Accelerated Computation of Free Energy Profile at ab initio QM/MM Accuracy via a Semi-Empirical Reference-Potential. III. Gaussian Smoothing on Density-of-States

Wenxin Hu,[†] Pengfei Li,[*,‡,⊥] Jia-Ning Wang,[‡] Yuanfei Xue,[‡] Yan Mo,[*,‡,¶,§] Jun Zheng,[†] Xiaoliang Pan,[‖] Yihan Shao,[‖] and Ye Mei[*,‡,¶,§]

†*The Computer Center, School of Data Science & Engineering, East China Normal University, Shanghai 200062, China*

‡*State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China*

¶*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

§*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

‖*Department of Chemistry and Biochemistry, University of Oklahoma, Norman OK 73019, United States of America*

⊥*Current address: Silicon Therapeutics (Suzhou) Co., Ltd., Suzhou, Jiangsu 215000, China*

E-mail: alan.pengfeili@gmail.com; ymo@phy.ecnu.edu.cn; ymei@phy.ecnu.edu.cn

Abstract

Calculations of free energy profile, aka potential of mean force (PMF), along a chosen collective variable (CV) are now routinely applied to the studies of chemical

processes, such as enzymatic reactions and chemical reactions in condensed phases. However, if the ab initio QM/MM level of accuracy is required for the PMF, it can be formidably demanding even with the most advanced enhanced sampling methods, such as umbrella sampling. To ameliorate this difficulty, we developed a novel method for the computation of free energy profile based on the reference-potential method recently, in which a low-level reference Hamiltonian is employed for phase space sampling and the free energy profile can be corrected to the level of interest (the target Hamiltonian) by energy reweighting in a nonparametric way. However, when the reference Hamiltonian is very different from the target Hamiltonian, the calculated ensemble averages, including the PMF, often suffer from numerical instability, which mainly comes from the overestimation of the density-of-states (DoS) in the low-energy region. Stochastic samplings of these low-energy configurations are rare events. If a low-energy configuration has been sampled with a small sample size N, the probability of visiting this energy region is $\sim 1/N$ (shall be exactly $1/N$ for a single ensemble), which can be orders-of-magnitude larger than the actual DoS. In this work, an assumption of Gaussian distribution is applied to the DoS in each CV bin, and the weight of each configuration is rescaled according to the accumulated DoS. The results show that this smoothing process can remarkably reduce the ruggedness of the PMF and increase the reliability of the reference-potential method.

Introduction

Molecular dynamics (MD) simulations are now playing a more and more important role in the studies of thermodynamic properties in complex systems with a large number of degrees of freedom, such as enzymatic reactions and chemical reactions in condensed phases. However, these simulations are formidably demanding even with contemporary supercomputers, due that a quantum mechanical description of bond-forming or breaking is imperative. For instance, it may take about 100 seconds of wall-clock time on an off-the-shelf computer node for a single step of *ab initio* quantum mechanical/molecular mechanical (aiQM/MM) MD propagation[1–4] with a medium-sized molecule in the QM region, and a simulation that is long enough to yield statistically meaningful observations may require 10^6 to 10^9 steps of propagation to reach timescales of nanoseconds to microseconds. To reduce the computational demand, a number of enhanced sampling techniques have been proposed by either lowering the barrier or increasing the temperature, or doing both.[5–7] Since the total CPU time equals to the number of steps of propagation multiplied by the CPU time for a single step, the total cost can be reduced if the cost for a single step of propagation gets cheaper. There has been a growing interest in employing the reference-potential methods, developed independently by Gao,[8] and by Warshel and co-workers.[9] In this approach, the difference in state free energies or free energy profile along a predefined collective variable (CV) under an accurate, and usually expensive, Hamiltonian is obtained by energy reweighing of the trajectories from an inexpensive but less reliable Hamiltonian. The latter is often referred to as the reference Hamiltonian and the former as the target Hamiltonian. Simulations using the reference Hamiltonian can remarkably boost the efficiency in exploring the phase space by reducing the cost for a single step of propagation. Probabilities (or weights) of the configurations are then reassigned based on single-point energy calculations under the target Hamiltonian. Ensemble averages can be obtained for any time-independent thermodynamic quantities, including the free energy difference between two states and the free energy profile as a function of an order parameter, can be obtained using the nonuniform weights.

This idea has been proved quite successful in several recent works.[10–33] However, the statistical efficiency depends on the similarity between the reference and the target Hamiltonians. For instance, in umbrella sampling (US),[34] the reference Hamiltonians used in simulations are the biased Hamiltonians, which are usually defined as the unbiased Hamiltonian plus some biasing terms. The target Hamiltonian is exactly the unbiased Hamiltonian. Therefore, the stability, for most cases, is pretty satisfactory, and thus US is usually not categorized as a reference-potential method. However, if the target Hamiltonian is very different from the reference Hamiltonian, the convergence of the reweighting process can be troublesome, as discussed in our recent work.[24] Further smoothing of the free energy curve, for instance, using Gaussian process regression[35] is indispensable. The numerical instability comes from the exponential term, of which the ensemble average often shows random fluctuations with nonnegligible amplitude. A celebrated example is the thermodynamic perturbation (TP) calculation of free energy difference between two thermodynamic states. To deal with this numerical difficulty, the delta method is usually employed by expanding the ensemble average as a series of polynomials.[36] If a certain approximation can be well applied to the distribution of this energy term that appears in the exponent, the convergence can be considerably accelerated. Truncation after the second-order term of the cumulant series accelerates the convergence of the TP calculations.[37]

In this work, we applied an assumption of Gaussian distribution to the density-of-states (DoS) in the calculation of potential of mean force (PMF) for two chemical reactions in aqueous solution using the reference-potential method. The results show that convergence can be remarkably improved.

Methods

Multistate Bennett Acceptance Ratio Formulation for the Reference-Potential Methods

In our previous work,[24] the reweighting process, termed the weighted thermodynamic perturbation (wTP), was separated from the weight calculations using the Multistate Bennett Acceptance Ratio (MBAR).[38] However, these two steps can be integrated into the MBAR formulation, which will be shown below.

Suppose K simulations have been conducted, each with a reduced potential energy function $u_k(\mathbf{r}) = \beta_k U_k(\mathbf{r})$. For umbrella sampling,[34] which entails the application of restraints $W_k(\mathbf{r})$ on the original Hamiltonian U_0 to enforce the sampling in specific regions in phase space, the potential energy function U_k reads

$$U_k(\mathbf{r}) = U_0(\mathbf{r}) + W_k(\mathbf{r}). \tag{1}$$

The temperatures of the coupled bath for the simulations may also differ from each other, for instance as in the temperature-replica exchange molecular dynamics (REMD).[39] For simplicity, we will use reduced potential energy functions in the following, and omit the prefix "reduced" for simplicity. The free energy f_i associated with the potential energy function u_i can be computed using the MBAR analysis

$$f_i = -\ln \sum_{n=1}^{N} \frac{\exp\left[-u_i(\mathbf{r}_n)\right]}{\sum_{k=1}^{K} N_k \exp\left[f_k - u_k(\mathbf{r}_n)\right]}, \quad \forall i = 1 \ldots, K \tag{2}$$

where N_k is the number of configurations collected from the kth simulation, and $N = \sum_k N_k$. Then these data can be used in an extrapolation to any other states, for instance, the

unbiased state ($u_0(\mathbf{r})$) in umbrella sampling, as

$$\langle \hat{\mathbf{A}} \rangle_0 = \sum_{n=1}^{N} \omega_0(\mathbf{r}_n) \hat{\mathbf{A}}(\mathbf{r}_n), \tag{3}$$

with the normalized weight

$$\omega_0(\mathbf{r}_n) = \frac{\exp\left[f_0 - u_0(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - u_k(\mathbf{r}_n)\right]}, \tag{4}$$

where $\hat{\mathbf{A}}$ is an arbitrary operator dependent only on the coordinate $\mathbf{r}$, and $\langle \cdot \rangle$ denotes the ensemble average with the state indicated by the subscript, and f_0 is the free energy of the unbiased state. The partition function Q_0, aka the normalization constant, is

$$Q_0 = e^{-f_0} = \sum_{n=1}^{N} \frac{\exp\left[-u_0(\mathbf{r}_n)\right]}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - u_k(\mathbf{r}_n)\right]}. \tag{5}$$

When $\hat{\mathbf{A}}$ is an indication function δ of some chosen CV

$$\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) = \begin{cases} 1, & \text{if } -\Delta\boldsymbol{\xi}/2 < \boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n) < \Delta\boldsymbol{\xi}/2 \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

we have the PMF

$$F_0(\boldsymbol{\xi}_m) = -\ln \sum_{n=1}^{N} \omega_0(\mathbf{r}_n) \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) \tag{7}$$

defined up to an additive constant. Specifically, for umbrella sampling, $u_k(\mathbf{r}) = u_0(\mathbf{r}) + w_k(\mathbf{r})$, the weight can be simplified as

$$\omega_0(\mathbf{r}_n) = \frac{\exp\left(f_0\right)}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}, \tag{8}$$

which does not depend explicitly on the unbiased potential energy function $u_0(\mathbf{r})$.

If we are interested in another (the target) Hamiltonian $u_t(\mathbf{r})$ other than the unbiased Hamiltonian, the normalized weight function becomes[40]

$$\omega_t(\mathbf{r}_n) = \frac{\exp\left[f_t - u_t(\mathbf{r}_n)\right]}{\sum_{k=1}^{K} N_k \exp\left[f_k - u_k(\mathbf{r}_n)\right]}, \tag{9}$$

with f_t being the free energy associated to this Hamiltonian

$$f_t = -\ln \sum_{n=1}^{N} \frac{\exp\left[-u_t(\mathbf{r}_n)\right]}{\sum_{k=1}^{K} N_k \exp\left[f_k - u_k(\mathbf{r}_n)\right]}. \tag{10}$$

By subtracting the unbiased potential energy function $u_0(\mathbf{r})$ from the exponential terms on both the numerator and denominator, it can be rewritten as

$$\omega_t(\mathbf{r}_n) = \frac{\exp\left[f_t - \Delta u_t(\mathbf{r}_n)\right]}{\sum_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}, \tag{11}$$

where $\Delta u_t(\mathbf{r}) = u_t(\mathbf{r}) - u_0(\mathbf{r})$ is the energy difference between the target Hamiltonian and the unbiased Hamiltonian.

By defining the PMF in the mth CV bin as the state free energy corresponding to the Hamiltonian $u_t(\mathbf{r})$ with a screening factor $\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}))$ on the configurations, the covariance matrix can be computed as[38]

$$\hat{\boldsymbol{\Theta}} = \mathbf{W}^T \left(\mathbf{I}_N - \mathbf{W}\mathbf{N}\mathbf{W}^T\right)^+ \mathbf{W}, \tag{12}$$

where $\mathbf{I}_N$ is the $N \times N$ identify matrix with $N = \sum_{k=1}^{K} N_k$ being the total number of samples from all the simulations, $\mathbf{W}$ is the weight matrix with a dimension of $N \times (K + M)$ with M being the number of CV bins, $\mathbf{N} = \mathrm{diag}(N_1, N_2, \ldots, N_K, 0_1, 0_2, \ldots, 0_M)$, and the subscript $^+$

denotes the pseudoinverse. The variance of the PMF for the mth bin can be calculated as

$$\delta^2 F(\boldsymbol{\xi}_m) = \hat{\boldsymbol{\Theta}}_{K+m,K+m} + \hat{\boldsymbol{\Theta}}_{K+1,K+1} - 2\hat{\boldsymbol{\Theta}}_{K+1,K+m}, \tag{13}$$

if $F(\boldsymbol{\xi}_1)$ is fixed to zero. The reweighting entropy[22] for the samples collected in the mth bin is defined as

$$\mathcal{S}(\boldsymbol{\xi}_m) = -\frac{\sum_{n=1}^{N} \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))\omega_t(\mathbf{r}_n) \ln \omega_t(\mathbf{r}_n)}{\ln \sum_{n=1}^{N} \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))}, \tag{14}$$

which serves as another reliability metric for the MBAR calculations.

Gaussian Smoothing of the Density-of-States

The normalized weight function can be rewritten as the product of three terms

$$\omega_t(\mathbf{r}_n) = P(\mathbf{r}_n) \cdot \exp\left[-\Delta u_t(\mathbf{r}_n)\right] \cdot \exp\left(f_t\right), \tag{15}$$

in which

$$P(\mathbf{r}_n) = \frac{1}{\sum_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}, \tag{16}$$

and only depends on the Hamiltonians $\{u_k\}$ used in configuration sampling. The partition function associated with the target Hamiltonian is

$$\begin{aligned} Q_t &= \sum_{n=1}^{N} P(\mathbf{r}_n) \cdot \exp\left[-\Delta u_t(\mathbf{r}_n)\right] \\ &= \int_{-\infty}^{\infty} \sum_{n=1}^{N} \delta(\Delta u - \Delta u_t(\mathbf{r}_n)) P(\mathbf{r}_n) \cdot \exp\left(-\Delta u\right) \mathrm{d}\Delta u. \end{aligned} \tag{17}$$

Therefore,

$$\begin{aligned}\Omega(\Delta u) =& \sum_{n=1}^{N} \delta(\Delta u - \Delta u_t(\mathbf{r}_n)) P(\mathbf{r}_n) \\ =& \sum_{n=1}^{N} \frac{\delta(\Delta u - \Delta u_t(\mathbf{r}_n))}{\sum_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}\end{aligned} \tag{18}$$

is analogous to DoS. Equation 17 serves as a conversion from the work-based expression for the partition function to the DoS-based one.[41] In the same way, the potential of mean force under the target Hamiltonian becomes

$$\begin{aligned}F_t(\boldsymbol{\xi}_m) =& -\ln \sum_{n=1}^{N} \omega_t(\mathbf{r}_n) \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) \\ =& -\ln \int_{-\infty}^{\infty} \sum_{n=1}^{N} \frac{\delta(\Delta u - \Delta u_t(\mathbf{r}_n))}{\sum_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]} \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)) \exp\left(-\Delta u\right) \mathrm{d}\Delta u + C \\ =& -\ln \int_{-\infty}^{\infty} \Omega(\Delta u)_{\boldsymbol{\xi}_m} \exp\left(-\Delta u\right) \mathrm{d}\Delta u + C\end{aligned} \tag{19}$$

where C is an arbitrary constant and

$$\Omega(\Delta u)_{\boldsymbol{\xi}_m} = \sum_{n=1}^{N} \frac{\delta(\Delta u - \Delta u_t(\mathbf{r}_n))}{\sum_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]} \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n)). \tag{20}$$

Asymptotically with an infinite number of configurations, $\Omega(\Delta u)_{\boldsymbol{\xi}}$ follows a Gaussian distribution. However, with a limited number of samples in each bin, usually $\sum_{n=1}^{N} \delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))$ is on an order of 10^3 to 10^4, the distribution may deviate more or less from being perfectly Gaussian. It can be easily seen from Eq. 19 that configurations with more negative Δu may have large contributions to the integrand, due to the existence of the exponential term $\exp\left(-\Delta u\right)$. Their contributions are attenuated by the DoS $\Omega(\Delta u)_{\boldsymbol{\xi}_m}$.

Kofke et al realized the bias in the free energy may come from the tail-truncation error,

and proposed a Π-metric for estimating the reliability of the free energy.[42–45] In their model, they assumed a perfect sampling beyond an energy threshold, below which no samples have been obtained. However, there is another possibility, which is exactly opposite to Kofke's concern, i.e., a low energy configuration has been harvested among a limited number of samples leading to an overestimated DoS ($\sim 1/N$). It is detrimental to the smoothness of the free energy profile, and frequently emerges in the calculations using the reference-potential methods, especially when the target Hamiltonian is relatively far away from the reference Hamiltonian.

In this work, the configurational weights under the target Hamiltonian are scaled to generate a smooth Gaussian distribution in energy space, and the smoothing is performed in each configuration bin. For the samples falling in the mth CV bin $[\boldsymbol{\xi}_m - \frac{1}{2}\Delta\boldsymbol{\xi}, \boldsymbol{\xi}_m + \frac{1}{2}\Delta\boldsymbol{\xi}]$, the average energy difference $\overline{\Delta u}_{\boldsymbol{\xi}_m}$ and the variance $\sigma_{\boldsymbol{\xi}_m}$ can be calculated as

$$\overline{\Delta u}_{\boldsymbol{\xi}_m} = \frac{\sum\limits_{n=1}^{N} \Delta u(\mathbf{r}_n) \frac{\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}}{\sum\limits_{n=1}^{N} \frac{\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}} \tag{21}$$

and

$$\sigma_{\boldsymbol{\xi}_m} = \sqrt{\frac{\sum\limits_{n=1}^{N} \left(\Delta u(\mathbf{r}_n) - \overline{\Delta u}_{\boldsymbol{\xi}_m}\right)^2 \frac{\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}}{\sum\limits_{n=1}^{N} \frac{\delta(\boldsymbol{\xi}_m - \boldsymbol{\xi}(\mathbf{r}_n))}{\sum\limits_{k=1}^{K} N_k \exp\left[f_k - w_k(\mathbf{r}_n)\right]}}}. \tag{22}$$

In a narrow range $\left[\Delta u - \frac{1}{2}\delta\Delta u, \Delta u + \frac{1}{2}\delta\Delta u\right]$ in energy space (e.g. $\delta\Delta u = 0.2$ in this work), the cumulative probability of the Gaussian distribution of the density-of-states is

$$\rho_G\left(\left[\Delta u - \frac{1}{2}\delta\Delta u, \Delta u + \frac{1}{2}\delta\Delta u\right]\right) = \frac{1}{\sqrt{2\pi}\sigma_m} \int_{\Delta u - \frac{1}{2}\delta\Delta u}^{\Delta u + \frac{1}{2}\delta\Delta u} \exp\left[-\frac{\left(\Delta u - \overline{\Delta u}_{\boldsymbol{\xi}_m}\right)^2}{2\sigma_m^2}\right] \mathrm{d}\Delta u. \tag{23}$$

By summing up the density-of-states of the pooled configurations, the cumulative probability

is

$$\rho_S\left(\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]\right)=\frac{\sum_{n=1}^{N}\frac{\delta(\boldsymbol{\xi}_m-\boldsymbol{\xi}(\mathbf{r}_n))\delta(\Delta u-\Delta u(\mathbf{x}_n))}{\sum_{k=1}^{K}N_k\exp\left[f_k-w_k(\mathbf{r}_n)\right]}}{\sum_{n=1}^{N}\frac{\delta(\boldsymbol{\xi}_m-\boldsymbol{\xi}(\mathbf{r}_n))}{\sum_{k=1}^{K}N_k\exp\left[f_k-w_k(\mathbf{r}_n)\right]}},\tag{24}$$

which unnecessarily agrees with ρ_G exactly. To smooth the density-of-states, scale factors are thus defined as

$$\alpha\left(\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]\right)=\frac{\rho_G\left(\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]\right)}{\rho_S\left(\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]\right)},\tag{25}$$

and the weights are modified by this factor as

$$\omega_t'(\mathbf{r}_n)=\frac{\exp\left[f_t-\Delta u_t(\mathbf{r}_n)\right]}{\sum_{k=1}^{K}N_k\exp\left[f_k-w_k(\mathbf{r}_n)\right]}\cdot\alpha\left(\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]\right),\tag{26}$$

if the configuration $\mathbf{r}_n$ falls into the mth CV bin $[\boldsymbol{\xi}_m-\frac{1}{2}\Delta\boldsymbol{\xi},\boldsymbol{\xi}_m+\frac{1}{2}\Delta\boldsymbol{\xi}]$ and the energy bin $\left[\Delta u-\frac{1}{2}\delta\Delta u,\Delta u+\frac{1}{2}\delta\Delta u\right]$. The potential of mean force is then computed with the new weights

$$F_t(\boldsymbol{\xi}_m)=-\ln\sum_{n=1}^{N}\omega_t'(\mathbf{r}_n)\delta(\boldsymbol{\xi}_m-\boldsymbol{\xi}(\mathbf{r}_n))\tag{27}$$

up to an additive constant.

It is worth emphasizing that Gaussian smoothing for the weights $\omega_0(\mathbf{r})$ under the unbiased Hamiltonian $u_0(\mathbf{r})$ is unnecessary because $\Delta u(\mathbf{r})=0$ for all the configurations, which can be inferred by comparing Eq. 8 and 11.

Setup of the Simulations

Two reactions were studied. One is glycine intramolecular proton transfer reaction from the neutral form to the zwitterion form (see Fig. 1), and the other is Claisen rearrangement reaction of allyl vinyl ether to 4-pentenal (see Fig. 4). The simulation details have

been introduced in our previous work,[24] and here we only provide a brief description. For each system, the solute molecule was surrounded by a TIP3P[46] water sphere with a radius of 25 Å. The water sphere was restrained by a soft half-harmonic potential with a force constant of $10\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}\cdot\mathrm{Å}^{-2}$ to avoid evaporation. The quantum mechanical regions contained the solute molecules only. The nonbonded interactions were fully counted without any truncation. The van der Waals (vdW) parameters were taken from the general AMBER force field.[47] The low-level (reference) and the high-level (target) Hamiltonians were PM3[48] and B3LYP/6-31G(d),[49,50] respectively. Phase space samplings were enhanced by US. The one-dimensional collective variable was defined as $\xi = d_{\mathrm{OH}} - d_{\mathrm{NH}}$ for the glycine intramolecular proton transfer reaction, where H is the hydrogen atom to be transferred. To ensure sufficient overlap between neighboring simulations, 61 windows centering on ξ ranging from -1.5 to $1.5\,\mathrm{Å}$ were applied. Force constant of the restraint ranged from 100 to $1350\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}\cdot\mathrm{Å}^{-2}$. For the Claisen rearrangement reaction, 95 windows simulations were carried out with the collective variable $\xi = d_{\mathrm{OC5}} - d_{\mathrm{C2C3}}$ ranging from -2.2 to $1.7\,\mathrm{Å}$. The force constant of the restraint ranged from 100 to $1600\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}\cdot\mathrm{Å}^{-2}$. For each US window simulation, the system was optimized for 500 steps using the steepest decent optimization method followed by 500 steps of the conjugate gradient method. The system was heated to 300 K in 50 ps and was equilibrated for 100 ps. Finally, a 1-ns production simulation was conducted in each window for free energy analysis. The integration time step was set to 0.5 fs for simulations of glycine protein transfer, and 1 fs for the simulations of the Claisen rearrangement reaction. The configurations were saved every 0.1 ps. The temperature was regulated at 300 K with the Andersen temperature coupling scheme.[51] All the simulations were performed by the AmberTools 16 program package[52] interfacing with the Q-Chem 4.3 package.[53] The MBAR analysis was carried out using our own code, which can be downloaded from https://github.com/samuelymei/MBAR. The Gaussian process regression[35] with a squared-exponential kernel was carried out using the scikit-learn package.[54]

Results and Discussion

Glycine Proton Transfer Reaction

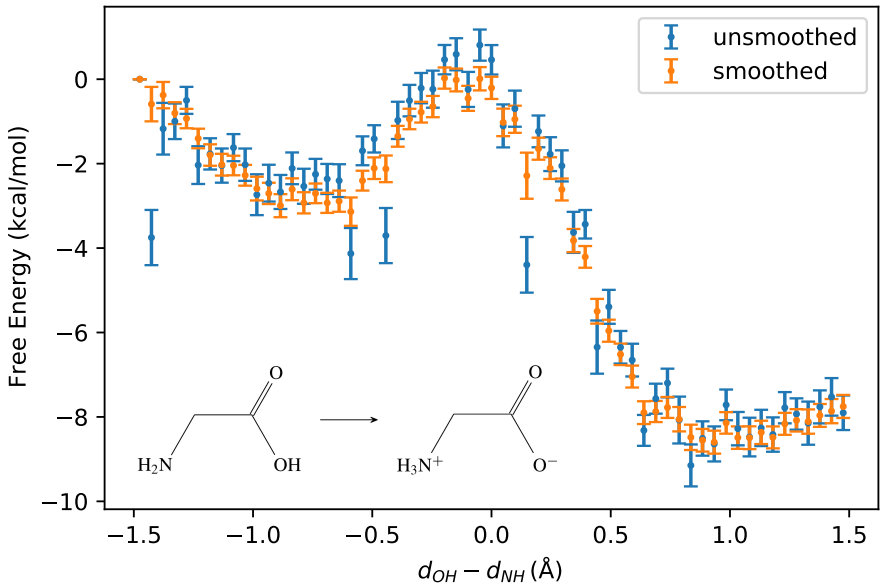


Figure 1: Free energy profiles of glycine proton transfer without (blue) and with (orange) Gaussian smoothing on the DoS of Δu.

The potential of mean force (PMF) at DFT level computed from the unsmoothed weights is shown in Fig. 7.(a) of Ref. 24, and is replotted in Fig. 1 for convenience. It clearly shows that sudden drops may occur somewhere along this curve, making the potential of mean force much lower than the neighboring values. For instance, the PMFs of the first and the third bins, with ξ centered at -1.475 Å and -1.377 Å, are 0 (which are fixed to zero) and -1.2 kcal/mol. While the PMF of the second bin, with $\xi = -1.426$ Å, is -3.8 kcal/mol. As shown in Fig. S13, the reweighting entropy S for the second bin is 2.0×10^{-3}, indicating that the PMF of this bin is determined by a small number of samples while contributions from other samples ($\omega_t(\mathbf{r}_n)$) are negligible. This observation also applies to the bins at $\xi = -0.590$, -0.443, 0.148, 0.443 and 0.836 Å.

The distribution of ΔU for the second bin, which contains 10002 configurations, is shown

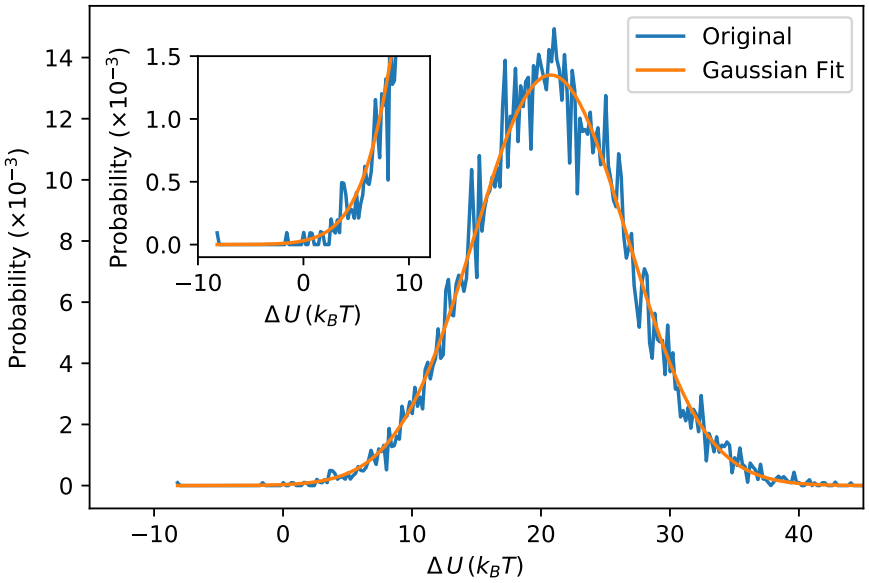


Figure 2: Density of state (blue) and its Gaussian interpolation (orange) in the bin centered at $\xi = -1.426\,\text{Å}$ for the glycine proton transfer. Inset: Enlarged view on the low energy tail of this distribution.

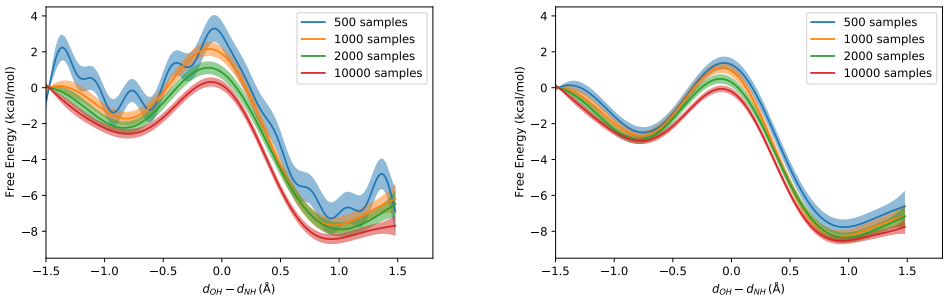
in Fig. 2. The distribution is generally Gaussian, which has a center at $\Delta U = 20.8\,\text{k}_\text{B}\text{T}$ and scale σ of $5.9\,\text{k}_\text{B}\text{T}$. However, at $\Delta U = -8.2\,\text{k}_\text{B}\text{T}$, the sampled probability is 9.3×10^{-5}, which is more than 3 orders larger than the predicted cumulative probability (9.5×10^{-8}) for a bin width $\delta\Delta u = 0.2\,\text{k}_\text{B}\text{T}$ from the fitted Gaussian distribution. This is the main cause for the local ruggedness of the sampled free energy profile. After Gaussian smoothing on the weights using Eq. 26, the PMF curve is shown in Fig. 1 as the orange curve with markedly suppressed ruggedness. The PMF of the second bin increases from $-3.8\,\text{kcal/mol}$ to $-0.5\,\text{kcal/mol}$, and the reweighting entropy S increases to 0.25, indicating a more reliable ensemble average. Similar observations can be seen for the energy distributions for the bins centered at $\xi = -0.590$, -0.443, 0.148, 0.443 and $0.836\,\text{Å}$, which can be found in Fig. S8 to S12 in SI. After Gaussian smoothing, the potential of mean force of these bins also increase, and stand well between their neighbors.

It will be interesting to investigate how these configurations with very negative ΔU are sampled. They should have relatively high energies under the reference Hamiltonian

(unbiased PM3) and, at the same time, relatively low energies under the target Hamiltonian (B3LYP/6-31G(d)). If the molecule under the target Hamiltonian is represented as coupled displaced-oscillators relative to the reference Hamiltonian in a simplified view, it is not unexpectedly some of the configurations may have very negative ΔUs. Besides, when the difference between these two Hamiltonians increases, the probability of encountering these configurations also increases. Some might assume that if we reduce the number of samples used in the MBAR analysis, these pathological configurations may disappear. However, this is not the case. It can be seen from Fig. S1 to S6 that low-ΔU configurations may appear randomly during the simulation. Even with 100 samples evenly extracted from the 10,000 samples of each window simulation, sudden drops in the PMF still occur. Moreover, when the number of samples from each window simulation increases from 2000 to 5000, the PMF at $\xi = 0.148$ drops by over $-4.2\,\mathrm{kcal/mol}$. Therefore, this numerical instability is inherent, and the discontinuity in the free energy profile is unavoidable with finite number of samples.

We then further smoothed the free energy profile curves by running Gaussian process regression on the unsmoothed and smoothed data. As shown in Fig. 3, Gaussian smoothing can accelerate the convergence of the PMF from GPR with respect to the number of samples. When 500 configurations were used per window simulation, the PMF undergoes large-amplitude fluctuations if Gaussian smoothing was not applied, and these is no definite location of the reactant and product states. In comparison, the counterpart with Gaussian smoothing has a single minimum at both of the reactant and the product states, and it is much closer to the PMF when larger number of configurations were used for analysis. In addition, Gaussian smoothing on the DoS can also shrink the 95% confidence interval of the PMF, although it does not significantly change the reaction barrier or the free energy difference between the product and the reactant when 10000 samples were used for analysis.

The cumulative distributions ρ_G and ρ_S in Eq. 23 and 24 depend on the resolution of $\delta\Delta u$. In Fig. S14 and S15, the differences in the free energy profiles are negligible with $\delta\Delta u$ ranging from $0.1\,\mathrm{k_B T}$ to $1.0\,\mathrm{k_B T}$, indicating that the impact of Gaussian smoothing on free



(a) unsmoothed (b) smoothed

Figure 3: Free energy profiles for glycine proton transfer reaction after Gaussian Process Regression, with 95% confidence interval also presented, using 500, 1000, 2000, and 10000 samples (a) without and (b) with Gaussian smoothing on the DoS.

energy profiles is insensitive to the exact value of $\delta\Delta u$, if it is below $1.0\,\mathrm{k_B T}$.

Claisen Rearrangement Reaction

The PMF at DFT level computed from the unsmoothed weights, shown in Fig. 9.(a) of Ref. 24, is replotted in Fig. 4 for convenience. Although the smoothness of the free energy profile looks much better than that of the glycine proton transfer, it still shows discontinuities in multiple bins along the CV. For the bin $\xi = -0.492$ Å near the peak of the free energy profile, the PMF is 26.0 kcal/mol, while those of its neighboring bins are 29.0 and 28.5 kcal/mol, respectively. This large decrease in PMF also comes from the overestimated probability of low-energy configurations, which can be seen from Fig. 5. The DoS distribution in this bin, by histogramming over 14196 configurations, is also approximately Gaussian, which centers at $\Delta U = -3.3\,\mathrm{k_B T}$ and has a scale σ of $5.5\,\mathrm{k_B T}$. However, the figure shows that at $\Delta U = -28.7\,\mathrm{k_B T}$, where the fitted Gaussian-distributed probability should be around 3.0×10^{-7}, the DoS has a cusp with a value of 1.3×10^{-4}, which is overestimated by nearly 3 orders. Similar observations are also seen for bins with $\xi = -1.573$, -0.343, 0.440, 0.626, and 0.961 Å (See Fig. S16 and S18 to S21 in SI for the energy distributions). Then the Gaussian smoothing was applied to the DoS in each bin, and the new free energy profile is much smoother as shown in Fig. 4. The PMF for the bin $\xi = -0.492$ Å is now 28.5 kcal/mol,

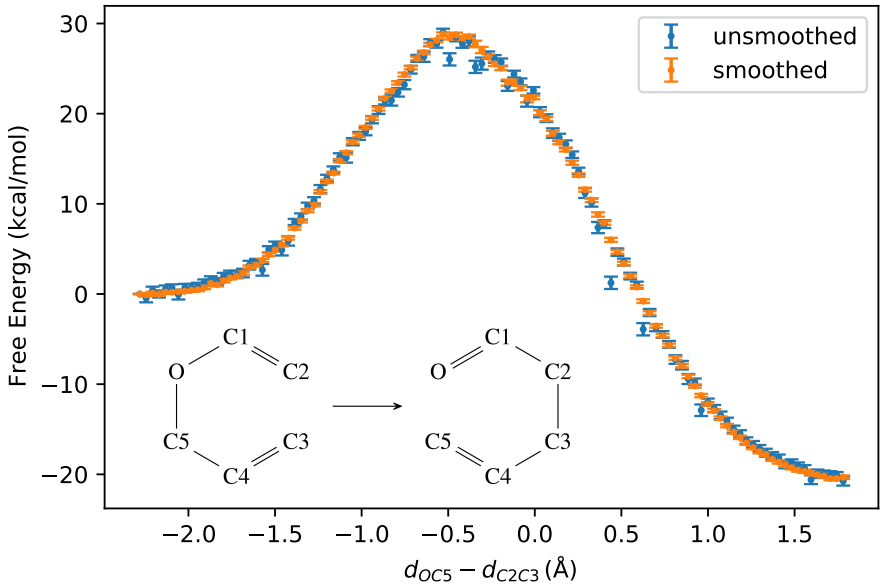


Figure 4: Free energy profiles of the Claisen rearrangement reaction without (blue) and with (red) Gaussian smoothing on the DoS of Δu.

jointing well with its neighbors. The reweighting entropy increases from 6×10^{-3} to 0.346, as shown in Fig. S22.

The GPR interpolated free energy profiles are shown in Fig. 6. Because the unsmoothed data show only small deviations from the smoothed data, these PMF curves have only small differences, with the Gaussian-smoothed ones converges slightly faster than the unsmoothed ones. This result shows that for numerically stable sampling Gaussian smoothing on the DoS does not deteriorate the ensemble average.

Conclusions

Free energy profiles calculations using reference-potential methods can significantly enhance computational efficiency. However, when the overlap in phase space between the reference Hamiltonian and the target Hamiltonian is small, numerical difficulties may arise and cause discontinuities in the free energy profiles. This numerical instability mainly comes from

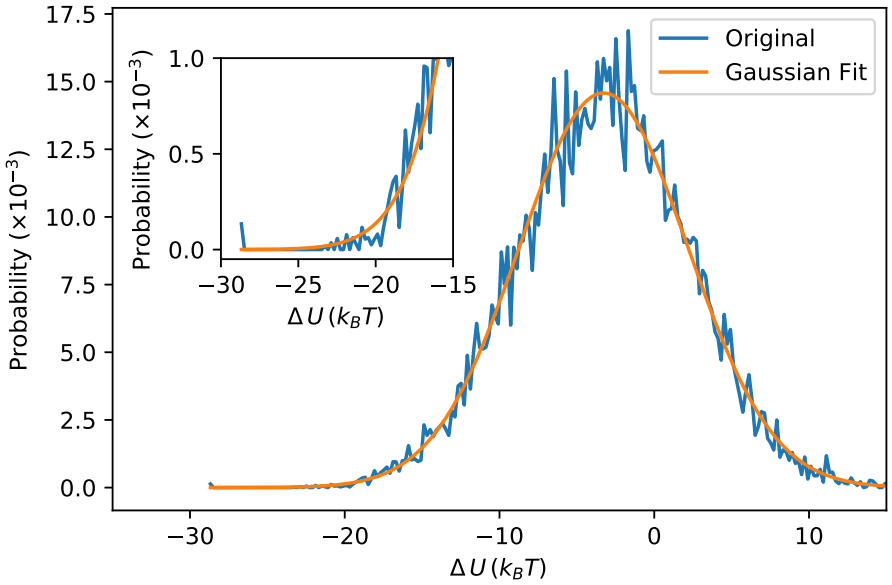
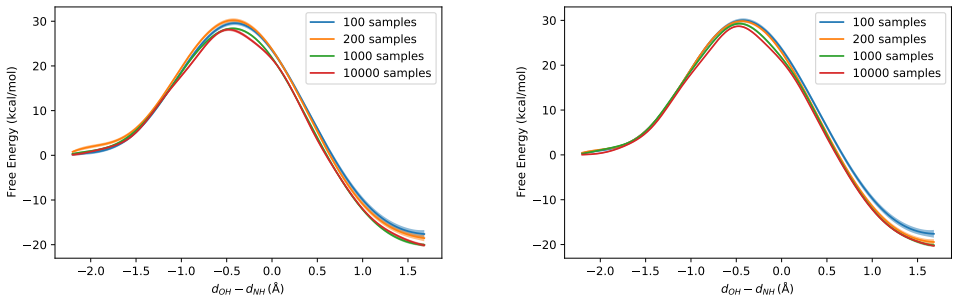


Figure 5: Density of state (blue) and its Gaussian interpolation (orange) in the bin centered at $\xi = -0.492\,\text{Å}$ for the Claisen rearrangement reaction. Inset: Enlarged view on the low energy tail of this distribution.

the overestimated density-of-states in some low-energy regions, which may introduce large noise into the potential of mean force via the exponential term in the working equations of Multistate Bennett Acceptance Ratio. In this work, Gaussian smoothing is applied to the density-of-states in each bin along the chosen collective variable (CV), and the results show that the numerical stability can be improved and the ruggedness in the free energy profile can be significantly suppressed. It is worth emphasizing that this smoothing process was carried out in all the bins one-by-one, not just the bins showing discontinuities along the free energy profiles. We also noticed that this Gaussian smoothing process is numerically robust in terms of the resolution $\delta\Delta u$ in computing the discrete density-of-states.

In this work, the free energy profiles are computed in a nonparametric way by summing up the weights of the configurations falling into each CV bin (or histogramming) independently and are further smoothed using Gaussian process regression. The numerical difficulty arises in the nonparametric summing-up step and may have an impact on the second step. We noticed some recent progress in Bayesian interference for free energy calculations.[55–59] The



(a) unsmoothed (b) smoothed

Figure 6: Free energy profiles for Claisen rearrangement reaction after Gaussian Process Regression, with 95% confidence interval also presented, using 100, 200, 1000, and 10000 samples (a) without and (b) with Gaussian smoothing on the DoS.

output free energy profiles are guaranteed to be smooth since continuous basis functions are used. How this Gaussian smoothing process will affect the free energy profile from these parametric fitting methods will be investigated in our future studies.

Acknowledgement

Y. Mei is supported by the National Natural Science Foundation of China (Grant No. 21773066). W.H. is supported by the Fundamental Research Funds for the Central Universities. Y. Mo is supported by the National Natural Science Foundation of China (Grant No. 21973030). Y.S. is supported by the National Institutes of Health (Grant No. R01GM135392). CPU time was provided by the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation No. 001).

References

(1) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(2) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with *ab initio* Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

(3) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2006**, *117*, 185.

(4) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Edit.* **2009**, *48*, 1198–1229.

(5) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

(6) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41–62.

(7) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.

(8) Gao, J. Absolute Free Energy of Solvation from Monte Carlo Simulations Using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem.* **1992**, *96*, 537–540.

(9) 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.

(10) Rod, T. H.; Ryde, U. Quantum Mechanical Free Energy Barrier for an Enzymatic Reaction. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(11) Beierlein, F. R.; Michel, J.; Essex, J. W. A Simple QM/MM Approach for Capturing Polarization Effects in Protein-Ligand Binding Free Energy Calculations. *J. Phys. Chem. B* **2011**, *115*, 4911–4926.

(12) König, G.; Boresch, S. Non-Boltzmann Sampling and Bennett's Acceptance Ratio Method: How to Profit from Bending the Rules. *J. Comput. Chem.* **2011**, *32*, 1082–1090.

(13) Heimdal, J.; Ryde, U. Convergence of QM/MM Free-Energy Perturbations Based on Molecular-Mechanics or Semiempirical Simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592–12604.

(14) Polyak, I.; Benighaus, T.; Boulanger, E.; Thiel, W. Quantum Mechanics/Molecular Mechanics Dual Hamiltonian Free Energy Perturbation. *J. Chem. Phys.* **2013**, *139*, 064105.

(15) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *J. Chem. Theory Comput.* **2014**, *10*, 1406–1419.

(16) Hudson, P. S.; Woodcock, H. L.; Boresch, S. Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems. *J. Phys. Chem. Lett.* **2015**, *6*, 4850–4856.

(17) Hudson, P. S.; White, J. K.; Kearns, F. L.; Hodoscek, M.; Boresch, S.; Woodcock, H. L. Efficiently Computing Pathway Free Energies: New Approaches Based on Chain-of-Replica and Non-Boltzmann Bennett Reweighting Schemes. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 944–953.

(18) Jia, X.; Wang, M.; Shao, Y.; König, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J. Chem. Theory Comput.* **2016**, *12*, 499–511.

(19) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.

(20) Dybeck, E. C.; König, G.; Brooks, B. R.; Shirts, M. R. Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials. *J. Chem. Theory Comput.* **2016**, *12*, 1466–1480.

(21) Kearns, F. L.; Hudson, P. S.; Woodcock, H. L.; Boresch, S. Computing Converged Free Energy Differences between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities. *J. Comput. Chem.* **2017**, *38*, 1376–1388.

(22) Wang, M.; Li, P.; Jia, X.; Liu, W.; Shao, Y.; Hu, W.; Zheng, J.; Brooks, B. R.; Mei, Y. Efficient Strategy for the Calculation of Solvation Free Energies in Water and Chloroform at the Quantum Mechanical/Molecular Mechanical Level. *J. Chem. Inf. Model.* **2017**, *57*, 2476–2489.

(23) Li, P.; Liu, F.; Jia, X.; Shao, Y.; Hu, W.; Zheng, J.; Mei, Y. Efficient Computation of Free Energy Surfaces of Diels-Alder Reactions in Explicit Solvent at Ab Initio QM/MM Level. *Molecules* **2018**, *23*, 2487.

(24) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 5583–5596.

(25) Wang, M.; Mei, Y.; Ryde, U. Predicting Relative Binding Affinity Using Nonequilibrium QM/MM Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 6613–6622.

(26) Hudson, P. S.; Boresch, S.; Rogers, D. M.; Woodcock, H. L. Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching. *J. Chem. Theory Comput.* **2018**, *14*, 6327–6335.

(27) König, G.; Brooks, B. R.; Thiel, W.; York, D. M. On the Convergence of Multi-Scale Free Energy Simulations. *Mol. Simulat.* **2018**, *44*, 1062–1081.

(28) Wang, M.; Mei, Y.; Ryde, U. Host-guest Relative Binding Affinities at Density-Functional Theory Level from Semiempirical Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 2659–2671.

(29) Pan, X.; Li, P.; Ho, J.; Pu, J.; Mei, Y.; Shao, Y. Accelerated Computation of Free Energy Profile at ab initio Quantum Mechanical/Molecular Mechanical Accuracy via a Semi-Empirical Reference Potential. II. Recalibrating Semi-Empirical Parameters with Force Matching. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20595–20605.

(30) Li, P.; Liu, F.; Shao, Y.; Mei, Y. Computational Insights into Endo/Exo Selectivity of the Diels-Alder Reaction in Explicit Solvent at Ab Initio Quantum Mechanical/Molecular Mechanical Level. *J. Phys. Chem. B* **2019**, *123*, 5131–5138.

(31) Hudson, P. S.; Woodcock, H. L.; Boresch, S. Use of Interaction Energies in QM/MM Free Energy Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 4632–4645.

(32) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM $\rightarrow$ QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.

(33) Piccini, G.; Parrinello, M. Accurate Quantum Chemical Free Energies at Affordable Cost. *J. Phys. Chem. Lett.* **2019**, *10*, 3727–3731.

(34) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(35) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press, 2006; pp 7–32.

(36) Oehlert, G. W. A Note on the Delta Method. *Am. Stat.* **1992**, *46*, 27–29.

(37) Chipot, C.; Pohorille, A. In *Free Energy Calculations*; Chipot, C., Pohorille, A., Eds.; Springer: Heidelberg, 2007; pp 199–247.

(38) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

(39) Hansmann, U. H. E.; Okamoto, Y. Prediction of Peptide Conformation by Multicanonical Algorithm: New Approach to the Multiple-Minima Problem. *J. Comput. Chem.* **1993**, *14*, 1333–1338.

(40) Shirts, M. R. Reweighting from the Mixture Distribution as a Better Way to Describe the Multistate Bennett Acceptance Ratio. arXiv.org, 2017; https://arxiv.org/abs/1704.00891.

(41) Kofke, D. A.; Frenkel, D. In *Handbook of Materials Modeling: Methods*; Yip, S., Ed.; Springer Netherlands: Dordrecht, 2005; pp 683–705.

(42) Lu, N.; Kofke, D. A. Accuracy of Free-Energy Perturbation Calculations in Molecular Simulation. I. Modeling. *J. Chem. Phys.* **2001**, *114*, 7303–7311.

(43) Wu, D.; Kofke, D. A. Model for Small-Sample Bias of Free-Energy Calculations Applied to Gaussian-Distributed Nonequilibrium Work Measurements. *J. Chem. Phys.* **2004**, *121*, 8742–8747.

(44) Wu, D.; Kofke, D. A. Phase-Space Overlap Measures. I. Fail-Safe Bias Detection in Free Energies Calculated by Molecular Simulation. *J. Chem. Phys.* **2005**, *123*, 054103.

(45) Schultz, A. J.; Kofke, D. A. Identifying and Estimating Bias in Overlap-Sampling Free-Energy Calculations. *Mol. Simul.* **2020**, DOI: 10.1080/08927022.2020.1758695.

(46) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(47) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(48) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

(49) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(50) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(51) Andrea, T. A.; Swope, W. C.; Andersen, H. C. The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water. *J. Chem. Phys.* **1983**, *79*, 4576–4584.

(52) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, N.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. AMBER 2016, University of California, San Francisco. 2016.

(53) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.;

Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184–215.

(54) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(55) Habeck, M. Bayesian Reconstruction of the Density of States. *Phys. Rev. Lett.* **2007**, *98*, 200601.

(56) Habeck, M. Bayesian Estimation of Free Energies From Equilibrium Simulations. *Phys. Rev. Lett.* **2012**, *109*, 100601.

(57) Lee, T.-S.; Radak, B. K.; Pabis, A.; York, D. M. A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 153–164.

(58) Ferguson, A. L. BayesWHAM: A Bayesian Approach for Free Energy Estimation, Reweighting, and Uncertainty Quantification in the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2017**, *38*, 1583–1605.

(59) Shirts, M. R.; Ferguson, A. L. Statistically Optimal Continuous Free Energy Surfaces from Biased Simulations and Multistate Reweighting. *J. Chem. Theory Comput.* **2020**, *16*, 4107–4125.

Graphical TOC Entry