

# Multireference Generalization of the Weighted Thermodynamic Perturbation Method

Timothy J. Giese, Jinzhe Zeng, and Darrin M. York\*



Cite This: *J. Phys. Chem. A* 2022, 126, 8519–8533



Read Online

ACCESS |



Metrics & More

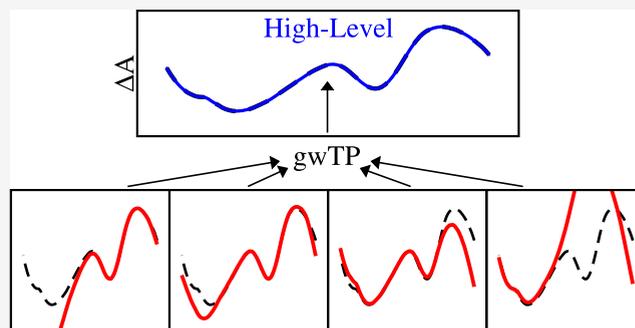


Article Recommendations



Supporting Information

**ABSTRACT:** We describe the generalized weighted thermodynamic perturbation (gwTP) method for estimating the free energy surface of an expensive “high-level” potential energy function from the umbrella sampling performed with multiple inexpensive “low-level” reference potentials. The gwTP method is a generalization of the weighted thermodynamic perturbation (wTP) method developed by Li and co-workers [*J. Chem. Theory Comput.* 2018, 14, 5583–5596] that uses a single “low-level” reference potential. The gwTP method offers new possibilities in model design whereby the sampling generated from several low-level potentials may be combined (e.g., specific reaction parameter models that might have variable accuracy at different stages of a multistep reaction). The gwTP method is especially well suited for use with machine learning potentials (MLPs) that are trained against computationally expensive ab initio quantum mechanical/molecular mechanical (QM/MM) energies and forces using active learning procedures that naturally produce multiple distinct neural network potentials. Simulations can be performed with greater sampling using the fast MLPs and then corrected to the ab initio level using gwTP. The capabilities of the gwTP method are demonstrated by creating reference potentials based on the MNDO/d and DFTB2/MIO semiempirical models supplemented with the “range-corrected deep potential” (DPRc). The DPRc parameters are trained to ab initio QM/MM data, and the potentials are used to calculate the free energy surface of stepwise mechanisms for nonenzymatic RNA 2'-O-transesterification model reactions. The extended sampling made possible by the reference potentials allows one to identify unequilibrated portions of the simulations that are not always evident from the short time scale commonly used with ab initio QM/MM potentials. We show that the reference potential approach can yield more accurate ab initio free energy predictions than the wTP method or what can be reasonably afforded from explicit ab initio QM/MM sampling.



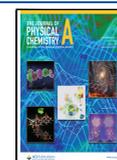
## INTRODUCTION

The study of chemical mechanisms of complex systems<sup>1</sup> has broad application in areas of heterogeneous, homogeneous, and enzyme catalysis,<sup>2</sup> synthetic chemistry,<sup>3</sup> and chemical education.<sup>4</sup> The chemical mechanism is often explored through calculation of a free energy surface (FES) in a reduced set of collective variables referred to as reaction coordinates.<sup>5</sup> The FES is used to characterize the location and free energy values of competitive reactive pathways connecting the reactant and product states. Numerous methods have been developed to calculate FESs from combined quantum mechanical/molecular mechanical (QM/MM) simulation. The approaches have been classified into three categories: methods based on the work of Jarzynski<sup>7</sup> which analyze nonequilibrium statistics,<sup>8–10</sup> methods that analyze equilibrium statistics generated from biased simulations (umbrella sampling),<sup>11–14</sup> and methods which introduce and integrate auxiliary degrees of freedom, such as  $\lambda$ -dynamics<sup>15–18</sup> and metadynamics.<sup>19,20</sup>

Umbrella sampling is a technique that applies an artificial bias to improve sampling efficiency. That is to say, a bias can

be chosen to enhance the sampling in the high-energy regions of the FES that would otherwise be difficult to sample on a reasonable time scale. If the FES was known a priori, then uniform sampling could be achieved by introducing a bias that exactly canceled the free energy. In practice, one instead performs a series of biased simulations that restrain the sampling to a particular region of reaction coordinates. The umbrella potentials are often chosen to be uncoupled harmonic oscillators, and the biased simulations differ by the choice of biasing force constants and equilibrium positions; however, the methods presented in the present work are general and do not assume a specific functional form for the biasing potential. For relatively simple mechanisms, the

**Received:** August 30, 2022  
**Revised:** October 14, 2022  
**Published:** October 27, 2022



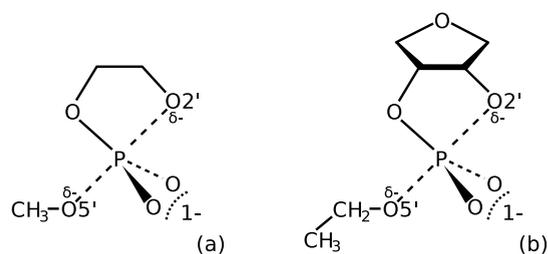
umbrella potentials can be chosen to span a predetermined range of reaction coordinates; however, methods have been developed to adaptively focus the sampling<sup>21–30</sup> and locate minimum free energy pathways.<sup>31–38</sup> There are several analysis techniques that can be used to reconstruct the FES from the biased simulations, including the weighted histogram analysis method (WHAM),<sup>39,40</sup> unbinned WHAM (UWHAM),<sup>41,42</sup> the multistate Bennett acceptance ratio method (MBAR),<sup>43</sup> the variational free energy profile (vFEP) method,<sup>44–46</sup> and umbrella integration.<sup>47–49</sup>

A series of recent studies have introduced the weighted thermodynamic perturbation (wTP) method.<sup>50–55</sup> This method estimates the FES of an expensive target potential from the umbrella sampling performed with a cost-effective reference potential. One can view the wTP method as being analogous to the reference potential method encountered in alchemical free energy applications,<sup>56–59</sup> and the accuracy of the wTP method similarly relies on close agreement between the target and reference potentials.<sup>60</sup> By performing simulations with an inexpensive reference potential, much greater sampling can be achieved through longer and/or more numerous independent simulations. Therefore, the wTP method is complimentary to other approaches that extend the sampling, such as multiple timestep integration.<sup>61</sup> Analysis of the extended simulations is useful for identifying unequilibrated sampling, and extended simulations become necessary in situations where multiple structural conformations are thermally accessible.

In the present work, we describe the generalized weighted thermodynamic perturbation (gwTP) method. The gwTP method estimates the FES of an expensive target potential from the aggregate umbrella sampling performed with multiple reference potentials. By combining the sampling, one no longer relies on a single reference potential to obtain good phase space overlap with the target potential throughout the entire FES.<sup>59,62,63</sup> Instead, only one of the reference potentials must have good phase space overlap in each region of the FES. If all of the potentials poorly overlapped with the target, then no reweighting strategy is likely to succeed from a limited number of target potential evaluations. Thus, it is important to have a quantitative measure of agreement between the reference and target potentials to gauge the reliability of the approach. The “reweighting entropy (RE)” is an index developed specifically for this purpose,<sup>64</sup> and we shall discuss it in more detail in subsequent sections. We were motivated to develop the gwTP method by the growing interest in supplementing semi-empirical QM/MM Hamiltonians with machine learning potentials (MLP) that are trained to reproduce ab initio QM/MM energies and forces.<sup>51,65–75</sup> If one had a priori confidence in the trained MLP, then one obviously can use it to estimate FESs without additional correction; however, if one is applying an MLP to a system that was not explicitly included in the training, then an MLP-corrected model naturally serves as an excellent reference potential to estimate the ab initio FES from reweighting. Furthermore, the active learning procedure used to train MLPs produces several neural network parameter sets (several potentials),<sup>76,77</sup> and the gwTP method provides a means to estimate the ab initio FES from the aggregate sampling performed with each potential. As multiple independent simulation runs are typically performed in order to produce robust averages and error estimates, the use of different reference potentials can often be accommodated for no added computational cost. The gwTP framework further

offers new possibilities in model design. Because the gwTP method can combine the sampling from several “specific reaction parameter” (SRP) models,<sup>78–80</sup> one can design several potentials that specialize their training to reproduce different chemical events embodied within a complex chemical mechanism, such as the general acid, general base, and phosphoryl transfer steps in RNA cleavage reactions.<sup>81</sup> Alternatively, one could parametrize several potentials that are trained to model the 2'-O-transesterification reaction in different RNA environments and then perform sampling with each potential when exploring the mechanism in an RNA environment not included in the training. In the above-mentioned discussion, the trained MLPs are used to perform the sampling; however, MLPs can also be utilized in another strategy whereby an uncorrected reference potential is sampled, and an MLP is trained to a small number of configurations to estimate the target potential energies required for thermodynamic perturbation.<sup>68,82–84</sup>

We apply the gwTP method to two nonenzymatic RNA 2'-O-transesterification model reactions shown in Figure 1. The



**Figure 1.** Nonenzymatic RNA 2'-O-transesterification model reactions explored in this work (atomic numbering of 2' and 5' positions in the model systems reflect their analogous positions in RNA). (a) Ethylene phosphate model reaction with a methoxide leaving group. (b) Native model reaction used in ref 88 with an ethoxide leaving group. The reaction coordinate used in this work is  $\xi_{PT} = R_{P-O5'} - R_{P-O2'}$ .

mechanism of closely related nonenzymatic phosphoryl transfer reactions have been explored with linear free energy relationships<sup>85,86</sup> and through the calculation of FESs.<sup>87,88</sup> These previous studies found that the pathway is correlated to the  $pK_a$  of the leaving group. Leaving groups with a  $pK_a < 11$  (“enhanced” leaving groups) proceed through a concerted mechanism containing a single, “early” ( $\xi_{PT} < 0$ ) transition state, whereas leaving groups with a  $pK_a > 12$  (“poor” leaving groups) proceed through two distinct barriers separated by a minimum. The methoxide and ethoxide leaving groups shown in Figure 1 are “poor” leaving groups, so the FESs are expected to contain an early transition state characterized by partial formation of the O2'-P bond and a second rate-controlling transition state characterized by partial cleavage of the O5'-P bond.

The ethylene phosphate reaction (Figure 1a) is used to demonstrate the gwTP method's ability to reconstruct a PBE0/6-31G\* target FES from the sampling obtained from four disparate reference potentials. For the purpose of providing a stringent test case for demonstration, the four reference potentials were specifically designed such that none of them accurately reproduce the target FES throughout the entire range of  $\xi_{PT}$  values. The reference potentials use the MNDO/d semiempirical Hamiltonian supplemented with a range-corrected deep potential<sup>76,88</sup> (DPRc) MLP. We trained

four ad hoc MNDO/d QM/MM + DPRc potentials using different target data to yield significantly different reference potentials. Each MNDO/d QM/MM + DPRc potential was trained to reproduce the PBE0/6-31G\* target energies and forces at different stages along the reaction coordinate (i.e., different subdomains of the  $\xi_{\text{PT}}$  reaction coordinate) such that the reference potentials disagree with each other and the target in the regions they were not trained. We will show that the wTP estimates of the target FES (based on a single reference potential) are inaccurate in the untrained regions, whereas the gwTP analysis of the aggregate sampling from all reference potentials accurately reproduces the PBE0/6-31G\* surface.

The native model reaction (Figure 1b) is used as a case study to emphasize the benefits offered by the reference potential approach. This reaction was one of the six nonenzymatic models used to parametrize the DFTB2/MIO QM/MM + DPRc potentials in ref 88. FESs of the PBE0/6-31G\* QM/MM, DFTB2/MIO QM/MM, and DFTB2/MIO QM/MM + DPRc potentials were obtained from identical simulation protocols. The DPRc correction was shown to improve the comparison with the ab initio results; however, some discrepancies between the PBE0/6-31G\* QM/MM and DFTB2/MIO QM/MM + DPRc native model profiles remained. In the present work, we revisit the calculation of the native model FES. We show that gwTP analysis of the DFTB2/MIO QM/MM + DPRc sampling performed in ref 88 results in excellent agreement with the PBE0/6-31G\* FES. Furthermore, we extend the DFTB2/MIO QM/MM + DPRc sampling from 100 ps/window to 1.2 ns/window of aggregate sampling—an amount of sampling well beyond what is routinely affordable using ab initio QM/MM simulation. The extended sampling made possible by the reference potential approach allows us to demonstrate that the native model reaction ab initio QM/MM free energy profiles presented in ref 88 are not converged. We show that a better estimate of the ab initio FES can be made using reference potentials with extended sampling than what can reasonably be afforded from ab initio QM/MM simulation.

## THEORY

**Multistate Bennett Acceptance Ratio Method.** In this section, we introduce our notation by reviewing the MBAR approach for calculating FESs. The description also serves to aid the reader's understanding of the differences between the MBAR, wTP, and gwTP methods. The FES of a system composed of 3N atomic coordinates  $\mathbf{r}$  is typically expressed in a reduced set of relevant collective variables  $\xi(\mathbf{r})$ , called reaction coordinates. Umbrella sampling is used to enhance the statistics in the high-energy regions of the FES by performing a series of  $K_h$  simulations of biased potential energy (PE) functions  $U_{hk}(\mathbf{r})$  that differ only by varying the bias  $W_{hk}(\mathbf{r})$  applied to the unbiased potential  $U_h(\mathbf{r})$ .

$$U_{hk}(\mathbf{r}) = U_h(\mathbf{r}) + W_{hk}(\mathbf{r}) \quad (1)$$

The subscripts  $h$  and  $k$  denote the unbiased PE function and biasing potential, respectively, and one can interpret  $(hk)$  as a combined index of biased states. The goal is to use MBAR to analyze the biased sampling performed with  $U_{hk}(\mathbf{r})$  to estimate the unbiased FES of  $U_h(\mathbf{r})$ . One could write the MBAR expressions without introducing the  $h$  subscript because only a single unbiased PE function is ever considered; however, the utility of its inclusion will become apparent when describing the wTP and gwTP methods in the ensuing sections. A

simulation performed with  $U_{hk}(\mathbf{r})$  at temperature  $T_{hk}$  produces an ensemble of  $N_{hk}$  structures. The 3N array of atomic coordinates of sample  $n$  in the ensemble of state  $hk$  is denoted as  $\mathbf{r}_{hkn}$ . We write the PE and biasing potential in reduced energy units:

$$u_{hk}(\mathbf{r}_{h'k'n}) = \beta_{h'k'} U_{hk}(\mathbf{r}_{h'k'n}),$$

$$u_h(\mathbf{r}_{h'k'n}) = \beta_{h'k'} U_h(\mathbf{r}_{h'k'n}), \quad \text{and} \quad w_{hk}(\mathbf{r}_{h'k'n}) = \beta_{h'k'} W_{hk}(\mathbf{r}_{h'k'n})$$

where  $\beta_{hk} = (k_B T_{hk})^{-1}$  and  $k_B$  is the Boltzmann constant.

The unbiased FES  $F_h(\xi)$  is, to within an additive constant, related to the probability of observing a sample at  $\xi$ ,  $\rho_h(\xi)$ .

$$F_h(\xi) = -\beta^{-1} \ln \rho_h(\xi) \quad (2)$$

In practice, the probability is approximated by discretizing  $\xi$  into histogram bins consisting of centers  $\xi_m$  and widths  $\Delta\xi$ . Let  $\delta(\xi_m - \xi(\mathbf{r}_{hkn}))$  denote the indicator function, which is 1 only if the sample is contained within bin  $m$ .

$$\delta(\xi_m - \xi(\mathbf{r}_{hkn})) = \begin{cases} 1, & \text{if } -\Delta\xi/2 < \xi_m - \xi(\mathbf{r}_{hkn}) < \Delta\xi/2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The probability of observing a sample in bin  $m$  is then given by eq 4, where  $\omega_h(\mathbf{r}_{hkn})$  is the weight of sample  $\mathbf{r}_{hkn}$  (see eq 6).

$$\rho_h(\xi_m) = \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_h(\mathbf{r}_{hkn}) \quad (4)$$

The MBAR expression for the FES of  $u_h(\mathbf{r})$  within histogram  $m$  is given by inserting eq 4 into eq 2.

$$F_h(\xi_m) = -\beta^{-1} \ln \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_h(\mathbf{r}_{hkn}) \quad (5)$$

If the sampling was performed with an unbiased PE, then the sample weights would be uniform and eqs 4 and 5 would merely count the fraction of samples observed within each bin. The situation is more complicated when umbrella sampling is performed because the distributions are skewed by the artificial bias. Reference 43 developed the equations to reweight trajectories using MBAR, and ref 50 specialized these equations to combine the biased sampling in FES applications. The result of these previous developments is shown in eq 6.

$$\omega_h(\mathbf{r}_{hkn}) = \frac{\exp[f_h - u_h(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - u_{hk'}(\mathbf{r}_{hkn})]} = \frac{\exp(f_h)}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - w_{hk'}(\mathbf{r}_{hkn})]} \quad (6)$$

The  $f_{hk}$  values are the free energies (in reduced energy units) of the  $K_h$  biased states obtained from solution to the MBAR/UWHAM equations.

$$f_{hl} = -\ln \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \frac{\exp[-u_{hl}(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - u_{hk'}(\mathbf{r}_{hkn})]}, \quad \forall l = 1, \dots, K_h \quad (7)$$

The  $f_h$  value is the free energy of unbiased PE  $u_h(\mathbf{r})$ . The numerator of eq 6 includes the factor  $\exp(f_h) = Q_h^{-1}$ ; this factor is the inverse of the unbiased state's partition function, which formally normalizes the weights.<sup>43,50</sup> In practice, the normalization constant shifts the entire FES by a constant.

$$f_h = -\ln \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \frac{\exp[-u_h(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - u_{hk'}(\mathbf{r}_{hkn})]} \quad (8)$$

### Weighted Thermodynamic Perturbation Method.

The wTP method has been developed over a series of recent articles.<sup>50–55</sup> The purpose of the method is to predict the FES of an expensive “target” PE function from the umbrella sampling performed with an inexpensive “reference” PE function. The unbiased target  $u_t(\mathbf{r}_{hkn})$  and reference  $u_h(\mathbf{r}_{hkn})$  PE functions of each sample are evaluated, and the energy differences  $\Delta u_{th}(\mathbf{r}_{hkn}) = u_t(\mathbf{r}_{hkn}) - u_h(\mathbf{r}_{hkn})$  are used to predict the target FES. In this sense, the wTP method is analogous to a “reference potential method” often encountered in alchemical free energy applications.

The wTP FES estimate of the unbiased target potential is given by eqs 9 and 10.

$$F_t(\xi_m) = -\beta^{-1} \ln \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_t(\mathbf{r}_{hkn}) \quad (9)$$

$$\omega_t(\mathbf{r}_{hkn}) = \frac{\exp[f_t - u_t(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - u_{hk'}(\mathbf{r}_{hkn})]} \quad (10)$$

The  $f_t$  quantity is the free energy of the unbiased target potential.

$$f_t = -\ln \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \frac{\exp[-u_t(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - u_{hk'}(\mathbf{r}_{hkn})]} \quad (11)$$

Multiplying the numerator and denominator of eq 10 by  $\exp[u_h(\mathbf{r}_{hkn})]$  allows one to rewrite the expression in terms of the energy difference.

$$\begin{aligned} \omega_t(\mathbf{r}_{hkn}) &= \frac{\exp[f_t - \Delta u_{th}(\mathbf{r}_{hkn})]}{\sum_{k'=1}^{K_h} N_{hk'} \exp[f_{hk'} - w_{hk'}(\mathbf{r}_{hkn})]} \\ &= \omega_h(\mathbf{r}_{hkn}) \exp[-\Delta u_{th}(\mathbf{r}_{hkn})] \exp(f_t - f_h) \end{aligned} \quad (12)$$

The second line of eq 12 suggests that one can interpret the wTP method as effectively performing “exponential averaging” between the reference and target potentials in each histogram bin upon reweighting the biased simulations. The  $\exp(f_t - f_h)$  factor is a ratio of partition functions that merely shifts the FES by a constant.

**Generalized Weighted Thermodynamic Perturbation Method.** The gwTP method extends the wTP approach by considering situations where umbrella sampling has been performed with  $N_{PE}$  reference potentials. We place no restrictions on the relationship between the sets of biasing potentials used to enhance the sampling of each reference potential, that is,  $K_h$  does not need to be the same as any other  $K_{h'}$ , and  $w_{hk}(\mathbf{r})$  does not need to be the same as any other  $w_{h'k'}(\mathbf{r})$ . It is for this precise reason why the biasing potentials are written with both  $h$  and  $k$  subscripts. With this freedom, there are a total of  $N_{sim} = \sum_{h=1}^{N_{PE}} K_h$  biased states, and expressions which sum all biased states must now be written either as a double summation  $\sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h}$  or by viewing  $hk$  as a combined index of biased simulations  $\sum_{(hk)=1}^{N_{sim}}$ . The MBAR/UWHAM equations for the solution of the  $N_{sim}$  biased free energies are shown in eq 13.

$$f_{il} = -\ln \sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \frac{\exp[-u_{il}(\mathbf{r}_{hkn})]}{\sum_{h'=1}^{N_{PE}} \sum_{k'=1}^{K_{h'}} N_{h'k'} \exp[f_{h'k'} - u_{h'k'}(\mathbf{r}_{hkn})]}, \quad (13)$$

$\forall (il) = 1, \dots, N_{sim}$

The target potential free energy (eq 14) and the target FES (eq 15) are similarly rewritten to account for the additional umbrella sampling.

$$f_t = -\ln \sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \frac{\exp[-u_t(\mathbf{r}_{hkn})]}{\sum_{h'=1}^{N_{PE}} \sum_{k'=1}^{K_{h'}} N_{h'k'} \exp[f_{h'k'} - u_{h'k'}(\mathbf{r}_{hkn})]} \quad (14)$$

$$F_t(\xi_m) = -\beta^{-1} \ln \sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_t(\mathbf{r}_{hkn}) \quad (15)$$

All that remains is to generalize the wTP expression for the target potential sample weights (eq 12). As previously stated, the wTP method can be interpreted as a weighted exponential averaging procedure within each histogram bin, so it may be perplexing how this could be adapted to treat multiple reference potentials because exponential averaging has traditionally been applied to the calculation of free energy differences between pairs of potentials. Our approach is to reinterpret the biased potentials as using a single reference (the “selected reference”) perturbed by an effective bias that accounts for the difference in unbiased energies. This is more clearly described by eq 16 which re-expresses the biased PE  $u_{h'k'}(\mathbf{r})$  in terms of the selected reference  $u_h(\mathbf{r})$  and the effective bias  $w_{k'}(\mathbf{r}) + \Delta u_{h'h}(\mathbf{r})$ , where  $\Delta u_{h'h}(\mathbf{r}) = u_{h'}(\mathbf{r}) - u_h(\mathbf{r})$ .

$$\begin{aligned} u_{h'k'}(\mathbf{r}_{hkn}) &= u_h(\mathbf{r}_{hkn}) + w_{k'}(\mathbf{r}_{hkn}) \\ &= u_h(\mathbf{r}_{hkn}) + [w_{k'}(\mathbf{r}_{hkn}) + \Delta u_{h'h}(\mathbf{r}_{hkn})] \end{aligned} \quad (16)$$

The decision of which potential to select as the reference is arbitrary; it has no effect on the predicted FES. Our convention is to select the potential which produced the sample. That is to say, the sample  $\mathbf{r}_{hkn}$  is a member of the ensemble generated from  $u_{hk}(\mathbf{r})$ , whose reference potential is  $u_h(\mathbf{r})$ ; thus, we choose  $u_h(\mathbf{r})$  to be the selected reference. The gwTP expression for the sample weights (eq 17) differs from eq 12 only by considering the additional simulated states and by replacing the biasing potential with the effective bias.

$$\begin{aligned} \omega_t(\mathbf{r}_{hkn}) &= \frac{\exp[f_t - \Delta u_{th}(\mathbf{r}_{hkn})]}{\sum_{h'=1}^{N_{PE}} \sum_{k'=1}^{K_{h'}} N_{h'k'} \exp[f_{h'k'} - w_{h'k'}(\mathbf{r}_{hkn}) - \Delta u_{h'h}(\mathbf{r}_{hkn})]} \end{aligned} \quad (17)$$

Multiplying the numerator and denominator by  $\exp[-u_h(\mathbf{r}_{hkn})]$  yields an expression that illustrates more clearly that the weights are independent of the selected reference.

$$\omega_t(\mathbf{r}_{hkn}) = \frac{\exp[f_t - u_t(\mathbf{r}_{hkn})]}{\sum_{h'=1}^{N_{PE}} \sum_{k'=1}^{K_{h'}} N_{h'k'} \exp[f_{h'k'} - u_{h'k'}(\mathbf{r}_{hkn})]} \quad (18)$$

The weights shown in eqs 6, 10, and 18 are very similar because they are all specialized forms of the general expressions for the MBAR weights developed by Shirts and Chodera.<sup>43</sup> In their work, an ensemble average of a target thermodynamic state is expressed using the samples obtained from a collection of explicitly simulated states (which may or may not include the target state). To apply the MBAR

approach in a novel way, one must define the target and sampled potentials in the context of the new application and describe how the weights are used. In the context of umbrella sampling, the target state corresponds to an unbiased potential, the sampled states are the biased simulations, and the weights are used to approximate the spatial distribution of the density (eq 4). The methods described in the present work differ only in their choice of target and sampled potentials.

The RE is a useful index to gauge the reliability of a calculated FES.<sup>64</sup> It can be interpreted as being a measure of “flatness” in the sample weights, such that it is 1 when the distribution of weights in a bin is uniform, and it approaches 0 when the sum of weights is dominated by only a few samples. The gwTP expression for the RE is given by eqs 19 and 20.

$$S_t(\xi_m) = - \frac{\sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_t(\mathbf{r}_{hkn}) \ln \frac{\omega_t(\mathbf{r}_{hkn})}{s_{tm}}}{\ln \sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn}))} \quad (19)$$

$$s_{tm} = \sum_{h=1}^{N_{PE}} \sum_{k=1}^{K_h} \sum_{n=1}^{N_{hk}} \delta(\xi_m - \xi(\mathbf{r}_{hkn})) \omega_t(\mathbf{r}_{hkn}) \quad (20)$$

## COMPUTATIONAL MODELS AND METHODS

All simulations described below were performed with a development version of SANDER.<sup>89</sup> The simulations were propagated with a 1 fs time step. All production sampling was performed in the canonical ensemble using the Langevin thermostat with a 5 ps<sup>-1</sup> collision frequency to maintain a temperature of 298 K.<sup>90</sup> The equilibration of the system densities was performed in the isothermal–isobaric ensemble using the Berendsen barostat to maintain an external pressure of 1 atm.<sup>91</sup> The long-range electrostatics were calculated with a particle mesh Ewald (PME) method using a 1 Å<sup>3</sup> regular grid spacing, 8 Å real-space cutoffs, and tinfoil boundary conditions.<sup>92</sup> Specifically, the semiempirical QM/MM simulations performed PME with Mulliken charges,<sup>93,94</sup> whereas the ab initio QM/MM calculations used the ambient-potential composite Ewald method.<sup>87</sup> The ab initio QM/MM simulations were performed with the HFDF software package developed within our group, which we interfaced to SANDER and described in ref 87. The Lennard–Jones potential was truncated at 8 Å, and a tail correction was applied to model the interactions beyond the cutoff.<sup>95</sup> The solute structures were initially prepared with the GAFF force field,<sup>96</sup> and the solvent was modeled with the TIP4P-Ew water model.<sup>97</sup> The MNDO/d QM/MM + DPRc and DFTB2/MIO<sup>98–100</sup> QM/MM + DPRc potentials include a nonelectrostatic MLP correction to the underlying QM/MM energy and forces. The DPRc potential is an extension of the DeepPot-SE model<sup>101</sup> that includes corrections for both the QM–QM and QM–MM interactions, and the neural network parameters are optimized to reproduce the target QM/MM energies and forces. The DPRc model has been previously described,<sup>76,88</sup> and additional details are provided in the Supporting Information.

**Simulations of the Ethylene Phosphate Reaction.** The ethylene phosphate (Figure 1a) FES was calculated from umbrella sampling performed at 48 values of  $\xi_{PT}$  ranging from –3.1 to 1.6 Å in steps of 0.1 Å that bias the PE with a spring force constant of 200 kcal mol<sup>-1</sup> Å<sup>-2</sup>. An initial structure for

the unimolecular reactant state was prepared by solvating the system in a truncated octahedron containing 2974 TIP4P-Ew water molecules. A 100 ps MNDO/d QM/MM simulation was performed in the NPT ensemble to equilibrate the system density, and the final real-space lattice vector lengths were 49.03 Å. A series of 3 ps NVT simulations were conducted to slowly progress the structure along the reaction coordinate (i.e., the brief simulation of window  $i + 1$  was restarted from simulation  $i$ ) to generate initial structures for each window. Each window was then equilibrated with MNDO/d QM/MM for 100 ps in the NVT ensemble. This was followed by an additional 20 ps of NVT equilibration with PBE0/6-31G\* QM/MM. Production sampling was performed with PBE0/6-31G\* QM/MM to obtain a target FES used to validate the wTP and gwTP reference potential methods. The PBE0/6-31G\* QM/MM production simulations were performed for 80 ps/(window-trial) and repeated four times to analyze the uncertainty in the FES values. The four trials differed only by changing the thermostat random number seed value. The aggregate amount of PBE0/6-31G\* sampling corresponds to 320 ps/window or 15.36 ns of sampling for the entire FES. The coordinates were saved every 10 fs for analysis.

The parametrization of MLPs often involves the use of an active learning procedure that results in several network parameter sets that produce similar corrections to the energies and forces for the data they were trained against. If the training data adequately represent the ensemble of structures observed in production sampling, then the trial network parameter sets can lead to similar FES values. For illustrative purposes, we have chosen to construct four ad hoc MNDO/d QM/MM + DPRc potentials that yield disparate FES values. These potentials were trained to reproduce different regions of the ethylene phosphate FES rather than the entire surface. Each potential well-reproduces the ab initio FES in the region where it was parametrized, but the FES values disagree with each other and the ab initio target elsewhere. We will show that the gwTP analysis from the combined sampling of all four potentials reproduces the ab initio FES over the full range. The MNDO/d QM/MM + DPRc potentials will be referred to as ML0, ML1, ML2, and ML3. The potentials were trained to reproduce different sets of target PBE0/6-31G\* energies and forces. Specifically, the ML0 potential was trained only using the sampling obtained from  $-0.1 \text{ \AA} \leq \xi_{PT} \leq 1.6 \text{ \AA}$ , the ML1 potential was trained to  $-1.3 \text{ \AA} \leq \xi_{PT} \leq 1.0 \text{ \AA}$ , the ML2 potential was trained to  $-2.5 \text{ \AA} \leq \xi_{PT} \leq -0.2 \text{ \AA}$ , and the ML3 potential was trained to  $-3.1 \text{ \AA} \leq \xi_{PT} \leq -1.4 \text{ \AA}$ . In the present work, we are making comparison with an FES obtained from explicit ab initio sampling; therefore, we have reused the sampling to serve as initial training data for each potential. Specifically, 2.5% of the PBE0/6-31G\* trajectory frames in the desired range were chosen at random to parametrize each potential. The optimization of the network parameters was performed with the DeePMD-kit and DP-GEN software.<sup>102,103</sup> The optimization consisted of 200k steps with initial and final learning rates of  $10^{-3}$  and  $5 \times 10^{-8}$ , respectively. The initial optimization was followed by nine cycles of active learning to search for additional training data.<sup>76</sup> Each active learning cycle performs four parameter optimizations to yield four trial parameter sets. One of the parameter sets is used to generate 20 ps of MNDO/d QM/MM + DPRc sampling for each window in the selected range of  $\xi_{PT}$  values. The atomic forces are then computed with each of the four trial parameter sets, and if the maximum standard deviation of atomic forces is

within the range 0.08 to 0.25 eV/Å, then the candidate structure is included in the next round of optimization. Upon completion of the active learning procedure, the “DP Compress” compression algorithm was applied to the MLP to improve computational performance during inference.<sup>104</sup> The production sampling of each potential, described below, was performed using only one of the four parameter sets. In hindsight, the initial parameterization to the ab initio sampling would have been sufficient for our purpose; each round of active learning only labeled six candidate structures, on average.

The MNDO/d QM/MM + DPRc production sampling of the ethylene phosphate system was performed for 250 ps/(window-potential-trial), and the coordinates were saved to a trajectory file every 50 fs. Three trials were performed with different random number seed values for error analysis. From these simulations, one can use MBAR or wTP to evaluate 12 surfaces corresponding to the 3 trials of the 4 MNDO/d QM/MM + DPRc potentials. One can further obtain four trial-averaged surfaces. Each of the 12 surfaces is analyzed from 12 ns/(potential-trial) of sampling, and each of the 4 trial-averaged surfaces are produced from 36 ns/potential of sampling. We use the notation  $\Delta A(\text{target}; \text{reference})$  to distinguish between the surfaces. This notation signifies that the curve is the FES of the target PE function estimated from the sampling performed with reference. When the target and reference are the same PE function, then the FES is calculated from the MBAR method. If the target FES is calculated from the sampling performed with a single reference other than the target, then the wTP method is used. If the target FES is estimated from multiple reference potentials, then the gwTP method is used. The trial-averaged estimates of the ML0, ML1, ML2, and ML3 potentials calculated from the MBAR method shall be labeled  $\Delta A(\text{ML}_i; \text{ML}_i)$  (where  $i \in [0, 3]$ ). The trial-averaged estimates of the PBE0/6-31G\* FES obtained from wTP analysis of the individual potentials are denoted as  $\Delta A(\text{PBE0}; \text{ML}_i)$ . The gwTP method analyzes the combined sampling from the four MNDO/d QM/MM + DPRc potentials. Therefore, one obtains three gwTP surfaces corresponding to the three trials; each surface consists of 48 ns/trial of sampling, and the trial-averaged gwTP FES includes 144 ns of aggregate sampling. The trial-averaged gwTP FES is referred to as  $\Delta A(\text{PBE0}; \text{ML}^*)$ . One can further make a “best estimate” of the ab initio FES using gwTP to analyze the 144 ns of MNDO/d QM/MM + DPRc sampling and the 15.36 ns of PBE0/6-31G\* sampling. The best estimate is denoted as  $\Delta A(\text{PBE0}; \text{All})$ .

**Simulations of the Native Model Reaction.** The native model reaction (Figure 1b) was previously examined in ref 88. It was one of the six nonenzymatic transesterification model reactions used to develop four DFTB2/MIO QM/MM + DPRc potentials trained to reproduce PBE0/6-31G\* QM/MM energies and forces. The four potentials were generated from an active learning procedure described in that work.<sup>88</sup> In the present work, we reuse the DFTB2/MIO QM/MM + DPRc potentials without modification. The five analogous systems included within the training differ from the native model by having replaced one or more oxygens with sulfurs at key positions. We do not reconsider the sulfur-substituted systems in the present work; instead, our interest in the native model reaction arises from noticeable discrepancies between the ab initio and DFTB2/MIO QM/MM + DPRc FES values. We will show that the native model reaction profiles presented in

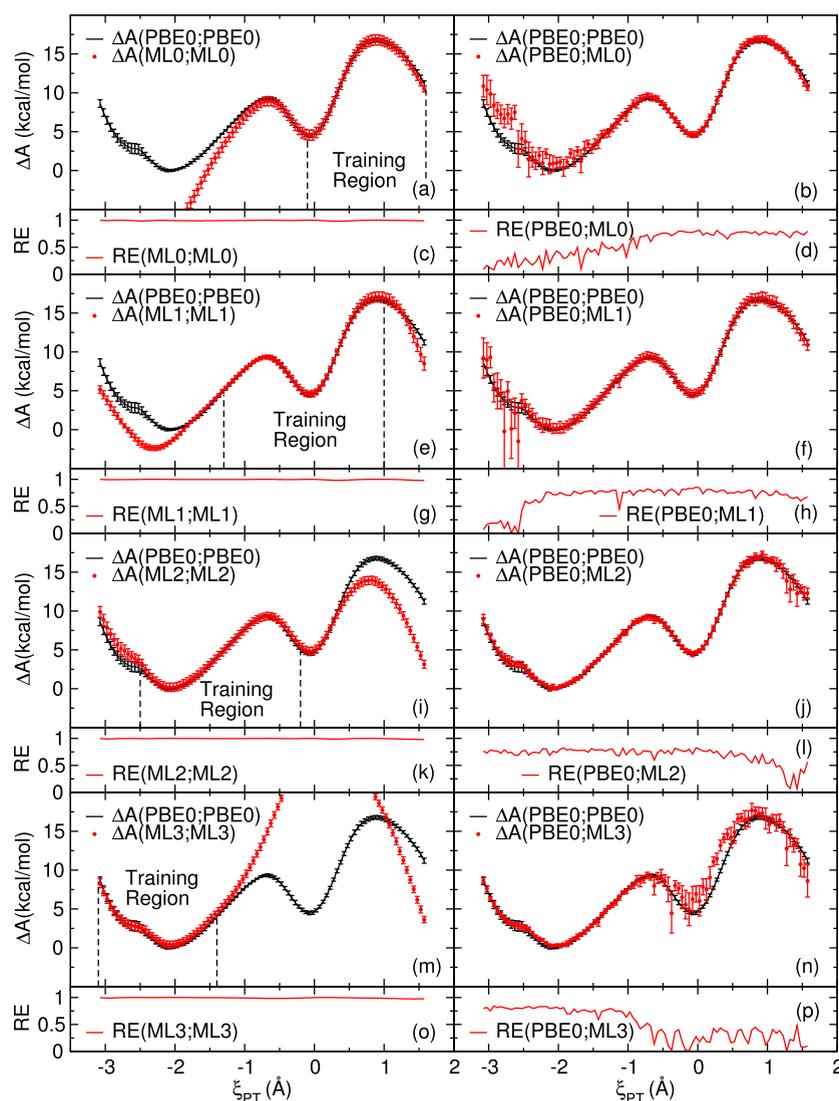
ref 88 are not fully converged. The high computational cost of ab initio QM/MM calculations, however, places practical limitations on the amount of sampling that can be reasonably achieved. We will demonstrate that gwTP analysis of extended sampling performed with the reference potentials produces a better estimate of the ab initio FES than what can be afforded from ab initio QM/MM simulation.

An initial structure for the native model unimolecular reactant state was prepared by solvating the system in a box of 1510 TIP4P-Ew waters and performing 200 ps of DFTB2/MIO QM/MM simulation in the NPT ensemble to equilibrate the system density. The final unit cell lattice vectors were approximately 35.8 Å. This was followed by a series of 2 ps DFTB2/MIO QM/MM simulations in the NVT ensemble that slowly incremented the value of the reaction coordinate to obtain initial structures for each window in the range  $-3.5 \text{ \AA} \leq \xi_{\text{PT}} \leq 5 \text{ \AA}$  in steps of 0.1 Å. The umbrella potentials biased the simulations with a spring force constant of  $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Each of the 86 windows were equilibrated with DFTB2/MIO QM/MM for 100 ps in the NVT ensemble, and this was followed by an additional 25 ps of PBE0/6-31G\* QM/MM NVT equilibration. The 25 ps/window of PBE0/6-31G\* equilibration is not included in the analysis presented in this work nor ref 88.

PBE0/6-31G\* production sampling of the native model reaction was run for 25 ps/(window-trial), and the coordinates were saved every 25 fs. Each simulation was repeated four times using different thermostat random number seed values to yield four FES estimates. Each surface is the analysis of 2.15 ns/trial of sampling. The trial-averaged FES, denoted as “ $\Delta A(\text{PBE0}; \text{PBE0}) 100 \text{ ps}$ ”, includes all 8.6 ns of production sampling (100 ps/window of aggregate sampling).

DFTB2/MIO QM/MM + DPRc production sampling of the native model reaction proceeded similarly. However, the active learning procedure yields four neural network parameter sets, so we performed 25 ps/(window-potential) of sampling with each potential. Therefore, the average DFTB2/MIO QM/MM + DPRc FES does not correspond to a single potential; instead, it is the mean of the four reference potentials. The average DFTB2/MIO QM/MM + DPRc FES includes 100 ps/window of aggregate sampling, and it is denoted as “ $\langle \Delta A(\text{ML}; \text{ML}) \rangle 100 \text{ ps}$ ”. The gwTP method can be used to estimate the PBE0/6-31G\* from the combined sampling of the DFTB2/MIO QM/MM + DPRc potentials, referred to as “ $\Delta A(\text{PBE0}; \text{ML}^*) 100 \text{ ps}$ ”.

To examine the behavior of the predicted native model reaction FES on a longer timescale, we extended the DFTB2/MIO QM/MM + DPRc simulations and performed multiple trials with each potential. The extended simulations were sampled for 100 ps/(window-potential-trial), and we performed three trials of each simulation with different random number seed values. The 3 trials of the 4 potentials yield 12 surfaces. Each surface is the analysis of 8.6 ns/(potential-trial) of sampling. The average FES is the mean of the 12 estimates, such that it considers 103.2 ns (1.2 ns/window) of aggregate sampling. The average DFTB2/MIO QM/MM + DPRc FES from the extended sampling shall be labeled “ $\langle \Delta A(\text{ML}; \text{ML}) \rangle 1.2 \text{ ns}$ ”. The sampling from the three trials produces three gwTP estimates of the ab initio surface. The PBE0/6-31G\* FES from the extended sampling is labeled “ $\Delta A(\text{PBE0}; \text{ML}^*) 1.2 \text{ ns}$ ”. An analogous set of profiles labeled “ $\langle \Delta A(\text{ML}; \text{ML}) \rangle$  last 600 ps” and “ $\Delta A(\text{PBE0}; \text{ML}^*)$  last 600 ps” were analyzed from the last 50 ps/(window-potential-trial) of sampling from



**Figure 2.** FESs ( $\Delta A$ ) of the ethylene phosphate model reaction (Figure 1a) calculated with MBAR and wTP analysis. REs are shown in the subplots below each FES. Parts (a,e,i,m) are the four MNDO/d QM/MM + DPRc potentials (ML0, ML1, ML2, and ML3) evaluated from MBAR analysis. Parts (b,f,j,n) are their corresponding wTP estimates of the PBE0/6-31G\* FES. Parts (c,d,g,h,k,l,o,p) are the REs of the corresponding MBAR and wTP analysis. The error bars are 95% confidence intervals.

each simulation. Finally, we shall refer to a profile labeled “ $\Delta A(\text{PBE0}; \text{All})$ ”. This FES includes all 103.2 ns of DFTB2/MIO QM/MM + DPRc sampling and the 8.6 ns of PBE0/6-31G\* sampling.

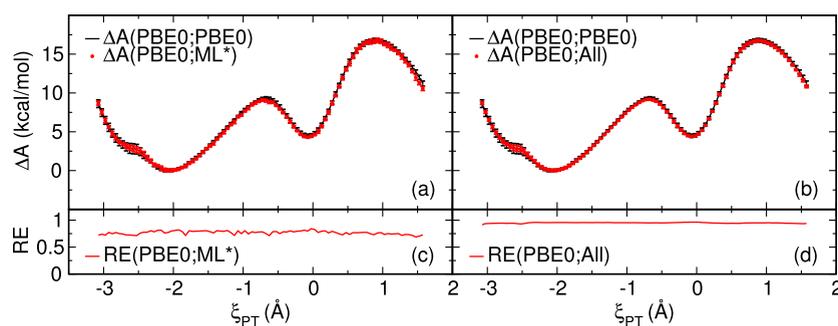
**Error Analysis.** The uncertainties in the free energy values can be estimated from the MBAR calculation of the sample weight covariance<sup>43</sup> between spatial bins, as was carried out in ref 52. This approach uses the fluctuations of the statistically independent samples within the observed ensembles to assign error values. Another strategy for estimating errors is the “ensemble average approach” which estimates the standard errors from the distribution of calculated results produced from independent simulations.<sup>105–108</sup> The ensemble average approach can be viewed as a strategy to gauge the uncertainty arising from limited finite sampling. In the present work, we incorporate the ideas that motivate both strategies by analyzing multiple trials with bootstrap analysis. Let  $F_t^{(i)}(\xi_m)$  be the free energy of target potential  $t$  in spatial histogram  $m$  from trial  $i$  ( $i \in [1, N_{\text{trial}}]$ ). The uncertainty of the FES value  $\delta F_t^{(i)}(\xi_m)$  is

estimated from circular moving block bootstrap error analysis,<sup>109</sup> where the block size of each simulation is chosen from the autocorrelation of the biasing potential time series  $w_{hk}(\mathbf{r}_{hkn})$ . The aggregate sampling from all trials can similarly be analyzed to yield free energy values  $F_t^{(a)}(\xi_m)$  and uncertainties  $\delta F_t^{(a)}(\xi_m)$ . To compute a trial-average  $\bar{F}_t(\xi_m)$ , we choose a set of constants  $C_i$  that uniformly shift each surface  $F_t^{(i)}(\xi_m) + C_i$  to minimize the weighted sum of squared differences with the  $F_t^{(a)}(\xi_m)$  values.

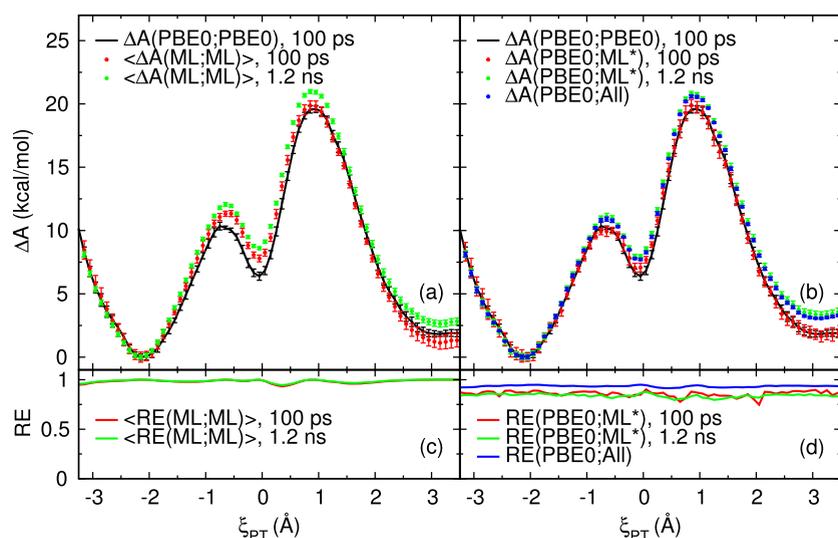
$$\frac{\partial}{\partial C_i} \sum_m w_m^{(i)} \{ [F_t^{(i)}(\xi_m) + C_i] - F_t^{(a)}(\xi_m) \}^2 = 0 \quad (21)$$

The weights are chosen to be inversely proportional to the bootstrap uncertainties.

$$w_m^{(i)} = \frac{\{ [\delta F_t^{(i)}(\xi_m)]^2 + [F_t^{(a)}(\xi_m)]^2 \}^{-1}}{\sum_m \{ [\delta F_t^{(i)}(\xi_m)]^2 + [F_t^{(a)}(\xi_m)]^2 \}^{-1}} \quad (22)$$



**Figure 3.** FESs ( $\Delta A$ ) of the ethylene phosphate model reaction (Figure 1) calculated with gwTP analysis. REs are shown in the subplots below each FES. (a) gwTP FES estimated from the combined umbrella sampling produced from the four MNDO/d QM/MM + DPRc potentials (ML0, ML1, ML2, and ML3). (b) PBE0/6-31G\* surface estimated from gwTP analysis of all available PBE0/6-31G\* and MNDO/d QM/MM + DPRc umbrella sampling. Parts (c,d) are the REs of the corresponding MBAR and gwTP analysis. The error bars are 95% confidence intervals.



**Figure 4.** FESs ( $\Delta A$ ) of the native model reaction (Figure 1). (a) Comparison of the PBE0/6-31G\* FES to the average FES of the four DFTB2/MIO QM/MM + DPRc potentials (ML0, ML1, ML2, and ML3). The listed times denote the total amount of sampling per umbrella window used in the FES calculation. (b) Comparison of the PBE0/6-31G\* FES to those predicted from gwTP analysis. The  $\Delta A(\text{PBE0}; \text{ML}^*)$  gwTP surfaces are evaluated using the sampling from all four DPRc potentials. The  $\Delta A(\text{PBE0}; \text{All})$  surface is calculated from the 1.2 ns/window of DPRc sampling and the 100 ps/window PBE0/6-31G\* sampling. The error bars are 95% confidence intervals. The RE values shown in parts (c,d) are the REs.

The average value  $\bar{F}_i(\xi_m)$  and its standard error  $\delta\bar{F}_i(\xi_m)$  are then given by eqs 23 and 24, respectively.

$$\bar{F}_i(\xi_m) = \frac{1}{N_{\text{trial}}} \sum_{i=1}^{N_{\text{trial}}} F_t^{(i)}(\xi_m) + C_i \quad (23)$$

$$\delta\bar{F}_i(\xi_m) = \left( \frac{1}{N_{\text{trial}} - 1} \sum_{i=1}^{N_{\text{trial}}} [F_t^{(i)}(\xi_m) + C_i - \bar{F}_i(\xi_m)]^2 + \frac{[\delta F_t^{(i)}(\xi_m)]^2}{N_{\text{trial}}} \right)^{1/2} \quad (24)$$

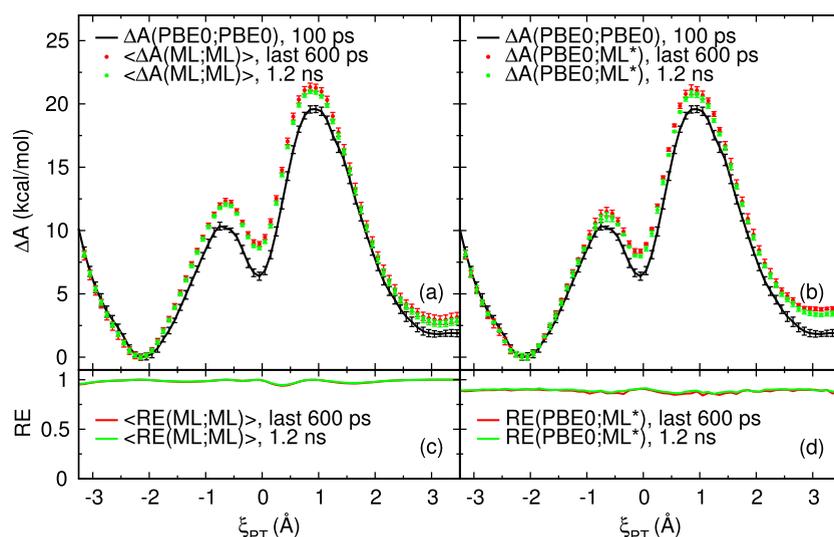
The first and second terms in eq 24 correspond to the unbiased sample variance between trials and the average bootstrap variance, respectively. The error bars shown in the figures are 95% confidence intervals ( $1.96 \delta\bar{F}_i(\xi_m)$ ).

## RESULTS AND DISCUSSION

**Ethylene Phosphate Reaction Profiles.** Figure 2 displays free energy profiles of the ethylene phosphate reaction calculated with the MBAR and wTP methods using each of the four MNDO/d QM/MM + DPRc reference potentials

(ML0, ML1, ML2, and ML3). As previously mentioned, the reference potentials were specifically designed to disagree with each other and the PBE0/6-31G\* target in various regions of the FES. The surfaces estimated with the gwTP method are shown in Figure 3. These figures demonstrate that the gwTP method can improve upon the wTP estimate by combining the sampling from multiple reference potentials.

Figure 2 parts (a, e, i, and m) show that the four MNDO/d QM/MM + DPRc potentials agree with PBE0/6-31G\* in the regions they were trained to reproduce the target energies and forces. Outside of these regions, the profiles disagree with ab initio and each other. Figure 2 parts (b, f, j, and n) are the wTP estimates of the target FES from each reference potential. The wTP surfaces generally improve the agreement with the target; however, they suffer from numerical noise and large uncertainties in the untrained regions. This is further emphasized by a corresponding decrease in the REs shown in parts (d, h, l, and p). The results shown in Figure 2 are consistent with previous work which found that reliable free energies are produced by the wTP method when the RE is larger than 0.6, but the results become questionable when the RE decreases below 0.3.<sup>50</sup> Every MNDO/d QM/MM + DPRc



**Figure 5.** FESs ( $\Delta A$ ) of the native model reaction (Figure 1). (a) Comparison of the PBE0/6-31G\* FES to the average FES of the four DFTB2/MIO QM/MM + DPRc potentials (ML0, ML1, ML2, and ML3). (b) Comparison of the PBE0/6-31G\* FES to those predicted from gwTP analysis. The surfaces labeled “last 600 ps” refer to aggregate sampling taken from the last 50 ps of the 12 DFTB2/MIO QM/MM + DPRc simulations. The error bars are 95% confidence intervals. The RE values shown in parts (c,d) are the REs.

potential shown in Figure 2 exhibits REs below 0.3 for some  $\xi_{PT}$  values.

The gwTP estimate of the ab initio FES is shown in Figure 3. Figure 3a uses the combined sampling from each MNDO/d QM/MM + DPRc potential, and the resulting FES agrees with the MBAR analysis of the ab initio sampling to within the uncertainties of the calculations throughout the entire range of  $\xi_{PT}$  values. Similarly, the REs shown in Figure 3c consistently remain above 0.6. The gwTP FES shown in Figure 3b includes sampling from both the ab initio and MNDO/d QM/MM + DPRc potentials and represents our best estimate of the surface. The inclusion of sampling from the target potential increases the RE values, but the FES values do not significantly change.

An approach for smoothing the density-of-states (DoS) in wTP calculations is presented in ref 52. The motivation for DoS smoothing is to effectively remove statistically unlikely samples from a distribution that are over-represented within the limited amount of finite sampling. We generalized the DoS smoothing procedure in the Supporting Information. Figures S1 and S2 are analogous to Figures 2 and 3 but include DoS smoothing. The use of DoS smoothing does not significantly improve the wTP estimates of the ab initio surface because the failure is produced from the large differences in reference and target potentials, not from the presence of a few, unlikely samples. The gwTP formalism does not introduce any mechanism to prevent the occurrence of over-represented samples. Although the DoS smoothing procedure is unnecessary in the present applications, we continue to regard it as a valuable technique which can help eliminate numerical noise from the predicted free energies. Application of DoS smoothing to the MBAR and gwTP analysis yields profiles that are indistinguishable from those shown in Figures 2 and 3.

**Native Model Reaction Profiles.** FESs of the native model reaction (Figure 1b) are shown in Figure 4 and 5. These figures illustrate that the extended sampling made possible using an affordable reference potential can result in a more accurate FES than what can be estimated from limited sampling performed with an expensive target potential.

The PBE0/6-31G\* and average DFTB2/MIO QM/MM + DPRc surfaces (the black and red surfaces, respectively) shown in Figure 4a are originally presented in ref 88. These surfaces include 100 ps/window of aggregate production sampling. Additional ab initio sampling was deemed too costly to extend, and the DFTB2/MIO QM/MM + DPRc sampling was chosen to mimic the protocol used to perform the ab initio simulations. The ab initio and reference potentials generally agree except in the range  $\xi_{PT} \approx [-1, 0.5 \text{ \AA}]$ , where the reference potential FES is noticeably higher in energy.

We used the gwTP method to estimate the ab initio FES from the 100 ps/window of DFTB2/MIO QM/MM + DPRc sampling, and the result is labeled “ $\Delta A(\text{PBE0}; \text{ML}^*)$ , 100 ps” in Figure 4b. This gwTP estimate agrees with the MBAR analysis of the ab initio sampling to within the uncertainties of the calculations, and the RE is larger than 0.6. This suggests that there is a small discrepancy in the reference potentials’ ability to reproduce the ab initio QM/MM energies and forces, which is corrected by reweighting the samples. However, if the reference potential simulations are extended to include 1.2 ns/window of aggregate sampling (the green surfaces shown in Figure 4a,b), the predicted activation energy increases by 1 kcal/mol, even though the REs remain largely unchanged. The large REs suggest that there is good phase space overlap between the reference and target potentials in both the 100 ps/window and 1.2 ns/window simulations. Similarly, the gwTP ab initio FES values are nearly 1 kcal/mol higher near the transition state when analyzing the combined sampling from both the PBE0/6-31G\* and extended DFTB2/MIO QM/MM + DPRc simulations (the blue surface in Figure 4b). These results suggest that another reason for the observed discrepancies between PBE0/6-31G\* and the DFTB2/MIO QM/MM + DPRc potentials may due to the limited amount of sampling. In other words, the ab initio simulations may not represent converged equilibrium ensembles. Consequently, the  $\Delta A(\text{PBE0}; \text{PBE0})$  MBAR estimate should not be trusted, and the  $\Delta A(\text{PBE0}; \text{All})$  gwTP result does not represent a “best estimate” of the FES because it is potentially skewed by nonequilibrium sampling.

**Table 1.** Native Model Reaction Free Energies (kcal/mol) of the First Transition State  $\Delta A(\xi_{\text{TS1}})$ , the Intermediate  $\Delta A(\xi_{\text{int}})$ , the Rate-Limiting Transition State  $\Delta A^\ddagger$ , and the Reaction Product  $\Delta A^a$ 

target	ref	sampling	$\Delta A(\xi_{\text{TS1}})$	$\Delta A(\xi_{\text{int}})$	$\Delta A^\ddagger$	$\Delta A$
PBE0	PBE0	100 ps	10.4 ± 0.2	6.4 ± 0.2	19.6 ± 0.2	1.5 ± 0.3
PBE0	ML*	100 ps	10.0 ± 0.2	7.0 ± 0.2	19.9 ± 0.3	1.2 ± 0.3
PBE0	ML*	last 600 ps	11.5 ± 0.2	8.3 ± 0.1	21.1 ± 0.2	3.6 ± 0.1

<sup>a</sup>The uncertainties are standard errors.

We performed forward and reverse analysis of the sampling<sup>110,111</sup> to examine the convergence of the extended simulations. The forward analysis generates a  $\langle \Delta A(\text{ML}; \text{ML}) \rangle$  FES from the first  $P$  % of samples, whereas reverse analysis calculates an FES from the last  $P$  % of the simulations. From these surfaces, we monitor the free energy values (and uncertainties) of the two transition states and intermediate relative to the unimolecular reactant state as a function of  $P$  %. The convergence behavior of these properties is shown in Figure S3 in the Supporting Information. In summary, the reverse analysis is stable with respect to  $P$  %, but the forward analysis display drifts in the free energy values. The forward and reverse analyses do not agree to be within their estimated uncertainties until the first 50% of the simulations are discarded. It is difficult to ascertain if the simulations have converged without having extended the sampling (see Figure S4 in the Supporting Information); when the data contain a slight drift, the reverse analysis artificially agrees with the forward analysis simply because it has not been given ample opportunity to approach an equilibrium value. The assertion that the ab initio simulations are not converged is further supported by umbrella integration analysis<sup>47–49</sup> (see Figure S5 in the Supporting Information). The extended simulations yield a smooth free energy gradient profile, whereas the gradient profiles obtained from the limited sampling are contaminated with noise. Figure 5 displays the  $\langle \Delta A(\text{ML}; \text{ML}) \rangle$  and  $\Delta A(\text{PBE0}; \text{ML}^*)$  surfaces after discarding the first half of each simulation. These surfaces represent our best estimate of the profiles. It is notable that the best estimate of the ab initio FES does not include any of the available PBE0/6-31G\* QM/MM sampling, and the best estimate of the rate-limiting transition state barrier is 1.5 kcal/mol higher in energy than the MBAR analysis of the PBE0/6-31G\* QM/MM simulations (see Table 1).

To reduce the number of target potential evaluations, one should identify correlation within the simulations and only consider the statistically independent samples. The “ $\langle \Delta A(\text{ML}; \text{ML}) \rangle$ , last 600 ps” and “ $\Delta A(\text{PBE0}; \text{ML}^*)$ , last 600 ps” surfaces displayed in Figure 5 were analyzed from the correlated samples, and the errors were estimated from block bootstrap analysis. Figure S6 in the Supporting Information compares these surfaces to the analysis performed with uncorrelated samples from the last half of each simulation. The uncorrelated data are the subset of samples extracted with a stride equal to the statistical inefficiency of the biasing energy time series. In summary, analysis of the uncorrelated samples produces surfaces that are visually indistinguishable from the results shown in Figure 5.

The obvious benefit of using gwTP (or wTP) is the reduced computational effort associated with performing the sampling with an affordable reference potential. The total cost of performing MBAR and gwTP estimates of the target FES is given by eqs 25 and 26, respectively.

$$C(t; t) = \frac{T}{\nu_t} \quad (25)$$

$$C(t; h) = \frac{T}{\nu_h} + (1 - f_{\text{eq}}) f_{\text{file}} f_{\text{indep}} f_{\text{eval}} \frac{T}{\nu_t} \quad (26)$$

where  $T$  is the aggregate simulation time (ps) of all windows, trials, and potentials (if applicable).  $\nu_t$  and  $\nu_h$  are the simulation rates (ps/day) using the target and reference potentials. Equation 25 is, therefore, the net cost of performing the sampling with the target potential, and the first term in eq 26 is the cost of performing the reference simulations. The second term in eq 26 is the additional cost of performing target potential evaluations of the saved samples.  $f_{\text{eq}}$  and  $f_{\text{file}}$  are the fraction of simulation steps excluded as equilibration and the fraction of samples written to file for analysis, respectively.  $f_{\text{indep}}$  is the fraction of statistically independent samples, and  $f_{\text{eval}}$  is the fraction of the statistically independent samples used to perform target potential evaluations. The ratio  $C(t; t)/C(t; h)$  denotes the computational savings from using the reference potential approach. If one assumes that the methods are performed with equal sampling, then the ratio is given by eq 27.

$$\frac{C(t; t)}{C(t; h)} = \left( \frac{\nu_t}{\nu_h} + (1 - f_{\text{eq}}) f_{\text{file}} f_{\text{indep}} f_{\text{eval}} \right)^{-1} \quad (27)$$

In practice, the goal is to use the information provided by the reference simulations to reduce the number of target potential evaluations, thus maximizing the savings. First, one should estimate  $f_{\text{eq}}$  by verifying that the reference simulations are converged to avoid unnecessary evaluations of unequilibrated data. In the present work, we chose  $f_{\text{eq}}$  by monitoring the energies of relevant stationary points from forward and reverse analysis of the time series.<sup>110,111</sup> Second, one should estimate the correlation in the time series, so only the statistically independent samples are evaluated with the target potential. In principle, one could estimate the correlation from the energy differences  $\Delta u_{t,h}$ , but this approach would defeat the purpose of reducing the number of target potential evaluations. Instead, we have chosen to estimate the correlation from the biasing potential  $w_{hk}$  time series. Finally, the value of  $f_{\text{file}}$  should generally be chosen such that the statistical inefficiency of the saved frames is close to 1 to avoid storing unnecessarily large files.

To estimate the time savings of the native model reaction, we measured the performance of the PBE0/6-31G\* QM/MM ( $\nu_t = 1.565$  ps/day) and DFTB2/MIO QM/MM + DPRc ( $\nu_h = 392.7$  ps/day) simulations on a single 2.10 GHz Intel Xeon Gold 6230 CPU core. We further note that the first half ( $f_{\text{eq}} = 1/2$ ) of each simulation was discarded as equilibration, the samples were saved every 50 steps ( $f_{\text{file}} = 1/50$ ), and the fraction of statistically independent samples was found to be  $f_{\text{indep}} = 0.763$ . Equation 27 suggests that the gwTP method is

**Table 2. Relative Energies of the Native Model Reaction Free Energy Surfaces Estimated from gwTP Analysis as a Function of the Number of Target Potential Evaluations<sup>a</sup>**

$N_{\text{eval}}$	$f_{\text{eval}}$	min RE	$\langle \text{RE} \rangle$	$\Delta A(\xi_{\text{TST}})$	$\Delta A(\xi_{\text{int}})$	$\Delta A^\ddagger$	$\Delta A$
5.0	0.013	0.42	0.78	12.0 ± 0.5	8.3 ± 0.4	21.2 ± 0.4	3.1 ± 0.4
10.0	0.026	0.56	0.80	11.7 ± 0.4	8.2 ± 0.4	21.0 ± 0.3	3.2 ± 0.3
15.0	0.039	0.64	0.81	11.7 ± 0.4	8.2 ± 0.4	21.1 ± 0.3	3.5 ± 0.2
25.0	0.066	0.70	0.83	11.8 ± 0.4	8.3 ± 0.3	21.2 ± 0.3	3.6 ± 0.2
49.7	0.130	0.75	0.85	11.7 ± 0.2	8.4 ± 0.2	21.5 ± 0.2	3.6 ± 0.1
97.9	0.257	0.71	0.86	11.5 ± 0.2	8.3 ± 0.2	21.1 ± 0.2	3.5 ± 0.1
381.4	1.000	0.82	0.88	11.5 ± 0.2	8.3 ± 0.2	21.1 ± 0.2	3.6 ± 0.1

<sup>a</sup> $N_{\text{eval}}$  is the number of statistically independent samples per simulation used to evaluate the target potential energies.  $f_{\text{eval}}$  is the fraction of the statistically independent samples selected for target potential evaluation. The uncertainties are the standard errors. “min RE” and  $\langle \text{RE} \rangle$  are the minimum and average RE observed during the gwTP analysis of the aggregate sampling.

86 times faster than performing an equivalent amount of ab initio QM/MM sampling for this system even if the full set of uncorrelated samples were used to evaluate the target potential ( $f_{\text{eval}} = 1$ ).

One may question how many of the uncorrelated samples are necessary to obtain a reasonable estimate of the target FES. The required number of target evaluations will generally depend on the agreement between the target and reference potentials. In the limit that the reference was the target potential, an accurate target FES would require the minimum number of samples needed to construct the reference FES. If the target potential differed from the reference, then one would expect this limit to be a lower bound on the number of target evaluations. Furthermore, the required number of samples will depend on the acceptable level of uncertainty. If one is comparing different mechanistic pathways, for example, then a high degree of certainty is not necessary to rule out a path exhibiting a large rate-limiting transition state. Table 2 shows how the free energy values, uncertainties, and REs behave when gwTP analysis is performed with a varying amount of uncorrelated samples. When  $f_{\text{eval}} = 1$ , the analysis is performed with all available uncorrelated samples (after removing the unequilibrated portion of simulations). On average, there are 381.4 samples/(trial·potential·window). In the remaining rows, we perform the analysis using the first  $N_{\text{eval}}$  uncorrelated samples from each simulation. For example,  $N_{\text{eval}} = 15$  signifies that the target potential was evaluated using the first 15 uncorrelated samples/(trial·potential·window), on average. In the present application, we have four reference potentials and each simulation was performed three times, so this corresponds to  $15 \times 4 \times 3 = 180$  evaluations/window. As  $f_{\text{eval}}$  increases, the value of  $N_{\text{eval}}$  ceases to be an integer because there are one or more simulations that had fewer samples than requested. From a practical perspective, we cannot recommend a precise value of  $N_{\text{eval}}$  that the reader should use, for reasons already discussed. Instead, we can offer guidance for a procedure that the reader can follow. First, we perform evaluations with a small number of samples and evaluate the FES. If the RE is less than some tolerance (e.g., 0.6) or the uncertainty in the free energy values is unacceptably large, then one should increase  $N_{\text{eval}}$  accordingly. One could even focus the evaluations in the particular region(s) of the FES where the RE was low. Table 2 suggests that  $N_{\text{eval}} = 15$  samples/simulation are required for the minimum observed entropy to exceed 0.6. If the activation energy must be estimated to be within an uncertainty of 0.2 kcal/mol, then the number of samples must be increased to  $N_{\text{eval}} = 50$ . The surfaces evaluated with  $N_{\text{eval}} = 15$  and  $N_{\text{eval}} = 50$  are compared in Figure S7 in the Supporting Information.

The reader may question if MLPs are required for their application or if it would suffice to perform wTP from uncorrected semiempirical QM/MM sampling. After all, the idea of training an MLP and also performing target potential evaluations to analyze the results may seem formidable. The gwTP method allows one to combine the MLP training and analysis so that all of the target potential evaluations used to train the MLP can also be used to estimate the target FES. For example, one could first perform standard semiempirical QM/MM sampling and estimate the target FES with the wTP method from a limited number of target potential evaluations. If the RE is high, then no refinement is necessary. If the entropy is low, one may deem it worthwhile to reuse the target evaluations to train an MLP correction. The trained MLP is then sampled, and a limited number of target potential evaluations are made to estimate the FES; however, the gwTP method allows one to analyze the FES using the combined sampling from the uncorrected and MLP-corrected simulations. This process can be iterated, such that one performs “active learning” in a manner which reuses the aggregate sampling from previous iterations to improve the predicted target FES while also generating new training data for the next iteration.

## CONCLUSIONS

We described a new reference potential approach for calculating FESs called the gwTP method. The gwTP method extends the formulation of wTP to include sampling from multiple reference potentials. This work was motivated by the growing interest in the use of MLPs being trained to correct semiempirical QM/MM energies and forces to match target ab initio values. The active learning procedure is often employed to train these corrections results in several neural network parameter sets. Each parameter set corresponds to a different reference potential, and the gwTP method can be used to analyze the aggregate sampling from each potential to estimate the target free energies.

We envision other strategies where gwTP analysis may be useful. One could train separate MLPs for different systems, such as RNA catalysis reactions in different enzyme environments, or create a single MLP that tries to model all systems. One can perform sampling with all of these potentials and use gwTP to make a best estimate of another system not included in the training. Alternatively, one could construct potentials that specialize their parametrizations to have high accuracy at different stages of a multistep enzyme reaction. In this way, the gwTP method can be used to combine the sampling from

several “specific reaction parametrizations”<sup>78–80</sup> and thus provide new opportunities in model design.

To this end, we created four ad hoc MNDO/d QM/MM + DPRc neural network parameter sets to calculate the FES of an ethylene phosphate model reaction. Each parametrization limited the training data such that the potentials well-reproduced PBE0/6-31G\* energies and forces in the regions they were trained, but the potentials disagree with each other and the ab initio target in all other regions. We showed that the wTP method fails to accurately predict the target free energies in the untrained regions. We further demonstrated that the gwTP method accurately reconstructs the ab initio surface from the aggregate sampling.

We used the gwTP method to examine the FES of a nonenzymatic RNA 2'-O-transesterification model with an ethoxide leaving group. This system was used as a case study to emphasize the benefits offered by the reference potential approach. Because the reference potentials are much more affordable, the sampling could be extended. The extended simulations allowed us to determine that the ab initio QM/MM sampling was not converged. Excluding the first half of each simulation from the analysis produced a rate-limiting transition state 1.5 kcal/mol higher than the value predicted from ab initio QM/MM sampling; however, the RE remained high. In other words, we made a better estimate of the ab initio FES using gwTP analysis than what direct ab initio QM/MM simulation can reasonably afford. For this system, the computational cost of estimating the ab initio surface from gwTP analysis of the DFTB2/MIO QM/MM + DPRc sampling was found to be 86 times higher than an equivalent amount of PBE0/6-31G\* QM/MM simulation. The cost savings are expected to increase for systems with larger QM regions due to the inherent differences in the formal computational scaling behaviors of ab initio and semiempirical Hamiltonians. The computational cost is minimized by choosing an efficient reference potential and reducing the number of target potential evaluations. The number of target potential evaluations is reduced by analyzing the reference potential simulations to identify and eliminate the unequilibrated portions of the simulations and by extracting the statistically independent samples.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.2c06201>.

Additional descriptions of the DPRc correction and the density-of-states smoothing procedure, comparison of the forward and reverse analysis, and analysis of statistically independent samples (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Darrin M. York** – Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; [orcid.org/0000-0002-9193-7055](https://orcid.org/0000-0002-9193-7055); Email: [Darrin.York@rutgers.edu](mailto:Darrin.York@rutgers.edu)

## Authors

**Timothy J. Giese** – Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; [orcid.org/0000-0002-0653-9168](https://orcid.org/0000-0002-0653-9168)

**Jinzhe Zeng** – Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; [orcid.org/0000-0002-1515-8172](https://orcid.org/0000-0002-1515-8172)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpca.2c06201>

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors are grateful for financial support provided by the National Institutes of Health (nos. GM62248 and GM107485) and also early-stage seed support from the Grossman Innovation Prize established by Alan Grossman. Computational resources were provided by the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey (specifically, the Amarel cluster and associated research computing resources), and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1548562<sup>112</sup> (specifically, the resources COMET and EXPANSE at SDSC through allocation TG-CHE190067). The authors also acknowledge the Texas Advanced Computing Center (TACC, <http://www.tacc.utexas.edu>) at The University of Texas at Austin for providing HPC resources, specifically the Frontera Supercomputer, that have contributed to the research results reported within this paper.

## ■ REFERENCES

- (1) Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J. Am. Chem. Soc.* **2014**, *136*, 528–546.
- (2) *Catalysis in Chemistry and Biology*; Wüthrich, K., Grubbs, R. H., de Bocarmé, T. V., De Wit, A., Eds.; World Scientific, 2018.
- (3) Cheng, G.-J.; Zhang, X.; Chung, L.; Xu, L.; Wu, Y.-D. Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chemical Reactions. *J. Am. Chem. Soc.* **2015**, *137*, 1706–1725.
- (4) Talanquer, V. Importance of Understanding Fundamental Chemical Mechanisms. *J. Chem. Educ.* **2018**, *95*, 1905–1911.
- (5) Christ, C.; Mark, A.; van Gunsteren, W. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569–1582.
- (6) Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932–942.
- (7) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (8) Adib, A. B. Free energy surfaces from nonequilibrium processes without work measurement. *J. Chem. Phys.* **2006**, *124*, 144111.
- (9) Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.* **2001**, *114*, 7330–7337.
- (10) Zuckerman, D. M.; Woolf, T. B. Theory of a Systematic Computational Error in Free Energy Differences. *Phys. Rev. Lett.* **2002**, *89*, 180602.
- (11) Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

- (12) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (13) McDonald, I. R.; Singer, K. Machine Calculation of Thermodynamic Properties of a Simple Fluid at Supercritical Temperatures. *J. Chem. Phys.* **1967**, *47*, 4766–4772.
- (14) McDonald, I. R.; Singer, K. Examination of the Adequacy of the 12–6 Potential for Liquid Argon by Means of Monte Carlo Calculations. *J. Chem. Phys.* **1969**, *50*, 2308–2315.
- (15) Knight, J. L.; Brooks, C. L., III  $\lambda$ -Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.
- (16) Kong, X.; Brooks, C. L., III  $\lambda$ -dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- (17) Liu, Z.; Berne, B. J. Method for accelerating chain folding and mixing. *J. Chem. Phys.* **1993**, *99*, 6071–6077.
- (18) Tidor, B. Simulated annealing on free energy surfaces by a combined molecular dynamics and Monte Carlo approach. *J. Phys. Chem.* **1993**, *97*, 1069–1073.
- (19) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (20) White, A. D.; Dama, J. F.; Voth, G. A. Designing Free Energy Surfaces That Match Experimental Data with Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2451–2460.
- (21) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin–Biotin Rupture Force. *Science* **1996**, *271*, 997–999.
- (22) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997**, *72*, 1568–1581.
- (23) Evans, E.; Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **1997**, *72*, 1541–1555.
- (24) Mezei, M. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. *J. Comput. Phys.* **1987**, *68*, 237–248.
- (25) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J. Chem. Phys.* **1992**, *97*, 6690–6694.
- (26) Bartels, C.; Karplus, M. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- (27) Bartels, C.; Karplus, M. Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B* **1998**, *102*, 865–880.
- (28) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- (29) Hansen, H. S.; Hünenberger, P. H. Using the local elevation method to construct optimized umbrella sampling potentials: calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J. Comput. Chem.* **2010**, *31*, 1–23.
- (30) Rjamani, R.; Naidoo, K. J.; Gao, J. Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. *J. Comput. Chem.* **2003**, *24*, 1775–1781.
- (31) Vanden-Eijnden, E. Some recent techniques for free energy calculations. *J. Comput. Chem.* **2009**, *30*, 1737–1747.
- (32) Weinan, E.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.
- (33) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120*, 7877–7886.
- (34) Weinan, E.; Ren, W.; Vanden-Eijnden, E. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **2005**, *109*, 6688–6693.
- (35) Weinan, E.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.
- (36) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
- (37) Sheppard, D.; Terrell, R.; Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **2008**, *128*, 134106.
- (38) Crehuet, R.; Field, M. J. A temperature-dependent nudged-elastic-band algorithm. *J. Chem. Phys.* **2003**, *118*, 9563–9571.
- (39) Kumar, S.; Rosenberg, D.; Bouzida, R. H.; Swendsen, P. A.; Kollman, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (40) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.
- (41) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.* **2012**, *136*, 144102.
- (42) Zhang, B. W.; Xia, J.; Tan, Z.; Levy, R. M. A Stochastic Solution to the Unbinned WHAM Equations. *J. Phys. Chem. Lett.* **2015**, *6*, 3834–3840.
- (43) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (44) 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.
- (45) Lee, T.-S.; Radak, B. K.; Huang, M.; Wong, K.-Y.; York, D. M. Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *J. Chem. Theory Comput.* **2014**, *10*, 24–34.
- (46) Giese, T. J.; Ekesan, Ş.; York, D. M. Extension of the Variational Free Energy Profile and Multistate Bennett Acceptance Ratio Methods for High-Dimensional Potential of Mean Force Profile Analysis. *J. Phys. Chem. A* **2021**, *125*, 4216–4232.
- (47) Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. *J. Chem. Phys.* **2005**, *123*, 144104.
- (48) Kästner, J.; Thiel, W. Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.* **2006**, *124*, 234106.
- (49) Kästner, J. Umbrella integration in two or more reaction coordinates. *J. Chem. Phys.* **2009**, *131*, 034109.
- (50) 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.
- (51) 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.
- (52) Hu, W.; Li, P.; Wang, J.-N.; Xue, Y.; Mo, Y.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 3. Gaussian Smoothing on Density-of-States. *J. Chem. Theory Comput.* **2020**, *16*, 6814–6822.
- (53) Wang, J.-N.; Liu, W.; Li, P.; Mo, Y.; Hu, W.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 4. Adaptive QM/MM. *J. Chem. Theory Comput.* **2021**, *17*, 1318–1325.
- (54) Jin, S.; Wang, J.-N.; Xue, Y.; Li, P.; Mei, Y. Selectivity of parvalbumin B protein binding to Ca<sup>2+</sup> and Mg<sup>2+</sup> at an ab initio

- QM/MM level using the reference-potential method. *Chin. J. Chem. Phys.* **2021**, *34*, 741–750.
- (55) Xue, Y.; Wang, J.-N.; Hu, W.; Zheng, J.; Li, Y.; Pan, X.; Mo, Y.; Shao, Y.; Wang, L.; Mei, Y. Affordable Ab Initio Path Integral for Thermodynamic Properties via Molecular Dynamics Simulations Using Semiempirical Reference Potential. *J. Phys. Chem. A* **2021**, *125*, 10677–10685.
- (56) 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.
- (57) Gao, J.; Xia, X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631–635.
- (58) 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.
- (59) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM → QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.
- (60) Cave-Ayland, C.; Skylaris, C.-K.; Essex, J. W. Direct validation of the single step classical to quantum free energy perturbation. *J. Phys. Chem. B* **2015**, *119*, 1017–1025.
- (61) Pan, X.; Van, R.; Epifanovsky, E.; Liu, J.; Pu, J.; Nam, K.; Shao, Y. Accelerating Ab Initio Quantum Mechanical and Molecular Mechanical (QM/MM) Molecular Dynamics Simulations with Multiple Time Step Integration and a Recalibrated Semiempirical QM/MM Hamiltonian. *J. Phys. Chem. B* **2022**, *126*, 4226–4235.
- (62) Piccini, G.; Parrinello, M. Accurate Quantum Chemical Free Energies at Affordable Cost. *J. Phys. Chem. Lett.* **2019**, *10*, 3727–3731.
- (63) Ito, S.; Cui, Q. Multi-level free energy simulation with a staged transformation approach. *J. Chem. Phys.* **2020**, *153*, 044115.
- (64) 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.
- (65) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The  $\Delta$ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (66) Pan, X.; Yang, J.; Van, R.; Epifanovsky, E.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-Learning-Assisted Free Energy Simulation of Solution-Phase and Enzyme Reactions. *J. Chem. Theory Comput.* **2021**, *17*, 5745–5758.
- (67) Kim, B.; Snyder, R.; Nagaraju, M.; Zhou, Y.; Ojeda-May, P.; Keeton, S.; Hege, M.; Shao, Y.; Pu, J. Reaction Path-Force Matching in Collective Variables: Determining Ab Initio QM/MM Free Energy Profiles by Fitting Mean Force. *J. Chem. Theory Comput.* **2021**, *17*, 4961–4980.
- (68) Bučko, T.; Gešvandtnerová, M.; Rocca, D. Ab Initio Calculations of Free Energy of Activation at Multiple Electronic Structure Levels Made Affordable: An Effective Combination of Perturbation Theory and Machine Learning. *J. Chem. Theory Comput.* **2020**, *16*, 6049–6060.
- (69) Ruth, M.; Gerbig, D.; Schreiner, P. R. Machine Learning of Coupled Cluster (T)-Energy Corrections via Delta ( $\Delta$ )-Learning. *J. Chem. Theory Comput.* **2022**, *18*, 4846–4855.
- (70) Lu, F.; Cheng, L.; DiRisio, R. J.; Finney, J. M.; Boyer, M. A.; Moonkaen, P.; Sun, J.; Lee, S. J. R.; Deustua, J. E.; Miller, T. F., 3rd; et al. Fast Near Ab Initio Potential Energy Surfaces Using Machine Learning. *J. Phys. Chem. A* **2022**, *126*, 4013–4024.
- (71) Liu, Y.; Li, J. Permutation-Invariant-Polynomial Neural-Network-Based  $\Delta$ -Machine Learning Approach: A Case for the HO<sub>2</sub> Self-Reaction and Its Dynamics Study. *J. Phys. Chem. Lett.* **2022**, *13*, 4729–4738.
- (72) Pham, C. H.; Lindsey, R. K.; Fried, L. E.; Goldman, N. High-Accuracy Semiempirical Quantum Models Based on a Minimal Training Set. *J. Phys. Chem. Lett.* **2022**, *13*, 2934–2942.
- (73) Dettmann, M. A.; Cavalcante, L. S. R.; Magdaleno, C.; Masalkovaitė, K.; Vong, D.; Dull, J. T.; Rand, B. P.; Daemen, L. L.; Goldman, N.; Faller, R.; et al. Comparing the Expense and Accuracy of Methods to Simulate Atomic Vibrations in Rubrene. *J. Chem. Theory Comput.* **2021**, *17*, 7313–7320.
- (74) Bösel, L.; Thürlmann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658.
- (75) Unzueta, P. A.; Greenwell, C. S.; Beran, G. J. O. Predicting Density Functional Theory-Quality Nuclear Magnetic Resonance Chemical Shifts via  $\Delta$ -Machine Learning. *J. Chem. Theory Comput.* **2021**, *17*, 826–840.
- (76) Zeng, J.; Giese, T. J.; Ekesan, Ş.; York, D. M. Development of Range-Corrected Deep Learning Potentials for Fast, Accurate Quantum Mechanical/Molecular Mechanical Simulations of Chemical Reactions in Solution. *J. Chem. Theory Comput.* **2021**, *17*, 6993–7009.
- (77) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.
- (78) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486–504.
- (79) Tejero, I.; González-Lafont, À.; Lluch, J. M. A PM3/s specific reaction parameterization for iron atom in the hydrogen abstraction catalyzed by soybean lipoxygenase-1. *J. Comput. Chem.* **2007**, *28*, 997–1005.
- (80) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. A semiempirical quantum model for hydrogen-bonded nucleic acid base pairs. *J. Chem. Theory Comput.* **2005**, *1*, 1275–1285.
- (81) Bevilacqua, P. C.; Harris, M. E.; Piccirilli, J. A.; Gaines, C.; Ganguly, A.; Kostenbader, K.; Ekesan, Ş.; York, D. M. An Ontology for Facilitating Discussion of Catalytic Strategies of RNA-Cleaving Enzymes. *ACS Chem. Biol.* **2019**, *14*, 1068–1076.
- (82) Chehaibou, B.; Badawi, M.; Bučko, T.; Bazhiron, T.; Rocca, D. Computing RPA Adsorption Enthalpies by Machine Learning Thermodynamic Perturbation Theory. *J. Chem. Theory Comput.* **2019**, *15*, 6333–6342.
- (83) Rizzi, A.; Carloni, P.; Parrinello, M. Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *J. Phys. Chem. Lett.* **2021**, *12*, 9449–9454.
- (84) Herzog, B.; Chagas da Silva, M.; Casier, B.; Badawi, M.; Pascale, F.; Bučko, T.; Lebègue, S.; Rocca, D. Assessing the Accuracy of Machine Learning Thermodynamic Perturbation Theory: Density Functional Theory and Beyond. *J. Chem. Theory Comput.* **2022**, *18*, 1382–1394.
- (85) Huang, M.; York, D. M. Linear free energy relationships in RNA transesterification: theoretical models to aid experimental interpretations. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15846–15855.
- (86) Chen, H.; Giese, T. J.; Huang, M.; Wong, K.-Y.; Harris, M. E.; York, D. M. Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions. *Chem. - Eur. J.* **2014**, *20*, 14336–14343.
- (87) Giese, T. J.; York, D. M. Ambient-Potential Composite Ewald Method for ab Initio Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2016**, *12*, 2611–2632.
- (88) Giese, T. J.; Zeng, J.; Ekesan, Ş.; York, D. M. Combined QM/MM, Machine Learning Path Integral Approach to Compute Free Energy Profiles and Kinetic Isotope Effects in RNA Cleavage Reactions. *J. Chem. Theory Comput.* **2022**, *18*, 4304–4317.
- (89) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; et al. *Amber 2022*; University of California: San Francisco, 2022.

- (90) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanine-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.
- (91) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (92) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (93) Nam, K.; Gao, J.; 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.
- (94) Kubař, T.; Welke, K.; Groenhof, G. New QM/MM implementation of the DFTB3 method in the gromacs package. *J. Comput. Chem.* **2015**, *36*, 1978–1989.
- (95) Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564–1582.
- (96) 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.
- (97) Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (98) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (99) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. Application of an approximate density-functional method to sulfur containing compounds. *J. Mol. Struct.: THEOCHEM* **2001**, *541*, 185–194.
- (100) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (101) Zhang, L.; Han, J.; Wang, H.; Saidi, W.; Car, R.; Weinan, E. End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems. *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc., 2018; pp 4436–4446.
- (102) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.
- (103) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; Weinan, E. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* **2020**, *253*, 107206.
- (104) Lu, D.; Jiang, W.; Chen, Y.; Zhang, L.; Jia, W.; Wang, H.; Chen, M. DP Compress: A Model Compression Scheme for Generating Efficient Deep Potential Models. *J. Chem. Theory Comput.* **2022**, *18*, 5559–5567.
- (105) Caves, L. S.; Evanseck, J. D.; Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **1998**, *7*, 649–666.
- (106) Loccisano, A. E.; Acevedo, O.; DeChancie, J.; Schulze, B. G.; Evanseck, J. D. Enhanced sampling by multiple molecular dynamics trajectories: carbonmonoxy myoglobin 10  $\mu$ s A0  $\rightarrow$  A1-3 transition from ten 400 picosecond simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 369–376.
- (107) Likic, V. A.; Gooley, P. R.; Speed, T. P.; Strehler, E. E. A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. *Protein Sci.* **2005**, *14*, 2955–2963.
- (108) Elofsson, A.; Nilsson, L. How consistent are molecular dynamics simulations?: Comparing structure and dynamics in reduced and oxidized Escherichia coli thioredoxin. *J. Mol. Biol.* **1993**, *233*, 766–780.
- (109) Lahiri, S. N. Comparison of Block Bootstrap Methods. *Resampling Methods for Dependent Data*; Springer: New York, 2003; pp 115–144.
- (110) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397–411.
- (111) Yang, W.; Bitetti-Putzer, R.; Karplus, M. Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.* **2004**, *120*, 2618–2628.
- (112) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gathier, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; et al. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62–74.

## NOTE ADDED AFTER ASAP PUBLICATION

This article originally published with an error in equation 17. The corrected equation published November 3, 2022.

## Recommended by ACS

### Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials

Andrea Rizzi, Michele Parrinello, et al.

SEPTEMBER 23, 2021  
THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS

READ 

### On Sampling Minimum Energy Path

Mouad Ramil, Jean-Bernard Maillet, et al.

SEPTEMBER 08, 2022  
JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

### Neural Network for Principle of Least Action

Beibei Wang, Mark Stevens, et al.

JULY 05, 2022  
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

### Grassmann Extrapolation of Density Matrices for Born–Oppenheimer Molecular Dynamics

Étienne Polack, Filippo Lipparini, et al.

OCTOBER 08, 2021  
JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Get More Suggestions >