another may be marked by

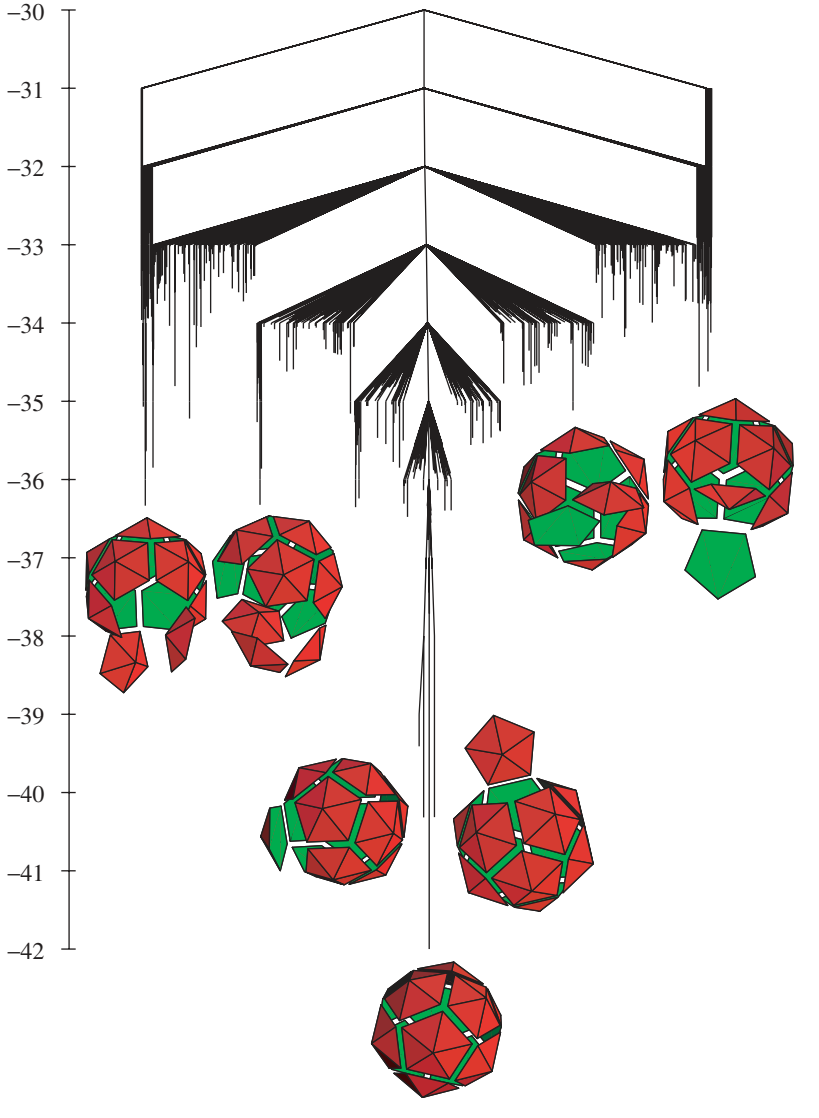

Figure 3. Disconnectivity graph for an icosahedral shell composed of twelve pentagonal pyramids [150]. Some of the low-lying minima are illustrated close to their corresponding branches in the graph. The energy is in reduced units [150].

a feature in the heat capacity [48, 49, 183, 184]. Additional palm tree features can lead to further relaxation components, and eventually to a glassy landscape, with a large number of low-lying structures all separated by relatively high barriers. Numerous examples have been described, including a three-colour, 46-bead model protein [59, 185–187] and binary salt clusters [188, 189].

Disconnectivity graphs have also been presented using enthalpy [190] and free energy [59, 191]. To use free energy we must consider how to treat groups of minima in superbasins and their connections. One approach is to use the mincut-maxflow theorem [192] empoyed in analysing network flows [191]. An alternative is to extend the superposition approach to consider groups of transition states [59, 63]. Consider two groups of local minima, J and L. The occupation probability and free energy of group J, for example, are written as

$$p_{\mathrm{J}}^{\mathrm{eq}}(T) = \sum_{\mathrm{j}\in\mathrm{J}} p_{\mathrm{j}}^{\mathrm{eq}}(T) \quad \text{and} \quad F_{\mathrm{J}}(T) = -k_B T \ln \sum_{\mathrm{j}\in\mathrm{J}} Z_{\mathrm{j}}(T). \tag{36}$$

The relative free energy of the set of transition states that connect a minimum in group J with a minimum in group L is given by

$$F_{\mathrm{JL}}(T) = -k_B T \ln \sum_{\mathrm{j}\in\mathrm{J}\leftrightarrow\mathrm{l}\in\mathrm{L}} Z_{\mathrm{jl}}(T), \tag{37}$$

where $Z_{\mathrm{jl}}(T)$ is the partition function for the transition state connecting l to j, with the degree of freedom corresponding to the reaction coordinate removed, as in equation (22). The sum is over all L to J connections and the rate constant from J to L becomes [193]

$$k_{\mathrm{LJ}}(T) = \sum_{\mathrm{j}\in\mathrm{J}\leftrightarrow\mathrm{l}\in\mathrm{L}} \frac{p_{\mathrm{j}}^{\mathrm{eq}}(T)}{p_{\mathrm{J}}^{\mathrm{eq}}(T)} k_{\mathrm{lj}}(T) = \frac{k_B T}{h} \exp(-[F_{\mathrm{JL}}(T) - F_{\mathrm{J}}(T)]/k_B T), \tag{38}$$

where $k_{\mathrm{lj}}(T)$ is the rate constant for a transition from minimum j to minimum l via a particular transition state. If there is more than one transition state connecting a given pair then we simply add the contributions. The entropy appears in these expressions both through the number of minima contributing to each region, and through the vibrational densities of states.

5. Results

5.1. *Lennard-Jones clusters*

The Lennard-Jones (LJ) potential is defined as [194]

$$V = 4\epsilon \sum_{i<j}^{N} \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right], \tag{39}$$

where ϵ is the pair well depth, $2^{1/6}\sigma$ is the equilibrium pair separation, and r_{ij} is the distance between atoms i and j. Clusters of N Lennard-Jones atoms, denoted LJ_N, display a rich variety of behaviour, and serve as benchmarks for global optimisation [33, 195–198], and analysis of thermodynamics [181, 182, 199] and dynamics [1, 2, 7, 51, 200, 201] in finite systems. For example, the permutational isomerisation of a two-dimensional LJ_7 cluster provided one of the first applications of the transition path sampling approach [200], which in turn enabled benchmark comparisons for DPS calculations [2].

Most LJ clusters composed of a hundred atoms or fewer have global potential energy minima that are based on icosahedral packing via an underlying Mackay icosahedron [202]. However, at certain sizes, where complete high-symmetry structures occur for alternative packing schemes, and the best incomplete Mackay icosahedron is relatively high in energy, different morphologies are observed [33, 203–208]. The corresponding potential energy landscapes exhibit a low entropy region of configuration space containing the global minimum, and a high entropy region containing icosahedrally based structures. The global minimum is therefore relatively difficult to locate, because it lies at the bottom of a narrow potential energy funnel, separated from the funnel corresponding to the much more numerous structures based on icosahedral packing by a large barrier. This potential energy barrier scales extensively with system size, so that ergodicity-breaking effects grow stronger for larger systems of this type. The change in global free energy minimum from a low potential energy to a high entropy region of configuration space often produces a small heat capacity feature for the corresponding solid-solid first-order-like transition [48, 49, 183, 184].

The LJ_{38} cluster exhibits a double-funnel PES [14, 48, 49, 149, 183, 184] where the global potential energy minimum is a truncated octahedron based upon face-centred-cubic (fcc) packing (figure 4). There is a heat capacity feature for this cluster at $k_BT/\epsilon \approx 0.11$ [48, 49, 183, 184], where the free energy global minimum changes from the fcc to the icosahedral region. The interconversion rates between the icosahedral (I_h) and fcc (O_h) regions of configuration space were calculated using DPS with fixed definitions of the A and B regions containing 5 (O_h) and 395 (I_h) minima. These sets were taken from a previous master equation study of the dynamics [183], which employed a stationary point database containing 6,000 minima and 8,633 transition states. The assignment of the A and B states was based upon the master equation eigenvector corresponding to the slowest relaxation, which contains coefficients for each minimum. The magnitude of the coefficient indicates how strongly the corresponding minimum is involved in this relaxation, and the sign enables us to separate the minima in different regions, since it indicates the direction of the flux. Recalculating both master equation and DPS rates for the new database as a function of the cutoff value used in assigning the I_h and O_h minima revealed good agreement over a wide range of values. In particular, DPS sampling located a path with a larger contribution than any obtained in the previous master equation study [183]. Agreement only begins to break down when the cutoff used to assign the minima becomes too large, so that structures with significant equilibrium occupation probabilities are included in the intervening set [2]. An Arrhenius fit, $k = a\exp(-\Delta/k_BT)$, to the rates calculated in the temperature

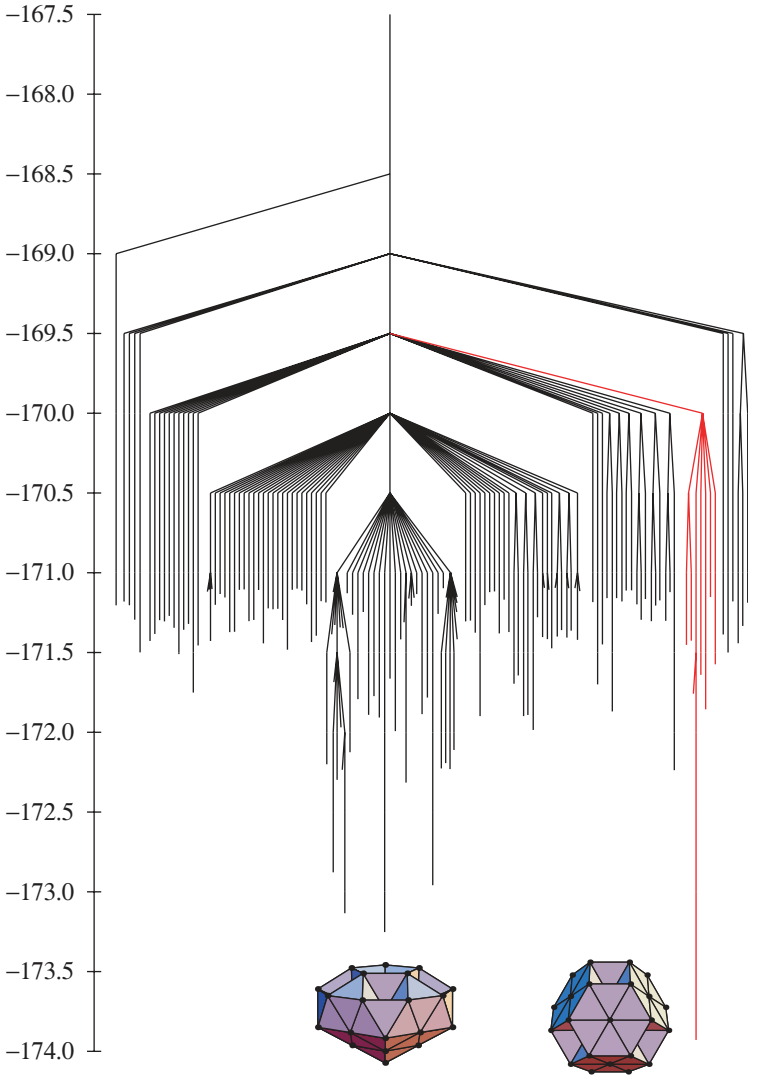

Figure 4. Disconnectivity graph for the LJ_{38} cluster including the lowest 200 local minima; permutation-inversion isomers are not distinguished. The global potential energy minimum lies at the bottom of the 'palm tree' feature highlighted in red. The lowest minimum based upon an incomplete Mackay icosahedron is also illustrated below the corresponding branch of the graph. The energy is in ϵ.

range $0.03 \leq k_B T/\epsilon \leq 0.4$ has coefficient of determination $R^2 = 0.9997$ with [7]

$$\begin{aligned} k_{I_h \leftarrow O_h}: &\quad a = 2.13 \times 10^6\, \nu_{\mathrm{LJ}}, \quad \Delta = 4.29\, \epsilon, \\ k_{O_h \leftarrow I_h}: &\quad a = 1.16 \times 10^3\, \nu_{\mathrm{LJ}}, \quad \Delta = 3.43\, \epsilon, \end{aligned} \tag{40}$$

where $\nu_{\mathrm{LJ}} = \sqrt{\epsilon/M\sigma^2}$ is the reduced unit of frequency, with M the atomic mass. The DPS calculation used to produce this fit was performed at $k_BT/\epsilon = 0.1$, and runs at different temperatures were found to give essentially the same results [2]. A two-state description is appropriate throughout this temperature range [183], as evidenced by the master equation eigenvalue spectrum and a net flow index analysis [51, 209]. For parameters appropriate to argon we find a rate of $55\,\mathrm{s}^{-1}$ at $T = 13.7\,\mathrm{K}$, where $k_{I_h \leftarrow O_h} \approx k_{O_h \leftarrow I_h}$. At $k_BT/\epsilon = 0.09$ one particular discrete path contributes about 45% of the total rate. When the temperature is raised to $k_BT/\epsilon = 0.12$ this contribution falls to about 23% while a two-step path contributes 6%. This result shows how discrete paths with higher potential energy barriers but more favourable entropies become increasingly competitive at higher temperature.

The next largest LJ cluster to exhibit a double-funnel landscape is LJ_{75}, where the global minimum is a Marks decahedron [212] (figure 5). This system exhibits a heat capacity feature at around $k_BT/\epsilon = 0.08$ [34, 47] where the global free energy minimum switches from the decahedral (D_{5h}) region to the icosahedral (I_h) region. For this cluster the fixed A and B sets were assigned based on a disconnectivity graph analysis of a stationary point database produced in earlier work [210]. The icosahedral and decahedral sets contained 909 and 91 minima, respectively. A new disconnectivity graph constructed following DPS runs is shown in figure 5. In this cluster the most important single discrete path contributes virtually 100% of the rate at $k_BT/\epsilon = 0.02$, which decreases to 48% at $k_BT/\epsilon = 0.13$, and 17% at $k_BT/\epsilon = 0.2$. An Arrhenius fit to the calculated rates in the range $0.02 \leq k_BT/\epsilon \leq 0.3$ has coefficient of determination $R^2 = 0.9995$ with [7]

$$\begin{aligned} k_{I_h \leftarrow D_{5h}}: &\quad a = 4.11 \times 10^8\, \nu_{\mathrm{LJ}}, \quad \Delta = 10.88\,\epsilon, \\ k_{D_{5h} \leftarrow I_h}: &\quad a = 5.97 \times 10^1\, \nu_{\mathrm{LJ}}, \quad \Delta = 9.61\,\epsilon. \end{aligned} \tag{41}$$

The largest potential energy barriers on the dominant path are 9.60 and $10.78\,\epsilon$, respectively, which are very close indeed to the values in the Arrhenius fit. The forward and backward rates from the fit are equal at $k_BT/\epsilon = 0.081$ where the value is about $5 \times 10^{-39}\mathrm{s}^{-1}$ for parameters appropriate to argon.

Permutational isomerisation has been considered for centre-to-surface atom migration in the 'magic number' Mackay icosahedral [202] clusters LJ_{13} and LJ_{55}, for which the global minima both have point group I_h. Figures 6 and 7 illustrate the effect of tagging one of the 13 atoms in LJ_{13} on the disconnectivity graph. Tagging makes one particular atom distinguishable from the others, which affects the point group assignment. Both graphs include 1,467 local minima, but in figure 6 all $2 \times 13!/120$ distinct permutation-inversion isomers of the icosahedron are included in the single branch corresponding to the global minimum. This graph clearly exhibits the palm tree motif, with a single potential energy funnel. The gap between the global minimum and the next-lowest structure is $2.64\,\epsilon$. In contrast, in figure 7 we distinguish $2 \times 12!/120$ minima with the tagged atom in the middle from the $2 \times 12!/10$ C_{5v} minimum with the tagged atom on the surface. Since $2 \times 12!/120 + 2 \times 12!/10 = 2 \times 13!/120$ the number of distinct minima is unchanged, but the net effect is that the branches corresponding to the global minimum represent

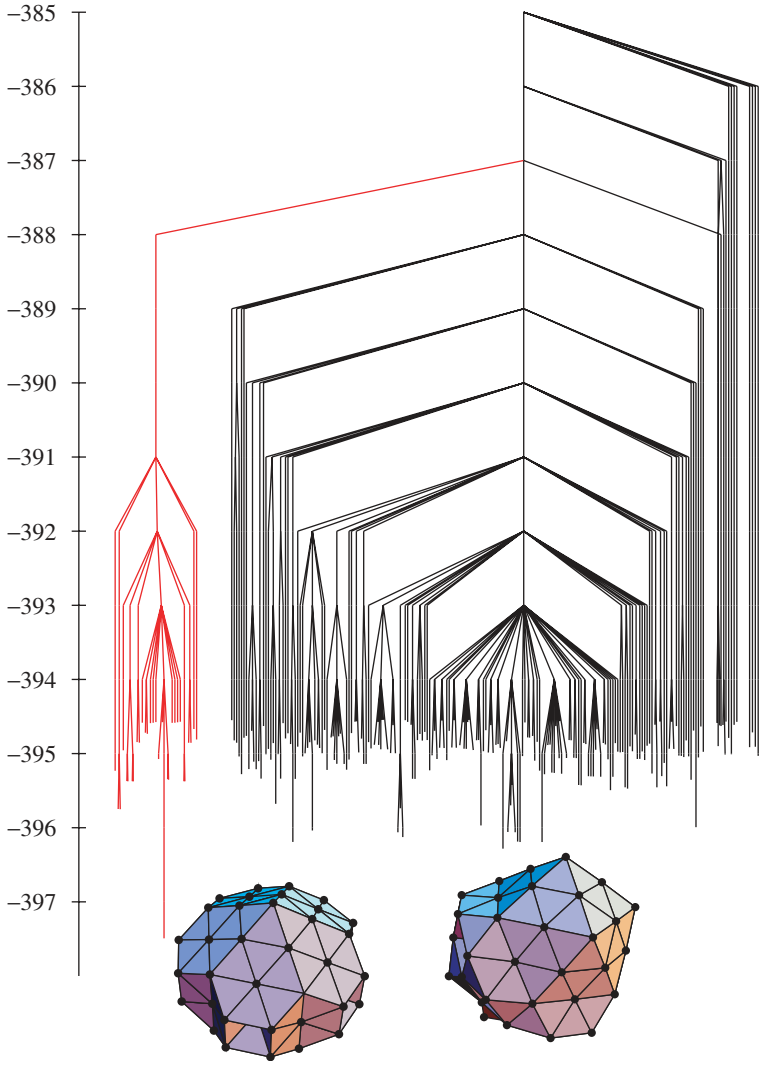

Figure 5. Disconnectivity graph for LJ_{75} in which permutation-inversion isomers are not distinguished. Only the 500 lowest minima are included for clarity, although the full DPS database of stationary points was used to establish the connectivities. Minima assigned to the decahedral region of configuration space are highlighted in red. The global minimum and the lowest energy structure based on icosahedral packing are illustrated next to the corresponding branches in the graph. The energy is in units of ϵ.

different numbers of permutation-inversion isomers in the ratio 1 : 12. In general, the number of distinct minima in figure 7 for a given structure is equal to the number of symmetry-distinct sites for the tagged atom in that configuration.

Relaxation from any high energy minimum to any permutation-inversion isomer of the global minimum is rapid over a wide temperature range for the palm tree surface in figure 6 [7, 56, 149]. If we now tag one atom and start from a non-equilibrium

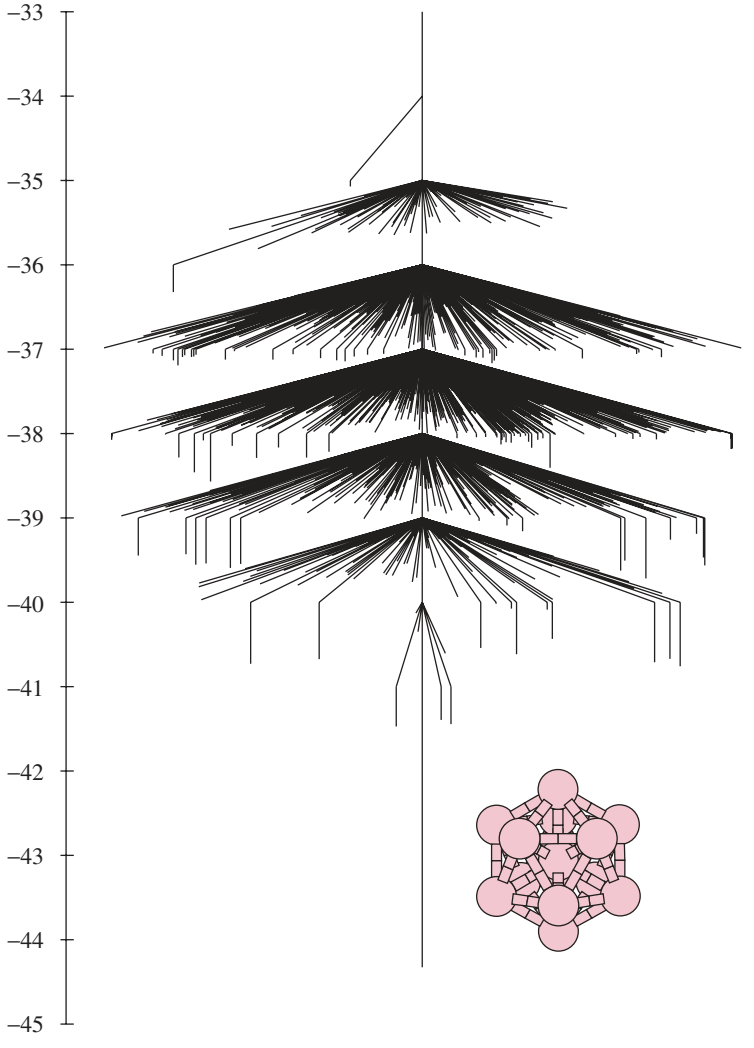

Figure 6. Disconnectivity graph for the LJ_{13} cluster including all the 1,467 local minima identified for this system in previous work [210]. The global minimum is a Mackay icosahedron [202], depicted using Xmakemol [211], while the next-lowest minima correspond to the three distinct capping sites when one atom is removed from the icosahedral shell and placed on the surface. In this graph all permutation-inversion isomers of each structure are grouped together. The energy is in units of ϵ.

distribution, in which only isomers with the tagged atom in the centre are occupied, then the slowest relaxation mode corresponds to transitions between the two funnels in figure 7 [7]. Two-state kinetics are observed over a wide range of temperature, with a clear separation between the slowest relaxation and the next-slowest in the master equation eigenvalue spectrum. The assignment of the slowest mode to relaxation between the two funnels was confirmed using the net-flow index [51, 209] for the eigenvector in question [7].

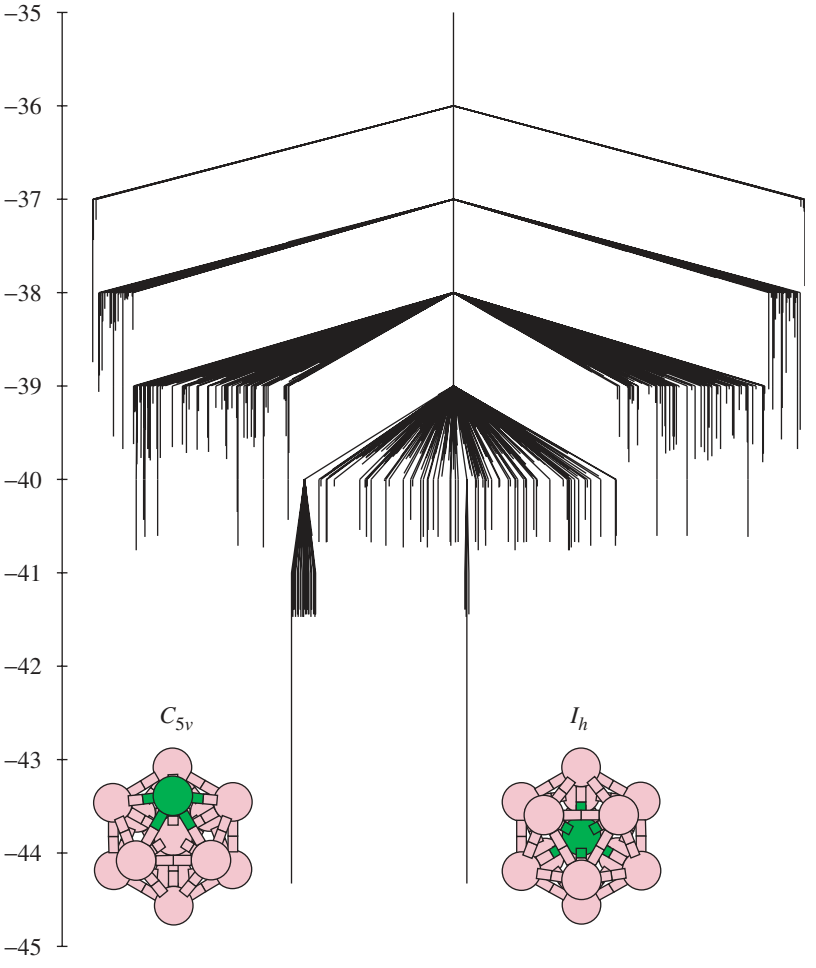

Figure 7. Disconnectivity graph for LJ_{13} in which permutation-inversion isomers of the shaded atom are distinguished. Permutation-inversion isomers of the other twelve atoms are grouped together for every minimum and transition state. The tagged atom can occupy either a central or a surface site in the global minimum, leading to two separate branches in the graph for isomers with I_h and C_{5v} symmetry. The database in question contained 1,608 minima and 6,115 transition states and the graph includes all the minima. The stationary points used to construct this graph were characterised in a DPS calculation of the isomerisation rate between the two distinct isomers of the global minimum [14]. The energy is in units of ϵ.

The rate constants for permutational isomerisation obtained from the MM, KMC and master equation approaches agree quantitatively over the temperature range $0.04 \leq k_B T/\epsilon \leq 0.3$ [7]. An Arrhenius fit has coefficient of determination $R^2 = 0.9996$ for

$$\begin{aligned} k_{\mathrm{AB}}: \quad & a = 1.95 \times 10^3 \, \nu_{\mathrm{LJ}}, \quad V = 4.63\,\epsilon, \\ k_{\mathrm{BA}}: \quad & a = 2.34 \times 10^4 \, \nu_{\mathrm{LJ}}, \quad V = 4.63\,\epsilon. \end{aligned} \tag{42}$$

The prefactors differ by a factor of twelve due to symmetry, and the barrier agrees quantitatively with the potential energy difference between the highest minimum on the most important discrete path. At $k_B T/\epsilon = 0.1$ the centre-to-surface rate corresponds to about $8 \times 10^{-5}\,\mathrm{s}^{-1}$ for parameters appropriate to argon.

LJ_{55} is the next size at which a complete Mackay icosahedron is the global minimum. The low energy region of the landscape again conforms to the palm tree pattern [14, 22, 149, 210], and the energy difference between the global minimum and the next-lowest is $2.64\,\epsilon$. The properties of this cluster have been investigated in many different studies [14, 20–22, 25, 149, 199, 210, 213–230]. There are four symmetry-distinct sites for a tagged atom in the 55-atom Mackay icosahedron: one in the centre, one in the middle shell, and two in the outer shell. Hence there are four different branches corresponding to these distinct isomers of the global minimum in figure 8. The total number of different permutation-inversion isomers for this structure if all the atoms are distinguishable is $2 \times 55!/120$. The branches in figure 8 correspond to $2 \times 54!/10$, $2 \times 54!/4$, $2 \times 54!/10$ and $2 \times 54!/120$ minima, from left to right. Hence the relative equilibrium populations are $12:30:12:1$, or $42:12:1$ for surface : middle shell : centre when we add the contributions for the two surface sites. The forward and reverse rates for centre-to-surface isomerisation therefore differ by a factor of 42 due to symmetry.

The kinetic analysis of the DPS stationary point database proved to be especially difficult for this system, due to the size of the potential energy barrier between the states of interest [7, 12]. For example, at the lowest temperatures considered the barrier is several hundred $k_B T$. Arrhenius fits for this isomerisation were revised in the light of calculations using the GT approach, which agree well with the values obtained for $k_{\mathrm{AB}}^{\mathrm{NSS}}$ and $k_{\mathrm{BA}}^{\mathrm{NSS}}$ when the local minima are regrouped according to a suitable energy threshold [12]. The resulting fit has coefficient of determination, $R^2 = 0.99999$ for

$$\begin{array}{ll} k_{\mathrm{AB}}: & a = 1.38 \times 10^8\,\nu_{\mathrm{LJ}}, \quad \Delta = 14.05\,\epsilon, \\ k_{\mathrm{BA}}: & a = 6.52 \times 10^9\,\nu_{\mathrm{LJ}}, \quad \Delta = 14.04\,\epsilon, \end{array} \tag{43}$$

where B represents the minima with the tagged atom in the surface, and A the minima with the tagged atom in the centre. The value of Δ agrees very well with the potential energy difference between the highest transition state on the lowest-energy discrete path between A and B. The ratio of prefactors deviates a little from 42 because additional minima are included in the A and B sets after regrouping [12]. At $k_B T/\epsilon = 0.1$ the centre-to-surface rate constant has order $10^{-40}\,\mathrm{s}^{-1}$ for parameters appropriate to argon.

5.2. *Water clusters*

Water clusters are of great interest because of the insight they may be able to provide into the fundamental aspects of intermolecular forces, solvation, and the properties of bulk water. As a consequence, the number of different model interaction potentials proposed for water molecules is enormous; here we concentrate on the rigid molecule TIP4P representation [232, 233]. Global optimisation of rigid-body water clusters is significantly more difficult than for most Lennard-Jones clusters with a comparable number of degrees of freedom, due to the interplay between the orientational and

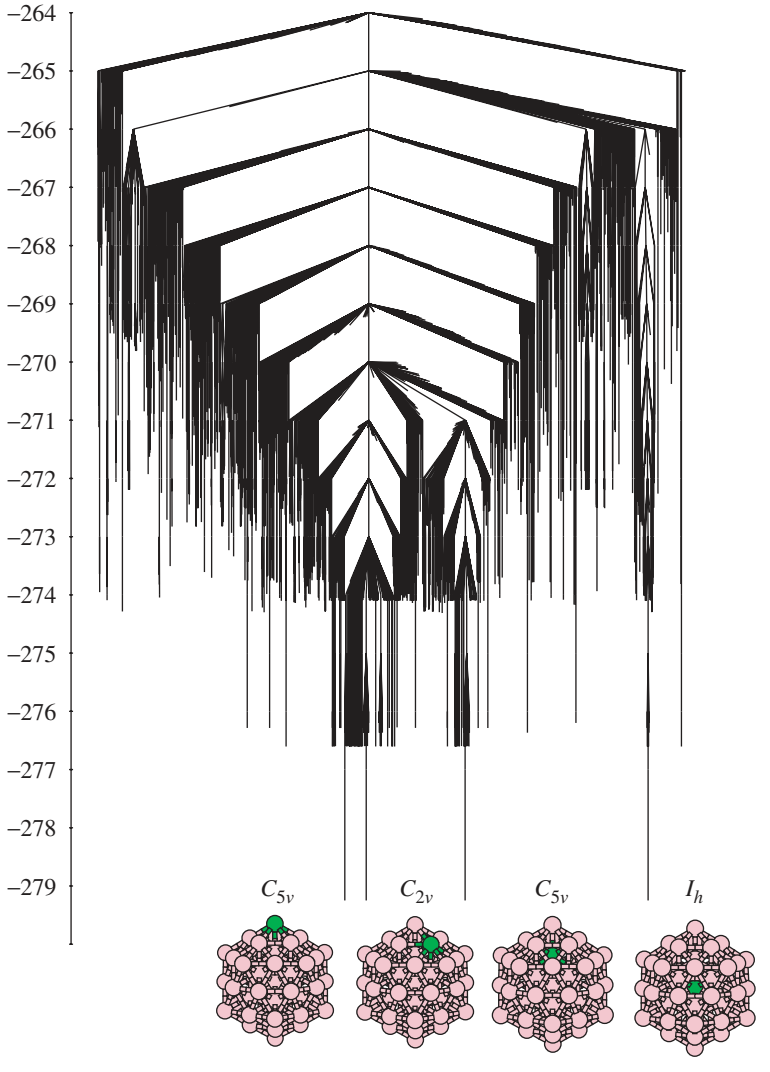

Figure 8. Disconnectivity graph for LJ_{55} in which permutation-inversion isomers of the shaded atom are distinguished. This tagged atom can occupy four distinct sites in the icosahedral global minimum, producing four separate branches. The stationary points used to construct this graph were characterised in a DPS calculation of the isomerisation rate between isomers of the global minimum where the tagged atom occupies central (region A) and surface (region B) sites [7]. The energy is in units of ϵ.

centre-of-mass coordinates [234, 235]. Nevertheless, there is no dispute over the global minima of the TIP4P clusters $(H_2O)_8$, $(H_2O)_9$ and $(H_2O)_{20}$ [234, 236] considered in the DPS studies described below.

For $(H_2O)_8$ there are two particularly low-lying minima with point groups D_{2d} and S_4, each based on a cube of water molecules [223, 237–242]. These structures have also been identified experimentally, both as free clusters [243] and bound to a benzene

molecule [244]. The water octamer has also been characterised in a crystal structure [245]. Microcanonical DPS rate constants were reported for the interconversion of these two isomers for comparison with previous transition path sampling (TPS) calculations [242]. The latter results were obtained for conversion of two specific permutation-inversion isomers, and the corresponding DPS rate constants were therefore considered. At a total energy of −253.13 kJ/mol one particular four-step discrete path was found to contribute 90% of the calculated rate constant. The corresponding path consists of four sequential donor–acceptor exchange rearrangements [240] (figure 9). The next-largest contribution (4%) comes from a three-step path (figure 10), which appears to correspond to the free energy transition state in the transition path sampling study, where rearrangements were restricted to a single face of the cube [242]. The rates calculated for contributions from three-step paths of this kind in the DPS and TPS studies agree to within a factor of about five [2]. Quantitative agreement would not be expected when harmonic densities of states are employed in the DPS calculations along with transition state theory rate constants for the minimum-to-minimum rearrangements.

A further comparison between DPS and TPS rates is available for $(H_2O)_9$. The rearrangement between two low-lying minima corresponds to a single-step discrete path (figure 11), i.e. to a pathway mediated by a single transition state, and there are no other significant contributions to the rate [2]. The transition state in question has an unusual three-centre hydrogen bond in which one hydrogen is coordinated to two different oxygens. The DPS [2] and TPS [246] rates for the fastest path are in quantitative agreement, indicating that harmonic transition state theory works well for this system.

Interconversion of competing morphologies for the $(H_2O)_{20}$ cluster is a much more demanding challenge. Previous disconnectivity graphs for this system revealed a more

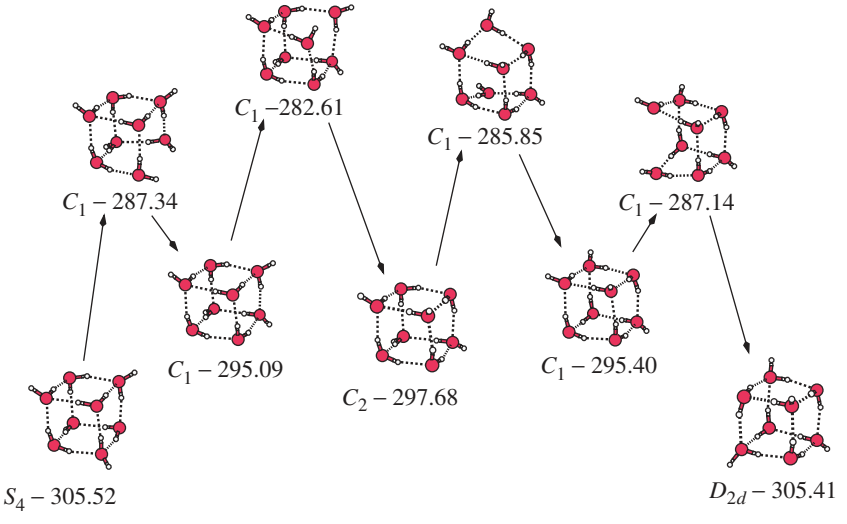

Figure 9. Four-step path linking S_4 and D_{2d} cuboids in the water octamer. The point group symmetries and energies (kJ/mol) for the TIP4P potential [231] are given below each stationary point.

hierarchical structure, reminiscent of the banyan tree pattern [14, 149]. In fact, there are three low-lying morphologies for this cluster separated by large potential energy barriers. The three structures are based on fused cubes (or 'box-kites') [22, 234, 240, 247], stacked pentagonal prisms (sharing pentagonal faces), and square-face-sharing pentagonal prisms (including the global minimum). An equilibrium simulation of this system that includes more than one morphology has not yet been possible, even using parallel tempering [39–41, 248]. The results of a DPS study to link the global minimum with the lowest box-kite structure illustrate this difficulty quite clearly [7]. The disconnectivity graph in figure 12 shows that a large potential energy barrier separates the two morphologies. One of the paths that makes a significant contribution to the rate is illustrated in figure 13. Most of the 19 elementary rearrangements can be described as donor–acceptor exchanges [239]. In this case a significantly faster rate would probably be obtained if we chose to link higher-lying minima belonging to the two morphologies in question. However, the barriers for interconverting minima with different hydrogen-bonding patterns in each class are themselves rather large. This effect illustrates the complexity introduced by the interplay between orientational and centre-of-mass coordinates: minima with essentially the same centre-of-mass coordinates but different hydrogen-bonding arrangements span a wide range of energy [234]. The profile in figure 13 suggests that a steady-state approximation is not likely to be very accurate, due to the presence of low-lying intervening minima. Not surprisingly, the calculated rate constants in the range $30\,\mathrm{K} \leq T \leq 150\,\mathrm{K}$ are very small, with barriers from Arrhenius fits in the region 35 to 40 kJ/mol [7].

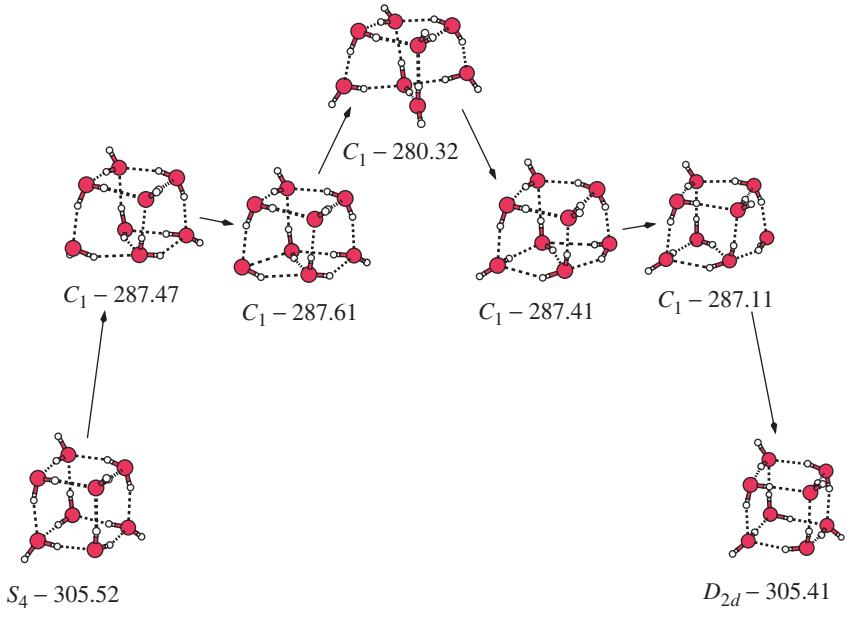

Figure 10. Three-step path linking S_4 and D_{2d} cuboids in the water octamer. The point group symmetries and energies (kJ/mol) for the TIP4P potential [231] are given below each stationary point.

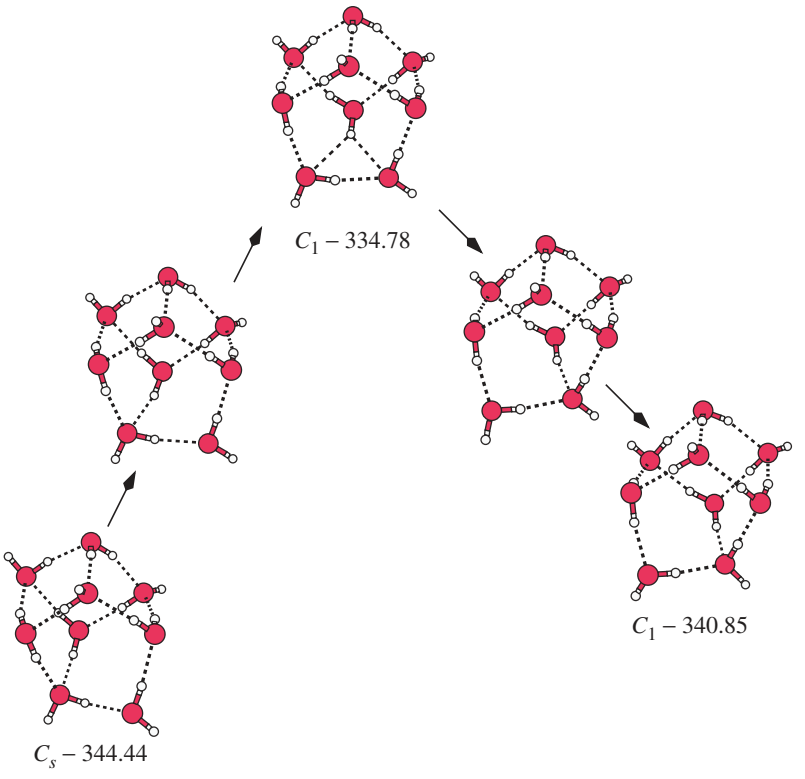

Figure 11. One-step path for isomerisation of $(H_2O)_9$; one intermediate geometry is included on each side of the transition state to clarify the mechanism [2]. The point group symmetries and energies (kJ/mol) for the TIP4P potential [231] are given below each stationary point.

5.3. *Peptides*

DPS studies have recently been performed for two peptides, namely the neurotransmitter pentapeptide met-enkephalin (NH_3^+-Tyr-Gly-Gly-Phe-Met-COO^-) [60] and the 16-amino acid 'GB1' peptide from residues 41–56 (the C-terminal fragment) of the B1 domain of protein G [62]. Both studies employed the CHARMM19 force field [249] and the EEF1 implicit solvation potential [250]. Symmetrised parameters were defined for the dihedral and improper dihedral angle terms in residues ASN, GLN and TYR, to ensure that rotamers have the same energies and geometries [251].

For met-enkephalin experiments indicate that the 'folded' conformation in aqueous solution populates an ensemble of different structures [252]. Numerous thermodynamic studies have used this system for benchmarking new sampling methods [32, 43, 253, 254]. Most of these studies identify two significant structures, namely, the lowest energy ground state, which is a type II′ β turn, and configurations with a type II β turn, which lie higher in energy. A free energy disconnectivity graph is shown in figure 14, where further grouping of minima was performed to simplify the

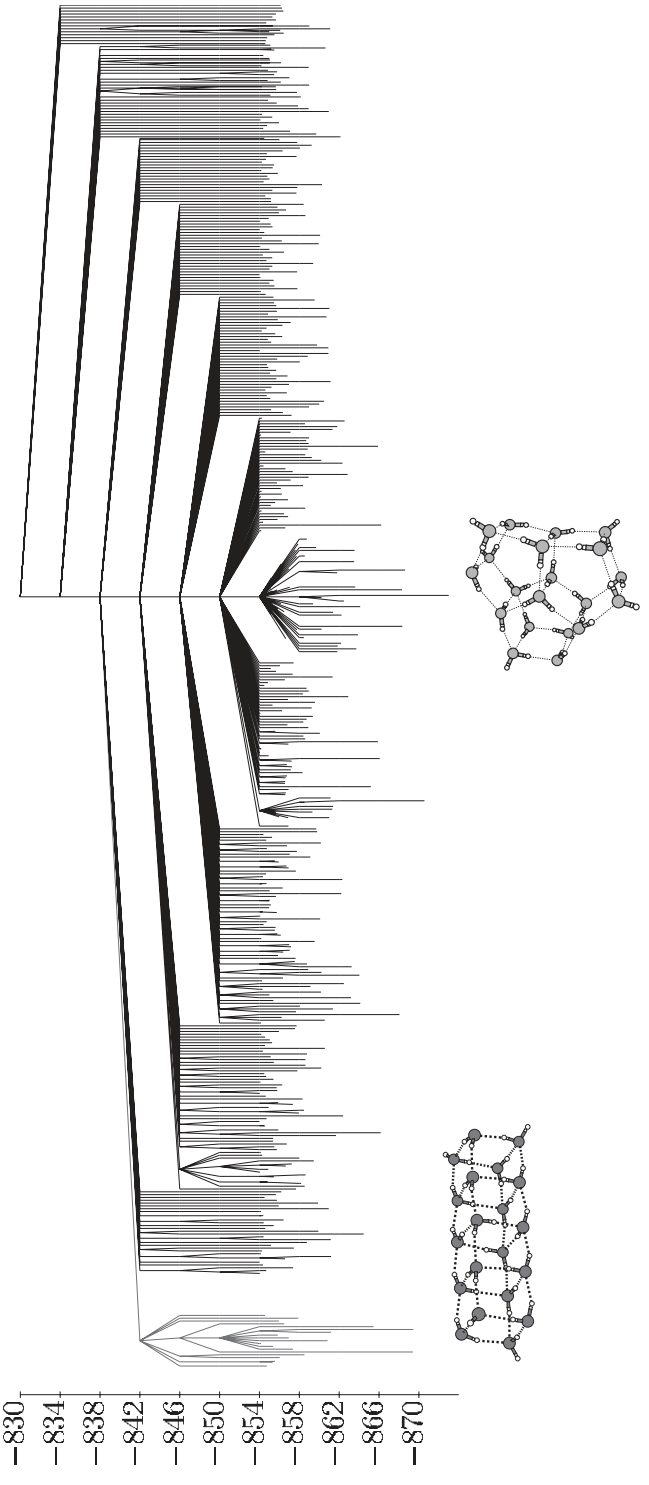

Figure 12. Disconnectivity graph for $(H_2O)_{20}$ modelled by the rigid molecule TIP4P potential. Only the 500 lowest minima are shown for clarity, but every stationary point in the DPS database was included in the disconnectivity analysis. Minima based on the box-kite structure are highlighted in red. The global minimum and the lowest energy structure based upon the box-kite are illustrated next to the corresponding branches in the graph. The energy is in units of kJ/mol.

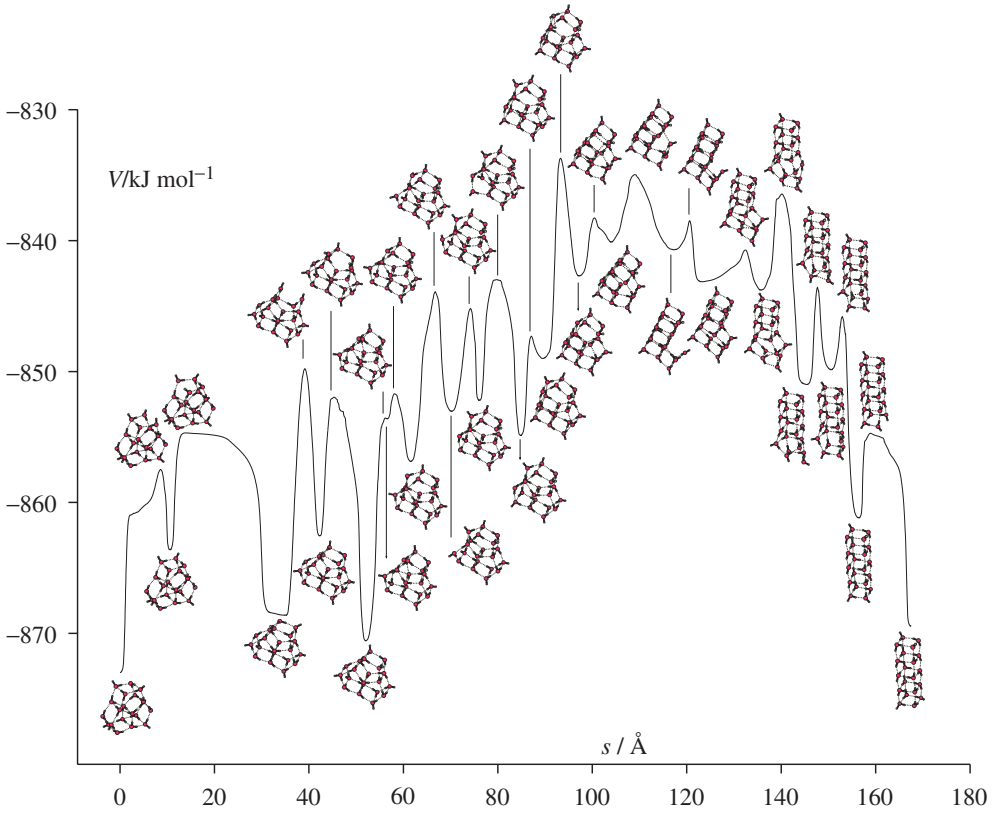

Figure 13. A nineteen-step path linking the TIP4P global minimum of $(H_2O)_{20}$ (left) to the lowest energy box-kite structure [7]. All the structures have point group C_1 except for the final box-kite minimum, which has S_4 symmetry. The graph indicates the potential energy, V, as a function of the integrated path length, s, calculated for approximate steepest-descent paths. The geometries of the transition states and minima are shown above and below this curve, respectively.

appearance [60]. This particular grouping scheme combines local minima if they are connected to a central low energy minimum by transition states that lie below a certain energy barrier [60]. The corresponding regions of configuration space are most likely to be in local equilibrium on the time scale of folding for an appropriate choice of barrier height. Grouping reduced the dimension of the global kinetic analysis problem sufficiently for a master equation analysis to be feasible.

The global potential energy minimum found from basin-hopping [32, 33] has a type II′ β turn, consistent with previous work [32, 255–257]. Both superposition and parallel tempering calculations identified the global free energy minimum at room temperature with configurations based upon the same structure as the global potential energy minimum [60]. Figure 15 shows the occupation probabilities of three particular states on relaxation from an initial non-equilibrium distribution. The groups containing extended configurations and configurations with a type II β turn both exhibit maximum occupation probabilities before the equilibrium distribution is achieved.

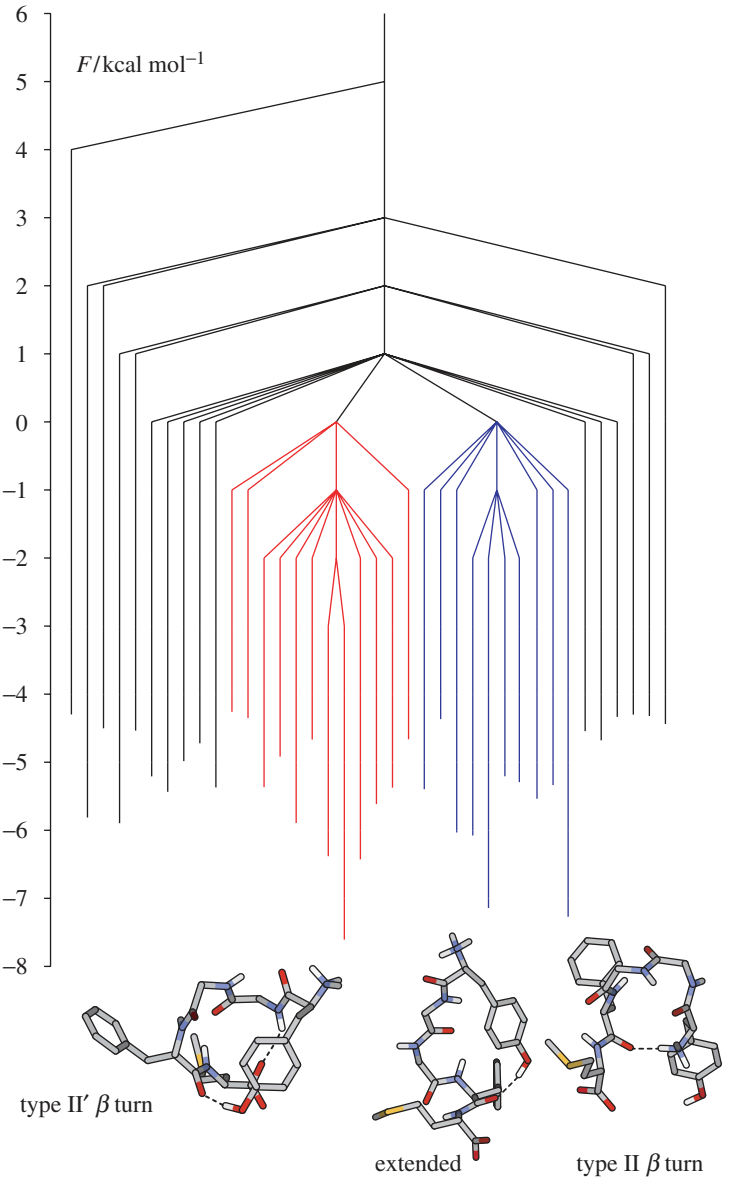

Figure 14. Free energy disconnectivity graph for met-enkephalin at 298 K. Each node represents a group of minima constructed as described in reference [60]. The lowest 38 groups are shown, as these are calculated to contain 90 % of the population. The energy is in units of kcal mol^{-1}. The low energy region of the graph contains two funnels, highlighted in red and blue.

Both could therefore be considered as intermediates on the folding path, although the rate-determining step in a two-state description is between the configurations with β turns. The corresponding DPS folding rate was calculated as $3.1 \times 10^7\,\mathrm{s}^{-1}$ at 298 K [60], in reasonable agreement with previous work [258–260].

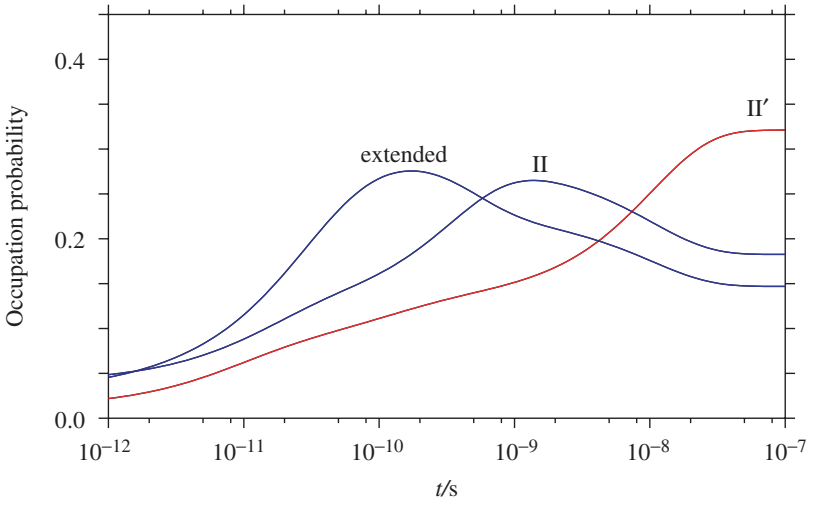

Figure 15. Results of master equation dynamics calculations for met-enkephalin at 298 K starting from a high temperature distribution (800 K) [60].

The GB1 peptide forms a β hairpin in the intact protein [261] and when isolated in solution [262]. Folding experiments based upon tryptophan fluorescence indicate two-state kinetics with a folding rate constant of $1/(6\,\mu s)$ at 297 K [263]. A simple model, where each residue can adopt a random coil or hairpin conformation, suggests that the first step in folding is formation of the turn region [263, 264]. Various simulation studies have also addressed this system [103, 265–271], and have generally found a β hairpin to be the dominant conformation in solution at room temperature, except for one replica-exchange study [272]. However, the equilibrium structure is not rigid, and most simulations find that only about half the native hydrogen bonds from the crystal structure are present at any one time [268, 273, 274], in agreement with NMR data [275]. There is also some evidence that existing empirical force fields may overbind the secondary structure [268].

For this system, the DPS stationary point database was constructed in a series of different runs, applying Dijkstra's algorithm [109, 110] to ensure that the discrete path making the largest contribution to the rate constants was considered. Minima were grouped according to a kinetic scheme based on downhill barrier heights [62], facilitating a master equation analysis, which was compared with the results for the ungrouped minima. The final calculated folding time was between 30 and 90 μs, which is about an order of magnitude slower than the experimental value of 6 μs. This result constitutes reasonable agreement given that harmonic vibrational densities of states are employed, as well as an implicit solvent description. A free energy disconnectivity graph is shown in figure 16, and the occupation probabilities for various groups as a function of time (calculated from KMC) are shown in figure 17. Representative structures from the different groups are also illustrated in these figures.

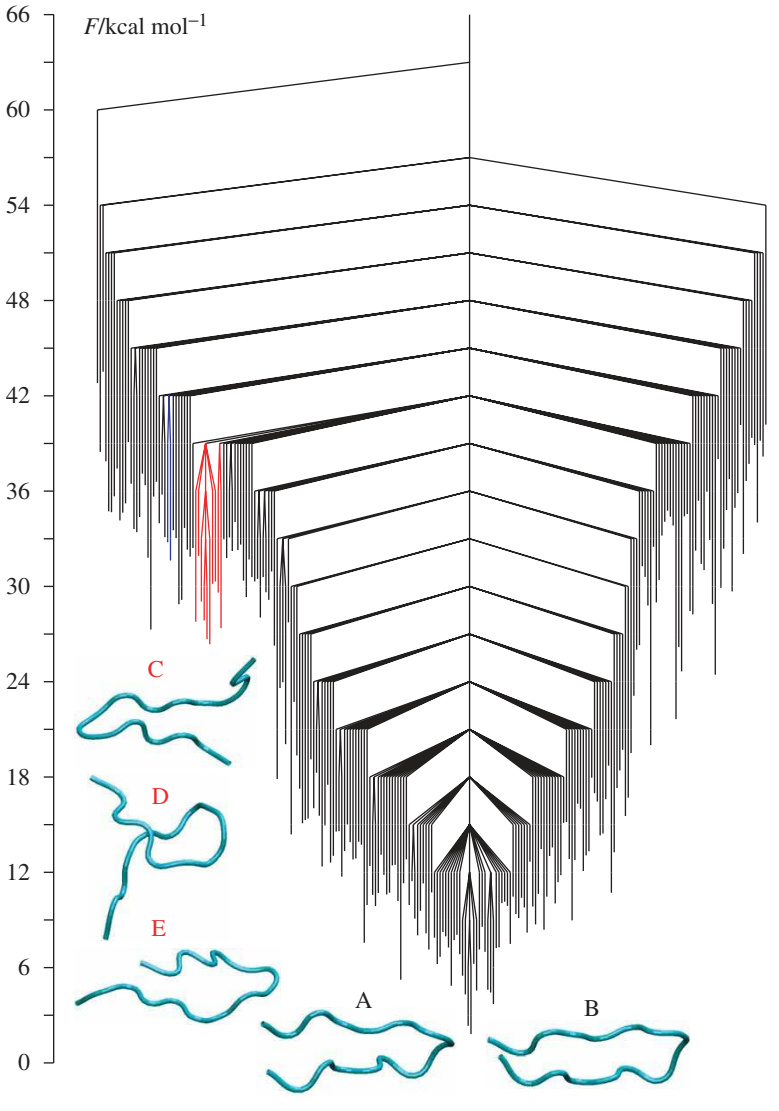

Figure 16. Free energy disconnectivity graph for GB1 at 298 K, grouped as discussed in the text [62].

For this system, an overall two-state description is not appropriate, as there are sets of local minima between the extended and native states that are significantly populated. A detailed analysis of the stationary point database identified two significant intermediates. The first of these states, which dominates up to a time of about 10^{-6} s, consists mostly of minima like F and G, with relatively loose structures. These configurations exhibit significant intramolecular hydrogen bonding, but there is no identifiable hydrophobic core. The second intermediate, whose occupation probability peaks around 10^{-5} s in figure 17, contains more compact structures like C, D and E.

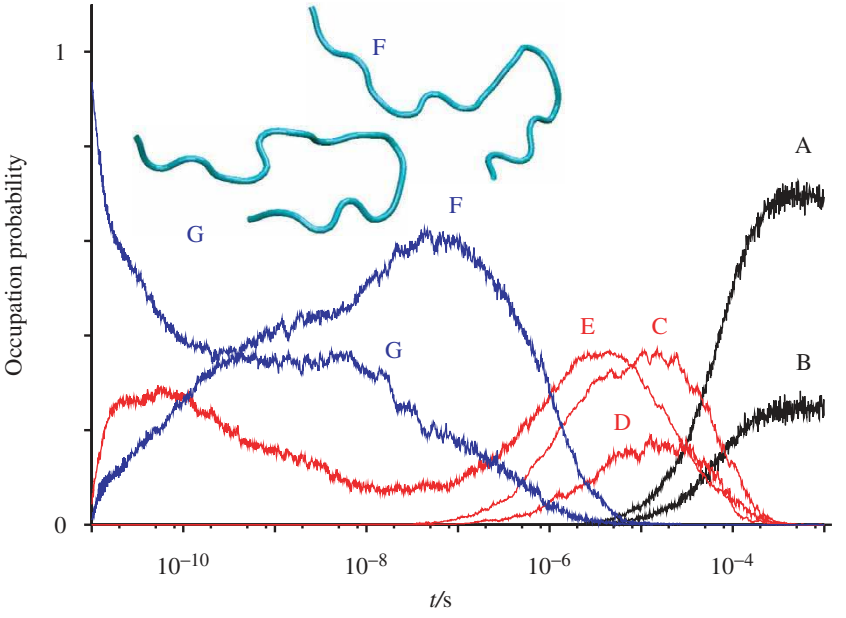

Figure 17. Time evolution of the average occupation probabilities of different groups of minima for GB1 calculated from KMC simulations [62].

Nevertheless, the folding kinetics can be described approximately by a single-exponential form, based upon the rate-determining step from the minima in groups C, D and E to the folded state.

The principal folding intermediate identified in a previous multicanonical Monte Carlo study was described as a free energy minimum (the 'H' state) with no native hydrogen bonds but a well-packed hydrophobic core [265]. Minima in groups C, D and E characterised in the DPS study do correspond to a small decrease in the radius of hydration of the hydrophobic core, defined by the non-hydrogen atoms in the sidechains of the tryptophan, tyrosine, phenylalanine and valine residues. However, a few native hydrogen bonds were also identified in some of these minima. It is possible that these hydrogen bonds become well defined in the DPS study simply because all the structures considered correspond to local minima.

The free energy surfaces calculated by Zhou *et al.* [268] suggest that the choice of order parameters can affect conclusions about the order of folding events. These authors did not find a significant 'H' intermediate. However, a parallel replica dynamics investigation [269] provided evidence for an intermediate 'H' state with some key hydrophobic contacts formed. A transition path sampling unfolding study also located a metastable 'H' intermediate, and concluded that the rate determining step for unfolding was from the folded state to 'H' [270].

It would be useful to conduct further benchmarking calculations for GB1 to try and resolve some of the above discrepancies. In particular, another recent analysis suggests that alternative structures may also be significant for this peptide [103]. Future DPS studies are therefore likely to revisit this system. In particular, new methods to construct

the stationary point database are currently under development. These schemes will probably be more efficient than the framework used to generate the results discussed in this review.

References

[1] D. Passerone and M. Parrinello, Phys. Rev. Lett. **87,** 108302 (2001).
[2] D. J. Wales, Mol. Phys. **100,** 3285 (2002).
[3] A. Laio and M. Parrinello, Proc. Nat. Acad. Sci. USA **99,** 12562 (2002).
[4] T. S. van Erp, D. Moroni and P. G. Bolhuis, J. Chem. Phys. **118,** 7762 (2003).
[5] C. Dellago, P. Bolhuis and P. L. Geissler, Adv. Chem. Phys. **123,** 1 (2002).
[6] P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, Annu. Rev. Phys. Chem. **53,** 291 (2002).
[7] D. J. Wales, Mol. Phys. **102,** 891 (2004).
[8] N. Singhal, C. D. Snow and V. S. Pande, J. Chem. Phys. **121,** 415 (2004).
[9] A. K. Faradjian and R. Elber, J. Chem. Phys. **120,** 10880 (2004).
[10] S. Sriraman, I. G. Kevrekidis and G. Hummer, J. Phys. Chem. B **109,** 6479 (2005).
[11] K. N. Kudin and R. Car, J. Chem. Phys. **122,** 114108 (2005).
[12] S. A. Trygubenko and D. J. Wales, Mol. Phys. **104,** 1497 (2006).
[13] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **121,** 9186 (2004).
[14] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge (2003).
[15] F. H. Stillinger and T. A. Weber, Science **225,** 983 (1984).
[16] D. J. Wales and J. P. K. Doye, J. Chem. Phys. **119,** 12409 (2003).
[17] D. J. McGinty, J. Chem. Phys. **55,** 580 (1971).
[18] J. J. Burton, J. Chem. Phys. **56,** 3133 (1972).
[19] M. R. Hoare, Adv. Chem. Phys. **40,** 49 (1979).
[20] G. Franke, E. R. Hilf and P. Borrmann, J. Chem. Phys. **98,** 3496 (1993).
[21] D. J. Wales, Mol. Phys. **78,** 151 (1993).
[22] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson and T. R. Walsh, Adv. Chem. Phys. **115,** 1 (2000).
[23] P. G. Mezey, Theo. Chim. Acta **58,** 309 (1981).
[24] P. G. Mezey, *Potential Energy Hypersurfaces*, Elsevier, Amsterdam (1987).
[25] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **102,** 9659 (1995).
[26] K. D. Ball and R. S. Berry, J. Chem. Phys. **109,** 8541 (1998).
[27] K. D. Ball and R. S. Berry, J. Chem. Phys. **109,** 8557 (1998).
[28] M. A. Miller, J. P. K. Doye and D. J. Wales, Phys. Rev. E **60,** 3701 (1999).
[29] F. Calvo, J. P. K. Doye and D. J. Wales, J. Chem. Phys. **115,** 9627 (2001).
[30] F. Starr, S. Sastry, E. L. Nave, A. Scala, H. Stanley and F. Sciortino, Phys. Rev. E **63,** 041201 (2001).
[31] F. Calvo, J. P. K. Doye and D. J. Wales, J. Chem. Phys. **114,** 7312 (2001).
[32] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84,** 6611 (1987).
[33] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101,** 5111 (1997).
[34] T. V. Bogdan, D. J. Wales and F. Calvo, J. Chem. Phys. **124,** 044102 (2006).
[35] F. Wang and D. P. Landau, Phys. Rev. Lett. **86,** 2050 (2001).
[36] F. Wang and D. P. Landau, Phys. Rev. E **64,** 056101 (2001).
[37] D. P. Landau, S.-H. Tsai and M. Exler, Am. J. Phys. **72,** 1294 (2004).
[38] S. F. Chekmarev and S. V. Krivov, Phys. Rev. E **57,** 2445 (1998).
[39] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57,** 2607 (1986).
[40] G. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E. K. Keramidas, p. 156. Inferface Foundation, Fairfax Station (1991).
[41] K. Hukushima and K. Nemoto, J. Phys. Soc. Japan **65,** 1604 (1996).
[42] M. C. Tesi, E. J. J. van Rensburg, E. Orlandini and S. G. Whittington, J. Stat. Phys. **82,** 155 (1996).
[43] U. H. E. Hansmann, Chem. Phys. Lett. **281,** 140 (1997).
[44] J. P. K. Doye and D. J. Wales, Phys. Rev. Lett. **80,** 1357 (1998).
[45] J. P. K. Doye, D. J. Wales and M. A. Miller, J. Chem. Phys. **109,** 8143 (1998).
[46] J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. **110,** 6896 (1999).
[47] J. P. K. Doye and F. Calvo, J. Chem. Phys. **116,** 8307 (2002).
[48] J. P. Neirotti, F. Calvo, D. L. Freeman and J. D. Doll, J. Chem. Phys. **112,** 10340 (2000).
[49] F. Calvo, J. P. Neirotti, D. L. Freeman and J. D. Doll, J. Chem. Phys. **112,** 10350 (2000).
[50] J. N. Murrell and K. J. Laidler, Trans. Faraday. Soc. **64,** 371 (1968).

[51] R. E. Kunz and R. S. Berry, J. Chem. Phys. **103,** 1904 (1995).
[52] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam (1981).
[53] R. E. Kunz, *Dynamics of First-Order Phase Transitions*, Deutsch, Thun (1995).
[54] K. D. Ball, R. S. Berry, R. E. Kunz, F.-Y. Li, A. Proykova and D. J. Wales, Science **271,** 963 (1996).
[55] K. D. Ball and R. S. Berry, J. Chem. Phys. **111,** 2060 (1999).
[56] M. A. Miller, J. P. K. Doye and D. J. Wales, J. Chem. Phys. **110,** 328 (1999).
[57] P. N. Mortenson and D. J. Wales, J. Chem. Phys. **114,** 6443 (2001).
[58] P. N. Mortenson, D. A. Evans and D. J. Wales, J. Chem. Phys. **117,** 1363 (2002).
[59] D. A. Evans and D. J. Wales, J. Chem. Phys. **118,** 3891 (2003).
[60] D. A. Evans and D. J. Wales, J. Chem. Phys. **119,** 9947 (2003).
[61] F. Calvo, F. Spiegelman and D. J. Wales, J. Chem. Phys. **118,** 8754 (2003).
[62] D. A. Evans and D. J. Wales, J. Chem. Phys. **121,** 1080 (2004).
[63] J. M. Carr and D. J. Wales, J. Chem. Phys. **123,** 234901 (2005).
[64] E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, Chem. Phys. Lett. **156,** 472 (1989).
[65] R. Czerminski and R. Elber, J. Chem. Phys. **92,** 5580 (1990).
[66] E. M. Sevick, A. T. Bell and D. N. Theodorou, J. Chem. Phys. **98,** 3196 (1993).
[67] R. Olender and R. Elber, J. Chem. Phys. **105,** 9299 (1996).
[68] S. Huo and J. E. Straub, J. Chem. Phys. **107,** 5000 (1997).
[69] A. F. Voter, J. Chem. Phys. **106,** 4665 (1997).
[70] A. F. Voter, Phys. Rev. Lett. **78,** 3908 (1997).
[71] S. F. Chekmarev and S. V. Krivov, Chem. Phys. Lett. **287,** 719 (1998).
[72] S. Huo and J. E. Straub, Proteins **36,** 249 (1999).
[73] C. Dellago, P. G. Bolhuis and D. Chandler, J. Chem. Phys. **110,** 6617 (1999).
[74] R. Elber, J. Meller and R. Olender, J. Phys. Chem. B **103,** 899 (1999).
[75] M. R. Sørensen and A. F. Voter, J. Chem. Phys. **112,** 9599 (2000).
[76] R. Elber and D. Shalloway, J. Chem. Phys. **112,** 5539 (2000).
[77] P. Eastman, N. Gronbech-Jensen and S. Doniach, J. Chem. Phys. **114,** 3823 (2001).
[78] S. V. Krivov, S. F. Chekmarev and M. Karplus, Phys. Rev. Lett. **88,** 038101 (2002).
[79] F. Montalenti and A. F. Voter, J. Chem. Phys. **116,** 4819 (2002).
[80] J. A. Rahman and J. C. Tully, J. Chem. Phys. **116,** 8750 (2002).
[81] J. A. Rahman and J. C. Tully, Chem. Phys. **285,** 277 (2002).
[82] R. Elber, A. Ghosh and A. Cardenas, Accounts Chem. Res. **35,** 396 (2002).
[83] J. E. Straub, J. Guevara, S. Huo and J. P. Lee, Accounts Chem. Res. **35,** 473 (2002).
[84] D. Chandler, J. Chem. Phys. **68,** 2959 (1978).
[85] J. B. Anderson, J. Chem. Phys. **58,** 4684 (1973).
[86] C. H. Bennett, in *Algorithms for Chemical Computations*, edited by R. E. Christofferson, pp. 63–97, American Chemical Society, Washington, D.C. (1977).
[87] J. Pitman, *Probability*, Springer-Verlag, New York (1993).
[88] B. E. Trumbo, *Relationship Between the Poisson and Exponential Distributions*, http://www.sci.csuhayward.edu/statistics/Resources/Essays/PoisExp.htm (1999).
[89] R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka and E. I. Shakhnovich, J. Chem. Phys. **108,** 334 (1998).
[90] C. D. Snow, E. J. Sorin, Y. M. Rhee and V. S. Pande, Ann. Rev. Biophys. Biomol. Struct. **34,** 43 (2004).
[91] A. B. Bortz, M. H. Kalos and J. L. Lebowitz, J. Comput. Phys. **17,** 10 (1975).
[92] D. A. Reed and G. Ehrlich, Surf. Sci. **105,** 603 (1981).
[93] A. F. Voter, Phys. Rev. B **34,** 6819 (1986).
[94] K. A. Fichthorn and W. H. Weinberg, J. Chem. Phys **95,** 1090 (1991).
[95] A. F. Voter, in *Radiation Effects in Solids*, pp. 1–22. Springer-Verlag (2005).
[96] M. A. Novotny, Phys. Rev. Lett. **74,** 1 (1995).
[97] W. Cai, M. H. Kalos, M. de Koning and V. V. Bulatov, Phys. Rev. E **66,** 046703 (2002).
[98] M. de Koning, W. Cai, B. Sadigh, T. Oppelstrup, M. H. Kalos and V. V. Bulatov, J. Chem. Phys. **122,** 074103 (2005).
[99] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, Academic Press, Orlando (1984).
[100] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN*, Cambridge University Press, 2nd edn. (1992).
[101] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe and H. van der Vorst (eds), *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia (2000).
[102] A. R. Fersht, Proc. Nat. Acad. Sci. USA **99,** 14122 (2002).
[103] S. V. Krivov and M. Karplus, Proc. Nat. Acad. Sci. USA **101,** 14766 (2004).
[104] N. J. Marianayagam, N. L. Fawzi and T. Head-Gordon, Proc. Nat. Acad. Sci. USA **102,** 16684 (2005).
[105] C. D. Snow, B. Zagrovic and V. S. Pande, J. Am. Chem. Soc. **124,** 14548 (2002).
[106] C. D. Snow, H. Nguyen, V. Pande and M. Gruebele, Nature **420,** 102 (2002).

[107] C. D. Snow, L. Qiu, F. Gai, S. J. Hagen and V. S. Pande, Proc. Natl. Acad. Sci. USA **101,** 4077 (2004).
[108] D. J. Wales, *OPTIM: A Program for Optimising Geometries and Calculating Pathways*, http://www-wales.ch.cam.ac.uk/software.html.
[109] E. W. Dijkstra, Numerische Math. **1,** 269 (1959).
[110] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Mass., 2nd edn (2001).
[111] H. Pelzer and E. Wigner, Z. Phys. Chem. **B15,** 445 (1932).
[112] W. F. K. Wynne-Jones and H. Eyring, J. Chem. Phys. **3,** 492 (1935).
[113] H. Eyring, Chem. Rev. **17,** 65 (1935).
[114] M. G. Evans and M. Polanyi, Trans. Farady Soc. **31,** 875 (1935).
[115] W. Forst, *Theory of Unimolecular Reactions*, Academic Press, New York (1973).
[116] K. J. Laidler, *Chemical Kinetics*, Harper & Row, New York (1987).
[117] H. Eyring, J. Chem. Phys. **3,** 107 (1935).
[118] M. G. Evans and M. Polanyi, Trans. Farady Soc. **33,** 448 (1937).
[119] R. G. Gilbert and S. C. Smith, *Theory of Unimolecular and Recombination Reactions*, Blackwell, Oxford (1990).
[120] D. T. Gillespie, J. Comput. Phys. **22,** 403 (1976).
[121] D. T. Gillespie, J. Phys. **81,** 2340 (1977).
[122] G. H. Gilmer, Science **208,** 335 (1980).
[123] T. J. Frankcombe and S. C. Smith, J. Comp. Chem. **21,** 592 (2000).
[124] S. A. Trygubenko and D. J. Wales, J. Chem. Phys., in press.
[125] J. Nocedal, Mathematics of Computation **35,** 773 (1980).
[126] D. Liu and J. Nocedal, Math. Prog. **45,** 503 (1989).
[127] L. J. Munro and D. J. Wales, Phys. Rev. B **59,** 3969 (1999).
[128] Y. Kumeda, L. J. Munro and D. J. Wales, Chem. Phys. Lett. **341,** 185 (2001).
[129] F. B. Hildebrand, *Methods of Applied Mathematics*, Dover, New York (1992).
[130] E. B. Wilson, J. C. Decius and P. C. Cross, *Molecular Vibrations*, Dover, New York (1980).
[131] M. Page and J. W. McIver, J. Chem. Phys. **88,** 922 (1988).
[132] D. J. Wales, J. Chem. Phys. **101,** 3750 (1994).
[133] D. J. Wales and J. Uppenbrink, Phys. Rev. B **50,** 12342 (1994).
[134] D. J. Wales and T. R. Walsh, J. Chem. Phys. **105,** 6957 (1996).
[135] R. Fletcher, *Practical Methods of Optimization: Volume 1—Unconstrained Optimization*, Wiley, New York (1980).
[136] J. Simons, P. Jørgenson, H. Taylor and J. Ozment, J. Phys. Chem. **87,** 2745 (1983).
[137] V. Bakken and T. Helgaker, J. Chem. Phys. **117,** 9160 (2002).
[138] K. Németh, O. Coulaud, G. Monard and J. G. Angyan, J. Chem. Phys. **113,** 5598 (2000).
[139] K. Németh, O. Coulaud, G. Monard and J. G. Angyan, J. Chem. Phys. **114,** 9747 (2001).
[140] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120,** 2082 (2004).
[141] R. Elber and M. Karplus, Chem. Phys. Lett. **139,** 375 (1987).
[142] R. Elber and M. Karplus, Chem. Phys. Lett. **311,** 335 (1999).
[143] G. Henkelman and H. Jónsson, J. Chem. Phys. **111,** 7010 (1999).
[144] G. Henkelman, B. P. Uberuaga and H. Jónsson, J. Chem. Phys. **113,** 9901 (2000).
[145] G. Henkelman and H. Jónsson, J. Chem. Phys. **113,** 9978 (2000).
[146] G. Henkelman and H. Jónsson, J. Chem. Phys. **115,** 9657 (2001).
[147] J. M. Carr, S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **122,** 234903 (2005).
[148] O. M. Becker and M. Karplus, J. Chem. Phys. **106,** 1495 (1997).
[149] D. J. Wales, M. A. Miller and T. R. Walsh, Nature **394,** 758 (1998).
[150] D. J. Wales, Phil. Trans. Roy. Soc. A **363,** 357 (2005).
[151] C. Levinthal, in *Mössbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, edited by P. DeBrunner, J. Tsibris and E. Munck, p. 22, Urbana (1969), University of Illinois Press.
[152] T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G. J. Rylance, R. L. Johnston and D. J. Wales, J. Chem. Phys. **122,** 084714 (2005).
[153] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31,** 5262 (1985).
[154] T. F. Middleton and D. J. Wales, Phys. Rev. B **64,** 024205 (2001).
[155] G. Chartrand, *Introductory Graph Theory*, Dover, New York (1985).
[156] P. E. Leopold, M. Montal and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89,** 8721 (1992).
[157] Y. Kumeda and D. J. Wales, Chem. Phys. Lett. **374,** 125 (2003).
[158] R. Zwanzig, A. Szabo and B. Bagchi, Proc. Natl. Acad. Sci. USA **89,** 20 (1992).
[159] R. Zwanzig, Proc. Natl. Acad. Sci. USA **92,** 9801 (1995).
[160] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **105,** 8428 (1996).

[161] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84,** 7524 (1987).
[162] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93,** 6902 (1989).
[163] J. D. Bryngelson and P. G. Wolynes, Biopolymers **30,** 177 (1990).
[164] H. Frauenfelder, S. G. Sligar and P. G. Wolynes, Science **254,** 1598 (1991).
[165] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, Proc. Natl. Acad. Sci. USA **92,** 3626 (1995).
[166] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, Proteins: Struct., Func. and Gen. **21,** 167 (1995).
[167] P. G. Wolynes, J. N. Onuchic and D. Thirumalai, Science **267,** 1619 (1995).
[168] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, Annu. Rev. Phys. Chem. **48,** 545 (1997).
[169] K. A. Dill and H. S. Chan, Nature Struct. Biol. **4,** 10 (1997).
[170] N. D. Socci, J. N. Onuchic and P. G. Wolynes, Proteins: Struct., Func. Gen. **32,** 136 (1998).
[171] C. M. Dobson, A. Šali and M. Karplus, Angew. Chem. Int. Ed. **37,** 868 (1998).
[172] J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine and N. D. Socci, Adv. Prot. Chem. **53,** 87 (2000).
[173] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101,** 1519 (1994).
[174] N. D. Socci, J. N. Onuchic and P. G. Wolynes, J. Chem. Phys. **104,** 5860 (1996).
[175] R. A. Goldstein, Z. Luthey-Schulten and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **89,** 4918 (1992).
[176] M. Karplus and A. Šali, Curr. Opin. Struct. Biol. **5,** 58 (1995).
[177] F. Crick and J. D. Watson, Nature **177,** 473 (1956).
[178] H. S. Chan and K. Dill, J. Chem. Phys. **99,** 2116 (1993).
[179] H. S. Chan and K. Dill, J. Chem. Phys. **100,** 9238 (1994).
[180] H. S. Chan and K. Dill, Proteins: Struct., Func. Gen. **30,** 2 (1998).
[181] R. S. Berry and D. J. Wales, Phys. Rev. Lett. **63,** 1156 (1989).
[182] D. J. Wales and R. S. Berry, Phys. Rev. Lett. **73,** 2875 (1994).
[183] J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. **110,** 6896 (1999).
[184] D. D. Frantz, J. Chem. Phys. **115,** 6136 (2001).
[185] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87,** 3526 (1990).
[186] J. D. Honeycutt and D. Thirumalai, Biopolymers **32,** 695 (1992).
[187] M. A. Miller and D. J. Wales, J. Chem. Phys. **111,** 6610 (1999).
[188] J. P. K. Doye and D. J. Wales, Phys. Rev. B **59,** 2292 (1999).
[189] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **111,** 11070 (1999).
[190] T. F. Middleton and D. J. Wales, J. Chem. Phys. **118,** 4583 (2003).
[191] S. V. Krivov and M. Karplus, J. Chem. Phys. **117,** 10894 (2002).
[192] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, *Network Flows*, Prentice-Hall, New Jersey (1993).
[193] J. S. Shaffer and A. K. Chakraborty, Macromolecules **26,** 1120 (1993).
[194] J. E. Jones and A. E. Ingham, Proc. R. Soc. A **107,** 636 (1925).
[195] D. J. Wales and H. A. Scheraga, Science **285,** 1368 (1999).
[196] L. T. Wille, in *Annual Reviews of Computational Physics VII*, edited by D. Stauffer, World Scientific, Singapore (2000).
[197] D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas, T. F. Middleton, The Cambridge Cluster Database, http://www-wales.ch.cam.ac.uk/CCD.html.
[198] For example, the Birmingham cluster web site at http://www.tc.bham.ac.uk/bcweb contains a variety of cluster global minima, global minima for Lennard-Jones clusters are tabulated at http://www.vcl.uh.edu/~cbarron/LJ_cluster/researchpot.html, and global optimisation packages using interval arithmetic are available from http://www.mscs.mu.edu/~globsol. Results for genetic algorithms and related methods can be found at http://panizzi.shef.ac.uk/cisrg/links/ea_bib.html.
[199] R. S. Berry, T. L. Beck, H. L. Davis and J. Jellinek, Adv. Chem. Phys. **70B,** 75 (1988).
[200] C. Dellago, P. G. Bolhuis and D. Chandler, J. Chem. Phys. **108,** 9236 (1998).
[201] W. E, W. Ren and E. Vanden-Eijnden, Phys. Rev. B **66,** 052301 (2002).
[202] A. L. Mackay, Acta Cryst. **15,** 916 (1962).
[203] S. Gómez and D. Romero, in *Proceedings of the First European Congress of Mathematics*, vol. III, pp. 503-509, Birkhauser, Basel (1994).
[204] J. Pillardy and L. Piela, J. Phys. Chem. **99,** 11805 (1995).
[205] J. P. K. Doye and D. J. Wales, Chem. Phys. Lett. **247,** 339 (1995).
[206] J. P. K. Doye, D. J. Wales and R. S. Berry, J. Chem. Phys. **103,** 4234 (1995).
[207] J. A. Niesse and H. R. Mayne, J. Chem. Phys. **105,** 4700 (1996).
[208] R. H. Leary, Abstract for the *International Workshop on Global Optimization* (1999).
[209] R. S. Berry and R. Breitengraser-Kunz, Phys. Rev. Lett. **74,** 3951 (1995).
[210] J. P. K. Doye, M. A. Miller and D. J. Wales, J. Chem. Phys. **111,** 8417 (1999).
[211] M. P. Hodges, *XMakemol: A Program for Visualising Atomic and Molecular Systems*, version 4.09 (1998).

[212] L. D. Marks, Phil. Mag. A **49,** 81 (1984).
[213] V. V. Nauchitel and A. J. Pertsin, Mol. Phys. **40,** 1341 (1980).
[214] J. Jellinek, T. L. Beck and R. S. Berry, J. Chem. Phys. **84,** 2783 (1986).
[215] J. D. Honeycutt and H. C. Andersen, J. Phys. Chem. **91,** 4950 (1987).
[216] T. L. Beck, J. Jellinek and R. S. Berry, J. Chem. Phys. **87,** 545 (1987).
[217] T. L. Beck and R. S. Berry, J. Chem. Phys. **88,** 3910 (1988).
[218] L. M. Amon and W. P. Reinhardt, J. Chem. Phys. **113,** 3573 (2000).
[219] P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65,** 1567 (1990).
[220] D. J. Wales and R. S. Berry, J. Chem. Phys. **92,** 4473 (1990).
[221] H.-P. Cheng, X. Li, R. L. Whetten and R. S. Berry, Phys. Rev. A **46,** 791 (1992).
[222] S. Weerasinghe and F. G. Amar, J. Chem. Phys. **98,** 4967 (1993).
[223] C. J. Tsai and K. D. Jordan, J. Chem. Phys. **99,** 6957 (1993).
[224] R. E. Kunz and R. S. Berry, Phys. Rev. E **49,** 1895 (1994).
[225] R. M. Lynden-Bell and D. J. Wales, J. Chem. Phys. **101,** 1460 (1994).
[226] D. J. Wales and J. P. K. Doye, J. Chem. Phys. **103,** 3061 (1995).
[227] R. B. McClurg, R. C. Flagan and W. A. Goddard, J. Chem. Phys. **102,** 3322 (1995).
[228] D. D. Frantz, J. Chem. Phys. **102,** 3747 (1995).
[229] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **102,** 9673 (1995).
[230] G. E. Lopez, J. Chem. Phys. **104,** 6650 (1996).
[231] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J. Chem. Phys. **79,** 926 (1983).
[232] W. L. Jorgensen, J. Am. Chem. Soc. **103,** 335 (1981).
[233] W. L. Jorgensen, J. Chem. Phys. **77,** 4156 (1982).
[234] D. J. Wales and M. P. Hodges, Chem. Phys. Lett. **286,** 65 (1998).
[235] B. Hartke, Phys. Chem. Chem. Phys. **5,** 275 (2003).
[236] H. Kabrede and R. Hentschke, J. Phys. Chem. B **107,** 3914 (2003).
[237] C. J. Tsai and K. D. Jordan, J. Chem. Phys. **95,** 3850 (1991).
[238] C. J. Tsai and K. D. Jordan, Chem. Phys. Lett. **213,** 181 (1993).
[239] D. J. Wales and I. Ohmine, J. Chem. Phys. **98,** 7257 (1993).
[240] D. J. Wales and I. Ohmine, J. Chem. Phys. **98,** 7245 (1993).
[241] J. M. Pedulla and K. D. Jordan, Chem. Phys. **239,** 593 (1998).
[242] D. Laria, J. Rodriguez, C. Dellago and D. Chandler, J. Phys. Chem. A **105,** 2646 (2001).
[243] U. Buck, I. Ettisher, M. Melzer, V. Buch and J. Sadlej, Phys. Rev. Lett. **80,** 2578 (1998).
[244] C. J. Gruenloh, J. R. Carney, F. C. Hagemeister, C. A. Arrington, T. S. Zwier, S. Y. Fredericks, J. T. Wood and K. D. Jordan, J. Chem. Phys. **109,** 6601 (1998).
[245] W. B. Blanton, S. W. Gordonwylie, G. R. Clark, K. D. Jordan, J. T. Wood, U. Geiser and T. J. Collins, J. Am. Chem. Soc. **121,** 3551 (1999).
[246] J. Rodriguez, G. Moriena and D. Laria, Chem. Phys. Lett. **356,** 147 (2002).
[247] C. J. Tsai and K. D. Jordan, J. Phys. Chem. **97,** 5208 (1993).
[248] G. J. Geyer, Stat. Sci. **7,** 437 (1992).
[249] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, J. Comp. Chem. **4,** 187 (1983).
[250] T. Lazaridis and M. Karplus, Proteins: Struct., Func. Gen. **35,** 133 (1999).
[251] http://www.trygub.com/charmm/rotamers/ (2004).
[252] W. H. Graham, E. S. Carter II and R. P. Hicks, Biopolymers **32,** 1755 (1992).
[253] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14,** 1333 (1993).
[254] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314,** 141 (1999).
[255] U. H. E. Hansmann, M. Masuya and Y. Okamoto, Proc. Natl. Acad. Sci. USA **94,** 10652 (1997).
[256] F. Eisenmenger and U. H. E. Hansmann, J. Phys. Chem. B **101,** 3304 (1997).
[257] U. H. E. Hansmann, Y. Okamato and J. N. Onuchic, Proteins: Struct., Func. and Gen. **34,** 472 (1999).
[258] M. Shen and K. F. Freed, Biophys. J. **82,** 1791 (2002).
[259] M. H. Zaman, M. Shen, R. S. Berry and K. F. Freed, J. Phys. Chem. B **107,** 1685 (2003).
[260] J. Dudowicz, K. F. Freed and M. Shen, J. Chem. Phys. **118,** 1989 (2003).
[261] A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield and G. M. Clore, Science **253,** 657 (1991).
[262] F. J. Blanco, G. Rivas and L. Serrano, Nature Struct. Biol. **1,** 584 (1994).
[263] V. Muńoz, P. A. Thompson, J. Hofrichter and W. A. Eaton, Nature **390,** 196 (1997).
[264] V. Muńoz, E. R. Henry, J. Hofrichter and W. A. Eaton, Proc. Natl. Acad. Sci. USA **95,** 5872 (1998).
[265] A. R. Dinner, T. Lazaridis and M. Karplus, Proc. Natl. Acad. Sci. USA **96,** 9068 (1999).
[266] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. USA **96,** 9062 (1999).
[267] A. E. García and K. Y. Sanbonmatsu, Proteins: Struct., Func. and Gen. **42,** 345 (2001).

[268] R. Zhou, B. J. Berne and R. Germain, Proc. Natl. Acad. Sci. USA **98,** 14931 (2001).
[269] B. Zagrovic, E. J. Sorin and V. Pande, J. Mol. Biol. **313,** 151 (2001).
[270] P. G. Bolhuis, Proc. Nat. Acad. Sci. USA **100,** 12129 (2003).
[271] G. Wei, P. Derreumaux and N. Mousseau, J. Chem. Phys. **119,** 6403 (2003).
[272] R. Zhou and B. J. Berne, Proc. Nat. Acad. Sci. USA **99,** 12777 (2002).
[273] D. Roccatano, A. Amadel, A. Di Nola and H. J. C. Berendsen, Protein Sci. **8,** 2130 (1999).
[274] B. Ma and R. Nussinov, J. Mol. Biol. **296,** 1091 (2000).
[275] F. J. Blanco and L. Serrano, Eur. J. Biochem. **230,** 634 (1995).