



Published in final edited form as:

*Methods Enzymol.* 2016 ; 577: 57–74. doi:10.1016/bs.mie.2016.05.012.

## Generalized Ensemble Sampling of Enzyme Reaction Free Energy Pathways

Dongsheng Wu<sup>1,†</sup>, Mikolai I. Fajer<sup>2,3,†</sup>, Liaoran Cao<sup>3</sup>, Xiaolin Cheng<sup>2,4,\*</sup>, and Wei Yang<sup>1,3,\*</sup>

<sup>1</sup>Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

<sup>2</sup>UT-ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830

<sup>3</sup>Department of Chemistry & Biochemistry, Florida State University, Tallahassee, Florida 32306

<sup>4</sup>Department of Biochemistry, Cellular and Molecular Biology, the University of Tennessee, Knoxville, Tennessee, 37996

### Abstract

Free energy path sampling plays an essential role in computational understanding of chemical reactions, particularly those occurring in enzymatic environments. Among a variety of molecular dynamics simulation approaches, the generalized ensemble sampling strategy is uniquely attractive for the fact that it not only can enhance the sampling of rare chemical events but also can naturally ensure consistent exploration of environmental degrees of freedom. In this review, we plan to provide a tutorial-like tour on an emerging topic: generalized ensemble sampling of enzyme reaction free energy path. The discussion is largely focused on our own studies, particularly ones based on the metadynamics free energy sampling method and the on-the-path random walk path sampling method. We hope that this mini presentation will provide interested practitioners some meaningful guidance for future algorithm formulation and application study.

### Keywords

Enzyme Reaction; Generalized Ensemble Sampling; Minimum Free Energy Path

### Introduction

Powered by thermal reservoir, chemical reaction systems aimlessly fluctuate in their surrounding media. Through molecular interactions, energy is channeled to the reaction center and ultimately leads to large amplitude of fluctuations that cause chemical bond breaking and formation. Because the overall probability for a system to be adequately activated and successfully form reactive configurations is low, a chemical reaction is usually orders of magnitudes slower than the elementary vibration that is directly responsible for the reactive event. Based on the transition state theory (TST)<sup>1,2</sup>, among all the possible regions, from which the system can barrierlessly proceed to the product basin, there is a

\*Correspondence: chengx@ornl.gov (X.L.C.); yyang2@fsu.edu (W.Y.), Tel: 850-645-6884, Fax: 850-644-7244.

†Equal contribution to this work

characteristic transition state (TS) region, the reaching of which from the reactant basin requires the least activation input, and therefore the TS region is expected to attract a majority of reaction fluxes to pass through. Based on such a simplified picture, understanding the mechanism of a chemical reaction largely means (a) elucidating how a chemical system proceeds along the most probable pathway, in particular from its reactant basin to the transition state region and (b) quantifying the free energy change along such a pathway. Correspondingly, free energy path calculation constitutes a major task in computational analysis of chemical reaction mechanisms, such as those occurring in enzymatic environment.

Enzyme reaction represents a special class of chemical reactions, the rates of which are usually significantly higher than their counterparts in aqueous solution; for instance enzymatic rate enhancement<sup>3</sup> can be as high as  $\sim 10^{17}$ . Applying molecular dynamics (MD) simulation methods to elucidate enzyme reaction mechanisms and understand the corresponding catalytic strategies has been a classical topic in computational chemistry and biophysics<sup>4–9</sup>. In the recent years, method development efforts for this topic, in particular towards quantitative depiction of free energy pathways, have been reviving, partly due to ever-increasing interest in enzyme designs and partly due to an increasing demand of predicting atomistic level details that can serve as meaningful hypothesis for experimental test. Generally speaking, the quality of an enzyme reaction free energy path calculation relies on both potential energy function accuracy and sampling adequacy. Constrained by ever limited computing power, tremendous algorithm developments have focused on (a) how to reduce energy and force evaluation cost while maintaining minimum accuracy loss and (b) how to more efficiently perform reaction free energy path sampling. Development efforts in the former aspect have been widely acknowledged, for instance as represented by the combined quantum mechanical and molecular mechanical (QM/MM) scheme<sup>10–11</sup> and recent multi-scale reactive force field models<sup>12–14</sup>. In contrast, advancing sampling methods for free energy path calculation only became flourishing about a decade ago. Notably, recent advancements have been mainly catalyzed by the development of generalized ensemble sampling techniques<sup>15–17</sup> and path optimization algorithms<sup>18–25</sup>. In this review, we focus our discussion on an emerging topic: generalized ensemble sampling of enzyme reaction free energy path. Our review is largely based on our earlier and recent studies; these studies respectively represent early applications of the metadynamics free energy sampling method<sup>26</sup> and the on-the-path random walk path sampling method<sup>24,27</sup> on enzyme reaction systems. We hope that our presentation will provide interested practitioners some meaningful guidance for future algorithm formulation and application study.

## Collective Variable and Reaction Order Parameter

Although enzymes reaction rates are accelerated, in comparison with timescales commonly accessible to MD simulations, these processes are still far too slow. Estimated based on the diffusion limit, commonly, enzyme reaction free energy barriers are higher than 11–12 kcal/mol<sup>28</sup>. Together with the fact that costly electronic calculation has to be included for chemical transformation treatment, there has never been a hope that a reactive trajectory can be obtained by a canonical ensemble QM/MM molecular dynamics simulation. Therefore, activation biases need to be introduced in order to sufficiently sample high free energy

regions, particularly regions that are  $> 2$  kT above the reactant basin; for instance, a simplest way to realize such activation is to restrain the system around to-be-activated regions. Following this requirement, a key technical question arises: which degrees of freedom need to be chosen for sampling enhancement.

In correspondence, often the first step in an enzyme reaction simulation study is to choose a set of essential collective variables (CVs) (commonly geometric variables) that is hopefully sufficient to describe the target reaction. In the context of free energy path sampling, being an “essential” collective variable indicates that free energy flattening along this CV can lead to random walk dynamics in certain portions of the reaction pathway, which is otherwise unachievable along another candidate CV. In an application study, collective variable selection is commonly dictated by chemical intuition, e.g. simulators’ educated guess or priori knowledge of the reaction mechanism. In practice, simulators often choose “reaction bond order parameters” (RBOP), which can distinguish the reactant basin, the product basin, and plausible metastable intermediates, as candidate CVs. Here is an example. Inosine monophosphate dehydrogenase (IMPDH) catalyzes two sequential chemical transformations: (1) a dehydrogenase reaction between IMP and  $\text{NAD}^+$  that produces a Cys319-linked intermediate E-XMP\* and NADH, and (2) a hydrolysis reaction that releases XMP. Before our study<sup>29</sup> on the hydrolysis step catalyzed by the IMPDH Arg418Gln variant, it was known that Tyr419 needs to be deprotonated; and thereby it can act as a general base to activate a water molecule so that it can become a better nucleophile to replace Cys319 on the XMP ring (Figure 1 A). Two reaction processes, proton transfer between water and Tyr419 and nucleophilic substitution between the activated water (if completely activated, a hydroxyl group) and Cys319, may occur either sequentially or concertedly; even within each chemical process, two sub-events, for instance proton detachment from water and protonation of Tyr419 during proton transfer, may also occur either sequentially or concertedly. Thus four pairs of distances, corresponding to four chemical bonds [O(water)-H(water), O(Tyr419)-H(water), O(water)-C(XMP), and S(Cys319)-C(XMP)] that form either in the reactant basin, the product basin, or at possible metastable intermediate states, were initially chosen as candidate CVs for this study. It should be noted that by definition, “order parameters” are defined to distinguish stable basins. As is generally known, order parameters are often insufficient to describe protein conformational changes, because between two conformational basins, slow re-organization changes are likely to occur along orthogonal degrees of freedom. In contrast, enzyme active sites are generally pre-organized for chemical steps and reactive events usually closely follow high-frequency fluctuations that directly involve chemical bond vibrations. Therefore unless chemical transformation occurs through a conformation that differs from that of the starting structure or the reaction involves slow conformational changes, the RBOP based CV identification strategy is generally effective; considering possibly complex electronic structural effects involved in enzyme reactions, sometimes bond formation/breaking angles can be essential CVs<sup>24</sup>.

## Traditional Importance Sampling versus Generalized Ensemble Sampling

Upon the identification of CV candidates, the next question is how to use them to sample enzyme reaction free energy pathways. Till about a decade ago, most related studies had

been carried out based on the traditional importance sampling methods: either the umbrella sampling method<sup>30</sup> or the blue-moon-ensemble method<sup>31</sup>, where independent MD simulations are performed with the CV either restrained around or constrained at a series of values that can cover the entire reaction span. Due to the computing power limitation, a majority of these sampling efforts were along a single CV. Despite that these studies had played a significant role in advancing the field of computational enzymology and deepening our understanding of enzymatic catalysis, their sampling deficiency is obvious. When a single CV is employed, it is very likely that it cannot sufficiently describe the whole reaction pathway. As shown in Figure 2, due to the missing of certain essential CVs, the hidden CV issue may lead the sampling CV to proceed transversely along the physical reaction path. Due to the restriction or the loss of dynamics along the sampling CV, accurately exploring the traverse region can be forbiddingly challenging to the umbrella sampling and the blue-moon-ensemble methods; for instance under the restraint treatment, transitions between A and B or between B and C (Figure 2) have to be through a region kinetically inaccessible to the physical process. It should be specially noted that unless missing essential CVs can be specifically guessed, the hidden CV issue is practically undetectable from such simulations themselves because there is no phase space connectivity information on samples obtained from independently restrained/constrained windows.

Different from the traditional importance sampling strategy, generalized ensemble (GE) methods<sup>15–17</sup> do not require any restriction of dynamics along the sampling CV. Instead, activation along the sampling CVs is enabled through the modification of the Hamiltonian, as follows:

$$H_m = H_o + f_m \left[ \vec{\theta}(X) \right], \quad (1)$$

where  $H_o$  represented the original Hamiltonian and  $f_m[\vec{\theta}(X)]$  stands for the biasing potential along the pre-chosen CV set  $\vec{\theta}(X)$ . Commonly, the target function of  $f_m[\vec{\theta}(X)]$  is set to be  $-G_o[\vec{\theta}(X)]$ , the negative of the free energy surface (FES) mapped along  $\vec{\theta}(X)$ .  $G_o[\vec{\theta}(X)]$  is the sampling target, which is unknown *a priori*. To adaptively obtain  $G_o[\vec{\theta}(X)]$ , three major recursion approaches have been developed, including the adaptive umbrella sampling method<sup>32</sup>, in which free energy estimations are based on probability distributions, the adaptive biasing force (ABF) method<sup>33</sup>, in which free energy estimations are based on the thermodynamic integration (TI) formula<sup>34–35</sup>, and the metadynamics<sup>26</sup>/local elevation<sup>36</sup> method, which is realized through continuous deposition of repulsive basis functions. Through GE sampling, target free energy surfaces are explored through single continuous trajectories. It allows moderate hidden CV problems to be possibly bypassed and severe hidden CV problems to be detectable. In addition, through random walks following a single trajectory, the remaining environment degrees of freedom that are not subject to biased activation can be consistently sampled along the reaction pathway.

Among the above recursion methods, metadynamics has attracted the most attention. In the past years, tremendous efforts have been made to improve its robustness and convergence

behavior. The original metadynamics is as simple as generating  $f_m[\vec{\theta}(X)]$  by continuously depositing Gaussian functions:

$$f_m[\vec{\theta}(X)] = \sum_i h \prod_{i=1}^m \exp\left(-\frac{(\theta_i - \theta_i^t)^2}{2w_i^2}\right), \quad (2)$$

where  $\theta_i$  stands for the  $i$ th collective variable and  $\theta_i^t$  stands for the value of the  $i$ th collective variable at the scheduled time  $t$ ;  $h$  is the height of the basis Gaussian function; and  $w_i$  is the width of the  $i$ th component of the basis Gaussian function. Realizing the fact that the error of

free energy estimation [ $G_o[\vec{\theta}(X)] = -\sum_t h \prod_{i=1}^m \exp\left(-\frac{(\theta_i - \theta_i^t)^2}{2w_i^2}\right)$ ] strongly depends on the size of the basis Gaussian function, our group introduced the first systematic improvement<sup>37</sup> for metadynamics by strategically reducing its size through the Wang-Landau flat-histogram procedure. Since this beginning, there have been several ingenious and rigorous improvements, such as well-tempered metadynamics<sup>38,39</sup>, transition-tempered metadynamics<sup>40</sup>, and recently very promising meta-basin metadynamics<sup>41</sup> etc., formulated. Among them, well-tempered metadynamics has become a widely applied method. As one can expect, the recent more elegant metadynamics methods<sup>40,41</sup> will soon prove their unique advantages for sampling enzyme reaction free energy paths.

## Dimensionality Limit

In theory, reaction free energy sampling can be performed in any number of dimensions. In practice, it is challenging to simultaneously sample more than three CVs. Such dimensionality limit is commonly considered being the result of the sampling manifold issue. As discussed earlier, an essential CV should play its sampling role in certain portions of the reaction pathway; in the other portions of the pathway, ideally it shouldn't be activated so as to confine sampling within a one-dimension reaction channel. In high-dimension GE sampling, CVs are unselectively activated even when they are not around their individual working regions. Therefore the sampling manifold is much larger than the size of the reaction channel. With the increase of the sampling dimensionality, the diffusion time in regions unrelated to the physical process is expected to grow drastically.

Indeed, the origin of dimensionality limit can be physical. For instance, gas phase reactions can be readily studied via three-dimensional metadynamics<sup>42</sup>. On contrary, based on our observation and experience, for enzyme reactions, it is likely that two dimensions are a common practical limit, while with a careful choice of CVs, higher-dimension GE sampling might be marginally possible. When multiple CVs are applied for GE sampling, lower-frequency collective motions inaccessible to the physical process are likely to be promoted. As one can imagine, a boost of 11–12 kcal/mol or above on these lower-frequency motions, which involve collective interplays of these sampling CVs, can be detrimental to the overall structural integrity and the system stability.

## One of the First Metadynamics-Based Enzyme Reaction Studies

In 2008, we reported one of the first metadynamics-based enzyme reaction studies<sup>29</sup>, which is on the IMPDH-catalyzed hydrolysis step. As discussed earlier, for the reaction catalyzed by the IMPDH Arg418Gln variant, we initially identified four candidate distance CVs [d1: O(water)-H(water); d2: O(Tyr419)-H(water); d3: O(water)-C(XMP); and d4: S(Cys319)-C(XMP)]. To reduce the CV dimensionality at least to two, we re-defined d4-d3 as CV1 ( $\theta_1$ ) to sample the nucleophilic substitution process and d2-d1 as CV2 ( $\theta_2$ ) to sample the proton transfer process. It is noted that taking CV difference is a common means to reduce CV dimensionality; however it should be applied with caution because possible high degeneracy may introduce large diffusion sampling overhead.

As shown in Figure 1B, the Wang-Landau metadynamics (flat-histogram metadynamics) simulation led to a detailed and nicely-converged free energy surface, on which besides the reactant and product basins, there is a metastable intermediate occurring between the proton transfer step and the nucleophilic substitution step. From the free energy surface, we could generate a string of CV( $\theta$ )-space points to describe the minimum free energy path (MFEP) between the centers of the reactant and product basins. These points satisfy the following

string condition<sup>20</sup>:  $\left[\frac{\partial G}{\partial \theta}\right]^\perp = 0$ , in which  $\frac{\partial G}{\partial \theta} \left(\frac{\partial G}{\partial \theta_1}, \frac{\partial G}{\partial \theta_2}\right)$  stands for the free energy gradient vector and  $\perp$  denotes the projection perpendicular to the minimum free energy curve. Along the MFEP, the two chemical events proceed in a stepwise manner. The free energy barrier of the second step is higher (about 17 kcal/mol); therefore the nucleophilic substitution process is the rate-limiting step. Taking into the account the free energy penalty for Tyr deprotonation, the overall free energy barrier is about 21–22 kcal/mol, which is in good accord with the barrier observed for the reactions of the IMPDH Arg418Gln and Arg418Ala variants (about 20–21 kcal/mol). The location of the transition state reveal that both the proton transfer step and the nucleophilic displacement step are concerted. As shown in Figure 3c, at the transition state of the rate-limiting step, the S(Cys319)-C(XMP) bond partially breaks and the O(water)-C(XMP) bond partially forms.

Back in 2008, obtaining an enzyme reaction free energy surface with the above quality was rare; it was impossible without our own implementation of the metadynamics method in the CHARMM program<sup>43</sup>. Interestingly, this early implementation has many worth-noting features. For instance, Gaussian functions are deposited to grids with their heights determined by the second-order spline function; in addition, we introduced a mechanism to uniformly delete Gaussian functions to prevent Gaussian functions from flooding outside pre-defined boundaries.

For this review, we did a careful literature search and found that one metadynamics-based enzyme reaction study<sup>44</sup> was published before our above study; in this work, reported by the Houk group in 2007, 1-D metadynamics sampling was employed to exam the direct decarboxylation mechanism catalyzed by the most proficient enzyme: Orotidine-5'-monophosphate Decarboxylase (ODCase). Since then, there have been only ~30 metadynamics-based enzyme reaction studies reported. Considering the popularity of metadynamics, this small number likely reflects the practical challenge in applying

metadynamics to explore enzyme reaction pathways. As mentioned earlier, recent advancement of the metadynamics method<sup>39-41</sup> will certainly lead to more successful applications. Nevertheless dimensionality limit still requires simulators' to creatively design low-dimension CVs<sup>15</sup> and often apply them in a trial-and-error manner.

## Generalized Ensemble Based String Optimization: The On-the-Path Random Walk Method

The above case study based on free energy surface sampling demonstrates the indirect reaction free energy path calculation strategy<sup>46</sup>. As an alternative, the chain-of-states (COS) path optimization strategy can be employed to directly obtain reaction pathways. In comparison with the free energy surface sampling based strategy, the path optimization strategy only requires one-dimension sampling and thus has no dimensionality limit issue; e.g. multiple candidate CVs  $\theta(X) = (\theta_1(X), \dots, \theta_m(X))$  can be employed to represent the path space. Among various COS algorithms, the string (FTS) method has attracted a great deal of attention, in particular recently for enzyme reaction mechanism studies<sup>47-55</sup>. Based on string

method, MFEP can be obtained according to the minimization criterion  $\left[ M \frac{\partial G}{\partial \theta} \right]^\perp = 0$  (M is the diffusion tensor matrix), which, in comparison with  $\left[ \frac{\partial G}{\partial \theta} \right]^\perp = 0$ , can more accurately reflect the curvilinear nature of CVs.

In common string method applications, sampling is performed on a series of non-communicating images between two pre-chosen end points  $Z^A = (z_1^A, \dots, z_m^A)$  and  $Z^B = (z_1^B, \dots, z_m^B)$ . Regarding the original string method, two sampling issues are worth noting: (a) because images are independently explored, the on-the-path continuity of the environmental degrees of freedom cannot be guaranteed; (b) the CV degrees of freedom are restricted from regular MD sampling and thus it is challenging for a being-optimized path that represents an unfavorable mechanism to be switched into the correct reaction channel. To overcome these issues, we developed a generalized ensemble sampling based string path optimization method: the on-the-path random walk (OTPRW) algorithm<sup>24,27</sup>. In OTPRW, the CV-space pathway is represented by a set of  $\lambda$ -dependent functions  $Z(\lambda) = (z_1(\lambda), \dots, z_m(\lambda))$ , in which  $Z(0) = Z^A$  and  $Z(1) = Z^B$ , where the progressing parameter  $\lambda$  is set equal to the percentage of the on-the-path distance of the corresponding state from the starting point  $Z^A$ . Dynamic propagation in OTPRW is based on the following extended-dynamics Hamiltonian,

$$H_\lambda = H_o + \frac{p_\lambda^2}{2m_\lambda} + \sum_{i=1}^m \frac{1}{2} K_i (\theta_i(X) - z_i(\lambda))^2 + f_m(\lambda), \quad (4)$$

where  $\lambda$  is treated as a one-dimension dynamic particle with a mass of  $m_\lambda$  and its momentum of  $p_\lambda$  and is propagated based on Langevin dynamics; via the

$\sum_{i=1}^m \frac{1}{2} K_i (\theta_i(X) - z_i(\lambda))^2$  term, the system is restrained on the path from the latest optimization update. Through the biasing function  $f_m(\lambda)$ , which can be adaptively obtained via either the metadynamics<sup>24</sup> or the ABF method<sup>27</sup>, the target system, instead of being constrained on non-communicating images, can randomly walk along the instantaneous path to collect samples for the following path optimization. Thereby, the structural continuity of the environmental degrees of freedom can be naturally ensured; notably at a joint image state between two pathways, the system can switch from an unfavorable pathway to a better reaction channel (Figure 3).

The OTPRW method has been successfully applied to the studies of the transformation between Chorismate and Prephenate<sup>24</sup>, where eight CVs were used, and the DNA base extrusion process<sup>27</sup>, where ten CVs were employed. In these studies, single-trajectory OTPRW simulations led to nicely converged MFEPs, which could be convincingly validated via the committor analysis. Interestingly in both studies, unexpected essential CVs were identified despite the fact that these systems had been immensely investigated. For instance, it was shown that besides RBOPs, bond breaking angles are intimately involved in the formation of the transition state between Chorismate and Prephenate<sup>24</sup>; and in the DNA base extrusion process, rather than commonly assumed base flipping, a base-plane-elongation event is responsible for the formation of the transition state and the energetic penalty at the transition state is mainly introduced by the stretching of the Watson-Crick base pair<sup>27</sup>.

## OTPRW Study of A Substrate-Assisted Glycosylation Reaction

Recently, we applied the OTPRW method to study the substrate-assisted glycosylation (SAG) reaction (Figure 4A) that is catalyzed by a  $\beta$ -Hexosaminidase protein, OfHex I. The SAG reaction involves two key chemical processes: “proton transfer” between the general acid (GluH) and glycosidic oxygen atoms and “nucleophilic substitution” around the central anomeric carbon. In this study, eight distance CVs were selected to describe the reaction path. Four of these CVs are reaction bond order parameters, corresponding to the chemical bonds directly involved in the bond forming and breaking events. To accurately describe geometrical constraints on the proton transfer process, the distance between the proton transfer donor and acceptor oxygen atoms and the distance between the to-be-transferred proton and the non-proton-donor oxygen of GluH were included in the CV set. In addition, two extra distance CVs around the nucleophilic substitution center were defined to describe possible bond formation and breaking angle changes during the reaction.

The OTPRW simulation began with a minimum energy path (MEP), along which the two chemical events are largely de-synchronized. As shown in Figure 4B, along this MEP, the proton transfer event proceeds earlier than the nucleophilic substitution event. Using the on-the-path ABF method<sup>27</sup>, we calculated the free energy profile along this initial-guess reaction path. As shown by the dotted line in Figure 4D, the “apparent” transition state corresponds to the state of  $\lambda = 0.371$ , which is right between the mid-points of the two chemical events (Figure 4B); and the overall free energy barrier is about 27.0 kcal/mol. Within 7 ns, the OTPRW simulation converged. As shown in Figure 4C, along the MFEP, the two chemical processes are precisely synchronous and the sub-events in each of the



processes are highly concerted. As shown by the solid line in Figure 4D, the transition state corresponds to the state of around  $\lambda = 0.49$ , which is also the midpoints of the two chemical processes. The corresponding free energy barrier is about 13.1 kcal/mol, which is in excellent agreement with the experimental value. This study clearly demonstrates the importance of the minimum free energy path sampling over the MEP calculation; obviously a MEP obtained based on a non-dynamic enzyme environment can be drastically different from the target MFEP. Based on this result, we strongly discourage any future attempt to perform free energy calculation along a MEP to estimate the reaction free energy barrier.

## Final Remarks

Free energy path sampling plays an essential role in computational understanding of chemical reactions, particularly ones occurring in enzymatic environments. Among a variety of molecular dynamics simulation approaches, the generalized ensemble sampling strategy is uniquely attractive for the fact that it not only can enhance the sampling of rare chemical events but also can naturally ensure consistent exploration of environmental degrees of freedom. In this review, we plan to provide a tutorial-like tour on an emerging topic: generalized ensemble sampling of enzyme reaction free energy path. The discussion is largely focused on our own studies, particularly ones based on the metadynamics free energy sampling method and the on-the-path random walk path sampling method. We hope that this mini-presentation will provide interested practitioners some meaningful guidance for future algorithm formulation and application study.

We would also like to point out that the generalized ensemble sampling strategy is still far from being adequate. Necessary human input on the pre-selection of essential collective variables is still greatly hindering enzyme reaction studies from reaching the predictive stage. In addition, there is still scarce of convincingly successful case study on enzyme reactions that couple with slow conformational transitions. Currently, we are actively working on further enriching the orthogonal space sampling scheme<sup>56-59</sup> to overcome these challenges.

## Acknowledgments

Funding support from the National Science Foundation (MCB1158284) and National Institute of Health (R01GM054403 and R01GM111886) is acknowledged. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

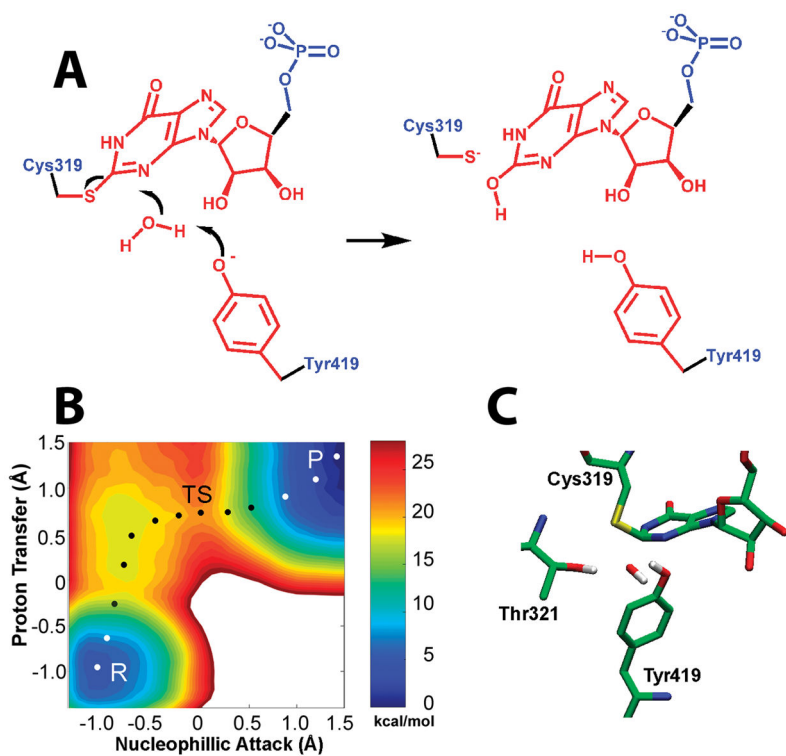
## References

1. Wigner E. The transition state method. *Trans Faraday Soc.* 1938; 34:29–41.
2. Chandler D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J Chem Phys.* 1978; 68:2959–2970.
3. Miller BG, Wolfenden R. Catalytic proficiency: The unusual case of OMP Decarboxylase. *Ann Rev Biochem.* 2002; 71:847–885. [PubMed: 12045113]
4. Gao JL, Truhlar DG. Quantum mechanical methods for enzyme kinetics. *Ann Rev Phys Chem.* 2002; 53:467–505. [PubMed: 11972016]
5. Gao JL, Ma SH, Major DT, Nam K, Pu JZ, Truhlar D. Mechanisms and free energies of enzymatic reactions. *Chem Rev.* 2006; 106:3188–3209. [PubMed: 16895324]

6. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM. Electrostatic basis for enzyme catalysis. *Chem Rev.* 2006; 106:3210–3235. [PubMed: 16895325]
7. Senn HM, Walter T. QM/MM studies of enzymes. *Curr Opin Chem Biol.* 2007; 11:182–187. [PubMed: 17307018]
8. Hu H, Yang WT. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Ann Rev Phys Chem.* 2008; 59:573–601. [PubMed: 18393679]
9. Orlando A, Jorgensen WL. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res.* 2010; 43:142–151. [PubMed: 19728702]
10. Warshel A, Levitt M. Theoretical studies of enzymic reactions – Dielectric, electrostatic and steric stabilization of carbocation-ion in reaction of lysozyme. *J Mol Biol.* 1976; 103:227–249. [PubMed: 985660]
11. Field MJ, Bash PA, Karplus M. A combine quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J Comput Chem.* 1990; 11:700–733.
12. Schmitt UW, Voth GA. Multistate empirical valence bond model for proton transport in water. *J Phys Chem B.* 1998; 102:5547–5551.
13. Voth GA. Computer simulation of proton solvation and transport in aqueous and biomolecular systems. *Acc Chem Res.* 2006; 39:143–150. [PubMed: 16489734]
14. Swanson JMJ, Maupin CM, Chen HN, Petersen MK, Xu JC, Wu YJ, Voth GA. Proton solvation and transport in aqueous and biomolecular systems: Insights from computer simulations. *J Phys Chem B.* 2007; 111:4300–4314. [PubMed: 17429993]
15. Okamoto Y. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J Mol Graphics Modell.* 2004; 22:425–439.
16. Yang W, Nymeryer H, Zhou HX, Berg BA, Brüschweiler R. Quantitative computer simulations of biomolecules: A snapshot. *J Comput Chem.* 2008; 29:668–672. [PubMed: 17708535]
17. Zuckerman DM. Equilibrium sampling in biomolecular simulation. *Annu Rev Biophys.* 2011; 40:41–62. [PubMed: 21370970]
18. Elber R, Karplus M. A method for determining reaction paths in large molecules: Application to myoglobin. *Chem Phys Lett.* 1987; 139:375–380.
19. Henkelman G, Uberuaga BP, Jonsson H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys.* 2000; 113:9901–9904.
20. EWN, Ren WQ, Vanden-Eijnden E. String method for the study of rare events. *Phys Rev B.* 2002; 66:052301.
21. EWN, Ren WQ, Vanden-Eijnden E. Finite temperature string method for the study of rare events. *J Phys Chem B.* 2005; 109:6688–6693. [PubMed: 16851751]
22. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: Minimum free energy paths and siocommittor surfaces. *J Chem Phys.* 2006; 125:024106.
23. Pan AC, Sezer D, Roux B. Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B.* 2008; 112:3432–3440. [PubMed: 18290641]
24. Chen ME, Yang W. On-the-path random walk sampling for efficient optimization of minimum free-energy path. *J Comput Chem.* 2009; 30:1649–1653. [PubMed: 19462399]
25. Dickson BM, Huang H, Post CB. Unrestrained computation of free energy along a path. *J Phys Chem B.* 2012; 116:11046–11055. [PubMed: 22816870]
26. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA.* 2002; 99:12562–12566. [PubMed: 12271136]
27. Cao L, Lv C, Yang W. Hidden conformation events in DNA base extrusions: A generalized-ensemble path optimization and equilibrium simulation study. *J Chem Theor Comput.* 2013; 9:3756–3768.
28. Alberty RA, Hammes GG. Application of the theory of diffusion-controlled reactions to enzyme kinetics. *J Phys Chem.* 1958; 62:154–159.
29. Min D, Josephine HR, Li HZ, Lakner C, MacPherson IS, Naylor GJ, ... Yang W. An enzymatic atavist revealed in dual pathways for water activation. *PLoS Biology.* 2008; 6:e206. [PubMed: 18752347]

30. Patey GN, Valleau JP. A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. *J Chem Phys.* 1975; 63:2334–2339.
31. Sprik M, Ciccotti G. Free energy from constrained molecular dynamics. *J Chem Phys.* 1998; 109:7737–7744.
32. Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain people. *J Comput Chem.* 1997; 18:1450–1462.
33. Darve E, Pohorille A. Calculating free energies using average force. *J Chem Phys.* 2001; 115:9169–9183.
34. Kirkwood JG. Statistical mechanics of fluid mixtures. *J Chem Phys.* 1935; 3:300–313.
35. Carter EA, Ciccotti G, Hynes JT, Kapral R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem Phys Lett.* 1989; 156:472–477.
36. Huber T, Torda AE, van Gunsteren WF. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol Des.* 1994; 8:695–708. [PubMed: 7738605]
37. Min D, Liu Y, Carbone I, Yang W. On the convergence improvement in the metadynamics simulations: A Wang-Landau recursion approach. *J Chem Phys.* 2007; 126:194104. [PubMed: 17523795]
38. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett.* 2008; 100:020603. [PubMed: 18232845]
39. Dama JF, Parrinello M, Voth GA. Well-tempered metadynamics converges asymptotically. *Phys Rev Lett.* 2014; 112:240602. [PubMed: 24996077]
40. Dama JF, Rotskoff J, Parrinello M, Voth GA. Transition-tempered metadynamics: Robust, convergent metadynamics via on-the-fly transition barrier estimation. *J Chem Theor Comput.* 2014; 10:3626–3633.
41. Dama JF, Hocky GM, Sun R, Voth GA. Exploring valleys without climbing every peak: More efficient and forgiving metabasin metadynamics via robust on-the-fly bias domain restriction. *J Chem Theor Comput.* 2015; 11:5638–5650.
42. Ensing B, Klein ML. Perspective on the reactions between  $F^-$  and  $CH_3CH_2F$ : The free energy landscapes of the E2 and  $SN_2$  reaction channels. *Proc Natl Acad Sci USA.* 2005; 102:6755–6759. [PubMed: 15863622]
43. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, ... Karplus M. CHARMM: The biomolecular simulation program. *J Comp Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
44. Stanton CL, Kuo IFW, Mundy CJ, Laino T, Houk KN. QM/MM metadynamics study of the direct decarboxylation mechanism for orotidine-5'-monophosphate decarboxylase using two different QM regions: Acceleration too small to explain rate of enzyme catalysis. *J Phys Chem B.* 2007; 111:12573–12581. [PubMed: 17927240]
45. McCullagh M, Saunders MG, Voth GA. Unraveling the mystery of ATP hydrolysis in actin filaments. *J Am Chem Soc.* 2014; 136:13053–13058. [PubMed: 25181471]
46. Ensing B, Laio A, Parrinello M, Klein ML. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J Phys Chem B.* 2005; 109:6676–6687. [PubMed: 16851750]
47. 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. [PubMed: 21539371]
48. Lans I, Medina M, Rosta E, Hummer G, Garcia-Viloca M, Lluch JM, Gonzalez-Lafont A. Theoretical study of the mechanism of the hydride transfer between Ferredoxin-NADP(+) reductase and NADP(+): The role of Tyr303. *J Am Chem Soc.* 2012; 134:20544–20553. [PubMed: 23181670]
49. Zinovijev K, Ruiz-Pernia JJ, Tunon I. Toward an automatic determination of enzymatic reaction mechanisms and their activation free energies. *J Am Chem Soc.* 2013; 9:3740–3749.
50. Aranda J, Zinovjev K, Roca M, Tunon I. Dynamics and reactivity in thermophilic N6-adenine methyltransferase. *J Am Chem Soc.* 2014; 136:16227–16239. [PubMed: 25347783]

51. Khavrutskii IV, Legler PM, Friedlander AM, Wallqvist A. A reaction path study of the catalysis and inhibition of the bacillus anthracis CapD gamma-glutamyl transpeptidase. *Biochemistry*. 2014; 53:6954–6967. [PubMed: 25334088]
52. Kumari M, Kozmon S, Kuhanek P, Stepan J, Tvaroska I, Koca J. Exploring reaction pathways for O-GlcNAc transferase catalysis. *J Phys Chem B*. 2015; 119:4371–4381. [PubMed: 25731954]
53. Zhang SX, Ganguly A, Goyal P, Bingaman JL, Bevilacqua PC, Hammes-Schiffer S. Role of the active site Guanine in the glmS ribozyme self-cleavage mechanism: Quantum mechanical/molecular mechanical free energy simulations. *J Am Chem Soc*. 2015; 137:784–798. [PubMed: 25526516]
54. Aranda J, Zinovjev K, Roca M, Tunon I. Dynamics and reactivity in thermos aquaticus N6-adenine methyltransferase. *J Am Chem Soc*. 2014; 136:16227–16239. [PubMed: 25347783]
55. Sanchez-Martinez M, Field M, Crehuet R. Enzymatic minimum free energy path calculations using swarms of trajectories. *J Phys Chem B*. 2015; 119:1103–1113. [PubMed: 25286154]
56. Zheng LQ, Chen ME, Yang W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Natl Acad Sci U S A*, 2008. 2008; 105:20227–20232.
57. Zheng LQ, Chen ME, Yang W. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *J Chem Phys*. 2009; 130:234105. [PubMed: 19548709]
58. Zheng LQ, Yang W. Practically efficient and robust free energy calculations: Double-integration orthogonal space tempering. *J Chem Theor Comput*. 2012; 8:810–823.
59. Lv C, Aitchison EW, Wu D, Zheng LQ, Cheng XL, Yang W. Comparative exploration of hydrogen sulfide and water transmembrane free energy surfaces via orthogonal space tempering free energy sampling. *J Comput Chem*. 2016; 37:567–574. [PubMed: 26119423]

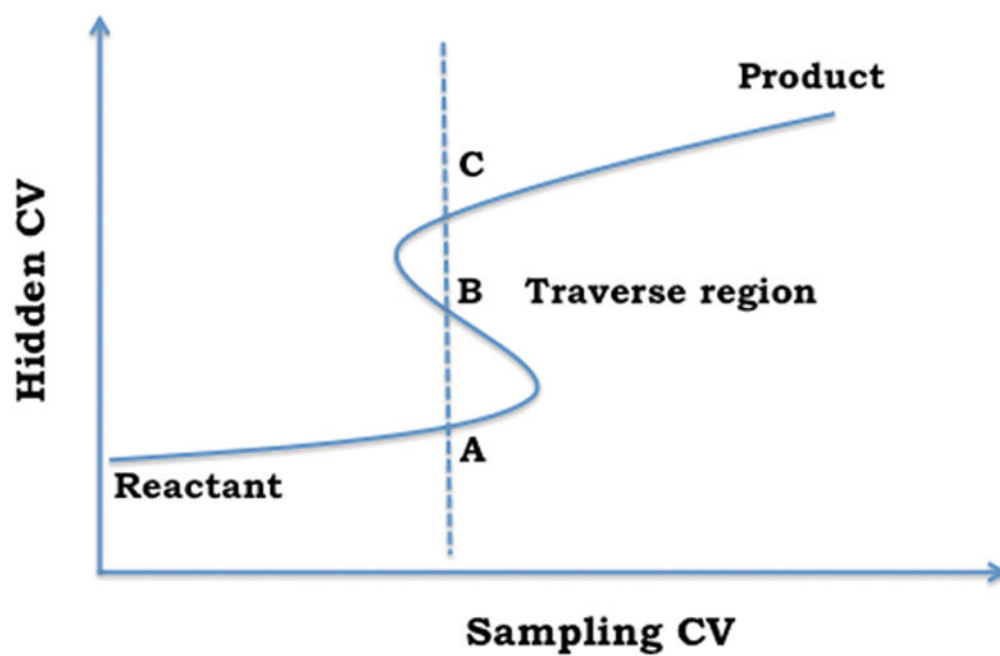


**Figure 1. Wang-Landau (flat-histogram) metadynamics simulation of the hydrolysis step in the IMPDH Arg418Gln variant. Figure 1 was originally published in PLoS Biology (open-access, doi:10.1371/journal.pbio.0060206.g003)**

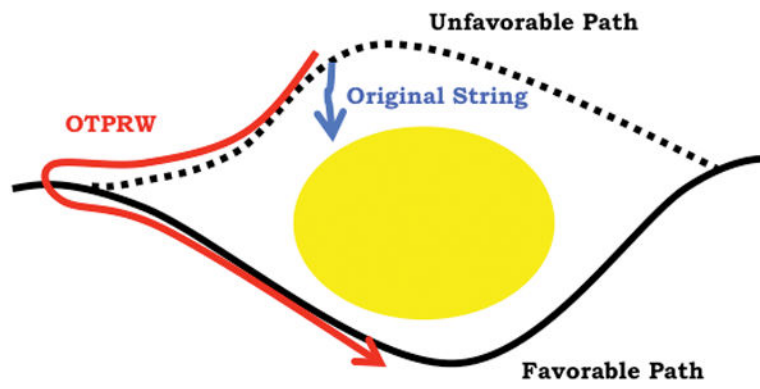
(A) The proposed mechanism on the hydrolysis of E-XMP\* with Tyr419 acting as the general base.

(B) The free energy landscape of the Tyr419 pathway in the Ar418Gln variant. P: product; R: reactant; and TS: transition state.

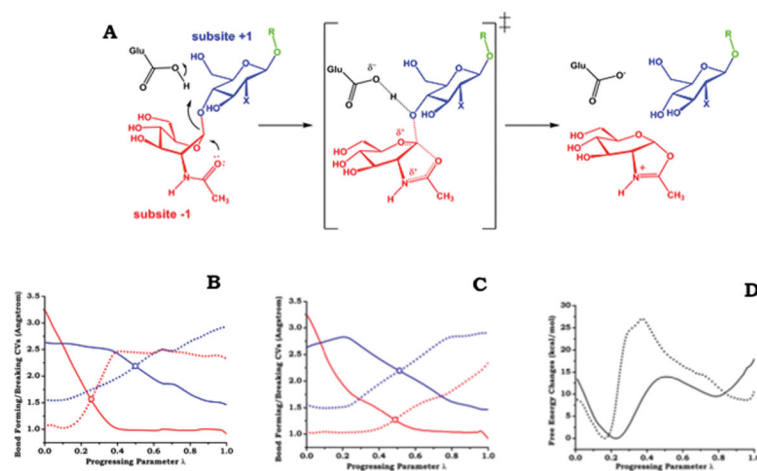
(C) The corresponding transition state structure.



**Figure 2.**  
The schematic illustration of the hidden CV issue.



**Figure 3.**  
The schematic illustration of the path switching mechanism in OTPRW.



**Figure 4. The OTPRW simulation of the substrate-assisted glycosylation reaction in OfHex1**  
 (A) The proposed mechanism on the substrate-assisted glycosylation reaction in OfHex1.  
 (B) The chemical order parameters changes along the initial minimum energy path. Red: The CVs for the proton transfer process. Blue: The CVs for the nucleophilic substitution process.  
 (C) The chemical order parameters changes along the OTPRW optimized minimum free energy path. Red: The CVs for the proton transfer process. Blue: The CVs for the nucleophilic substitution process.  
 (D) The free energy changes along the initial minimum energy path (the dotted line) and along the OTPRW optimized minimum free energy path (the solid line).