

Reaction Path Optimization and Sampling Methods and Their Applications for Rare Events

Peng Tao*, Joseph D. Larkin and Bernard R. Brooks
*Laboratory of Computational Biology,
National Heart, Lung and Blood Institute,
National Institutes of Health, Bethesda, Maryland
USA*

1. Introduction

Reaction mechanisms are an important tool for chemists in the determination of thermodynamic and kinetic properties of chemical reactions. (Hånggi & Borkovec 1990; Heidrich 1995; March 1992; Tolman 1925) The mechanisms are integral in the understanding of detailed molecular or chemical transitions from one equilibrium state (reactant) to another equilibrium state (product). In computational chemistry, the reaction mechanism is often represented as a reaction path on the Born-Oppenheimer potential energy surface (PES) of the system of interest through construction of a potential energy function of the nuclear coordinates. (Bader & Gangi 1975; Lewars 2011; Mezey 1987; Truhlar 2001; Wales 2003) The PES serves as an important theoretical construct to provide a framework to describe the transition between different states in detail. The equilibrium states correspond to local minimum on the PES with zero first order derivatives (gradient) in all directions and all positive eigenvalues of the second order derivative (Hessian) matrix, excluding rotation and translation degrees of freedom. The transition states (TSs), based on the transition state theory (TST), (Doll 2005; Eyring 1935; Laidler & King 1983; Pechukas 1981; Truhlar et al. 1983; Wigner 1938; Yamamoto 1960) are the first order saddle points with zero gradient and only one negative eigenvalue of the Hessian matrix. The equilibrium states are often easy to identify through experimental or computational studies. Understanding the detailed transition process between equilibrium states is of more interest in research, but unfortunately is very difficult to study experimentally. On a given PES, one can imagine that there could exist an infinite number of possible routes connecting two predefined states on that surface. However, not every route has the same weight in elucidating of reaction mechanisms. In the static point of view, the minimum energy path (MEP) is the route that needs the least amount of potential energy for the system to undergo the transition. The MEP connecting two local minima must go through one or more TSs, and is identified as a representative reaction path. In the statistical point of view, the minimum free energy path (MFEP) is the most probable transition path connects two metastable states. The simulation of the systems either through molecular dynamics (MD) or Monte Carlo (MC) sampling on the PES could generate an ensemble of transition paths, from which the MFEP can be identified. Both an MEP and an MFEP can be used to predict important properties, such as a reaction's kinetic isotope effect.

Although one could generate a reduced PES for complex systems with selected reaction coordinates,(Klähn et al. 2005; Shi et al. 2008; Tao et al. 2009a) it is rarely practical to construct a reduced PES of a system of interest to identify a reaction path connecting two minima. Moreover, reducing a complex multi-dimensional system to a simplified pathway is a form of data reduction. This reduction is non-unique and a choice imposed in this reduction will affect the quality and applicability of the results. It is more feasible to search for the MEP or MFEP directly on a given PES. Given the complexity and high degree of freedom of most systems of interest in chemistry, molecular biology and materials science, there has been a rapid development in methodologies for reaction pathway identification in large systems. As an attempt to reflect the current development and to better understand the consequences of specific methodological choices, this chapter reviews the recent progress of various methods to identify reaction paths with or without knowing the TS(s) *a priori*. Specifically, it emphasizes the applications of these methods in macromolecular systems.

It should be noted that identifying saddle points on a PES alone is not the focus of this report. This report does not cover the geometry optimization including equilibrium and transition structures,(Farkas & Schlegel 2003; Henkelman et al. 2000a; Olsen et al. 2004; Schlegel 1982, 2003, 2011) conformational sampling,(Beusen 1996; Leach 1991; Parish 2002) or global minimum search methodologies,(Floudas & Pardalos 1996; Horst 1995, 2000; Torn 1989) which are all important for the studies of computational chemistry and biology. Other related topics, including enhanced sampling methods,(Earl & Deem 2005; Hamelberg et al. 2004; Lei & Duan 2007; Okur et al. 2006; Sugita 1999; Swendsen & Wang 1986; Thomas et al. 2005; Wen et al. 2004) simulation of nonequilibrium states,(Bair et al. 2002; Cummings & Evans 1992; Hoover 1983; Hoover & Hoover 2005; Kjelstrup & Hafskjold 1996; Li et al. 2008; Mundy et al. 2000) and minimization methods,(Bonnans 2003; Dennis & Schnabel 1996; Fletcher 2000; Gill 1982; Haslinger & Mäkinen 2003; Nocedal 2006; Scales 1985) are not covered in this chapter either. The curious readers are welcome to read cited references for more information.

2. PES walking methods

Without the intention to generate a complete PES, it is logical to develop methods to explore the PES by walking along the surface from certain starting points using local information of the PES, such as the energy, gradient and even the Hessian.(Hratchian & Schlegel 2005a; Schlegel 2003, 2011) In this way, one can start from somewhere on the PES, either reactant, product, or TS, and walk uphill or downhill, depending on the starting points to reach the adjacent stationary points. The walking trajectories, after successfully reaching these stationary points, are the reaction pathways that describe the mechanism of transitions. For smaller systems, walking methods may be sufficient to fully understand a given reaction mechanism. However, for larger and more complex systems, pathways explored by such walking mechanisms are often not reversible, and can show significant hysteresis that results in a poor representation of the reaction.

2.1 Reaction path following methods

When using mass-weighted Cartesian coordinates, a steepest descent path from the TS down to the reactant and product is referred to as the intrinsic reaction coordinate (IRC) path.(Fukui 1981; Quapp & Heidrich 1984; Tachibana & Fukui 1980; Yamashita et al. 1981) The steepest descent pathway is given by the differential equation

$$\frac{d\mathbf{x}(s)}{ds} = -\frac{\mathbf{g}(\mathbf{x})}{|\mathbf{g}(\mathbf{x})|} \quad (1)$$

where \mathbf{x} is the vector of Cartesian coordinates, s is the step size of the path, and \mathbf{g} is the energy gradient at \mathbf{x} . The path obtained by solving this equation is the IRC, when \mathbf{x} is mass-weighted. Numerous methods were developed to locate the TS on a PES. (Baker 1986; Banerjee et al. 1985; Bell & Crighton 1984; Cerjan 1981; Ionova & Carter 1993, 1995; Jensen 1983; Müller & Brown 1979; Peng et al. 1996; Simons & Nichols 1990) The IRC could be optimized based on its variational nature. (Bofill & Quapp 2011; Quapp 2008) However, it is more applicable for many systems to construct the IRC by solving Eq. (1) from a TS.

Gonzalez and Schlegel developed the implicit trapezoid method (GS-IRC) for reaction path following at second order accuracy. (Gonzalez & Schlegel 1989, 1990) In their initial development, the points along the target reaction path are constructed by constrained optimization using internal degrees of freedom of the molecules. For each step of optimization along the path, the new point is constructed and optimized so that the gradient at each point is tangent to the path. Therefore, the resulting path is both continuous and differentiable. This initial method is correct to second order in the limit of small step size. The same method was later developed up to sixth order accuracy. (Gonzalez & Schlegel 1991) The GS-IRC method is generally efficient for small systems.

To improve the computational efficiency, the velocity Verlet algorithm (Verlet 1967) to propagate a classical dynamics trajectory was applied to integrate the IRC with a magnitude of the velocity damping for each step. (Hratchian & Schlegel 2002) This method is referred as the damped velocity Verlet (DVV) algorithm. The time step for each integration step is adjusted to ensure that the damped trajectory stays close to the IRC. The DVV-IRC method can be considered as running downhill along the PES from TS in a slow motion (by damping the velocity at each step). It enjoys the stability of the Verlet integrator and low cost of computation since the Hessian does not need to be calculated.

In their later work, Hratchian and Schlegel introduced an approach using a Hessian based predictor-corrector (HPC) integrator to solve Eq. (1). (Hratchian & Schlegel 2004) The HPC integrator comprises two steps: the predictor step and the corrector step. The gradient \mathbf{g} and Hessian \mathbf{H} of the system PES are used to calculate the predictor step with second order accuracy. Then, the correction of the predicted step is calculated through a modified Bulirsch-Stoer algorithm based on the gradient information at the predicted step. (Bulirsch & Stoer 1964, 1966a, b) Although, the HPC-IRC method is comparable to GS-IRC with fourth order accuracy, calculation of the Hessian at each step can be rather expensive for large systems. This bottleneck was resolved by applying a Hessian updating scheme in their later development. (Hratchian & Schlegel 2005b) For each step of an IRC calculation, the Hessian is not calculated *de novo*, but updated from the Hessian of the previous step and the change of the gradient and step size between two steps. With this scheme, the Hessian only needs to be calculated once at the TS, and then is updated at each step of the IRC calculation. This HPC-IRC method with Hessian updating has been applied successfully in large protein systems using a combined quantum mechanical and molecular mechanical (QM/MM) method. (Tao et al. 2010; Tao et al. 2009b; Zhou et al. 2010) In these studies, the inhibition mechanism of matrix metalloproteinase 2 (MMP2) by its potent inhibitor, was elucidated in great detail using QM/MM methods. The TS of the key reaction in the active site of MMP2 was identified. The IRC of the reaction including the protein environment was calculated to confirm that the reactant and product are connected through the identified TS (Fig. 1).

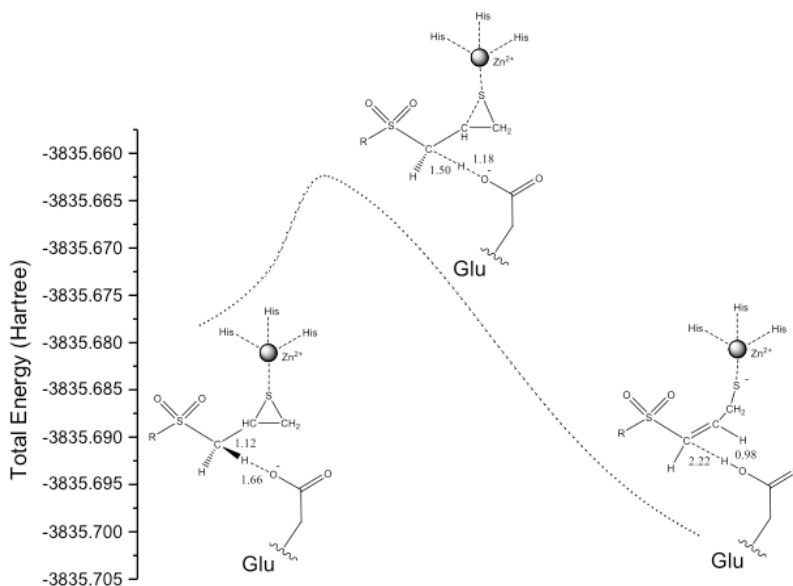


Fig. 1. IRC profile for SB-3CT in the MMP2 active site at the ONIOM(B3LYP/6-31G(d):AMBER) level of theory. Key bond lengths are in angstroms. (Reprinted with permission from ref. (Zhou et al. 2010). Copyright 2010 American Chemical Society.)

Very recently, Hratchian and Schlegel applied the Euler (first-order) predictor and corrector method (EulerPC) using an Euler explicit integrator in the calculation of predicted step. (Hratchian et al. 2010) This method avoids the expensive Hessian calculation at the TS and updating afterwards. By repeating the evaluation and correction several steps after the prediction, the error of the calculation is greatly reduced. The newly developed EulerPC method shows comparable accuracy with HPC but with much less computational cost, and is tested on several rather large enzymatic systems. (Hratchian & Frisch 2011)

As a summary, the IRC calculations are becoming practical even for large enzyme systems using the QM/MM approach. However, to apply any of the IRC methods listed in this section, a well-defined TS structure is necessary to serve as the starting point. For a large system, e.g. an enzymatic reaction system, using QM/MM methods may take substantial effort in identifying a TS.

2.2 Uphill walking methods

By walking uphill from a minimum, one could reach an adjacent TS. Applying a reaction path following method on the obtained TS could yield another minimum corresponding to a product or intermediate state and a complete reaction pathway could be formed. Simons and coworkers developed methods that walk on the PES toward the selected direction (either uphill or downhill) using local gradient and Hessian information. (Nichols et al. 1990; Simons & Nichols 1990; Taylor & Simons 1985) By applying a local quadratic approximation, the PES close to a starting structure \mathbf{x}_0 can be written as

$$E(\mathbf{x}) = E_0 + \mathbf{x}\mathbf{F}_0 + \frac{1}{2}\mathbf{x}\mathbf{H}_0\mathbf{x} . \quad (2)$$

where E_0 , \mathbf{F}_0 and \mathbf{H}_0 are the energy, gradient and Hessian at \mathbf{x}_0 , respectively. Vector \mathbf{x} is in the region around \mathbf{x}_0 , in which the local quadratic approximation is valid. In the attempt to walk uphill, the vector with the lowest Hessian eigenvalue will be chosen to define the moving direction. The potential energy increases along the chosen direction, but remains at minima along the other eigenvectors. For the best computational efficiency, the step size for each round of searching is controlled using eigenvalues of the Hessian matrix. After leaving the region in which the local quadratic approximation is valid, the Hessian matrix \mathbf{H} can be recomputed or updated for further calculations. Along the walk, the Hessian eigenvalue of the eigenvector, which is being followed may cross with other eigenvalues. This is likely to occur when the starting geometry is not within the quadratic approximation region of the TS. In such situations, a decision needs to be made to either keep track of the original eigenvector, or follow the eigenvector with the lowest eigenvalue after crossing. This decision may significantly affect the final TS.

An algorithm was developed by Ohno and Maeda to find reaction pathways on a PES systemically.(Maeda 2003; Ohno 2004; Ohno & Maeda 2006; Yang et al. 2005) In their method, the PES around the equilibrium structure (ES) is expanded using reduced normal coordinates in terms of normal coordinates \mathbf{Q}_i with eigenvalues λ_i ,

$$\mathbf{q}_i = \lambda_i^{\frac{1}{2}} \mathbf{Q}_i . \quad (3)$$

The degrees of freedom for translation and rotation are projected out from the normal coordinates. In such representation, any constant energy around the ES within the limit of harmonic potential gives spherical hypersurface (hypersphere) (Fig. 2). Any structure represented on a hypersphere is somewhat distorted from the ES around which the hypersphere is constructed. This common feature is referred to as anharmonic downward distortions following (ADDF), and is used for a reaction path search by walking uphill toward the direction with the local maxima of ADDF. Through a series of hyperspheres with different sizes but common origin, different reaction paths may be identified by following the local maxima of ADDF on each hypersphere. TS regions or dissociation channels (DC) can be identified through the variation of the first order derivatives along these reaction paths. Further calculations need to be carried out to precisely determine the real TS structures, which may not be on any hypersphere. The reaction path following methods can be applied to the TSs to find new ESs. The whole procedure can be repeated for new TSs. This is referred as scaled hypersphere search (SHS) method. Theoretically, the SHS method can be repeated until all the stationary points and reaction channels are identified for a given system. This method provides a means to systematically explore the PES and is referred to as global reaction route mapping (GRRM). By applying GRRM, the PESs of several small organic molecules were explored with numerous ESs and TSs identified for each system.(Maeda & Ohno 2007; Yang et al. 2005)

Recently, a new GRRM method was developed to search reaction pathways for large flexible systems using a microiteration-ADDF (μ -ADDF) technique.(Maeda et al. 2009) The microiteration scheme was originally developed for the QM/MM method.(Svensson et al. 1996; Vreven et al. 2006b; Vreven et al. 2006a; Vreven et al. 2003) For large systems, it may not be practical to follow all the ADDF maxima. Instead, two other methods were developed by the same authors to follow only large ADDF pathways (lADDF),(Maeda & Ohno 2007) or to follow the reverse direction from a point on a very large hypersphere to the sphere center with decreasing hypersphere radius (double-ended ADDF, or dADDF).(Maeda et al. 2009)

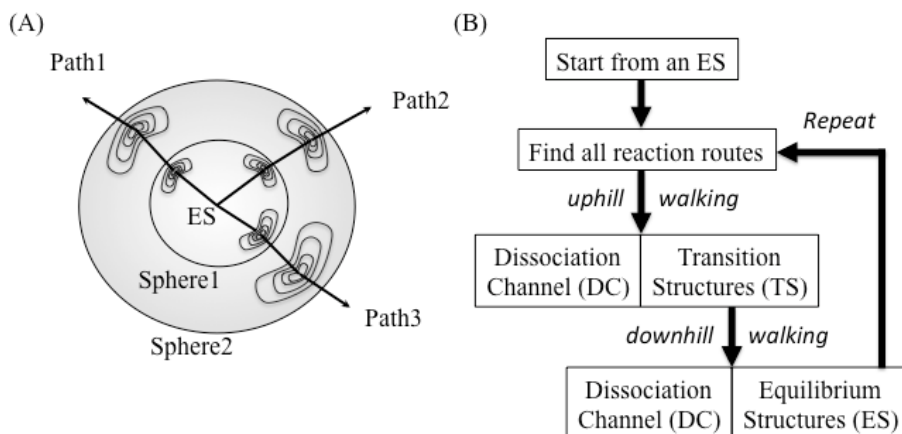


Fig. 2. Schematic illustrations of computational procedures for GRRM by the SHS method: (A) Though the reference harmonic potential has a constant energy on the scaled hypersphere surface, the real potential has some minima on the same surface, which correspond to the anharmonic downward distortions indicating the symptoms of chemical reactions. Following those minima on different sizes of scaled hypersphere (spheres 1 and 2), reaction routes (paths 1-3) can be traced from the equilibrium structure (ES) by the SHS method as shown by arrows. (B) Starting from an ES, find all reaction routes as energy minima on the scaled hypersphere (maxima of anharmonic downward distortions), then continue uphill walking to reach DC or TS, and then downhill walking from the TS region to DC or another ES. From each new ES, the above procedures for finding DC, TS, and ES should be repeated until no new ES is found. This one-after-another approach in the SHS method can be automated, and it enables us to perform GRRM within finite processes. (Reprinted with permission from ref. (Ohno & Maeda 2006). Copyright 2006 American Chemical Society.)

Both methods were implemented with the μ -ADDF technique. In this new GRRM method, a large system is divided into reaction-center and nonreaction-center parts. All the path following calculations, including minimizations on scaled hyperspheres, ES and TS optimization, and IRC following, are carried out in macroiteration steps. The positions of nonreaction-center atoms are optimized during microiteration steps after macroiteration. All movements of reaction-center atoms are treated by the GRRM method as in the case without microiterations with either exact or updated Hessian. $(\text{H}_2\text{CO})(\text{H}_2\text{O})_{100}$ and $(\text{Si}_6)(\text{C}_{21}\text{H}_{17})_6$ were examples used to test the stability and efficiency of the new method. Multiple reaction pathways were identified in both cases. Thus, the GRRM method with μ -ADDF could serve as a powerful tool to explore PESs of reactions in large molecular systems. However, the GRRM has not been reported to be applied on protein systems.

2.3 Combined method for determining reaction path, minima, and TSs

It is worth noting that Schlegel *et al.* developed a method to determine TSs, minima and the reaction path in a single procedure without calculating the Hessian matrix. (Ayala & Schlegel 1997) In this method, a starting approximate path is constructed as several (5 to 7) structures on PES, and iteratively relaxed until two endpoints reach minima, and one of the middle points reaches a TS. The final path is a second order approximation of the steepest descent

path. However, this method became impractical for large biological systems, such as proteins, because of the computational cost of eigenvector following searches for TSs. Nevertheless, this method provides a rigorous means of identifying reaction paths and TSs simultaneously.

2.4 Reduced Hessian methods

Hessian calculations are important for TS calculations, as well as reaction path following methods. However it may not be practical or necessary to calculate a full Hessian for macromolecular systems, since a large amount of degrees of freedom are unrelated to the reaction of interest. Therefore, these degrees of freedom can be somewhat disregarded. The partial Hessian vibrational analysis (PHVA),(Li & Jensen 2002) was developed to diagonalize only a subblock of the Hessian matrix to yield vibrational frequencies for partially optimized systems. In a recent development of vibrational subsystem analysis (VSA), the complexity of Hessian calculation can be reduced by separating a large system into an active “subsystem” with the remainder of the system defining the “environment”. The environment is kept at a minimum energy with respect to the motion of the subsystem, thus an effective Hessian involving only the subsystem needs to be considered.(Woodcock et al. 2008) The VSA is an improvement over PHVA, but does entail higher computational costs. These reduced Hessian approaches can only work if an adequate subsystem including all important atoms for the reaction can be identified.

By applying the methods introduced in this section, one could locate an IRC on any given PES with well-defined TSs. There are certain limitations of these methods, especially for large systems with high degrees of freedom. For large systems, the uphill walking methods should not be one’s first choice since PESs of such systems are rather complicated and rugged, with numerous local saddle points. The downhill walking methods worked very well for the listed studies. However, a TS needs to be identified before applying IRC calculations. There is always a possibility that the IRC calculated from identified TS does not reach the desired reactant or product. It is even more likely that there are multiple TSs that exist between the reactant or product. In consideration of these factors, there might be more interest to obtain information about reaction pathway rather than TSs, especially for large biomolecular systems. Accordingly, the so-called chain-of-states methods were developed to obtain a reaction pathway without identifying TSs.

3. Chain-of-states methods

In chain-of-states methods, a number of replicas (i.e. states) of a system are used to connect two endpoints, and are subject to minimization simultaneously. The first and the last replicas usually correspond to the reactant and product, and are often fixed during the minimization. For large complex systems, chain-of-states methods can be used to address issues relating to hysteresis, free energy, reaction rates, and multiple pathways. In this section, various chain-of-states methods to build reaction paths are surveyed.

3.1 Line integral methods

Elber and Karplus (EK) developed a method using a line integral representation of a discretized path subject to optimization.(Elber 1987) In their method, the objective function subject to optimization reads as

$$S^{EK}(\mathbf{R}_0, \dots, \mathbf{R}_M)_L = \frac{1}{\sqrt{\sum_{j=1}^M (\Delta_j)^2}} \sum_{j=1}^M V(\mathbf{R}_j) \Delta_j + \sum_{j=1}^M \lambda \left(\Delta_j - \sqrt{\frac{\sum_{j=1}^M (\Delta_j)^2}{M}} \right)^2 \quad (4)$$

where \mathbf{R}_j is the coordinates of replica j , M is the total number of steps from the starting replica \mathbf{R}_0 to the final replica \mathbf{R}_M , $\Delta_j = \sqrt{[(\mathbf{R}_j - \mathbf{R}_{j-1})^2]}$, and $V(\mathbf{R}_j)$ is the potential energy of the system at replica \mathbf{R}_j . The degrees of freedom of the rigid body, i.e. translation and rotation, are projected out from the minimization for replicas with reference to the end points of the path. The objective function is subject to non-linear optimization for the final reaction path. The method was applied to several systems including the conformational change of myoglobin. (Elber 1987)

Czermanski and Elber then developed the self-penalty walk (SPW) method (Czermanski & Elber 1990a) based on original EK formulation. The main development of SPW is the addition of repulsion terms for each replica j :

$$\rho \sum_{i>j+1}^M \exp \left(-\frac{\Delta_{i,j}^2}{(\lambda \Delta)^2} \right) \quad (5)$$

where $\Delta_{i,j} = |\mathbf{R}_i - \mathbf{R}_j|$, $\Delta = \sqrt{\sum_{j=1}^M \frac{(\Delta_j)^2}{M}}$, λ and ρ determine the range and maximal value of the

repulsion between replica i and j . These repulsive terms can help to prevent the aggregation of replicas in the neighborhood of two endpoints where the energies of replicas are lower than those close to the TS region. As discussed in their paper, the repulsive terms reflect the stiffness of the reaction path and mimic the effect of kinetic energy on the classical trajectory. The reaction path calculation of alanine dipeptide isomerization using SPW displayed a convergence rate that is 10 times faster than the one using the EK method. The conformational change of isobutyryl-ala₃-NH-methyl (IAN) between the helix and extended chain was also studied. By using the optimal values of parameters in this method, a reaction path that is very close to the MEP was obtained for IAN from the calculations with a straight line as initial path. In another study, the SPW method was applied to study the diffusion of carbon monoxide through leghemoglobin. (Nowak et al. 1991) Three similar but distinct diffusion pathways were identified and compared. The barrier heights calculated for the three pathways were in the agreement with the proposed model.

Ulitsky and Elber (UE) proposed a locally updated planes (LUP) method to calculate steepest descent paths (SDP) in flexible polyatomic systems. (Ulitsky & Elber 1990) For a series of replicas $\{\mathbf{r}^k\}_{k=1,M}$, \mathbf{s}^k is the unit vector along the gradient for replica \mathbf{r}^k . The SPD satisfies that $\nabla V_{proj} = \nabla V - (\nabla V \cdot \mathbf{s}^k) \mathbf{s}^k = 0$, where V is the potential energy. For a discretized path, the vector \mathbf{s}^k is approximated as $(\mathbf{r}^{k+1} - \mathbf{r}^{k-1}) / |\mathbf{r}^{k+1} - \mathbf{r}^{k-1}|$. To refine the path of each round, the coupled differential equations of all the replicas $\{(\partial/\partial t)\mathbf{r}^k(t) = \nabla V_{proj}\}_{k=1,M}$ were solved by a fifth order Adams predictor-corrector algorithm. (Gear 1971) The SDP could be reached in the limit of $t \rightarrow \infty$. Choi and Elber later improved the LUP method by a

gradient updating scheme.(Choi & Elber 1991) The local gradient vector \mathbf{s}^k is calculated based on the initial path, and updated based on new reaction path after every M steps of optimization. The authors found that $M=20$ to be efficient for the helix formation of tetrapeptides under their studies. As Choi and Elber pointed out, the final results depend on the initial guess. If multiple MEPs exist between two endpoints and the initial guess is not in the radius of convergence of a single path, the result may be discontinuous and contain segments from different MEPs.

3.2 Nudged elastic band (NEB) methods

As pointed by Jónsson, Mills and Jacobesen,(Jónsson et al. 1998) the line integral methods suffer from the “corner cutting” problems in which the final paths bypass the TS region, leading to overestimated barriers. This problem originates from the fact that elastic forces added to replicas have non-zero components perpendicular to the path. The optimization of objective functions that include elastic forces will have a tendency to pull replicas off from the MEP. In addition, replicas along the final path tend to aggregate around endpoints where the potential energies are smaller than the TS region, which is underrepresented in the chain. This is because the actual forces of the path pull the replicas downhill against elastic forces. To solve these problems, the NEB methods were developed to project out perpendicular components of elastic forces and parallel components of the true force with respect to the path under minimization.(Henkelman et al. 2000a; Jónsson et al. 1998)

For a reaction path with $N+1$ replicas $[\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_N]$, the NEB method is implemented as the following: The tangent at the replica i , $\hat{\boldsymbol{\tau}}_i$, can be estimated based on adjacent replicas $i-1$ and $i+1$:

$$\hat{\boldsymbol{\tau}}_i = \frac{\mathbf{R}_{i+1} - \mathbf{R}_{i-1}}{|\mathbf{R}_{i+1} - \mathbf{R}_{i-1}|} \quad (6)$$

The force acting on a replica subject to optimization is

$$\mathbf{F}_i^{NEB} = \mathbf{F}_i^\perp + \mathbf{F}_i^{s//} \quad (7)$$

where \mathbf{F}_i^\perp is the sum of the true force perpendicular to the tangent:

$$\mathbf{F}_i^\perp = -\nabla V(\mathbf{R}_i) + \nabla V(\mathbf{R}_i) \bullet \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \quad (8)$$

and $\mathbf{F}_i^{s//}$ is the elastic force along the tangent,

$$\mathbf{F}_i^{s//} = k(|\mathbf{R}_{i+1} - \mathbf{R}_i| - |\mathbf{R}_i - \mathbf{R}_{i-1}|) \bullet \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \quad (9)$$

V is the potential energy of the system, k is the elastic force constant. By projecting out the elastic force perpendicular to the path, the final path should relax to the MEP in principle.

One potential problem that the NEB method may encounter is producing kinks along the path, mainly in regions where the parallel component of force is large compared with the perpendicular component. A new NEB method was developed by Henkelman and Jónsson to eliminate the kinks along the path.(Henkelman & Jónsson 2000) In the new implementation, the tangent is defined as

$$\boldsymbol{\tau}_i = \begin{cases} \mathbf{R}_{i+1} - \mathbf{R}_i & \text{if } V_{i+1} > V_i > V_{i-1} \\ \mathbf{R}_i - \mathbf{R}_{i-1} & \text{if } V_{i+1} < V_i < V_{i-1} \end{cases} \quad (10)$$

where V_i is the potential energy of replica i , $V(\mathbf{R}_i)$. However, when the replica i is at a minimum ($V_{i+1} > V_i < V_{i-1}$) or at a maximum ($V_{i+1} < V_i > V_{i-1}$), the tangent is estimated as

$$\boldsymbol{\tau}_i = \begin{cases} (\mathbf{R}_{i+1} - \mathbf{R}_i)\Delta V_i^{\max} + (\mathbf{R}_i - \mathbf{R}_{i-1})\Delta V_i^{\min} & \text{if } V_{i+1} > V_i \\ (\mathbf{R}_{i+1} - \mathbf{R}_i)\Delta V_i^{\min} + (\mathbf{R}_i - \mathbf{R}_{i-1})\Delta V_i^{\max} & \text{if } V_{i+1} < V_i \end{cases} \quad (11)$$

where $\Delta V_i^{\max} = \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$, and $\Delta V_i^{\min} = \min(|V_{i+1} - V_i|, |V_{i-1} - V_i|)$. This implementation was motivated by the calculations of finding the MEP from a TS. It is always easier to find a MEP by following the PES downhill from a TS rather than going uphill from a minimum. Therefore, the local tangent of the path is always defined by the higher energy replicas nearby to improve the stability of the calculation. The updated NEB method is reported to behave well and remove kinks from the reaction paths. (Henkelman & Jónsson 2000)

Henkelman, Uberuaga and Jónsson (Henkelman et al. 2000b) developed climbing image NEB (CI-NEB) by defining the force of the replica with the highest energy i_{\max} as

$$\mathbf{F}_{i_{\max}} = -\nabla V(\mathbf{R}_{i_{\max}}) + 2\nabla V(\mathbf{R}_{i_{\max}}) \bullet \hat{\boldsymbol{\tau}}_{i_{\max}} \hat{\boldsymbol{\tau}}_{i_{\max}} \quad (12)$$

The basic idea is to invert the component of force along the path. Therefore, the minimization movement based on a given force will lead this replica toward the saddle point, which has the maximum energy along the path but minimum in all other directions. In addition, the elastic force constant along the path is scaled linearly on the energy of the replicas. Therefore, the replicas around the TS region would be connected through stronger elastic bands than those closer to the endpoints. Compared with the regular NEB method, the CI-NEB method generates reaction pathways with TS regions better represented, and leads to more accurate estimates of reaction barriers.

Maragakis *et al.* (Maragakis et al. 2002) proposed the adaptive nudged elastic band approach (ANEBA) to search for saddle points. Instead of starting with a large number of replicas, three movable replicas are added for the initial round of the NEB calculation. After obtaining reasonable convergence, the two replicas that are adjacent to the one with the highest energy will serve as endpoints for the next round of NEB calculations after inserting new movable replicas in between. This process could be repeated until a well-defined TS is identified. The major goal of ANEBA is TS searching, but a reaction path that approximates the MEP will be obtained after finding the TS.

Trygubenko and Wales proposed a doubly nudged elastic band (DNEB) method by retaining a portion of perpendicular component of elastic force. (Trygubenko & Wales 2004b; Trygubenko & Wales 2004a) The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) quasi-Newton optimizer was implemented with the DNEB method. One of the strategies adapted is gradually projecting the perpendicular component of elastic force until the full path is reasonably stable. Then replicas with potential energies above their adjacent replicas are subject to an eigenvector-following calculation for TSs.

Chu *et al.* (Chu et al. 2003) developed the first superlinear minimizer for the NEB method that was based on expanding the adopted basis Newton-Raphsion (ABNR) method (Brooks

et al. 1983), and this is available in CHARMM.(Brooks et al. 2009) Other improved optimizers,(Alfonso & Jordan 2003; Bergonzo et al. 2009; Galván & Field 2008) and other software packages, such as AMBER,(Case et al. 2005) have efficient implementations of NEB as well.

In general, the NEB methods have been applied in numerous studies, including diffusion processes on surfaces,(Ágoston & Albe 2010; Sørensen et al. 1996; Uberuaga et al. 2000; Villarba & Jonsson 1994; Yang et al. 2009) nucleation process,(Lutsko 2008) stability of nanoparticle,(Vélez et al. 2008) and dissociative adsorption of a molecule on a surface.(Mills 1995)

Xie, Liu and Yang adapted the NEB method for enzymatic reactions.(Xie et al. 2004) The major development of their method is carefully choosing the degrees of freedom that are essential for the reaction. The chosen degrees of freedom were subject for the calculation of distances between replicas. The large number of soft and floppy spectator degrees of freedoms are, therefore, excluded from the elastic force computation in NEB. The potential problem of such selection is discontinuity of contributions from spectator degrees of freedoms to the total potential energy. One solution is starting from relatively rigid reference systems and keeping position restraints to the environmental atoms throughout the NEB calculations. The third alteration is cutting off the elastic band for intermediate states from multiple reaction steps to allow them to relax to the minimum. They applied this modified NEB method to study the mechanism of enzyme class A β -lactamase, and obtained the MEP of the reaction at the active site. With further improvement,(Cisneros et al. 2005; Liu et al. 2004) the NEB methods were applied in more enzyme systems to map out the detailed MEP of their mechanisms.(Cisneros et al. 2009; Zhao & Liu 2008)

3.3 Zero temperature string (ZTS) methods

Similar to NEB, the ZTS method was developed to find the reaction pathway connecting two minima on a PES.(Cameron et al. 2010; E et al. 2002, 2007; Ren 2003) In the ZTS method, a series of evenly distributed initial states between two endpoints are minimized towards a MEP using path gradients and tangent information. Instead of adding spring forces on the path to maintain appropriate distances between states, a step is added to evenly redistribute the states along the path after each step of minimization. The basics of the ZTS method can be summarized as the following. For transitions between two metastable states A and B , the MEP is a smooth curve φ^* between A and B which satisfies

$$(\nabla V)^\perp(\varphi^*) = 0 \quad (13)$$

where $(\nabla V)^\perp$ is the component of ∇V perpendicular to φ^* . From an arbitrary string φ connecting A and B , searching MEP can be realized by evolving φ through

$$u^\perp = -(\nabla V)^\perp(\varphi) \quad (14)$$

where u^\perp is the perpendicular component of force with reference to φ . For simplicity, φ is parametrized by normalized arc length a , for which $a=0$ at A and $a=1$ at B . The Eq. (13) can be rewritten as

$$\varphi_\tau = -(\nabla V)^\perp + r\hat{\tau} \quad (15)$$

where

$$(\nabla V)^\perp = \nabla V - (\nabla V, \hat{\tau}) \hat{\tau} \quad (16)$$

$\hat{\tau}$ is the unit tangent factor of φ and scalar field $r(\alpha, t)$ is a Lagrange multiplier determined by parameterization. In real applications, the string is discretized into a series of replicas, which move according to the first term of the right hand side of Eq. (15). Eq. (15) can be solved using ordinary differential equation (ODE) solvers to evolve the string towards a MEP. The reparameterization can be carried out after certain number of evolving steps of the string to redistribute the replicas to enforce the equal arc length (distance) between adjacent replicas. In a further development, (E et al. 2007) the Eq. (15) is rewritten as

$$\varphi_t = -(\nabla V) + \bar{r} \hat{\tau} \quad (17)$$

where $\bar{r} = r + (\nabla V, \hat{\tau})$. Eq. (17) is equivalent to Eq. (15), but avoids a force projection operation in the string evolving step for better numerical stability. It is noted that the ZTS method does not require the location of minima beforehand to generate an initial string. The final converged string should connect two minima, as long as the two endpoints of the initial string lie in the attraction basins of the minima (Fig. 3). Based on the framework of the ZTS method, the transition path theory was developed to sample the minimum free energy path. This theory will be introduced in section 4.3.

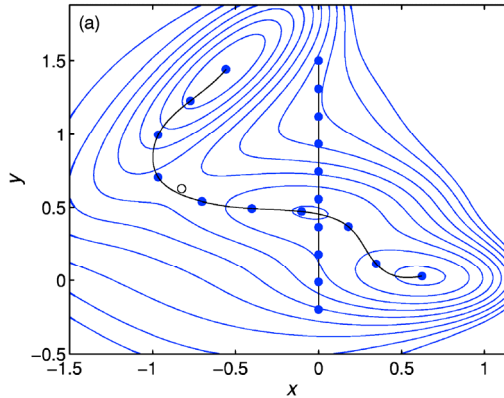


Fig. 3. Initial string and calculated MEP using the string method with ten images. The empty circle indicates the saddle point identified by combining the string method with the climbing image technique. (Reprinted with permission from ref. (E et al. 2007). Copyright 2007, American Institute of Physics.)

3.4 Growing string methods

So far, the chain-of-states algorithms introduced in section 3.1–3.3 require an initial path composed by series of replicas. The initial paths are usually constructed by linear interpolation between two minima. There are two potential drawbacks of these methods when applying them in studies involving expensive quantum mechanics (QM) calculations. First, the expensive QM calculation needs to be carried out for every replica in each round of

optimization from the very beginning of the reaction path calculation until convergence. Second, the initial linear interpolation path may include structures with severely overlapping atoms, which can cause failure of the QM calculations. The first drawback means intensive computation of reaction path optimization. The second drawback simply means immediate failure of the chain-of-states calculation. Growing string methods (GSM) were proposed (Peters et al. 2004) and are under continuous development to address these issues.(Goodrow et al. 2008, 2009, 2010; Quapp 2005, 2009)

The basic idea is to gradually grow two strings separately from two minima by continuously adding replicas to each string until both strings merge. The GSM generally comprises five steps.(Goodrow et al. 2008) In step one, two replicas are added along the linear synchronous transit path, and placed close to reactant and product, which defines the linear path. The reactant and product plus each added replica close to them serve as starting segments of two separate strings. In step two, the two string segments are minimized until the norm of the orthogonal force of each replica converges to a specified tolerance. After step two, the two string segments lie on the steepest descent direction from the frontier replica of each segment. In step three, replicas on both string segments are redistributed uniformly in terms of arc length, similar to the ZTS. The arc length of the string can be calculated by integrating the cubic spline fitted to all the replicas. The rigid body movement needs to be projected out for all the replicas with reference to the reactant and product after each round of the minimization. In step four, a new replica is added to each string segment along the fitted cubic spline. Steps two through four are repeated until the arc length between two frontier replicas from the string segments is small enough so that the two string segments can be considered a united string. In step five, the joined string is minimized and reparametrized with a fixed number of replicas until the sum of the norm of the orthogonal forces on all nodes falls under predefined tolerance. Step five is analogous to the ZTS method. One or more replicas with maximum energies among their neighbors can be used for TS search calculations. The general GSM is illustrated in Fig. 4.

In various developments, Newton projector,(Quapp 2005) internal coordinate,(Goodrow et al. 2008) conjugate gradient method,(Goodrow et al. 2008) and different TS search strategies (Goodrow et al. 2009, 2010) are implemented in GSM for better performance.

The major goal of GSM is improving the calculation efficiency of the ZTS framework for systems involving expensive QM calculations. By growing strings stepwise from two endpoints instead of minimizing from an initial path with potentially bad geometries, a large amount of calculations can be saved from carrying out QM calculations to move the fixed number of replicas from initial path to MEP iteratively. The GSM has been applied in one study using a model system representation of an enzymatic system (Maresh et al. 2008) and some other studies using small models.(Goodrow & Bell 2008; Zheng & Bell 2008a, b) The application of GSM directly on protein system has yet to be reported.

3.5 Conjugate peak refinement (CPR) method

The CPR method (Fischer 1992; Noé et al. 2006) focuses on finding saddle points along a path that connects a predefined reactant (\mathbf{r}) and product (\mathbf{p}). At the beginning of the CPR calculation, the replica with maximum potential energy, \mathbf{y}_1 , is searched along an initial path, e.g. linear interpolation between \mathbf{r} and \mathbf{p} (Fig. 5). Then, the energy is minimized along the direction of each conjugate vector to reach the next replica, \mathbf{x}_1 . This replica is added to the path, which is represented as $(\mathbf{r}, \mathbf{x}_1, \mathbf{p})$. The procedure is repeated for the partial path $(\mathbf{r}, \mathbf{x}_1)$ and $(\mathbf{x}_1, \mathbf{p})$. If there is only one TS that exists between \mathbf{r} and \mathbf{p} , the maximum may not

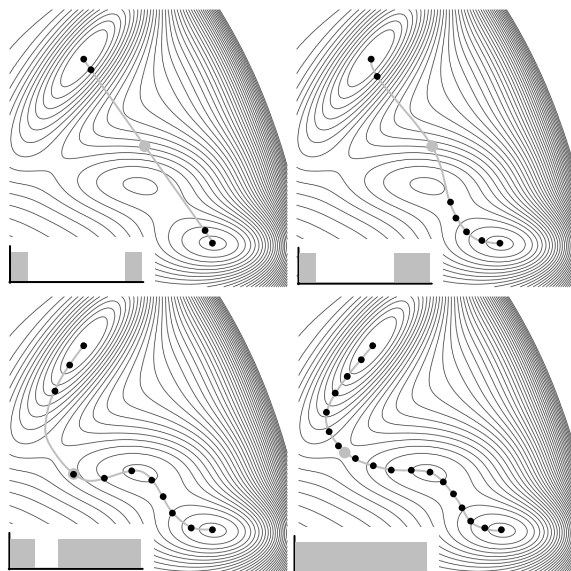


Fig. 4. Four snapshots of a growing string on the Muller-Brown potential energy surface. (Reprinted with permission from ref. (Peters et al. 2004). Copyright 2004, American Institute of Physics.)

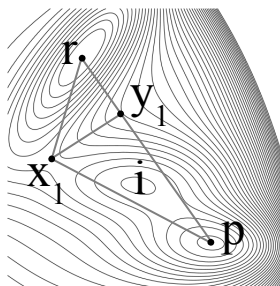


Fig. 5. Illustration of the conjugate peak refinement (CPR) method. The reactant, product and intermediate wells are indicated with r , p and i . (Reprinted from ref. (Fischer 2004), Copyright 2004, with permission from Elsevier.)

exist for one path segment. Therefore the procedure will not be carried out for this specific path segment. The procedure is carried out recursively on each new path segment defined by newly added replicas, until the root mean square (RMS) of the gradient falls under a predefined tolerance for the saddle points by refinement along conjugate vectors. In practice, the linear interpolation often produces very poor structures with severe overlapping atoms. This could result in extremely high energies and gradients. Therefore, certain limitations need to be imposed on atomic movements during refinement. The CPR method can possibly identify one or more TSs between r and p through recursive searches and refinement. However, a smooth path will be not a direct result from the CPR procedure. From identified TSs, steepest descent calculations can be carried out to generate a final MEP

between \mathbf{r} and \mathbf{p} . Implicitly, it is assumed that TSs identified through the CPR method lie on the same MEP. Otherwise, the steepest descent calculations from TSs do not necessarily generate a continuous path between \mathbf{r} and \mathbf{p} . The CPR method has been applied to study the ligand-binding pathways in myoglobin,(Golden & Olsen 2008) several small molecule systems (Fischer et al. 1994; Fischer et al. 1995; Verma et al. 1996) as well as large molecules, such as proteins.(Blondel et al. 1999; Caflisch et al. 1997; Dutzler et al. 2002; Fischer et al. 1993; Gruia et al. 2005; Santos et al. 2000)

3.6 Reference path methods

In some applications, especially for macromolecules, the convergence to the MEP using chain-of-states methods could require a large number of iterations. Reference path methods were developed to generate reference reaction paths that are a good approximation to true MEPs with fast convergence rates and evenly distributed replicas along the path. Such reference paths can be used as an expansion to calculate free energies, reaction rates, or kinetic isotope effects.

3.6.1 Hyperplane projection methods

Czermanski and Elber (Czermanski & Elber 1989) proposed applying holonomic constraints on replica \mathbf{R} along a reaction path connecting replicas \mathbf{R}_1 and \mathbf{R}_2 :

$$(\mathbf{R} - \mathbf{R}_\alpha) \cdot (\mathbf{R}_1 - \mathbf{R}_2) = 0 \quad (18)$$

where $\mathbf{R}_\alpha = [(1 - \alpha)\mathbf{R}_1 + \alpha\mathbf{R}_2]$, and α is a parameter that varies from 0 (reactant) to 1 (product) with small steps. For each α value, the potential energy of the system $V(\mathbf{R}_\alpha)$ is minimized while keeping the system in the hyperplane defined by Eq. (18). It is necessary to project out the rigid body motions from the optimization using either \mathbf{R}_1 or \mathbf{R}_2 as reference structures. The optimized path satisfies the condition that at any point along the path, only one direction may have negative energy curvature, because energy is minimized for all degrees of freedom except the direction perpendicular to the hyperplane defined by Eq. (18). The Powell conjugate gradient algorithm was applied to optimize the reaction path with linear constraints in this method. The conformational change of IAN between the helix and extended chain was studied using this method. The lowest energy path of the transition identified by the calculation was presented. In addition, the 257 direct transitions between 112 minima of IAN were also identified and subject to statistical analysis. The same authors later introduced another constraint to the path using parameter η :

$$\frac{|\mathbf{R} - \mathbf{R}_1|}{|\mathbf{R} - \mathbf{R}_1| + |\mathbf{R} - \mathbf{R}_2|} - \eta = 0 \quad (19)$$

where η is the parameter between 0 (reactant) and 1 (product).(Czermanski & Elber 1990b) Combining application of α and η constraints, more comprehensive results were presented for conformational changes of IAN.

3.6.2 RPATH/restraint method

Woodcock *et al.* implemented the reaction path (RPATH) method with restraints of both best-fit root mean square distances (RMSD) and angles on replicas along the pathway. (Woodcock et al. 2003) The RMSD restraint forces are defined as

$$E_{rms} = \sum_{i=1}^N \frac{1}{2} k_{rms} (r_i - \bar{r})^2 \quad (20)$$

where N is the number of replicas, k_{rms} is the force constant used to restrain distance between adjacent replicas along the reaction pathway, r_i is the best-fit RMSD between replica i and $i+1$, and \bar{r} is the average distance between adjacent replicas. The angle energetic penalty term reads

$$E_{angle} = \sum_{i=1}^N \frac{1}{2} k_{ang} (\text{COSMAX} - \cos(\theta)_i)^2 \quad \text{COSMAX} > \cos(\theta)_i, \quad (21)$$

$$E_{angle} = 0 \quad \text{COSMAX} \leq \cos(\theta)_i$$

The angle θ , illustrated in Fig. 6, defines the deviation of the pathway from linearity. The force constant k_{ang} controls the rigidity of the pathway. Constant COSMAX determines the value of $\cos(\theta)$ subject to the angle forces.

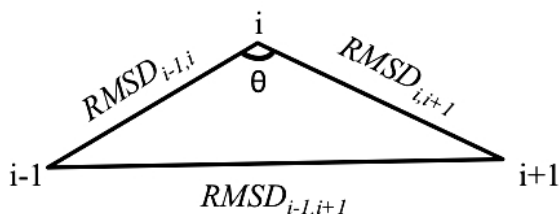


Fig. 6. Definition of angle θ for replica i in RPATH calculation. $\text{RMSD}_{i-1,i}$ is the distance between replica $i-1$ and i . It is similar to $\text{RMSD}_{i,i+1}$ and $\text{RMSD}_{i-1,i+1}$.

The RPATH/restraint method was applied to study the mechanism of chorismate mutase. (Woodcock et al. 2003) All protein residues or water molecules that had any atom within 6 \AA of the substrate were replicated 21 times to build the reaction path. All other atoms within the system served as the environment of all the replicas. Starting from a linear interpolation pathway, the optimization of the RPATH/restraint calculation was converged within 400 steps. The reaction barrier obtained from RPATH/restraint calculations was also in good agreement with experiments.

3.6.3 RPATH/constraint method

Recently, Chu and coworkers (Brokaw et al. 2009) applied an equal distance holonomic constraint in the RPATH framework. Given two states of a molecular system with N atoms, r^0 and r^k , a chain of $K+1$ replicas can be constructed to connect these two states. The distance between each pair of adjacent replicas is set to be equal to each other:

$$\Delta l^0 = \dots = \Delta l^i = \dots = \Delta l^{K-1} = \bar{\Delta l} \quad (22)$$

Here, Δl^i is the distance between replica i and $i+1$. $\bar{\Delta l}$ is the average distance between adjacent replicas. The distance (Δl) can be in any form, including best-fit RMS distance. To apply the equal distance constraint defined in Eq. (22), the following scheme is used to propagate the reaction path with two ends, r^0 and r^k , fixed at the initial coordinates.

- i. Set up and calculate initial average distance, $\overline{\Delta l}$, for replicas $r^{0(0)}$ through $r^{k(0)}$. The superscript "(0)" indicates the optimization iteration step.
- ii. To maintain the equal distance, a set of K coefficients, $(\lambda^i)^{(n)}(i=0, K-1)$, are used to update the coordinates of each replica i :

$$\left(\mathbf{r}^i\right)^{(n+1)} = \left(\mathbf{r}^i\right)^{(n)} + \left(\lambda^{i-1}\right)^{(n)} \left(\frac{\partial \Delta l^{i-1}}{\partial \mathbf{r}^i}\right)^{(n)} + \left(\lambda^i\right)^{(n)} \left(\frac{\partial \Delta l^i}{\partial \mathbf{r}^i}\right)^{(n)} \quad (23)$$

- iii. Solve $(\lambda^i)^{(n)}$ by setting the first order Taylor expansion of each of $((\lambda^i)^{(n+1)} - \overline{\Delta l})(i=0, K-1)$ with respect to $(\lambda^i)^{(n)}$ to zero:

$$-\left(\Delta l^i\right)^{(n)} - \overline{\Delta l} = \sum_{j=i-1}^{i+1} \left(\frac{\partial \Delta l^i}{\partial \lambda^j}\right)_{\lambda^j} \left(\lambda^j\right)^{(n)} \quad (24)$$

- iv. If any of the values of $|\Delta l^i - \overline{\Delta l}| (i=0, K-1)$ is greater than a selected tolerance, then repeat steps (ii) and (iii).

After convergence, the RPATH calculation leads to a reaction path composed by K+1 equal distance replicas connecting states r^0 and r^k .

In the framework of using constraints with RPATH, a kinetic energy potential can be added to the potential energy with overall objective functions as a Hamiltonian for path optimization. Therefore the optimized path is the so-called minimum Hamiltonian path (MHP) instead of MEP. The kinetic energy component in the potential prevents kinks, therefore helping maintain the smoothness of the path. This smoothness comes at the cost of deviating from the MEP, resulting in higher reaction barriers, but with the benefit as a better starting point for free energy studies.

The RPATH/constraint was applied to study the helix-to-sheet transition of a GNNQQNY heptapeptide. (Brokaw et al. 2009) An initial reaction path with 100 replicas was generated. Then a stable intermediate was chosen to divide the pathway into two segments in which the number of replicas was doubled. Using this divide-and-conquer strategy, a smooth reaction pathway with 464 replicas was obtained to describe the transition (Fig. 7).

4. Free energy sampling of reaction paths for large systems

The methods introduced in Section 3 provide a framework for determining a minimum potential energy pathway connecting two equilibrium states. For biological processes the minimum *free* energy pathway (MFEP) is often more desirable to compute, ensuring that the results obtained for the PES are converged, in a thermodynamic sense. Experimental observables can be transformed more readily to free energies of the reactant and product states. (Hu et al. 2008) Biased sampling methods, such as umbrella sampling (Bartels & Karplus 1997; Kästner 2011; Torrie & Valleau 1977; Torrie & Valleau 1974) and metadynamics (Bussi et al. 2006; Laio 2002; Laio & Gervasio 2008; Raiteri et al. 2006) have been widely applied to sample free energies of the transition path for large systems. Due to their high computational cost, the applications of these methods have been limited in studies of enzymatic reaction mechanisms using QM/MM methods. However, powerful modern computing systems and advanced methodologies make the dynamical simulation of enzymatic reactions for free energy profiles much more feasible than a decade ago. In this

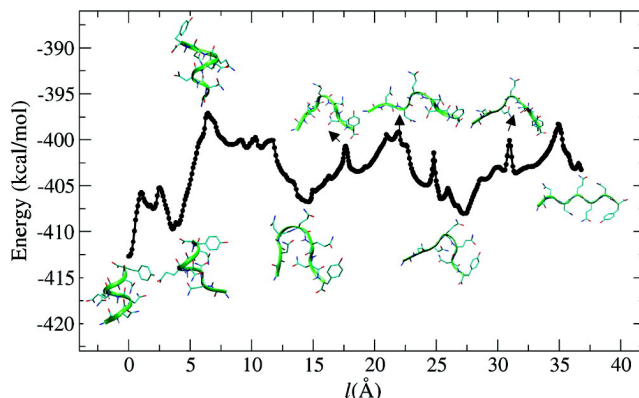


Fig. 7. An optimized path of the helix-to-sheet transition of a solvated GNNQQNY heptapeptide RPATH/constraint. The path contains a total of 464 replicas. Structures of selected replicas are shown to illustrate the nature of the transition. (Reprinted with permission from ref. (Brokaw et al. 2009). Copyright 2009 American Chemical Society.)

section, methods for direct simulation of protein systems using QM/MM methods are surveyed.

4.1 QM/MM free energy path (FE) and minimum free energy (MFEP) path methods

QM/MM-FE methods have been pioneered by Yang and coworkers (Zhang et al. 2000) and are sequential based approaches for first optimizing an equilibrium geometry on a reaction potential energy surface, followed by the calculation of free energies between two states, or along a path. (Hu & Yang 2008) The first innovation of their approach involves separating the QM and MM parts of the system and optimizing them independently to allow for faster convergence. There is no concurrent optimization of the QM and MM energy functions making it effective in reducing the number of expensive QM energy and gradient evaluations. Starting from a given structure, the coordinates of the MM portion are frozen while the QM coordinates are optimized. The MM portion is subsequently minimized while the QM portion is held fixed and the process is repeated until convergence is reached for each partition. Further simplification is made by using an approximate QM/MM energy function for the MM minimization where only Coulombic interactions between the point charges of the MM atoms and the QM electrostatic potential (ESP) fitted charges are accounted for:

$$\left\langle \Psi \left| -\sum_1^{N_{\text{off}}} \sum_{\beta \in \text{MM}} \frac{q_{\beta}}{r_{\beta i}} + \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Z_{\alpha}^* q_{\beta}}{r_{\alpha \beta}} \right| \Psi \right\rangle = \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Q_{\alpha}^* q_{\beta}}{r_{\alpha \beta}} \quad (25)$$

In order for the approximation to yield a monotonic convergence of the QM/MM energy, the MM minimization must be able to approximate the PES closely. It is important to fully optimize the reactant and product structure at the same level of theory applied on QM/MM-FE calculations.

From the converged PES obtained through the sequential optimizations, a free energy of perturbation is performed:

$$\begin{aligned} \Delta F_{QM/MM}^{(A \rightarrow B)} &= F_{QM/MM}(R_c^B) - F_{QM/MM}(R_c^A) \\ &= -\frac{1}{\beta} \ln \left\langle \exp \left\{ -\beta \left[E_{QM/MM} \left(r_{QM}^{\min} \left(R_c^B \right) \right) - E_{QM/MM} \left(r_{QM}^{\min} \left(R_c^A \right) \right) \right] \right\} \right\rangle_{MM,A} \end{aligned} \quad (26)$$

R_c^A and R_c^B are different points along the reaction coordinate and $\langle \dots \rangle_{MM,A}$ represents an ensemble average over the MM sub-system with the QM region frozen to the optimized coordinates. (Zhang et al. 2000) A major advantage of the QM/MM-FE method over previous QM-FE approaches (Chandrasekhar & Jorgensen 1985; Chandrasekhar et al. 1985; Jorgensen 1989) is the inclusion of the enzymatic environment in the QM optimization. Yang and coworkers (Hu & Yang 2008) have extended the QM/MM-FE method to overcome the challenge of having the optimized path be influenced by the initial starting conformation. This influence is problematic when solvent effects may be prominent in the reaction system, e.g. for solvent exposed active sites. Instead of using MM minimization in the sequential optimization scheme, MD sampling is used for obtaining the MM subsystem minima. Optimization of the QM region is then performed with a fixed MM conformational ensemble. Another key distinction of this so-called QM/MM-MFEP over the original QM/MM-FE method is that the optimization function for the QM region is the QM potential of mean force

$$A^{(n)}(\mathbf{r}_{QM}) = A_{ref} - \frac{1}{\beta} \ln \left\{ \frac{1}{N} \sum_{\tau=1}^N \exp\{-\beta \Delta E\} \right\} \quad (27)$$

instead of the total energy surface. The associated gradient with respect to the i -th QM coordinate

$$\frac{\partial A^{(n)}(\mathbf{r}_{QM})}{\partial \mathbf{r}_{QM,i}} = \frac{\sum_{\tau=1}^N \partial \tilde{E}(\mathbf{r}_{QM}, \mathbf{r}_{MM}^{(n)}(\tau)) / \partial \mathbf{r}_{QM,i} \exp\{-\beta \Delta E\}}{\sum_{\tau=1}^N \exp\{-\beta \Delta E\}} \quad (28)$$

provides a convenient way to compute the free energy as an ensemble average over the QM atoms. The efficiency of the method derives from having a fixed size ensemble of MM conformations instead of repetitive sampling at each step. This is achieved through the updated QM reference structures of each iteration cycle. A precise potential of mean force (PMF) is obtained through optimization with classical numerical schemes. Convergence is equally quick, often achieved within 10 steps.

4.2 Transition path sampling (TPS) methods

First pioneered by Pratt (Pratt 1986) and developed much further by Chandler and his collaborators, (Bolhuis et al. 2002; Bolhuis et al. 2000; Dellago & Bolhuis 2004; Dellago & Bolhuis 2009; Dellago et al. 2002) TPS methods are a set of computational techniques to determine the transition process between two metastable states of complex systems. TPS does not require a reaction path *a priori*, because no specific or "typical" reaction pathway is pursued in TPS methods at all. Instead, a large number of transitions between two metastable states (say A and B in Fig. 8) need to be sampled to represent a transition path

ensemble (TPE) between these states. The TPE comprises all the dynamical trajectories that start from one state and end up in the other, which are called reactive trajectories. It is impractical to sample the TPE by starting a dynamics simulation from either A or B in many trials and hoping that enough number of trials can overcome the barrier and reach the other state. This is simply because the transition between A and B is a rare event, and can only be observed in an extremely limited number of trajectories if at all. Instead of this direct sampling strategy, the importance sampling methods, such as generalized MC procedures, were applied to enhance the sampling efficiency and generate a collection of reactive trajectories with a frequency proportional to their probability weight in the TPE.

Starting from a reactive trajectory, new trajectories can be generated through so-called *shooting moves* (Fig. 8). From initial trajectory, a random frame is selected as the *shooting point*. Then, the velocity of each particle in the shooting point is perturbed. From the shooting point with perturbed velocity, the equations of motion are integrated forward and backward to obtain a complete trajectory. In a simplified formalism, the new trajectory is accepted based on the acceptance probability:

$$P_{acc}[x^{(o)}(T) \rightarrow x^{(n)}(T)] = h_A[x_0^{(n)}]h_B[x_T^{(n)}] \quad (29)$$

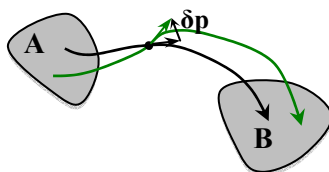


Fig. 8. In the shooting algorithm for deterministic dynamics a new path (green) is generated from an old one (black) by first randomly selecting one point on the old path, the *shooting point*. Then, the particle momenta at that point are modified by addition of a small perturbation δp . From the point with perturbed momenta the equations of motion are integrated forward and backward to obtain a complete trajectory. For small perturbations, the new trajectory will be close to the old one near the shooting point but will then rapidly diverge from it due to the chaoticity of the underlying dynamics. (With kind permission from Springer Science+Business Media: Fig. 8 in ref. (Dellago & Bolhuis 2009).)

Function $h_A[x_0^{(n)}]$ indicates that whether the system resides in state A at time 0 in the new trajectory: $h_A[x_0^{(n)}]$ is 1 for $x_0 \in A$, and 0 otherwise. Function $h_B[x_T^{(n)}]$ is defined analogously to indicate whether the system resides in state B at time T in the new trajectory. According to this expression, any new trajectory that is reactive will be accepted. The perturbation added to the shooting point can be adjusted to control the acceptance probability and optimize the TPE sampling efficiency. (Dellago et al. 1999) To begin path sampling, an initial reactive trajectory is required as the starting point. In most studies of rare events, this may not be achieved by simply running a long molecular dynamics trajectory starting from either metastable state. In practice, the system can be driven from one state to the other artificially. Such a trajectory may not carry large statistical weight in the TPE, but will be sufficient as a starting point for the sampling procedure.

A key factor for a successful TPS calculation is the sampling efficiency in the path space. For a simple system with only one MEP and one TS in between, the path sampling could

converge quickly. However, the PES of a complex system may have multiple reaction pathways, with one or more stable intermediate states presenting on each pathway. To enhance the sampling efficiency for such a complex system, the TPS with path replica exchange (Bolhuis 2008) and multi states transition TPS (Rogal & Bolhuis 2008) were developed. Based on the TPS framework, a transition interface sampling (TIS) method (Moroni et al. 2004; Van Erp et al. 2003; Vanerp & Bolhuis 2005) was developed to measure the positive flux through a series of hypersurfaces in phase space for the calculation of reaction rate constants. The TPS and TIS methods were applied to simulate the β -hairpin folding, (Bolhuis 2003) base-pair binding of DNA, (Hagan 2003) Trp-cage folding in explicit solvent, (Juraszek & Bolhuis 2006) and the reaction mechanism of lactate dehydrogenase. (Quaytman & Schwartz 2007)

4.3 Transition path theory (TPT)

Most of the reaction path theories are based on the TST, (Eyring 1935; Wigner 1938; Yamamoto 1960) with major emphasis on identifying a first order saddle point on the PES as the TS. The TS concept is also the foundation for the methods introduced in section 2 and 3. The TPS method bypasses the TS identification by sampling the reactive trajectories directly. (Bolhuis et al. 2002) Vanden-Eijnden and his coworkers advanced further by proposing TPT. (E et al. 2005b; E et al. 2005a; E & Vanden-Eijnden 2006; Maragliano et al. 2006; Metzner et al. 2006; Ren et al. 2005; Vanden-Eijnden ; Vanden-Eijnden & Venturoli 2009) The TPT is a framework to study the transition trajectories between two metastable states and the probability density functions of these reactive trajectories. In the TPT framework, the reactive trajectories are viewed as a portion of a long trajectory that oscillates between the metastable states A and B (Fig. 9). Two important quantities in TPT are probability density function and probability current function, which can be defined for a hypothetical long trajectory of an overdamped system with friction γ_i and white noise η_i on component x_i , as the following. For a given position x (neither in A nor in B) and t , the function $q(x)$ defines the probability that the trajectory reaches first B rather than A . The probability density to observe a reactive trajectory at point x and time t is

$$Z^{-1}e^{-\beta V(x)}q(x)(1-q(x)) \quad (30)$$

where Z is the partition function of the whole trajectory, $V(x)$ is the potential function of the system at x . Another quantity of interest is the probability flux of reactive trajectories across an isoprobability surface, or isocommittor surface, dividing A and B through x . The flux is the vector field with i th component reads

$$J_{AB,i}(x) = Z^{-1}e^{-\beta V(x)}k_B T \gamma_i^{-1} \frac{\partial q(x)}{\partial x_i} \quad (31)$$

where k_B is the Boltzmann factor, T is the temperature. It should be noted that for a system in thermodynamic equilibrium, the net flux through any surface is zero. The overall reaction rate from A to B can be calculated by integrating the reactive flux through an isocommittor surface S dividing A and B :

$$v_R = \int_S \hat{\mathbf{n}}_S(z) \cdot J_R(z) d\sigma_S(z) \quad (32)$$

where $\hat{\mathbf{n}}_S(z)$ is the unit vector perpendicular to S toward B . The TPT is a general theory about the rare transition events between metastable states based on reactive trajectories

but in configurational space. Similar to the TPS, the concept of reactive trajectories does not require the identification of the TS, which is not well defined for many complex systems.

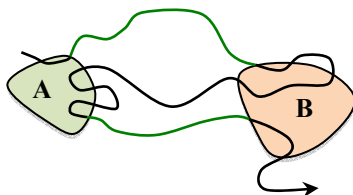


Fig. 9. Schematic illustration of a long trajectory oscillating between the reactant state A and the product state B. The reactive pieces of this trajectory, during which the system travels from A to B, are shown in green. (Reprinted with permission from ref. (E & Vanden-Eijnden 2010). Copyright 2010 Annual Reviews.)

The method to apply TPT in practice is finite temperature string (FTS) method,(E et al. 2005b) which is a finite temperature generalization of the ZTS method introduced in section 3.3. In the FTS method, the concept of string ensemble $\varphi^\omega(a)$ is introduced to have a mean value as the string $\varphi(a)$, where a is a parameter between $[0,1]$. The string ensemble can evolve by the equation

$$\dot{\varphi}_t^\omega = -\left(\nabla V(\varphi^\omega)\right)^\perp + \left(\eta^\omega\right)^\perp + r\hat{\tau} \quad (33)$$

where $\hat{\tau}$ is the unit tangent vector along φ , η^ω is a white noise with covariance

$$\left\langle \eta_j^\omega(\alpha, t) \eta_k^\omega(\alpha', 0) \right\rangle = \begin{cases} 2k_B T \delta_{jk} \delta(t) & \text{if } \alpha = \alpha' \\ 0 & \text{otherwise} \end{cases} \quad (34)$$

Using isocommitter surface, a transition tube can be defined as certain region around the most probable reactive trajectory through isocommitter surfaces between A and B. These regions have a significant probability to be visited by reactive trajectories. Of course, there could exist more than one transition tube between two metastable states. These may correspond to the multiple reaction pathways or mechanisms observed on the PES.

The free energy associated with isocommitter surface $S(a)$ is given as

$$F(\alpha) = -k_B T \ln \int_{S(\alpha)} e^{-\beta V(q)} d\sigma \quad (35)$$

In this presentation, these hyperplanes serve as reaction coordinates of the transition between metastable states. The free energies are a minimum at A and B. There may exist one or more maxima between the two metastable states.

The FTS method has been applied in a couple of biomolecular systems very recently. In one study,(Rosta et al. 2011) the mechanism of cleavage of the RNA backbone catalyzed by ribonuclease H is elucidated through QM/MM simulations. The converged strings for the entire reaction are reported with estimated free energies along the path (Fig. 10). In another study,(Ovchinnikov et al. 2011) the conformational change between the prepowerstroke and rigor states of myosin VI was simulated using a replica exchange umbrella sampling

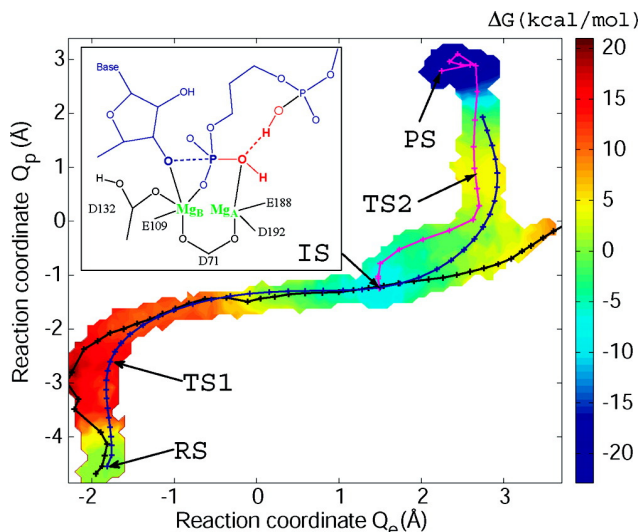


Fig. 10. 2D free energy surface and reaction strings for the complete catalytic reaction with a protonated Asp132, obtained by projection onto coordinates $Q_e = \text{ET}$ probing bond formation and breaking and $Q_p = (\text{PT1a} + \text{PT1b})/2 + \text{PT2}$ probing PT steps. The black curve shows the initial string; the blue and purple curves show the converged strings for the entire reaction and for the final step, respectively. The inset shows a schematic of the active-site coordination in the intermediate state. (Reprinted with permission from ref. (Rosta et al. 2011). Copyright 2011 American Chemical Society.)

algorithm in a string method framework. The free energy and reaction rate of the transition were calculated and found to be consistent with the experimental observation. (De La Cruz et al. 2001)

The harmonic Fourier bead (HFB) approximation proposed by Khavrutskii and his coworkers (Khavrutskii et al. 2006; Khavrutskii et al. 2008a; Khavrutskii & Mccammon 2007) possesses many similarities with FTS method. The main difference is the representation and parametrization of the path by

$$q_i(\alpha) = q_i(0) + (q_i(1) - q_i(0))\alpha + \sum_{n=1}^P b_n^i \sin(n\pi\alpha) \quad (36)$$

where α is single progress parameter between 0 and 1, q_i is the i th bead, which is the i th component of the configuration vector $\mathbf{R} = (q_1, \dots, q_{3N})$, b_n^i are the amplitudes of the Fourier basis functions. The string or reaction path in HFB can evolve under certain constraints or reparametrization, similar to ZTS and FTS, to obtain MEP and MFEP. The HFB method was applied to study the conformational transition of the signature peptide of the KcsA ion selectivity filter. (Khavrutskii et al. 2008b) A novel hypothesis of the ion selectivity mechanism was proposed based on the HFB simulation results.

4.4 Action based methods

There is a group of methods that are based on minimization of action functional of transition paths. According to the Onsager-Machlup (Onsager & Machlup 1953) and Freidlin-Wentzell

theories,(E et al. 2004; Freidlin 1984) the most probable path for the random dynamical system is the *minimum action path* (MAP). The MAP is a more general definition than MEP, for MEP is undefined for nongradient systems. For gradient systems with energy landscapes, the MAP degenerates with MEP in the high friction limit. Therefore, MEP can be obtained through optimization of path action. Olender and Elber (Olender 1997) first proposed to minimize a functional, which is called “scalar work” for a trajectory:

$$W_{scalar} = \int |\nabla V| |dl| \quad (37)$$

where V is the potential energy of the system, dl is a segment of the trajectory. For the steepest descent path, W_{scalar} has a minimum value, where ∇V is parallel to dl along the path. More recently, the geometric minimum action method (Vanden-Eijnden & Heymann 2008) and adaptive minimum action method (Zhou et al. 2008) were developed as two action based methods for MEP optimization.

Some action-based methods are developed to simulate the reaction path. Doniach and his co-workers (Eastman et al. 2001) proposed reaction path annealing methods to simulate protein folding trajectories which are distributed according to the Onsager-Machlup action functional:

$$P[\mathbf{X}(t)] \propto \exp\left[\frac{-S[\mathbf{X}(t)]}{k_B T}\right] \quad (38)$$

where the action $S[\mathbf{X}(t)]$ can be calculated for an overdamped system with friction γ in discretization form as

$$S[\mathbf{X}(t)] = \frac{1}{2\gamma} \sum_i \left(\gamma \mathbf{M} \left(\frac{\mathbf{X}_i - \mathbf{X}_{i-1}}{\Delta t} \right) + \nabla V(\mathbf{X}_i) \right)^2 \cdot \Delta t \quad (39)$$

\mathbf{M} is the mass matrix, V is the potential energy function, and Δt is the time step of the trajectories. The reaction path annealing method was employed to simulate the conformational changes among the α -helix, β -sheet and the random coil of a seven residue peptide from prion-like protein. (Lipfert et al. 2005)

Faccioli and co-workers developed the dominant reaction path (DRP) method, and applied it on numerous systems. (Autieri et al. 2009; Faccioli 2008; Faccioli et al. 2006; Mazzola et al. 2011; Sega et al. 2007) Their method is based on the Fokker-Planck equation, which defines the probability of finding a particle at position x and time t :

$$\frac{\partial}{\partial t} P(\mathbf{X}, t) = D \frac{\partial}{\partial \mathbf{X}} \left(\frac{1}{k_B T} \frac{\partial U(\mathbf{X})}{\partial \mathbf{X}} P(\mathbf{X}, t) \right) + D \frac{\partial^2}{\partial \mathbf{X}^2} P(\mathbf{X}, t) \quad (40)$$

where D is the diffusion constant, U is the potential energy function for the particle. The Boltzmann distribution of probability $P(\mathbf{X}) \sim \exp(-V(\mathbf{X})/k_B T)$ is a stationary solution of Eq. (40). Using a substitution

$$P(\mathbf{X}, t) = e^{-\left(\frac{1}{k_B T}\right)U(\mathbf{X})} \psi(\mathbf{X}, t) \quad (41)$$

the Fokker-Planck equation can be written in the form of a time dependent Schrödinger equation,

$$-\frac{\partial}{\partial t}\psi(\mathbf{X},t) = \hat{H}_{\text{eff}}\psi(\mathbf{X},t) \quad (42)$$

where the effective Hamiltonian operator is defined as

$$\hat{H}_{\text{eff}} = -D\nabla^2 + V_{\text{eff}}(\mathbf{X}) \quad (43)$$

and the effective potential is defined as

$$V_{\text{eff}}(\mathbf{X}) = \frac{D}{4(k_B T)^2} \left[(\nabla U(\mathbf{X}))^2 - 2k_B T \nabla^2 U(\mathbf{X}) \right] \quad (44)$$

A path-integral representation can be obtained as the solution of Eq. (42) subject to the boundary conditions as that the system is in state *A* at time 0, and state *B* at time *T*. this presentation also allows a switch from the time-dependent Newtonian description to the energy-dependent Hamilton-Jacobi (HJ) description. In the HJ framework, the most probable trajectory can be obtained by minimizing the target HJ functional

$$S_{\text{HJ}}(\mathbf{X}) = \int_A^B dl \sqrt{\frac{\gamma}{k_B T} [E_{\text{eff}} + V_{\text{eff}}(\mathbf{X})]} \quad (45)$$

where $dl = \sqrt{d\mathbf{X}^2}$, γ is the friction, and E_{eff} is an effective energy parameter, which determines the total time of trajectory. In very recent studies, the simulation of reaction pathways connecting two boundary states were carried out to search for the dominant reaction pathway for folding of a small peptide, (Faccioli et al. 2010) an organic reaction from cyclobutene to butadiene, (Beccara et al. 2010) and the folding of tetraalanine at semiempirical level of theory. (Beccara et al. 2011) Although the DPR method is still in its early stage of development, the current progress demonstrates the efficiency of this method in sampling the reaction pathways for rare transition events.

5. Summary

The development of reaction path methods to study the mechanism of rare events of high dimensional systems has been an active research field for more than two decades. Many efficient methods have been developed and applied in numerous applications. The IRC is a well defined reaction pathway of the MEP. It plays an important role in explaining reactions in complicated systems, especially when the reaction coordinates are not simple order parameters. However, one can only generate an IRC after obtaining a TS. In many studies, identifying a TS is the bottleneck of the research. In some cases, the IRC generated from a TS may not connect with the target reactant and product. For large systems, there may exist multiple reaction pathways between the reactant and product. To map out multiple reaction pathways through IRC, it is required to identify multiple TSs *a priori*, which is a very challenging task. The chain-of-states methods were developed to simultaneously optimize a series of replicas as the representation of the reaction pathway connecting two equilibrium states. Starting from an initial guess, the path optimization could lead to a reaction pathway that closely resembles the MEP between two equilibrium states. One advantage of the chain-of-states methods is that no TS is required to generate a reaction path. The replicas with

highest energy from reaction path calculations are often subject to a TS search for better estimation of the reaction barriers.

For large systems, special caution is needed when carrying out the IRC calculations or chain-of-states calculations. To ensure the continuity of the reaction pathways that connect the reactant and product, large amounts of spectator degrees of freedom need to be controlled, or frozen, during the calculations. This treatment can make the working PES cleaner, but has the drawback of ignoring important degrees of freedom that contribute to the free energies, which can be directly compared with the experimental observations. Recently, more research has been reported to develop methods that sample the reaction pathway directly in the transition path space, rather than the phase space. With these developments, multiple reaction pathways could potentially be sampled without obtaining TSs for each pathway *a priori*. For rare events with high thermodynamic or kinetic barriers, the efficiency of sampling in the reaction path space is higher than the one of sampling in the phase space. The free energies could be estimated for each pathway with the sampled ensembles. Currently, the reaction path sampling method is still in its early stage of development for both theory and sampling techniques. It is clear that advanced computational power, theories and sampling techniques are in sight to make the computational study of the mechanisms of rare events for large systems more practical and appealing.

6. Abbreviations

μ -ADDF, microiteration-ADDF;
ABNR, adopted basis Newton-Raphsion;
ADDF, anharmonic downward distortions following;
ANEBA, adaptive nudged elastic band approach;
CI-NEB, climbing image NEB;
CPR, conjugate peak refinement;
dADDF, double-ended ADDF;
DC, dissociation channels;
DNEB, doubly nudged elastic band;
DRP, dominant reaction path;
DVV, damped velocity Verlet;
EK, Elber and Karplus;
ES, equilibrium structure;
ESP, electrostatic potential;
EulerPC, Euler predictor and corrector method;
FTS, finite temperature string;
GRRM, global reaction route mapping;
GS-IRC, Gonzalez and Schlegel IRC;
GSM, Growing string method;
HFB, harmonic Fourier bead;
HJ, Hamilton-Jacobi;
HPC, Hessian based predictor-corrector;
IAN, isobutyryl-ala₃-NH-methyl;
IRC, intrinsic reaction coordinate;
lADDF, large ADDF;
L-BFGS, limited-memory Broyden-Fletcher-Goldfarb-Shanno;

LUP, locally updated planes;
MAP, minimum action path;
MC, Monte Carlo;
MD, molecular dynamics;
MEP, minimum energy path;
MFEP, minimum free energy path;
MHP, minimum Hamiltonian path;
MMP2, matrix metalloproteinase 2;
NEB, nudged elastic band;
ODE, ordinary differential equation;
PES, potential energy surface;
PHVA, partial Hessian vibrational analysis;
PMF, potential of mean force;
QM, quantum mechanics;
QM/MM, combined quantum mechanical and molecular mechanical method;
RMS, root mean square;
RMSD, root mean square distances;
RPATH, reaction path;
SDP, steepest descent path;
SHS, scaled hypersphere search;
SPW, self-penalty walk;
TPE, transition path ensemble;
TPS, transition path sampling;
TPT, transition path theory;
TS, transition state;
TST, transition state theory;
UE, Ulitsky and Elber;
VSA, vibrational subsystem analysis;
ZTS, zero temperature string.

7. References

- Ágoston, P., Albe, K. (2010). Ab initio modeling of diffusion in indium oxide. *Phys. Rev. B*, Vol. 81, No. 19, pp. 195205.
- Alfonso, D. R., Jordan, K. D. (2003). A flexible nudged elastic band program for optimization of minimum energy pathways using ab initio electronic structure methods. *J. Comput. Chem.*, Vol. 24, No. 8, pp. 990-996.
- Autieri, E., Faccioli, P., Sega, M., Pederiva, F., Orland, H. (2009). Dominant reaction pathways in high-dimensional systems. *J. Chem. Phys.*, Vol. 130, No. 6, pp. 064106.
- Ayala, P. Y., Schlegel, H. B. (1997). A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.*, Vol. 107, No. 2, pp. 375-384.
- Bader, R. F. W., Gangi, R. A. (1975). Ab initio calculation of potential energy surfaces, In: *Theoretical Chemistry*. Dixon RN, Thomson C ed. pp. 1-65, Royal Society of Chemistry, ISBN 9780851867649, Cambridge.
- Bair, S., McCabe, C., Cummings, P. (2002). Comparison of Nonequilibrium Molecular Dynamics with Experimental Measurements in the Nonlinear Shear-Thinning Regime. *Phys. Rev. Lett.*, Vol. 88, No., pp. 058302.

- Baker, J. (1986). An algorithm for the location of transition states. *J. Comput. Chem.*, Vol. 7, No. 4, pp. 385-395.
- Banerjee, A., Adams, N., Simons, J., Shepard, R. (1985). Search for stationary points on surfaces. *J. Phys. Chem.*, Vol. 89, No. 1, pp. 52-57.
- Bartels, C., Karplus, M. (1997). Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J. Comput. Chem.*, Vol. 18, No. 12, pp. 1450-1462.
- Beccara, A. S., Faccioli, P., Sega, M., Pederiva, F., Garberoglio, G., Orland, H. (2011). Dominant folding pathways of a peptide chain from ab initio quantum-mechanical simulations. *J. Chem. Phys.*, Vol. 134, No. 2, pp. 024501.
- Beccara, S. A., Garberoglio, G., Faccioli, P., Pederiva, F. (2010). Communications: Ab initio dynamics of rare thermally activated reactions. *J. Chem. Phys.*, Vol. 132, No. 11, pp. 111102.
- Bell, S., Crighton, J. S. (1984). Locating transition states. *J. Chem. Phys.*, Vol. 80, No. 6, pp. 2464-2475.
- Bergonzo, C., Campbell, A. J., Walker, R. C., Simmerling, C. (2009). A Partial Nudged Elastic Band Implementation for Use With Large or Explicitly Solvated Systems. *Int. J. Quantum Chem.*, Vol. 109, No. 15, pp. 3781-3790.
- Beusen, D. (1996). Systematic search strategies in conformational analysis. *Drug Discovery Today*, Vol. 1, No. 10, pp. 429-437.
- Blondel, A., Renaud, J.-P., Fischer, S., Moras, D., Karplus, M. (1999). Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. *J. Mol. Biol.*, Vol. 291, No. 1, pp. 101-115.
- Bofill, J. M., Quapp, W. (2011). Variational nature, integration, and properties of Newton reaction path. *J. Chem. Phys.*, Vol. 134, No. 7, pp. 074101.
- Bolhuis, P. G. (2003). Transition-path sampling of β -hairpin folding. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 100, No. 21, pp. 12129-12134.
- Bolhuis, P. G. (2008). Rare events via multiple reaction channels sampled by path replica exchange. *J. Chem. Phys.*, Vol. 129, No. 11, pp. 114108.
- Bolhuis, P. G., Dellago, C., Geissler, P. L., Chandler, D. (2000). Transition path sampling: throwing ropes over mountains in the dark. *J. Phys.: Condens. Matter*, Vol. 12, No. 8A, pp. A147-A152.
- Bolhuis, P. G., Chandler, D., Dellago, C., Geissler, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, Vol. 53, No. 1, pp. 291-318.
- Bonnans, J. (2003). *Numerical optimization: theoretical and practical aspects*. Springer, ISBN 9783540001911, Berlin.
- Brokaw, J. B., Haas, K. R., Chu, J.-W. (2009). Reaction Path Optimization with Holonomic Constraints and Kinetic Energy Potentials. *J. Chem. Theory Comput.*, Vol. 5, No. 8, pp. 2050-2061.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, Vol. 4, No. 2, pp. 187-217.
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. and others (2009). CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, Vol. 30, No. 10, pp. 1545-1614.

- Bulirsch, R., Stoer, J. (1964). Fehlerabschätzungen und Extrapolation mit rationalen Funktionen bei Verfahren vom Richardson-Typus. *Numer. Math.*, Vol. 6, No. 1, pp. 413-427.
- Bulirsch, R., Stoer, J. (1966a). Numerical treatment of ordinary differential equations by extrapolation methods. *Numer. Math.*, Vol. 8, No. 1, pp. 1-13.
- Bulirsch, R., Stoer, J. (1966b). Asymptotic upper and lower bounds for results of extrapolation methods. *Numer. Math.*, Vol. 8, No. 2, pp. 93-104.
- Bussi, G., Gervasio, F. L., Laio, A., Parrinello, M. (2006). Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.*, Vol. 128, No. 41, pp. 13435-13441.
- Cafilisch, A., Fischer, S., Karplus, M. (1997). Docking by Monte Carlo minimization with a solvation correction: Application to an FKBP-substrate complex. *J. Comput. Chem.*, Vol. 18, No. 6, pp. 723-743.
- Cameron, M., Kohn, R. V., Vanden-Eijnden, E. (2010). The String Method as a Dynamical System. *J. Nonlinear Sci.*, Vol. 21, No. 2, pp. 193-230.
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., Woods, R. J. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.*, Vol. 26, No. 16, pp. 1668-1688.
- Cerjan, C. J. (1981). On finding transition states. *J. Chem. Phys.*, Vol. 75, No. 6, pp. 2800-2806.
- Chandrasekhar, J., Jorgensen, W. L. (1985). Energy profile for a nonconcerted SN2 reaction in solution. *J. Am. Chem. Soc.*, Vol. 107, No. 10, pp. 2974-2975.
- Chandrasekhar, J., Smith, S. F., Jorgensen, W. L. (1985). Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.*, Vol. 107, No. 1, pp. 154-163.
- Choi, C., Elber, R. (1991). Reaction path study of helix formation in tetrapeptides: Effect of side chains. *J. Chem. Phys.*, Vol. 94, No. 1, pp. 751-760.
- Chu, J.-W., Trout, B. L., Brooks, B. R. (2003). A super-linear minimization scheme for the nudged elastic band method. *J. Chem. Phys.*, Vol. 119, No. 24, pp. 12708-12717.
- Cisneros, G. A., Liu, H. Y., Lu, Z. Y., Yang, W. T. (2005). Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods. *J. Chem. Phys.*, Vol. 122, No. 11, pp. 114502.
- Cisneros, G. a. S., Perera, L., Schaaper, R. M., Pedersen, L. C., London, R. E., Pedersen, L. G., Darden, T. A. (2009). Reaction Mechanism of the ϵ Subunit of E. coli DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.*, Vol. 131, No. 4, pp. 1550-1556.
- Cummings, P. T., Evans, D. J. (1992). Nonequilibrium molecular dynamics approaches to transport properties and non-Newtonian fluid rheology. *Ind. Eng. Chem. Res.*, Vol. 31, No. 5, pp. 1237-1252.
- Czermanski, R., Elber, R. (1989). Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 86, No. 18, pp. 6963-6967.
- Czermanski, R., Elber, R. (1990a). Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *Int. J. Quantum Chem.*, Vol. 38, No. S24, pp. 167-185.

- Czerminski, R., Elber, R. (1990b). Reaction path study of conformational transitions in flexible systems: Applications to peptides. *J. Chem. Phys.*, Vol. 92, No. 9, pp. 5580-5601.
- De La Cruz, E. M., Ostap, E. M., Sweeney, H. L. (2001). Kinetic Mechanism and Regulation of Myosin VI. *J. Biol. Chem.*, Vol. 276, No. 34, pp. 32373-32381.
- Dellago, C., Bolhuis, P. (2004). Activation Energies from Transition Path Sampling Simulations. *Mol. Simul.*, Vol. 30, No. 11, pp. 795-799.
- Dellago, C., Bolhuis, P. G. (2009). Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events, In: *Advanced Computer Simulation Approaches for Soft Matter Sciences III*. Holm C, Kremer K ed. pp. 167-233, Springer, ISBN 9783540877059, Berlin, Heidelberg.
- Dellago, C., Bolhuis, P. G., Chandler, D. (1999). On the calculation of reaction rate constants in the transition path ensemble. *J. Chem. Phys.*, Vol. 110, No. 14, pp. 6617-6625.
- Dellago, C., Bolhuis, P. G., Geissler, P. L. (2002). Transition path sampling, In: *Advances in Chemical Physics, Vol 123*. pp. 1-78, John Wiley & Sons Inc, ISBN 9780471231509, New York.
- Dennis, J., Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. Society for Industrial and Applied Mathematics, ISBN 9780898713640, Philadelphia.
- Doll, J. D. (2005). A Modern Perspective on Transition State Theory, In: *Handbook of Materials Modeling*. Yip S ed. pp. 1573-1583, Springer Netherlands, ISBN 9781402032875, Dordrecht.
- Dutzler, R., Schirmer, T., Karplus, M., Fischer, S. (2002). Translocation Mechanism of Long Sugar Chains across the Maltoporin Membrane Channel. *Structure*, Vol. 10, No. 9, pp. 1273-1284.
- E, W., Vanden-Eijnden, E. (2006). Towards a Theory of Transition Paths. *J. Stat. Phys.*, Vol. 123, No. 3, pp. 503-523.
- E, W., Vanden-Eijnden, E. (2010). Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.*, Vol. 61, No. 1, pp. 391-420.
- E, W., Ren, W., Vanden-Eijnden, E. (2002). String method for the study of rare events. *Phys. Rev. B*, Vol. 66, No. 5, pp. 052301.
- E, W., Ren, W., Vanden-Eijnden, E. (2004). Minimum action method for the study of rare events. *Comm. Pure Appl. Math.*, Vol. 57, No. 5, pp. 637-656.
- E, W., Ren, W., Vandeneijnden, E. (2005a). Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.*, Vol. 413, No. 1-3, pp. 242-247.
- E, W., Ren, W., Vanden-Eijnden, E. (2005b). Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B*, Vol. 109, No. 14, pp. 6688-6693.
- E, W., Ren, W., Vanden-Eijnden, E. (2007). Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, Vol. 126, No. 16, pp. 164103.
- Earl, D. J., Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, Vol. 7, No. 23, pp. 3910-3916.
- Eastman, P., Grønbech-Jensen, N., Doniach, S. (2001). Simulation of protein folding by reaction path annealing. *J. Chem. Phys.*, Vol. 114, No. 8, pp. 3823-3841.

- Elber, R. (1987). A method for determining reaction paths in large molecules: Application to myoglobin. *Chem. Phys. Lett.*, Vol. 139, No. 5, pp. 375-380.
- Eyring, H. (1935). The Activated Complex in Chemical Reactions. *J. Chem. Phys.*, Vol. 3, No. 2, pp. 107-115.
- Faccioli, P. (2008). Characterization of Protein Folding by Dominant Reaction Pathways. *J. Phys. Chem. B*, Vol. 112, No. 44, pp. 13756-13764.
- Faccioli, P., Lonardi, A., Orland, H. (2010). Dominant reaction pathways in protein folding: A direct validation against molecular dynamics simulations. *J. Chem. Phys.*, Vol. 133, No. 4, pp. 045104.
- Faccioli, P., Sega, M., Pederiva, F., Orland, H. (2006). Dominant Pathways in Protein Folding. *Phys. Rev. Lett.*, Vol. 97, No. 10, pp. 108101.
- Farkas, Ö., Schlegel, H. B. (2003). Geometry optimization methods for modeling large molecules. *THEOCHEM*, Vol. 666-667, No. 1, pp. 31-39.
- Fischer, S. (1992). Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.*, Vol. 194, No. 3, pp. 252-261.
- Fischer, S., Michnick, S., Karplus, M. (1993). A mechanism for rotamase catalysis by the FK506 binding protein (FKBP). *Biochemistry*, Vol. 32, No. 50, pp. 13830-13837.
- Fischer, S., Dunbrack, R. L., Karplus, M. (1994). Cis-Trans Imide Isomerization of the Proline Dipeptide. *J. Am. Chem. Soc.*, Vol. 116, No. 26, pp. 11931-11937.
- Fischer, S., Grootenhuys, P. D. J., Groenen, L. C., Van Hoorn, W. P., Van Veggel, F. C. J. M. V., Reinhoudt, D. N., Karplus, M. (1995). Pathways for Conformational Interconversion of Calix[4]arenes. *J. Am. Chem. Soc.*, Vol. 117, No. 5, pp. 1611-1620.
- Fletcher, R. (2000). *Practical methods of optimization*. Wiley, ISBN 9780471494638, Chichester.
- Floudas, C. A., Pardalos, P. M. (1996). *State of the art in global optimization: computational methods and applications*. Kluwer Academic, ISBN 9780792338383.
- Freidlin, M. (1984). *Random perturbations of dynamical systems*. Springer, ISBN 9780387908588, New York.
- Fukui, K. (1981). The path of chemical reactions - The IRC approach. *Acc. Chem. Res.*, Vol. 14, No. 12, pp. 363-368.
- Galván, I. F., Field, M. J. (2008). Improving the efficiency of the NEB reaction path finding algorithm. *J. Comput. Chem.*, Vol. 29, No. 1, pp. 139-143.
- Gear, C. (1971). In: *Numerical initial value problems in ordinary differential equations.*, Prentice-Hall, ISBN 9780136266068, Englewood Cliffs N.J.
- Gill, P. (1982). *Practical optimization*. Academic, ISBN 9780122839528, London.
- Golden, S., Olsen, K. (2008). Use of the Conjugate Peak Refinement Algorithm for Identification of Ligand-Binding Pathways in Globins. *Methods in Enzymology*, Vol. 437, No. 1, pp. 417-437.
- Gonzalez, C., Schlegel, H. B. (1989). An improved algorithm for reaction path following. *J. Chem. Phys.*, Vol. 90, No. 4, pp. 2154-2161.
- Gonzalez, C., Schlegel, H. B. (1990). Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.*, Vol. 94, No. 14, pp. 5523-5527.
- Gonzalez, C., Schlegel, H. B. (1991). Improved algorithms for reaction path following: Higher-order implicit algorithms. *J. Chem. Phys.*, Vol. 95, No. 8, pp. 5853-5860.

- Goodrow, A., Bell, A. T. (2008). A Theoretical Investigation of the Selective Oxidation of Methanol to Formaldehyde on Isolated Vanadate Species Supported on Titania. *J. Phys. Chem. C*, Vol. 112, No. 34, pp. 13204-13214.
- Goodrow, A., Bell, A. T., Head-Gordon, M. (2008). Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *J. Chem. Phys.*, Vol. 129, No. 17, pp. 174109.
- Goodrow, A., Bell, A. T., Head-Gordon, M. (2009). Transition state-finding strategies for use with the growing string method. *J. Chem. Phys.*, Vol. 130, No. 24, pp. 244108.
- Goodrow, A., Bell, A. T., Head-Gordon, M. (2010). A strategy for obtaining a more accurate transition state estimate using the growing string method. *Chem. Phys. Lett.*, Vol. 484, No. 4-6, pp. 392-398.
- Gruia, A. D., Bondar, A.-N., Smith, J. C., Fischer, S. (2005). Mechanism of a Molecular Valve in the Halorhodopsin Chloride Pump. *Structure*, Vol. 13, No. 4, pp. 617-627.
- Hagan, M. F. (2003). Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 100, No. 24, pp. 13922-13927.
- Hamelberg, D., Mongan, J., Mccammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, Vol. 120, No. 24, pp. 11919-11929.
- Hänggi, P., Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, Vol. 62, No. 2, pp. 251-341.
- Haslinger, J., Mäkinen, R. a. E. (2003). *Introduction to Shape Optimization: Theory, Approximation, and Computation*. SIAM, ISBN 9780898715361, Philadelphia, PA.
- Heidrich, D. (1995). *The reaction path in chemistry: current approaches and perspectives*. Kluwer Academic Publishers, ISBN 9780792335894, Boston, MA.
- Henkelman, G., Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, Vol. 113, No. 22, pp. 9978-9985.
- Henkelman, G., Jóhannesson, G., Jónsson, H. (2000a). Methods for Finding Saddle Points and Minimum Energy Paths, In: *Theoretical Methods in Condensed Phase Chemistry*, Schwartz SD ed. pp. 269-302, Kluwer Academic Publishers, ISBN 9780792366874, New York.
- Henkelman, G., Uberuaga, B. P., Jónsson, H. (2000b). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, Vol. 113, No. 22, pp. 9901-9904.
- Hoover, W. G. (1983). Nonequilibrium Molecular Dynamics. *Annu. Rev. Phys. Chem.*, Vol. 34, No. 1, pp. 103-127.
- Hoover, W. G., Hoover, C. G. (2005). Nonequilibrium molecular dynamics. *Condensed Matter Physics*, Vol. 8, No. 2, pp. 247-260.
- Horst, R. (1995). *Handbook of global optimization*. Kluwer Academic Publishers, ISBN 9780792331209, Dordrecht; Boston.
- Horst, R. (2000). *Introduction to global optimization*. Kluwer Academic Publishers, ISBN 9780792365747, Dordrecht; Boston.
- Hratchian, H. P., Schlegel, H. B. (2002). Following Reaction Pathways Using a Damped Classical Trajectory Algorithm. *J. Phys. Chem. A*, Vol. 106, No. 1, pp. 165-169.

- Hratchian, H. P., Schlegel, H. B. (2004). Accurate reaction paths using a Hessian based predictor-corrector integrator. *J. Chem. Phys.*, Vol. 120, No. 21, pp. 9918-9924.
- Hratchian, H., Schlegel, H. B. (2005a). Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces, In: *Theory and Applications of Computational Chemistry*. pp. 195-249, Elsevier, ISBN 9780444517197.
- Hratchian, H. P., Schlegel, H. B. (2005b). Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method. *J. Chem. Theory Comput.*, Vol. 1, No. 1, pp. 61-69.
- Hratchian, H. P., Frisch, M. J. (2011). Integrating steepest-descent reaction pathways for large molecules. *J. Chem. Phys.*, Vol. 134, No. 20, pp. 204103.
- Hratchian, H. P., Frisch, M. J., Schlegel, H. B. (2010). Steepest descent reaction path integration using a first-order predictor-corrector method. *J. Chem. Phys.*, Vol. 133, No. 22, pp. 224101.
- Hu, H., Yang, W. (2008). Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.*, Vol. 59, No. 1, pp. 573-601.
- Hu, H., Lu, Z., Parks, J. M., Burger, S. K., Yang, W. (2008). Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: Sequential sampling and optimization on the potential of mean force surface. *J. Chem. Phys.*, Vol. 128, No. 3, pp. 034105.
- Ionova, I. V., Carter, E. A. (1993). Ridge method for finding saddle points on potential energy surfaces. *J. Chem. Phys.*, Vol. 98, No. 8, pp. 6377-6386.
- Ionova, I. V., Carter, E. A. (1995). Direct inversion in the iterative subspace-induced acceleration of the ridge method for finding transition states. *J. Chem. Phys.*, Vol. 103, No. 13, pp. 5437-5441.
- Jensen, A. (1983). Location of transition states and stable intermediates by MINIMAX/MINIMI optimization of synchronous transit pathways. *Theor. Chim. Acta*, Vol. 63, No. 4, pp. 269-290.
- Jónsson, H., Mills, G., Jacobsen, K. W. (1998). Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions, In: *Classical and Quantum Dynamics in Condensed Phase Simulations*. Berne BJ, Ciccotti G, Coker DF ed. pp. 385-404, World Scientific, ISBN 9789810234980, Singapore.
- Jorgensen, W. L. (1989). Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.*, Vol. 22, No. 5, pp. 184-189.
- Juraszek, J., Bolhuis, P. G. (2006). Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 103, No. 43, pp. 15859-15864.
- Kästner, J. (2011). Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 1, No. 6, pp. 932-942.
- Khavrutskii, I. V., Mccammon, J. A. (2007). Generalized gradient-augmented harmonic Fourier beads method with multiple atomic and/or center-of-mass positional restraints. *J. Chem. Phys.*, Vol. 127, No. 12, pp. 124901.
- Khavrutskii, I. V., Arora, K., Brooks, C. L. (2006). Harmonic Fourier beads method for studying rare events on rugged energy surfaces. *J. Chem. Phys.*, Vol. 125, No. 17, pp. 174108.

- Khavrutskii, I. V., Dzubiella, J., Mccammon, J. A. (2008a). Computing accurate potentials of mean force in electrolyte solutions with the generalized gradient-augmented harmonic Fourier beads method. *J. Chem. Phys.*, Vol. 128, No. 4, pp. 044106.
- Khavrutskii, I. V., Fajer, M., Mccammon, J. A. (2008b). Intrinsic Free Energy of the Conformational Transition of the KcsA Signature Peptide from Conducting to Nonconducting State. *J. Chem. Theory Comput.*, Vol. 4, No. 9, pp. 1541-1554.
- Kjelstrup, S., Hafskjold, B. R. (1996). Nonequilibrium Molecular Dynamics Simulations of Steady-State Heat and Mass Transport in Distillation. *Ind. Eng. Chem. Res.*, Vol. 35, No. 11, pp. 4203-4213.
- Klähn, M., Braun-Sand, S., Rosta, E., Warshel, A. (2005). On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B*, Vol. 109, No. 32, pp. 15645-15650.
- Laidler, K. J., King, M. C. (1983). Development of transition-state theory. *J. Phys. Chem.*, Vol. 87, No. 15, pp. 2657-2664.
- Laio, A. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 99, No. 20, pp. 12562-12566.
- Laio, A., Gervasio, F. L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, Vol. 71, No. 12, pp. 126601.
- Leach, A. R. (1991). A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules, In: *Reviews in Computational Chemistry, Vol 2*. Lipkowitz KB, Boyd DB ed. pp. 1-55, John Wiley & Sons, Inc., ISBN 9780470125793, Hoboken, NJ, USA.
- Lei, H., Duan, Y. (2007). Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.*, Vol. 17, No. 2, pp. 187-191.
- Lewars, E. G. (2011). The Concept of the Potential Energy Surface, In: *Computational Chemistry*. Lewars EG ed. pp. 9-43, Springer Netherlands, ISBN 9789048138609, Dordrecht.
- Li, H., Jensen, J. H. (2002). Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.*, Vol. 107, No. 4, pp. 211-219.
- Li, S., Sheng, N., Liu, X. (2008). A non-equilibrium multiscale simulation paradigm. *Chem. Phys. Lett.*, Vol. 451, No. 4-6, pp. 293-300.
- Lipfert, J., Franklin, J., Wu, F., Doniach, S. (2005). Protein Misfolding and Amyloid Formation for the Peptide GNNQQNY from Yeast Prion Protein Sup35: Simulation by Reaction Path Annealing. *J. Mol. Biol.*, Vol. 349, No. 3, pp. 648-658.
- Liu, H., Lu, Z., Cisneros, G. a. S., Yang, W. (2004). Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.*, Vol. 121, No. 2, pp. 697-706.
- Lutsko, J. F. (2008). Density functional theory of inhomogeneous liquids. III. Liquid-vapor nucleation. *J. Chem. Phys.*, Vol. 129, No. 24, pp. 244501.
- Maeda, S. (2003). A new method for constructing multidimensional potential energy surfaces by a polar coordinate interpolation technique. *Chem. Phys. Lett.*, Vol. 381, No. 1-2, pp. 177-186.

- Maeda, S., Ohno, K. (2007). Structures of Water Octamers (H₂O)₈: Exploration on Ab Initio Potential Energy Surfaces by the Scaled Hypersphere Search Method. *J. Phys. Chem. A*, Vol. 111, No. 20, pp. 4527-4534.
- Maeda, S., Ohno, K., Morokuma, K. (2009). An Automated and Systematic Transition Structure Explorer in Large Flexible Molecular Systems Based on Combined Global Reaction Route Mapping and Microiteration Methods. *J. Chem. Theory Comput.*, Vol. 5, No. 10, pp. 2734-2743.
- Maragakis, P., Andreev, S. A., Brumer, Y., Reichman, D. R., Kaxiras, E. (2002). Adaptive nudged elastic band approach for transition state calculation. *J. Chem. Phys.*, Vol. 117, No. 10, pp. 4651-4658.
- Maragliano, L., Fischer, A., Vanden-Eijnden, E., Ciccotti, G. (2006). String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, Vol. 125, No. 2, pp. 024106.
- March, J. (1992). *Advanced organic chemistry: Reactions, mechanisms and structures*. Wiley, ISBN 9780471601807, New York N.Y.
- Maresh, J. J., Giddings, L.-A., Friedrich, A., Loris, E. A., Panjikar, S., Trout, B. L., Stöckigt, J., Peters, B., O'connor, S. E. (2008). Strictosidine Synthase: Mechanism of a Pictet-Spengler Catalyzing Enzyme. *J. Am. Chem. Soc.*, Vol. 130, No. 2, pp. 710-723.
- Mazzola, G., Beccara, S. A., Faccioli, P., Orland, H. (2011). Fluctuations in the ensemble of reaction pathways. *J. Chem. Phys.*, Vol. 134, No. 16, pp. 164109.
- Metzner, P., Schütte, C., Vanden-Eijnden, E. (2006). Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, Vol. 125, No. 8, pp. 084110.
- Mezey, P. (1987). *Potential energy hypersurfaces*. Elsevier, ISBN 9780444428875, Amsterdam; New York.
- Mills, G. (1995). Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.*, Vol. 324, No. 2-3, pp. 305-337.
- Moroni, D., Vanerp, T., Bolhuis, P. (2004). Investigating rare events by transition interface sampling. *Physica. A*, Vol. 340, No. 1-3, pp. 395-401.
- Müller, K., Brown, L. D. (1979). Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta*, Vol. 53, No. 1, pp. 75-93.
- Mundy, C. J., Balasubramanian, S., Bagchi, K., Tuckerman, M. E., Martyna, G. J., Klein, M. L. (2000). Nonequilibrium Molecular Dynamics, In: *Reviews in Computational Chemistry*, Vol. 14. Lipkowitz KB, Boyd DB ed. pp. 291-397, John Wiley & Sons, Inc., ISBN 9780470125915, Hoboken, NJ, USA.
- Nichols, J., Taylor, H., Schmidt, P., Simons, J. (1990). Walking on potential energy surfaces. *J. Chem. Phys.*, Vol. 92, No. 1, pp. 340-346.
- Nocedal, J. (2006). *Numerical optimization*. Springer, ISBN 9780387400655, New York.
- Noé, F., Krachtus, D., Smith, J. C., Fischer, S. (2006). Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *J. Chem. Theory Comput.*, Vol. 2, No. 3, pp. 840-857.
- Nowak, W., Czerminski, R., Elber, R. (1991). Reaction path study of ligand diffusion in proteins: application of the self penalty walk (SPW) method to calculate reaction coordinates for the motion of CO through leghemoglobin. *J. Am. Chem. Soc.*, Vol. 113, No. 15, pp. 5627-5637.

- Ohno, K. (2004). A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem. Phys. Lett.*, Vol. 384, No. 4-6, pp. 277-282.
- Ohno, K., Maeda, S. (2006). Global Reaction Route Mapping on Potential Energy Surfaces of Formaldehyde, Formic Acid, and Their Metal-Substituted Analogues. *J. Phys. Chem. A*, Vol. 110, No. 28, pp. 8933-8941.
- Okur, A., Wickstrom, L., Layten, M., Geney, R. E., Song, K., Hornak, V., Simmerling, C. (2006). Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.*, Vol. 2, No. 2, pp. 420-433.
- Olender, R. (1997). Yet another look at the steepest descent path. *THEOCHEM*, Vol. 398-399, No. 1, pp. 63-71.
- Olsen, R. A., Kroes, G. J., Henkelman, G., Arnaldsson, A., Jónsson, H. (2004). Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.*, Vol. 121, No. 20, pp. 9776-9792.
- Onsager, L., Machlup, S. (1953). Fluctuations and Irreversible Processes. *Phys. Rev.*, Vol. 91, No. 6, pp. 1505-1512.
- Ovchinnikov, V., Karplus, M., Vanden-Eijnden, E. (2011). Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *J. Chem. Phys.*, Vol. 134, No. 8, pp. 085103.
- Parish, C. (2002). A comparison of the Low Mode and Monte Carlo conformational search methods. *J. Mol. Graphics Modell.*, Vol. 21, No. 2, pp. 129-150.
- Pechukas, P. (1981). Transition State Theory. *Annu. Rev. Phys. Chem.*, Vol. 32, No. 1, pp. 159-177.
- Peng, C., Ayala, P. Y., Schlegel, H. B., Frisch, M. J. (1996). Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.*, Vol. 17, No. 1, pp. 49-56.
- Peters, B., Heyden, A., Bell, A. T., Chakraborty, A. (2004). A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, Vol. 120, No. 17, pp. 7877-7886.
- Pratt, L. R. (1986). A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.*, Vol. 85, No. 9, pp. 5045-5048.
- Quapp, W. (2005). A growing string method for the reaction pathway defined by a Newton trajectory. *J. Chem. Phys.*, Vol. 122, No. 17, pp. 174106.
- Quapp, W. (2008). Chemical reaction paths and calculus of variations. *Theor. Chem. Acc.*, Vol. 121, No. 5-6, pp. 227-237.
- Quapp, W. (2009). The growing string method for flows of Newton trajectories by a second-order method. *J. Theor. Comput. Chem.*, Vol. 8, No. 1, pp. 101-117.
- Quapp, W., Heidrich, D. (1984). Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems. *Theor. Chim. Acta*, Vol. 66, No. 3-4, pp. 245-260.
- Quaytman, S. L., Schwartz, S. D. (2007). Reaction coordinate of an enzymatic reaction revealed by transition path sampling. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 104, No. 30, pp. 12253-12258.

- Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C., Parrinello, M. (2006). Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B*, Vol. 110, No. 8, pp. 3533-3539.
- Ren, W. (2003). Higher order string method for finding minimum energy paths *COMM. MATH. SCI.*, Vol. 1, No. 2, pp. 377-384.
- Ren, W., Vanden-Eijnden, E., Maragakis, P., E, W. (2005). Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, Vol. 123, No. 13, pp. 134109.
- Rogal, J., Bolhuis, P. G. (2008). Multiple state transition path sampling. *J. Chem. Phys.*, Vol. 129, No. 22, pp. 224107.
- Rosta, E., Nowotny, M., Yang, W., Hummer, G. (2011). Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.*, Vol. 133, No. 23, pp. 8934-8941.
- Santos, J. S.-D. O., Fischer, S., Guilbert, C., Lewit-Bentley, A., Smith, J. C. (2000). Pathway for Large-Scale Conformational Change in Annexin V. *Biochemistry*, Vol. 39, No. 46, pp. 14065-14074.
- Scales, L. (1985). *Introduction to non-linear optimization*. Springer-Verlag, ISBN 9780387912523, New York.
- Schlegel, H. B. (1982). Optimization of equilibrium geometries and transition structures. *J. Comput. Chem.*, Vol. 3, No. 2, pp. 214-218.
- Schlegel, H. B. (2003). Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.*, Vol. 24, No. 12, pp. 1514-1527.
- Schlegel, H. B. (2011). Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 1, No. 5, pp. 790-809.
- Sega, M., Faccioli, P., Pederiva, F., Garberoglio, G., Orland, H. (2007). Quantitative Protein Dynamics from Dominant Folding Pathways. *Phys. Rev. Lett.*, Vol. 99, No. 11, pp. 118102.
- Shi, Q., Meroueh, S. O., Fisher, J. F., Mobashery, S. (2008). Investigation of the Mechanism of the Cell Wall DD-Carboxypeptidase Reaction of Penicillin-Binding Protein 5 of Escherichia coli by Quantum Mechanics/Molecular Mechanics Calculations. *J. Am. Chem. Soc.*, Vol. 130, No. 29, pp. 9293-9303.
- Simons, J., Nichols, J. (1990). Strategies for walking on potential energy surfaces using local quadratic approximations. *Int. J. Quantum Chem.*, Vol. 38, No. S24, pp. 263-276.
- Sørensen, M., Jacobsen, K., Jónsson, H. (1996). Thermal Diffusion Processes in Metal-Tip-Surface Interactions: Contact Formation and Adatom Mobility. *Phys. Rev. Lett.*, Vol. 77, No. 25, pp. 5067-5070.
- Sugita, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, Vol. 314, No. 1-2, pp. 141-151.
- Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. (1996). ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.*, Vol. 100, No. 50, pp. 19357-19363.
- Swendsen, R. H., Wang, J.-S. (1986). Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.*, Vol. 57, No. 21, pp. 2607-2609.

- Tachibana, A., Fukui, K. (1980). Novel variational principles of chemical reaction. *Theor. Chim. Acta*, Vol. 57, No. 1, pp. 81-94.
- Tao, P., Gatti, D. L., Schlegel, H. B. (2009a). The Energy Landscape of 3-Deoxy-D-mannooctulosonate 8-Phosphate Synthase. *Biochemistry*, Vol. 48, No. 49, pp. 11706-11714.
- Tao, P., Fisher, J. F., Shi, Q., Mobashery, S., Schlegel, H. B. (2010). Matrix Metalloproteinase 2 (MMP2) Inhibition: DFT and QM/MM Studies of the Deprotonation-Initialized Ring-Opening Reaction of the Sulfoxide Analogue of SB-3CT. *J. Phys. Chem. B*, Vol. 114, No. 2, pp. 1030-1037.
- Tao, P., Fisher, J. F., Shi, Q., Vreven, T., Mobashery, S., Schlegel, H. B. (2009b). Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue. *Biochemistry*, Vol. 48, No. 41, pp. 9839-9847.
- Taylor, H., Simons, J. (1985). Imposition of geometrical constraints on potential energy surface walking procedures. *J. Phys. Chem.*, Vol. 89, No. 4, pp. 684-688.
- Thomas, G. L., Sessions, R. B., Parker, M. J. (2005). Density guided importance sampling: application to a reduced model of protein folding. *Bioinformatics*, Vol. 21, No. 12, pp. 2839-2843.
- Tolman, R. C. (1925). The mechanism of chemical reaction. *J. Am. Chem. Soc.*, Vol. 47, No. 6, pp. 1524-1553.
- Torn, A. (1989). *Global optimization*. Springer-Verlag, ISBN 9783540508717, Berlin; New York.
- Torrie, G., Valleau, J. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, Vol. 23, No. 2, pp. 187-199.
- Torrie, G. M., Valleau, J. P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.*, Vol. 28, No. 4, pp. 578-581.
- Truhlar, D. G. (2001). Potential Energy Surfaces, In: *Encyclopedia of Physical Science and Technology*, 3rd ed. Meyers RA ed. pp. 9-17, Academic Press, ISBN 9780122274107, New York.
- Truhlar, D. G., Hase, W. L., Hynes, J. T. (1983). Current status of transition-state theory. *J. Phys. Chem.*, Vol. 87, No. 15, pp. 2664-2682.
- Trygubenko, S. A., Wales, D. J. (2004a). Erratum: A doubly nudged elastic band method for finding transition states [J. Chem. Phys. 120, 2082 (2004)]. *J. Chem. Phys.*, Vol. 120, No. 16, pp. 7820.
- Trygubenko, S. A., Wales, D. J. (2004b). A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.*, Vol. 120, No. 5, pp. 2082-2094.
- Uberuaga, B., Leskovar, M., Smith, A., Jónsson, H., Olmstead, M. (2000). Diffusion of Ge below the Si(100) Surface: Theory and Experiment. *Phys. Rev. Lett.*, Vol. 84, No. 11, pp. 2441-2444.
- Ulitsky, A., Elber, R. (1990). A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.*, Vol. 92, No. 2, pp. 1510-1511.
- Van Erp, T. S., Moroni, D., Bolhuis, P. G. (2003). A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, Vol. 118, No. 17, pp. 7762-7774.
- Vanden-Eijnden, E. (2006). Transition Path Theory, In: *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Vol. 1*. Ferrario M, Ciccotti G, Binder K ed. pp. 453-493, Springer, ISBN 9783540352709, Berlin, Heidelberg.

- Vanden-Eijnden, E., Heymann, M. (2008). The geometric minimum action method for computing minimum energy paths. *J. Chem. Phys.*, Vol. 128, No. 6, pp. 061103.
- Vanden-Eijnden, E., Venturoli, M. (2009). Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, Vol. 130, No. 19, pp. 194103.
- Vanerp, T., Bolhuis, P. (2005). Elaborating transition interface sampling methods. *J. Comput. Phys.*, Vol. 205, No. 1, pp. 157-181.
- Vélez, P., Dassie, S. A., Leiva, E. P. M. (2008). When do nanowires break? A model for the theoretical study of the long-term stability of monoatomic nanowires. *Chem. Phys. Lett.*, Vol. 460, No. 1-3, pp. 261-265.
- Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, Vol. 159, No. 1, pp. 98-103.
- Verma, C. S., Fischer, S., Caves, L. S. D., Roberts, G. C. K., Hubbard, R. E. (1996). Calculation of the Reaction Pathway for the Aromatic Ring Flip in Methotrexate Complexed to Dihydrofolate Reductase. *J. Phys. Chem.*, Vol. 100, No. 7, pp. 2510-2518.
- Villarba, M., Jonsson, H. (1994). Diffusion mechanisms relevant to metal crystal growth: Pt/Pt(111). *Surf. Sci.*, Vol. 317, No. 1-2, pp. 15-36.
- Vreven, T., Morokuma, K., Farkas, O., Schlegel, H. B., Frisch, M. J. (2003). Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.*, Vol. 24, No. 6, pp. 760-769.
- Vreven, T., Frisch, M. J., Kudin, K. N., Schlegel, H. B., Morokuma, K. (2006a). Geometry optimization with QM/MM methods II: Explicit quadratic coupling. *Mol. Phys.*, Vol. 104, No. 5-7, pp. 701-714.
- Vreven, T., Byun, K. S., Komaromi, I., Dapprich, S., Montgomery, J. A., Jr., Morokuma, K., Frisch, M. J. (2006b). Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.*, Vol. 2, No. 3, pp. 815-826.
- Wales, D. (2003). *Energy landscapes*. Cambridge University Press, ISBN 9780521814157, Cambridge, UK; New York.
- Wen, E. Z., Hsieh, M.-J., Kollman, P. A., Luo, R. (2004). Enhanced ab initio protein folding simulations in Poisson-Boltzmann molecular dynamics with self-guiding forces. *J. Mol. Graphics Modell.*, Vol. 22, No. 5, pp. 415-424.
- Wigner, E. (1938). The transition state method. *Trans. Faraday Soc.*, Vol. 34, No. 1, pp. 29-41.
- Woodcock, H. L., Hodošček, M., Sherwood, P., Lee, Y. S., Schaefer III, H. F., Brooks, B. R. (2003). Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.*, Vol. 109, No. 3, pp. 140-148.
- Woodcock, H. L., Zheng, W., Ghysels, A., Shao, Y., Kong, J., Brooks, B. R. (2008). Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J. Chem. Phys.*, Vol. 129, No. 21, pp. 214109.
- Xie, L., Liu, H., Yang, W. (2004). Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes. *J. Chem. Phys.*, Vol. 120, No. 17, pp. 8039-8052.
- Yamamoto, T. (1960). Quantum Statistical Mechanical Theory of the Rate of Exchange Chemical Reactions in the Gas Phase. *J. Chem. Phys.*, Vol. 33, No. 1, pp. 281-289.

- Yamashita, K., Yamabe, T., Fukui, K. (1981). IRC approach to chemical dynamics: toward mode-selective chemical reactions. *Chem. Phys. Lett.*, Vol. 84, No. 1, pp. 123-126.
- Yang, J., Hu, W., Chen, S., Tang, J. (2009). Surface Self-Diffusion Behavior of a Pt Adatom on Wulff Polyhedral Clusters. *J. Phys. Chem. C*, Vol. 113, No. 52, pp. 21501-21505.
- Yang, X., Maeda, S., Ohno, K. (2005). Global Investigation on the Potential Energy Surface of CH₃CN: Application of the Scaled Hypersphere Search Method. *J. Phys. Chem. A*, Vol. 109, No. 32, pp. 7319-7328.
- Zhang, Y., Liu, H., Yang, W. (2000). Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.*, Vol. 112, No. 8, pp. 3483-3492.
- Zhao, Z., Liu, H. (2008). A Quantum Mechanical/Molecular Mechanical Study on the Catalysis of the Pyridoxal 5'-Phosphate-Dependent Enzyme L-Serine Dehydratase. *J. Phys. Chem. B*, Vol. 112, No. 41, pp. 13091-13100.
- Zheng, X., Bell, A. T. (2008a). Theoretical Analysis of the Mechanism for the Oxidative Carbonylation of Toluene to p-Toluic Acid by Rhodium Complexes. *J. Phys. Chem. C*, Vol. 112, No. 6, pp. 2129-2136.
- Zheng, X., Bell, A. T. (2008b). A Theoretical Investigation of Dimethyl Carbonate Synthesis on Cu-Y Zeolite. *J. Phys. Chem. C*, Vol. 112, No. 13, pp. 5043-5047.
- Zhou, J., Tao, P., Fisher, J. F., Shi, Q., Mobashery, S., Schlegel, H. B. (2010). QM/MM Studies of the Matrix Metalloproteinase 2 (MMP2) Inhibition Mechanism of (S)-SB-3CT and its Oxirane Analogue. *J. Chem. Theory Comput.*, Vol. 6, No. 11, pp. 3580-3587.
- Zhou, X., Ren, W., E, W. (2008). Adaptive minimum action method for the study of rare events. *J. Chem. Phys.*, Vol. 128, No. 10, pp. 104111.