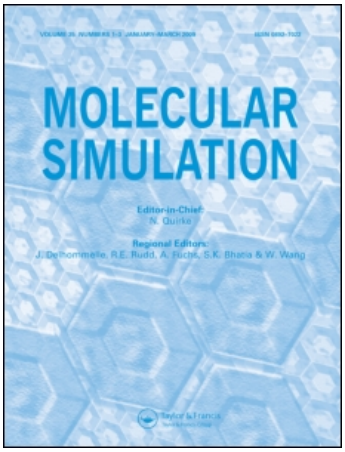

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Recent advances in transition path sampling: accurate reaction coordinates, likelihood maximisation and diffusive barrier-crossing dynamics

Baron Peters[ab]
[a] Department of Chemical Engineering, University of California, Santa Barbara, CA, USA [b] Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA, USA

First published on: 05 November 2010

To cite this Article Peters, Baron(2010) 'Recent advances in transition path sampling: accurate reaction coordinates, likelihood maximisation and diffusive barrier-crossing dynamics', Molecular Simulation, 36: 15, 1265 — 1281, First published on: 05 November 2010 (iFirst)
To link to this Article: DOI: 10.1080/08927020903536382
URL: http://dx.doi.org/10.1080/08927020903536382

Molecular Simulation
Vol. 36, No. 15, December 2010, 1265–1281

Taylor & Francis
Taylor & Francis Group

Recent advances in transition path sampling: accurate reaction coordinates, likelihood maximisation and diffusive barrier-crossing dynamics

Baron Peters*

Department of Chemical Engineering, University of California, Santa Barbara, CA 93106, USA; Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA

(*Received 3 September 2009; final version received 7 December 2009*)

Because transition states are too rare and transient to observe in experiments, simulations are among the best sources of molecular-level insight on transition states and reaction mechanisms. A first step in many rare-event simulations is to identify a reaction coordinate. For reactions that break and make strong bonds, the reaction coordinate is the unstable eigenmode at a saddle point on the potential energy surface. However, processes such as nucleation and protein folding may disrupt and reorganise thousands of interactions, so identifying a reaction coordinate is a major challenge. An accurate reaction coordinate should result in a free energy profile that is consistent with the projected dynamics. Ten years ago, this intuitive requirement led to committor analysis, a trial-and-error procedure for testing putative reaction coordinates. Since then, our ability to identify accurate reaction coordinates has dramatically improved. First, new versions of committor analysis are quantitative and more efficient. Second, likelihood maximisation can systematically and efficiently identify accurate reaction coordinates from thousands of candidates using only the shooting point data from a path sampling simulation. Finally, the new aimless shooting version of transition path sampling retains a high sampling efficiency even for systems with highly diffusive barrier-crossing dynamics.

Keywords: transition path sampling; likelihood maximisation; reaction coordinates; diffusive dynamics; rare events

1. Introduction

Activated processes such as chemical reactions, nucleation and electron transfer lead to important changes in materials and living organisms. At the molecular level, these 'rare events' begin as reactants, pass through transition states and end as products in just a few fleeting moments when compared to typical lifetimes of the stable reactant and product states. Even though transition states are only briefly and rarely occupied, they are important because they determine the lifetimes of the reactants and products. These lifetimes give rise to technologically important properties from catalytic turnover frequencies [1] to polymorph selectivity in crystallisation [2]. Because transition states are rarely and briefly occupied, they cannot be directly observed in experiments. Thus, theory and computation are valuable sources of molecular-level insight on transition states. When carefully formulated with links to observable phenomena, theory and computation provide a way to test hypotheses about mechanisms and transition states at the molecular level.

For reactions that break and make strong bonds, theory and computation have reached an extraordinary level of sophistication. For example, electronic structure theory is routinely used to screen new metal alloys and oxides for more active and cost-effective catalysts. Theory has so frequently made accurate predictions about catalysts ahead of experiments that patents have been filed and granted on the basis of computational evidence alone! [1]. The success of theory and computation in catalysis and bond-breaking chemistry has been enabled by advances in three key areas.

(1) Quantum chemistry research matured to a level that potential energy landscapes for many strong bond-breaking and bond-making processes can be described with sufficient accuracy to compare mechanisms, activation barriers and kinetic selectivities [3,4].
(2) Dynamical bottlenecks of processes that break or make a few strong bonds coincide with saddle points on the potential energy surface (PES) [5,6]. Thus, saddle-point search algorithms that emerged in the 1980s and 1990s (Cerjan–Miller [7], nudged elastic band [8], etc. [9–13]) enabled the identification of transition states and reaction coordinates for strong bond-breaking and bond-making processes.
(3) Corrections for tunnelling and recrossing are sometimes needed, but transition state theory (TST) is an accurate approximation for many reactions that break and make strong bonds [14].

The third point requires justification. To break or make a strong bond, key atoms must move just a few Angstroms.

*Email: baronp@engineering.ucsb.edu

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020903536382
http://www.informaworld.com

Because of the short distance to cross the barrier, ballistic motion along the reaction coordinate persists long enough for trajectories to reach the product state. Once in the product state, recrossing can still occur due to 'energy diffusion limitations' [15], but this phenomenon mostly occurs for isomerisations [16] and reactions between small molecules in the gas phase [17].

Studies of strong bond-breaking and bond-making processes have been remarkably successful because the transition-state search algorithms enable simulations to take advantage of a separation of timescales. For all rare-event processes, the characteristic waiting time between barrier-crossing events is much longer than the timescales on which the system 'forgets' an initial condition within the reactant basin and establishes a local equilibrium. A conventional molecular dynamics (MD) or Monte Carlo (MC) simulation would take a very long time to escape from basins A to B if the barrier between A and B is high. Even if a trajectory initiated in basin A does eventually cross the barrier, the fraction of time spent at the top of the barrier is miniscule relative to the overall simulation time. The separation of timescales is illustrated in Figure 1. For systems with dynamics characterised by rare events, it is essential to develop algorithms that focus simulation effort on understanding the bottleneck regions.

In contrast to the breaking and making of strong bonds, activated processes such as nucleation, self-assembly, protein folding and electron transfer involve the disruption and reorganisation of many weak bonds and interactions. The individual disrupted molecular interactions may only be comparable to k_BT in energy, but these processes can still exhibit a separation of timescales when so many weak non-bonded interactions must be simultaneously disrupted that the overall barrier remains large. Such processes are commonly described as having a rugged PES because the pathway to the bottleneck traverses many small barriers [18]. The rugged nature of the PES tends to disrupt ballistic motion in the barrier-crossing trajectories, so activated processes such as nucleation [19], self-assembly [20], protein folding [21] and electron transfer [22] often show highly diffusive dynamics. Because coordinates of individual atoms are almost always irrelevant for these processes, collective variables such as the nucleus size [19], the vertical energy gap [23] and numbers of native contacts [24] are needed.

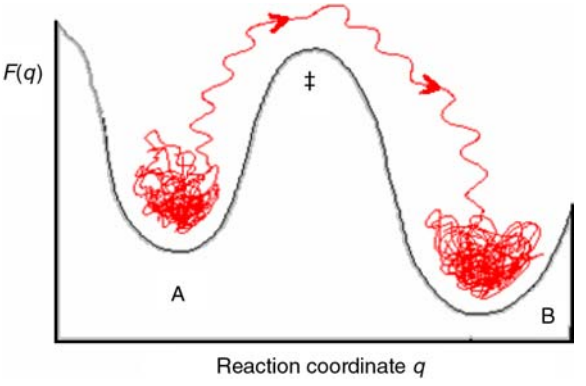

Figure 1. Schematic free energy (F) as a function of the reaction coordinate. The reactant basin is denoted as A, the product as B and the transition state is denoted by ‡. When compared to the short time required to relax from an initial condition to a quasi-equilibrium within one of the basins, the red trajectory spends a long waiting time in each basin between escapes (colour online).

A frequent question is whether variational TST can address these types of problems. Simple models of these processes, for example Langevin dynamics on a model free energy surface (FES) in two dimensions, foster a naive impression that variational TST is practical for activated processes on rugged PESs. In principle, if given an FES from projection of the free energy onto a subspace of *a priori* specified collective variables, $q_1, q_2, \ldots, q_n$, and if also given an efficient way to average the Jacobian at each point on the FES, the variational TST dividing surface can be identified by finding the surface parameterised by $q(q_1, q_2, \ldots, q_n) = q^*$ that maximises [25]

$$F(q^*) - k_BT \ln(\langle|\nabla q|\rangle_{q^*}/\langle|\nabla q|\rangle_A). \quad (1)$$

The Jacobian correction has been explained in numerous references [25–27], including an example of its use on a multi-dimensional free energy landscape [28].

Assuming the practical challenges in these details can be surmounted, an enormous obstacle remains: the variational optimisation will only identify the optimal dividing surface *in the subspace of a priori specified variables*. Clearly, variational TST applied in some subspaces will result in better dividing surfaces than in other subspaces. For example, collective variables to quantify nucleus size can be defined in many different ways. It is not *a priori* obvious whether nucleus cluster size is best determined on the basis of proximity of monomers regardless of their relative orientations, or whether some solid-like local structure should be required before including a monomer when counting the cluster size. Furthermore, we must decide which of many possible solid-like local structural characteristics should be used in defining a critical nucleus size. The number of collective variables to be investigated is growing rapidly, and we have not yet introduced entire classes of other potentially important variables: shape, surface area, etc. The list of potentially important variables is even longer for protein folding.

Thus, variational TST is useful when the relevant coordinates are obvious, or when a multistep process can be analysed as a series of separate activated elementary steps [29,30]. Otherwise, there are many reasons why

variational TST is not tractable for activated processes on a rugged PES where thousands of weak interactions are disrupted.

(1) It is not initially clear which collective variables among perhaps hundreds or thousands of possible choices may be important. Yet to compute an FES, the first step in variational TST, one must choose a small set of collective variables. The cost of an FES calculation grows exponentially with the number of variables retained in the projection, so even FES calculations with three collective variables are rare in the literature [31]. The variational TST approach thus faces an important limitation in the number of collective variables that can be used to define the optimal dividing surface.

(2) Because of the point raised in (1), a variational TST approach is probably limited to optimising the dividing surface for pairs of collective variables. To test all pairs of collective variables with variational TST, the number of 2D FES calculations grows as the square of the number of collective variables that might potentially be important. Realistically, finding the optimal pair is not tractable when many collective variables could potentially be components of the optimal coordinate. Even computing the 1D FES for hundreds of potentially important collective variables is impractical with today's computing power.

Ideas that variational TST is practical for optimising the dividing surface of activated processes characterised by a rugged PES, probably, stem from common 'toy-models' of these processes. For example, Langevin dynamics with forces computed from the model FES is often used to represent the free energy and dynamics from a projection onto one or two collective variables. However, *such models represent a highly idealised case because the dynamics and the FES are consistent by construction*. Experience shows that consistency between the dynamics and the FES is highly unlikely in a real physical problem even when coordinates are carefully chosen with the best intuition. Thus, the true essence of the physical problem is finding coordinates for which the barrier-crossing dynamics are consistent with the FES.

Many methods are available for computing rates and free energies in systems with rugged potential energy landscapes. A complete list is beyond the scope of this article. For example, we do not discuss a group of clever methods to accelerate the transitions between stable states [32–35]. We also omit important methods that parse configuration space into basins in a manner that maximises the degree to which transitions between basins appear to follow simple master equations [36–38]. Finally, we omit transition path theory, an exciting recent development that includes a description of both the scalar committor probability field and the vector flux field along transition pathways [39–41]. Instead, we begin with some of the most widely used importance sampling methods for computing free energy landscapes before discussing the main focus of this article, path sampling methods.

The blue-moon sampling technique [42,43] applies a rigid constraint to the dynamics. As the system evolves on a reaction coordinate isosurface according to the constrained dynamics, the force required to preserve the constraint is averaged. The mean force is computed in this manner for a series of foliations in phase space that correspond to different values of the assumed reaction coordinate. The mean force on each foliation is then integrated to obtain the free energy barrier [44], $\mathrm{d}F = \langle f \rangle_q \, \mathrm{d}q$, where $F(q)$ is the free energy and $\langle f \rangle_q$ is the mean force at the q-isosurface [42].

Umbrella sampling [45] methods typically use a harmonic bias potential of the form $+k(q - q_0)^2/2$ to force the system to sample configurations with values of the assumed reaction coordinate q near q_0. A series of these simulations are performed with different values of q_0 so that the distributions of q-values generated in each simulation overlap. The overlapping distributions provide a local free energy near the target q_0 from $\beta F(q) = -\ln P(q) - k(q - q_0)^2/2$ [45]. Local free energy curves at the different values of q_0 are then self-consistently spliced into a complete free energy profile using a weighted histogram analysis [46].

Still other methods adaptively adjust the bias potential to optimise the sampling efficiency along an assumed reaction coordinate. For example, Wang–Landau methods [47] use the history of states sampled during the calculation to adaptively build a bias potential that is the inverse of the free energy profile. Similar ideas were invoked in the development of metadynamics [48]. For these methods, the FES is obtained from the negative of the accumulated bias potential after convergence.

The nudged elastic band [8] idea has been generalised for use with collective variables by Maragliano et al. [49]. In the generalised version, correlations between motions along the collective variables modify the direction of the reaction pathway on the FES. The minimum free energy path (MFEP, analogous to the MEP from nudged elastic band) follows the equation [49]

$$\mathrm{d}\mathbf{q}(s)/\mathrm{d}s = -M(\mathbf{q}(s))\nabla_{\mathbf{q}} F(\mathbf{q}(s)), \qquad (2)$$

where the path $\mathbf{q}(s)$ which passes through the saddle point on the FES is parameterised by its own arclength s and the matrix M includes the effects of correlated motions between the components of $\mathbf{q}$. Again, the components of the reaction coordinate, i.e. the elements of $\mathbf{q}$, must be specified *a priori* to compute the MFEP [49]. Also, the free energy $F(\mathbf{q})$ is a free energy at a specific point $\mathbf{q}(s)$ along the path in the space of collective variables [49]. To obtain

a rate constant, the free energy from the MFEP must be further projected onto a single coordinate. Transition path theory [39,49] provides a prescription for a mapping onto the generalised arclength coordinate s, but to our knowledge this has not yet been done for any applications. The MFEP is a significant advance because, when the relevant set of coordinates are known *a priori*, the MFEP approach rigorously converts the high-dimensional variational optimisation over collective variables into a tractable problem involving the optimisation of a curve in the space of many collective variables. In this sense, MFEP is an enormously important development that can circumvent many of the problems encountered by variational TST.

Transition path sampling (TPS) [50–55], unlike all of the methods above, does not require *a priori* knowledge about the reaction coordinate or component variables in the reaction coordinate. This article discusses several recent advances in the TPS methodology.

(1) Aimless shooting: for efficiently sampling diffusive barrier-crossing trajectories [56,57].
(2) Quantitative version of committor analysis with simultaneously improved efficiency [58].
(3) Genetic neural network (GNN; square error minimisation) for identifying reaction coordinates [59] and
(4) Likelihood maximisation: accurate reaction coordinates directly from shooting point data [56,57].

TPS has been extensively reviewed [60,61] by the original developers, but we review the central ideas here to emphasise the recent extensions. Before embarking on a review of TPS, we motivate the search for accurate reaction coordinates by highlighting examples where assumed reaction coordinates turned out to be incorrect.

2. Why seek accurate reaction coordinates?

Many problems can arise from an inaccurate reaction coordinate. Variational TST shows that the free energy of activation will be erroneously low if the dividing surface is oblique to the true separatrix [62,63]. The transmission coefficient κ should decrease as the free energy barrier decreases to preserve the rate constant. Thus, as the reaction coordinate becomes more inaccurate, the transmission coefficient becomes smaller, and eventually an intractable number of trajectories are required to compute the rate constant and κ accurately [25,64]. In a detailed analysis, van Erp [65] showed how accurate reaction coordinates facilitate accurate calculations of free energy barriers as well as accurate transmission coefficients. More severe problems can also result from reaction coordinate error. For example, Figure 2 depicts how reaction coordinate error can cause the free energy to exhibit an apparent hysteresis.

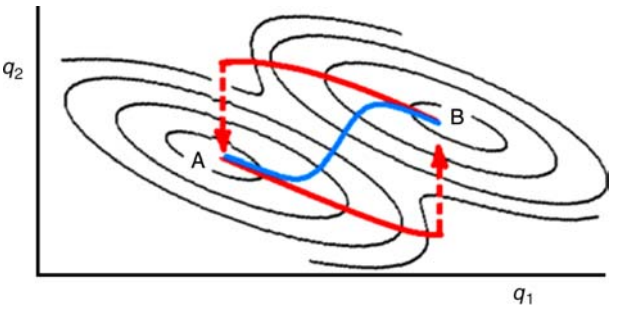

Figure 2. FES as a function of two coordinates q_1 and q_2 with two stable basins, A and B. Because q_1 incurs a large change between the reactant (A) and the product state (B), one might assume q_1 is the reaction coordinate. However, a free energy calculation from A to B along q_1 would result in a different free energy profile than a free energy calculation from B to A along q_1. In both cases, the system fails to follow the true (blue) pathway when forced along the coordinate q_1. Instead, sampling along q_1 traces the red pathways resulting in hysteresis errors in the free energy barrier (colour online).

Chandler and co-workers illustrated that reaction coordinates can be surprisingly difficult to identify, even in systems where the relevant coordinates would seem obvious. For instance, in the dissociation reaction of a Na^+ and a Cl^- ion in an aqueous solution, the intuitive reaction coordinate would be the distance between the two ions. However, Geissler et al. [55] showed that the distance between the ions is not an accurate reaction coordinate. Instead, they suggested, as anticipated by Onsager [66], that solvent degrees of freedom are important components of the correct reaction coordinate. Efforts in follow-up work to our knowledge have yet to definitively identify an accurate 1D coordinate from a mixture of solvent coordinates and a Na^+–Cl^- distance [67].

Another example in which the reaction coordinate turned out to be surprisingly complicated is the isomerisation of the alanine dipeptide. Decades of previous studies [68–75] assumed that the Ramachandran angles were accurate reaction coordinates, but Bolhuis et al. [76] showed that an accurate reaction coordinate could not be constructed from Ramachandran angles alone. Later, Ma and Dinner [59] showed that solvent degrees of freedom were essential components of the reaction coordinate for the alanine dipeptide isomerisation.

TPS provided a way of computing rate constants without knowing the reaction coordinate [60]. However, Dimelow et al. [77] compared the TPS procedure for computing a reaction rate without any knowledge of the reaction coordinate to a rate calculation from the potential of mean force along the reaction coordinate. In their comparison using a reaction coordinate was over 20 times faster than the coordinate-free approach outlined in [77]. Thus, while TPS has provided an elegant coordinate-free

approach to computing rates, using an accurate reaction coordinate is more efficient whenever an accurate coordinate can be identified. These examples illustrate the importance of accurate reaction coordinates, both for obtaining accurate free energy barriers and for efficient calculations of rates.

3. Transition path sampling

TPS was designed to compute rate constants and to harvest unbiased dynamical barrier-crossing trajectories without the need to assume a reaction coordinate [60]. The use of unbiased dynamics sets TPS apart from all previous importance sampling methods for rare events in which sampling or dynamics was biased along an assumed coordinate. TPS is a MC procedure to generate a sequence of trajectories in trajectory space according to the statistical weight of each reactive trajectory [60]. TPS generates a sequence of reactive trajectories by modifying each trajectory and then accepting or rejecting the new trajectory based on its statistical weight in the space of reactive trajectories. The statistical weight of a trajectory is the probability to start at the initial point of the trajectory (from the Boltzmann distribution) multiplied by a series of transition probabilities for following the trajectory from one timeslice to the next, starting from time 0 to the total trajectory duration T. TPS accepts trajectories according to their probability in trajectory space with the constraint that accepted trajectories must connect basins A and B [60].

To generate new trajectories, the original versions of TPS employed shooting and shifting moves [61]. Shooting moves create a new trajectory from an old trajectory by changing the momentum slightly at a random time slice t along the previous trajectory. Then, the dynamics are propagated forward and backward in time to $t = 0$ and T. If the momentum perturbations are too large, then shooting from most of the points along an old trajectory will result in a rejected new trajectory. Therefore, in most applications, the momentum perturbation δp must be very small [61]. The size of δp should be reduced until the average acceptance rate becomes 40–50%. Shifting moves cut a length of time δt from one end of the old trajectory and then propagate the dynamics from the other end for an additional time δt to regain a trajectory of duration T [61]. Shifting moves allow for relaxation of those reactive trajectories that barely have time to reach one basin after spending the vast majority of the trajectory duration in the other basin. Shooting and shifting moves are illustrated in Figure 3.

Most formulations of TPS generate new trajectories in shooting and shifting using dynamics that conserve the ensemble in which the reaction is being studied. With this choice of dynamics, the acceptance rule for trajectories from shooting and shifting moves becomes very simple:

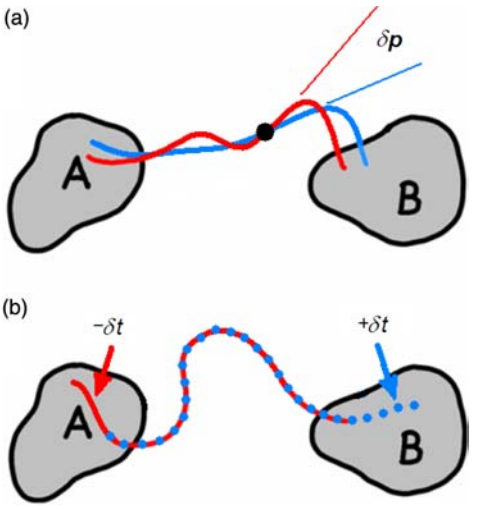

Figure 3. Generation of new trajectories (red) from old trajectories (blue) in the original shooting and shifting version of TPS. (a) Shooting begins with the selection of a random timeslice t, a shooting point, along the old trajectory. The momenta are perturbed slightly at the shooting point and the dynamics are propagated forward and backward in time to $t = 0$ and T. The resulting new trajectory is accepted if it connects basins A and B. (b) Shifting moves cut a length of time δt from one end of the old trajectory and then propagate the dynamics from the other end for an additional time δt to regain a trajectory of duration T. Again, the new trajectory is accepted if it connects basins A and B (colour online).

new trajectories are accepted if they connect states A and B and rejected otherwise [60].

TPS has been applied to a wide range of systems, such as conformational changes in model peptides [76], the folding of small proteins [78], conformational transitions in biomolecules [79] and sugars [77], micelle formation [80], hydrophobic polymer collapse [81], nanoparticle assembly [82], ion association in solution [55], auto-ionisation of water [83], the nucleation of hexagonal ice [84], nucleation in the lattice gas model [85], structural changes in cadmium-selenide nanocrystals [86,87], solid–solid polymorph transformations in organic crystals [88,89], hydrolysis reactions in solution [90], C—C bond formation in a chabazite catalyst [91] and DNA hybridisation [92].

Bolhuis noted that applications of the original TPS algorithm to diffusive systems such as crystal nucleation [93] and protein folding [94] exhibited very low acceptance rates. The commitment time for these types of rare events is usually on the nanosecond timescale (or greater) because of the rugged PES between the top of the free energy barrier and the adjacent stable basins. The problems with diffusive dynamics have inspired alternatives to TPS such as partial path sampling [95] and transition interface sampling [96,97]. Additionally, there have been efforts to develop TPS algorithms that can be applied to diffusive dynamics.

For example, one version of TPS uses only weak stochastic perturbations from the Anderson thermostat to generate new trajectories [93], one version uses shooting moves with submachine precision [98] and one version uses trajectory generation moves with double-ended constraints in the path interior [99]. In each of these methods, the strategy for generating accepted trajectories with diffusive dynamics is to generate new trajectories that are highly similar to much of the old trajectory. However, each trajectory is costly – especially for highly diffusive systems where relaxing from the barrier top into a stable basin may take nanoseconds. Therefore, it is preferable to generate trajectories that rapidly diverge from each other in the trajectory sequence while maintaining reasonable acceptance rates. A method that accomplishes these goals simultaneously, aimless shooting [56,57], is discussed in the following section.

4. Aimless shooting

The new 'aimless shooting' version of TPS replaces the shooting and shifting moves from the original TPS algorithm with a single new type of hybrid move. Aimless shooting retains all of the formal properties of the original TPS algorithm, while also resulting in some very useful new capabilities [56,57]. In aimless shooting, the momenta are drawn from the Boltzmann distribution for each new trajectory, so trajectories de-correlate more quickly in aimless shooting than in conventional TPS [100]. In addition, aimless shooting generates most of the shooting points near the stochastic separatrix even though we have no *a priori* knowledge of the separatrix location [56]. Other advantages are that there is only one adjustable parameter, δt, and that the algorithm is easily used with MD packages such as CHARMM [101] and NAMD [102]. Each aimless shooting trajectory has three segments [57]:

(1) a 'backward' trajectory from $x(t = 0)$ to $x(t = -t/2)$;
(2) a 'connector' trajectory from $x(t = 0)$ to $x(t = \delta t)$ and
(3) a 'forward' trajectory from $x(t = \delta t)$ to $x(t = \delta t + t/2)$.

These segments are illustrated in Figure 4. To initiate aimless shooting, a small time δt is chosen such that $\delta t \ll t$ along the initial pathway of length t. Typically, we have found that $\delta t = 0.01t$ yields favourable acceptance rates. Along the initial trajectory, a configuration is selected that is close to the transition state region. This initial configuration is typically found by firing several randomly seeded trajectories from points along the initial pathway. From here, the aimless shooting algorithm, illustrated in Figure 4, works as follows:

(1) from the previous trajectory, select $x(t = 0)$ or $x(t = \delta t)$ as the shooting point with 50% probability for the two choices. Save the shooting point as $x_{\text{new}}(t = 0)$;

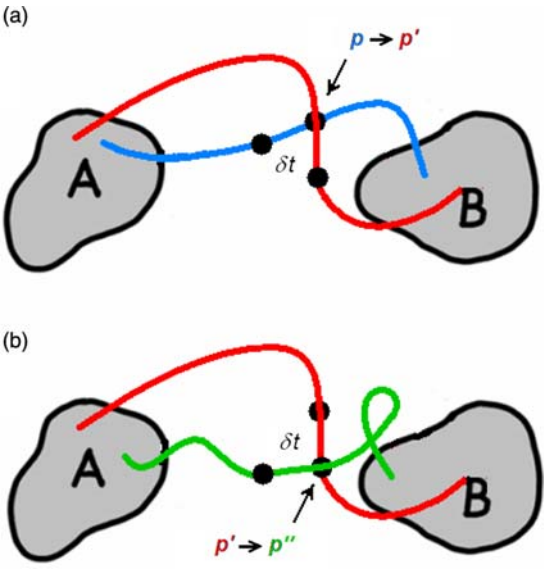

Figure 4. Showing how accepted aimless shooting moves generate new candidate shooting points. (a) The original trajectory is shown in blue and the new trajectory is shown in red. On both trajectories, the connector segment is labelled with δt. New momenta are drawn from the Boltzmann distribution to generate the red trajectory. (b) For the next iteration of aimless shooting, the shooting point is selected randomly between $x(t = 0)$ and $x(t = \delta t)$ configurations on the red trajectory. Again, the momenta to generate the green trajectory will be drawn from the Boltzmann distribution (colour online).

(2) draw new velocities from the Boltzmann distribution at the chosen shooting point;
(3) propagate the dynamics backwards from 0 to time $-t/2$, i.e. reverse the momenta and run a forward trajectory for time $t/2$;
(4) propagate the dynamics forwards in time by δt from the shooting point, and save the configuration as $x_{\text{new}}(t = \delta t)$;
(5) continue the forward trajectory for an additional time $+t/2$;
(6) accept the new trajectory if it joins the reactant and product states, and reject the new trajectory if it does not, just as in the original TPS algorithm.

'Inconclusive' trajectories have one or both endpoints that do not reach either basin. As in the original TPS algorithm, frequent inconclusive trajectories indicate that:

(1) the trajectory length is insufficient or
(2) the basin definitions are too narrow to include fluctuations within the basin;
(3) a third state separate from A and B may be trapping the trajectories.

Aimless shooting has some very useful properties for diffusive barrier-crossing dynamics and for identifying reaction coordinates [56,57]. First, each shooting move is

an independent realisation of the committor probability because the momenta are chosen fresh from the Boltzmann distribution [57]. Second, Peters and Trout [56] showed that the shooting points themselves are approximately distributed according to

$$\rho(x_{sp}) = p(\mathrm{TP}|x_{sp})p(x_{sp}|\mathrm{TP}), \tag{3}$$

where x_{sp} is the configuration of the system at a shooting point (sp), $\rho(x_{sp})$ is the distribution of shooting point configurations, $p(\mathrm{TP}|x_{sp})$ is the probability of being on a transition path given a shooting point configuration and $p(x_{sp}|\mathrm{TP})$ is the probability of being at a shooting point configuration given being on a transition path. This distribution of shooting points is confined to the transition pathway by the factor $p(x|\mathrm{TP})$. The distribution of shooting points is also confined to the zone along the transition pathway where the factor $p(\mathrm{TP}|x)$ is non-zero. Because of this property, aimless shooting automatically maintains a high acceptance probability, even without using small momentum perturbations. Furthermore, the zone where $p(\mathrm{TP}|x)$ is large also corresponds to the region where $p_B(x)$ is near 1/2. Having many shooting points in this region provides data to help identify accurate reaction coordinate and the separatrix [56]. Figure 5 depicts the difference in the typical location of shooting points for aimless shooting and shooting–shifting strategies.

Aimless shooting has been applied to understand highly diffusive processes such as the nucleation in the lattice gas model [56], polymorph transitions in terephthalic acid [88] and nucleation of Lennard-Jonesium crystals from the melt [103]. Figure 6 shows three snapshots from an aimless shooting trajectory linking the Lennard-Jones liquid and a closest-packed Lennard-Jones solid. In the highly diffusive Lennard-Jones systems, aimless shooting generated 1200 accepted trajectories from a total of 3000 generated trajectories. Thus, aimless shooting trajectories rapidly diverge from one another at the shooting point while still maintaining an acceptance rate of 40%.

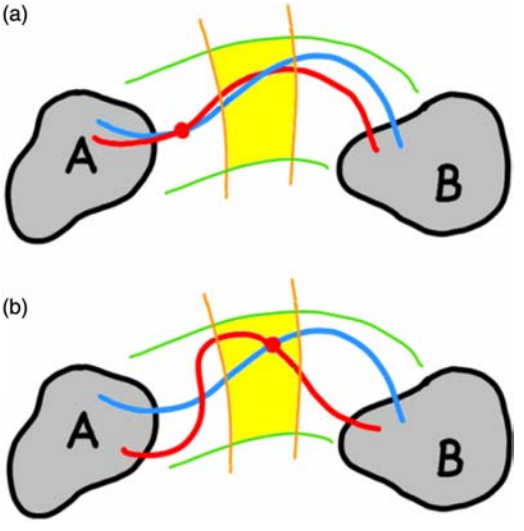

Figure 5. (a) The shooting and shifting version of TPS generates trajectories and shooting points that are confined to the transition pathway between the green boundaries. The shooting points can be anywhere along the transition pathway. (b) Aimless shooting also generates trajectories in the transition pathway, but now the shooting points are automatically generated in the yellow region, which contains the stochastic separatrix (colour online).

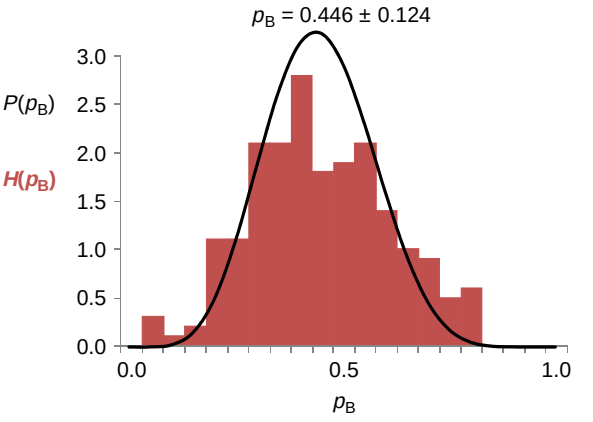

Figure 6. Aimless shooting in the NPT ensemble was used to generate 3000 trajectories with approximately 40% acceptance for the highly diffusive process of crystal nucleation from a supercooled Lennard-Jones liquid. The calculations provided shooting point data that were used to rank scores of various reaction coordinates that have been used to simulate Lennard-Jones nucleation in previous studies. The most accurate 1D reaction coordinate for nucleation is not nucleus size, but instead a product of nucleus size and a measure of structure within the nucleus. The histogram is shown in red with a beta distribution approximant to the true committor distribution. The beta distribution has mean and variance given by Equations (4) and (5) [103] (colour online).

In summary, aimless shooting retains all features of the original TPS algorithm and offers these additional advantages [57]:

(1) each shooting point represents an independent realisation of the committor probability p_B, a feature that will be discussed in the following sections on identifying reaction coordinates;
(2) shooting points are automatically distributed over $0 < p_B < 1$ with a distribution that is peaked near the stochastic separatrix at $p_B = 1/2$. This feature gives high acceptance, even for diffusive barrier-crossing dynamics;
(3) aimless shooting trajectories diverge from each other more rapidly than those from previous shooting–shifting schemes [100];
(4) aimless shooting is easily implemented with existing MD packages and has just one adjustable parameter, δt.

Our aimless shooting simulations have used 1% of the total trajectory length for the parameter δt. However, we can suggest a method for optimising the choice of δt. For this purpose, it is important to remember that the shooting points move in configuration space during the aimless shooting procedure. Each time a new trajectory is accepted, a new shooting point is generated. An optimum δt value should maximise the rate per shooting move at which the successful shooting points diffuse in configuration space. Here, we define a successful shooting point as the shooting point that results in an accepted reactive trajectory. If δt is too small, the acceptance rate will be high, but the shooting points will still move slowly in configuration space. The successful shooting points will also diffuse slowly if δt is so large that it exceeds the typical time to descend from the top of the barrier. In that case, the new shooting points generated by aimless shooting will already have relaxed towards $p_B = 0$ or 1, and these will be unlikely to give new accepted trajectories.

5. Improved version of committor analysis: quantifying reaction coordinate error

The committor probability, denoted as $p_B(x)$, is the probability that a given configuration x will commit to the product basin B when integrated over initial velocities from the Boltzmann distribution. $p_B(x)$ takes the value of 1 for configurations in the product basin, 0 for configurations in the reactant basin and 1/2 for points on the stochastic separatrix. The committor probability $p_B(x)$ depends on the full configuration space, but it can be used to test the accuracy of a physical, collective variable as a reaction coordinate in the following way.

An accurate reaction coordinate q should enable a projection of the dynamics in the full phase space x onto a 1D free energy profile $F(q)$ and a simple description of the dynamics that is consistent with the free energy profile. For example, the dynamics after projection onto an accurate reaction coordinate should resemble Langevin dynamics with forces given by $-\mathrm{d}F(q)/\mathrm{d}q$. Such Langevin descriptions result in Fokker–Planck equations that (after integration of any initial velocity dependence) describe the probabilistic evolution from an initial condition in the projected space $q(t = 0) = q_0$ to a distribution $P(q, t)$ in the projected space at later times. Descriptions of an evolving probability distribution in a projected landscape have been used directly by Hummer and Kevrekidis [104] to infer long-time dynamics from short-duration trajectories. *However, for a projection of the dynamics onto a single coordinate q to be useful, all atomistic configurations x with the same projected initial condition* $x(t = 0) \rightarrow q_0$ *should give a similar distribution* $P(q, t)$ *at later times.*

The histogram test, also known as a *p*-fold test [105] and committor analysis [60], measures the extent to which different configurations x on an isosurface of a coordinate q evolve similarly at later times. The evolution from different configurations is compared at the simplest level: the fraction of trajectories that eventually commit to state B instead of state A. First, one identifies a putative reaction coordinate $q(x)$ and a putative transition state location q^* along $q(x)$. A Boltzmann distributed sample of points are generated on the q^*-isosurface. Each point sampled on the q^*-isosurface has a committor probability value. For a coordinate that accurately captures the dynamics, a histogram of the committor values should be sharply peaked around $p_B = 1/2$. Most investigators only validate the reaction coordinate for a single isosurface that approximates $p_B = 1/2$ [60,61,76,84,85,91,105,106]. In principle, other isosurfaces of q should also be tested to ensure that the values of q and typical values of the committor probability on the family q-isosurfaces change together in a monotonic fashion. This property indicates that when averaged over initial momenta, all configurations x with $q(x) = q^*$ evolve in time similarly to the extent that they all have similar probabilities of reaching the product basin B. Figure 7 depicts the histogram test for an inaccurate reaction coordinate and also for an accurate reaction coordinate.

Histogram tests that have been performed in the literature [60,61,76,84,85,91,105,106] use approximately 100,000 trajectories, often with 100 trajectories per p_B-estimate and 1000 p_B-estimates. The extraordinary cost of each histogram test combined with the unreliable trial-and-error strategy for proposing putative reaction coordinates has motivated the development of systematic methods to identify accurate coordinates. The computational costs of committor analysis have also motivated the development of a more efficient [58] and quantitative [58] version of committor analysis.

Committor analysis as described above is a qualitative indicator of reaction coordinate accuracy because protocol-dependent sampling errors add noise to the p_B-estimates. Peters [58] deconvoluted the sampling error in the histogram to quantify the mean and variance in the actual distribution of p_B-values on the dividing surface. The actual p_B-distribution and the histogram of p_B-estimates are contrasted schematically in Figure 8.

The mean and variance of the actual committor distribution can be obtained from the mean and variance of the histogram. Here, we give the relations without proof, but the derivations as well as analyses of error propagation are given by Peters [58]:

$$\mu_p = \mu_H, \tag{4}$$

$$\sigma_p^2 = \frac{N\sigma_H^2}{N-1} - \frac{\mu_H(1-\mu_H)}{N-1}, \tag{5}$$

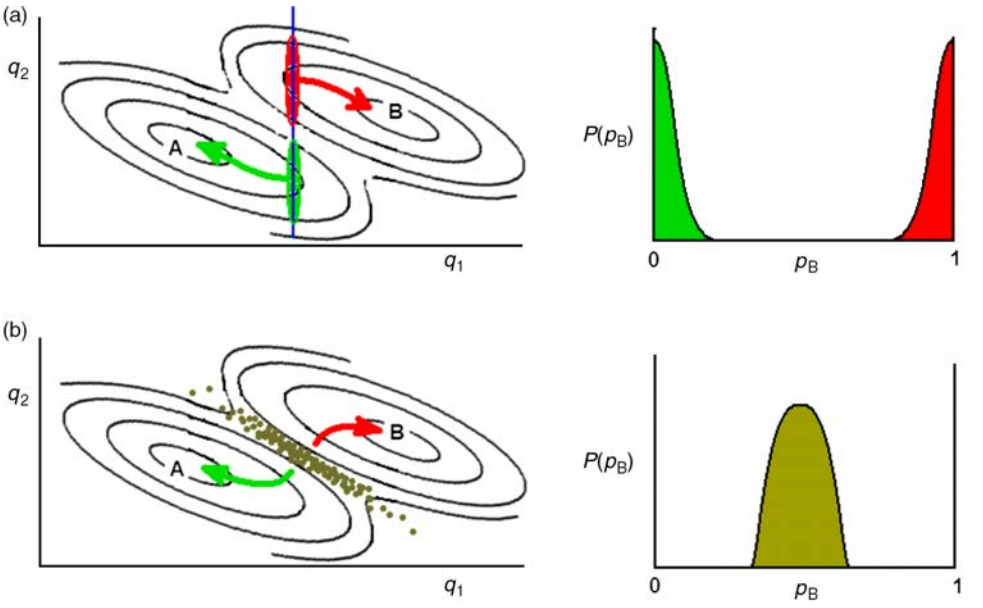

Figure 7. Schematic showing the histogram test. (a) The coordinate q_1 omits the important role of coordinate q_2. A histogram test of q_1 reveals that the dynamics cannot be projected onto q_1 because some configurations on q_1^* are committed reactants and other configurations on q_1^* are committed products. (b) An appropriate mixture of coordinates q_1 and q_2 gives an isosurface for which all points have the same value of the committor probability. For this coordinate, the histogram test gives a sharply peaked unimodal histogram.

where μ_H is the histogram mean, σ_H is the histogram variance, μ_p is the intrinsic mean, σ_p is the intrinsic variance and N is the number of trajectories per p_B-estimate. These formulae provide a quantitative range of committor probability values on a predicted transition state surface:

$$p_B = \mu_p \pm \sigma_p. \tag{6}$$

Peters [58] showed that μ_p and σ_p can be accurately computed in a relatively inexpensive calculation. For reaction coordinates with errors of $\sigma_p > 0.15$, μ_p and σ_p can be estimated to within 10% of their values with 2000 trajectories partitioned among 200 p_B-estimates. Thus, the quantitative measure of reaction coordinate error often costs less than a tenth of the cost of the original committor analysis procedure [58].

6. Systematic methods to identify accurate reaction coordinates

TPS generates an ensemble of trajectories that connect the reactant and product basins, and also rate constants with no *a priori* knowledge of the reaction coordinate. However, for gaining insight about the mechanism of an activated process, the reaction coordinate is the key insight from a simulation. From the reaction coordinate, one can identify the transition state ensemble, and these two pieces of information constitute a nearly complete understanding of the reaction mechanism. Since the beginning of TPS, there has been a focus on systematic ways to identify accurate reaction coordinates. Section 5 outlined committor analysis procedures. Later developments included the GNN least-squares optimisation procedure by Ma and Dinner [59] and the likelihood maximisation approach by Peters and Trout [56] and Peters et al. [57].

6.1 The GNN method

The first method to find the reaction coordinate in a more systematic manner was the GNN method of Ma and Dinner [59]. The GNN method begins with an ensemble of trajectories from TPS. $p_B(x)$ estimates are computed at points along the transition paths and a training set of p_B-estimates is selected so that the estimates are evenly distributed with p_B-values between 0 and 1. Multilevel

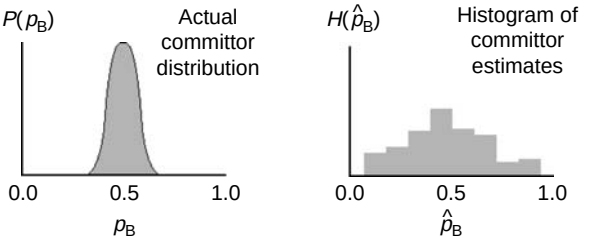

Figure 8. (Schematic) The intrinsic committor distribution $P(p_B)$ corresponds to the distribution of infinitely accurate committor probability estimates. Variance in the distribution $P(p_B)$ is entirely due to reaction coordinate error. The histogram $H(p_B)$ corresponds to the distribution of p_B-estimates in a histogram test. The histogram always has a larger variance because of binomial sampling error in the p_B-estimation process.

neuron models of the reaction coordinate are constructed from a database of possible components of the reaction coordinate. A genetic algorithm sorts through many possible perceptron models and finds the best model reaction coordinate by comparing the square error scores. Each score is determined by minimising the square error between the training set p_B-estimates and the perceptron predictions. A neural network model of the committor probability can be quite complex, as shown in Figure 9, but also quite flexible to describe very complicated reaction coordinates. The scoring function to compare reaction coordinate models is given in Equation (7):

$$S = \sum_{k=1}^{\text{training set}} (\hat{p}_B(x_k) - p_M(x_k))^2. \quad (7)$$

In Equation (7), $\hat{p}_B(x_k)$ is the committor probability estimate at point x_k in the training set, and $p_M(x_k)$ is the neural network model prediction of the committor probability.

The GNN approach provided the first reaction coordinate to explain the role of solvent dynamics in isomerisation of an alanine dipeptide [59]. Specifically, GNN found that the local orientation of solvent dipoles must preorganise to enable the conformational change. Because the solvent preorganisation is dynamically slower than fluctuations in the Ramachandran angles, the solvent orientations are an integral part of the reaction coordinate [59]. GNN's success for the alanine dipeptide was remarkable because this problem had thwarted all previous attempts to find a good reaction coordinate using Ramachandran angles and other internal coordinates.

The GNN approach was a vast improvement over trial-and-error approaches, but the alanine dipeptide example revealed some aspects of the GNN method that needed further improvement. GNN still required many p_B-estimates to construct the training set from the transition path ensemble. Even with a training set constructed from 100,000 trajectories, the histogram test

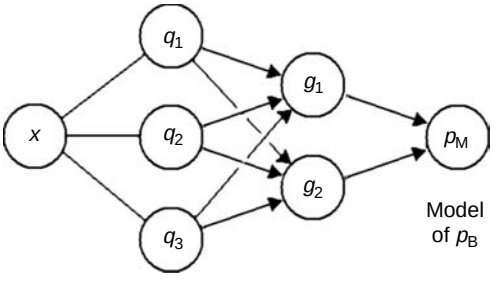

Figure 9. The GNN method uses multilevel neurons to fit a model for the reaction coordinate (p_M) to p_B-estimates computed at points drawn from the transition path ensemble. The inputs to the neurons are subsets of many possible components of the reaction coordinate ($q_1 \ldots q_N$).

confirmed that some components of the reaction coordinate were still missing. Additionally, the reaction coordinate identified by GNN gave a sharply peaked p_B-histogram at $p_B = 1/2$ only when data beyond the range of the training set were omitted [59].

6.2 *Likelihood maximisation*

Peters and Trout [56] and Peters et al. [57] developed likelihood maximisation to obtain reaction coordinates directly from the shooting point information in a path sampling simulation. All previous methods for identifying reaction coordinates relied on separately computed committor probability estimates, an expensive post-processing step beyond the TPS simulation. Peters and Trout [56] and Peters et al. [57] recognised that the shooting points accumulated during TPS along with the shooting point outcomes contain information about the committor probability that had been discarded by all earlier methods. The outcome at each shooting point is a trajectory that commits to either A or B, a binary realisation of the committor probability [56,57]. With no additional trajectory calculations, this binary data-set can be used to screen thousands of model reaction coordinates by likelihood maximisation.

Likelihood maximisation works as follows. First, a transition path ensemble is collected with aimless shooting. Aimless shooting makes each of the binary outcomes statistically independent by drawing the momenta fresh from the Boltzmann distribution at each shooting point. Thus, every shooting point from aimless shooting can be used in the likelihood maximisation step [56]. (Successive outcomes from the original TPS algorithm are correlated due to the use of small momentum perturbations between trajectories.) During aimless shooting, the outcome (A or B) of each forward shooting point is saved regardless of whether the shooting move is accepted or rejected, and the configuration at each shooting point is also saved. Each order parameter that may (or may not) be part of the reaction coordinate is evaluated at each shooting point. This results in a data file with one row for each of the n shooting points as shown below [56,57].

Shooting point	Forward outcome	Candidate order parameter values			
$\boldsymbol{x}_1$	B	$q_1(\boldsymbol{x}_1)$	$q_2(\boldsymbol{x}_1)$	...	$q_N(\boldsymbol{x}_1)$
$\boldsymbol{x}_2$	A	$q_1(\boldsymbol{x}_1)$	$q_2(\boldsymbol{x}_1)$	...	$q_N(\boldsymbol{x}_1)$
$\boldsymbol{x}_3$	A	$q_1(\boldsymbol{x}_1)$	$q_2(\boldsymbol{x}_1)$	...	$q_N(\boldsymbol{x}_1)$
...	...			...	
$\boldsymbol{x}_{N\text{sp}}$	B	$q_1(\boldsymbol{x}_1)$	$q_2(\boldsymbol{x}_1)$	...	$q_N(\boldsymbol{x}_1)$

The tabulated data are then used to optimise models of the committor probability. One constraint on model reaction coordinates is that the committor probability

should approach 0 at small values of the reaction coordinate, and it should approach 1 for large values of the reaction coordinate. We use a single sigmoid function

$$p_{\mathrm{B}}(r) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}(r), \tag{8}$$

where p_{B} models can also be constructed from the hyperbolic tangent function as suggested in the original paper [56] or from highly flexible multilevel neurons as used by Ma and Dinner [59]. In the single-level perceptron models, the argument of the sigmoid function should be some monotonic function of the candidate collective variables. In the simplest case, $r(q)$ may be a linear combination of M collective variables as follows:

$$r(q) = \alpha_0 + \sum_{k=1}^{k=M} \alpha_k q_k. \tag{9}$$

The linear combination model is actually quite flexible because the collective variables themselves are nonlinear and nonlinear combinations of collective variables may be added to the list of candidates. For example, in Lennard-Jones nucleation simulations, Beckham and Peters [103] included products of nucleus size coordinates and nucleus structure coordinates. If the model reaction coordinate from Equations (8) and (9) is correct, then the likelihood of observing the data in the table is

$$L = \prod_{x_k \to \mathrm{B}} p_{\mathrm{B}}(r(x_k)) \prod_{x_k \to \mathrm{A}} (1 - p_{\mathrm{B}}(r(x_k))). \tag{10}$$

Of course, most model reaction coordinates are not correct and these will give very small likelihood scores. In Equation (10), $x_k \to \mathrm{B}$ denotes a product over all shooting points (x_k) that result in forward trajectories that committed to basin B and the other product is over shooting points where trajectories committed to $x_k \to \mathrm{A}$. The coefficients in the model reaction coordinate of Equation (9) are varied to maximise the likelihood score in Equation (10) [56,57]. Starting from $M = 1$, likelihood maximisation is performed for each M-component reaction coordinate model, i.e. for a total of $N!/(M!(N - M)!)$ models. The resulting M-component reaction coordinate models are ranked in terms of their log-likelihood scores to identify the best M-component model. Models with M components can then be compared to models with $M+1$ components, but the effect of an additional adjustable parameter from Equation (9) must be included in the comparison. The best model with $M+1$ component variables will always have a higher likelihood score than the best model with M-component variables. However, the improvement may not be physically significant. Peters and Trout suggested the Bayesian information criterion (BIC) to decide whether an improvement was significant. The BIC is given by $(1/2)\ln(N_{\mathrm{sp}})$, where N_{sp} is the number of aimless shooting points used to compute the likelihood score. If the likelihood scores for two models with a different number of components differ by more than the BIC, the additional component variable in the model is probably a physically significant improvement. The BIC test is thus useful in determining how many component variables from the list of candidates are needed to adequately describe the reaction coordinate [56,57].

Figure 10 shows that aimless shooting generates shooting points near the separatrix ($p_{\mathrm{B}} = 1/2$) and within the transition pathway. The shooting points on the Muller–Brown PES [107] are coloured red or green depending on whether the forward trajectory from the shooting point committed to B or to A, respectively. There is an easily discernible trend in the colouration that is detected by a large likelihood score for the coordinate that points in this direction [57].

The example above provides an illustration of the method, but as emphasised in Section 1, such models obscure the essence of the true physical problem. The relevant subset of coordinates and the FES in the model system were given; but in a physical problem, the relevant coordinates must be identified without having the FES in any of the many coordinate subspaces that may be important. Indeed, many methods can identify an optimal dividing surface from an FES in the subspace of pre-identified coordinates. Only GNN, likelihood maximisation and a related method by Borrero and Escobedo [108] developed for use with forward flux

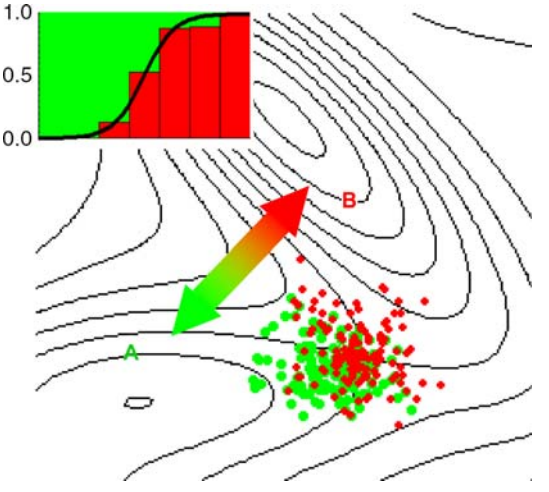

Figure 10. Aimless shooting points on a model PES. The points are coloured red if the resulting trajectory went to B and green if the resulting trajectory went to A. From the shooting point data, likelihood maximisation detects the direction and lengthscale on which the committor probability changes from 0 to 1 (inset). Note that the shooting points are clustered in the reaction pathway and near the transition state location. The Muller–Brown surface with Langevin dynamics was used in this calculation [57] (colour online).

sampling [109] can identify optimal coordinates from the dynamics. Figure 11 illustrates how likelihood maximisation with aimless shooting data can identify an accurate reaction coordinates in high-dimensional systems where the dynamics are best described by collective variables.

To ensure that likelihood maximisation has identified an appropriate reaction coordinate, the predicted reaction coordinate should still be subjected to a histogram test [56]. This is necessary because the aimless shooting data are harvested from a $p(x|\mathrm{TP})$-weighted transition path ensemble and not from the equilibrium ensemble. By contrast, the committor probability distribution on a reaction coordinate isosurface reflects an *equilibrium distribution of states on the reaction coordinate isosurface*. Because of the different ensembles, aimless shooting and likelihood maximisation may identify a coordinate that accurately describes dynamics along the reaction channel, but fails to separate progress along the reaction pathway from excursions into an off pathway branch of the stable basin A. Problems like this have not been encountered in previous applications, but still the final reaction coordinate should be rigorously tested for accuracy. In performing the final test of reaction coordinate accuracy, the improvements by Peters [58] can substantially reduce the computational demands of the histogram test.

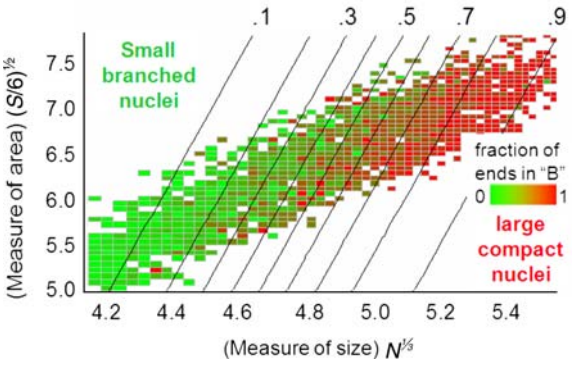

Figure 11. Aimless shooting points have been projected onto coordinates $N^{1/3}$ and $(S/6)^{1/2}$ that measure nucleus size and nucleus surface area in a simulation of the lattice gas model. Because of the projection, each bin may contain several shooting point outcomes. The bins have been coloured according to the fraction of trajectories from each bin that committed to the stable phase 'B'. The diagonal lines show the predicted isocommittor surfaces from likelihood maximisation with each isocommittor labelled with the predicted value of p_B. The $p_B = 0.5$ surface reveals that critical nuclei can be smaller than the typical size if they are very compact or larger than the typical size if they are branched. When only nucleus size was included as a candidate variable, likelihood maximisation identified the exact same critical size ($N* = 115$) that was identified by Pan and Chandler in an earlier study using the traditional committor analysis (colour online).

7. From an accurate reaction coordinate to the rate constant

Many prevalent rate theories are formulated as corrections to the TST rate constant [26]. For elementary reaction steps that break and make strong bonds, TST rate constants have been sufficiently accurate to perform sophisticated analyses of reaction networks [1]. As explained in Section 1, friction between the reaction coordinate and the bath tends to be sufficiently weak that the initial barrier crossing occurs successfully. Then, for a high barrier corresponding to strong bond breaking and making, even weak friction dissipates enough of the large amount of energy in the reaction coordinate to quench the trajectory in the product state [15]. Thus, TST for the breaking of strong bonds, even in the complex environment of an enzyme active site [110–112], tends to be a good approximation. For such reactions, path sampling methods are not usually necessary because harmonic TST with dynamical corrections and tunnelling corrections for hydrogen transfer reactions [113] will be less expensive and just as accurate. An optimal strategy for these systems is to focus computational effort on an accurate quantum mechanical description of the bond-breaking and bond-making energetics [114], and on accurately identifying all relevant saddle points leading out of each stable state [115–118].

7.1 Moderate to high friction

In contrast to reactions that break and make a few strong bonds, activated processes that alter many weak interactions tend to exhibit more recrossing due to friction between motion along the reaction coordinate and the bath [15]. Bennett–Chandler and related methods [64,119,120] can provide a transmission coefficient that accounts for barrier recrossing in the rate constant. The Bennett–Chandler method correlates the flux through the dividing surface at the initial moment of crossing with the requirement that trajectories carrying the initial flux remain in the product state some time later. The reactive flux approaches the TST flux at the initial time $t = 0$, and decays to a plateau as the trajectories relax from the barrier top into the stable basins. When normalised by the TST flux, the reactive flux correlation function decays from a value of 1 at $t = 0$ to a value of κ at the plateau [64]

$$\kappa(t) = \frac{\langle \dot{q}(0)\delta[q(0) - q^*]h[q(t) - q^*]\rangle}{\langle |\dot{q}(0)|\delta[q(0) - q^*]/2\rangle}. \qquad (11)$$

For extremely long time, i.e. the timescale between reaction events, the reactive flux correlation function decays to zero. The plateau value is the transmission coefficient, a dynamical correction to the rate constant [64]

$$k = \kappa(\tau_{\mathrm{plateau}})k_{\mathrm{TST}}. \qquad (12)$$

For every strong friction, i.e. for every small κ, methods similar to that of Ruiz-Montero et al. [121] or Vanden-Eijnden and Tal [25] are needed to accurately compute the transmission coefficient. The transmission coefficient also serves as a correction for any residual errors in the reaction coordinate [65].

The transmission coefficient calculation involves tracking the evolution of a swarm of trajectories initiated from the dividing surface at time $t = 0$. For an accurate reaction coordinate, this evolution should be consistent with the features of the FES. An example of the correspondence between the FES and the dynamics for an accurate reaction coordinate is shown in Figure 12. The example also serves as a counterexample to the common (but incorrect) view that optimal dividing surfaces must rigidly be identified with the $p_B = 1/2$ surface. The optimum location depends on how the dividing surface will be used to compute a rate constant. $p_B = 1/2$ is a good choice for a mean first passage time calculation [122], but for computing a transmission coefficient correction to the TST rate constant, $p_B = 1/2$ may not be an optimal surface.

Figure 12 [28] shows that p_B-isosurfaces other than $p_B = 1/2$ may be preferable when $p_B = 1/2$ corresponds to a stable intermediate between A and B. In this example, $q = 0$ is the best estimate for $p_B \approx 1/2$ by symmetry, but the free energy is approximately $2kT$ higher at the red dot, where $p_B \approx 0.65$ and where the transmission coefficient is $\kappa = 0.24$. If the transmission coefficient was evaluated at the plane of symmetry instead, we can anticipate that the transmission coefficient would be reduced by a factor of approximately $\exp[-2]$ to preserve the overall rate constant.

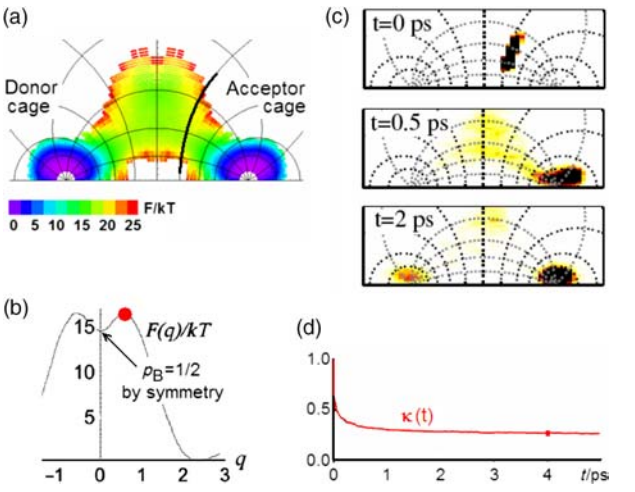

Figure 12. (a) FES for methane molecule hopping between two cages of a methane hydrate. The thick black line denotes the dividing surface used to calculate the TST rate constant and the transmission coefficient. By symmetry, the chosen surface is not at $p_B = 1/2$, but instead at $p_B = 0.65$. This choice was made to avoid the broad shallow basin at $p_B = 1/2$, the plane of symmetry which would yield a very low transmission coefficient, κ. (b) The free energy projected onto a single (spherical bipolar) coordinate reveals a shallow intermediate of depth approximately $2kT$. The red dot corresponds to the blackened isosurface in (a). (c) Snapshots of the swarm of trajectories evolving in time. Initially, the swarm splits with half of the trajectories immediately committing to the acceptor minimum and half of the trajectories entering the shallow intermediate basin. Then, approximately half of the trajectories in the intermediate basin escape to A and half escape to B. (d) The reactive flux correlation function indicating the plateau value of $\kappa = 0.24$ (colour online).

7.2 The high-friction limit: diffusive barrier-crossing dynamics

Barrier-crossing dynamics of nucleation processes and conformational changes of solvated macromolecules are likely to fall in the intermediate to high-friction Kramers' regime [21,123,124]. Computing rates in the high-friction limit is remarkably simple when a 1D reaction coordinate can be identified [125]. Now with methods such as GNN [59] and likelihood maximisation [56,57], this first step is feasible. In the high-friction limit, motion along the reaction coordinate resembles a 1D diffusion process. For diffusive dynamics from the top of a locally parabolic barrier,

$$\beta F(q) \approx \beta F(q^*) - \frac{1}{2}\omega^2 (q - q^*)^2, \qquad (13)$$

the Smoluchowski equation [122] with a delta-function initial condition at the top of the parabolic barrier can be used to show that the diffusivity D_{q^*} along the reaction coordinate is related to the mean-squared displacement by [126]

$$\langle \omega^2 (q(t) - q^*)^2 \rangle = \exp[2\omega^2 D_{q^*} t] - 1. \qquad (14)$$

To leading order this result is identical to the Einstein relation [127], but higher-order terms lead to a mean-squared displacement that exceeds the Einstein prediction. Figure 13 shows the free energy projected onto an accurate reaction coordinate for a model of nucleation from solution [126]. Figure 14 shows the mean-squared displacement for trajectories initiated at the top of the nucleation barrier as a function of time. In Figure 14, Equation (14) has been rearranged to enable a simple linear regression for the diffusivity along the reaction coordinate. The line in Figure 14 is not perfectly straight for two reasons. First, the barrier is not exactly parabolic. Second, the diffusivity along the reaction coordinate slowly varies with position along the reaction coordinate. Because diffusivity along a nucleus size coordinate is an attachment frequency, it should scale with nucleus surface area [128].

Equation (14) provides an accurate way to compute the diffusivity, which is tantamount to knowing the transmission

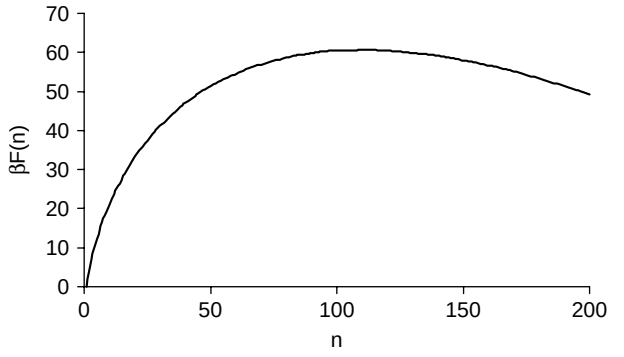

Figure 13. Free energy as a function of nucleus size, an accurate reaction coordinate for this lattice model of nucleation from solution [126].

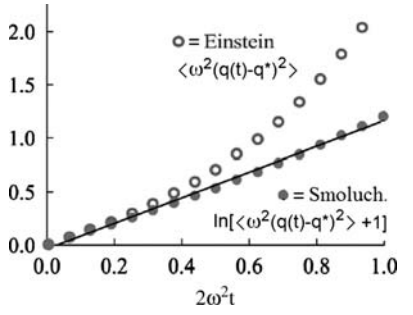

Figure 14. The mean-squared displacement for motion along the reaction coordinate was computed for simulations initiated at the top of the nucleation barrier in Figure 13. The mean-squared displacement depends nonlinearly on time because trajectories drift away from the barrier top at increasing rates as distance from the barrier top increases. Equation (14) predicts a function of the mean-squared displacement that does remain linear over longer times. The dots in the figure are actual simulation data [126].

coefficient for a highly diffusive barrier crossing. The diffusivity enters the high-friction rate constant as a prefactor:

$$k = D_{q^*}\left[\int_{\cup} \exp[-\beta F(q)]\mathrm{d}q \int_{\cap} \exp[+\beta F(q)]\mathrm{d}q\right]^{-1}. \tag{15}$$

The notations $\int_{\cup}$ and $\int_{\cap}$ were borrowed from Berezhkovskii and Szabo [125] to indicate that only the bottom of the well and the top of the barrier contribute to the two integrals, respectively. In a forthcoming paper, we will show how detailed knowledge of the barrier-crossing dynamics, as embodied by Equation (14), may be used to help test theories about laser-induced nucleation. We anticipate that barrier-crossing dynamics may be important in some other contexts as well.

8. Conclusions

This review gave particular emphasis to new ways of obtaining accurate reaction coordinates for activated processes with diffusive barrier-crossing dynamics. Although diffusive barrier crossings have been theoretically understood since the work of Kramers [129] and Langer [124], they have posed significant challenges for molecular simulation. These challenges have been twofold. First, coordinate-independent path sampling methods have struggled to sample path space efficiently for activated processes with diffusive dynamics. Second, accurate reaction coordinates for processes with diffusive dynamics have been difficult to identify among the many collective variables that might potentially be important.

Aimless shooting and likelihood maximisation address both of those challenges. Aimless shooting is able to maintain high acceptance rates for highly diffusive dynamics, and likelihood maximisation uses shooting point data from aimless shooting to identify accurate reaction coordinates. We illustrated these calculations for a model FES and for several nucleation problems.

The reaction coordinate itself is one of the key insights that one hopes to gain from simulating an activated process. From the reaction coordinate, one can also identify the transition state ensemble. These two pieces of information, the reaction coordinate and the nature of transition states, are tantamount to a complete understanding of the reaction mechanism. These advances in the identification of reaction coordinates should also facilitate many other methods that require *a priori* specification of a reaction coordinate or reaction coordinate components.

We also demonstrated how the FES becomes consistent with the dynamics when projected onto an accurate reaction coordinate. This internal consistency imparts confidence about the computed activation barriers and facilitates the calculation of accurate rate constants. Once an accurate 1D reaction coordinate is identified, computing rate constants in the high-friction limit is reduced to computing the free energy profile, estimating the diffusion constant for motion along the reaction coordinate at the top of the barrier, and straightforward integration to compute a mean first passage time. For reactions that exhibit ballistic motion over the barrier, the rate can be estimated from the TST with a transmission coefficient to correct for any barrier recrossing or tunnelling. New replica exchange versions of TPS that efficiently sample multiple reaction channels suggest that TPS methods will continue to improve [130].

Dr Gregg Beckham has developed scripts to implement aimless shooting in CHARMM and NAMD. These scripts are available for download, along with code that performs likelihood maximisation at www.engineering.ucsb.edu/ ~ baronp/codes.html.

Acknowledgements

We thank Peter Bolhuis, Eric Vanden-Eijnden, Giovanni Ciccotti, Aaron Dinner, David Chandler, Berend Smit and Ravi Radhakrishnan for years of stimulating interactions. We thank Graeme Henkelman for discussions about variational TST. We also thank Gregg Beckham, Bernhardt Trout, Mike Doherty and Brandon Knott for central contributions to the ideas and examples in this work. We acknowledge the support of The American Chemical Society Petroleum Research Fund.

References

[1] C.H. Christensen and J.K. Norskov, *A molecular view of heterogeneous catalysis*, J. Chem. Phys. 128 (2008), p. 182503.

[2] D. Erdemir, A.Y. Lee, and A.S. Myerson, *Polymorph selection: The role of nucleation, crystal growth, and molecular modeling*, Curr. Opin. Drug Discov. Dev. 10 (2007), pp. 746–755.

[3] M.H. Head-Gordon, *Quantum chemistry and molecular processes*, J. Phys. Chem. 100 (1996), pp. 13213–13225.

[4] W. Kohn, A.D. Becke, and R.G. Parr, *Density functional theory of electronic structure*, J. Phys. Chem. 100 (1996), pp. 12974–12980.

[5] W.H. Miller, *Spiers memorial lecture: Quantum and semiclassical theory of chemical reaction rates*, Faraday Discuss. 110 (1998), pp. 1–21.

[6] W.H. Miller, N.C. Handy, and J.E. Adams, *Reaction path Hamiltonian for polyatomic molecules*, J. Chem. Phys. 72 (1990), pp. 99–112.

[7] C.J. Cerjan and W.H. Miller, *On finding transition states*, J. Chem. Phys. 75 (1981), pp. 2800–2806.

[8] H. Jonsson, G. Mills, and K.W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions*, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, B.J. Berne, G. Ciccotti, and D.F. Coker, eds., World Scientific, Singapore, 1998.

[9] J. Baker, *An algorithm for the location of transition states*, J. Comp. Chem. 7 (1986), pp. 385–395.

[10] G. Henkelman and H. Jonsson, *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives*, J. Chem. Phys. 111 (1999), pp. 7010–7022.

[11] E. Weinan, W.Q. Ren, and E. Vanden-Eijnden, *String method for the study of rare events*, Phys. Rev. B 66 (2002), p. 052301.

[12] B. Peters, A. Heyden, A.T. Bell, and A.K. Chakraborty, *A growing string method for determining transition states: Comparison to the nudged elastic band and string methods*, J. Chem. Phys. 120 (2004), pp. 7877–7886.

[13] A. Heyden, A.T. Bell, and F.J. Keil, *Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method*, J. Chem. Phys. 123 (2005), p. 224101.

[14] D.G. Truhlar, B.C. Garrett, and S.J. Klippenstein, *Current status of transition state theory*, J. Phys. Chem. 100 (1996), pp. 12771–12800.

[15] W.T. Coffey, Y.P. Kalmykov, and J.T. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry, and Electrical Engineering*, 2nd ed., World Scientific, Singapore, 2004.

[16] B. Peters, A.T. Bell, and A.K. Chakraborty, *Rate constants from the reaction path Hamiltonian I: Reactive flux simulations for dynamically correct rates*, J. Chem. Phys. 121 (2004), pp. 4453–4460.

[17] L.P. Ju, K.L. Han, and J.Z.H. Zhang, *Global dynamics and transition state theories: Comparative study of reaction rate constants for gas-phase chemical reactions*, J. Comp. Chem. 30 (2009), pp. 305–316.

[18] J.N. Onuchic, Z. Luthey-Schulten, and P. Wolynes, *Theory of protein folding: The energy landscape perspective*, Ann. Rev. Phys. Chem. 48 (1997), pp. 545–600.

[19] P.R. Ten Wolde, M.J. Ruiz-Montero, and D. Frenkel, *Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling*, J. Chem. Phys. 104 (1996), pp. 9932–9947.

[20] S.C. Glotzer, M.J. Solomon, and N.A. Kotov, *Self-assembly: From nanoscale to microscale colloids*, AIChE J. 50 (2004), pp. 2978–2985.

[21] K. Schulten, A. Schulten, and A. Szabo, *Dynamics of reactions involving diffusive barrier crossing*, J. Chem. Phys. 74 (1981), pp. 4426–4432.

[22] D.F. Calef and P. Wolynes, *Classical solvent dynamics and electron transfer. 1. Continuum theory*, J. Phys. Chem. 87 (1983), pp. 3387–3400.

[23] A. Warshel, *Dynamics of reactions in polar solvents: Semiclassical trajectory studies of electron transfer and proton transfer reactions*, J. Phys. Chem. 86 (1982), pp. 2218–2224.

[24] R. Best and G. Hummer, *Reaction coordinates and rates from transition paths*, Proc. Natl Acad. Sci. USA 102 (2005), pp. 6732–6737.

[25] E. Vanden-Eijnden and F.A. Tal, *Transition state theory: Variational formulation, dynamical corrections, and error estimates*, J. Chem. Phys. 123 (2005), p. 184103.

[26] P. Hanggi, P. Talkner, and M. Borkovec, *Reaction rate theory – 50 years after Kramers*, Rev. Mod. Phys. 62 (1990), pp. 251–341.

[27] G.K. Schenter, B.C. Garrett, and D.G. Truhlar, *Generalized transition state theory in terms of the potential of mean force*, J. Chem. Phys. 119 (2003), pp. 5828–5833.

[28] B. Peters, N.E.R. Zimmerman, G.T. Beckham, J.W. Tester, and B.L. Trout, *Path sampling calculation of methane diffusivity in natural gas hydrates from a water-vacancy assisted mechanism*, J. Am. Chem. Soc. 130 (2008), pp. 17342–17350.

[29] L. Xu, G. Henkelman, C.T. Campbell, and H. Jonsson, *Pd diffusion on MgO(100): Energetics, structure, and role of defects*, Surf. Sci. 600 (2006), pp. 1351–1362.

[30] J.R. Murdoch, *What is the rate limiting step of a multistep reaction?*, J. Chem. Edu. 58 (1) (1981), pp. 32–36.

[31] B. Ensing and M.L. Klein, *Perspective on the reactions between F^- and CH_3CH_2F: The free energy landscape of the E2 and S_N2 reaction channels*, Proc. Natl Acad. Sci. USA 102 (2005), pp. 6755–6759.

[32] A.F. Voter, *Hyperdynamics: Accelerated molecular dynamics of infrequent events*, Phys. Rev. Lett. 78 (1997), pp. 3908–3911.

[33] M.R. Sorenson and A.F. Voter, *Temperature accelerated dynamics for simulation of infrequent events*, J. Chem. Phys. 112 (2000), pp. 9599–9606.

[34] A.F. Voter, *A method for accelerating the molecular dynamics simulation of infrequent events*, J. Chem. Phys. 106 (1997), pp. 4665–4677.

[35] A. Miron and K.A. Fichthorn, *Accelerated molecular dynamics with the bond-boost method*, J. Chem. Phys. 119 (2003), pp. 6210–6216.

[36] F. Noe, I.S. Horenko, C. Schütte, and J.C. Smith, *Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states*, J. Chem. Phys. 126 (2007), p. 155102.

[37] J.D. Chodera, N. Singhal, V.S. Pande, K.A. Dill, and W.C. Swope, *Automatic discovery of Metastable states for the construction of Markov models of macromolecular conformational dynamics*, J. Chem. Phys. 126 (2007), p. 155101.

[38] N.V. Buchete and G. Hummer, *Coarse master equations for peptide folding dynamics*, J. Phys. Chem. B 112 (2008), pp. 6057–6069.

[39] E. Vanden-Eijnden, *Transition path theory*, in *Computer Simulations in Condensed Matter: From Materials for Chemical Biology*, G.C.M. Ferrario and K. Binder, eds., Vol. 2, Springer-Verlag, Berlin, 2006.

[40] P. Metzner, C. Schutte, and E. Vanden-Eijnden, *Illustration of transition path theory on a collection of simple examples*, J. Chem. Phys. 125 (2006), p. 084110.

[41] P. Metzner, C. Schutte, and E. Vanden-Eijnden, *Transition path theory for Markov jump processes*, Multiscale Model Simul. 7 (2009), pp. 1192–1219.

[42] E.A. Carter, G. Ciccotti, J.T. Hynes, and R. Kapral, *Constrained reaction coordinate dynamics for the simulation of rare events*, Chem. Phys. Lett. 156 (1989), pp. 472–477.

[43] D.J. Tobias and C.L. Brooks, *Calculation of free energy surfaces using the methods of thermodynamic perturbation theory*, Chem. Phys. Lett. 142 (1987), pp. 472–476.
[44] J.G. Kirkwood, *Statistical mechanics of fluid mixtures*, J. Chem. Phys. 3 (1935), pp. 300–313.
[45] G.M. Torrie and J.P. Valleau, *Non-physical sampling distributions in Monte-Carlo free energy estimation – Umbrella sampling*, J. Comp. Phys. 23 (1977), pp. 187–199.
[46] A.M. Ferrenberg and R.H. Swendsen, *Optimized Monte-Carlo data-analysis*, Phys. Rev. Lett. 63 (1989), pp. 1195–1198.
[47] F. Wang and D.P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*, Phys. Rev. Lett. 86 (2001), pp. 2050–2053.
[48] A. Laio and M. Parrinello, *Escaping free energy minima*, Proc. Natl Acad. Sci. USA 99 (2002), pp. 12562–12566.
[49] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *String method in collective variables: Minimum free energy path and isocommittor surfaces*, J. Chem. Phys. 125 (2006), p. 024106.
[50] L.R. Pratt, *A statistical method for identifying transition states in high dimensional problems*, J. Chem. Phys. 85 (1986), pp. 5045–5048.
[51] P.G. Bolhuis, C. Dellago, and D. Chandler, *Sampling ensembles of deterministic pathways*, Faraday Discuss. 110 (1998), pp. 421–436.
[52] C. Dellago, P.G. Bolhuis, and D. Chandler, *Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements*, J. Chem. Phys. 108 (1998), pp. 9236–9245.
[53] C. Dellago, P.G. Bolhuis, F.S. Csajka, and D. Chandler, *Transition path sampling and the calculation of rate constants*, J. Chem. Phys. 108 (1998), pp. 1964–1977.
[54] C. Dellago, P.G. Bolhuis, and D. Chandler, *On the calculation of rate constants in the transition path ensemble*, J. Chem. Phys. 110 (1999), pp. 6617–6625.
[55] P.G. Geissler, C. Dellago, and D. Chandler, *Kinetic pathways of ion pair dissociation in water*, J. Phys. Chem. B 103 (1999), pp. 3706–3710.
[56] B. Peters and B.L. Trout, *Obtaining reaction coordinates by likelihood maximization*, J. Chem. Phys. 125 (2006), p. 054108.
[57] B. Peters, G.T. Beckham, and B.L. Trout, *Extensions to the likelihood maximization approach for finding reaction coordinates*, J. Chem. Phys. 127 (2007), p. 034109.
[58] B. Peters, *Using the histogram test to quantify reaction coordinate error*, J. Chem. Phys. 125 (2006), p. 241101.
[59] A. Ma and A.R. Dinner, *Automatic method for identifying reaction coordinates in complex systems*, J. Phys. Chem. B 109 (2005), pp. 6769–6779.
[60] P.G. Bolhuis, D. Chandler, C. Dellago, and P.G. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, Ann. Rev. Phys. Chem. 53 (2002), pp. 291–318.
[61] C. Dellago, P.G. Bolhuis, and P.G. Geissler, *Transition path sampling*, Adv. Chem. Phys. 123 (2001), pp. 1–78.
[62] B.C. Garrett and D.G. Truhlar, *Variational transition state theory*, Ann. Rev. Phys. Chem. 35 (1984), pp. 159–189.
[63] J. Keck, *Variational theory of reaction rates*, Adv. Chem. Phys. 13 (1967), pp. 85–121.
[64] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, CA, 2002.
[65] T.S. van Erp, *Efficiency analysis of reaction rate calculation methods using analytical models I: The two-dimensional sharp barrier*, J. Chem. Phys. 125 (2006), p. 174106.
[66] L. Onsager, *Initial recombination of ions*, Phys. Rev. 54 (1938), pp. 554–557.
[67] T.A. McCormick and D. Chandler, *Grid-flux method for learning the solvent contribution to the mechanisms of reactions*, J. Phys. Chem. B 107 (2003), pp. 2796–2801.
[68] B.M. Pettitt and M. Karplus, *The potential of mean force surface for the alanine dipeptide in aqueous solution – a theoretical approach*, Chem. Phys. Lett. 121 (1985), pp. 194–201.
[69] T. Lazaridis, D.J. Tobias, C.L. Brooks, and M.E. Paulaitis, *Reaction paths and free-energy profiles for conformational transitions – in internal coordinate approach*, J. Chem. Phys. 95 (1991), pp. 7612–7625.
[70] D.J. Tobias and C.L. Brooks, *Conformational equilibrium in the alanine dipeptide in the gas-phase and aqueous solution – a comparison of theoretical results*, J. Phys. Chem. 96 (1992), pp. 3864–3870.
[71] T.J. Marrone, M.K. Gilson, and J.A. McCammon, *Comparison of continuum and explicit models of solvation: Potentials of mean force for alanine dipeptide*, J. Phys. Chem. 100 (1996), pp. 1439–1441.
[72] C. Bartels and M. Karplus, *Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations*, J. Comp. Chem. 18 (1997), pp. 1450–1462.
[73] M. Scarsi, J. Apostolakis, and A. Caflisch, *Comparison of a GB solvation model with explicit solvent simulations: Potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane*, J. Phys. Chem. B 102 (1998), pp. 3637–3641.
[74] J. Apostolakis, P. Ferrara, and A. Caflisch, *Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water*, J. Chem. Phys. 110 (1999), pp. 2099–2108.
[75] A. Laio and F.L. Gervasio, *Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry, and material science*, Rep. Prog. Phys. 71 (2008), pp. 126601–126622.
[76] P.G. Bolhuis, C. Dellago, and D. Chandler, *Reaction coordinates of biomolecular isomerization*, Proc. Natl Acad. Sci. USA 97 (2000), pp. 5877–5882.
[77] R.J. Dimelow, R.A. Bryce, A.J. Masters, I.H. Hillier, and N.A. Burton, *Exploring reaction pathways with transition path and umbrella sampling: Application to methyl maltoside*, J. Chem. Phys. 124 (2006), p. 114113.
[78] J. Juraszek and P.G. Bolhuis, *Rate constant and reaction coordinate of Trp-Cage folding in explicit water*, Biophys. J. 95 (2008), pp. 4246–4257.
[79] J. Hu, A. Ma, and A.R. Dinner, *A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT*, Proc. Natl Acad. Sci. USA 105 (2008), pp. 4615–4620.
[80] R. Pool and P.G. Bolhuis, *Sampling the kinetic pathways of a micelle fusion and fission transition*, J. Chem. Phys. 126 (2007), p. 244703.
[81] P.R. Ten Wolde and D. Chandler, *Drying-induced hydrophobic polymer collapse*, Proc. Natl Acad. Sci. USA 99 (2002), pp. 6539–6543.
[82] A.P. Willard and D. Chandler, *The role of solvent fluctuations in hydrophobic assembly*, J. Phys. Chem. B 112 (2008), pp. 6187–6192.
[83] P.G. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, *Autoionization in liquid water*, Science 291 (2001), pp. 2121–2124.
[84] R. Radhakrishnan and B.L. Trout, *Nucleation of hexagonal ice (I-h) in liquid water*, J. Am. Chem. Soc. 125 (2003), pp. 7743–7747.
[85] A.C. Pan and D. Chandler, *Dynamics of nucleation in the Ising model*, J. Phys. Chem. B 108 (2004), pp. 19681–19686.
[86] M. Grunwald and C. Dellago, *Transition state analysis of solid–solid transformations in nanocrystals*, J. Chem. Phys. 131 (2009), p. 154116.
[87] M. Grunwald and C. Dellago, *Nucleation and growth in structural transitions in nanocrystals*, Nano Lett. 9 (2009), pp. 2099–2102.
[88] G.T. Beckham, B. Peters, N. Variankaval, C. Starbuck, and B.L. Trout, *Surface mediated nucleation in the solid state polymorph transformation of terephthalic acid*, J. Am. Chem. Soc. 129 (2007), pp. 4714–4723.
[89] G.T. Beckham, B. Peters, and B.L. Trout, *Evidence for size-dependent nucleation mechanism in solid state polymorph transformations*, J. Phys. Chem. B 112 (2008), pp. 7460–7466.
[90] D. Zahn, *Exploring the mechanisms of reactions in solution from transition path sampling molecular dynamics simulations*, J. Chem. Theory Comp. 2 (2006), pp. 107–114.
[91] C. Lo, R. Radhakrishnan, and B.L. Trout, *Application of transition path sampling methods in catalysis: A new mechanism*

for C–C bond formation in the methanol coupling reaction in chabazite, Catal. Today 105 (2005), pp. 93–105.
[92] M.F. Hagan, A.R. Dinner, D. Chandler, and A.K. Chakraborty, *Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA*, Proc. Natl Acad. Sci. USA 100 (2003), pp. 13922–13927.
[93] P.G. Bolhuis, *Transition path sampling on diffusive barriers*, J. Phys. Cond. Matt. 15 (2003), pp. S113–S120.
[94] P.G. Bolhuis, *Transition-path sampling of beta-hairpin folding*, Proc. Natl Acad. Sci. USA 100 (2003), pp. 12129–12134.
[95] D. Moroni, P.G. Bolhuis, and T.S. van Erp, *Rate constants for diffusive processes by partial path sampling*, J. Chem. Phys. 120 (2004), pp. 4055–4065.
[96] T.S. van Erp, D. Moroni, and P.G. Bolhuis, *A novel path sampling method for the calculation of rate constants*, J. Chem. Phys. 118 (2003), pp. 7762–7774.
[97] T.S. van Erp and P.G. Bolhuis, *Elaborating transition interface sampling methods*, J. Comp. Phys. 205 (2005), pp. 157–181.
[98] M. Grunwald, C. Dellago, and P.G. Geissler, *Precision shooting: Sampling long transition pathways*, J. Chem. Phys. 129 (2008), p. 194101.
[99] T.F. Miller and C. Predescu, *Sampling diffusive transition paths*, J. Chem. Phys. 126 (2007), p. 144102.
[100] T.S. van Erp, *Reaction rate calculation by parallel path swapping*, Phys. Rev. Lett. 98 (2007), p. 268301.
[101] A.D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus, *An all-atom empirical energy function for the simulation of nucleic acids*, J. Am. Chem. Soc. 117 (1995), pp. 11946–11975.
[102] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable molecular dynamics with NAMD*, J. Comp. Chem. 26 (2005), pp. 1781–1802.
[103] G.T. Beckham and B. Peters (in preparation).
[104] G. Hummer and I.G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations*, J. Chem. Phys. 118 (2003), pp. 10762–10773.
[105] R. Du, V.S. Pande, A.Y. Grosberg, T. Tanaka, and E.S. Shakhnovich, *On the transition coordinate for protein folding*, J. Chem. Phys. 108 (1998), pp. 334–350.
[106] D. Moroni, P.R. Ten Wolde, and P.G. Bolhuis, *Interplay between structure and size in a critical crystal nucleus*, Phys. Rev. Lett. 94 (2005), p. 235703.
[107] K. Muller and L.D. Brown, *Location of saddle points and minimum energy paths by a constrained simplex optimization procedure*, Theor. Chem. Acta 53 (1979), pp. 75–93.
[108] E.E. Borrero and F.A. Escobedo, *Reaction coordinates and transition pathways of rare events via forward flux sampling*, J. Chem. Phys. 127 (2007), p. 164101.
[109] R.J. Allen, D. Frenkel, and P.R. Ten Wolde, *Simulating rare events in equilibrium or nonequilibrium stochastic systems*, J. Chem. Phys. 124 (2006), p. 024102.
[110] A. Warshel, P.K. Sharma, M. Kato, Y. Xiang, H. Liu, and M.H.M. Olsson, *Electrostatic basis for enzyme catalysis*, Chem. Rev. 106 (2006), pp. 3210–3235.
[111] M. Garcia-Viloca, J. Gao, M. Karplus, and D.G. Truhlar, *How enzymes work: Analysis by modern rate theory and computer simulations*, Science 303 (2004), pp. 186–195.
[112] H. Hu and W. Yang, *Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods*, Ann. Rev. Phys. Chem. 59 (2008), pp. 573–601.
[113] B. Peters, A.T. Bell, and A.K. Chakraborty, *Rate constants from the reaction path Hamiltonian II: Nonseparable semiclassical transition state theory*, J. Chem. Phys. 121 (2004), pp. 4461–4466.
[114] F. Jensen, *Introduction to Computational Chemistry*, Wiley, New York, 1999.
[115] B. Peters, W.-Z. Liang, A.T. Bell, and A.K. Chakraborty, *Biasing a transition state search to locate multiple reaction pathways*, J. Chem. Phys. 118 (2003), pp. 9533–9541.
[116] F. El-Mellouhi, N. Mousseau, and L.J. Lewis, *Kinetic activation–relaxation technique: An off-lattice self-learning kinetic Monte Carlo algorithm*, Phys. Rev. B 78 (2008), p. 153202.
[117] D.J. Wales and T.V. Bogdan, *Potential energy and free energy landscapes*, J. Phys. Chem. B 110 (2006), pp. 20765–20776.
[118] L. Xu and G. Henkelman, *Adaptive kinetic Monte Carlo for first-principles accelerated dynamics*, J. Chem. Phys. 129 (2008), p. 114104.
[119] B. Berne, M. Borkovec, and J.E. Straub, *Classical and modern methods in reaction rate theory*, J. Phys. Chem. 92 (1988), pp. 3711–3725.
[120] R.A. Kuharski, D. Chandler, J.A. Montgomery, F. Rabii, and S.J. Singer, *Stochastic molecular dynamics study of cyclohexane isomerization*, J. Phys. Chem. 92 (1988), pp. 3261–3267.
[121] M.J. Ruiz-Montero, D. Frenkel, and J.J. Brey, *Efficient schemes to compute diffusive barrier crossing rates*, Mol. Phys. 90 (1997), pp. 925–941.
[122] A. Nitzan, *Chemical Dynamics in Condensed Phases*, Oxford University Press, Oxford, 2006.
[123] D. Kaschiev, *Nucleation: Basic Theory and Applications*, Butterworth-Heineman, Burlington, MA, 2000.
[124] J.S. Langer, *Statistical theory of the decay of metastable states*, Ann. Phys. 54 (1969), pp. 258–275.
[125] A. Berezhkovskii and A. Szabo, *One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions*, J. Chem. Phys. 122 (2005), p. 014503.
[126] B. Knott, N. Duff, M.F. Doherty, and B. Peters, *Estimating diffusivity along a reaction coordinate in the high friction limit: Insights on pulse times in laser induced nucleation*, J. Chem. Phys 131 (2009), p. 224112.
[127] G.E. Uhlenbeck and L.S. Ornstein, *On the theory of the Brownian motion*, Phys. Rev. 36 (1930), pp. 823–841.
[128] S. Auer and D. Frenkel, *Quantitative prediction of crystal-nucleation rates for spherical colloids: A computational approach*, Ann. Rev. Phys. Chem. 55 (2004), pp. 333–361.
[129] H.A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica 7 (1940), pp. 284–304.
[130] P.G. Bolhuis, *Rare events via multiple reaction channels sampled by path replica exchange*, J. Chem. Phys. 129 (2008), p. 114108.