JCTC Journal of Chemical Theory and Computation

Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy

Donghong Min,[†] Lianqing Zheng,[†] William Harris,[‡] Mengen Chen,[†] Chao Lv,[‡] and Wei Yang*,[†,‡]

Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306

Received January 19, 2010

Abstract: The difference between free energy changes occurring at two chemical states can be rigorously estimated via alchemical free energy (AFE) simulations. Traditionally, most AFE simulations are carried out under the classical energy potential treatment; then, accuracy and applicability of AFE simulations are limited. In the present work, we integrate a recent second-order generalized ensemble strategy, the orthogonal space random walk (OSRW) method, into the combined quantum mechanical/molecular mechanical (QM/MM) potential based AFE simulation scheme. Thereby, within a commonly affordable simulation length, accurate QM/MM alchemical free energy simulations can be achieved. As revealed by the model study on the equilibrium of a tautomerization process of hydrated 3-hydroxypyrazole and by the model calculations of the redox potentials of two flavin derivatives, lumichrome (LC) and riboflavin (RF) in aqueous solution, the present OSRW-based scheme could be a viable path toward the realization of practically efficient QM/MM AFE simulations.

I. Introduction

In the recent decades, the alchemical free energy (AFE) simulation approach[1−12] has evolved to be a key technique of predicting free energy changes associated with various chemical and biophysical events, such as solvation, protein−ligand binding, protein−protein binding, protonation/deprotonation, electron transfer, etc. AFE simulations allow rigorous estimation of the difference between free energy changes occurring at two chemical states via calculations of free energy values along alchemical directions rather than physical directions. For instance, the solvation free energy difference between two chemical species A and B [$\Delta\Delta G = \Delta G_{\text{gas}\rightarrow\text{solution}}(\text{B}) - \Delta G_{\text{gas}\rightarrow\text{solution}}(\text{A})$] can be obtained as the difference of the alchemical free energy changes in two target environments [$\Delta\Delta G = \Delta G_{\text{solution}}(\text{A}\rightarrow\text{B}) - \Delta G_{\text{gas}}(\text{A}\rightarrow\text{B})$].[3] Thereby, visits of the phase regions that could be challenging to sample, such as the ones associated with gas-solution interface transitions, are naturally avoided. In an AFE simulation, an expanded ensemble is usually constructed so as to allow end chemical states (for instance, the states A and B) to be connected in the same phase space; for the above purpose, in a prebuilt hybrid potential energy function,

$$U_{\text{o}} = U_{\text{s}}(\lambda) + U_{\text{e}} \tag{1}$$

the constraints of $U_{\text{s}}(0) = U_{\text{s}}^{\text{A}}$ and $U_{\text{s}}(1) = U_{\text{s}}^{\text{B}}$ should be set to realize the energy terms (U_{s}^{A} and U_{s}^{B}) unique in two end chemical states (respectively represented by $\lambda = 0$ and $\lambda = 1$). In eq 1, the common environmental energy terms shared by two end states are labeled as U_{e}. One of the simplest forms of eq 1 is the linear function

$$U_{\text{o}} = (1 - \lambda)U_{\text{s}}^{\text{A}} + \lambda U_{\text{s}}^{\text{B}} + U_{\text{e}} \tag{2}$$

On the basis of a constructed hybrid potential energy function, molecular simulations (molecular dynamics simulations or Monte Carlo simulations) are required to collect

* Corresponding author e-mail: yyang2@fsu.edu.
[†] Institute of Molecular Biophysics.
[‡] Department of Chemistry and Biochemistry.

10.1021/ct100033s
Published on Web 07/08/2010

samples for the estimation of the free energy difference between two end states A ($\lambda = 0$) and B ($\lambda = 1$); it should be noted that specific procedures of AFE simulations vary with the employed free energy simulation algorithms.[13−35]

In AFE simulations, target systems are commonly described molecular mechanically (MM). As generally known, classical force fields are limited in treating intricate molecular interactions in terms of accuracy and applicability; for instance, redox potential calculations, tautomerization free energy predictions, and metal-associated ligand binding affinity estimations are classical examples. Therefore, developing AFE simulation methods[23,35−59] in the combined quantum mechanical and molecular mechanical (QM/MM) paradigm, where the interaction center is treated quantum mechanically and the rest are treated classically,[60−64] is a natural next step with the evolution of the field. Despite almost two decades' development, due to the apparent efficiency issue, there are only a few nontrivial AFE simulations that have been performed under the QM/MM treatment. Very recently, more attention has been paid to the practical issue on how to improve AFE simulation techniques so as to achieve converged QM/MM AFE simulations using commonly available computing resources.[23,50−54,56−59] Synergistic with our general interest in improving free energy simulation efficiency, we have been actively exploring the above issue in the framework of the generalized ensemble (GE) sampling.[64−67] Particularly motivated by a recent development, the orthogonal space random walk (OSRW) algorithm,[34,68] which can lead to significant efficiency improvements in both free energy simulation and conformational sampling, we generalized the usage of the OSRW strategy in QM/MM AFE simulations, and the presently reported studies reveal the fact that this OSRW strategy could be a viable path to the realization of practically efficient QM/MM AFE simulations.

In the present paper, the OSRW-based QM/MM AFE simulation method is introduced. Its efficiency is demonstrated by a model study on the equilibrium of a tautomerization process of hydrated 3-hydroxypyrazole and a model calculation of the redox potentials of two flavin derivatives: lumichrome (LC) and riboflavin (RF) in aqueous solution.

II. Theoretical Designs

II.A. Brief Introduction to QM/MM AFE Simulation Methods. For the purpose of calculating the free energy difference $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ between two chemical states (for instance, A and B), both of which are described by the QM/MM potential, there have been two schemes of the related method developments: the "direct" scheme and the "indirect" scheme.

In the "direct" scheme, each QM/MM AFE calculation is performed via alchemical changes directly from the starting state $A_{QM/MM}$ to the ending state $B_{QM/MM}$. Here, alchemical transitions can be realized directly through mechanical switching, where alchemical mixing is implemented on the basis of two independent QM/MM electronic structural calculations as reflected in the following equation:

$$U_o = (1 - \lambda)U_s^{A,QM/MM} + \lambda U_s^{B,QM/MM} + U_e \qquad (3)$$

where U_s^A and U_s^B in eq 2 are described by the corresponding QM/MM energy terms $U_s^{A,QM/MM}$ and $U_s^{B,QM/MM}$; the other common environmental energy terms shared by two end states are labeled as U_e. As a comparison, if alchemical mixing is implemented at the level of the QM/MM electronic structural description,[42,55] where intermediate states are defined in electronic structural Hamiltonians, simulation efficiency could be doubled for the fact that, with electronic structural switching, only one QM/MM force calculation is needed at each time step. The recent fractional electron strategy is a promising example of such a design.[55] The "direct" scheme allows rigorous calculation of AFE values when two end states share a similar molecular structure, for instance, in the redox potential calculations.[36,37,44,46,49−51,55,57−59] When two end states possess different numbers of atoms (even the same number of atoms but with distinct structures), the chaperoned approach has been proposed to realize the "direct" scheme.[44,47−49,52,59] Here, two sets of structures are required so that for each of the end states, there is one structure to take care of the physical nature of the corresponding chemical state, while there is a second one to maintain the structural integrity of the vanishing atoms. It should be noted that when two end states are chemically distinct, it is extremely challenging to rigorously apply the "direct" scheme due to the numerical singularity problem that occurs in the energy derivative calculations at the atom-annihilation end state; because of the complication of the QM/MM treatment and the numerical singularity problem, in this situation, end point contributions can only be empirically estimated via an extrapolation strategy. Clearly, the "direct" scheme is more appropriate for the redox potential or the excitation free energy[69] type of calculations, where two end states have very similar nucleus configurations but different electronic configurations; in this case, eq 3 can be directly employed. Notably, in pK_a calculations, if the deprotonated states do not interact strongly with the surrounding environment, an effective approximation is to keep the van der Waals interactions associated with the acidic proton in QM/MM free energy simulations and then annihilate the acidic proton in a separate set of simulations.[70]

In comparison with the "direct" scheme, the "indirect" scheme,[38−41,43,52,54] which was pioneered by Gao et al., can robustly deal with the cases where target end states are drastically different. In the "indirect" scheme, the target free energy difference $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ is typically calculated through three steps which are respectively responsible for estimating $\Delta G_{A,MM\rightarrow A,QM/MM}$, $\Delta G_{A,MM\rightarrow B,MM}$, and $\Delta G_{B,MM\rightarrow B,QM/MM}$; thereafter, $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ can be calculated via the following equation:

$$\Delta G_{A,QM/MM\rightarrow B,QM/MM} = \Delta G_{B,MM\rightarrow B,QM/MM} - \Delta G_{A,MM\rightarrow A,QM/MM} + \Delta G_{A,MM\rightarrow B,MM} \qquad (4)$$

The advantage of the "indirect" scheme lies in the fact that QM/MM calculations are only involved in the transitions between two resolutions of energy function descriptions on the same molecule, such as the one based on the hybrid potential function

$$U_o = (1 - \lambda)U_s^{A,MM} + \lambda U_s^{A,QM/MM} + U_e \quad (5)$$

where the other common energy terms shared by two end states are labeled as U_e, for the calculation of $\Delta G_{A,MM\rightarrow A,QM/MM}$, and then possible atom annihilation or large chemical configuration change is only required in the $\Delta G_{A,MM\rightarrow B,MM}$ calculation, where the end point singularity problem can be readily taken care of by the soft-core potential treatment.[71−73] Clearly, the "indirect" scheme is more appropriate for the free energy calculations where two end states have different chemical configurations.

II.B. The Sampling Issue in QM/MM AFE Simulations. Due to the fact that QM/MM force calculations are time-demanding, how to achieve decent free energy convergence within a short simulation length has been a bottleneck problem in the QM/MM AFE simulation method development. The sampling issue in QM/MM AFE simulations is by nature the same as that in classical AFE simulations. It concerns two interrelated aspects:[35,74] (1) how to efficiently collect samples to fill the phase space gap between two target states and (2) how to accurately collect samples to achieve certain ensemble averages as required by the employed free energy theory. The former aspect is usually called the "overlap sampling" issue, and the latter aspect is usually called the "conformational sampling" issue. Apparently, the difficulty level of an AFE simulation depends on (1) how well the phase regions of two target chemical states overlap and (2) how rough the energy landscape is in the region connecting two target states. Specifically, in the "direct" scheme of QM/MM AFE simulations (for instance, in the redox potential calculations), the sampling challenge lies in the fact that in the environmental portion (for instance, solvent or protein), nontrivial structural reorganization might occur between chemical state transitions. In the "indirect" scheme of QM/MM AFE simulations, any discrepancy between the employed MM and QM/MM potentials could lead to nontrivial environmental reorganization; it is particularly true when "generic" or any fast-made MM parameter is utilized, although theoretically, employed MM intermediate states in eq 5 should not influence the target free energy difference $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$, provided that adequate sampling is ensured.

Recently, the sampling issue in QM/MM AFE simulations has attracted several research efforts.[23,50−54,56−59] Most of these developments were carried out in the framework of the generalized ensemble sampling.[23,52−54] Up to now, two specific strategies have been used to realize the first-order generalized ensemble based QM/MM AFE simulations; one is the replica exchange based strategy,[52,54] and the other is the simulated scaling based strategy.[23,53] In the simulated scaling strategy, a biasing potential (or a biasing weight function) is employed as in the following equation:

$$U_m = U_o + f_m(\lambda) = U_s(\lambda) + U_e + f_m(\lambda) \quad (6)$$

where $f_m(\lambda)$ is targeted as $-G_o(\lambda)$; here, $G_o(\lambda)$ is the λ-dependent free energy profile in the canonical ensemble with U_o as the potential energy function. The simulated scaling strategy requires λ to be dynamically coupled with the system motions; for this purpose, we can employ either the hybrid Monte Carlo based methods[21,22] or the λ-dynamics approach.[19,20] As shown in our recent works, the simulated scaling strategy can alleviate the above sampling problems in QM/MM AFE simulations in both the "direct" scheme,[23] where the hybrid potential is constructed according to eq 3, and the "indirect" scheme,[52] where the hybrid potential as described by eq 4 is used.

As discussed above, the sampling issue in QM/MM AFE simulations largely occurs in the phase region orthogonal to the order parameter λ, namely, the environmental portion; to ensure efficient free energy convergence, environmental relaxation needs to be synergistically sampled with the λ move. From this point of view, the first-order generalized ensemble treatment is limited because of its sole focus on the removal of explicit free energy barriers along λ. Recently, the concept of a second-order generalized ensemble was introduced to additionally remove the hidden free energy barriers that are responsible for slow environmental relaxation. In this novel scheme, the developed orthogonal space random walk (OSRW) technique has shown intriguing capability in both free energy simulation and general conformational sampling.[35,68] In the present work, we are adopting the OSRW strategy to deal with the sampling issue in QM/MM AFE simulations. The OSRW method and its generalization for QM/MM AFE simulation are detailed in II.C.

II.C. The Orthogonal Space Random Walk Based QM/MM AFE Simulation Method. As discussed in our previous work, on the top of the first-order generalized ensemble treatment [namely, with a biasing energy term $f_m(\lambda)$], a second-order generalized ensemble simulation method can be designed with one more biasing energy term $F_m[\lambda,h(\lambda)]$ so as to further flatten free energy surfaces along λ and $h(\lambda)$. Ideally, for every target λ state, $h(\lambda)$ represents the corresponding order parameter that describes necessary structural relaxations in the space orthogonal to the λ direction; thereby, required environmental sampling can be accelerated. In our earlier works,[35] $\partial U_o/\partial\lambda$ was identified as such a function form; this order parameter function is particularly robust for the systems in which orthogonal space structural transitions strongly couple with the move of the order parameter λ. The second-order generalized ensemble method, where the "hidden barrier" order-parameter function $h(\lambda)$ is represented as $(\partial U_o)/(\partial\lambda)$, was named as the orthogonal space random walk (OSRW) algorithm.

Accordingly, in the OSRW-based QM/MM AFE simulations, the target energy function should be

$$U_m = U_o + f_m(\lambda) + F_m\left[\lambda, \frac{\partial U_o}{\partial\lambda}\right] = U_s(\lambda) + U_e + f_m(\lambda) + F_m\left[\lambda, \frac{\partial U_o}{\partial\lambda}\right] \quad (7)$$

where U_o is the employed QM/MM alchemical hybrid potential based on either eq 3 (in the direct scheme) or eq 5 (in the indirect scheme), $f_m(\lambda)$ is recursively updated toward $-G_o(\lambda)$ [$G_o(\lambda)$ is the λ-dependent free energy profile in the canonical ensemble with U_o as the potential energy function], and $F_m(\lambda,\partial U_o/\partial\lambda)$ is recursively updated toward $-G'_o(\lambda,\partial U_o/\partial\lambda)$

[$G'_o(\lambda,\partial U_o/\partial\lambda)$ is the $(\lambda,\partial U_o/\partial\lambda)$ space free energy profile corresponding to the canonical ensemble with $U_o - G_o(\lambda)$ as the potential energy function].

To obtain the biasing energy terms, $f_m(\lambda)$ [toward $-G_o(\lambda)$] and $F_m(\lambda,\partial U_o/\partial\lambda)$ [toward $-G'_o(\lambda,\partial U_o/\partial\lambda)$ (eq 7), the same simulation design as in the original OSRW algorithm is employed here. First, the dynamic coupling of the scaling parameter λ with the rest of the system is realized via the λ-dynamics treatment,[19,20] e.g., the scaling parameter λ is treated as a one-dimension particle moving between 0 and 1, as commonly applied in the extended Hamiltonian framework.[75] Second, the recursion procedure, by which $f_m(\lambda)$ and $F_m(\lambda,\partial U_o/\partial\lambda)$ are adaptively achieved, has two components: a "recursion kernel" to estimate $F_m(\lambda,\partial U_o/\partial\lambda)$ directly from simulations and a "recursion slave" to estimate $f_m(\lambda)$ based on $F_m(\lambda,\partial U_o/\partial\lambda)$.

Specifically on the "recursion kernel", we employ a similar recursion strategy as the one in the metadynamics method.[76,77] $F_m(\lambda,\partial U_o/\partial\lambda)$ is obtained as the sum of a relatively small Gaussian-shaped repulsive potential:

$$h\exp\left(-\frac{|\lambda-\lambda(t_i)|^2}{2w_1^2}\right)\exp\left(-\frac{\left|\frac{\partial U_o}{\partial\lambda}-\frac{\partial U_o}{\partial\lambda}(t_i)\right|^2}{2w_2^2}\right) \quad (8)$$

which is centered at $[\lambda(t_i),(\partial U_o/\partial\lambda)(t_i)]$, and thereby discourages the system from often visited configurations. Here, t_i represents the time step when the ith update is scheduled. With this procedure repeated, the overall biasing potential

$$F_m\left(\lambda,\frac{\partial U_o}{\partial\lambda}\right)=\sum_{t_i}h\exp\left(-\frac{|\lambda-\lambda(t_i)|^2}{2w_1^2}\right)\times \exp\left(-\frac{\left|\frac{\partial U_o}{\partial\lambda}-\frac{\partial U_o}{\partial\lambda}(t_i)\right|^2}{2w_2^2}\right) \quad (9)$$

builds up and eventually flattens the underlying curvature of the target free energy surface. Then, like in traditional metadynamics simulations,[76] the free energy profile along the orthogonal space $(\lambda,\partial U_o/\partial\lambda)$ can be estimated as $-F_m(\lambda,\partial U_o/\partial\lambda)$. Thus, for any state λ', the free energy profile along its generalized force space can be estimated as $-F_m(\lambda',(\partial U_o/\partial\lambda)|_{\lambda'})$; correspondingly, the energy derivative probability should be proportional to $\exp[\beta F_m(\lambda',(\partial U_o/\partial\lambda)|_{\lambda'})]$. Then, the free energy derivative, as the energy derivative ensemble average at the state λ', should be estimated as

$$\left.\frac{\partial G_o}{\partial\lambda}\right|_{\lambda'}=\left\langle\frac{\partial U_o}{\partial\lambda}\right\rangle_{\lambda'} =\frac{\int_{\frac{\partial U_o}{\partial\lambda}}\frac{\partial U_o}{\partial\lambda}\exp\left\{\beta\left[F_m\left(\lambda,\frac{\partial U_o}{\partial\lambda}\right)\right]\right\}\delta(\lambda-\lambda')}{\int_{\frac{\partial U_o}{\partial\lambda}}\exp\left\{\beta\left[F_m\left(\lambda,\frac{\partial U_o}{\partial\lambda}\right)\right]\right\}\delta(\lambda-\lambda')} \quad (10)$$

Following the thermodynamic integration formula,[13] the free energy change between the initial state with λ_i, which is the lower bound of the scaling parameter range, and any target state with the scaling parameter λ can be estimated as a function of λ:

$$G_o(\lambda)=\int_{\lambda_i}^{\lambda}\left.\frac{\partial G_o}{\partial\lambda}\right|_{\lambda'}d\lambda' \quad (11)$$

Therefore, eqs 9 and 10 can act as the "recursion slave" to obtain $-G_o(\lambda)$, which is the recursion target of $f_m(\lambda)$.

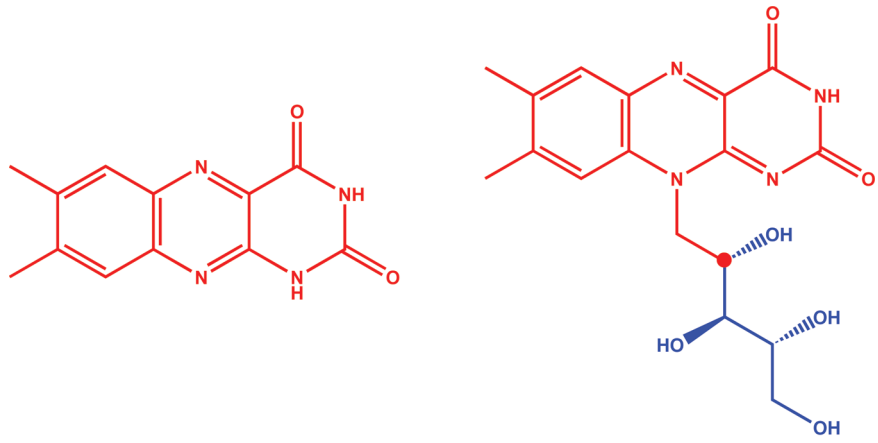
In summary, our recursion procedure for the purpose of obtaining a decent biasing potential $f_m(\lambda) + F_m(\lambda,\partial U_o/\partial\lambda)$ has two components. First, $f_m(\lambda)$ is updated via the "recursion slave" based on eqs 10 and 11. It should be noted that in eqs 10 and 11, the estimation of $G_o(\lambda)$ for the $f_m(\lambda)$ update is not directly dependent on the instantaneous $f_m(\lambda)$ but only dependent on individual one-dimensional (along $\partial U_o/\partial\lambda$) free energy profiles $-F_m(\lambda',(\partial U_o/\partial\lambda)|_{\lambda'})$ at each state λ'. Therefore, the above procedure can ensure the convergence of $f_m(\lambda)$ to the target function of $-G_o(\lambda)$. When $f_m(\lambda)$ converges to $-G_o(\lambda)$, this recursion procedure effectively becomes a two-dimensional metadynamics simulation along the directions of λ and $\partial U_o/\partial\lambda$ with $U_o - G_o(\lambda)$ as the potential energy function. Thereafter, through the "recursion kernel" described by eq 10, $F_m(\lambda,\partial U_o/\partial\lambda)$ will converge to the target $-G'_o(\lambda,\partial U_o/\partial\lambda)$. To understand how the above procedure works, model study 1 in ref 35 can be a good illustrative example.

It should be noted that, in all of the above equations, $\partial U_o/\partial\lambda$ is equal to $(U_s^{B,QM/MM} - U_s^{A,QM/MM})$ for the direct scheme based QM/MM AFE simulations (eq 3), and $\partial U_o/\partial\lambda$ is equal to, for instance, $U_s^{A,QM/MM} - U_s^{A,MM}$ to calculate $\Delta G_{A,QM/MM\rightarrow A,MM}$, in the indirect scheme (eq 5).

III. Computational Details

The present OSRW-based QM/MM AFE simulation method was implemented in our customized CHARMM program.[78,79] Presently, it is implemented with the SCCDFTB (self-consistent charge density functional tight binding)/CHARMM treatment.[80] We are actively working on the implementation of the present method for higher level electronic structural descriptions. Notably, the present direct scheme implementation is realized via the mechanical mixing method rather than the electronic structural switching method;[55] the latter scheme requires only one electronic structural force calculation at each time step, and therefore it is another target for our future development. To illustrate the presently proposed algorithm in both the direct scheme and the indirect scheme, two model studies were performed.

In OSRW simulations, the introduction of the energy term $F_m(\lambda,\partial U_o/\partial\lambda)$ may bring some computational overhead compared with canonical MD simulations. However, our implementation regarding this energy term benefit from two unique treatments: (1) like the implementation in a presently employed metadynamics package,[81] unit Gaussian functions are stored in grids; (2) the force vectors related to $\partial U_o/\partial\lambda$ are stored in the middle of the force calculations on U_o. Therefore, the force calculation in each step has less than 0.3% overhead, even for small-sized systems like ones in model study 1. The other source of the overhead involves the recursion slave estimations based on eqs 9 and 10; it is also less than 0.1% of 100 step force calculations, which



(a) Lumichrome (LC) (b) Riboflavin (RF)

Figure 1. Molecular structures of two flavin derivatives: lumichrome (a) and riboflavin (b). The portions colored in red are described quantum mechanically; the portion colored in blue is described molecular mechanically. The atom labeled by the red dot represents the interface atom described by the generalized hybrid orbital (GHO) treatment.

are commonly used by us as the interval of the recursion slave update.

To illustrate the presently proposed algorithm in both the direct scheme and the indirect scheme, two model studies were performed.

III.A. Model Study 1: The Calculations of the Redox Potentials of Two Hydrated Flavin Derivatives. To illustrate the direct scheme, model study 1 was performed to calculate the redox potentials of two flavin derivatives: lumichrome (LC) and riboflavin (RF) (Figure 1), in aqueous solution. The model calculation on RF was designed to illustrate a situation when an across-bond boundary between the quantum mechanical (QM) portion and the molecular mechanical (MM) portion is defined (as shown in Figure 1b, where the carbon atom labeled by the red dot is the interface atom described by the generalized hybrid orbital (GHO) treatment[82−84]). It is noted that such a boundary definition is not obligatory for predicting the RF redox potential; it was set up only to demonstrate our implementation, which naturally permits the GHO boundary treatment. Moreover, the long-range electrostatics were treated by the recent integrated SCCDFTB-based lattice Ewald method.[49,85]

Seven sets of independent OSRW-based QM/MM AFE simulations, which were started with different random seeds, were performed on two model systems to assess the efficiency of the present method. The statistical error values were calculated among seven simultaneously estimated free energy values, which were obtained at the same simulation length. In each of these alchemical switchings, which are based on eq 3, the starting state ($\lambda = 0$) was set as the neutral state (LC or RF), and the ending state ($\lambda = 0$) was set as the reduced radical state ($LC^{\bullet-}$ or $RF^{\bullet-}$). Therefore, the obtained free energy changes correspond to the reduction potentials of two model systems. It is noted that the utilization of these two model systems was motivated by a recent work, which was to develop an efficient minimum energy path based redox potential estimation algorithm.[56]

In order to compare the present method with the classical approaches, a set (11) of canonical ensemble MD simulations was performed on the RF model system; these MD simulations were based on the energy function described by eq 3 with the scaling parameter λ set as 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Starting with $\lambda = 0.0$, sequentially, each simulation was run with a simulation length of 280 ps. The reduction potential of the RF model system was estimated on the basis of an advanced thermodynamic integration (TI) procedure; specifically based on the samples collected in the 11 canonical ensemble simulations, the free energy derivatives of 101 equally spaced λ states were estimated via the maximal likelihood based λ-WHAM method, the details of which can be found in ref 86. It should be noted that the λ-WHAM approach can ensure better free energy convergence than the classical TI method.

III.B. Model Study 2: The Equilibrium of a Tautomerization Process of Hydrated 3-Hydroxypyrazole. To illustrate the indirect scheme, model study 2 was performed to calculate the free energy change of a tautomerization process of hydrated 3-hydroxypyrazole (Figure 2), which was a target in a recent prediction contest. Following the related discussions in section II.A (eq 5), two indirect-scheme-based QM/MM AFE simulations are required to respectively calculate $\Delta G_{B,MM\rightarrow B,QM/MM}$ and $\Delta G_{A,MM\rightarrow A,QM/MM}$, and one classical AFE simulation is required to calculate $\Delta G_{A,MM\rightarrow B,MM}$. As shown in Figure 2, $\Delta G_{A,MM\rightarrow B,MM}$ can be estimated by the following:

$$\Delta G_{A,MM\rightarrow B,MM} = \Delta G_{A,MM\rightarrow A',MM} + \Delta G_{A',MM\rightarrow B',MM} + \Delta G_{B',MM\rightarrow B,MM} \quad (12)$$

where the A′ and B′ states are defined as the modified A and B states with one extra dummy atom in each; the addition of these dummy atoms allows the calculation of $\Delta G_{A',MM\rightarrow B',MM}$ to be more efficiently handled. Specifically, in the A′ state, a dummy atom, which is alchemically switched to a hydroxyl hydrogen during the $\Delta G_{A',MM\rightarrow B',MM}$ AFE simulations, is linked to the corresponding oxygen via a bond and an angle energy term; in the B′ state, a dummy atom, which is alchemically switched to a hydrogen in a five-member ring during the $\Delta G_{A',MM\rightarrow B',MM}$ AFE

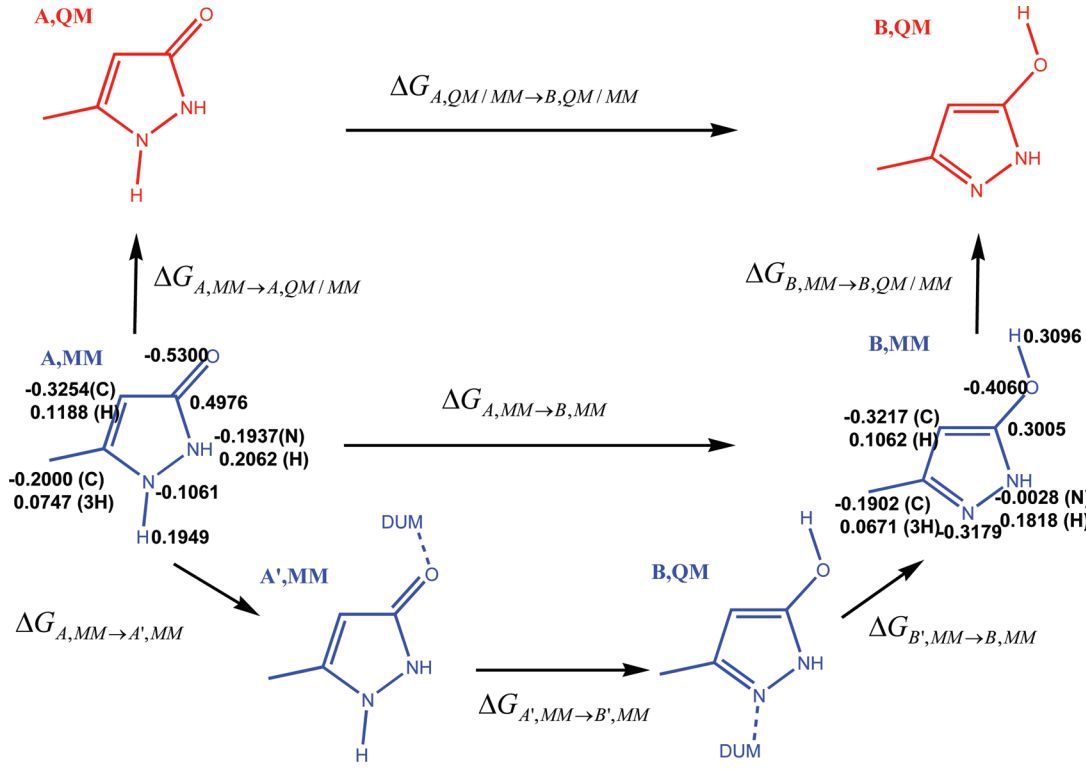


Figure 2. Thermodynamic cycle employed to estimate the free energy change of a tautomerization process of hydrated 3-hydroxypyrazole based on the QM/MM potential. The portions colored in red are described quantum mechanically; the portions colored in blue are described molecular mechanically. The charges in the charge set 1 of the MM parameters are labeled around their corresponding atoms.

Table 1. Restraint Setups for the Calculation of $\Delta G_{A',MM\rightarrow B',MM}$ in Model Study 2[a]

	bond restraints (Å, kcal/mol/Å^2)	angle restraints (rad, kcal/mol/rad^2)	torsional restraints (rad, kcal/mol/rad^2)
dummy atom in A′	(0.96, 1090.0)	(0.52, 242.0)	NA
dummy atom in B′	(1.03, 900.0)	(2.12, 60.0)	(3.14, 5)

[a] All the restraint potentials are based on the harmonic equation, $U_{Restraint} = 1/2K_\xi(\xi - \xi_o)^2$, where ξ_o represents a reference value and K_ξ represents the corresponding force constant. The dummy atom in B′ is torsionally restrained on the same plane as the heterocyclic ring.

simulations, is linked to the corresponding nitrogen atom via a bond, an angle, and a dihedral energy restraint energy term. The dummy atom setup is based on the general principle of the "virtual bond algorithm" (VBA):[87] (1) Each dummy atom has only one bond, angle, and dihedral term so that in the regime of the harmonic oscillator and the rigid rotor approximations, its free energy contribution is separable from the rest of the system and only analytically depends on these restraint terms. (2) The restraint energy terms are set as the same to the corresponding ones at the other end state of $\Delta G_{A',MM\rightarrow B',MM}$ AFE simulations to maximize the phase space overlap between two end chemical states. In the present model study, the restraint energy term parameters of the A′ and B′ state dummy atoms are listed in Table 1. Moreover, the dummy atom in the $\Delta G_{A',MM\rightarrow B',MM}$ AFE simulations is not torsionally restrained; this design is to take advantage of the generalized ensemble tunneling mechanism so as to facilitate the conformational sampling of its alchemical partner.[23,24,75] $\Delta G_{A',MM\rightarrow B',MM}$ was directly calculated via the MM-based OSRW AFE simulation approach.[35] On the basis of the VBA analysis,[87] $\Delta G_{A,MM\rightarrow A',MM} + \Delta G_{B',MM\rightarrow B',MM}$ is estimated to be −0.02 kcal/mol. Due to the introduction of the dummy atoms, the separation-shifted type[71] of the soft-core potentials was employed. The same as in our previous AFE simulations,[35,74] the soft-core potential treatments are simultaneously applied to the van der Waals portion and the real-space electrostatic portion. It should be noted that this soft-core electrostatic potential switching has never been published; however, this treatment has been in the CHARMM program for many versions and has been extensively utilized.[79] A similar simultaneous soft-core treatment can also be found

in the recent literature.[88] The soft-core shifting parameters are set as 5 for both the van der Waals switching and the electrostatic switching.

Two sets of QM/MM AFE simulations on this model system were performed to assess the sampling robustness of the indirect scheme based method. All of the MM state bond, angle, dihedral, and van der Waals parameters were set up according to the CHARMM parameter set. In the first simulation set, the MM charge of each atom was set the same as the charge derived from the SCCDFTB calculation based on the corresponding gas phase minimum energy structure. These charges are shown in Figure 2. In the second

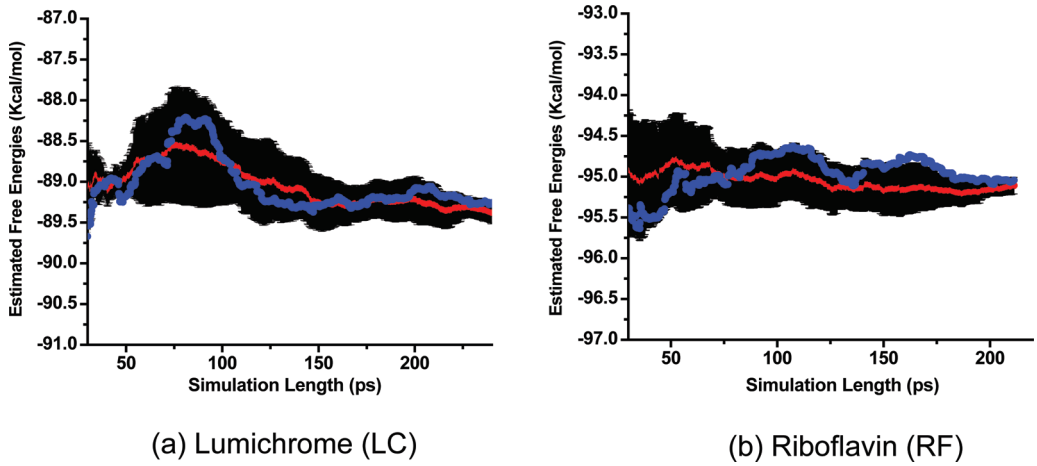


Figure 3. Time-dependent estimated reduction potentials on lumichrome (a) and riboflavin (b). The averages of seven simulation results are shown by the red lines. On each system, the results of a randomly picked free energy simulation are shown by the blue dots, and the statistical errors among seven simulation results are shown around their average results.

simulation set, the MM charge of each atom was set as a half of the charge of the corresponding atom in the first simulation set. Between two simulation sets, the first simulation set represents a case that the MM parameters are well-prepared, and the other simulation set represents the case that the MM parameters are casually or badly prepared.

In all of the OSRW setups, the height of the Gaussian function h (eq 9) was set as 0.01 kcal/mol; the widths of the Gaussian function, ω_1 and ω_2, were set as 0.02 and 4 kcal/mol, respectively. In the recursion kernel, $F_m(\lambda,\partial U_o/\partial\lambda)$ was updated every 10 time steps; in the recursion slave, $f_m(\lambda)$ was updated every 100 time steps. Regarding the choice of the unit Gaussian function, like that discussed in the metadynamics method,[89,90] a greater Gaussian height (or smaller widths) may lead to a larger error of the estimated $-G'_o(\lambda,\partial U_o/\partial\lambda)$; however, a lower Gaussian height (or smaller widths) may cost extra recursion time.[89,90] The set of the Gaussian parameters employed in this work has been extensively tested in various earlier OSRW-based alchemical free energy simulations and robustly applied in our earlier work. Certainly they can be optimized to further improve the efficiency for the present model systems, for instance, via the Wang−Laudau updating scheme;[90] we will leave such discussions for future work.

III.C. General Simulation Setups. In terms of the molecular dynamics simulation setup, all the target systems are respectively solvated in the cubic water (with the TIP3P water model[91]) box with an initial size of 31 × 31 × 31 Å^3, and the linear Ewald method[85] was applied to take care of the long-range columbic interactions while the short-range interactions were switched starting at 8 Å and were totally off at 12 Å. The Nose−Hoover method[92] was employed to maintain a constant temperature at 300 K, and the Langevin piston algorithm[93,94] was used to maintain the constant pressure at 1 atm. The time step was set as 1 fs.

IV. Results and Discussions

IV.A. Model Study 1: The Calculations of the Redox Potentials of Two Hydrated Flavin Derivatives. All of the OSRW AFE simulations on LC and RF show robust and efficient convergence behaviors. As shown in Figure 3, the average of seven independent free energy simulation results in LC reached a superior convergence with a statistical error of 0.13 kcal/mol in about 240 ps; then, the reduction potential of LC was estimated as −89.37 kcal/mol (−3.88 eV). The average of seven independent free energy simulations on RF also reached a superior convergence with a statistical error of 0.10 kcal/mol in about 210 ps; then the reduction potential of RF was estimated as −95.11 kcal/mol (−4.12 eV). As introduced in section III.A, each statistical error value was calculated among seven simultaneously estimated free energy values that were obtained at the same simulation length. Therefore, such statistical error should reflect the magnitude of the divergence of a single free energy prediction from the average at each particular simulation length; i.e., it reflects the robustness of a single OSRW AFE simulation prediction. In terms of the absolute values, the estimated reduction potential of LC, −3.88 eV, is in good agreement with the experimental range −4.3 to −3.7 eV,[95] and the estimated reduction potential of RF, −4.12 eV, is also in a good agreement with the experimental range −4.5 to −3.9 eV.[95] The uncertainty of the experimental values is due to the limitation of the reference reduction potential measurement, which was carried out on the basis of the standard hydrogen electrode (SHE).[96−100] The difference between two estimated reduction potentials of RF and RL is −0.24 eV (−5.53 kcal/mol), which also agrees well with the corresponding measured difference −0.21 eV (−4.84 kcal/mol).[95] Besides the superior convergence, such an excellent agreement should also be attributed to many other factors, for instance, the QM/MM treatment, the inclusion of the EWALD treatment, the possible error cancelation, and even the employment of the GHO boundary condition etc. Although, in an earlier section, the possible limitation of the employed SCCDFTB description was acknowledged and the emphasis of the present work on the efficiency of AFE simulations was clarified, the quantitative prediction power on the LC and RF redox potentials based on the present package of the technical treatments, as demonstrated here, is still practically encouraging.

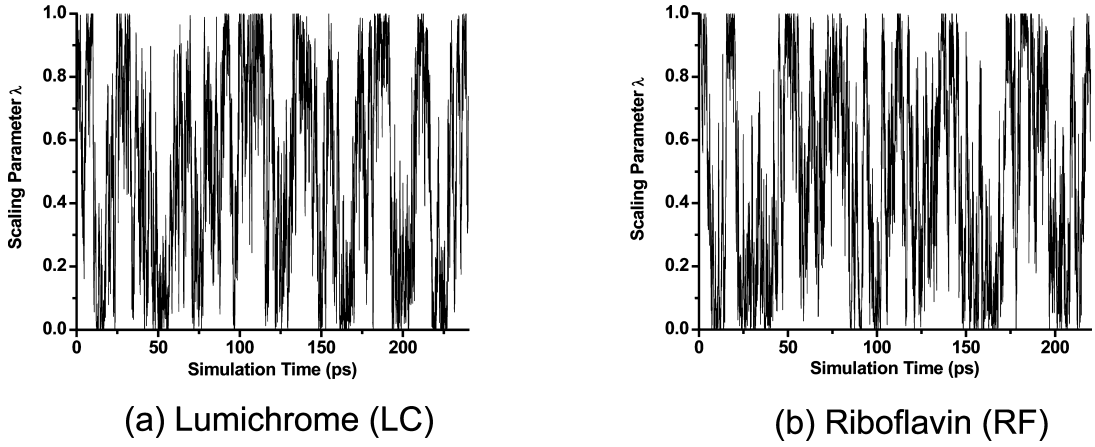


Figure 4. Time-dependent scaling parameter changes of the representative runs on lumichrome (a) and riboflavin (b).

To understand the convergence behavior and assess the efficiency of a single OSRW alchemical FES, let us take a close look at one randomly chosen case from each of the two simulation sets (on LC and RF). As shown in Figure 4a, which reveals the time-dependent scaling parameter changes in a representative run on LC, a complete round trip of the scaling parameter between two end states was realized in about 30 ps. As shown in Figure 3a, at a simulation length of 30 ps, the free energy value (shown by the blue dots) estimated by this single OSRW simulation is −89.64 kcal/mol, which is different from the finally converged value −89.37 kcal/mol only by 0.27 kcal/mol; at the same time, the average free energy value of seven OSRW simulations is −89.07 kcal/mol and the statistical error is 0.52 kcal/mol. As the matter of fact, in the following 20 ps, the convergence robustness of a single OSRW simulation was further improved; this can be reflected by the smaller magnitudes of the time-dependent free energy value changes (the blue dots in Figure 3a) and the reduced statistical errors (Figure 3a). It should be noted that there are two major types of environmental relaxation that strongly couple with the electron transfer process in LC: the water molecule reorganization and the bending and stretching motions of the LC ring. In about 50 ps of simulation, the necessary environmental water reorganization was sufficiently activated by the energy surface flattening along the orthogonal space order parameter $\partial U_o/\partial\lambda$ so that a fast random walk of the scaling parameter was achieved (Figure 4a). As shown by the first plot of Figure 5a, although at 50 ps the height of $-F_m(\lambda,\partial U_o/\partial\lambda)$ was only about 0.6 kcal/mol, the free energy derivative $\partial G_o/\partial\lambda$ curve (the black line) had then been well obtained. Obviously, the free energy result obtained at around 50 ps has been quite acceptable for quantitative prediction purposes. When this OSRW simulation was continued, further free energy flattening started accelerating the LC high-frequency bending and stretching motions and in particular their coupled water reorganization, which could also influence the equilibrium of the target electron transfer process. Then, the contribution of the second type of environmental relaxation, which commonly requires nontrivial canonical sampling to capture, is included as well. The additional recursion, by which the biasing potential term $-F_m(\lambda,\partial U_o/\partial\lambda)$ was further updated to realize possible acceleration of such environmental reorganization, took about another 25 ps of simulation length; as shown in Figure 4a, in this relaxation activation recursion period, another complete scaling parameter round trip ended at 75 ps, when the statistical error among seven independent predictions increased to a peak value, 0.71 kcal/mol. Thereafter, the further recursion allowed the refinement of the following free energy estimations. At a simulation length of 100 ps, when the height of $-F_m(\lambda,\partial U_o/\partial\lambda)$ was increased to 1.2 kcal/mol (the second plot of Figure 5a), the statistical error was reduced to 0.52 kcal/mol (Figure 3a). At a simulation length of 150 ps, when the height of $-F_m(\lambda,\partial U_o/\partial\lambda)$ was increased to 1.6 kcal/mol (the third plot of Figure 5a), the statistical error was reduced to 0.34 kcal/mol; then, the averaged estimated value was −89.25 kcal/mol, which is already very close to the finally estimated value −89.37 kcal/mol. As shown by the fourth plot of Figure 5b, at around 200 ps, a high resolution kernel biasing potential $-F_m(\lambda,\partial U_o/\partial\lambda)$ was well obtained; on the basis of eqs 11 and 12, such a high resolution kernel biasing potential naturally leads to a nicely converged free energy value.

In terms of the RF case, Figure 4b reveals the time-dependent scaling parameter changes of a representative run, where a complete round trip of the scaling parameter between two end states was realized in about 40 ps. As shown in Figure 3b, at a simulation length of 40 ps, the free energy value (shown by the blue dots) estimated by this single OSRW simulation is −95.52 kcal/mol, which is different from the finally converged value −95.11 kcal/mol by only 0.41 kcal/mol; at the same time, the average free energy value from seven OSRW simulations is −94.98 kcal/mol, and the statistical error is 0.62 kcal/mol. When this OSRW simulation was continued, the following recursion simulation allowed the free energy estimations to be further refined. For instance, at a simulation length of 50 ps, when the height of $-F_m(\lambda,\partial U_o/\partial\lambda)$ was around 1.0 kcal/mol (the first plot of Figure 5b), the statistical error was reduced to 0.54 kcal/mol (Figure 3a). At a simulation length of 100 ps, when the height of $-F_m(\lambda,\partial U_o/\partial\lambda)$ was increased to 1.6 kcal/mol (the second plot of Figure 5b), the statistical error was reduced to 0.30 kcal/mol. At a simulation length of 200 ps, the

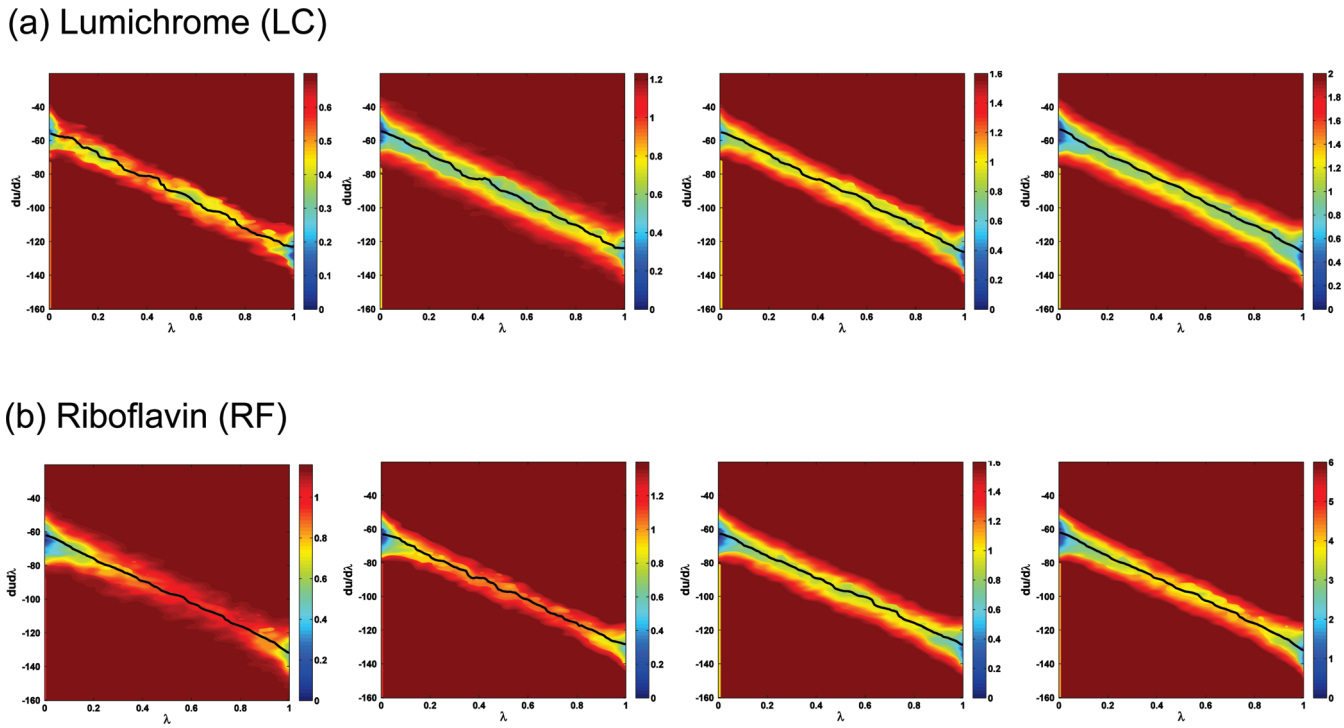


Figure 5. Contour plots on the estimated free energy potential $G'_o(\lambda,\partial U_o/\partial\lambda)$ (colored contours) and the plots on the λ-dependent free energy derivatives (black lines). The depth of each estimated free energy potential $G'_o(\lambda,\partial U_o/\partial\lambda)$ contour plot reveals the activation level in the $\partial U_o/\partial\lambda$ direction. These results were obtained from the representative simulations on lumichrome (a) and riboflavin (b). On each system, the results from the 50 ps simulation are shown as the first, the results from the 100 ps simulation are shown as the second, the results from the 150 ps simulation are shown as the third, and the results from the 200 ps simulation are shown as the fourth.

statistical error was reduced to 0.11 kcal/mol, and then, the averaged estimated value was −95.11 kcal/mol, which is identical to the finally estimated value.

IV.B. The Comparison Study on the RF System Based on the λ-WHAM TI Method. As described in the Computational Details section, the comparison study on the RF system is based on 11 canonical ensemble MD simulations, which were sequentially performed with a length of 280 ps on each λ state; all of the free energy analysis is based on an improved TI (λ-WHAM) procedure.[86]

The lines in Figure 6 show the changes of the estimated free energies with an increase of the simulation length on each single λ state. It is noted that the computing cost should be calculated on the basis of the overall sampling length; for instance, when the simulation length on each λ state is 10 ps, the overall simulation length is 110 ps (from all 11 state simulations). The red line reveals the fact that it took each of the simulations about 180 ps (with overall 1.9 ns simulation length) to reach a "pseudo-convergence", and then the free energy value was estimated as −95.28 kcal/mol. Considering possible contamination of the nonequilibrium data, the samples generated during the first 50 ps of each simulation (with overall 550 ps simulation length) were removed; then, the estimated free energy value shifts up to −95.24 kcal/mol, which converges around 180 ps (also with overall 1.9 ns simulation length; blue line). If the samples generated during the first 150 ps of each simulation (with overall 1.65 ns simulation length) are removed, then the estimated free energy value further shifts to −95.13 kcal/mol, which converges around 220 ps (with overall 2.42 ns simulation length) with a statistical error of 0.52 kcal/mol (green line). If the samples generated during the first 200 ps of each simulation (with overall 2.2 ns simulation length) are removed, then the estimated free energy value further

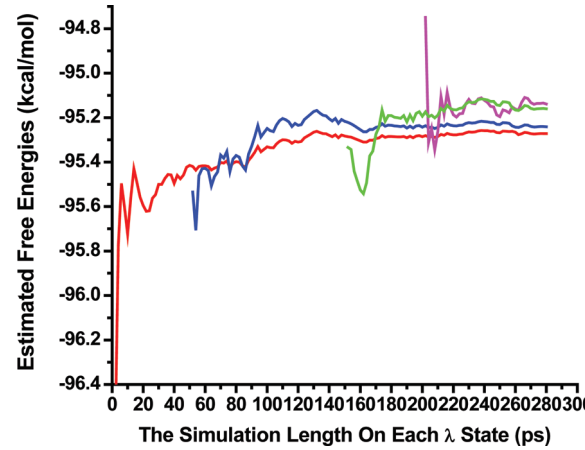


Figure 6. Free energy simulation results from the comparison study on the riboflavin model system. The lines show the changes of the estimated free energies with the increase of the simulation length on each single λ state. The red line shows the time-dependent results based on the samples without the removing of any equilibration data. The blue line shows the time-dependent results when the first 50 ps samples in each λ state simulation are removed. The green line shows the time-dependent results when the first 150 ps samples in each λ state simulation are removed. The pink line shows the time-dependent results when the first 200 ps samples in each λ state simulation are removed.

Table 2. Free Energy Components for the Calculation of $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ in Model Study 2[a]

	$\Delta G_{A',MM\rightarrow B',MM}$ (kcal/mol)	$\Delta G_{A,MM\rightarrow B,MM}$ (kcal/mol)	$\Delta G_{A,MM\rightarrow A,QM/MM}$ (kcal/mol)	$\Delta G_{B,MM\rightarrow B,QM/MM}$ (kcal/mol)	$\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ (kcal/mol)
charge set 1	6.72	6.70	−10 761.54	−10 771.91	−3.67
charge set 2	1.90	1.88	−10 794.92	−10 800.55	−3.75

[a] $\Delta G_{A,MM\rightarrow B,MM}$ was estimated on the basis of eq 12; on the basis of the VBA analysis, $\Delta G_{A,MM\rightarrow A',MM} + \Delta G_{B',MM\rightarrow B',MM}$ is estimated to be −0.02 kcal/mol. $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ was estimated on the basis of eq 5. The experimental value of $\Delta G_{A,QM/MM\rightarrow B,QM/MM}$ is −3.40 kcal/mol, which was measured on the basis of the pK_a value difference between two target molecules.

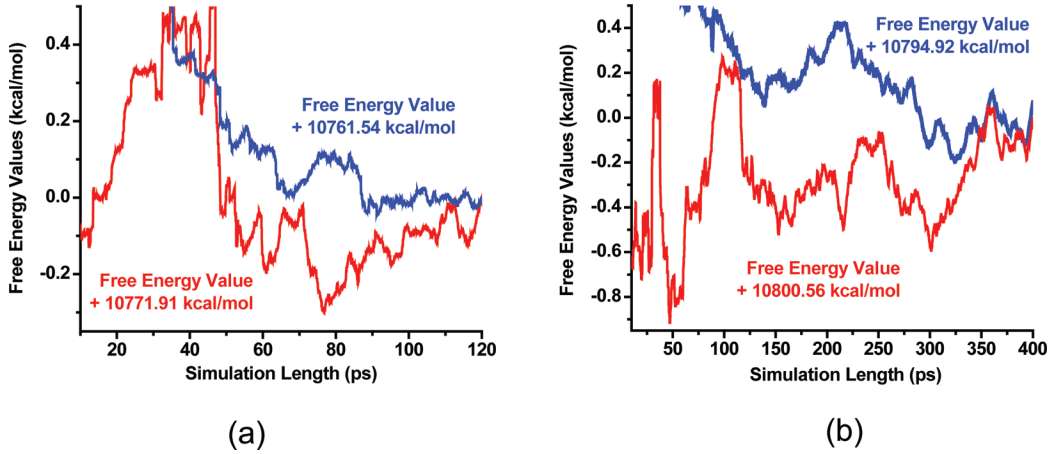


Figure 7. Time-dependent predicted free energy values in model study 2. The free energy results from the simulations where the charge set 1 of the MM parameters was employed are shown in a, and the free energy results from the simulations where the charge set 2 of the MM parameters was employed are shown in b. The free energy results on molecule A are shown in red, and the free energy results on molecule B are shown in blue. To show the results clearly, the values from each simulation are shown in the shifted fashion; for instance, the red line in Figure 6a shows the change of the free energy value plus 10771.91 kcal/mol.

shifts to −95.12 kcal/mol, which also converges around 220 ps (with overall 2.42 ns simulation length) with a statistical error of 0.61 kcal/mol (pink line). It should be noted that, certainly, each of the 11 simulations may have different equilibration lengths. However, the present analysis is performed only for a qualitative efficiency comparison purpose, and a quantitative efficiency comparison requires more elaborate analysis.

Clearly, the comparison study demonstrates the fact that the OSRW approach and the classical methods produce the same free energy results, and the OSRW approach has superior sampling efficiency over the employed λ-WHAM-based TI procedure. Notably, in this study on the RF system, the intrinsic environmental (water molecules) relaxation is relatively fast; it can be anticipated that in more complex systems, where environmental responses are slower, more efficiency gains based on the OSRW method could be obtained.

IV.C. Model Study 2: The Equilibrium of a Tautomerization Process of Hydrated 3-Hydroxypyrazole. As shown in Table 2, two sets of the indirect scheme based AFE simulations led to quantitatively the same predictions (−3.67 and −3.75 kcal/mol) on the equilibrium free energy change of the tautomerization process of the hydrated 3-hydroxypyrazole that is illustrated in Figure 2. These estimations are in great agreement with the experimental measured result, −3.40 kcal/mol; this experimental value was obtained on the basis of the pK_a difference (2.50) of two target molecules.[101] As emphasized in section IV.A, although the purpose of this work is to demonstrate the efficiency of the present QM/MM AFE scheme, the quantitative prediction power of the present package of implementations, which include the OSRW sampling strategy, the EWALD treatment, and even the coupling of the constant pressure treatment etc., shows the necessity of an elaborate development of QM/MM-based AFE simulation facilities.

In the indirect scheme of the simulations based on charge set 1 of the MM parameters, both of the QM/MM AFE simulations reached decent convergence within 50 ps (Figure 7a), when the predicted values were within 0.3 kcal/mol from the converged free energy values (obtained after 500 ps); this is consistent with the fact that at around 50 ps, the first scaling parameter round-trips had been realized (as shown in Figure 7a). The free energy refinements, which were obtained by more frequent scaling parameter space random walks after the initial scaling parameter round-trips (Figure 8a), resulted in smaller deviation of the estimated results from the converged free energy values; for instance, at round 100 ps, the predicted values are only within 0.1 kcal/mol from the converged free energy results. The $\Delta G_{A',MM\rightarrow B',MM}$ prediction was achieved via a 300 ps classical OSRW simulation. It is noted that from the sampling viewpoint, the QM/MM AFE simulations on charge set 1 of the MM parameters are less challenging for the fact that charge set 1 was obtained on the basis of the SCCDFTB gas phase calculations; therefore the QM/MM and the MM end states in each QM/MM AFE simulation have reasonable phase-space overlaps. In comparison, the QM/MM AFE simulations on

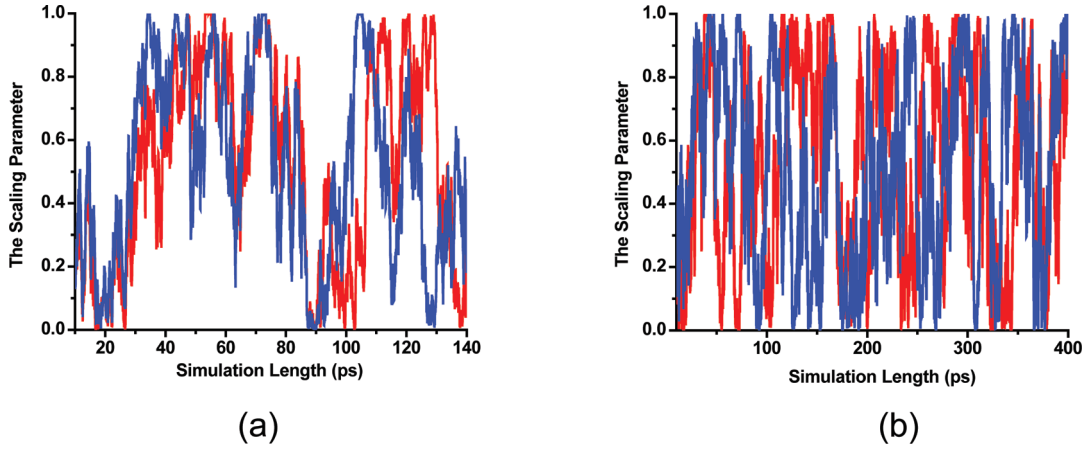


Figure 8. Time-dependent scaling parameter changes in model study 2. The results from the simulations where the charge set 1 of the MM parameters was employed are shown in a, and the results from the simulations where the charge set 2 of the MM parameters was employed are shown in b. The results on molecule A are shown in red, and the results on molecule B are shown in blue.

charge set 2 of the MM parameters show slower convergence behaviors; this observation is also consistent with the fact that charge set 2 was set to represent a poorly prepared charge set with each atomic charge equal to the half of its corresponding value in charge set 1. Their initial scaling parameter round trips were obtained until about 100 ps (Figure 8b), after the predicted values could be 0.5 kcal/mol from the converged free energy values (Figure 7b). More frequent scaling parameter random walks after the first 100 ps of simulation length (Figure 8b) naturally led to better free energy convergence; for instance, at 350 ps, the predicted values are only within 0.1 kcal/mol from the converged free energy results (Figure 7b). On charge set 2, the $\Delta G_{A',MM\rightarrow B',MM}$ prediction was achieved via a 300 ps classical OSRW simulation as well.

V. Concluding Remarks

The difference of free energy changes occurring at two chemical states can be rigorously estimated via alchemical free energy (AFE) simulations. Traditionally, most AFE simulations are carried out under the classical energy potential treatment; then, accuracy and applicability of AFE simulations could be limited; for instance, tautomerization free energy estimations and metal associated ligand binding affinity estimations are outstanding examples.

Due to the fact that QM/MM force calculations are time-demanding, how to achieve decent free energy convergence within a short simulation length has been a bottleneck problem in the QM/MM AFE simulation method development. In the present work, we integrate a recent second-order generalized ensemble strategy, the orthogonal space random walk (OSRW) method, into the combined quantum mechanical/molecular mechanical (QM/MM) calculation based AFE simulation scheme. Thereby, within a commonly affordable simulation length of time, accurate QM/MM alchemical free energy simulations can be achieved. As revealed by the model study on the equilibrium of a tautomerization process of hydrated 3-hydroxypyrazole and by the model calculations of the redox potentials of two flavin derivatives: lumichrome (LC) and riboflavin (RF) in aqueous solution, the present OSRW based scheme could be a viable path toward the realization of practically efficient QM/MM AFE simulations.

Acknowledgment. This paper is dedicated to the 80th birthday of Professor Martin Karplus, whose encouragement on the development of practically efficient QM/MM AFE simulation methods has been pivotal to the success of the present work. We would also like to thank Drs. Jingzhi Pu, Kwangho Nam, Guishan Zheng, Qiang Cui, and Darrin York for many helpful discussions. We would like to acknowledge the National Science Foundation (MCB 0919983) for the funding support. We also thank the Florida State University High Performance Computing Center and the Institute of Molecular Biophysics computing facility for their computing support.

References

(1) Warshel, A. Simulating the energetics and dynamics of enzymatic reactions. *Pont. Acad. Sci. Scr. Var.* **1983**, *55*, 59–81.

(2) Tembe, B. L.; McCammon, J. A. Ligand receptor interactions. *Comput. Chem.* **1984**, *8*, 281–283.

(3) Jorgensen, W. L.; Ravimohan, C. Monte-Carlo simulation of differences in free-energies of hydration. *J. Am. Chem. Soc.* **1985**, *83*, 3050–3054.

(4) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Free-energy calculations by computer-simulation. *Science* **1987**, *236*, 564–568.

(5) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins - A molecular-dynamics analysis. *Science* **1989**, *244*, 1069–1072.

(6) Jorgensen, W. L. Free-energy calculations - A breakthrough for modeling organic-chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(7) Beveridge, D. L.; Dicapua, F. M. Free-energy via molecular simulation - Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.

(8) Straatsman, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(9) Kollman, P. Free-energy calculations - Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(10) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.

(11) Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(12) Jorgensen, W. L.; Thomas, L. L. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.

(13) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.

(14) Zwanzig, R. W. High-temperature equation of state by a perturbation method. 1. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(15) Bennett, C. H. Efficient estimation of free-energy differences from Monte-Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(16) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling g with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.

(17) Shirts, M. R.; Bair, E.; Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.

(18) Lu, N. D.; Kofke, D. A.; Woolf, T. B. Improving the efficiency and reliability of free energy calculations using overlap sampling methods. *J. Comput. Chem.* **2004**, *25*, 28–39.

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

(20) Knight, J. L.; Brooks, C. L. Lambda-dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.

(21) Tidor, B. Simulated annealing on free-energy surfaces by a combined molecular-dynamics and Monte-Carlo approach. *J. Phys. Chem.* **1993**, *97*, 1069–1073.

(22) Pitera, J.; Kollman, P. Designing an optimum guest for a host using multimolecule free energy calculations: Predicting the best ligand for Rebek's "tennis ball". *J. Am. Chem. Soc.* **1998**, *120*, 7557–7567.

(23) Li, H. Z.; Fajer, M.; Yang, W. Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J. Chem. Phys.* **2007**, *126*, 024106.

(24) Min, D. H.; Yang, W. Energy difference space random walk to achieve fast free energy calculations. *J. Chem. Phys.* **2008**, *128*, 191102.

(25) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.

(26) Bitetti-Putzer, R.; Yang, W.; Karplus, M. Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett.* **2003**, *377*, 633–641.

(27) Fasnacht, M.; Swendsen, R. H.; Rosenberg, M. Adaptive integration method for Monte Carlo simulations. *Phys. Rev. E* **2004**, *69*, 056704.

(28) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. Comparison of free energy methods for molecular systems. *J. Chem. Phys.* **2006**, *126*, 184114.

(29) Pomes, R.; Eisenmesser, E.; Post, C. B.; Roux, B. Calculating excess chemical potentials using dynamic simulations in the fourth dimension. *J. Chem. Phys.* **1999**, *111*, 3387–3395.

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

(31) Woods, C. J.; Essex, J. W.; King, M. A. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.

(32) Lu, N. D.; Wu, D.; Woolf, T. B.; Kofke, D. A. Using overlap and funnel sampling to obtain accurate free energies from nonequilibrium work measurements. *Phys. Rev. E* **2004**, *69*, 057702.

(33) Christ, C. D.; van Gunsteren, W. F. Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *J. Chem. Phys.* **2007**, *126*, 184110.

(34) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.* **2006**, *125*, 074115.

(35) Zheng, L. Q.; Chen, M. G.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.

(36) Warshel, A. Dynamics of reactions in polar-solvents - Semiclassical studies of electron-transfer and proton-transfer reactions. *J. Phys. Chem.* **1982**, *86*, 2218–2224.

(37) King, G.; Warshel, A. Investigation of the free-energy functions for electron-transfer reactions. *J. Chem. Phys.* **1990**, *93*, 8682–8692.

(38) Gao, J. L.; Xia, X. F. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631–635.

(39) Luzhkov, V.; Warshel, A. Microscopic models for quantum-mechanical calculations of chemical processes in solutions - LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J. Comput. Chem.* **1992**, *13*, 199–213.

(40) Gao, J. L.; Luque, F. J.; Orozco, M. Induced dipole-moment and atomic charges based on average electrostatic potentials in aqueous-solution. *J. Chem. Phys.* **1993**, *98*, 2975–2982.

(41) Wesolowski, T.; Warshel, A. Ab-initio free-energy perturbation calculations of solvation free-energy using the frozen density-functional approach. *J. Phys. Chem.* **1994**, *98*, 5183–5187.

(42) Stanton, R. V.; Little, L. R.; Merz, K. M. Quantum free-energy perturbation study within a PM3-MM coupled potential. *J. Phys. Chem.* **1995**, *99*, 483–486.

(43) Gao, J. L.; Freindorf, M. Hybrid ab initio QM/MM simulation of N-methylacetamide in aqueous solution. *J. Phys. Chem. A* **1997**, *101*, 3182–3188.

(44) Li, G. H.; Zhang, X. D.; Cui, Q. Free energy perturbation calculations with combined QM/MM potentials: Complications, simplifications, and applications to redox potential calculations. *J. Phys. Chem. B* **2003**, *107*, 8643–8653.

(45) Li, G. H.; Cui, Q. pKa calculations with QM/MM free energy perturbations. *J. Phys. Chem. B* **2003**, *107*, 14521–14528.

(46) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. Frozen density functional free energy simulations of redox proteins: Computational studies of the reduction potential of plastocyanin and rusticyanin. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(47) Yang, W.; Bitetti-Putzer, R.; Karplus, M. Chaperoned alchemical free energy simulations: A general method for QM, MM, and QM/MM potentials. *J. Chem. Phys.* **2004**, *120*, 9450–9453.

(48) Hu, H.; Yang, W. T. Dual-topology/dual-coordinate free-energy simulation using QM/MM force field. *J. Chem. Phys.* **2005**, *123*, 041102.

(49) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; Guo, H.; Elstner, M.; Cui, Q. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

(50) Blumberger, J.; Tavernelli, I.; Klein, M. L.; Sprik, M. Diabatic free energy curves and coordination fluctuations for the aqueous Ag^{+}/Ag^{2+} redox couple: A biased Born-Oppenheimer molecular dynamics investigation. *J. Chem. Phys.* **2006**, *124*, 064507.

(51) Blumberger, J.; Sprik, M. Quantum versus classical electron transfer energy as reaction coordinate for the aqueous Ru^{2+}/Ru^{3+} redox. *Theor. Chem. Acc.* **2006**, *115*, 113–126.

(52) Li, H. Z.; Yang, W. Sampling enhancement for the quantum mechanical potential based molecular dynamics simulations: A general algorithm and its extension for free energy calculation on rugged energy surface. *J. Chem. Phys.* **2007**, *126*, 114104.

(53) Zheng, L.; Li, H.; Yang, W. In *From Computational Biophysics to Systems Biology (CBSB08)*; Hansmann, U. H. E., Meinke, J., Mohanty, S., Nadler, W., Zimmerman, O., Eds.; NIC: Jülich, Germany, 2008; NIC Series Vol. 36, pp 57−64.

(54) Woods, C. J.; Manby, F. R.; Mulholland, A. J. An efficient method for the calculation of quantum mechanics/molecular mechanics free energies. *J. Chem. Phys.* **2008**, *128*, 014109.

(55) Zeng, X. C.; Hu, H.; Hu, X. Q.; Cohen, A. J.; Yang, W. T. Ab initio quantum mechanical/molecular mechanical simulation of electron transfer process: Fractional electron approach. *J. Chem. Phys.* **2008**, *128*, 124510.

(56) Rosta, E.; Haranczyk, M.; Chu, Z. T.; Warshel, A. Accelerating QM/MM free energy calculations: Representing the surrounding by an updated mean charge distribution. *J. Phys. Chem. B* **2008**, *112*, 5680–5692.

(57) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in Ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **2009**, *113*, 1253–1272.

(58) Zeng, X. C.; Hu, H.; Hu, X. Q.; Yang, W. T. Calculating solution redox free energies with ab inito quantum mechanical/molecular mechanical minimum free energy path method. *J. Chem. Phys.* **2009**, *130*, 164111.

(59) Cheng, J.; Sulpizi, M.; Sprik, M. Redox potentials and pKa for benzoquinone from density functional theory based molecular dynamics. *J. Chem. Phys.* **2009**, *131*, 154504.

(60) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions - Dielectric, electrostatic and steric stabilization of carbounium-ion in reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(61) 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.

(62) Gao, J. L. Hybrid quantum and molecular mechanical simulations: An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* **1996**, *29*, 298–305.

(63) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(64) Monard, G.; Merz, K. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.* **1999**, *32*, 904–911.

(65) Berne, B. J.; Straub, J. E. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181–189.

(66) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96–123.

(67) Okamoto, Y. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 425–439.

(68) Zheng, L. Q.; Chen, M. G.; 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.

(69) Gao, J. L.; Li, N. Q.; Freindorf, M. Hybrid QM/MM simulations yield the ground and excited state pK(a) difference: Phenol in aqueous solution. *J. Am. Chem. Soc.* **1996**, *118*, 4912–4913.

(70) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; Guo, H.; Elstner, M.; Cui, Q. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

(71) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling: A new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(72) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(73) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.

(74) Min, D. H.; Li, H. Z.; Li, G. H.; Bitetti-Putzer, R.; Yang, W. Synergistic approach to improve "alchemical" free energy calculation in rugged energy surface. *J. Chem. Phys.* **2007**, *126*, 144109.

(75) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.

(76) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(77) Ensing, B.; De Vivo, M.; Liu, Z. W.; Moore, P.; Klein, M. L. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* **2006**, *39*, 73–81.

(78) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(79) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Calfisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(80) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **2001**, *105*, 569–585.

(81) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinello, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *10*, 1961–1972.

(82) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

(83) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. L. The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: Formulation and tests of the analytical derivatives. *Theor. Chem. Acc.* **2000**, *104*, 336–343.

(84) Pu, J. Z.; Gao, J. L.; Truhlar, D. G. Combinding self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A* **2004**, *108*, 5454–5463.

(85) Nam, K.; Gao, J. L.; York, D. M. An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.

(86) Li, H.; Yang, W. Forging the missing link in free energy estimations: λ-WHAM in thermodynamic integration, overlap histogramming, and free energy perturbation. *Chem. Phys. Lett.* **2007**, *440*, 155–159.

(87) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(88) Anwar, J.; Heyes, D. M. Robust and accurate method for free-energy calculation of charged molecular systems. *J. Chem. Phys.* **2005**, *122*, 224117.

(89) Min, D. H.; Liu, Y. S.; Carbone, I.; Yang, W. On the convergence improvement in the metadynamics simulations: A Wang-Landau recursion approach. *J. Chem. Phys.* **2007**, *126*, 194104.

(90) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714.

(91) 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.

(92) Nose, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(93) Nose, S.; Klein, M. L. Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(94) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-pressure molecular-dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(95) Wardman, P. Reduction potentials of one-electron couples involving free-radicals in aqueous-solution. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1637–1755.

(96) Donald, W. A.; Leib, R. D.; O'Brien, J. T.; Bush, M. F.; Williams, E. R. Absolute standard hydrogen electrode potential measured by reduction of aqueous nanodrops in the gas phase. *J. Am. Chem. Soc.* **2008**, *130*, 3371–3381.

(97) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J. Phys. Chem. B* **2006**, *110*, 16066–16081.

(98) Reiss, H. The Fermi level and the redox potential. *J. Phys. Chem.* **1985**, *89*, 3783–3791.

(99) Gomer, R.; Tryson, G. Experimental-determination of absolute half-cell EMFS and single ion free-energies of solvation. *J. Chem. Phys.* **1977**, *66*, 4413–4424.

(100) Marcus, Y. Thermodynamics of solvation of ions. 5. Gibbs free-energy of hydration at 298.1 K. *J. Chem. Soc. Faraday Trans.* **1991**, *87*, 2995–2999.

(101) Parchment, O. G.; Green, D. V. S.; Taylor, P. J.; Hillier, I. H. The prediction of tautomer equilibria in hydrated 3-Hydroxypyrazole: A challenge to theory. *J. Am. Chem. Soc.* **1993**, *115*, 2352–2356.

CT100033S