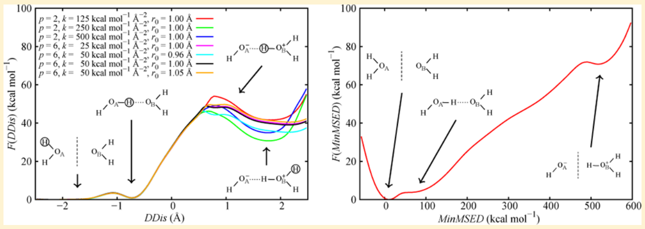
THE JOURNAL OF PHYSICAL CHEMISTRY B

Topologically Invariant Reaction Coordinates for Simulating Multistate Chemical Reactions

Letif Mones* and Gábor Csányi

Engineering Laboratory, University of Cambridge, Cambridge, United Kingdom

ABSTRACT: Evaluating free energy profiles of chemical reactions in complex environments such as solvents and enzymes requires extensive sampling, which is usually performed by potential of mean force (PMF) techniques. The reliability of the sampling depends not only on the applied PMF method but also the reaction coordinate space within the dynamics is biased. In contrast to simple geometrical collective variables that depend only on the positions of the atomic coordinates of the reactants, the E_{gap} reaction coordinate (the energy difference obtained by evaluating a suitable force field using reactant and product state topologies) has the unique property that it is able to take environmental effects into account leading to better convergence, a more faithful description of the transition state ensemble and therefore more accurate free energy profiles. However, E_{gap} requires predefined topologies and is therefore inapplicable for multistate reactions, in which the barrier between the chemically equivalent topologies is comparable to the reaction activation barrier, because undesired "side reactions" occur. In this article, we introduce a new energy-based collective variable by generalizing the E_{gap} reaction coordinate such that it becomes invariant to equivalent topologies and show that it yields more well behaved free energy profiles than simpler geometrical reaction coordinates.



INTRODUCTION

Molecular dynamics (MD) simulation is an established tool to investigate chemical reactions in solutions and enzymatic environments. During the cleavage and formation of the chemical bonds the charge density distribution can change significantly, and therefore the description of these phenomena typically requires a quantum mechanical (QM) representation, where at least a part of the electronic degrees of freedom, e.g., the valence electrons of atoms included in the reaction, is treated explicitly. On the other hand, evaluating free energy profiles often needs an extensive sampling. Since the required step size for integrating the equations of motion in all-atom MD simulation is comparatively small (e.g., 0.5−1.0 fs), real time simulation of chemical reactions typically occurring on millisecond time scales is not feasible, or indeed efficient. From a practical point of view, due to the low Boltzmann weight of high energy states and their concomitant low probability of occupancy, even if the computational resources were available, performing realtime simulation of reactions would be a waste of resources. The sampling of such high energy states, which in the end determine the rate of reactions, can be enhanced by potential of mean force (PMF) techniques like umbrella integration,[1] steered molecular dynamics,[2] constrained dynamics,[3] metadynamics,[4] and adaptive biasing force.[5] The common feature of PMF methods is that they apply a bias in one or more collective variables (CVs), artificially enhancing the occupancy of high energy states. Because the applied bias is known, the unbiased free energy profile can be determined from the biased dynamics. This can be done either as a postprocessing step from the biased (but well sampled) free energy profile[1−3] or directly during the dynamics (on-the-fly) like in the case of history-dependent nonequilibrium sampling methods.[4,5]

The PMF and the corresponding free energy thus become functions of these CVs. It is generally accepted that in some sense the "ideal" reaction coordinate is the forward (or backward) configurational committor function, defined as follows.[6,7] Let us consider a system with a single reactant and a single product state which are disjoint subsets of configuration space. From each configuration one can initiate MD trajectories using velocities drawn from the appropriate Maxwell−Boltzmann distribution and sooner or later some of these trajectories will cross the boundary of either the reactant or the product states. Here in this article we are going to define the forward committor function as a function of configuration (i.e., all atomic positions) and the temperature as the proportion of the trajectories reaching the product state before the reactant state. (Note that there are other definitions of the committor function in the literature where it is a function of both the positions and momenta.[8,9]) By construction, the value of the committor for the configurations belonging to the a priori defined reactant and product states are 0 and 1, respectively. The committor function straightforwardly captures the transition state (TS) ensemble: a configuration is considered to be a TS if and only if its committor value is 0.5. In complex systems (e.g., reactions in the condensed phase) the TS ensemble is defined by not only the geometry of the reactant

Received: August 2, 2012
Revised: October 19, 2012
Published: December 6, 2012

dx.doi.org/10.1021/jp307648s

molecules: the committor is a function of all atomic coordinates of the system, including the solvent.

However, calculating the committor function in systems with a large number of degrees of freedom and especially with a QM model is unfeasible, let alone using it as a collective variable for a PMF technique. Therefore, in practice one usually resorts to using one or more empirical reaction coordinates. The choice of a good collective variable has several criteria. The most basic requirement is that the variable should accurately distinguish the reactant and product states and the transition should be continuous between the end points. Continuity is essential for evaluating forces on the atoms during the biased simulation when using a PMF technique. From the free energy profile as a function of the CVs, the stationary points can be identified: local minima correspond to reactant, product and intermediate states, while local maxima (and possible saddle points on multidimensional free energy surfaces) represent the possible transition states. Another feature of a good CV is to allow the determination of TS ensemble. While it is unlikely that the TS of an empirical CV will be identical to the true TS ensemble as defined above, the performance of a CV can be assessed by a posteriori committor analysis for the purported TS configurations.[10,11]

Often it is not easy to find a single CV that satisfies the above criteria and so additional techniques were introduced to improve the sampling and TS determination. One way is to apply multiple reaction coordinates instead of trying to find the most appropriate single one and evaluate the free energy surface in the space spanned by many CVs. For example, metadynamics is capable of calculating a multidimensional free energy surface efficiently.[4] Exploring more dimensions increases the simulation time and practically limits the maximum number of CVs to 4 or 5, so it is worthwhile to try and improve CVs so that fewer are needed. There have been recent attempts to develop single approximate reaction coordinates in a systematic way using a linear combination of several relevant CVs.[12] The method is straightforward if one has an appropriate set of CVs, from which the final reaction coordinate can be built up. Transition path sampling method[13] and the finite temperature string method[14] try to describe the reaction in a CV-free manner. However, their application is usually very time consuming for complex systems and too expensive for QM simulations.

Because of their relatively easy implementation and interpretation, the most commonly used CVs are geometrical coordinates like bond distance, angle, torsion, distance differences, coordination numbers, etc. However, finding the important degrees of freedom to be biased is very difficult and there is still no systematic way to explore them. The reason is that many reactions proceed by a concerted motion of not only the reactants but also the solvent molecules or the atoms of the environment around the reactants. This fact by itself is a strong argument against using simple geometrical variables, because they do not include these degrees of freedom.

A promising direction to construct good reaction coordinates that takes environmental effects also into account originates from Marcus.[15,16] Marcus theory was introduced for electron-transfer reactions and the solvent reorganization free energy was calculated based on the parabolas representing the two hypothetical diabatic electronic states. The representation of the electronic reactant and product states was later extended for reactions in solutions and enzymatic environment in the framework of empirical valence bond theory (EVB),[17,18] and the general reaction coordinate E_{gap} was also introduced as the difference between the predefined diabatic classical energy functions. The concept of E_{gap} is illustrated in Figure 1. The schematic diagram shows a hypothetical system having one relevant degree of freedom and two diabatic states that have approximately harmonic shape around their minima. The applicability of E_{gap} was later extended for ab initio QM/MM simulations[11] as well, and for a very simple symmetric nucleophilic reaction in water solution using the QM(PM3)/MM model it was demonstrated that in contrast to the generally applied geometrical coordinates, E_{gap} captures the transition state correctly and its committor function is also more reasonable.[11] For most elementary reactions in solutions and enzymatic environment the initial and final states have single and well-defined covalent connectivity and E_{gap} can be applied with confidence for these reactions. From now on we will refer the atom index-dependent connectivity as *topology*, the customary term used by the biomolecular simulation community.

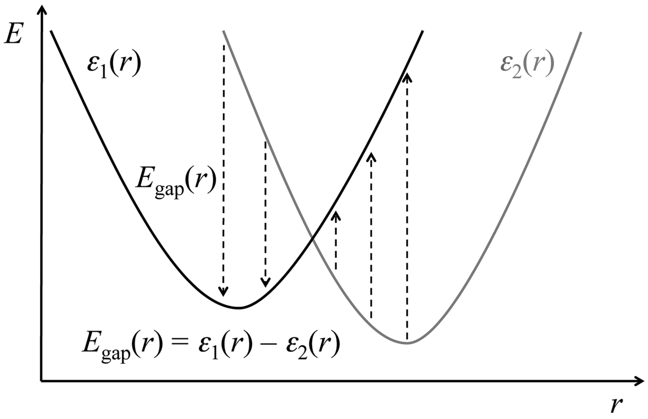


Figure 1. Schematic diagram of a hypothetical two-state system that has one relevant degree of freedom that distinguishes between the two states. The energy functions ε_1 and ε_2 correspond to the covalent topologies of the two states, and the dashed arrows represent the signed magnitude of E_{gap}.

Besides the single-state reactions, there are also reactions where there exist many possible chemically equivalent initial and/or final states that differ only topologically. The reason for the existence of multiple states is that topology definition represents only a single possible bond network of the system and it is not invariant with respect to the permutation of identical elements. In the case of most reactions this is not a problematic issue because the barrier of both the forward and reverse reactions for permutation of elements are far higher than the barriers under investigation and a single-topology description can be applied without observing spurious “side reactions” leading to other topologies. However, if the barrier of either direction is sufficiently low, the single-topology coordinate is not able to control the sampling anymore. Typical examples for such multistate reactions are the proton-transfer (PT) reactions, especially in water solution: self-ionization of water, deprotonation of weak acids and polyprotic acids, and protonation of weak bases. In such cases if a simple geometrical coordinate (e.g., distance or distance difference) or E_{gap} between two chosen states is used, undesired side reactions can take place during the simulation. For example, in the case of simulating the deprotonation of a weak acid, another proton of the acceptor water molecule can easily occupy the vacant position on the donor atom of the acid. There are several possible approaches to overcome this problem. One can use restraints or constraints for all other "interfering" bonds. Another possibility is to use special geometrical reaction coordinates that are invariant with respect to the different states, e.g., coordination number.[19] We investigate these solutions below, and also introduce new collective variables inspired by the two-state E_{gap}, the extended empirical valence bond model (extended-EVB)[20] and the multistate empirical valence bond model (MS-EVB).[21,22] Both methods were originally developed for investigating proton-transfer in water solution and their resultant reactive potential energy surface is the lowest eigenvalue of an EVB Hamiltonian that contains several, even chemically equivalent valence bond states to describe all reasonable possibilities. The MS-EVB method was successfully used to create several reactive classical potential energy surfaces for systems having multiple protonated states.[23] The new energy-based reaction coordinates presented here have the ability to describe multistate reactions on higher level of theory without applying any restraints or constraints. We also demonstrate that these collective variables capture the TS ensemble better than the geometrical coordinates.

■ METHODS

Model Systems. We investigated two proton-transfer reactions in water solution: the symmetric proton-transfer reaction between a hydronium ion and a water molecule, and water autoprotolysis:

$$H_3O_A{}^+(aq) + H_2O_B(aq) \rightarrow H_2O_A(aq) + H_3O_B{}^+(aq) \quad (1)$$

$$H_2O_A(aq) + H_2O_B(aq) \rightarrow O_AH^-(aq) + H_3O_B{}^+(aq) \quad (2)$$

The simulations were performed using the Amber 9 program,[24] while the PMF was calculated using the PMFlib library package[11,25] linked to Amber 9. The reactants were immersed into a cubic water box with a side of 26.5 Å (approximately 640 water molecules) using the Leap program of the Amber 9 program package. In the QM/MM simulations only the reactants were represented quantum mechanically and we used the PM3-PDDG semiempirical method.[26] Although in this work our purpose was to introduce new topologically invariant energy-based collective variables and compare them to geometrical ones, rather than analyzing the mechanisms and reproducing the experimental free energies, we have chosen the PM3-PDDG method because out of the available semiempirical methods in Amber 9 the barrier computed using this model for the hydronium−water reaction[27] was the closest to the experimental value. The classical water molecules were represented by the flexible TIP3P model.[28] van der Waals parameters of the oxygen and hydrogen atoms of the reactants were the same as those in the flexible TIP3P water model. Periodic boundary conditions were applied, MM−MM, QM−MM, and QM−QM long-range electrostatic interactions were calculated by the particle mesh Ewald method[29] with 9 Å direct interaction cutoff. The total charge of the system and the quantum region was +1 and 0 for the hydronium−water and autoprotolysis reactions, respectively, and the former was neutralized by applying a uniform background charge of −1 during the dynamics.

Simulation Protocol. *MM Simulations.* For both systems the relaxation procedure and the preparation of the initial configurations for the PMF calculations followed the same protocol described below. After 5000 steps of steepest descent minimization, the systems were gradually heated up to 300 K in 25 ps then held at this temperature for a further 25 ps. The density of the system was adjusted in a subsequent 200 ps long *NpT* simulation. After an additional 200 ps long equilibration in the canonical (*NVT*) ensemble, configurations were collected from a 1 ns long simulation at every 100 ps (10 configurations altogether). In all MM and QM/MM MD simulations 0.5 fs time step was used and the temperature was regulated by a Langevin thermostat[30] with a friction coefficient of 5 ps^{-1}.

QM/MM Simulations. Each of the 10 MM equilibrated configurations was equilibrated for 100 ps using the QM/MM model. The free energy profiles were then evaluated using the adaptive biasing force (ABF) technique[5] accelerated by the multiple walker method[31] using 10 walkers, one for each starting configuration. For each collective variable the samples were collected into 100 bins within a fixed range. The simulation length of each replica was 0.5 ns long with data exchange frequency of 50 fs between the replicas, giving an overall sampling time of 5 ns. When the convergence of the free energy profiles was investigated, the simulation length of each replica was 1 ns long (and the overall sampling time was 10 ns). For *Dis* and *DDis* coordinates the free energy profiles were calculated using umbrella integration (UI)[1,32] as well. These simulations were carried out in 100 windows in the same interval of reaction coordinates as was used in the ABF simulations. The windows were sampled sequentially, and each sampling was preceded by a 5 ps "transition period" in which the restraint position of the reaction coordinates were gradually displaced to its new value. In each sampling window, 50 ps long simulation was performed restraining the coordinates to the corresponding value by a force constant of 200 kcal mol^{-1} Å^{-2}. Samples were collected at intervals of 50 fs and the first 10% of the data were discarded and the derivative of the PMF in each window was calculated at the average value of the reaction coordinate. To prevent the reactants from getting too far from each other, a weak harmonic restraint with 12.5 kcal mol^{-1} Å^{-2} force constant was applied for the distance between the two oxygen atoms beyond 4.5 Å. For the committor analysis, from each of the 10 initial configurations 200 ps long constrained dynamics was performed keeping fixed the reaction coordinates at the value corresponding to the TS and altogether 1000 snapshots were collected. From each of these TS configurations, 100 unconstrained simulations were initiated with different random Maxwell−Boltzmann velocity distribution with initial temperature of 300 K. The length of these trajectories was relatively short (0.5 ps), but it was still sufficient for the system to slide either to the reactant or the product states. The probability density functions were calculated using the histogram-free method of Berg and Harris.[33]

Reaction Coordinates. We tested the widely used geometrical reaction coordinates based on distances and coordination numbers. Table 1 gives a summary, including the newly proposed reaction coordinates. This section gives the detailed definition of all reaction coordinates used, with our shorthand name in parentheses.

Geometrical Reaction Coordinates. The distance between the proton acceptor O_B atom and one of the hydrogen atoms initially bonded to O_A atom (*Dis*). The distance difference between a given hydrogen−O_A atoms and the same hydrogen−O_B atoms (*DDis*). The rational coordination number of the

Table 1. List of Collective Variables Tested in This Work[a]

collective variable	note
Dis	$\|\mathbf{r}_{\mathrm{H_AO_B}}\|$
DDis	$Dis(\mathbf{r}_{\mathrm{H_AO_A}}) - Dis(\mathbf{r}_{\mathrm{H_AO_B}})$
RCN	eq 3
DRCN	$RCN(\{\mathbf{r}_{\mathrm{HO_A}}\}) - RCN(\{\mathbf{r}_{\mathrm{HO_B}}\})$
EwMSED	eq 8
MeanMSED	$\lim_{\beta\to 0} EwMSED(\beta,\mathbf{x})$
MinMSED	$\lim_{\beta\to \infty} EwMSED(\beta,\mathbf{x})$

[a] For a detailed explanation of each reaction coordinate, see the text.

acceptor $\mathrm{O_B}$ atom (*RCN*). We also introduced a symmetrized version of *RCN*, the difference of the rational coordination number of the donor and acceptor oxygen atoms (*DRCN*). The functional form of the rational coordination number was

$$RCN(\{\mathbf{r}_{\mathrm{HO_B}}\}) = \sum_{i}^{\mathrm{H\,atoms}} \frac{1-\left(\frac{r_i}{r_0}\right)^{\alpha}}{1-\left(\frac{r_i}{r_0}\right)^{\beta}} \tag{3}$$

with $\alpha = 6$, $\beta = 18$, and reference distance $r_0 = 1.6$ Å.

Energy-Based Reaction Coordinates. To start building the energy-based reaction coordinates, let us consider a system with n_{R} chemically relevant reactant and n_{P} product topologies. We describe each of these states by MM potential energy functions according to their specific bond topology:

$$\varepsilon_i(\mathbf{x}) = \sum_{q}^{N_m} U^i_{\mathrm{Morse},q} + \sum_{j}^{N_b} U^i_{\mathrm{bond},j} + \sum_{k}^{N_a} U^i_{\mathrm{angle},k} + \sum_{l}^{N_t} U^i_{\mathrm{torsion},l} + \sum_{m\neq n}^{N_{nb}} (U^i_{\mathrm{el},mn} + U^i_{\mathrm{vdW},mn}) \tag{4}$$

where $\mathbf{x}$ denotes all atomic position vectors and the first term is the Morse potential for bonds to be broken or formed, the second, third, and fourth terms are the harmonic bond, angle and torsion energies for covalently bonded atoms, respectively, and the last term is the nonbonded electrostatic and van der Waals interactions.

In the case of $n_{\mathrm{R}} = 1$ and $n_{\mathrm{P}} = 1$ (single reactant− and product−topology system), a special reaction coordinate called energy gap (E_{gap}) can be defined as the difference between the force field-like energy functions of the initial (ε^{R}) and final (ε^{P}) states:[18]

$$E_{\mathrm{gap}}(\mathbf{x}) = \varepsilon^{\mathrm{R}}(\mathbf{x}) - \varepsilon^{\mathrm{P}}(\mathbf{x}) \tag{5}$$

The argument $\mathbf{x}$ is shown to emphasize that the energy functions using the reactant and product topologies are evaluated for the set of atomic positions. The power of E_{gap} lies in the fact that it is a general collective variable that includes all relevant degrees of freedom of the system weighted by the corresponding force field parameters. The effect of the wider environment of the reactants is taken into account mostly via the electrostatic and van der Waals terms.

The force field parameters of the classical states applied for the energy-based reaction coordinates were the following. To approximate the QM(PM3-PDDG) potential of the reactants, the bonded parameters for hydronium, water, and hydroxide were derived from gas-phase PM3-PDDG quantum calculations. Morse-type bond stretching and harmonic bending parameters were fitted based on gas-phase optimization. To avoid large harmonic angle energies/forces when bonds are being broken, the angle terms were coupled to the associated Morse potential of bonds. The charges of the atoms of reactants were the Mulliken charges obtained from geometrically optimized structures in the gas phase. van der Waals parameters of the flexible TIP3P model[28] were applied for reactant atoms as well. To avoid large repulsion between bonded atoms, the approximate exponential-6 form of the Lennard−Jones function was applied.[34] The surrounding solvent molecules were described by the flexible TIP3P model. For the nonbonded interactions between the reactants and solvent molecules, simple distance cutoff of 9 Å was used. However, the force field parameters of the energy-based reaction coordinates do not need to be set precisely, as we show below.

Since E_{gap} is a single topology coordinate, it is not suitable for simulating multistate reactions ($n_{\mathrm{R}} > 1$ or $n_{\mathrm{P}} > 1$). The basic problem is that ε^{R} and ε^{P} are MM potentials that depend on the indices in the list of atoms and not the chemical identity. Therefore, they are able to describe only specific valence states based on the topology and not the chemistry. In multistate reactions, however, there are many competing connectivities belonging to the same valence state. As the simulation progresses, different topological states become dominant. Unfortunately, even if these states belong to the same valence state, E_{gap} does not describe this invariance.

For example, in the symmetrical proton-transfer reaction between the hydronium and water, there are 20 chemically equivalent states: 10 “reactant” states when $\mathrm{O_A}$ is the hydronium oxygen (out of 5 hydrogen atoms the all possible variations of 3 bonded to this oxygen atom is $\binom{5}{3}$) and 10 “product” states when $\mathrm{O_B}$ is the central oxygen of the hydronium. Similarly, for the water autoprotolysis reaction there are 6 reactant and 8 product states. The most obvious extension of the two-state E_{gap} is the linear combination of all states with the same absolute value of the coefficients, whose sign is positive for reactant and negative for product states:

$$MeanMSED(\mathbf{x}) = \frac{1}{n_{\mathrm{R}}}\sum_{i}^{n_{\mathrm{R}}} \varepsilon_i^{\mathrm{R}}(\mathbf{x}) - \frac{1}{n_{\mathrm{P}}}\sum_{j}^{n_{\mathrm{P}}} \varepsilon_j^{\mathrm{P}}(\mathbf{x}) \tag{6}$$

We call this coordinate “mean multistate energy difference” (*MeanMSED*) as it includes the energy of all states. As will be shown below, *MeanMSED* is not able to distinguish well between general initial and final states. Another candidate is the “minimum energy multistate energy difference” (*MinMSED*) defined as

$$MinMSED(\mathbf{x}) = \min\{\varepsilon_i^{\mathrm{R}}(\mathbf{x})\}_i^{n_{\mathrm{R}}} - \min\{\varepsilon_j^{\mathrm{P}}(\mathbf{x})\}_j^{n_{\mathrm{P}}} \tag{7}$$

For each configuration during the dynamics, *MinMSED* checks the actual energy value of the possible reactant and product states and selects the lowest ones from both sets. In principle, all possible topologies should be included, but in practice simple geometric criteria can be safely used to only check a few configurations. This results in a single topology, which is dynamically adapted to the chemically most sensible connectivity, instead of being a function of fixed atom indices. Although *MinMSED* defined in eq 7 is a continuous function of the coordinates, its derivatives are not. This can cause problems in some PMF methods (e.g., if using constrained dynamics). For these situations we introduce the energy-weighted version (energy-weighted multistate energy difference, *EwMSED*), which is a Boltzmann-weighted linear combination of all states and so its first derivatives are also continuous functions

$$EwMSED(\beta, \mathbf{x}) = \frac{\sum_{i}^{n_{\mathrm{R}}} \mathrm{e}^{-\beta \varepsilon_i^{\mathrm{R}}(\mathbf{x})} \varepsilon_i^{\mathrm{R}}(\mathbf{x})}{\sum_{k}^{n_{\mathrm{R}}} \mathrm{e}^{-\beta \varepsilon_k^{\mathrm{R}}(\mathbf{x})}} - \frac{\sum_{j}^{n_{\mathrm{P}}} \mathrm{e}^{-\beta \varepsilon_j^{\mathrm{P}}(\mathbf{x})} \varepsilon_j^{\mathrm{P}}(\mathbf{x})}{\sum_{l}^{n_{\mathrm{P}}} \mathrm{e}^{-\beta \varepsilon_l^{\mathrm{P}}(\mathbf{x})}} \tag{8}$$

where β is a fictitious inverse temperature. It is easy to see that as β goes to zero or infinity, the *EwMSED* becomes *MeanMSED* and *MinMSED*, respectively. Again, in principle, the sum is to be taken over all possible topologies; however, only a finite number needs to be computed because simple geometric criteria can be used to exclude those topologies which would yield such a high energy that the associated Boltzmann weight is negligible.

■ RESULTS AND DISCUSSION

For each collective variable we calculated the free energy profiles at 300 K using multiple walker accelerated adaptive biasing force (ABF) molecular dynamics simulations. The simulations were started from pre-equilibrated reactant state configurations, in which there is a hydrogen bond between one of the hydrogen atoms of the donor oxygen atom (H_A) and the acceptor oxygen atom (O_B). Although all the possible reactant states (in which O_A is the hydronium oxygen atom or donor water oxygen atom, for the hydronium−water and the water−water reactions, respectively) are chemically equivalent, *Dis* and *DDis* are not invariant to these states as they depend rather on the atomic indices in the topology file than the chemical elements. To examine the effect of this on the free energy profile, two separate simulations were carried out for the hydronium−water case, which differed in the way the atoms in the initial configurations were indexed: in the first case the H and O atoms referred to in the reaction coordinate were the ones also participating in the hydrogen bond, while in the second case, all possible assignments were generated, one for each of the initial configurations of the multiple walker ABF scheme. For the latter and more subtle case, we also investigated the convergence of the free energy profiles. Since the ABF method estimates the derivative of the free energy on-the-fly during the simulation, it is more straightforward to compare the derivatives rather than the integrated profiles. Therefore, we defined the following error function that estimates the overall deviation of the free energy derivatives at time t compared to the final one obtained at the end of the simulation

$$\delta(t, t_{\mathrm{max}}) = \int_{\xi_{\mathrm{min}}}^{\xi_{\mathrm{max}}} \left| \frac{\partial F(t)}{\partial \xi} - \frac{\partial F(t_{\mathrm{max}})}{\partial \xi} \right| \mathrm{d}\xi \tag{9}$$

where ξ is the collective variable and $[\partial F(t)]/[\partial \xi]$ and $[\partial F(t_{\mathrm{max}})]/[\partial \xi]$ are the free energy derivatives at time t and at the end of the simulation, respectively. The errors of the PMF gradients are shown in Figure 2 for the different collective variables. In the beginning of the dynamics, *DRCN* converges the most rapidly but after 100 ps there is no significant difference between the convergence of *RCN*, *DRCN*, and *MinMSED* and the final profiles of these coordinates can be considered converged with an overall error of less than 0.1 kcal mol^{-1}. In contrast to these coordinates, *Dis* and *DDis* have a very slow convergence behavior. This is in accord with previous results[11] that revealed an insufficient convergence of free energy profiles of these geometrical coordinates due to the large hysteresis using another on-the-fly PMF method (Metadynamics). Similarly to those observations, we found that further simulations result in some changes but do not alter significantly the shape of the entire profiles we are going to characterize. Nevertheless, this nonconvergence could be the failure of the applied on-the-fly PMF methods for these coordinates as the number of walkers is finite, and in this case ABF does not necessarily converge.[35] To provide evidence that the nonconvergence is due to the choice of collective variables, in addition to ABF we evaluated the free energy profiles for *Dis* and *DDis* using the umbrella sampling (US) technique and starting the simulations from the 10 topologically different initial configurations.

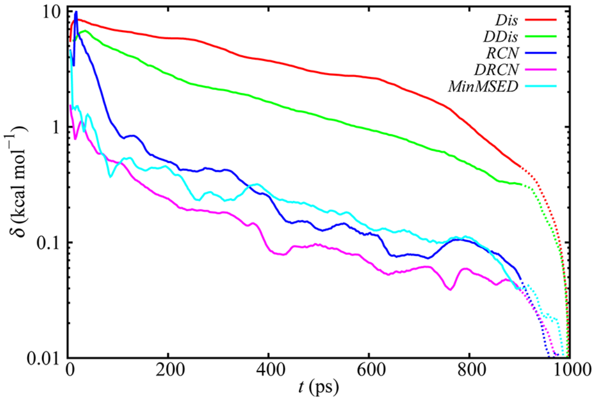


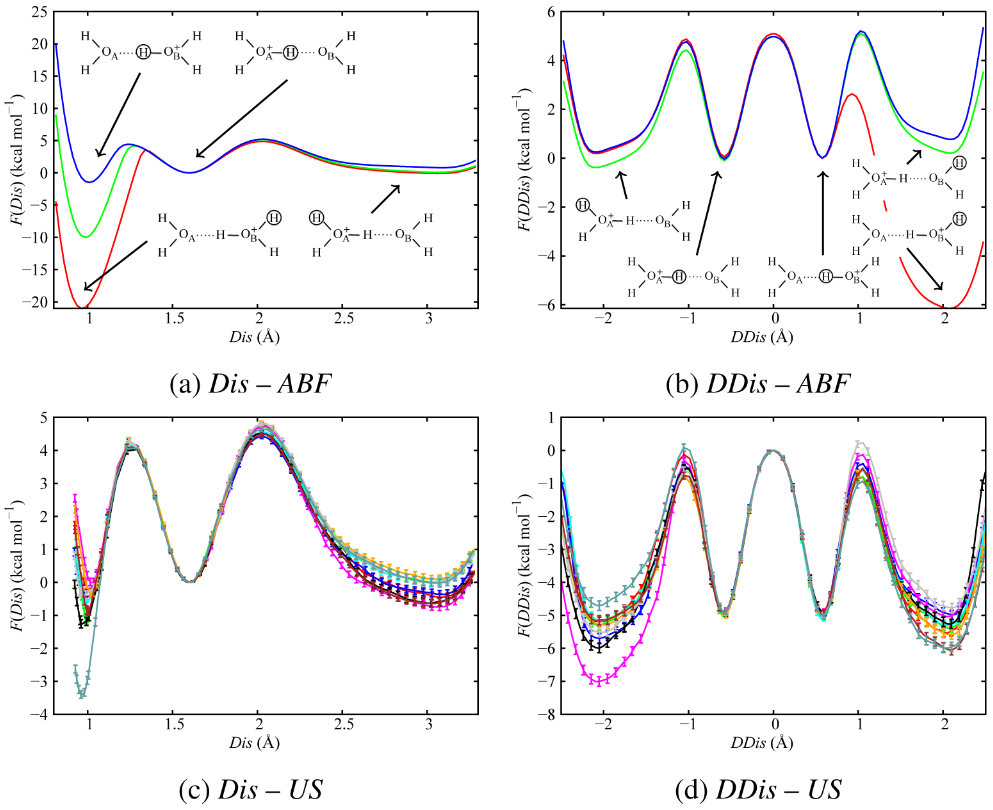
Figure 2. Integrated error of free energy derivatives of the investigated collective variables as function of simulation time. The last 80 ps are shown in dotted style to indicate that the rapid decrease in apparent error is an artifact of using the profiles from the end of the simulations as references.

Topologically Noninvariant Geometrical Coordinates.

Figure 3a and 3b show the ABF profiles using *Dis* and *DDis* for the hydronium−water reaction. The red and green profiles correspond to using different and identical topologies for the initial configuration, respectively. In the case of *Dis*, sampling of all possible initial topologies results in a dramatically lower minimum for the reactant state, which is an artifact of including all the topologies that do not match the reaction coordinate. Even using identical topologies does not make the profile symmetric, because the barrier for topological rearrangements is not very large, and the sampling does visit some of these configurations. The effect is much smaller but still significant for *DDis*, this time particularly near the product state minimum.

For comparison the free energy profiles of this reaction were obtained using umbrella sampling simulations as well. The free energy profiles obtained by initiating the sampling from different topological states are shown in Figure 3c and 3d, for *Dis* and *DDis*, respectively. For both coordinates there is an interval where the profiles match pretty well ([1.1,1.8] and [−0.8,0.8] Å for *Dis* and *DDis*, respectively). However, out of these regions the curves start diverging, indicating a similar sampling problem as was observed for the ABF curves, although the deviation is not as large as for the ABF profiles.

The usual remedy to prevent the undesirable side reactions or rotation of the reactants is to impose stiff restraints on O−H bonds that are *not* part of the reaction coordinate. The blue profile on the figure shows the result of using a one-sided harmonic (p = 2) restraint beyond r_0 = 1.0 Å and spring constant k = 500 kcal mol^{-1} Å^{-2}



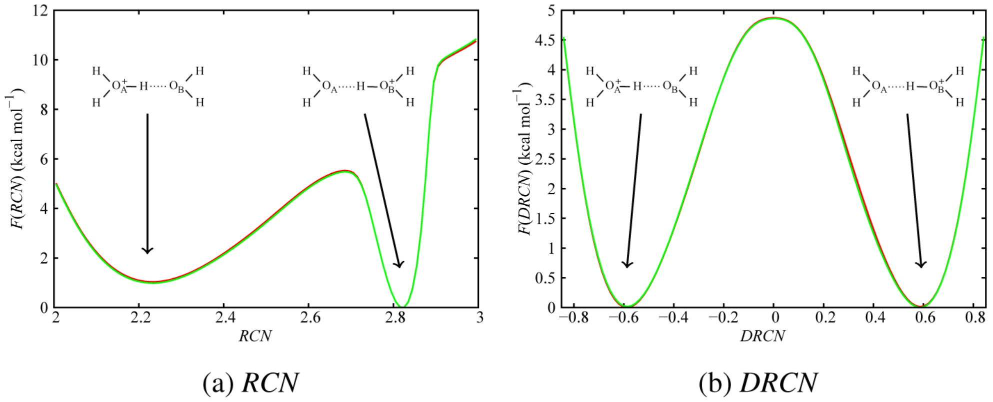
(a) *Dis – ABF*

(b) *DDis – ABF*

(c) *Dis – US*

(d) *DDis – US*

Figure 3. Free energy profiles obtained from ABF (panels a and b) and US (panels c and d) simulations using the *Dis* and *DDis* collective variables for the proton-transfer reaction of the hydronium−water system. In the case of ABF simulations, different curves correspond to different ways of handling the permutation problem of the reaction: different initial topologies (red line), identical initial topologies (green line), and identical initial topologies using restraints during the dynamics (blue line). Dominant topological structures of the reactants at the minima are shown (the hydrogen atom involved in the collective variables is circled). For US simulations, different curves correspond to simulations initiated from different initial topologies.



(a) *RCN*

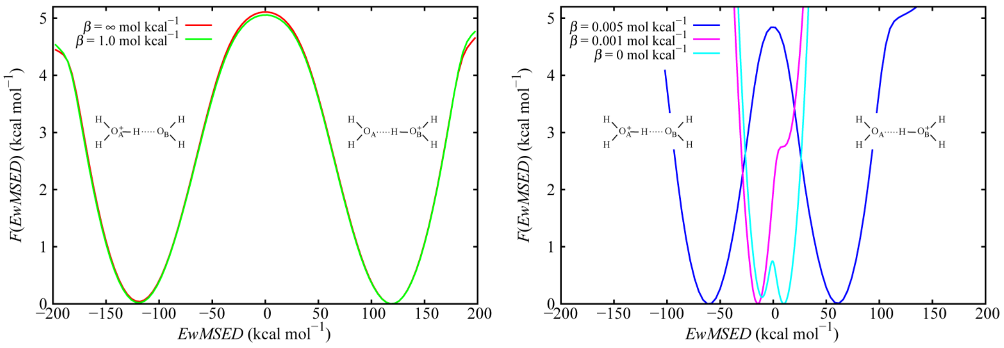
(b) *DRCN*

Figure 4. ABF free energy profiles using the *RCN* and *DRCN* collective variables for proton-transfer reaction of the hydronium−water system. Different curves correspond to different ways of handling the permutation problem of the reaction: different initial topologies (red line) and identical initial topologies (green line). Dominant topological structures of the reactants at the minima are shown (as all hydrogen atoms are involved in these variables, no marking was used).

$$V_{\text{restraint}} = \begin{cases} k(r - r_0)^p & \text{if } r > r_0 \\ 0 & \text{else} \end{cases} \tag{10}$$

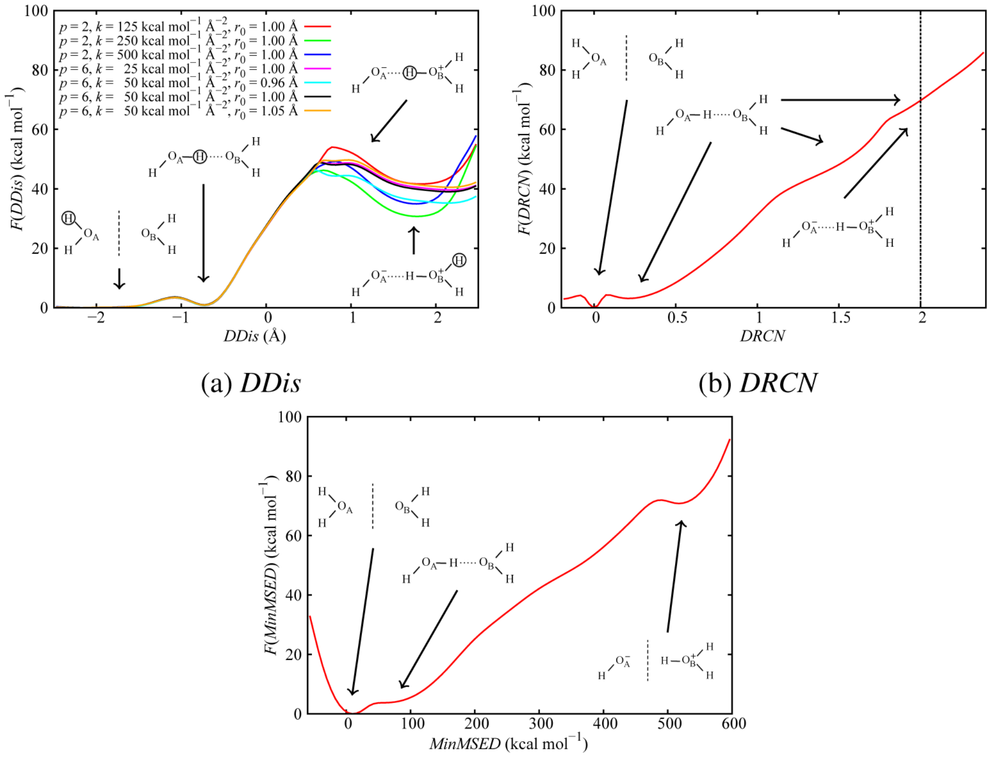
As can be expected, this leads to a considerable improvement for this symmetric reaction, especially when using the symmetric *DDis* reaction coordinate. In order to investigate the stability of this approach, we chose a more challenging system, the autoprotolysis of water, and varied the parameters of the restraints. Note again that the restraints in question are not those that enforce the value of reaction coordinate but the additional restraints needed to prevent the side reactions.

The profiles as functions of *DDis* with different parameters are shown in Figure 6a. Each profile has three well-defined minima: the first wide minimum in the range of [−2.5,−1.5] Å corresponds to the separately solvated reactants, and the second one represents the hydrogen-bonded complex, while the third minimum is the hydrogen-bonded products. The



(a) *EwMSED* with large *β* values (b) *EwMSED* with small *β* values

Figure 5. ABF free energy profiles using the *EwMSED* collective variable for the proton-transfer reaction of the hydronium−water system using various *β* values. Note that, when *β* = ∞, *EwMSED* corresponds to *MinMSED*.



(a) *DDis* (b) *DRCN*

(c) *MinMSED*

Figure 6. ABF free energy profiles using the *DDis*, *DRCN*, and *MinMSED* collective variables for water autoprotolysis reaction. Dominant topological structures of the reactants at the minima are shown. In the case of *DRCN* coordinate a vertical dashed line at *DRCN* = 2 indicates the region where the state of the separated products is expected.

depth of the third minimum (and the height of the barrier) strongly depends on the chosen parameters. For the widely used one-sided harmonic restraint (p = 2) we have significantly different profiles as we change the force constant between 125 and 500 kcal mol^{-1} Å^{-2}. Profiles generated with the much tighter p = 6 restraint are less sensitive to the force constant; however, placing the position of the restraint potential again has significant impact. An additional effect of this type of very stiff restraint is the appearance of an artificial small minimum close to the transition state.

Topologically Invariant Geometrical Coordinates. It is thus clear that describing multistate reactions with topologically noninvariant coordinates with or without extra restraints is problematic. Rather, the invariance of the reactants and products with respect to relabeling of atoms should be built into the reaction coordinate. For proton-transfer reactions a widely used reaction coordinate is the *rational coordination number* that satisfies this criterion and is given in eq 3. Following the analogy of the relation between *Dis* and *DDis* we investigated the free energy curves of both *RCN* and the symmetrized version *DRCN* for the hydronium−water reaction (Figure 4). The first observation that can be made is that in the case of both *RCN* and *DRCN* the profiles are almost identical, regardless of how the initial configurations are labeled. This follows from the invariance property of the collective variables. It can also be seen that *RCN* does not provide zero reaction free energy (ΔF ∼ 1.0 kcal mol^{-1}), indicating that it is not an appropriate coordinate to describe the proton-transfer reaction.

However, *DRCN* leads to a symmetric profile (ΔF = 0.0 kcal mol^{-1}) with an activation barrier of $\Delta F^{\ddagger}$ = 4.9 kcal mol^{-1}.

We also investigated the free energy profile of the nonsymmetric proton-transfer reaction using *DRCN* (Figure 6b). In this case the reaction coordinate's value at the reactant state is 0 while in the product state it should be around 2. In contrast to *DDis*, this curve does not have any product-like minimum. An analysis of the trajectories reveals that, although there are configurations where the hydronium and hydroxide ions are formed, the contact ion pair never gets separated during the simulations, and even for a *DRCN* value near 2, there are configurations in which the products have not formed. This lack of clear separation of reactants and products results from the reaction coordinate not being able to clearly distinguish between breaking and forming covalent bonds, and slightly shorter than normal hydrogen bonds.

Energy-Based Coordinates. We show the free energy profiles generated using the topologically invariant *EwMSED* reaction coordinate for a range of β values in Figure 5. As β increases, the profiles converge to a symmetrical curve with $\Delta F^{\ddagger}$ = 5.1 kcal mol^{-1}. For small β values, the reaction coordinate mixes together all possible topologies using similar weights including topologies that represent a connectivity far from the current one, and so this leads to unphysical profiles. In the limit of large β, *EwMSED* becomes *MinMSED* that selects the state with the lowest energy from all possible reactant states and similarly the one from the product set. Another way to express this is that the reaction coordinate is the best possible E_{gap} value given the current atomic positions.

As Figure 6c shows, *MinMSED* is able to describe correctly the water autoprotolysis reaction as well. The profile has only two minima, the separated initial and final states. It is worth noting that the prereactive complex can also be identified in the profile as a plateau around *MinMSED* = 60−70 kcal mol^{-1} and the relative free energy difference between the separated minima is 71.9 kcal mol^{-1}. This value is at least 30 kcal mol^{-1} higher than the free energy differences in the case of *DDis*. However, here we emphasize again that the final minimum of the *DDis* curves does not correspond to the product state but rather can be considered as an artifact of the applied restraints.

Force Field Parameter Dependence of the Energy-Based Coordinates. As we described in the Methods section, the model parameters for the energy-based reaction coordinates were developed by using MM parameters for reactants derived from the gas-phase PM3-PDDG surface, with some modifications to allow bonds to break and form smoothly. However, since the force field is only used to build a reaction coordinate, and not as a potential energy surface, it is natural to expect that the free energy profiles would not depend strongly on the force field parameters. To test this, we calculated the free energy profile for the symmetric reaction with eight different parameter sets, each of which differed substantially from the original. In the first two sets we modified only the bonded parameters, while the charges were kept from the original parameter set: in set 1 the bond and angle force constants were doubled and the corresponding equilibrium distances and angles were increased by 10% compared to the original ones, while in set 2 the force constants were halved and the equilibrium constants were decreased by 10%. In the second, two sets only the charges were changed: the charge polarization of the reactants was increased and decreased by 20% for set 3 and set 4, respectively. For sets 5−8 we combined the different modifications from sets 1 to 4. The descriptions of the different sets are shown in Table 2. Figure 7 shows the free energy profiles generated with the various force field parameter sets, as well as that corresponding to the original parameters. To aid comparison, the free energy profiles were normalized such that in each case the reactant state was at a reaction coordinate value of −1.0 while product state was at +1.0. The normalized profiles look very similar, with the maximum difference between the reaction free energies and activation barriers being 0.17 and 0.18 kcal mol^{-1}, respectively. This demonstrates that the force field parameters do not have to be fine-tuned in order to construct a good reaction coordinate.

Table 2. List of Investigated Sets with Different Force Field Parameters

	bonded parameters		
no. of set	k_{bond}, k_{angle}	r_0, φ_0	charge polarization
1	+100%	+10%	original
2	−50%	−10%	original
3	original	original	+20%
4	original	original	−20%
5	+100%	+10%	+20%
6	−50%	−10%	+20%
7	+100%	+10%	−20%
8	−50%	−10%	−20%

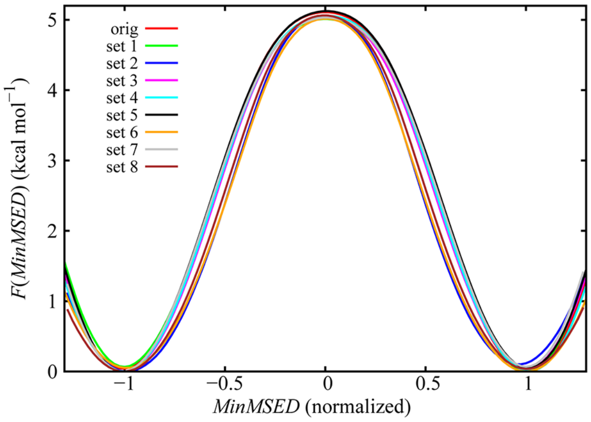


Figure 7. ABF free energy profiles of normalized *MinMSED* using different force field parameters. For the description of the different sets see text and Table 2.

Comparing TS Ensembles. In order to characterize how well the transition state ensemble is captured by the different reaction coordinates, we carried out committor analysis[10] for *DDis*, *DRCN*, and *MinMSED* for the hydronium−water reaction. The calculated probability density functions of the trajectories committed to the product state (P_{PS}) are presented in Figure 8. For the ideal reaction coordinate (the committor function) this distribution would be a delta function at 0.5. As expected and also shown in a previous work,[11] *DDis* is not able to capture the TS ensemble at all and leads to an asymmetric distribution with multiple local maxima (none of them at P_{PS} = 0.5). The profile of *DRCN* has a single maximum approximately at 0.5, but is relatively flat compared to that of *MinMSED*.

Both *DRCN* and *MinMSED* are topologically invariant coordinates, and therefore the difference between the committor results is due to the ability of the latter in taking the environmental effects into account. To show this, we performed the committor analysis of *MinMSED* but this time

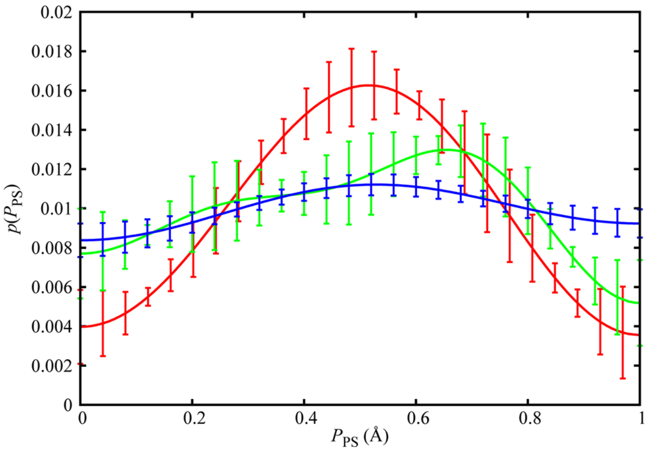


Figure 8. Committor analysis of the *DDis* (green line), *DRCN* (blue line), and *MinMSED* (red line) collective variables for the proton-transfer reaction of the hydronium−water system.

using a zero cutoff distance for the solute−solvent electrostatic and van der Waals interactions in the force field, so that the reaction coordinate only takes the geometry of the reactants into account and completely neglects the solvent. Figure 9 shows the comparison between this and the original *MinMSED*. Although the reactants-only *MinMSED* gives the same activation barrier and reaction free energy difference as the original, it has a flat committor probability density function similar to that of *DRCN*.

CONCLUSION

In summary, in this article we introduced a new energy-based collective variable that can be used to bias the dynamics when using potential of mean force techniques. It is a generalized form of the E_{gap} reaction coordinate (the energy difference between the reactant and product valence states) and can therefore be used for simulations on arbitrary potential energy surfaces, including QM/MM and fully quantum mechanical models. In the case of reactions that have multiple chemically equivalent initial or final configurations, neither traditional geometrical reaction coordinates such as bond lengths nor E_{gap} is applicable, because undesired "side reactions" occur. We demonstrated that restraining or constraining bonds to avoid these side reactions is not a good strategy because the free energy profiles thus obtained will strongly depend on the force constant and the position of the restraint/constraint. The key idea of the generalization of the E_{gap} reaction coordinate for such multistage reactions is similar to how the EVB[17,18] method is generalized to multiple states in the extended-EVB[20] and MS-EVB scheme.[21] We add up the energies of all possible competing valence states instead of just two, but weight them according to their energy. The resulting reaction coordinate is the energy-weighted multistate energy difference (*EwMSED*).

In practice it is not necessary to calculate the energy of all possible chemically equivalent states, because all but a handful have negligible weight and can be excluded based on simple geometric criteria. For those cases when there is no requirement for the derivative of the reaction coordinate to be continuous, a simpler limit of *EwMSED* can be used, in which the lowest energy state is selected from among the competing reactant and product states with the weight of all other configurations set to zero, leading to the minimum energy multistate energy difference (*MinMSED*) reaction coordinate. We demonstrated that similarly to E_{gap}, the variants of the new reaction coordinates also possess a good sampling behavior. Although there are simple geometrical reaction coordinates that can be also made invariant to permutation of identical species, we showed using committor analysis that the energy-based reaction coordinate captures the transition state ensemble significantly better. By varying the cutoff used for the nonbonded interactions included in the energy model of our reaction coordinate, this good feature was shown to be linked to the inclusion of the wider environment of the reacting atoms in the reaction coordinate. This underscores the importance of the concerted motion of atoms in chemical reaction pathways.

AUTHOR INFORMATION

Corresponding Author

*E-mail: lam81@cam.ac.uk.

Notes

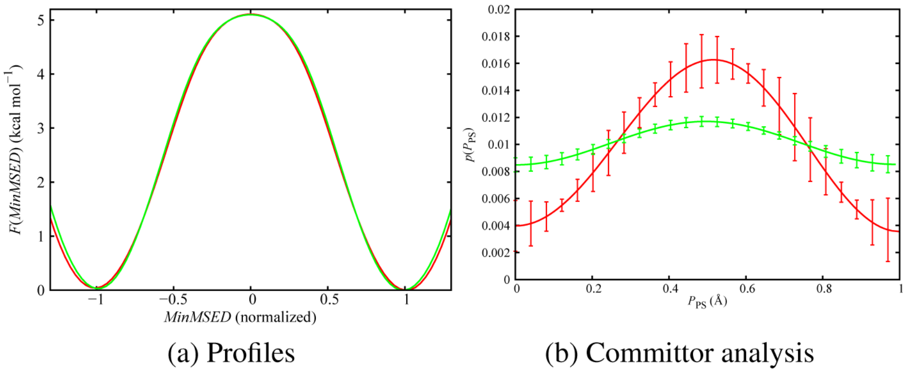
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Noam Bernstein for insightful discussions. L.M. was supported by EU-FP7-NMP grant 229205 ADGLASS.

REFERENCES

(1) Kastner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.
(2) Isralewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.
(3) Carter, E.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472−477.



(a) Profiles

(b) Committor analysis

Figure 9. ABF free energy profiles and committor analysis using *MinMSED* with different nonbonded cutoff distances: r_{cutoff} = 9 Å (red line) and r_{cutoff} = 0 Å (green line). The zero value cutoff corresponds to an *MinMSED* that depends only on the coordinates of reactants.

(4) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(5) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(6) Metzner, P.; Schutte, C.; Vanden-Eijnden, E. *J. Chem. Phys.* **2006**, *125*, 084110.
(7) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. *J. Chem. Phys.* **2008**, *129*, 174102.
(8) Vanden-Eijnden, E. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer-Verlag: Berlin Heidelberg, 2006; Vol. *1*, pp 453−493.
(9) E, W.; Vanden-Eijnden, E. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.
(10) Geissler, P. L.; Dellago, C.; Chandler, D. *J. Phys. Chem. B* **1999**, *103*, 3706−3710.
(11) Mones, L.; Kulhánek, P.; Simon, I.; Laio, A.; Fuxreiter, M. *J. Phys. Chem. B* **2009**, *113*, 7867−7873.
(12) Rosta, E.; Woodcock, H. L.; Brooks, B. R.; Hummer, G. *J. Comput. Chem.* **2009**, *30*, 1634−1641.
(13) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964−1977.
(14) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(15) Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.
(16) Marcus, R. A. *J. Chem. Phys.* **1965**, *43*, 679−701.
(17) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(18) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218−2224.
(19) Sprik, M. *Faraday Discuss.* **1998**, *110*, 437−445.
(20) Vuilleumier, R.; Borgis, D. *J. Mol. Struct.* **1997**, *436*−*437*, 555−565.
(21) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547−5551.
(22) Čuma, M.; Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. A* **2001**, *105*, 2814−2823.
(23) Swanson, J. M. J.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4300−4314.
(24) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; et al. *AMBER 9*; University of California: San Francisco, CA, 2006.
(25) Kulhánek, P.; Fuxreiter, M.; Koča, J.; Mones, L.; Střelcová, Z.; Petřek, M. PMFLib: A Toolkit for Free Energy Calculations. https://lcc.ncbr.muni.cz/whitezone/development/pmflib.
(26) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.
(27) Luz, Z.; Meiboom, S. *J. Am. Chem. Soc.* **1964**, *86*, 4768−4769.
(28) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(29) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(30) Adelman, S. A.; Doll, J. D. *J. Chem. Phys.* **1976**, *64*, 2375−2388.
(31) Lelievre, T.; Rousset, M.; Stoltz, G. *J. Chem. Phys.* **2007**, *126*, 134111.
(32) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187−199.
(33) Berg, B. A.; Harris, R. C. *Comput. Phys. Commun.* **2008**, *179*, 443−448.
(34) Lim, T.-C. *J. Math. Chem.* **2003**, *33*, 279−285.
(35) Lelièvre, T.; Rousset, M.; Stoltz, G. *Nonlinearity* **2008**, *21*, 1155.