JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Toward an Automatic Determination of Enzymatic Reaction Mechanisms and Their Activation Free Energies

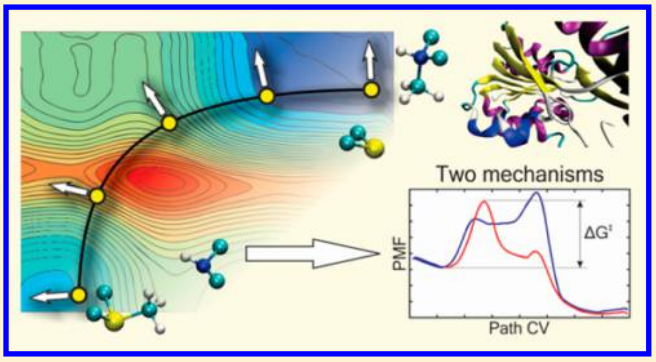
Kirill Zinovjev,[‡,§] J. Javier Ruiz-Pernía,[§] and Iñaki Tuñón*[,‡]

[‡]Departament de Química Física, Universitat de València, 46100 Burjassot, Spain

[§]Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain

Supporting Information

ABSTRACT: We present a combination of the string method and a path collective variable for the exploration of the free energy surface associated to a chemical reaction in condensed environments. The on-the-fly string method is employed to find the minimum free energy paths on a multidimensional free energy surface defined in terms of interatomic distances, which is a convenient selection to study bond forming/breaking processes. Once the paths have been determined, a reaction coordinate is defined as a measure of the advance of the system along these paths. This reaction coordinate can be then used to trace the reaction Potential of Mean Force from which the activation free energy can be obtained. This combination of methodologies has been here applied to the study, by means of Quantum Mechanics/Molecular Mechanics simulations, of the reaction catalyzed by guanidinoacetate methyltransferase. This enzyme catalyzes the methylation of guanidinoacetate by S-adenosyl-L-methionine, a reaction that involves a methyl transfer and a proton transfer and for which different reaction mechanisms have been proposed.



1. INTRODUCTION

The correct description of enzymatic reactivity is a challenge for computational chemistry. In contrast to gas phase reactivity of small molecules, where the size of the system allows to perform high-level *ab initio* or Density Functional calculations, yielding results within chemical accuracy;[1,2] enzymes are huge systems, making unaffordable the explicit treatment of all the electrons of the system and thus limiting the suite of applicable techniques to more approximate ones, such as hybrid quantum mechanics/molecular mechaniscs (QM/MM) Hamiltonians.[3−5] On the other hand, gas phase reactions can be often described by means of the analysis of a Potential Energy Surface (PES) function of a few degrees of freedom and by the characterization of the stationary structures located on this PES (reactants, products, intermediates, and transition structures). However, the study of processes in condensed environments, as enzymatic reactions, must be carried out sampling a myriad of conformations that contribute to the observed macroscopic properties. Finite-temperature simulations are needed to sample the corresponding thermodynamic ensemble, and from these simulations, it is possible to reconstruct the free energy surface (FES) as a function of relevant coordinates that control the chemical process under study. The FES is related to the probability of finding the system with the desired values of the selected coordinates and then can be related to equilibrium and kinetic properties of the system such as the equilibrium and rate constants.[6,7]

An additional difficulty for the description of enzymatic reactions is the high number of degrees of freedom and, therefore, the complexity of the free energy surface. A single geometric reaction coordinate controlling the whole process from reactants to products can be guessed only in simple cases, such as bond breaking/forming or transfer reactions in which a distance or combination of distances can be used to trace the Potential of Mean Force (PMF).[8−11] In more complicated cases, when several chemical bonds are broken and formed simultaneously, a simple selection will probably fail, resulting in hysteresis problems. In these cases, the explicit construction of 2D or even 3D free energy surfaces are needed for a satisfactory description of the reaction.[10,12,13] This gives rise to the so-called 'curse of dimensionality'—the computational cost rapidly increases with the number of reaction coordinates used. While techniques such as metadynamics,[14] Replica Exchange Umbrella Sampling,[15] and Temperature Accelerated Molecular Dynamics[16] can significantly reduce the time needed to construct the free energy surface in comparison to conventional Umbrella Sampling method,[17] the computational cost still scales exponentially, thus prohibiting the use of more than just a few reaction coordinates.

A way to overcome this issue comes from the fact that one is not usually interested in reconstructing the whole free energy hypersurface but to localize the most probable transition

Received: February 28, 2013
Published: June 17, 2013

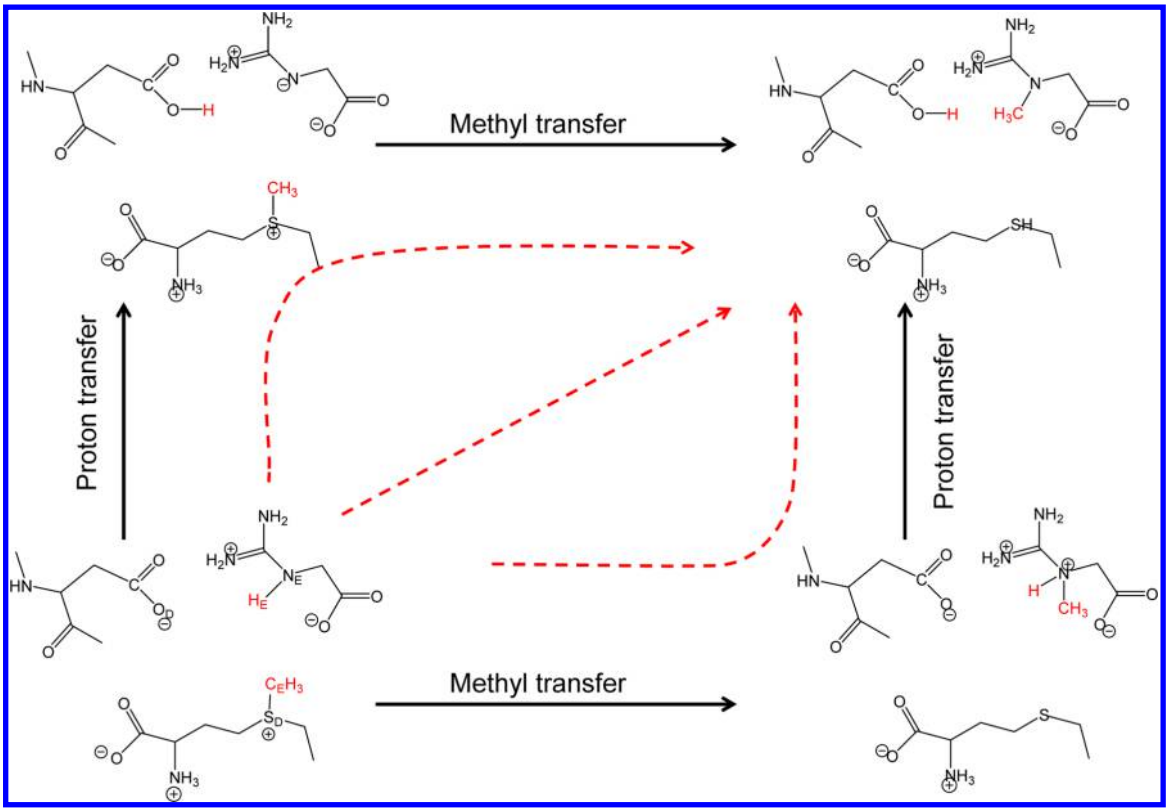


Figure 1. Schematic representation of possible reaction mechanisms catalyzed by GAMT. The arrows indicate a concerted path (diagonal) and two possible stepwise paths; the upper arrow corresponds to the methyl transfer preceding the proton transfer and the lower one to the path in which the proton transfer precedes the methyl transfer.

pathways for the process of interest. Large regions of such a surface are energetically unfavorable and, therefore, almost never visited in the unbiased ensemble. In fact, in most cases it can be assumed that there is a narrow tube around the minimum free energy path (MFEP) where most of reactive trajectories take place—the so-called "reaction tube".[18] Since the system rarely leaves the reaction tube, the correct energetics can be obtained ignoring other regions of the free energy surface. Therefore, the problem of reconstruction of multi-dimensional FES is substituted by the problem of finding a one-dimensional MFEP and sampling the free energy around it.

There are several computational strategies to find the MFEP in complex systems. A path-metadynamics method for determining the average transition path, that under the single, well-defined reaction tube approximation is close to the MFEP, was recently presented.[19] This method uses the sampling along planes orthogonal to the path to guide its evolution and to simultaneously reconstruct the free energy profile along this coordinate with metadynamics.[14] Another approach published by Bohner et al.[20] uses Umbrella Integration[21] to calculate the Hessian of the free energy. This is used to perform Newton–Raphson optimization to reach the transition state and then the MFEP is found by following the gradient of the free energy.

A simple, mathematically rigorous and rapid approach to obtain the MFEP without explicitly having the FES is the string method proposed by Vanden-Eijnden et al.[22,23] The main idea of the string method is to use several replicas of the system that correspond to different, equidistant points between reactants and products basins in a given reaction coordinate space. When these replicas are allowed to relax, maintaining the equidistant parametrization, this 'string' eventually arrives to the MFEP. Once the string is converged the free energy profile along the given path can be calculated without any additional simulations.[22,23] One of the greatest advantages of the string method and related techniques is that the computational cost is practically independent of the number of dimensions used, making affordable the study of very complicated processes.[24,25] Two main versions of the string method exist, the zero temperature string method[22] (and its on-the-fly version[23]) and the finite temperature string method,[18] aimed to obtain the MFEP and the transition principal curve, respectively. The main difference between these methods is the way that the system is maintained at a given position along the path. In the zero temperature version a multidimensional harmonic constraint is used, so that each replica never goes far from some point on the string. The update of the path is done by following the approximated gradient of the free energy at these points. In case of the finite temperature version the system is maintained in one of the Voronoi cells, defined by a series of points along the path. In this way, the system is free to move in directions orthogonal to the path and the principal curve can be recovered. The main advantage of the principal curve is that it is smooth even if the underlying free energy surface is rugged. In this method, the sampling within the cells depends on a delicate balance between their size and the free energy slope. Large barrier processes may require small cells where the trajectory of the system should be continuously perturbed to keep it within the limits of the cells. Obtaining good sampling with short simulation times is definitely a critical point for the study of enzymatic reactions where expensive QM/MM potentials must be employed. Since in the case of enzymatic reactions important differences between the principal curve and

the MFEP are not expected, especially when adequate collective variables are employed, the zero temperature string method has been selected to obtain the path. It must be also considered that, up to now, the zero temperature version was applied to a wide variety of practical problems,[24−28] proving its robustness and utility. Another approach, an on-the-path random walk sampling algorithm,[29] was proposed to obtain the MFEP using a combination of the string method, λ-dynamics,[30] and metadynamics.

The main goal of this paper is the adaptation of the string method for the analysis of enzymatic reactivity. Analysis of the rate of chemical processes is usually based in the use of Transition State Theory (TST).[7,31] According to this, the rate constant of a chemical reaction can be expressed in terms of a one-dimensional free energy profile or PMF provided that some corrections are considered when the coordinate selected to trace the profile is curvilinear.[32] It must be stressed that the free energy profile obtained from the string method is not a PMF, since the shape of the free energy surface in orthogonal directions to the string is neglected. In other words, D degrees of freedom are not sampled in the zero-temperature string method, while only one degree of freedom (the advance along the reaction coordinate) should be restrained for a correct calculation of the free energy barrier. This issue can be solved by defining a path collective variable (CV) along the converged string. Branduardi et al. introduced two path CVs as functions of series of Cartesian structures selected along a particular path.[33] One coordinate (the s coordinate) describes the advance along the path and the other (the z coordinate) the distance to the path. This approach was applied to study conformational transitions in proteins,[34,35] enzyme inhibition,[36] ion conduction,[37] as well as to describe chemical transformations.[38,39] These s and z functions can be straightforwardly redefined in the space of internal coordinates instead of Cartesian ones[40] or in an arbitrary space of any collective variables one could be interested in. When the path is defined by the MFEP, the PMF can be traced as a function exclusively of the advance along the path (the s coordinate). Explicit consideration of a CV taking into account deviations from the MFEP are not needed because, as said, this lies in the center of a tube where the studied transition is more likely to occur.

In this work, we present a combination of these two techniques, string method and path CVs, to explore and determine enzymatic reaction mechanisms and their associated activation free energy without *a priori* assumptions of the preferred reaction path. The only choice to be done by the user is the set of coordinates that define the multidimensional FES to be explored. This choice can be done from the analysis of the geometry at the reactants and product states. The advantage of such an approach is that after the path CV is defined, any suitable technique might be used to calculate the PMF along it, such as umbrella sampling,[41] metadynamics,[14] etc. On the other hand, having a well-defined analytical expression for the reaction coordinate allows to apply the above-mentioned corrections due to the curvilinear nature of the coordinate.[32]

The methodology will be illustrated with the analysis of the reaction mechanism in Guanidinoacetate methyltransferase (GAMT), an enzyme involved in the catalysis of the last step in creatine biosynthesis.[42,43] GAMT catalyzes the methylation of guanidinoacetate (GAA) by S-adenosyl-L-methionine (SAM), forming creatine and the corresponding S-adenosyl-L-homocysteine (SAH), as shown in Figure 1.[44,45] The reaction involves a methyl transfer from the sulfur atom (S_D) of SAM to the N_E nitrogen of GAA and a proton transfer (H_E) from this nitrogen atom to one of the oxygen atoms (O_D) of the carboxylate group of Asp134 (see Figure 1). The experimental k_{cat} has been determined to be 3.8 ± 0.2 min^{-1} at 296 K, from which a phenomenological activation free energy of 19.0 $kcal{\cdot}mol^{-1}$ can be calculated.[44] Previous theoretical studies have shown the possibility of different reaction mechanisms for this process. In a first study carried out using Density Functional Theory (DFT) calculations on a reduced model of the active site, Himo and co-workers[46] characterized a concerted but asynchronous reaction mechanism where methyl transfer precedes proton transfer and where no stable intermediate was located. However, being the net process a charge migration, the results can dramatically depend on environmental effects. Using self-consistent-charge density functional tight binding/molecular mechanics and single point DFT/MM calculations, Bruice and co-workers[47] found two possible reaction pathways: a stepwise mechanism in which proton transfer precedes methyl transfer and a concerted process where methyl transfer is slightly more advanced than proton transfer.

In the present work, we illustrate the advantages of the combination of the string method and a path CV examining the FES and finding the possible MFEPs for the GAMT reaction by means of the string method. The PMF is then traced as a function of the advance along the strings using a path CV and the activation free energy determined. These calculations were performed using a hybrid QM/MM method that considered the influence of the protein and its flexibility.

2. THEORY AND METHODS

2.1. The String Method. Here, we briefly describe the on-the-fly string method. Interested readers can find details in ref 23. The idea of the method is to define a curve $\{\boldsymbol{z}(s,t)\colon s \in [0,1]\}$ in a space of CVs $\boldsymbol{\theta}(\boldsymbol{R}) = (\theta_1(\boldsymbol{R}),\theta_2(\boldsymbol{R}),...,\theta_D(\boldsymbol{R}))$ that describes a transition between two free energy minima in this space, $\boldsymbol{R}$ being a set of $3N$ Cartesian coordinates of all the atoms in the system (x_i, i = 1, $3N$). A series of equidistant points $\{\boldsymbol{z}(s_1),...,\boldsymbol{z}(s_N)\}$ (the string nodes) is taken along this curve. Two coupled dynamics are done for each node:

1. A normal molecular dynamics, with an extra harmonic potential V_i that keeps the system close to the node i in the CV space:

$$V_i(\mathbf{R}_i(t)) = \frac{k}{2}\,|\boldsymbol{\theta}(\mathbf{R}_i(t)) - \mathbf{z}(s_i, t)|^2 \tag{1}$$

where k corresponds to the force constant of the harmonic potential, $\boldsymbol{R}_i$ is the replica of the system for the string node i and $\boldsymbol{\theta}(\boldsymbol{R}_i)$ represents the vector of current values of all CVs (|···| - Euclidean length of the vector).

2. A dynamic of the string node itself, governed by the following differential equation:

$$\gamma \dot{\mathbf{z}}(s_i, t) = k\tilde{\mathbf{M}}(\mathbf{R}_i(t))(\boldsymbol{\theta}(\mathbf{R}_i(t)) - \mathbf{z}(s_i, t)) \tag{2}$$

where γ is the friction acting on the string node and $\tilde{\mathbf{M}}$ is

$$\tilde{M}_{ab}(\mathbf{R}_i(t)) = \sum_{k=1}^{3N} \frac{\partial\theta_a(\mathbf{R}_i(t))}{\partial x_k}\frac{1}{m_k}\frac{\partial\theta_b(\mathbf{R}_i(t))}{\partial x_k} \tag{3}$$

where N is the number of atoms in the system and m_k is the mass of the atom corresponding to coordinate x_k. In practice, eq 2 is discretized and positions of string nodes are updated every time-step:

$$\mathbf{z}(s_i, t + \Delta t) = \mathbf{z}(s_i, t) - \frac{k}{\gamma}\tilde{\mathbf{M}}(\mathbf{R}_i(t))(\mathbf{z}(s_i, t) - \boldsymbol{\theta}(R_i(t)))\Delta t \quad (4)$$

Since the nodes follow the gradient of the free energy, they tend to drift to the local free energy minima. To avoid this, the string is reparameterized after each step maintaining the nodes equidistant. This can be done by simple linear interpolation as in ref 23 or by some other more precise technique such as cubic splines interpolation, which is the procedure selected in this work. As mentioned by Maragliano et al., the reparameterization does not have to be done each time step; it is sufficient to perform it at given time-intervals so that during this period nodes do not drift too much.[23]

The convergence of the string can be checked by observing the mean displacement of the nodes from their initial positions:

$$d(t) = \frac{1}{N}\sum_{i=1}^{N} |\mathbf{z}(s_i, t) - \mathbf{z}(s_i, 0)| \quad (5)$$

When the string is converged, the free energy profile along it can be obtained as follows:

$$F(s) - F(0) = \int_0^s \frac{dF(\mathbf{z}(s'))}{ds'} ds' = \int_0^s \frac{d\mathbf{z}(s')}{ds'} \frac{dF(\mathbf{z}(s'))}{d\mathbf{z}} ds' \quad (6)$$

The vector $dF(\boldsymbol{z}(s'))/d\boldsymbol{z}$ can be approximately calculated at each node:

$$\frac{dF(\mathbf{z}(s_i))}{d\mathbf{z}} \approx \frac{1}{t_{total} - t_{conv}} \int_{t_{conv}}^{t_{total}} k(\mathbf{z}(s_i, t) - \boldsymbol{\theta}(\mathbf{R}_i(t)))dt \quad (7)$$

where t_{conv} is the simulation time at which the string has converged and t_{total} the total time of the simulation. During string evolution, the values $(\boldsymbol{z}(s_i,t) - \boldsymbol{\theta}(\boldsymbol{R}_i(t)))$ are calculated at each step (eq 4) and have to be saved for a posteriori calculation of the free energy gradient (eq 7). The values of $dF(\boldsymbol{z}(s_i))/d\boldsymbol{z}$ should be interpolated to get the continuous vector function to be used in eq 6. Again, this can be done by simple linear interpolation or by using the cubic splines, as was done in the present work. The vector $d\boldsymbol{z}(s)/ds$ can be obtained by averaging the positions of the nodes $\boldsymbol{z}(s_i)$ over the period $[t_{conv};t_{total}]$, interpolating the averaged string to get $\boldsymbol{z}(s)$ and then calculating the derivative of each component of $\boldsymbol{z}(s)$ with respect to s.

2.2. Path Collective Variable. The s coordinate defined by Branduardi et al.[33] describes the advance of the system along some particular pathway defined, in principle, by the Cartesian coordinates of a series of structures lying on the path. This coordinate can be generalized for a path defined in the space of arbitrary CVs in the following way:[40]

$$s(\mathbf{R}) = \lim_{\lambda\to\infty} \frac{\int_0^1 t\, e^{-\lambda|\boldsymbol{\theta}(\mathbf{R}) - \mathbf{z}(t)|} dt}{\int_0^1 e^{-\lambda|\boldsymbol{\theta}(\mathbf{R}) - \mathbf{z}(t)|} dt} \quad (8)$$

where $\boldsymbol{\theta}(\boldsymbol{R}) = (\theta_1(\boldsymbol{R}),\theta_2(\boldsymbol{R}),...,\theta_D(\boldsymbol{R}))$ is the set of D collective variables employed to define the path; $\boldsymbol{z}(t)$ is a vector whose argument t goes from 0 to 1 defining some path in the space of collective variables $\{\boldsymbol{\theta}\}$ and $|\boldsymbol{\theta}(\boldsymbol{R}) - \boldsymbol{z}(t)|$ is some distance between two points in the space $\{\boldsymbol{\theta}\}$. The simplest choice for this would be the squared Euclidean distance:

$$|\boldsymbol{\theta}(\mathbf{R}) - \mathbf{z}(t)| = \sum_{i=1}^{D} (\theta_i(\mathbf{R}) - z_i(t))^2 \quad (9)$$

The coordinates we use here have the same dimensionality, and we assumed them not to be dynamically coupled, so the squared Euclidean distance was used as in ref 40, ensuring that the isosurfaces of the coordinate are orthogonal to the path. However, in cases when coordinates of different nature are used simultaneously (for instance, distances and angles) or are strongly dynamically coupled, this approach could be not applicable. In this case, the definition can be straightforwardly generalized by using some other distance metric, such as, for example, weighted Euclidean or one including more sophisticated metric tensor.

For practical applications, eq 8 should be discretized:

$$s(\mathbf{R}) = \frac{1}{N-1} \frac{\sum_{i=1}^{N} (i-1) e^{-\lambda|\boldsymbol{\theta}(\mathbf{R}) - \mathbf{z}(s_i)|}}{\sum_{i=1}^{N} e^{-\lambda|\boldsymbol{\theta}(\mathbf{R}) - \mathbf{z}(s_i)|}} \quad (10)$$

where $\boldsymbol{z}(s_i)$ is series of N discrete, equidistant structures along the path and λ is the reciprocal distance between two successive $\boldsymbol{z}(s_i)$.We kept the same notation as was used to describe the string method. A sequence of equidistant points $\boldsymbol{z}(s_i)$ can be generated from the converged string $\boldsymbol{z}(s)$ to be used in eq 10. If the number of string nodes and desired number of points coincide, the nodes of the string can be taken directly, since they are by definition equidistant. In this case, the same distance metric should be used in the reparameterization of the string and the definition of the path CV $s(\boldsymbol{R})$. Normally, if the number of the string nodes is relatively small (to keep the method computationally cheap), it would be desirable to interpolate more structures to be used in eq 10 to get more detailed description of the path.

The umbrella sampling protocol is then used to obtain the PMF. An harmonic energy term is added in each simulation window to the potential energy of the system:

$$V_s(\mathbf{R}) = \frac{1}{2}K_s(s(\mathbf{R}) - s_0)^2 \quad (11)$$

The positions (s_0) and force constants (K_s) are selected using the free energy profile provided by the string method to provide an approximately uniform sampling along the path CV. It must be pointed out that if the path is strongly curved, the sampling would not be orthogonal to the path. However, this will happen only if the deviation of the system from the path is larger than the radius of curvature and since we use the MFEP to define the path CV, we expect that the sampling never goes too far from the path. In any case, if the path presents a high degree of curvature, it would be then desirable to select a different space of CVs.

2.3. Setup of the System and Simulations. The initial coordinates of the system were taken from the X-ray crystal structure 1XCL of GAMT.[44] The enzyme was crystallized in the reactant state, where the methyl group is bonded to the cofactor. Protonation state of titratable residues at pH = 7 was determined by means of PROPKA2.0 program.[48,49] Addition of hydrogen atoms and minimization of the resulting structure was carried out using fDYNAMO tools.[50,51] The system was then placed in a cubic box of 55.8 Å/side of pre-equilibrated water molecules. Those water molecules closer than 2.8 Å to any

heavy atom of the protein, substrate, or crystallization water molecules were removed to ensure a correct solvation of the system and to avoid a possible overlapping among molecules. Therefore, the full system was finally composed by the substrate and the cofactor (SAM and GAA), the protein (with 229 residues), 82 crystallization water molecules, and 4475 solvation water molecules. A total number of 17324 atoms formed the simulated model. The substrate and the cofactor, together with Asp134 were included in the quantum subsystem while the rest of the system constituted the MM subsystem. The QM part was described by the AM1 Hamiltonian,[52] while the protein was described by OPLS-AA force field[53] and the water molecules by the TIP3P potential.[54] During the QM/MM minimizations and simulations, periodic boundary conditions and a cutoff radius of 17.5 Å for all kinds of interactions were employed. After minimization, those MM atoms beyond 25 Å from the substrate or cofactor were frozen, allowing the rest to move. The system was then heated and equilibrated by Molecular Dynamic simulations carried out in the NVT ensemble at 300 K, using the Langevin–Verlet algorithm and a time step of 1 fs. The total length of the simulations was of 1.5 ns.

Since the reaction catalyzed by GAMT consists of two transfer processes, the four distances corresponding to breaking/forming of chemical bonds were selected as the active space of CVs for the string method: S_D-C_E and C_E-N_E for the methyl transfer and N_E-H_E and H_E-O_D for the proton transfer. Thirty string nodes were used for each path. The values employed for the force constant (eq 1) and friction (eq 2) in the string method were $k \approx 3000$ kJ·mol^{-1}·Å^{-2} and $\gamma \approx 45$ s·kJ·mol^{-1}·Å^{-2}. The value of the force constant was determined in such a way that the mean distance of the system from the corresponding node $\langle |\boldsymbol{\theta}(\boldsymbol{R}_i(t)) - \boldsymbol{z}(s_i,t)| \rangle$ become <0.025 Å. γ was then calculated so that the mean displacements of the node during one step of dynamics were ca. 0.001 Å. Sixty picoseconds of Langevin dynamics at 300 K with a time step of 1 fs were performed for each node. String reparameterization was performed each 10 fs. The averaged positions of the string nodes were determined over the last 50 ps, once the strings were converged, and employed to calculate the derivatives for eq 6 and to define the path CV. Different initial guesses were used in the string method to explore all the possible reaction mechanisms schematically presented in Figure 1, but as explained below, only two different converged strings were obtained.

A set of 70 structures was interpolated from both the converged strings to be used in the definition of the path CV given by eq 10. The coordinate was parametrized so that the values 0 and 1 correspond to the end points of the converged string. Negative values correspond to reference structures extrapolated beyond the reactants and values larger than unity to extrapolated structures beyond the reaction products. This extrapolation procedure is used to have well-defined reactants and products valleys in the PMF. For each converged string, the parameters for 50 umbrella sampling[41] windows were adjusted to flatten the free energy profile obtained from the string method with force constants ranging between 10000 and 40000 kJ·mol^{-1}. Initial structures for each window were taken from the final structures of the closest string node. One picosecond of relaxation followed by 20 ps of production dynamics with a time step of 0.5 fs was performed for each simulation window. Then, the PMFs were integrated from the obtained histograms using WHAM.[17]

3. RESULTS

3.1. String Method. Initial guesses are needed for the string method. We selected those paths corresponding to (i) a geometrically synchronous path (diagonal of Figure 1); (ii) an ideal asynchronous path where methyl transfer occurs first with formation of a tetra-substituted N_E atom, and (iii) an ideal asynchronous path where proton transfer precedes methyl transfer. Thirty structures were linearly interpolated from reactants to products for each of these initial guesses. These could be then directly used as input for the string method or improved by means of a zero-temperature version of string method,[55] as done also in this work. In this case, the structures were updated with forward Euler method with a step of 0.0001 Å^2·mol·kJ^{-1}. The use of zero-temperature string method improves the convergence of the finite temperature method but does not change the final results. As explained in the previous section, 60 ps of Langevin dynamics were performed for each of the nodes. Evolution of the nodes, as monitored by the mean displacement (eq 5) typically converges in 10 ps (see Figure S1 in Supporting Information), while the last 50 ps are used to define the average properties and compute the free energy along the string. Once converged, fluctuations of the string have amplitudes of ca. 0.1 Å (Figure S1). The amplitude of the fluctuation could be reduced by using larger friction values but increasing the convergence time.

In our simulations, the string method converged to two different solutions. The geometrical description of the two solutions found is provided in Figure 2. This plot shows the evolution of the four distances that define the multidimensional

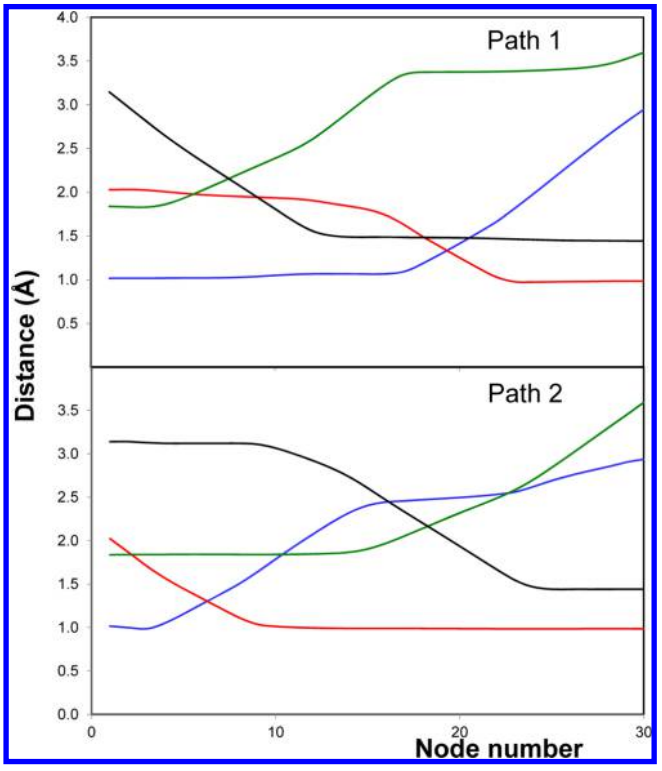


Figure 2. Evolution of average distances along the two strings found for the reaction catalyzed by GAMT: H_E-N_E (blue line), H_E-O_D (red line), C_E-S_D (green line), and C_E-N_E(black line). Path 1 (upper figure) corresponds to a mechanism in which methyl transfer precedes proton transfer while in Path 2 (lower figure) proton transfer precedes methyl transfer.

dx.doi.org/10.1021/ct400153r | J. Chem. Theory Comput. 2013, 9, 3740–3749

FES from reactants to products as a function of the node number. Note that while the free energy is determined in a four-dimensional space, the coordinates of all those atoms found at a distance smaller than 25 Å from the substrate or SAM are sampled during the simulations. The first of the converged strings corresponds to a reaction path in which the methyl is first transferred from the sulfur atom of the cofactor to the N_E atom of the substrate. This process is accompanied by a reduction of the hydrogen bond distance between the N_E atom of the substrate and the carboxylate group of Asp134, which finally leads to the proton transfer from the former to the O_D atom of the latter. The second converged string corresponds also to a process that starts with the approach of Asp134 to the substrate and with the proton transfer from N_E to O_D. Then, the reaction is completed with the transfer of the methyl group from SAM to the unprotonated substrate.

The performance of the string method to converge to the MFEPs could be tested obtaining the four-dimensional FES. As this is a computationally unaffordable task, we instead traced the AM1/MM FES as a function of two coordinates: the antisymmetric combination of distances that define the proton $((H_E-N_E)-(H_E-O_D))$ and the methyl $((C_E-S_D)-(C_E-N_E))$ transfers (details are given in the Supporting Information). The projection of the converged strings on this reduced FES confirms the success of the string method locating the MFEPs (see Figure 3). It should be stressed that the computational cost

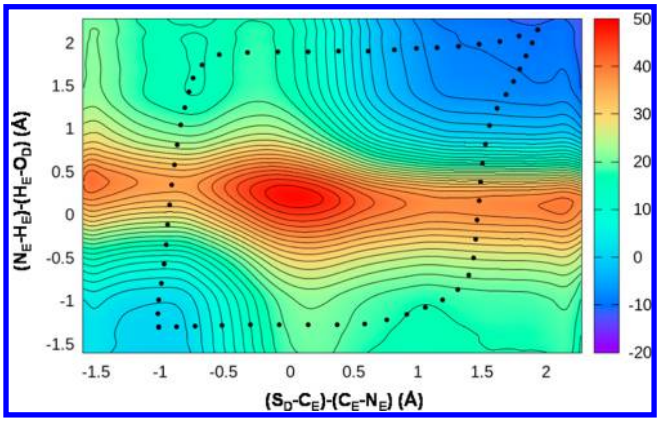


Figure 3. Free energy surface for the GAMT catalyzed reaction obtained as a function of the two antisymmetric combinations of distances defining the proton and the methyl transfer. Isocontour lines are drawn in kcal·mol^{-1}. Projected on the surface (black dots) are the nodes corresponding to the two converged MFEPs obtained with the string method.

of a two-dimensional FES is about 10^2 higher than the cost of the string method. Obviously, the cost dramatically increases with the number of dimensions explicitly incorporated in the FES while this factor practically does not alter the computational time needed for the string method.

The AM1/MM free energy profiles obtained along the two converged MFEPs are shown in Figure 4. These profiles are obtained from application of eq 6 to the data obtained from the string method without any additional simulation. Note, as said in the introduction, that these free energy profiles correspond to the change along the MFEP determined on the four-dimensional active space and thus, this is not the same that the free energy change along the reaction coordinate (the reaction PMF), which is the magnitude to be used in the determination of the rate constant and that is obtained in the next section. In

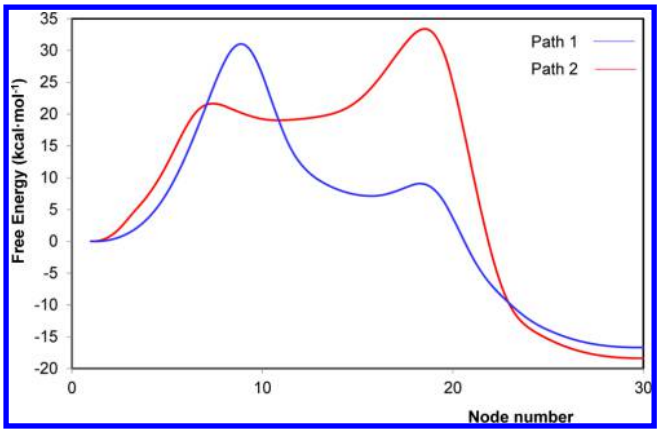


Figure 4. Free energy profiles along the two MFEPs found with the string method for the reaction catalyzed by GAMT: Path 1 (methyl transfer followed by proton transfer) and Path 2 (proton transfer followed by the methyl transfer).

any case, the free energy profiles presented in Figure 4 are illustrative of the reaction mechanism in GAMT. In both cases, the MFEPs correspond to asynchronous stepwise processes. In Path 1, the intermediate corresponds to a tetra-substituted N_E atom, which is stabilized by means of hydrogen bond interactions with Asp134 and with the lone pair of S_D. In Path 2, the intermediate is the unprotonated substrate, where a N_E atom bearing a formal negative charge is stabilized again by Asp134 (acting now as proton donor) and the positive charge of the methyl group bonded to the cofactor. In both cases, the rate-limiting-step is the methyl transfer.

3.2. Path Collective Variable and Potential of Mean Force. We obtained the PMFs associated to the two reaction mechanisms described above using the path CV $s(\boldsymbol{R})$ given by eq 10, where the reference $\boldsymbol{z}(s_i)$ values are interpolated from the two average strings. The force constants to be used in the umbrella sampling were determined from the free energy profiles shown in Figure 4, providing a good uniform sampling along the path CV. Figure 5 shows the PMFs obtained for the two reaction mechanisms as a function of the path CV $s(\boldsymbol{R})$ (dashed lines) together with the free energy profiles along the converged MFEPs (continuous lines). Averaged geometries and relative values of the PMF for the stationary states of both mechanisms are given in Table 1.

Figure 5 shows that the free energy profiles along the MFEPs are, in this case, a very good approximation to the PMFs. Note that this can be not true for all the cases because it depends on the change of the curvature of the free energy surface around the MFEP. If the curvature orthogonal to the path remains essentially constant, then the free energy profile would be approximately equal to the PMF. While the free energy barriers determined from the PMFs are close to those obtained from the free energy profiles along the MFEPs, the free energy corresponding to the products are lower in the former. This means that the entropic contribution due to the three degrees of freedom orthogonal to the path (which are ignored in the free energy change estimated from the MFEPs) is higher in the products than in the reactants. It is also interesting to note that the relative value of the PMF for the products obtained for the two mechanisms differ in only 0.6 kcal·mol^{-1}. This small difference can be partly due to finite sampling and integration errors. However, it should be noted that, in contrast to the free energy profiles, the PMFs do not have to coincide in the

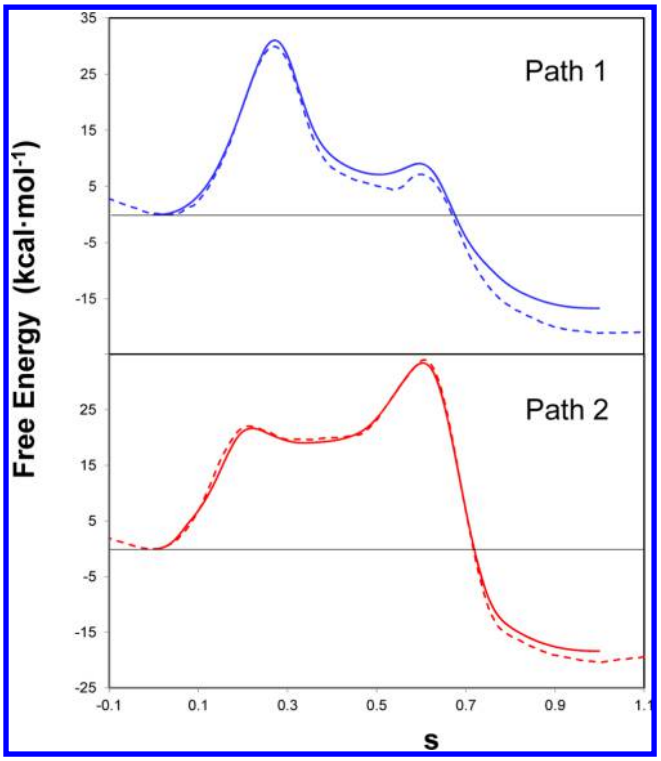


Figure 5. AM1/MM Free energy profiles along the MFEPs (continuous line) and PMFs (dashed lines) as a function of the path CV for the two possible mechanisms in GAMT: Path 1 (blue) and Path 2 (red).

products, since isohypersurfaces of path CVs that correspond to the same local minimum are, in general, different, when different reference pathways are used.

The geometrical description of the reaction along the PMFs (Table 1) is very similar to the results obtained from the string method (Figure 2). Path 1 describes an asynchronous stepwise process, where the first TS (TS1) corresponds to the methyl transfer. At TS1 the methyl group is slightly closer to the target nitrogen atom of the substrate than to the donor sulfur atom. An intermediate with a positively charged N_E atom is formed with higher free energy than the reactants (4.3 kcal·mol^{-1}). At this intermediate, the substrate presents a strong hydrogen bond with Asp134 (the average H_E-O_D distance being 1.82 Å). From this intermediate, the reaction continues through a small free energy barrier (2.9 kcal·mol^{-1}) transferring the proton from the substrate to Asp134. In Path 2, the process begins with the proton transfer from the N_E atom of the substrate to the carboxylate group of Asp134. At TS1, the process is quite advanced, being the H_E-O_D distance shorter than the H_E-N_E one (1.19 and 1.38 Å, respectively). After a shallow intermediate that corresponds to an unprotonated substrate stabilized by a strong hydrogen bond with Asp134 (N_E-H_E distance equal to 1.79 Å), the reaction proceeds with the methyl transfer from SAM to the substrate. At TS2, which is the highest in free energy (33.9 kcal·mol^{-1} relative to the reactants), the methyl group is found almost halfway between the donor and the acceptor atoms (the distances to C_E being 2.19 and 2.15 Å, respectively).

3.3. Activation Free Energies. The main purpose of this work is to show that a combination of the string method and a path CV (defined in a space of interatomic distances) provides a valuable theoretical tool to explore the mechanism of complex chemical reactions in condensed environments. Obviously, the quality of the free energy changes along the reaction coordinates for the reaction explored here are limited by the quality of the energy function employed in the simulations. This can be improved using higher level quantum methods or by means of an appropriate parametrization of the energy function. In our case, the semiempirical AM1 Hamiltonian clearly overestimates the free energy barriers when compared to experimental derived values. While this is not the main goal of the paper, we illustrate in this section a fast method to improve the quality of the results and also the consideration of several correction terms to the PMF change that should be considered in the evaluation of activation free energies.

In order to reduce the errors associated to the low-level of theory (LL) description employed to obtain the PMFs, we recalculated them applying a correction method developed in our group.[56,57] The method considers the inclusion of an interpolated correction energy term obtained as a function of the reaction coordinate $s(\boldsymbol{R})$:

$$E = E_{QM}^{LL} + E_{QM/MM}^{LL} + E_{MM} + \mathrm{Spl}[\Delta E_{LL}^{HL}(s)] \qquad (12)$$

where Spl denotes a cubic-spline whose argument is the correction term ΔE_{LL}^{HL}, obtained as the difference between the energy provided by the low level method (LL) and a single-point calculation at a higher level of theory (HL) for configurations of the QM subsystem obtained along the reaction coordinate. In this case, the LL single-point energies were obtained using the AM1 Hamiltonian, while the HL energies were obtained at the two different levels, the hybrid functional M06-2x/6-31+G(d,p)[1,58] and the *ab initio* method MP2/6-31+G(d,p).[59] The configurations selected to evaluate

Table 1. Free Energies (in kcal·mol^{-1}) for the Stationary Structures Corresponding to the PMFs for the Two Possible Reaction Mechanism in GAMT, the Corresponding Average Distances (in Å) and Their Standard Deviations (in Parentheses)

		reactants	TS1	intermediate	TS2	products
Path 1	free energy	0.0	30.0	4.3	7.2	−21.0
	$d(N_E-H_E)$	1.02 (0.03)	1.04 (0.03)	1.07 (0.03)	1.22 (0.03)	2.89 (0.16)
	$d(H_E-O_D)$	2.06 (0.14)	1.96 (0.11)	1.82 (0.09)	1.44 (0.03)	0.99 (0.03)
	$d(S_D-C_E)$	1.84 (0.04)	2.28 (0.05)	3.48 (0.11)	3.54 (0.21)	3.69 (0.30)
	$d(C_E-N_E)$	3.00 (0.01)	1.97 (0.04)	1.49 (0.03)	1.48 (0.03)	1.44 (0.02)
Path 2	free energy	0.0	22.0	19.6	33.9	−20.4
	$d(N_E-H_E)$	1.02 (0.03)	1.38 (0.03)	1.79 (0.06)	2.58 (0.31)	2.99 (0.29)
	$d(H_E-O_D)$	2.06 (0.02)	1.19 (0.03)	1.02 (0.03)	0.99 (0.03)	0.98 (0.03)
	$d(S_D-C_E)$	1.84 (0.04)	1.84 (0.05)	1.84 (0.05)	2.19 (0.05)	3.60 (0.06)
	$d(C_E-N_E)$	3.08 (0.15)	3.16 (0.15)	3.20 (0.18)	2.15 (0.05)	1.44 (0.02)

dx.doi.org/10.1021/ct400153r | J. Chem. Theory Comput. 2013, 9, 3740−3749

this correction term were obtained optimizing the system at different values of the path CV $s(\mathbf{R})$.

The PMFs obtained using this correction term at the M06-2X level are shown in Figure 6. Those profiles obtained using

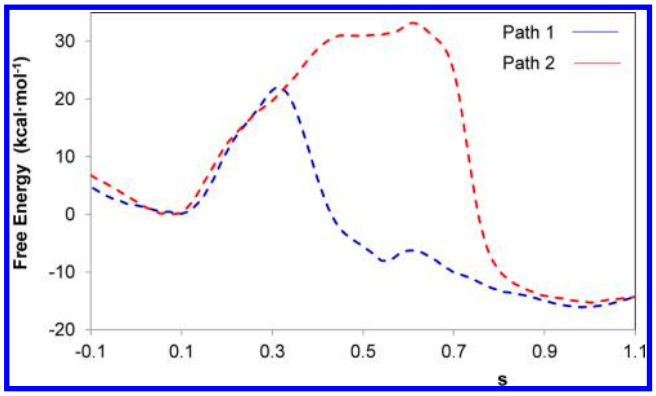


Figure 6. PMFs along the path CV corrected at the M06-2X/6-31+G(d,p) level.

corrections at the MP2 level are almost identical and are provided as Supporting Information (see Figure S2). In the corrected PMFs, the first mechanism, in which methyl transfer precedes proton transfer, is clearly favored presenting a PMF barrier of about 21.9 kcal·mol^{-1}, while in the alternative mechanism, where the proton transfer precedes methyl transfer, the PMF barrier is about 33.2 kcal·mol^{-1}. Another noticeable difference with respect to the AM1/MM results is that both mechanisms are now described as asynchronous but concerted. Thus, according to our simulations, the reaction mechanism in GAMT is probably a concerted asynchronous process where methyl transfer precedes proton transfer.

Finally, in order to get an estimation of the activation free energy, we must consider two contributions in addition to the PMF change between the TS and the reactants state (R). The activation free energy is properly given by[32]

$$\begin{aligned}\Delta G(s) &= \Delta W(s) - \Delta G_{s^R} + W_{curv}(s)\\ &= [W(s) - W(s^R)] - \Delta G_{s^R} + W_{curv}(s)\end{aligned} \quad (13)$$

where the first term is the difference between the value of the PMF at the maximum ($s^\ddagger$) and the reactants minimum (s^R), the second term accounts for the work associated to setting the reaction coordinate to the value of the PMF minimum, and the last one accounts for the curvilinear nature of the reaction coordinate. The work due to setting the reactants to a particular value of the reaction coordinate (s^R) has been evaluated as the contribution of one classical vibrational degree of freedom with frequency ν_s that correspond to the motion along the reaction coordinate at the reactants:[32]

$$\Delta G_{S^R} = -RT \ln \frac{kT}{h\nu_s} \quad (14)$$

The frequency ν_s was obtained from the force constant derived from the PMF variation around the reactants minima and the mass associated to this coordinate (obtained from the application of the equipartition principle to the velocity along the coordinate). For the reactants state of Path 1, the frequency was 215 cm^{-1}, and then, the free energy term in eq 14 amounts to only 0.02 kcal·mol^{-1}. The correction due to the curvilinear nature of the reaction coordinate has been evaluated as[32]

$$W_{curv}(s^\ddagger) = -RT \ln \frac{\langle (\sum_{i=1}^{3N} m_i^{-1}(\partial s/\partial x_i))^{1/2} \rangle_{s=s^\ddagger}}{\langle (\sum_{i=1}^{3N} m_i^{-1}(\partial s/\partial x_i))^{1/2} \rangle_{s=s^R}} \quad (15)$$

This quantity is equal to −0.14 kcal·mol^{-1} for Path 1. Then, our best estimate of the activation free energy using the M06-2X corrected PMF is 21.9 − 0.02 − 0.14 ∼ 21.7 kcal·mol^{-1}, close to the phenomenological activation free energy derived from the experimental rate constant (19.0 kcal·mol^{-1}).[44]

The corrections due to the motion along the reaction coordinate at the reactants and the curvilinear nature of the coordinate are slightly larger in the second mechanisms (0.70 and 0.43 kcal·mol^{-1}, respectively), but this mechanism is discarded because of its much larger free energy barrier (33.2 − 0.70 + 0.43 ∼ 32.9 kcal·mol^{-1}).

4. CONCLUSIONS

A combination of the on-the-fly string method and a path CV has been used to explore the multidimensional free energy surface of an enzymatic reaction. The string method was employed to locate the MFEPs, and a path CV is then used to trace the corresponding PMF as a function of the advance along the path. This combination of methodologies has been implemented to work in a multidimensional space of CVs defined as interatomic distances, which is more appropriate for the exploration of chemical reactions because bond breaking and forming processes can be better expressed as a function of interatomic distances.

This computational methodology has been implemented in a local version of the program fDynamo, and the performance tested in the study of the methylation reaction of guanidinoacetate catalyzed by GAMT, using hybrid QM/MM simulations. The analysis of the results show that, depending on the initial guess employed, the string method converges to two possible MFEPs. The two reaction mechanisms observed at the AM1/MM level are stepwise processes where the relative ordering between methylation and proton transfer is reversed. The converged strings perfectly match the MFEPs observed on a reduced two-dimensional FES. Obviously, the computational cost of the string method is significantly smaller than the needed for an exploration of the complete FES.

The PMFs along the MFEPs can be then traced using a CV that measures the advance along the path. The PMFs obtained do not differ much from the free energy profiles directly obtained from the string calculation. Although the two magnitudes are not equal, this observation could allow determining approximate activation free energies from string calculations, without explicit calculation of the free energy as a function of the reaction coordinate. Other contributions associated with the path CV in the determination of the activation free energy (the curvilinear correction and the work associated to setting the reactants at a particular value of the coordinate) have been shown to be small in this case, although more experience could be needed before neglecting them in every case.

■ ASSOCIATED CONTENT

Supporting Information

Figures showing the RMSD evolution of the string and MP2 corrected PMFs and computational details about the calculation of the free energy surface. This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ignacio.Tuñón@uv.es.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from MEC (project CTQ2012-36253-C03) J.J.R.P. thanks Ministerio Ciencia e Innovación for a 'Juan de la Cierva' contract and Generalitat Valenciana (project GV/2012/044) and Bancaixa (project P1·1A2010-08). K.Z. acknowledges a fellowship of the Universidad de Valencia ('Atracció de Talent'), a FPU fellowship from Ministerio de Educación and a Geronimo Forteza contract at Universitat Jaume I from the Generalitat Valenciana (ref FPA/2012/052). The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València in the 'Tirant' supercomputer.

REFERENCES

(1) Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(2) Zhao, Y.; Truhlar, D. G. Applications and validations of the Minnesota density functionals. *Chem. Phys. Lett.* **2011**, *502*, 1−13.

(3) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(4) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(5) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(6) Chandler, D. *Introduction To Modern Statistical Mechanics*. Oxford University Press: Oxford, U.K., 1987.

(7) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current status of transition-state theory. *J. Phys. Chem.* **1996**, *100*, 12771−12800.

(8) Roca, M.; Marti, S.; Andres, J.; Moliner, V.; Tuñón, I.; Bertran, J.; Williams, I. H. Theoretical modeling of enzyme catalytic power: Analysis of "cratic" and electrostatic factors in catechol O-methyltransferase. *J. Am. Chem. Soc.* **2003**, *125*, 7726−37.

(9) Rajamani, R.; Naidoo, K. J.; Gao, J. Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. *J. Comput. Chem.* **2003**, *24*, 1775−81.

(10) Ruiz Pernia, J. J.; Tuñón, I.; Williams, I. H. Computational simulation of the lifetime of the methoxymethyl cation in water. A simple model for a glycosyl cation: When is an intermediate an intermediate? *J. Phys. Chem. B* **2010**, *114*, 5769−74.

(11) Nam, K.; Gao, J.; York, D. M. Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J. Am. Chem. Soc.* **2008**, *130*, 4680−91.

(12) Kanaan, N.; Roca, M.; Tuñón, I.; Marti, S.; Moliner, V. Theoretical study of the temperature dependence of dynamic effects in thymidylate synthase. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11657−64.

(13) Ensing, B.; Laio, A.; Gervasio, F. L.; Parrinello, M.; Klein, M. L. A minimum free energy reaction path for the E2 reaction between fluoro ethane and a fluoride ion. *J. Am. Chem. Soc.* **2004**, *126*, 9492−9493.

(14) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826−843.

(15) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.

(16) Maragliano, L.; Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **2006**, *426*, 168−175.

(17) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

(18) Vanden-Eijnden, E.; Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **2009**, *130*, 194103.

(19) Díaz Leines, G.; Ensing, B. Path finding on high-dimensional free energy landscapes. *Phys. Rev. Lett.* **2012**, *109*, 020601.

(20) Bohner, M. U.; Kastner, J. An algorithm to find minimum free-energy paths using umbrella integration. *J. Chem. Phys.* **2012**, *137*, 034105.

(21) Kastner, J. Umbrella integration in two or more reaction coordinates. *J. Chem. Phys.* **2009**, *131*, 034109−8.

(22) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 24106.

(23) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182−190.

(24) Miller, T. F.; Vanden-Eijnden, E.; Chandler, D. Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14559−14564.

(25) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *J. Chem. Phys.* **2011**, *134*, 085103.

(26) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8934−8941.

(27) Stober, S. T.; Abrams, C. F. Energetics and mechanism of the normal-to-amyloidogenic isomerization of β2-microglobulin: On-the-fly string method calculations. *J. Phys. Chem. B* **2012**, *116*, 9371−9375.

(28) Matsunaga, Y.; Fujisaki, H.; Terada, T.; Furuta, T.; Moritsugu, K.; Kidera, A. Minimum free energy path of ligand-induced transition in adenylate kinase. *PLoS Compu.t Biol.* **2012**, *8*, e1002555.

(29) Chen, M.; Yang, W. On-the-path random walk sampling for efficient optimization of minimum free-energy path. *J. Comput. Chem.* **2009**, *30*, 1649−1653.

(30) Kong, X.; Brooks, C. L., III Lambda-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414−2423.

(31) Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **1935**, *3*, 107−115.

(32) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.* **2003**, *119*, 5828−5833.

(33) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.

(34) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 13938−13944.

(35) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling protein motion and solvent effect during ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1467−1472.

(36) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5411−5416.

(37) Ceccarini, L.; Masetti, M.; Cavalli, A.; Recanatini, M. Ion conduction through the hERG potassium channel. *PLoS ONE* **2012**, *7*, e49017.

(38) Branduardi, D.; De Vivo, M.; Rega, N.; Barone, V.; Cavalli, A. Methyl phosphate dianion hydrolysis in solution characterized by path collective variables coupled with DFT-based enhanced sampling simulations. *J. Chem. Theory Comput.* **2011**, *7*, 539−543.

(39) Lodola, A.; Branduardi, D.; De Vivo, M.; Capoferri, L.; Mor, M.; Piomelli, D.; Cavalli, A. A catalytic mechanism for cysteine N-terminal nucleophile hydrolases, as revealed by free energy simulations. *PLoS ONE* **2012**, *7*, e32397.

(40) Zinovjev, K.; Martí, S.; Tuñón, I. A collective coordinate to obtain free energy profiles for complex reactions in condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 1795−1801.
(41) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(42) Im, Y. S.; Chiang, P. K.; Cantoni, G. L. Guanidoacetate methyltransferase. Purification and molecular properties. *J. Biol. Chem.* **1979**, *254*, 11047−50.
(43) Ogawa, H.; Ishiguro, Y.; Fujioka, M. Guanidoacetate methyltransferase from rat liver: Purification, properties, and evidence for the involvement of sulfhydryl groups for activity. *Arch. Biochem. Biophys.* **1983**, *226*, 265−75.
(44) Komoto, J.; Yamada, T.; Takata, Y.; Konishi, K.; Ogawa, H.; Gomi, T.; Fujioka, M.; Takusagawa, F. Catalytic mechanism of Guanidinoacetate methyltransferase: Crystal structures of Guanidinoacetate methyltransferase ternary complexes. *Biochemistry* **2004**, *43*, 14385−14394.
(45) Fujioka, M.; Konishi, K.; Takata, Y. Recombinant rat-liver guanidinoacetate methyltransferase−Reactivity and function of sulfhydryl-groups. *Biochemistry* **1988**, *27*, 7658−7664.
(46) Velichkova, P.; Himo, F. Theoretical study of the methyl transfer in guanidinoacetate methyltransferase. *J. Phys. Chem. B* **2006**, *110*, 16−9.
(47) Zhang, X.; Bruice, T. C. Reaction mechanism of guanidinoacetate methyltransferase, concerted or step-wise. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16141−6.
(48) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins* **2005**, *61*, 704−21.
(49) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pK_a values for protein−ligand complexes. *Proteins* **2008**, *73*, 765−83.
(50) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press: Cambridge, 1999.
(51) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000**, *21*, 1088−1100.
(52) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum-mechanical molecular-models 0.76. AM1: A new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(53) Jorgensen, W. L.; Tiradorives, J. The OPLS potential functions for proteins—Energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(55) E, W.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.
(56) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. Hybrid QM/MM potentials of mean force with interpolated corrections. *J. Phys. Chem. B* **2004**, *108*, 8427−8433.
(57) Ruiz-Pernia, J. J.; Silla, E.; Tunon, I.; Marti, S. Hybrid quantum mechanics/molecular mechanics simulations with two-dimensional interpolated corrections: Application to enzymatic processes. *J. Phys. Chem. B* **2006**, *110*, 17663−70.
(58) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(59) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618−622.