Steered Molecular Dynamics Methods Applied to Enzyme Mechanism and Energetics

C.L. Ramírez*, M.A. Martí*, A.E. Roitberg†,1
*FCEN, UBA, Buenos Aires, Argentina
†University of Florida, Gainesville, FL, United States
1Corresponding author: e-mail address: roitberg@ufl.edu

Contents

1. Introduction 2
1.1 Computer Simulation of Chemical Reactions: The Birth of QM/MM Methods 2
1.2 Reaction Coordinate, Biased Sampling, and Jarzynski's Relationship 4
1.3 Quasi Reversible FE Profile of the Chorismate to Prephenate Conversion Reaction in Solution 7
1.4 FE Profiles for the Chorismate to Prephenate Conversion in Solution Using MSMD and Jarzynski's Relationship 7
1.5 The Hybrid Differential Relaxation Algorithm 10
1.6 HyDRA Scheme Improves Both Barrier and Reaction Free Energies Accuracy Determinations in MSMD–JR Simulations 13
2. Discussion 17
3. Methodology 18
3.1 Simulation Details 18
References 18

Abstract

One of the main goals of chemistry is to understand the underlying principles of chemical reactions, in terms of both its reaction mechanism and the thermodynamics that govern it. Using hybrid quantum mechanics/molecular mechanics (QM/MM)-based methods in combination with a biased sampling scheme, it is possible to simulate chemical reactions occurring inside complex environments such as an enzyme, or aqueous solution, and determining the corresponding free energy profile, which provides direct comparison with experimental determined kinetic and equilibrium parameters. Among the most promising biasing schemes is the multiple steered molecular dynamics method, which in combination with Jarzynski's Relationship (JR) allows obtaining the equilibrium free energy profile, from a finite set of nonequilibrium reactive trajectories by exponentially averaging the individual work profiles. However, obtaining

Methods in Enzymology
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2016.05.029

statistically converged and accurate profiles is far from easy and may result in increased computational cost if the selected steering speed and number of trajectories are inappropriately chosen. In this small review, using the extensively studied chorismate to prephenate conversion reaction, we first present a systematic study of how key parameters such as pulling speed, number of trajectories, and reaction progress are related to the resulting work distributions and in turn the accuracy of the free energy obtained with JR. Second, and in the context of QM/MM strategies, we introduce the Hybrid Differential Relaxation Algorithm, and show how it allows obtaining more accurate free energy profiles using faster pulling speeds and smaller number of trajectories and thus smaller computational cost.

1. INTRODUCTION

1.1 Computer Simulation of Chemical Reactions: The Birth of QM/MM Methods

Understanding the underlying principles of chemical reactions is one of the main goals of chemistry. At the atomic and molecular level a reaction is understood in terms of the reaction pathways that describe it, and the thermodynamics that govern it. In other words, a chemical reaction is described in terms of the atomic motions that drive the system from the reactant to the product state, and the free energy profile along the reaction path. Experimentally, chemical dynamics is inferred from chemical kinetic studies that help in defining a reaction mechanism; ie, a scheme showing all bond-breaking/making steps, including proposals for the transition states, and relative free energy values of the key involved states. To move beyond the mechanistic proposal and really understand the process at an atomic level, one has to inevitably employ quantum mechanics (QM)-based molecular dynamics (MD) simulations.

QM methods in chemistry were a natural extension of early 20th century atomic physics moving to larger molecules, and gained momentum with the developments of Hartree–Fock (HF), post HF, and density functional theory (DFT)-based methods, as well as the birth of modern electronics. However, although computer power has been ever increasing, they are still computationally demanding and are restricted—in their pure form—to simulations of a few hundred atoms and MD runs for few picoseconds. This drawback presents itself as a barrier to true understanding, since most interesting reactions occur in condensed phases, like aqueous solvent, or inside enzymes, which requires simulations of systems of thousands of atoms.

Enzyme reactions are particularly interesting, since enzymes are not only exceptional catalysts, achieving rate enhancement factors of up to 1×10^{15} times but also are highly substrate and reactant/product specific. To really understand an enzyme mechanism, we need to describe not only how the free energy changes as the atoms move along the reaction (reaction mechanism), but also how the protein framework lowers the transition state (TS) free energy, with respect to the same reaction mechanism occurring in solution. In other words, how—structurally and energetically speaking—do enzymes achieve their rate enhancement.

To overcome the computational cost problem, while nonetheless being able to simulate a reaction inside an enzyme or in bulk solvent, recent nobel laureates Warshel, Levitt, and Karplus (Warshel & Karplus, 1972; Warshel & Levitt, 1976) introduced in the 1970s the concept of multiscale modeling, in particular the hybrid quantum mechanical/molecular mechanical (QM/MM) method. QM/MM methods allow an accurate description of the enzyme active site (or the solute) where the chemical reaction takes place, which is described at the QM level of theory, while considering the environment using a less expensive MM—or classical type of—force field level of theory. Key to these methods is the coupling between the QM and MM regions, which must properly describe the environment electrostatic as well as the steric effects of the environment on the reactive subsystem.

Among the most successful QM/MM coupling schemes are the additive ones, which as shown by Eqs. (1) and (2), add an additional coupling term ($H_{QM/MM}$) to the Hamiltonian. This term is composed by an electrostatic term that computes the interaction between the QM electrostatic potential and the classical point charges, and a Van der Waals term that computes steric interaction between QM and MM atoms, usually using a Lennard–Jones type of potential. Most important is that in the QM Hamiltonian (H_{QM}), the potential exerted by the classical point charges is self-consistently added to the nuclei potential an thus the solved electronic structure corresponds to the actual hybrid QM/MM potential.

$$H_{TOT} = H_{QM} + H_{MM} + H_{QM-MM} \tag{1}$$

$$H_{QM-MM} = \sum_{i=1}^{C} q_i \int \frac{\rho(r)}{|r - \tau_i|} dr + \sum_{i=1}^{C} \sum_{j=1}^{Q} \frac{q_i Z_j}{|R_j - \tau_i|} + E_{QM-MM}^{LJ} \tag{2}$$

where the sum in i is for all the classical particles (C) and in j is for the quantum particles (Q), q_i are the punctual charges in the MM region, $\rho(r)$ is the

charge density of the QM region, τ_i the position of the classical nuclei, and R_j the position of the quantum nuclei.

Although QM/MM methods allow a description of the dynamics of large molecular systems and the study of chemical reactions, in plain MD simulations at room temperature, the system is unable to cross even moderately high barriers ($>kT$), such as those presented by enzyme reactions, thus remaining trapped in the initial (reactant) state, unless driven up the energy hill. Therefore, to simulate chemical dynamics process the system must be driven along the reaction.

1.2 Reaction Coordinate, Biased Sampling, and Jarzynski's Relationship

To drive any chemical system along a selected pathway, first a proper reaction coordinate (RC) needs to be defined. The RC consists of a combination of structural parameters, usually distances (or angles) between atoms, that provide a description of the reaction progress and that therefore can be used to force the system along. Proper sampling of the system along the desired reaction coordinate allows obtaining the corresponding free energy (FE) profile—also called potential of mean force—which describes how the free energy changes along the reaction, providing the corresponding barrier ($\Delta G^{\ddagger}$) and reaction free energy (ΔG^{o}), both of which can be directly compared with experimental data. Moreover, by comparing the FE profiles obtained with different RC or varying conditions, alternative mechanistic proposals can be analyzed.

The key for obtaining accurate, and thus meaningful, FE profiles is to properly sample configurational space. The first, and most commonly used method, to achieve this is umbrella sampling (US) (Kumar, Rosenberg, Bouzida, Swendsen, & Kollman, 1992). In US, a harmonic potential is added to the system hamiltonian, and its equilibrium position is varied along the RC in discrete windows, until the whole RC width is evenly sampled. The resulting FE profile is obtained using Boltzmann weighting of the sampled configurations with respect to the RC. More recently, other strategies, such as metadynamics (Laio & Parrinello, 2002), adaptive biasing force (Hénin & Chipot, 2004), free energy perturbation (Zwanzig, 1954), and orthogonal space random walk (Zheng, Chen, & Yang, 2008), have been developed and applied with a varying degree of success.

Jarzynski (1997) published an outstanding work, where he showed that the free energy—an equilibrium property—could be obtained from

nonequilibrium dynamics, according to what has been called since then, Jarzynski's relationship (JR), shown in Eq. (3).

$$G(\lambda) = -\beta^{-1} \ln \left\langle e^{-\beta W_i(\lambda)} \right\rangle_0 \tag{3}$$

where $G(\lambda)$ represents the free energy as a function of the reaction coordinate $\lambda, \beta^{-1} = k_B T$, where k_B is Boltzmann's constant, and T is the system temperature, W_i represents the (nonreversible) work performed to drive the system along the RC, and brackets denote configurational average, where the subscript zero denotes that the initial structures of the ensemble are equilibrated for $\lambda = 0$.

In other words, the JR shows that for any system driven externally from an initial equilibrium state to a final state, at any speed, the free energy between them can nonetheless be obtained if the works of many individual trajectories are exponentially averaged (Crooks, 1998, 2000). A clear bottleneck—as will be shown later—is that the exponential average needs to be properly converged.

The JR provides a very simple strategy to obtain FE profile from steered molecular dynamics (SMD) simulations. First, an external force is added to the system, usually, in the form of a harmonic potential whose equilibrium position moves along the chosen RC at a given velocity (v), according to Eq. (4). The external force thus steers the system along the RC moving it from reactants to products.

$$F(t) = -k(\lambda - \lambda_0 - vt) \tag{4}$$

$$w(\lambda) = \int_0^{\lambda} F(\lambda) d\lambda \tag{5}$$

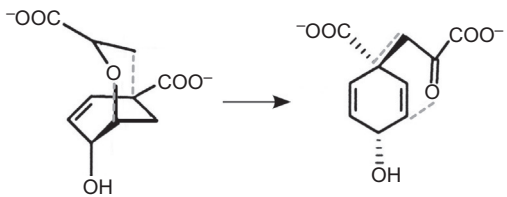
Individual steered trajectories are started from equilibrium configurations, and for each, the corresponding work profile along the RC is computed by simply integrating the external force. Finally, multiple works are exponentially averaged using JR to obtain the corresponding FE profile. The method, which is also called multiple steered molecular dynamics (MSMD), was initially used by Shulten and coworkers to study alanine decapeptide folding and compared with US (Park, Khalili-Araghi, Tajkhorshid, & Schulten, 2003), showing similar accuracy and potential for a reduction in computational cost. Moreover, in over a decade, it has been applied to a variety of systems and received multiple computational and experimental validations (Collin et al., 2005; Liphardt, Dumont,

Smith, Tinoco, & Bustamante, 2002; Saira et al., 2012; Xiong, Crespo, Marti, Estrin, & Roitberg, 2006).

Our group was among the first to use the MSMD approach and JR in QM/MM simulations to study an enzymatic reaction, in this case the chorismate to prephenate conversion as performed by *Bacillus subtilis* chorismate mutase (BsCM) (Crespo, Marti, Estrin, & Roitberg, 2005).

This particular reaction is involved in the biosynthesis of aromatic amino acids in bacteria, fungi, and plants, and there is extensive experimental and theoretical data in the literature. The enzyme catalyzes an intramolecular Claisen rearrangement. Kast et al. (1997) reported the experimental activation parameters for the enzyme-catalyzed reaction, $\Delta G^{\ddagger} = 15.4\,\text{kcal/mol}$, $\Delta H^{\ddagger} = 12.7\,\text{kcal/mol}$ and $\Delta S^{\ddagger} = -9.1\,\text{kcal/mol}$. Mulholland's group (Ranaghan et al., 2004), Lee Woodcock and coworkers (Lee Woodcock et al., 2003; Woodcock, Hodoscek, & Brooks, 2007), and Crespo (Crespo et al., 2005) have reported theoretical activation parameters for chorismate mutase, at different theory levels.

The reaction can be nicely described and driven forward using as RC a combination of the distance between the carbons dC–C that create the new bond and the distance of the bond that is broken dO–C (Scheme 1). We had previously computed the same reaction profile using minimum energy path, which allows obtaining the potential energy instead of free energy, using the level of theory DFT with the PBD Exchange Correlation functional. Although the obtained value of 5.3 kcal/mol is below the experimental determination of 12.7 kcal/mol, we correctly predicted the energy difference related to the catalytic enhancement, since the barrier in aqueous solvent was 8.5 kcal/mol higher (Crespo et al., 2003). With the MSMD method we were able to obtain a well-converged FE profile performing 30 runs, with a 2 Å/ps speed and a RC span of ca. 4 Å, totaling 60 ps of simulation time. The results were the same as those



Scheme 1 Reaction from chorismate to prephenate. In *gray dotted lines* are the distances dC–C and dO–C that conform the reaction coordinate.

obtained with US using 12 windows of 5 ps each, thus the same overall computational time (Crespo et al., 2005).

Over the last decade there has been an intense debate as to whether the MSMD strategy, combined with JR, really provides a significant reduction in computational cost—yielding nonetheless accurate FE profiles—beyond the simple fact that each nonequilibrium trajectory can be run in an independent computer and thus the method is easily parallelized. Also it is not clear which of the available biased sampling method provides the best FE profile at the lowest cost. We have decided to calculate a reference set which can be considered as converged in terms of FE profile, so we can then benchmark changes in several parameters affecting the JR convergence. For this purpose, we have used the density functional tight binding (DFTB) level of theory as the QM method to determine the chorismate to prephenate conversion reaction in aqueous solution.

1.3 Quasi Reversible FE Profile of the Chorismate to Prephenate Conversion Reaction in Solution

Prior to the analysis of the impact on the FE profile of the pulling speed and the number of SMD employed, we created a reference FE profile, in order to have a comparison between the different profiles. This was achieved by performing three SMD at a pulling speed of 0.02 Å/ps. Both values were chosen because they were the ones that showed a difference, between work profiles, in less than 1 kcal/mol, which points to a quasi-static path. With this result we were able to obtain a FE profile, employing JR, that was used as the reference profile for the reaction in solution (bold line in Fig. 1A and B).

1.4 FE Profiles for the Chorismate to Prephenate Conversion in Solution Using MSMD and Jarzynski's Relationship

Fig. 1A shows the typical behavior of the work profiles obtained for a finite set of MSMD trajectories, as well as their relation with the corresponding FE profile obtained with JR and the reference FE profile. The figure nicely shows how the width of the distribution of the work values increases as the reaction moves toward products. Also the JR estimate of the FE profile increasingly overestimates the reference FE profile. Both these effects are a consequence of the fact that the further the system is driven forward by the external force, the further it moves from equilibrium. The distribution of work values is wider as one moves away from equilibrium and as a consequence, the exponential average is harder to converge and is shifted to larger values, thus overestimating the free energy. This is a consequence of the

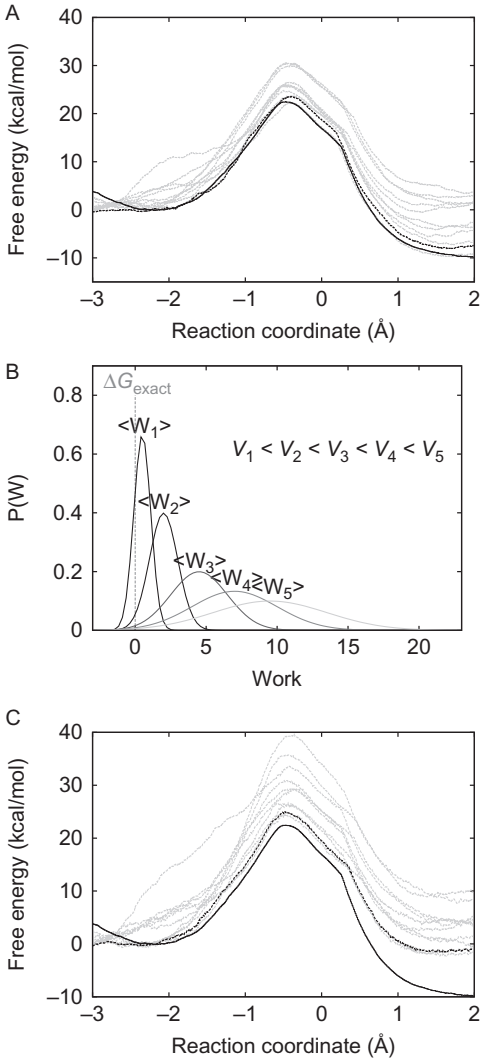


Fig. 1 (A) Work profiles (*dotted gray lines*) used to obtain a FE profile with JR (*dotted black line*), and a comparison with the reference profile employed (*continuous black line*) for the reaction of chorismate to prephenate in solution. All *dotted lines* correspond to a pulling speed of 0.5 Å/ps, while the reference has a puling speed of 0.02 Å/ps. (B) Gaussian distribution of work for different pulling speeds shown in a schematic form. It can be seen that the slower the pulling speed, the narrower the work distribution. (C) Same notation as in panel (A), but for a pulling speed of 1 Å/ps.

mathematical nature of the exponential average as can easily be seen for a Gaussian like work distribution, as shown schematically in Fig. 1B. Fig. 1B also shows why it is very important to have at least one trajectory whose work profile is close to the reference FE profile, since otherwise the exponential average will be overestimated. Indeed, Pohorille, Jarzynski, and Chipot (2010) estimated that if the standard deviation of the work distribution is around $1\,k_BT$, about one in six trajectories samples values close to the real free energy. The value decreases to 1 in 40 trajectories for $\sigma = 2k_BT$ and to 1/300,000 for $\sigma = 5k_BT$, strongly biasing the obtained FE profile if only a limited number of trajectories is performed. In other words, to have a well-converged exponential average, the works that need to be sampled are those in the lower tail of the work distribution which therefore have low occurrence probabilities.

To understand the relationship between the width of the work distribution and the pulling speed, it is interesting to analyze the data from Fig. 1C, where the work profiles, at twice the pulling speed as those shown in Fig. 1A, are shown together with the reference and obtained FE profile. The results clearly show how faster pulling results in wider distributions, which are also shifted to higher values (see also Fig. 1B) and therefore a significant overestimation of the free energy. On the contrary, extremely slow pulling speed would result in a distribution of work values which resemble a delta function, and thus the exponential average, the average, and the external work all display the same values (within the $1\,k_BT$) and thus equal the free energy, as required by the second law of thermodynamics.

Having analyzed the role of MSMD key parameter, which is the pulling speed, we will now analyze how it translates into the accuracy of the obtained FE profile, in relation to the number of trajectories and therefore the overall relative computational cost. The accuracy will be measured in terms of the overestimation of both the barrier and reaction free energy, and the total computational cost in relation to that required to obtain the reference FE profile. The results are presented in Table 1.

The results show that, for this particular system, the number of trajectories have less impact in the final FE profile that the pulling speed. This can be appreciated by the fact that the same computational cost is needed to obtain a FE profile with 20 trajectories at a pulling speed of 1 Å/ps as for one obtained using 10 trajectories at a pulling speed of 0.5 Å/ps, but the second one is more accurate.

In summary, when using MSMD combined with JR to obtain FE profile profiles, it is important to consider that since the exponential average is

Table 1 Relevant Values of the FE (in kcal/mol) Profile for the Reaction from Chorismate to Prephenate in Solution for Different Pulling Speeds (in Å/ps) and Number of Trajectories

Pulling Speed	Trajectories	$\Delta G^{\ddagger}$	ΔG^{o}	Total time	Relative Time
0.02	3	21.0	−10.0	787.5	100
0.5	10	24.6	−7.0	105.0	13.3
1.0	5	25.1	−0.6	26.3	3.3
1.0	10	25.9	−0.9	52.5	6.6
1.0	15	25.6	−0.9	78.8	10.0
1.0	20	25.2	−1.6	105.0	13.3
2.0	10	33.8	8.4	26.3	3.3

The total time (in ps) is calculated considering the time performed per simulation, multiplied for the number of trajectories performed. The relative time (in %) is calculated considering that the total time employed in performing the reference profile is 100%.

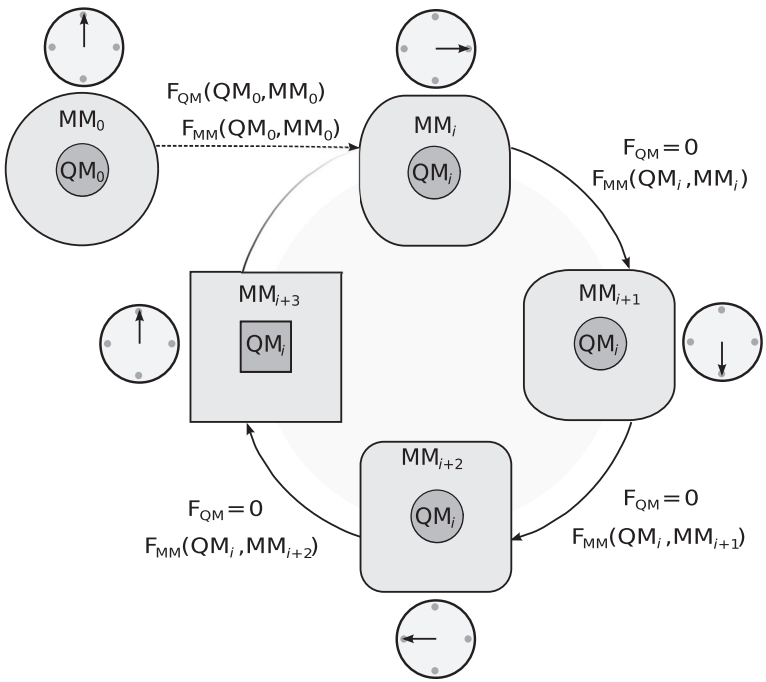
computed over a work values distribution, the more narrow (smaller SD) the more accurate the FE profile. The width of the distribution is directly proportional to the pulling speed and increases also as the system moves forward and thus further from equilibrium. For wider distributions an exponentially increasing number of trajectories is needed to sample values in the lower tail of the distribution, which are the ones required to obtain accurate FE profile. Therefore, the main drawback with JR is that in order to obtain a well-converged average, and thus accurate FE profile, either very large number of trajectories and/or very low pulling speeds are needed. These facts results in a high computational cost (only marginally smaller than that required for quasi-static transitions) and even sometimes in an insurmountable problem that prevents accurate convergence of the FE profile.

1.5 The Hybrid Differential Relaxation Algorithm

While the SMD strategy is formally correct, there are some serious convergence problems that can hinder its wider applicability. Several methods have been presented to deal with this these issues, mostly in the context of classical force fields (MM) (Ozer, Valeev, Quirk, & Hernandez, 2010). The differential relaxation strategy was inspired by the multiple time step ideas proposed by Tuckerman, Berne, and Martyna (1992) and Tuckerman and Parrinello (1994) which use different masses and time steps for the QM and MM regions. These early studies showed that while similar trajectories

as those in standard MD simulations are obtained, faster (about two times) convergence of the average force can be achieved. Taken this idea forward we implemented a similar multiple time step scheme, without mass scaling, that in the context of MSMD and JR allows obtaining accurate FE profiles at a significant lower computational cost.

The method works as follows (Scheme 2), after a joint QM/MM step where both QM and MM atoms are moved a time step forward, the QM region is frozen while the MM region is allowed to move (relax) for a given number of pure classical steps, after which a new joint QM/MM step is performed. Key to hybrid differential relaxation algorithm (HyDRA) success is that the work from the external force is only accumulated in the QM/MM steps, since in the pure MM steps the reaction coordinate—which should be part only of the QM region—remains frozen and thus no force is applied. The pure classical steps allow the MM region to relax to the perturbed QM region, and therefore during the joint steps—which are the ones that matter for the FE profile—closer to equilibrium work values are obtained. In the language of the JR, only the steps where work is accumulated



Scheme 2 Hybrid Differential Relaxation Algorithm scheme for a DRAr of 4.

(the QM/MM steps) count toward the free energy estimator. The steps where the QM region is frozen are heat (not work), and hence, they are irrelevant to the free energy calculation. The ratio between the MM and the joint QM/MM steps is called the differential relaxation algorithm ratio (DRAr). A DRAr ratio of 1 is equivalent to conventional QM/MM simulations, while with a DRAr of 10, for each joint step 10 pure classical relaxation steps are performed.

Technically implementing HyDRA is straight forward for any additive QM/MM Hamiltonian (Eq. 1) since it relies on the adequate calculation of the forces and atom coordinates for each type of step. When the simulation starts (step 0) both regions are synchronized and force contributions are computed as usual, thus forces acting on the QM atoms stem from the quantum contribution in the step 0 configuration and the QM/MM interaction contribution derived from the Lennard–Jones potential (third term in Eq. 2), which we will refer as $F_{QM}(QM_0)$ and $F_{QM}(MM_0)$. Similar reasoning applies to the forces acting for this step on the MM atoms, $F_{MM}(QM_0)$ and $F_{MM}(MM_0)$. With these forces the system is able to perform its first dynamic step, which leads to conformation *i* for both QM and MM regions. Now the DRA cycle starts. The pure classical and QM/MM forces acting on the MM atoms are computed as usual for conformation *i*, and now only the MM region moves a time step forward achieving conformation $i+1$. The QM region remains frozen in conformation *i*—which is achieved by zeroing the forces acting on the QM atoms. $F_{QM}(MM)$ forces are nonetheless stored since they will be used later. With new MM coordinates (configuration $i+1$) a new set of forces $F_{MM}(QM_i, MM_{i+1})$ acting on them is computed and the MM region is updated. After three of these steps the systems arrives at configuration $[QM_i, MM_{i+3}]$. Now the systems performs a new joint step to reach conformation $[QM_{i+1}, MM_{i+4}]$. The forces for this step are determined using the $i+3$ conformation for the MM contributions, and the *i* conformation for the quantum contributions. Considering the structures, it is important to note that in our implementation, only the conformations obtained after a joint step are meaningful and equivalent to those obtained in a standard QM/MM molecular dynamics simulations.

The method has been implemented in the AMBER (Case et al., 2012) computer simulation package and is available upon request to the author. HyDRA works with any of the available QM methods and was thoroughly tested with DFTB (Cui, Elstner, Kaxiras, Frauenheim, & Karplus, 2001; de M Seabra, Walker, Elstner, Case, & Roitberg, 2007) and pure DFT

implementation of the Perdew, Burke, and Ernzerhof (Perdew, Burke, & Ernzerhof, 1996) exchange correlation functional (DFP-PBE) (Nitsche, Ferreria, Mocskos, & González Lebrero, 2014) available under request to the author.

1.6 HyDRA Scheme Improves Both Barrier and Reaction Free Energies Accuracy Determinations in MSMD–JR Simulations

To test the potential of the HyDRA we first analyzed how increasing the DRAr affects the shape and accuracy of the obtained FE profiles for the chorismate to prephenate conversion at a fixed pulling speed (in this case 100 times faster than that required to obtain the reference FE profile) both in solution and in BsCM using 10 trajectories. The results are presented in Fig. 2 and summarized in Table 2.

The results show that the standard MSMD (DRAr = 1) significantly overestimates both the barrier and reaction free energy. In particular, in

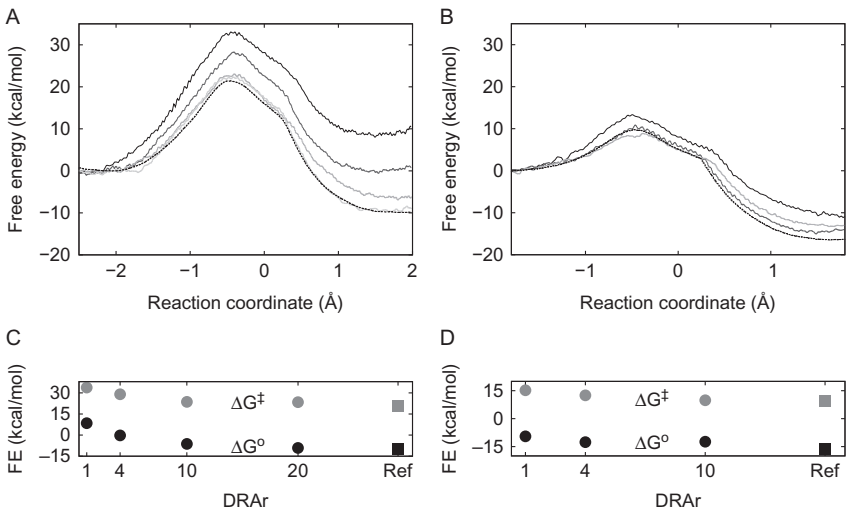


Fig. 2 FE profiles for the chorismate to prephenate reaction, obtained through 10 non-equilibrium dynamics with a pulling speed of 2 Å/ps, and reference FE profile. (A) Free energy profiles for the reaction in solution with a DRA ratio of 1, 4, 10, and 20 shown in scale of *gray*, decreasing color when the DRAr increases. (B) FE profiles in the BsCM enzyme, employing a DRAr of 1, 4, and 10, with the same color code as in (A). (C) $\Delta G^{\ddagger}$ (*gray*) and ΔG^{o} (*black*) for the FE profile shown in panel (A). *Circles* represent values for FE profile with different DRAr, and the *squares* are the values for the reference profile. (D) $\Delta G^{\ddagger}$ (*gray*) and ΔG^{o} (*black*) for the reaction in the enzyme.

Table 2 Relevant Values of the FE Profile (in kcal/mol) for the Reaction from Chorismate to Prephenate in Solution and in the Enzyme BsCM

	FE Profile	DRAr	$\Delta G^{\ddagger}$	ΔG^{o}
Solution	Reference	1	21	−10
	MSMD–JR	1	33.8	8.4
		4	29	−0.2
		10	23.6	−6.3
		20	23.4	−9.1
BsCM	Reference	1	9.6	−16.5
	MSMD–JR	1	15.2	−9.5
		4	12.5	−12.6
		10	9.9	−12.4

The pulling speed in all the cases is 2 Å/ps, and the number of trajectories is 10.

solution, plain MSMD yields a positive reaction free energy, when the reference value is −10 kcal/mol and overestimates the barrier in more than 10 kcal/mol. The HyDRA scheme significantly improves the FE profile accuracy, and the higher the DRAr, the closer the estimate to the reference FE profile. For a DRAr of 10, the obtained values for the barriers are overestimated in less than 10% of the barrier magnitude, and the free energy yields the correct trend. In general, the effect of HyDRA is more notorious for the reaction in solution than in the enzyme. This is not unexpected since the enzyme's active site is less flexible compared to the solvent and thus requires less relaxation along the reaction.

As mentioned earlier, one of the key aspects of enzyme reactions is to determine the catalytic efficiency, which corresponds to the change in the reaction free energy barrier for the reaction in the enzyme with respect to the observed in aqueous solution, ie, the $\Delta\Delta G^{\ddagger}$. For BsCM the experimental determined $\Delta\Delta G^{\ddagger}$ value is 9.1 kcal/mol, making BsCM a quite efficient enzyme. Using the reference FE profiles we obtain an efficiency of 11.4 kcal/mol for BsCM which thus is close to the experimental value. Plain MSMD yields a significantly overestimated $\Delta\Delta G^{\ddagger}$ value (18.6 kcal/mol) which HyDRA partially corrects, yielding a $\Delta\Delta G^{\ddagger}$ value of 13.5 kcal/mol.

To understand the effect of the HyDRA scheme in the context of MSMD–JR, and the opening discussion, it is interesting to analyze the work distributions obtained for increasing values of DRAr. Fig. 3 shows the work

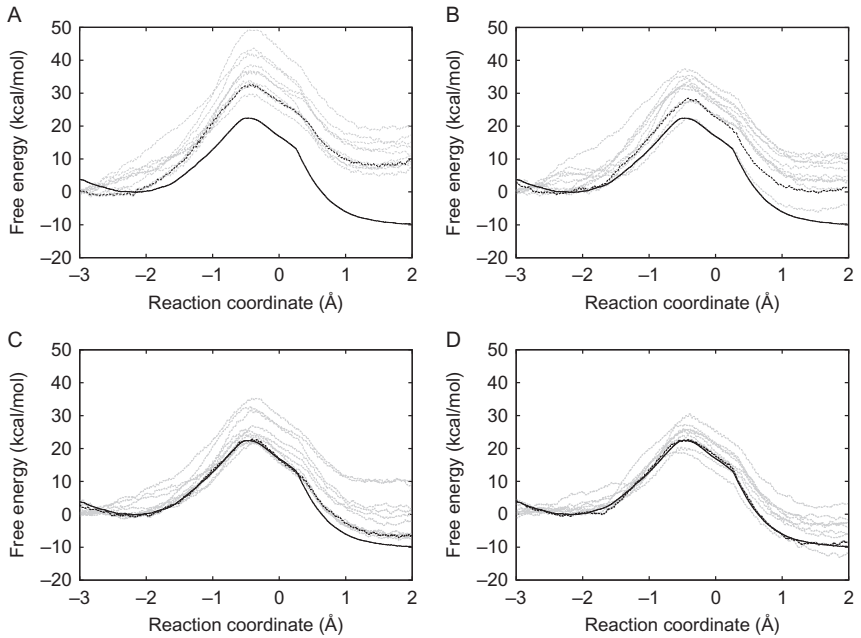


Fig. 3 Work distributions (in *gray*), Jarzynski's estimator (*dotted black line*), and reference FE profile (*solid black line*) for the chorismate to prephenate reaction in solution. In all cases the pulling speed is 2 Å/ps, and the number of trajectories is 10. Panels (A)–(D) correspond to a DRAr of 1, 4, 10, and 20, respectively.

distributions (in gray) for 10 trajectories with DRAr of 1,4, 10, and 20 for the reaction in solution, together with the Jarzynski's estimator (dotted black line) and the reference FE profile (solid black line). The results show clearly how with increasing DRAr the width of the work distribution becomes more narrow, and therefore, the Jarzynski estimate is closer to the reference profiles. Indeed for DRAr of 1 and 4 sampling is clearly inadequate since there are not any work profiles that sample the correct free energy value in the barrier and later regions. For DRAr of 10 there are already several profiles close to the reference FE profile in the transition state region, while for DARr of 20 there is even one case that goes below it. As already discussed, being able to sample these trajectories is essential for getting accurate FE profiles.

Another interesting point of analysis is related to how the HyDRA scheme impacts in the pulling speed and number of runs required to obtain accurate FEP. Table 3 shows how the obtained barrier and reaction free energy values depend on the pulling speed and number of trajectories using

Table 3 Relevant Values of the FE Profile Obtained Through JR for the Reaction of Chorismate to Prephenate in Solution

	Pulling Speed	Number of Trajectories	$\Delta G^{\ddagger}$		ΔG^{o}	
Reference FE profile	0.02	3	21		−10	
DRAr			1	10	1	10
Effect of pulling speed	0.5	10	24	22	−7	−8.1
	1	10	25.9	21.6	−0.9	−9.5
	2	10	33.8	23.6	8.4	−6.3
Effect of the number of trajectories	1	5	25.1	21.1	−0.6	−9.9
	1	10	25.9	21.6	−0.9	−9.5
	1	15	25.6	21.1	−0.9	−9.9
	1	20	25.2	21.1	−1.6	−9.9

the HyDRA scheme (using DRAr of 10). The results show that HyDRA effect on both parameters is dramatic, without HyDRA only using the slowest tests pulling speed (25 times faster than that required for the reference FEP) reasonable value for the reaction free energy and the barrier is slightly overestimated, HyDRA not only shields more accurate results at these slow speed, but it yields reasonable results even for a pulling speed which is 100 times faster than that used to obtain the reference FEP, where standard MSMD FEP is completely at odds. Concerning the number of trajectories, little difference is observed for all tested values. The results from standard MSMD are not accurate enough (especially for the reaction free energy) which possibly means that many more (in the order of hundred) simulations are required to obtain better FEPs. For HyDRA, however, the results with only five trajectories already yield accurate results, which are within 1 kcal/mol with respect to the reference FEP for both the barrier and reaction free energy.

In summary, the results presented earlier clearly show that the HyDRA scheme allows, in the context of MSMD–JR free energy profile determinations, to obtain more accurate profiles using significantly faster pulling speeds and smaller number of trajectories compared to standard QM/MM MD simulations. Moreover, the results show by comparing FE profile of the reactions in solution and enzyme accurate estimates of the enzyme catalytic power can be determined.

2. DISCUSSION

Since the introduction of Jarzynki's relationship in 1996 many theoretical studies (Crooks, 1998, 2000; Ozer et al., 2010) have taken advantage of its simplicity and easy parallelization to obtain free energy profiles of different processes at the molecular level. However, the declaimed reduction in computational cost in comparison to other schemes, such as US, and the accuracy of the obtained FE profile was more a matter of faith, than a fact. This is specially so in QM/MM simulations of enzymatic reactions where the complexity of the systems usually prevents the application of different schemes for comparative purposes, and where often no experimental accurate value for direct comparison with the FE profile is available.

The presented HyDRA scheme not only provides a method which has the potential of delivering more accurate FE profiles at significant lower computational cost but also highlight some interesting facts about the relation between the pulling speed, the number of trajectories, the resulting work profiles, and the accuracy of the resulting FE profiles. The present data add up to our previous experience in the subject and clearly show that it is very difficult to obtain accurate FE profile when the pulling speed is too fast. More important, too fast speed cannot be corrected by performing more trajectories (unless possible huge numbers are considered) and that usually small set—in the order of 10—is enough to have a good estimate of the work distribution. Therefore, when performing standard MSMD–JR FE profile estimates, the work value distribution should display deviations in the order of $k_B T$ to obtain meaningful results. HyDRA allows the system to remain closer to the equilibrium during the pulling steps—which are the joint QM/MM steps—and thus more narrow work distributions and more accurate FE profiles are obtained.

Although the method was only recently developed, during the last year we applied it to study the reaction of two essential Zn hydrolases of *Mycobacterium tuberculosis*, namely, MshB (Rv1170) and MA-amidase (Rv3717) (Romero, Martin, Ramirez, Dumas, & Marti, 2015). Both enzymes hydrolyze the difficult C–N bond of an amide. Our results allowed obtaining of accurate FE profiles for both reaction steps, the first involving Zn coordinated hydroxide nucleophilic attack to the scissile amide carbonyl that leads to meta-estable tetrahedral intermediate. And the second which involves breaking of the C–N bond and is promoted by active site base to amide nitrogen proton transfer. The obtained barriers and mechanism were similar

to those reported for other Zn hydrolases like carboxypeptidase-A (Xu & Guo, 2009), thermolysis (Blumberger, Lamoureux, & Klein, 2007), angiotensin converting enzyme (Zhang, Wu, & Xu, 2013), and the anthrax lethal factor peptidase (Smith, Smith, Yang, Xu, & Guo, 2010) pointing toward a conserved (or converged) mechanism for these type of enzymes which belong to different protein families. Most important, analysis of the required QM/MM steps for obtaining the corresponding FE profiles in each case shows that HyDRA scheme resulted in 5–20 times potential reduction in the overall computational cost, highlighting the tremendous power of this HyDRA strategy.

3. METHODOLOGY

3.1 Simulation Details

For pure classical equilibration simulations and the MM part in the QM–MM dynamics all classical parameters were taken from the Amber force field, ff99SB (Hornak et al., 2006), and TIP3P model was used for the water molecules (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). Parameters for the chorismate were taken from previous works from our group (Crespo et al., 2003, 2005). Simulations were performed at constant temperature using the Berendsen thermostat as implemented in SANDER, using periodic boundary conditions and either constant volume or constant pressure. Ewald Sums were used to treat long-range electrostatics.

For the QM–MM simulations the DFTB (Cui et al., 2001; de M Seabra et al., 2007) as implemented in the sander module of AMBER (Case et al., 2012) was used.

REFERENCES

Blumberger, J., Lamoureux, G., & Klein, M. L. (2007). Peptide hydrolysis in thermolysin: Ab initio QM/MM investigation of the Glu143-assisted water addition mechanism. *Journal of Chemical Theory and Computation*, *3*(5), 1837–1850. http://dx.doi.org/10.1021/ct7000792.

Case, D. A., Darden, T. A., Cheatham, T. E. I. I. I., Simmerling, C. L., Wang, J., Duke, R. E., … Kollman, P. A. (2012). *AMBER 12*. San Francisco: University of California.

Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., & Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, *437*(7056), 231–234. http://dx.doi.org/10.1038/nature04061.

Crespo, Alejandro, Marti, M. A., Estrin, D. A., & Roitberg, A. E. (2005). Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *Journal of the American Chemical Society*, *127*(19), 6940–6941. http://dx.doi.org/10.1021/ja0452830.

Crespo, A., Scherlis, D. A., Martí, M. A., Ordejón, P., Roitberg, A. E., & Estrin, D. A. (2003). A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase. *Journal of Physical Chemistry. B*, *107*(49), 13728–13736.

Crooks, G. E. (1998). Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, *90*(5–6), 1481–1487. http://dx.doi.org/10.1023/A:1023208217925.

Crooks, G. E. (2000). Path-ensemble averages in systems driven far from equilibrium. *Physical Review E*, *61*(3), 2361–2366. http://dx.doi.org/10.1103/PhysRevE.61.2361.

Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T., & Karplus, M. (2001). A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *The Journal of Physical Chemistry. B*, *105*(2), 569–585. http://dx.doi.org/10.1021/jp0029109.

de M Seabra, G., Walker, R. C., Elstner, M., Case, D. A., & Roitberg, A. E. (2007). Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package. *The Journal of Physical Chemistry. A*, *111*(26), 5655–5664. http://dx.doi.org/10.1021/jp070071l.

Hénin, J., & Chipot, C. (2004). Overcoming free energy barriers using unconstrained molecular dynamics simulations. *The Journal of Chemical Physics*, *121*(7), 2904–2914. http://dx.doi.org/10.1063/1.1773132.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, *65*(3), 712–725. http://dx.doi.org/10.1002/prot.21123.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*(14), 2690–2693. http://dx.doi.org/10.1103/PhysRevLett.78.2690.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. Retrieved from, http://www.scopus.com/inward/record.url?eid=2-s2.0-0004016501&partnerID=40&md5=7af48df275648024bbd5981b9583c129.

Kast, P., Tewari, Y. B., Wiest, O., Hilvert, D., Houk, K. N., & Goldberg, R. N. (1997). Thermodynamics of the conversion of chorismate to prephenate: Experimental results and theoretical predictions. *The Journal of Physical Chemistry. B*, *101*(50), 10976–10982. http://dx.doi.org/10.1021/jp9725011.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, *13*(8), 1011–1021. http://dx.doi.org/10.1002/jcc.540130812.

Laio, A., & Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12562–12566. http://dx.doi.org/10.1073/pnas.202427399.

Lee Woodcock, H., Hodošček, M., Sherwood, P., Lee, Y. S., Schaefer, H. F., III, & Brooks, B. R. (2003). Exploring the quantum mechanical/molecular mechanical replica path method: A pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theoretical Chemistry Accounts*, *109*(3), 140–148. http://dx.doi.org/10.1007/s00214-002-0421-3.

Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I., & Bustamante, C. (2002). Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science (New York, N.Y.)*, *296*(5574), 1832–1835. http://dx.doi.org/10.1126/science.1071152.

Nitsche, M. A., Ferreria, M., Mocskos, E. E., & González Lebrero, M. C. (2014). GPU accelerated implementation of density functional theory for hybrid QM/MM simulations. *Journal of Chemical Theory and Computation*, *10*, 959–967.

Ozer, G., Valeev, E. F., Quirk, S., & Hernandez, R. (2010). Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide Y. *Journal of Chemical Theory and Computation*, *6*(10), 3026–3038. http://dx.doi.org/10.1021/ct100320g.

Park, S., Khalili-Araghi, F., Tajkhorshid, E., & Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics*, *119*(6), 3559. http://dx.doi.org/10.1063/1.1590311.

Perdew, J., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, 77(18), 3865.

Pohorille, A., Jarzynski, C., & Chipot, C. (2010). Good practices in free-energy calculations. *The Journal of Physical Chemistry. B*, *114*(32), 10235–10253. http://dx.doi.org/10.1021/jp102971x.

Ranaghan, K. E., Ridder, L., Szefczyk, B., Sokalski, W. A., Hermann, J. C., & Mulholland, A. J. (2004). Transition state stabilization and substrate strain in enzyme catalysis: Ab initio QM/MM modelling of the chorismate mutase reaction. *Organic & Biomolecular Chemistry*, *2*(7), 968–980. http://dx.doi.org/10.1039/b313759g.

Romero, J. M., Martin, M., Ramirez, C. L., Dumas, V. G., & Marti, M. A. (2015). Efficient calculation of enzyme reaction free energy profiles using a hybrid differential relaxation algorithm: Application to mycobacterial zinc hydrolases. *Advances in Protein Chemistry and Structural Biology*, *100*, 33–65. http://dx.doi.org/10.1016/bs.apcsb.2015.06.006.

Saira, O.-P., Yoon, Y., Tanttu, T., Möttönen, M., Averin, D. V., & Pekola, J. P. (2012). Test of the Jarzynski and Crooks fluctuation relations in an electronic system. *Physical Review Letters*, *109*(18), 180601. http://dx.doi.org/10.1103/PhysRevLett.109.180601.

Smith, C. R., Smith, G. K., Yang, Z., Xu, D., & Guo, H. (2010). Quantum mechanical/molecular mechanical study of anthrax lethal factor catalysis. *Theoretical Chemistry Accounts*, *128*(1), 83–90. http://dx.doi.org/10.1007/s00214-010-0765-z.

Tuckerman, M., Berne, B. J., & Martyna, G. J. (1992). Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, *97*(3), 1990. http://dx.doi.org/10.1063/1.463137.

Tuckerman, M. E., & Parrinello, M. (1994). Integrating the Car–Parrinello equations. II. Multiple time scale techniques. *The Journal of Chemical Physics*, *101*(2), 1316. http://dx.doi.org/10.1063/1.467824.

Warshel, A., & Karplus, M. (1972). Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *Journal of the American Chemical Society*, *94*(16), 5612–5625. http://dx.doi.org/10.1021/ja00771a014.

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, *103*, 227–249.

Woodcock, H. L., Hodoscek, M., & Brooks, B. R. (2007). Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms. *The Journal of Physical Chemistry. A*, *111*(26), 5720–5728. http://dx.doi.org/10.1021/jp0714217.

Xiong, H., Crespo, A., Marti, M., Estrin, D., & Roitberg, A. E. (2006). Free energy calculations with non-equilibrium methods: Applications of the Jarzynski relationship. *Theoretical Chemistry Accounts*, *116*(1–3), 338–346. http://dx.doi.org/10.1007/s00214-005-0072-2.

Xu, D., & Guo, H. (2009). Quantum mechanical/molecular mechanical and density functional theory studies of a prototypical zinc peptidase (carboxypeptidase A) suggest a general acid-general base mechanism. *Journal of the American Chemical Society*, *131*(28), 9780–9788. http://dx.doi.org/10.1021/ja9027988.

Zhang, C., Wu, S., & Xu, D. (2013). Catalytic mechanism of angiotensin-converting enzyme and effects of the chloride ion. *The Journal of Physical Chemistry. B*, *117*(22), 6635–6645. http://dx.doi.org/10.1021/jp400974n.

Zheng, L., Chen, M., & Yang, W. (2008). Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20227–20232. http://dx.doi.org/10.1073/pnas.0810631106.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, *22*(8), 1420. http://dx.doi.org/10.1063/1.1740409.