

Combined QM/MM, Machine Learning Path Integral Approach to Compute Free Energy Profiles and Kinetic Isotope Effects in RNA Cleavage Reactions

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



Cite This: *J. Chem. Theory Comput.* 2022, 18, 4304–4317



Read Online

ACCESS |



Metrics & More

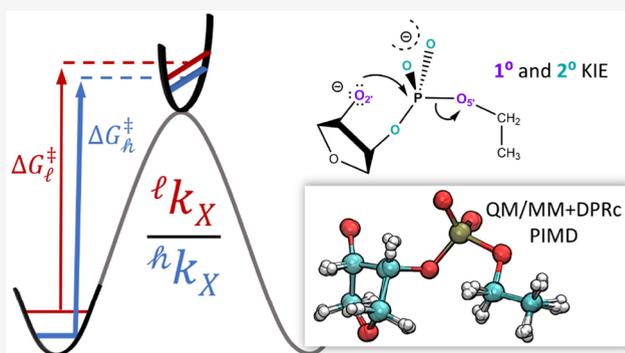


Article Recommendations



Supporting Information

ABSTRACT: We present a fast, accurate, and robust approach for determination of free energy profiles and kinetic isotope effects for RNA 2'-O-transphosphorylation reactions with inclusion of nuclear quantum effects. We apply a deep potential range correction (DPRc) for combined quantum mechanical/molecular mechanical (QM/MM) simulations of reactions in the condensed phase. The method uses the second-order density-functional tight-binding method (DFTB2) as a fast, approximate base QM model. The DPRc model modifies the DFTB2 QM interactions and applies short-range corrections to the QM/MM interactions to reproduce *ab initio* DFT (PBE0/6-31G*) QM/MM energies and forces. The DPRc thus enables both QM and QM/MM interactions to be tuned to high accuracy, and the QM/MM corrections are designed to smoothly vanish at a specified cutoff boundary (6 Å in the present work). The computational speed-up afforded by the QM/MM+DPRc model enables free energy profiles to be calculated that include rigorous long-range QM/MM interactions under periodic boundary conditions and nuclear quantum effects through a path integral approach using a new interface between the AMBER and i-PI software. The approach is demonstrated through the calculation of free energy profiles of a native RNA cleavage model reaction and reactions involving thio-substitutions, which are important experimental probes of the mechanism. The DFTB2+DPRc QM/MM free energy surfaces agree very closely with the PBE0/6-31G* QM/MM results, and it is vastly superior to the DFTB2 QM/MM surfaces with and without weighted thermodynamic perturbation corrections. ¹⁸O and ³⁴S primary kinetic isotope effects are compared, and the influence of nuclear quantum effects on the free energy profiles is examined.



1. INTRODUCTION

RNA strand cleavage by 2'-O-transphosphorylation is ubiquitous in biology^{1–3} and has far-reaching implications for medicine.⁴ There is thus great interest in obtaining a predictive understanding of the mechanisms of these reactions and the nature of the transition states that control their rates. The most sensitive experimental probes that report on changes in the structure and bonding in the transition state are the measurement of kinetic isotope effects (KIE).^{5–7} These experiments are challenging, often requiring painstaking synthetic efforts and careful, highly sensitive measurements to realize.⁵ For complex systems, the interpretation of these measurements in terms of structure and bonding requires recourse into computational simulations.^{8–10}

Ab initio combined quantum mechanical/molecular mechanical (QM/MM) simulations with rigorous treatment of long-range electrostatic interactions under periodic boundary conditions^{11,12} affords a powerful tool to gain insight into the pathways of these reactions, their transition states and intermediates, and environmental factors that modulate reactivity.^{13,14} However, *ab initio* QM/MM methods are

extremely computationally intensive, and hence the elucidation of the free energy surfaces (FESs) for complex reactions that require extensive sampling is often not practical. Methods that enable the treatment of nuclear quantum effects, despite great advances in new methods and software for their computation, nonetheless considerably exacerbate the computational cost making all but the most modest calculations prohibitive.

Machine learning potentials offer a potential mechanism to improve the accuracy and efficiency of QM/MM simulations, and they have had considerable impact in the development of methods to study chemical reactions.^{15–21} Herein, we develop an approach whereby we employ a recently described deep-potential range correction (DPRc) model²² to enhance the accuracy of a fast, approximate base QM/MM model to

Received: February 14, 2022

Published: June 16, 2022



reproduce the energies and forces of a much more computationally costly target QM/MM model. The new model parametrizes the DPRc potential using a machine learning neural network training procedure²² to correct the second-order density-functional tight-binding (DFTB2) semiempirical method^{23–25} to reproduce the PBE0/6-31G* energies and forces in explicit solvent QM/MM calculations. We describe a framework for introducing nuclear quantum effects into the calculations using path integral molecular dynamics (PIMD) through an interface between AMBER20²⁶ and i-PI²⁷ software packages. Together, this enables the calculation of QM/MM+DPRc free energy profiles^{28–30} with and without nuclear quantum effects and the prediction of primary and secondary KIE values. The methods are demonstrated and tested using a well-characterized native nonenzymatic RNA transphosphorylation reaction model for which KIE values are available experimentally, as well as a series of chemically modified (thio-substituted) reactions that are often used experimentally in mechanistic studies.^{31,32}

Figure 1 shows the specific reactions examined in this work. The “native” model reaction is a hydroxyalkyl phosphate ester

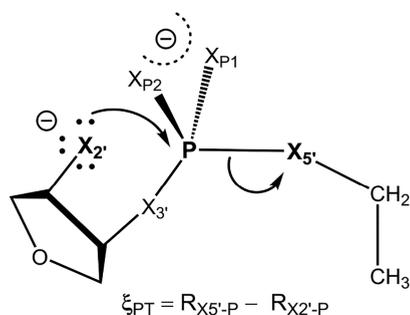


Figure 1. Model nonenzymatic phosphoryl transesterification reaction in solution. In a native RNA system, the positions labeled “X” correspond to oxygen positions (canonical RNA numbering scheme is used to identify atomic positions). However, chemically modified variants involving thio-substitution at one or more of these positions are commonly used in experimental mechanistic studies of RNA-cleaving enzymes.⁴⁵ In addition to the native model system, we will consider single thio-substitutions at the 2', 3', 5', and OP1 positions, as well as a double thio-substitution at the OP1 and OP2 positions. We designate these reactions as S2', S3', S5', S1P, and S12, respectively. The phosphoryl transfer reaction coordinate, $\xi_{PT} = R_{X5'-P} - R_{X2'-P}$, is a difference between bond-breaking and bond-forming bond lengths $R_{X5'-P}$ and $R_{X2'-P}$, respectively.

that undergoes phosphoryl transesterification with an ethoxide leaving group. The other systems explored in this work differ only by thio-substitution(s) at key oxygen positions. Similar model reactions with various leaving groups have been studied using implicit solvent *ab initio* calculations and linear free energy relationships (LFER).^{9,33} It was found that the mechanistic pathway is correlated to the pK_a of the leaving group. “Enhanced” leaving groups (leaving groups with a $pK_a < 11$) were found to proceed through a concerted mechanism containing a single, “early” ($\xi_{PT} < 0$) transition state. The reaction mechanism of “poor” leaving groups (leaving groups with a $pK_a > 12$) was found to be associative and proceeds through two distinct barriers separated by a minimum.^{9,33} The early transition state is characterized by partial formation of the O2'–P bond, whereas the second rate-controlling transition state is “late” ($\xi_{PT} > 0$) and is characterized by partial cleavage

of the O5'–P bond. Two-dimensional FESs of model phosphoryl transfer reactions with ethoxide (poor) and phenoxide (enhanced) leaving groups have been performed with *ab initio* umbrella sampling in explicit solvent.¹¹ The phenoxide FES was found to possess one, early transition state, and the ethoxide FES possessed two barriers with a late, rate limiting transition state.¹¹ These results are consistent with the LFER analysis.³³ The stability of the intermediate encountered with poor leaving groups should be sensitive to the solvation environment, that is, by the chosen solvation radius used in implicit solvent calculations or the Lennard-Jones radii employed in explicit solvent simulations. Furthermore, estimation of the expected lifetime of the intermediate requires one to also consider a mechanistic pathway that involves pseudorotation of the pentacoordinate structure.^{34–37} The model reactions in the present work involve an ethoxide leaving group, and the FES explored by the ξ_{PT} reaction coordinate is not intended to explore a pseudorotation mechanism; therefore, the native reaction model is expected to display two barriers with a late, rate limiting transition state. If the 2' nucleophilic or 5' leaving group oxygen is thio-substituted, however, then the asymmetric electron affinities may result in a single transition state.

The development of machine learning correction potentials is not intended to supplant the use of free energy perturbation reweighting approaches. Methods such as weighted thermodynamic perturbation^{38–40} (wTP) can be a cost-effective means to estimate the FES of an expensive, target Hamiltonian by reweighting the umbrella sampling performed with a reference potential; however, the success of reweighting methods heavily relies upon the degree of agreement between the reference and target potentials. A particularly useful metric to gauge the reliability of the free energy estimate is the “reweighting entropy”.⁴¹ Machine learning corrections are particularly well suited to situations where there is significant disagreement between the reference and target potentials, which corresponds to a low reweighting entropy. The reactions examined in the present work are not well modeled by the DFTB2/MIO semiempirical method^{25,42,43} because the P–S repulsive potentials were not specifically trained to reproduce the chemistry of small phosphates.^{43,44} We will show that reweighting of the DFTB2/MIO umbrella sampling yields poor estimates of the *ab initio* FESs, which motivates our development of the DPRc correction.

2. METHODS

2.1. Reaction Models and Initial System Setup. The ML network parameters within the DFTB2 QM/MM+DPRc method were parametrized to reproduce PBE0/6-31G* QM/MM total energies and forces^{46,47} for a series of RNA-like nonenzymatic model systems, as shown in Figure 1.

The initial solute structures were generated from SMILES using Open Babel.⁴⁸ The QM region was defined as the entire RNA-like solute molecule. The solute was solvated by 1510 TIP4P/Ew waters.⁴⁹ Simulations were performed with the SANDER program within AMBER20²⁶ using a 1 fs time step. The systems were equilibrated by performing DFTB2 (“MIO” parameter set^{25,42,43}) QM/MM simulations that gradually heated the system from 0 to 298 K over the course of 100 ps.⁵⁰ This was followed by 100 ps of density equilibration in the isothermal–isobaric ensemble at 1 atm and 298 K using the Berendsen barostat and Langevin thermostat.⁵¹ The Lennard-Jones potential was truncated at 8 Å, and a long-range tail

correction was used to model the interactions beyond the cutoff.⁵² The long-range electrostatics were evaluated with the QM/MM particle mesh Ewald method using a 1 Å³ grid spacing, a 8 Å real-space cutoff, and tinfoil boundary conditions.^{53–55}

2.2. *Ab Initio* Quantum Mechanical/Molecular Mechanical Simulations. PBE0/6-31G* QM/MM umbrella sampling was performed to obtain reference free energy surfaces and to provide structures, energies, and forces to initiate the DFTB2 QM/MM+DPRc network parameter optimization. The RNA cleavage reaction⁵⁶ is described by the collective variable (progress variable) for phosphoryl transfer (PT), ξ_{PT} , defined as the difference between the 5'–P and 2'–P bond distances.

$$\xi_{\text{PT}} = R_{\text{X}5'-\text{P}} - R_{\text{X}2'-\text{P}} \quad (1)$$

$R_{\text{X}5'-\text{P}}$ and $R_{\text{X}2'-\text{P}}$ are the bond-breaking and bond-forming bond lengths in Figure 1, respectively, and X is either O or S, depending on the thio-substitution. The FESs were reconstructed from a series of umbrella sampling simulations that vary the location of the restraining potential from –3.5 to 5 Å in steps of 0.1 Å. The umbrella force constants were 100 kcal mol^{–1} Å^{–2}. The simulations were performed with TIP4P/Ew rigid water in the canonical ensemble for 25 ps at 298 K. The ambient potential composite Ewald method was used to compute the *ab initio* QM/MM electrostatic interactions using a 1 Å³ regular grid spacing, a 8 Å real-space cutoff, and tinfoil boundary conditions.¹¹ The value of ξ_{PT} was written to file during the simulation, and the FESs were generated by analyzing the ξ_{PT} time series using the variational free energy profile (vFEP) method^{28,29} implemented in the FE-ToolKit software.³⁰

2.3. Range-Corrected Deep Potential (DPRc). The DFTB2 QM/MM+DPRc method modifies the DFTB2 QM/MM total energy by the inclusion of a machine learning potential which provides a nonelectronic correction to the QM-QM and nearby QM-MM interactions

$$E(\mathbf{R}; \mathbf{P}) = E_{\text{QM}}(\mathbf{R}; \mathbf{P}) + E_{\text{QM/MM}}(\mathbf{R}; \mathbf{P}) + E_{\text{MM}}(\mathbf{R}) + E_{\text{ML}}(\mathbf{R}) \quad (2)$$

where \mathbf{R} is an $N \times 3$ array of atomic coordinates, and \mathbf{P} is the QM Hamiltonian's single particle density matrix. $E_{\text{QM}}(\mathbf{R}; \mathbf{P})$ and $E_{\text{MM}}(\mathbf{R})$ are the QM and MM energies, respectively. $E_{\text{QM/MM}}(\mathbf{R}; \mathbf{P})$ contains the electrostatic and Lennard-Jones (or other nonelectrostatic-nonbonded model) interactions between the QM and MM regions. The MIO parameter set^{25,42,43} is used in both the DFTB2 and DFTB2 QM/MM+DPRc models. $E_{\text{ML}}(\mathbf{R})$ is the DPRc potential.²² In brief, the DPRc potential is an extension of the DeepPot-SE model⁵⁷ that includes corrections for the QM-QM and QM-MM interactions. The key features of the correction are that it does not modify the interactions between MM atoms, the MM atoms do not contribute a constant to the total energy, and the correction between the QM and MM atoms smoothly approaches zero as the distance between the QM and MM

atoms nears the correction cutoff (a parameter that can be adjusted). By constructing the correction potential to have these properties, the system total energy is conserved as the MM atoms approach or diffuse from the vicinity of the QM region during the course of simulation. The DPRc correction does not directly effect the calculation of the long-range interactions, which continue to be evaluated with the DFTB2 QM/MM PME method using a 1 Å³ reciprocal space grid, a 8 Å real space cutoff, and tinfoil boundary conditions.^{53–55} Our in-house *ab initio* software¹¹ also performs PME, which greatly simplifies the training of the network parameters; the reference PBE0/6-31G* QM/MM energies and forces of the entire periodic unit cell are computed without needing to truncate the system.^{12,19} The DPRc energy correction is a sum of atomic contributions.

$$E_{\text{ML}} = \sum_{i=1}^N E_i \quad (3)$$

The atomic energy E_i is the output layer of a “fitting network” (denoted \mathcal{N}_2) composed of 3 hidden layers using 240 neurons/layer.

$$E_i = \begin{cases} \mathcal{N}_2(\mathcal{D}_i), & i \in \text{QM} \\ \mathcal{N}_2(\mathcal{D}_i) - \mathcal{N}_2(\mathbf{0}), & i \in \text{MM} \end{cases} \quad (4)$$

The input layer of \mathcal{N}_2 is the atomic descriptor $\mathcal{D}_i \in \mathbb{R}^{M_1 \times M_2}$, and $\mathbf{0} \in \mathbb{R}^{M_1 \times M_2}$ is a matrix of zeros with the same shape as \mathcal{D}_i .

$$\mathcal{D}_i = (\mathcal{G}_{i1})^T \mathcal{R}_i (\mathcal{G}_{i2})^T \quad (5)$$

The embedding matrix $\mathcal{G}_{i1} \in \mathbb{R}^{N \times M_1}$ is the output layer of a “filtering network” \mathcal{N}_1 composed of 2 hidden layers with 13 and 26 neurons, respectively, and an output layer consisting of 52 neurons ($M_1 = 52$). The columns of the truncated embedding matrix $\mathcal{G}_{i2} \in \mathbb{R}^{N \times M_2}$ correspond to the first $M_2 = 12$ columns of \mathcal{G}_{i1} . By decreasing the M_2 filter axis, one improves the computational efficiency by reducing the size of the descriptor \mathcal{D}_i .

$$(\mathcal{G}_{i1})_j = \mathcal{N}_1(s(\mathcal{R}_{ij})) \quad (6)$$

The “coordinate matrix” $\mathcal{R}_i \in \mathbb{R}^{N \times 4}$ contains 4 elements for each row.

$$(\mathcal{R}_i)_j = \left\{ s(\mathcal{R}_{ij}), \frac{s(\mathcal{R}_{ij})X_{ij}}{R_{ij}}, \frac{s(\mathcal{R}_{ij})Y_{ij}}{R_{ij}}, \frac{s(\mathcal{R}_{ij})Z_{ij}}{R_{ij}} \right\} \quad (7)$$

X_{ij} , Y_{ij} , and Z_{ij} are the Cartesian components of the difference vector $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$, and R_{ij} is the internuclear separation. The $s(\mathcal{R}_{ij})$ quantity causes the correction to the QM-MM interactions to smoothly approach zero from $R_{ij} = R_{\text{on}}$ to $R_{ij} = R_{\text{off}}$. In the present work, $R_{\text{on}} = 1$ Å, $R_{\text{off}} = 6$ Å, and there is no correction cutoff between the QM and other QM atoms.

$$s(R_{ij}) = \begin{cases} 0, & \text{if } i, j \in \text{MM} \\ \frac{1}{R_{ij}}, & \text{if } i, j \in \text{QM} \\ \frac{1}{R_{ij}}, & \text{if } i \in \text{QM} \wedge j \in \text{MM} \text{ or } i \in \text{MM} \wedge j \in \text{QM}: \\ \frac{1}{R_{ij}}, & \text{if } R_{ij} \leq R_{\text{on}} \\ \frac{1}{R_{ij}} \left\{ \left(\frac{R_{ij} - R_{\text{on}}}{R_{\text{off}} - R_{\text{on}}} \right)^3 \left[-6 \left(\frac{R_{ij} - R_{\text{on}}}{R_{\text{off}} - R_{\text{on}}} \right)^2 + 15 \frac{R_{ij} - R_{\text{on}}}{R_{\text{off}} - R_{\text{on}}} - 10 \right] + 1 \right\}, & \text{if } R_{\text{on}} < R_{ij} < R_{\text{off}} \\ 0, & \text{if } R_{ij} \geq R_{\text{off}} \end{cases} \quad (8)$$

The filtering and fitting networks are both activated by the *tanh* function, and their parameters vary by element. The role of the filtering network is to provide a description of an atom and its surroundings, whereas the fitting network outputs an energy correction for an atom, given the atom's description. The underlying DeepPot-SE model, on which the DPRc potential is based, uses "atom type" assignments to describe the environment around each atom. For the QM atoms, we use the atomic number. For MM atoms, we prepend the atomic number by the letter "m", so that the MM atoms can be corrected differently than the QM atoms. We have implemented the DPRc method into the DeePMD-kit software package⁵⁸ and created an interface with a development version of AMBER's SANDER program.²⁶

The network parameters are optimized using an active learning approach described in detail elsewhere.²² The parameter optimizations were performed with the DP-GEN software,⁵⁹ and the DP Compress algorithm^{60,61} was applied to the trained models to improve computational performance and reduce the memory requirements. The active learning procedure involves 3 stages: training, exploration, and labeling. The training stage uses the available reference data to perform a stochastic "small-batch" minimization of the loss function⁶² to reproduce the difference between the DFTB2 and PBE0/6-31G* QM/MM energies and forces. The training is performed 4 times using different random number seeds, yielding 4 trial parameter sets. The exploration stage searches for relevant structures that should be used to retrain the correction. The search is made by performing DFTB2 QM/MM+DPRc umbrella sampling with one of the trial parameter sets for 25 ps/window. Every 50 steps, the DFTB2 QM/MM+DPRc energies and forces are re-evaluated using each of the 4 trial parameter sets, and the standard deviation (between DPRc corrections) of atomic forces is calculated. If the maximum standard deviation is in the range 0.1–0.25 eV/Å, then the trajectory frame is saved. The labeling stage evaluates the DFTB2 and PBE0/6-31G* QM/MM energies and forces of the periodic unit cell (with inclusion of long-range electrostatics). The differences between the methods are used as additional reference data in the next active learning cycle's training stage.

The active learning procedure is terminated when the 4 parameter sets agree for 99.8% of the frames in the current cycle. The 99.8% termination criteria is an empirical optimization threshold intended to strike a balance between the computational resources required to continue the active

learning procedure and the current uncertainty in the model fit. One can always imagine that including additional active learning iterations may discover new structures where the trial parameters yield disagreement; however, this concern is reduced if the number of new structures converges toward zero as the active learning cycle progresses. Table S2 in the Supporting Information lists the number of new structures encountered at each active learning iteration, which monotonically decays toward zero. A 2-parameter exponential fit results in a 0.97 Pearson correlation coefficient, which could be used to predict the number of new structures one may find in future iterations. The parameters of the fit will likely depend on the complexity of the systems being trained and the discrepancies between the base and target Hamiltonians. The correlation between the DPRc atomic force correction and the difference between the PBE0/6-31G* and DFTB2 forces is shown in Figure S1 of the Supporting Information.

In the present work, we perform PBE0/6-31G* umbrella sampling to make a comparison; therefore, we reused these PBE0/6-31G* trajectories to initiate the DPRc training procedure. Table S1 in the Supporting Information provides the number of PBE0/6-31G* structures used from each system to initiate the training. In principle, one could begin the first active learning cycle from DFTB2 QM/MM umbrella sampling. If the training was initiated from the semiempirical base model, the active learning procedure will likely need more cycles to reach convergence than what was required in the present work. Furthermore, if the base and target Hamiltonians significantly differed, then the initial training may include irrelevant high-energy structures; however, each active learning cycle begins with retraining, so each successive exploration phase should progressively approach an ensemble of relevant structures.

2.4. Path Integral Molecular Dynamics (PIMD) Simulations. After training the DFTB2 QM/MM+DPRc network parameters to the Born–Oppenheimer umbrella sampling, we applied the final 4 parameter sets to the calculation of FESs that consider nuclear quantum effects by performing umbrella sampling with PIMD. The PIMD simulations were performed by interfacing the i-PI software²⁷ to a development version of AMBER's SANDER program.²⁶ The i-PI software is responsible for dynamically evolving the system through time, and it sends requests (either via Linux sockets or Internet protocol) for energy and force evaluations. In this respect, SANDER is merely used as a calculator of the QM/MM total energy and force of the system, which is

returned to i-PI via the Internet protocol interface. The QM/MM total energy includes the particle mesh Ewald treatment of electrostatics,^{11,53–55} described above for the classical MD simulations, for all methods (DFTB2, PBE0/6-31G*, and DFTB2 QM/MM+DPRc). When launching SANDER, new command line options were introduced that supply the Internet protocol address of the computer running i-PI and the network port number to communicate messages. Rather than performing classical MD, SANDER enters an infinite loop in which it listens for messages on the specified port until a message is received to stop execution. A PIMD simulation requires multiple energy and force evaluations at each time step, which can be easily parallelized by launching multiple instances of SANDER, each of which may further parallelize the energy and force evaluation.

The PIMD dynamical motion was propagated with 6 beads (replicas) at a 0.25 fs time step at 298 K using the PIGLET quantum thermostat;^{63,64} therefore, up to 6 SANDER instances can be launched. The parameters for the PIGLET thermostat were taken from the GLE4MD Web site.^{65,66} The parameters were chosen to reproduce the quantum fluctuations at 298 K and span a range of frequencies up to 4142 cm⁻¹. The DFTB2 QM/MM+DPRc PIMD umbrella sampling was performed with the q-SPC/Fw flexible water model.^{67,68} The systems were prepared by replacing the TIP4P/Ew solvent with q-SPC/Fw, and the system density was re-equilibrated at 298 K and 1 atm pressure for 100 ps using classical MD. The final structure of the classical QM/MM density equilibration was used as the initial structure in the PIMD simulations. The PIMD restraint potentials were applied to the centroid positions, rather than the positions within each bead, to avoid an artificial perturbation of the nuclear wavepackets. The umbrella potentials were therefore applied via i-PI, as opposed to SANDER, by making use of the i-PI interface to PLUMED.^{69,70} An extended discussion regarding the application of restraints can be found in the Supporting Information. Each simulation was performed 16 times for 20 ps (0.25 fs time step) using a 200 kcal mol⁻¹ Å⁻² force constant. The 16 runs correspond to 4 simulations initiated from different thermostat random seeds for each of the 4 network parameter sets. During the course of simulation, the value of ξ_{PT} (as calculated from the centroid positions) was written to file, and the vFEP method was used to produce 16 FESs for each nonenzymatic system. At each point along ξ_{PT} , we report the average and standard deviation of the 16 FESs. An alternative approach for calculating FESs from PIMD simulations⁴⁰ is to use a semiempirical reference potential, which is a viable strategy when the reference and target potentials have good phase space overlap.^{38–40}

For comparison, we performed classical MD DFTB2 QM/MM+DPRc umbrella sampling with SPC/Fw solvent.⁶⁷ Four sets of classical MD umbrella sampling simulations were performed with SANDER, corresponding to the 4 sets optimized network parameters. Each simulation was run for 100 ps using a 1 fs time step, and each set of simulations was independently analyzed with vFEP to produce 4 FESs. At each point along ξ_{PT} , we report the average and standard deviation of the 4 FESs. Note that the FESs obtained from classical MD were performed with the SPC/Fw water model, whereas the PIMD simulations were performed with the q-SPC/Fw water model. The SPC/Fw water model was designed for use in classical MD simulations,⁶⁷ whereas the q-SPC/Fw water model has slightly modified parameters tuned to improve the

properties of bulk water in PIMD simulations.⁶⁸ We compare the FESs obtained from classical MD simulations using both water models in Figure S2 of the Supporting Information. In brief, the minor differences between the two water models do not significantly impact the calculated FESs; the FESs agree to within the uncertainties of the calculations. Figure S4 in the Supporting Information also provides a comparison between FESs calculated from classical MD simulations performed with SANDER and classical MD simulations performed through i-PI (in which case SANDER is used to evaluate the energies and forces). The resulting FESs obtained from both programs agree to within the uncertainties of the calculations.

2.5. Kinetic Isotope Effects. We applied the parametrized DFTB2 QM/MM+DPRc models to the calculation of OS' and O2' heavy atom kinetic isotope effects (KIE). KIE values are the ratios of reaction rate constants upon isotopic substitution; that is, $\eta = k_L/k_H$, where k_L and k_H are the rate constants of the light and isotopically substituted (heavy) systems, respectively. There are many methods for approximating the KIE,^{71–76} including the Bigeleisen-Mayer equation,^{77,78} which uses light and heavy normal mode vibrational frequencies of the reactant ($\{\omega_L\}$ and $\{\omega_H\}$) and the transition state ($\{\omega_L^\ddagger\}$ and $\{\omega_H^\ddagger\}$).

$$\eta^{\text{BM}} = \frac{\omega_{L,1}^\ddagger}{\omega_{H,1}^\ddagger} \frac{\prod_{i=2}^{3N-6} \frac{\omega_{L,i}^\ddagger \sinh(\hbar\beta\omega_{H,i}^\ddagger/2)}{\omega_{H,i}^\ddagger \sinh(\hbar\beta\omega_{L,i}^\ddagger/2)}}{\prod_{i=1}^{3N-6} \frac{\omega_{L,i} \sinh(\hbar\beta\omega_{H,i}/2)}{\omega_{H,i} \sinh(\hbar\beta\omega_{L,i}/2)}} \quad (9)$$

Eq 9 assumes there are 3N-6 internal degrees of freedom, and $\omega_{L/H,1}^\ddagger$ is the sole imaginary frequency. An alternate approach is to approximate the rate constants from Kramers' theory in the high-friction limit.^{79–82}

$$k_{L/H} = \frac{\omega_{L/H,1}^\ddagger}{\gamma} k_{L/H}^{\text{TST}} \quad (10)$$

γ is a mass weighted friction coefficient, and k^{TST} is the conventional transition state theory rate constant. In this work, we estimate the KIE from eq 11, where ΔA_{TS} and ΔA_{Min} are the free energies associated with changing the light isotope mass to a heavy isotope mass in the transition state and reactant minimum, respectively, and $\eta^{\text{TST}} = \exp[\beta(\Delta A_{\text{TS}} - \Delta A_{\text{Min}})]$.

$$\begin{aligned} \eta &= \left\langle \frac{\omega_{L,1}^\ddagger}{\omega_{H,1}^\ddagger} \right\rangle \eta^{\text{TST}} \\ &= \left\langle \frac{\omega_{L,1}^\ddagger}{\omega_{H,1}^\ddagger} \right\rangle e^{\beta(\Delta A_{\text{TS}} - \Delta A_{\text{Min}})} \end{aligned} \quad (11)$$

The ΔA_{TS} and ΔA_{Min} values are calculated from PIMD simulations using the thermodynamic free energy perturbation (TD-FEP) method,⁸³ described in the next subsection. The leading ratio of frequencies is estimated from the normal-mode analysis of many, independently optimized transition states starting from different initial structures. The brackets in eq 11 indicate that we use the average ratio obtained from the independent estimates. The transition states used to obtain the estimate are also used to calculate KIE values from eq 9 for comparison. The computational details of the geometry optimizations are described below.

PIMD TD-FEP Approach. The TD-FEP method calculates ΔA_{TS} and ΔA_{Min} from free energy perturbation of the atomic

masses.⁸³ This makes the TD-FEP method conceptually similar to the path integral free energy perturbation theory presented by Gao.⁸⁴ In brief, the free energy of transforming a mass from m to m' is estimated from eq 12

$$\Delta A = \frac{d}{2\beta} \log \frac{m'}{m} - \int_m^{m'} \frac{\langle T(\mu) \rangle}{\mu} d\mu \quad (12)$$

where d is the dimensionality of the problem, and $\langle T(\mu) \rangle$ is the ensemble average quantum kinetic energy of the atom with mass μ . In principle, one could evaluate the integral by performing a series of PIMD simulations that differ only by the chosen atomic mass μ and writing the atom's quantum kinetic energy to file during the course of simulation. From the series of PIMD simulations, the $\langle T(\mu) \rangle$ averages can be computed, and the integral could be performed either by numerical quadrature or by integrating a spline. This would be analogous to performing multiple thermodynamic integration simulations at intermediate states that connect two physical states. An advantage of the TD-FEP method is its ability to provide estimates of $T(\mu)$ for many values of μ from a *single* PIMD simulation performed using the light atomic masses. In this respect, the TD-FEP method is similar to the free energy perturbation umbrella sampling method (PI-FEP/UM) developed by Major and Gao.⁷² We instruct i-PI to calculate $T(\mu)$ of the selected atoms at 11 uniformly spaced mass values that range from the most abundant isotopic mass to the second most abundant isotopic mass. In this work, we calculate KIE values that transform ^{16}O to ^{18}O at the 2' and 5' positions. The thio-substituted model systems transform ^{32}S to ^{34}S when the selected positions have been replaced with sulfur atoms. The integral appearing in eq 12 is calculated by representing the $\langle T(\mu) \rangle$ values as an Akima spline and interpolating its values at 1000 uniformly spaced masses for trapezoidal-rule numerical integration.

Because ΔA_{TS} and ΔA_{Min} are both required, one must perform PIMD simulations of both the reactant and transition state. To perform simulations of these states, we restrain ξ_{PT} to the appropriate value, based on the location of the minimum and transition state identified on the FES by applying a 200 kcal mol⁻¹ Å⁻² harmonic restraint to the centroid positions. The simulations were performed with the SPC/Fw water model for 10 ps (0.25 fs time step) at 298 K using a 6-bead PIGLET quantum thermostat.^{63,64} Each DFTB2 QM/MM +DPRc simulation was repeated 4 times, corresponding to the use of the 4 active learning parameter sets, yielding 4 independent estimates of the KIE. We also performed an analogous set of DFTB2 QM/MM PIMD simulations (without the DPRc correction) with different random seeds to provide 4 estimates of the KIE for comparison. Furthermore, we performed PBE0/6-31G* QM/MM PIMD simulations to estimate the KIE; however, due to the much higher cost of the *ab initio* QM/MM PIMD simulations, only 1 estimate of each KIE was made. The uncertainty is estimated by calculating the standard error of the KIE from the analysis of 4 nonoverlapping segments of the trajectory.

Local Minimum Harmonic Approximation. Bigeleisen-Mayer Equation. An alternative approach to estimating KIE values is their calculation from vibrational analysis at geometrical stationary points using the Bigeleisen-Mayer equation.^{77,78} The application of this approach to condensed phase systems with explicit waters is complicated by the myriad possible stationary points. To make a comparison with this

approach, we attempted to find many minima and transition state structures starting from structures encountered during classical QM/MM simulations. The stationary point geometry optimizations were performed by interfacing SANDER to the DL-Find geometry optimization software.⁸⁵

A set of reaction minimum structures was found by performing classical QM/MM simulations near the minimum of the FES, extracting configurations from the trajectory file, removing the harmonic restraints, and optimizing the geometry for a minimum using the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (LBFGS). Only the solute and the nearby 6 Å of SPC/Fw solvent molecules were allowed to change their atomic positions. The atomic positions of the remaining system were fixed; however, the energy and forces are computed for the entire unit cell. After the stationary point was found, the Hessian was calculated from finite differentiation of the atomic gradients. The dimensions of the Hessian are proportional to the number of optimizable atoms; however, the forces are obtained from the full periodic system. The vibrational frequencies were obtained from diagonalization of the mass weighted Hessian upon removal of the translational and rotational degrees of freedom.

The vibrational frequencies of the transition state structures were performed analogously; however, the optimization of the transition state structures used the following procedure: (1) A structure near the transition state was taken from a stored trajectory file. (2) The solute and nearby 6 Å of solvent molecules were selected for optimization. (3) A series of restrained geometry minimizations were performed, each differing by the position of a ξ_{PT} harmonic restraint. (4) The unrestrained potential energy of each structure was monitored, and the structure with the maximum potential energy was selected for further optimization. (5) The solute and nearby 3 Å of solvent molecules were allowed to move during transition state search using the partitioned rational function algorithm.

We have described a procedure for calculating KIE values using the Bigeleisen-Mayer equation, and we apply this procedure to obtain an average KIE value (and uncertainty) from independent estimates generated from different initial conditions. In principle, one could use multistructural variational transition state theory (MS-VTST)^{86–88} or multipath variational transition state theory (MP-VTST)^{87–89} to treat multiple stationary points. The VTST methods have seen success when applied in conjunction with the equilibrium solvation path approximation⁹⁰ or microsolvation.⁷⁴ The application of any approach that incorporates Boltzmann averaging of geometry optimized periodic, condensed-phase systems, however, is complicated by the myriad possible local minimum that may exist far from the chemical event of interest.

3. RESULTS AND DISCUSSION

3.1. *Ab Initio* QM/MM and QM/MM+DPRc Free Energy Profiles. Figure 2 compares FESs produced by DFTB2 QM/MM, PBE0/6-31G* QM/MM, and DFTB2 QM/MM+DPRc classical MD simulations in the TIP4P/Ew solvent. The DFTB2 QM/MM simulations were run 4 times initiated from different random number seeds, yielding 4 estimates of the FES. The figure plots the mean, and the vertical bars are the standard error of the mean. The DFTB2 QM/MM+DPRc simulations were similarly run 4 times, once for each set of optimized network parameters. The PBE0/6-31G* simulations were run once, except for the native nonenzymatic model,

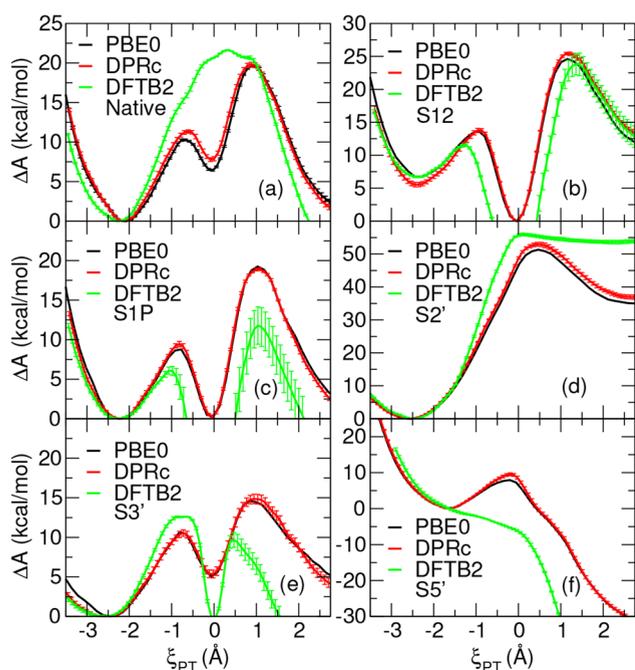


Figure 2. Comparison of DFTB2 QM/MM, PBE0/6-31G* QM/MM, and DFTB2 QM/MM+DPRc FESs generated from classical QM/MM simulations.

which was run 4 times. The relevant collective variable coordinates and free energies describing the free energy surfaces are summarized in Table 1.

The uncorrected DFTB2 model differs significantly from PBE0/6-31G*. When one (S1P) or both (S12) nonbridge oxygens are replaced with sulfurs, the DFTB2 model is likely to result in a nonphysical bond-breakage of the nonbridge P–S bond, resulting in a S^{2-} ion drifting into solution. This can occur in the range $-0.90 \leq \zeta_{PT} \leq 0.80$ Å for S12 and $-1.30 \leq \zeta_{PT} \leq 0.90$ Å for S1P. Regardless, we show these DFTB2 FESs to emphasize the significant degree of correction the DPRc model must undertake to achieve agreement with PBE0/6-31G*. The DFTB2 S5' stationary points are not included in Table 1 because it does not display any local minima or transition states. In comparison, the DFTB2 QM/MM+DPRc free energy surfaces agree much more closely with the PBE0/6-31G* results. Whereas the mean absolute error (MAE) for activation and reaction free energies for DFTB2 with respect to PBE0/6-31G* reference values are 3.4 and 9.4 kcal/mol, respectively, the corresponding MAE values for the DFTB2+DPRc are 1.0 kcal/mol. For DFTB2, the MAE of the reaction coordinate at the minimum is fairly small (0.08 Å), whereas for the transition state, it is significantly larger (0.56 Å). The DFTB2+DPRc model, on the other hand, has corresponding MAE values of 0.02 and 0.03 kcal/mol, respectively. Thus, overall the DFTB2+DPRc model is able to very closely reproduce the PBE0/6-31G* results, whereas DFTB2 exhibits significant differences. Table S4 in the Supporting Information compares the PBE0/6-31G* penta-coordinate intermediate free energies calculated from umbrella sampling to results obtained from geometry optimizations within continuum implicit solvent. Both approaches predict the presence of a penta-coordinate intermediate for all model systems containing 2' and 5' oxygens. Both approaches also

Table 1. Summary of the Free Energy Surfaces Shown in Figure 2^a

system	method	ξ_{PT}^{Min} (Å)	ξ_{PT}^{TS} (Å)	ΔA^\ddagger (kcal mol ⁻¹)	ΔA (kcal mol ⁻¹)
Native	DFTB2	-2.27	0.32	21.65 ± 0.04	-6.3 ± 0.6
	PBE0/ 6-31G*	-2.11	0.94	19.68 ± 0.13	1.5 ± 0.4
S12	DPRc	-2.13	0.88	19.82 ± 0.28	1.1 ± 0.1
	DFTB2	-2.39	1.40	17.15 ± 1.65	2.5 ± 1.7
	PBE0/ 6-31G*	-2.38	1.15	17.94	5.0
S1P	DPRc	-2.40	1.16	19.92 ± 0.20	6.0 ± 0.5
	DFTB2	-2.25	1.05	11.79 ± 2.36	5.2 ± 2.4
	PBE0/ 6-31G*	-2.20	1.03	19.25	2.6
S2'	DPRc	-2.21	1.05	18.92 ± 0.17	1.7 ± 0.5
	DFTB2	-2.47	0.11	55.93 ± 0.43	55.2 ± 0.6
	PBE0/ 6-31G*	-2.58	0.47	51.33	35.6
S3'	DPRc	-2.57	0.47	52.98 ± 0.60	37.5 ± 0.4
	DFTB2	-2.54	-0.68	12.58 ± 0.05	-9.2 ± 0.7
	PBE0/ 6-31G*	-2.47	0.86	14.59	5.2
S5'	DPRc	-2.53	0.95	14.85 ± 0.60	3.6 ± 0.6
	DFTB2
	PBE0/ 6-31G*	-1.62	-0.21	7.87	-33.2
MAE	DPRc	-1.64	-0.18	9.55 ± 0.40	-33.4 ± 0.3
MAE	DFTB2	0.08	0.56	3.37	9.4
MAE	DPRc	0.02	0.03	1.01	1.0

^aListed are the location of the minimum ξ_{PT}^{Min} and rate-controlling transition state ξ_{PT}^{TS} (where $\xi_{PT} = R_{X5'-P} - R_{X2'-P}$), the activation free energy ΔA^\ddagger , and the reaction free energy ΔA . Mean absolute errors (MAEs) for DFTB2 and DFTB2 QM/MM+DPRc models with respect to PBE0/6-31G* reference curves are shown at the bottom.

predict that thio-substitution at these positions results in only one transition state.

Figure 3 compares the *ab initio* FESs to the estimates made from reweighting the DFTB2 umbrella sampling used to prepare Figure 2. The reweighting was performed with the wTP method described in ref 38, which we have implemented in the NDFES program³⁰ distributed within the FE-ToolKit package.⁹¹ A 0.2 kT histogram was used to apply Gaussian smoothing on the density-of-states, and the reaction coordinate bin width was set to 0.05 Å. Figure 3 also shows the reweighting entropy of each bin. The entropy provides a measure for the “flatness” of the reweighting factor’s distribution; it approaches 1 when the reweighting factors are uniform and 0 when a small number of samples contribute a non-negligible weight.⁴¹ Previous applications of the wTP method found that target FESs were well reproduced when the reweighting entropy is greater than 0.6, but the uncertainty in the estimate becomes very large when the entropy decreases below 0.3.³⁹ The reweighting entropies shown in Figure 3 are around 0.1, indicating that there is poor phase space overlap between the *ab initio* and DFTB2 reference potentials. This result motivates our development of the DPRc correction potential.

3.2. QM/MM+DPRc Free Energy Profiles with and without Nuclear Quantum Effects. In order to gain insight into the effect of nuclear quantum effects on the FES, we compared simulations using two closely related flexible water models, SPC/Fw⁶⁷ and q-SPC/Fw,⁶⁸ that were designed for

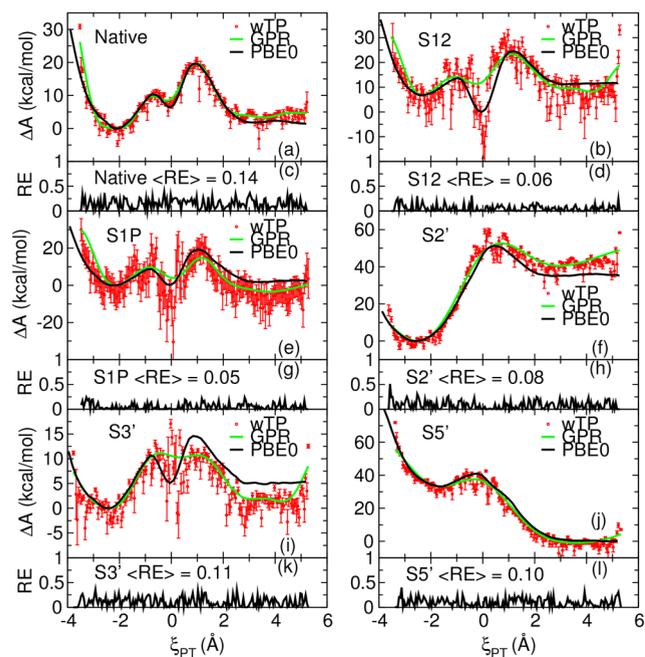


Figure 3. Comparison of the PBE0/6-31G* QM/MM free energy surfaces to those predicted from wTP analysis of the DFTB2 QM/MM umbrella sampling. The green line is a Gaussian Process Regression (GPR) fit to the wTP estimated free energy surface using a radial basis function kernel. The vertical bars are 95% confidence intervals in the wTP estimate obtained from bootstrap error analysis. Parts (c), (d), (g), (h), (k), and (l) are the reweighting entropies (RE) of the Native, S12, S1P, S2', S3', and S5' model reactions, respectively. The “<RE>” values are the average reweighting entropies for each system.

classical MD and PIMD simulations, respectively (see [Methods](#) for details). The classical MD simulations using SPC/Fw were overall very similar to those using the rigid TIP4P/Ew model ([Figure S3](#) of the Supporting Information).

[Figure 4](#) compares DFTB2 QM/MM+DPRc FESs produced by classical MD and PIMD simulations. The classical MD and PIMD curves are the average of 4 and 16 FESs, respectively, and the vertical bars are the standard error of the mean. The classical MD simulations were performed using the SANDER module of AMBER20, whereas the PIMD simulations were performed with i-PI interfaced with SANDER to calculate energies and forces. In our approach for calculating the FESs, we must choose an arbitrary zero of free energy, which we choose as the lowest local free energy minimum. That is, we do not directly obtain a nuclear quantum correction to the free energy as an absolute quantity. Nevertheless, the practical consequences of including nuclear quantum corrections on the FESs can be seen in [Figure 4](#). In all cases, the PIMD FES activation energies are smaller than those calculated from classical MD by 1.48 kcal/mol on average. Specifically, the PIMD activation energies are lowered by 2.10 (Native), 1.48 (S12), 1.32 (S1P), 1.65 (S2'), 1.14 (S3'), and 1.20 (S5') kcal/mol. This reflects, among other things, differences in the vibrational environments of the reactant minimum and transition state, the latter being overall more “loose”,⁵ thus leading to smaller zero point energies in the transition state relative to the reactant state and resulting in lower barriers when nuclear quantum effects are included. As will be seen in the next section, these differences are important in the

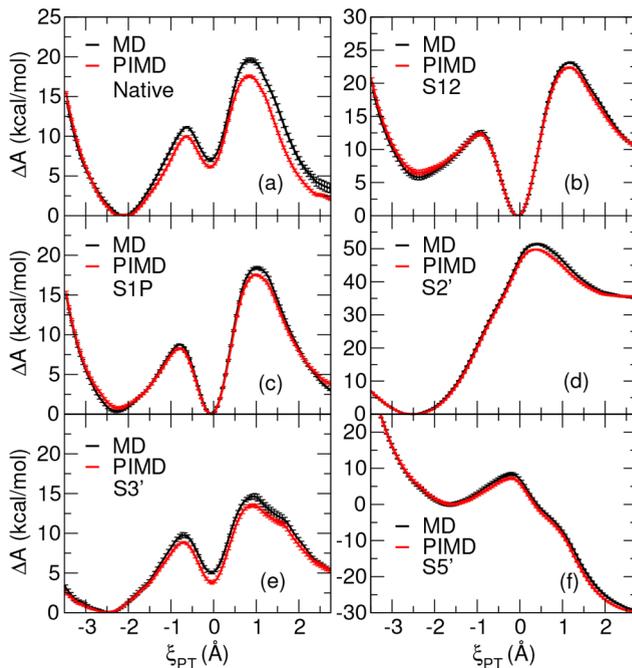


Figure 4. Comparison of DFTB2 QM/MM+DPRc FESs generated from classical MD with SPC/Fw and PIMD with q-SPC/Fw.

calculation of KIE values to aid in the interpretation of experimental measurements and provide insight into the structure of the transition state and the mechanism of the chemical reaction.^{8,92–94}

3.3. KIE Values Calculated with TD-FEP and the Bigeleisen-Mayer Equation. [Table 2](#) compares DFTB2 and DFTB2 QM/MM+DPRc KIE estimates of the native and thio-substituted nonenzymatic models using the TD-FEP method to those computed from *ab initio* PIMD simulations and available experimental values. The DFTB2 and DFTB2 QM/MM+DPRc simulations were performed 4 times, and the results shown in the table are the average of the 4 KIE values and the standard error of the mean. The PBE0/6-31G* QM/MM simulations were performed only once, and the uncertainties are the standard error of the mean upon analyzing 4 nonoverlapping segments of the trajectory. The experimental values are for UpG in alkaline conditions, and the uncertainties are reported to be estimates from numerical fitting errors to the HPLC chromatogram data.⁸ The DFTB2 and DFTB2 QM/MM+DPRc mean absolute errors (MAEs) are relative to the PBE0/6-31G* values. The DFTB2MAE values exclude the S5' KIE values because they do not predict a barrier separating the products from the reactant. Inclusion of the DPRc correction reduces the O2' KIE errors from 1.4% to 0.2%, and the O5' KIE errors are reduced from 2.4% to 0.2%. The DFTB2 QM/MM+DPRc MAE values are similar in size to the uncertainties in the PBE0/6-31G* results. To put the magnitude of the DFTB2MAE values into perspective, the experimental O2' and O5' KIE values of a 2'-O-transphosphorylation reaction catalyzed by RNase A are 0.994 ± 0.002 and 1.014 ± 0.003 , respectively.⁸ These KIE values suggest that the mechanism catalyzed by RNase A is associative ($\text{KIE O2}' \leq 1 \leq \text{KIE O5}'$); however, the KIE values differ by only 0.02. As discussed in previous work,⁹⁵ the calculated phosphoryl transfer reaction KIE values should agree to within 1% of the experimental values to reasonably

Table 2. Comparison of the KIE Values (Eq 11) Computed from PIMD Simulations to Available Experimental Measurements (Ref 8)^a

system	atom	DFTB2	DPRc	PBE0	expt
Native	O2'	0.988 ± 0.000	0.993 ± 0.001	0.993 ± 0.003	0.984 ± 0.003
	O5'	1.046 ± 0.001	1.040 ± 0.001	1.038 ± 0.002	1.034 ± 0.004
S12	O2'	0.984 ± 0.001	0.996 ± 0.001	0.993 ± 0.001	...
	O5'	1.059 ± 0.002	1.026 ± 0.002	1.027 ± 0.002	...
S1P	O2'	0.964 ± 0.001	0.994 ± 0.000	0.995 ± 0.001	...
	O5'	1.063 ± 0.001	1.033 ± 0.001	1.037 ± 0.002	...
S2'	S2'	1.000 ± 0.000	1.001 ± 0.001	1.002 ± 0.001	...
	O5'	1.047 ± 0.003	1.033 ± 0.002	1.037 ± 0.003	...
S3'	O2'	1.019 ± 0.001	0.991 ± 0.001	0.995 ± 0.002	...
	O5'	0.991 ± 0.001	1.034 ± 0.002	1.036 ± 0.003	...
S5'	O2'	No TS	1.029 ± 0.001	1.031 ± 0.003	...
	S5'	No TS	1.001 ± 0.000	1.001 ± 0.000	...
MAE	O2'	>0.014	0.002		
MAE	O5'	>0.024	0.002		

^aThe rows labeled "MAE" provide the mean absolute error relative to the PBE0/6-31G* values.

Table 3. Comparison of DFTB2+ML KIE Values Computed from Path Integral Simulations to Those Computed from the Bigeleisen-Mayer Equation^a

system	atom	TD-FEP		Bigeleisen				
		η^{TST}	η	$\langle \eta^{\text{BM}} \rangle$	N	σ	Min.	Max.
Native	O2'	0.989 ± 0.001	0.993 ± 0.001	0.991 ± 0.001	26	0.005	0.980	0.998
	O5'	1.024 ± 0.000	1.040 ± 0.001	1.038 ± 0.002	26	0.008	1.023	1.051
S12	O2'	0.993 ± 0.001	0.996 ± 0.001	0.996 ± 0.001	19	0.005	0.989	1.010
	O5'	1.013 ± 0.000	1.026 ± 0.002	1.024 ± 0.002	19	0.009	1.004	1.035
S1P	O2'	0.990 ± 0.000	0.994 ± 0.000	0.993 ± 0.001	62	0.005	0.983	1.005
	O5'	1.018 ± 0.000	1.033 ± 0.001	1.032 ± 0.001	62	0.009	1.010	1.051
S2'	S2'	0.997 ± 0.000	1.001 ± 0.001	1.001 ± 0.001	30	0.007	0.996	1.028
	O5'	1.021 ± 0.000	1.033 ± 0.002	1.032 ± 0.002	30	0.014	0.993	1.054
S3'	O2'	0.988 ± 0.001	0.991 ± 0.001	0.994 ± 0.001	19	0.005	0.986	1.005
	O5'	1.021 ± 0.001	1.034 ± 0.002	1.037 ± 0.002	19	0.010	1.016	1.050
S5'	O2'	1.008 ± 0.001	1.029 ± 0.001	1.026 ± 0.001	111	0.012	0.988	1.051
	S5'	1.000 ± 0.000	1.001 ± 0.000	1.003 ± 0.000	111	0.002	0.998	1.024

^aThe columns N , σ , "Min.", and "Max." are the number of KIE values, the standard deviation of the KIE distribution, and minimum and maximum KIE values, respectively.

discern between "large inverse" (less than 0.97), "inverse" (0.97–0.99), "near unity" (0.99–1.01), "normal" (1.01–1.03), or "large normal" (greater than 1.03) isotope effects.⁵

Table 3 compares DFTB2 QM/MM+DPRc KIE values computed from the Bigeleisen-Mayer equation applied to an ensemble of stationary structures and the TD-FEP KIE values calculated from PIMD simulations. As we have described, the Bigeleisen-Mayer equation is applied many times corresponding to a collection of stationary structures obtained by geometry optimizing trajectory frames either for a minimum or a transition state. The table lists the mean KIE value, the standard error of the mean, the number of KIE values (the number transition state structures that we found), the standard deviation of the KIE values, and the minimum and maximum value in the observed KIE distribution.

The average O5' (or S5') KIE values obtained from the Bigeleisen-Mayer equation $\langle \eta^{\text{BM}} \rangle$ are in good agreement with the estimates made from PIMD/TD-FEP simulations using eq 11. The two methods agree to within 0.003 for all systems. For comparison, Table 3 also shows the η^{TST} KIE values computed from PIMD/TD-FEP (These values assume the ratio of imaginary frequencies is unity.). The η^{TST} values are often 0.01 smaller than those obtained from the Bigeleisen-Mayer

equation. Table 3 emphasizes the importance of performing multiple transition state searches when attempting to apply the Bigeleisen-Mayer equation to explicitly solvated condensed phase systems. The average difference between the maximum and minimum η^{BM} estimates (the "Max." and "Min." columns in Table 3) is 0.03. Therefore, caution should be exercised when using the harmonic approximation in explicit solvent if only 1 structure is used.

3.4. QM/MM+DPRc Isotope Effects along the Reaction Coordinate. Enzymatic transition states are often interpreted using a "kinetic isotope effect approach" that combines experimental measurements with quantum mechanical calculations.⁷ In this approach, model transition states are systematically constructed until the calculated KIE value matches the experimental measurement. The corresponding transition state model is then used to infer properties of the enzymatic transition state⁷ such as structure, bonding, and electrostatics. This information can be used to gain a predictive understanding of the catalytic mechanism or act as a guide to design drugs based on transition state analogues.⁷ Ideally, a complete model system of the full enzyme environment can be constructed, and the catalytic reaction path(s) and rate-controlling transition state(s) can be determined. This can be a

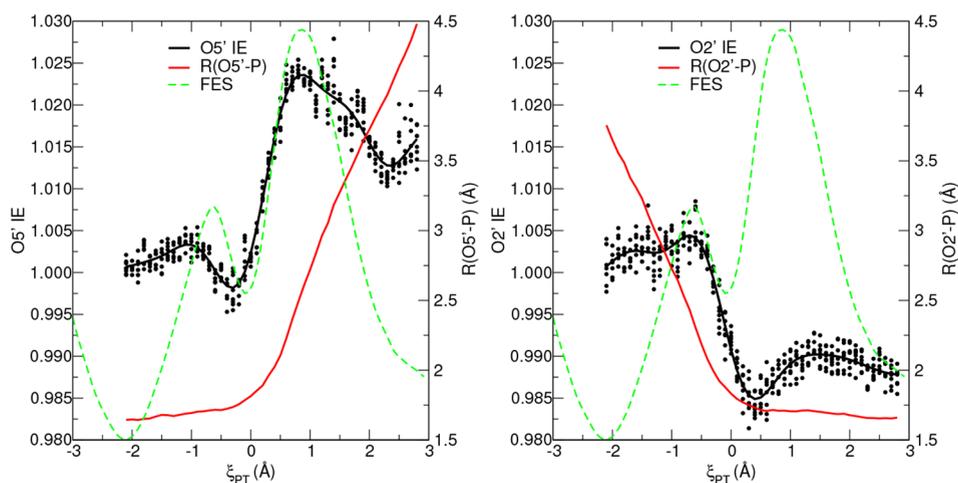


Figure 5. O5' and O2' TD-FEP estimate of η^{TST} IE values for the native reaction as a function of the collective variable defining the position of the effective transition state. The solid black line is a fit through the IE values. The red line is the average value of the covalent bond length. The green line is an overlay of the FES.

complex and computationally intensive endeavor that does not always lead to quantitative agreement with the experimental KIE values. In the event the predicted KIE values do not agree with measurements, it is useful to analyze trends in the isotope effects on the free energy at different positions along the reaction path to gain insight into the likely structure and bonding of the true transition state. With this knowledge, one can then use reverse engineering to explore what changes in the enzyme environment might lead to the predicted altered transition state. The use of KIE approaches to enzyme transition states has been reviewed elsewhere;^{7,96–100} here, we extend some of the ideas to the application of PIMD in the context of QM/MM+DPRc free energy profiles to examine isotope effects (IEs) along the reaction coordinate.

One can use PIMD TD-FEP simulations to examine the isotope effect on the free energy surface as a function of the reaction coordinate. Whereas the KIE values discussed in section 3.3 are measurable quantities, the isotope effects (IEs) along the reaction coordinate are generally not; rather their purpose is for the analysis and prediction described above. Figure 5 illustrates the ^{18}O primary (O2' and O5') η^{TST} IE values of the native nonenzymatic reaction as a function of ξ_{PT} . The IE values were generated by restraining the centroid collective variable position with a harmonic restraint and treating the sampled state as the effective transition state ensemble when computing η^{TST} , whereas the reactant state ensemble was simulated at the FES minimum. Each of the 4 ML parameter sets was simulated twice with different random number seeds, yielding 8 IE estimates per ξ_{PT} value. The dots in Figure 5 are the raw η^{TST} IE value estimates, and the black line is a Gaussian process regression fit through the scattered data. The red line is the ensemble average of the centroid ξ_{PT} value. The dashed green line is an overlay of the FES to aid the reader in orienting the relationship of the ξ_{PT} values to the location of the reaction minimum and transition state. It is noteworthy that the maximum O2' and O5' IE values are normal ($\eta^{\text{TST}} > 1$) and occur at the corresponding early and late transition states, respectively, where the bonding environments are the most “loose” (lower frequency local vibrational modes), as expected for associative phosphoryl transfer mechanisms.^{5,101}

Table 4. Computational Cost (Core Days) to Evaluate 4 Estimates of the Free Energy Surface for Each of the 6 Model Systems^a

method	param	sampling	analysis	net
PBE0/6-31G*	...	34,883	...	34,883
wTP	...	60	698	758
DFTB2 QM/MM+DPRc	323	139	...	462

^aAll timing estimates are based on single core performance measurements using a 2.10 GHz Intel Xeon Gold 6230 CPU with 28MB of cache. The umbrella sampling consists of 25 ps using a 1 fs timestep, and the wTP analysis of the PBE0/6-31G* energies is performed once every 50 steps.

3.5. Performance. Table 4 compares the computational cost of performing Born-Oppenheimer umbrella sampling with PBE0/6-31G* QM/MM, DFTB2 QM/MM+DPRc, and the wTP analysis of the DFTB2 QM/MM simulations. The net computational cost is decomposed into parametrization (param), sampling, and analysis. The DFTB2 QM/MM+DPRc parametrization cost is the overhead associated with performing the active learning procedure to yield 4 DPRc parameter sets. Table S2 in the Supporting Information further decomposes this cost into training, exploration, and labeling stages. In summary, 8% of the parametrization cost is spent performing the training. Only 16% of the effort is dedicated to labeling. The majority of the parametrization is associated with the exploration stage (81%). The “sampling” column is the cost of generating 4 estimates of the 6 free energy surfaces from QM/MM umbrella sampling. The wTP method performs the sampling with DFTB2 QM/MM, which is a factor 2.3 faster than DFTB2 QM/MM+DPRc on a single CPU core (909 ps/day versus 390 ps/day). This is in close agreement with timings reported in related works, which found that ML corrections reduce the semiempirical QM/MM performance by a factor of 2-to-4, depending on the implementation and depth of the neural network.^{18,19} Our implementation of the DPRc correction also supports GPU acceleration via the tensorflow library¹⁰² and several custom operators.^{60,61,103,104} The DFTB2 QM/MM+DPRc performance is only 26% slower than DFTB2 QM/MM when the correction is evaluated on a V100 NVIDIA GPU. In comparison to *ab initio* sampling, the

DFTB2 QM/MM+DPRc method is faster by a factor of 251. Table S3 in the Supporting Information provides further comparisons of computational cost. The analysis cost is only relevant to the wTP method, which reweights the DFTB2 QM/MM trajectories. It is important to note that the sampling and analysis costs both depend on the amount of production simulation, whereas the DPRc training cost is a constant overhead. For example, in addition to performing classical umbrella sampling, we also applied the DFTB2 QM/MM+DPRc method to PIMD umbrella sampling and TD-FEP calculations. If one wants to make a single estimate of an FES, then trajectory reweighting is certainly a cost-effective approach; however, if one wants to perform multiple, independent simulations to estimate uncertainties or perform PIMD, then the savings gained from the DFTB2 QM/MM+DPRc sampling quickly outweighs the parametrization overhead.

4. CONCLUSION

We developed a new DFTB2 QM/MM+DPRc model (DFTB2 with a DPRc machine learning potential correction) and parametrized it to reproduce the PBE0/6-31G* QM/MM energies and forces of native and thio-substituted non-enzymatic models of RNA 2'-O-transphosphorylation. The DPRc potential applies corrections to the QM interactions and the QM/MM interactions within 6 Å of the QM region. The parametrized model was found to significantly improve the prediction of free energy surfaces computed from umbrella sampling in explicit solvent. The DFTB2 QM/MM+DPRc simulations were found to be 2 orders of magnitude faster than the *ab initio* QM/MM simulations and 2.3 times slower than the uncorrected DFTB2 QM/MM simulations.

We interfaced the SANDER with the i-PI software package to perform path integral molecular dynamics umbrella sampling. The resulting free energy surfaces were compared to those obtained from classical QM/MM molecular dynamics. It was found that nuclear quantum effects lower the activation free energy of these reactions by 1.5 kcal/mol, on average.

The new interface between i-PI and SANDER allowed us to compute primary ¹⁸O KIE values using the TD-FEP method. Although there were minor discrepancies between the DFTB2 QM/MM+DPRc and experimental KIE values, the parametrized model was found to reproduce the PBE0/6-31G* KIE values to within the uncertainties of the calculations. The agreement with experiment can likely be improved by training the model to better *ab initio* reference data. We compared the KIE values using two methods: TD-FEP and the Bigeleisen-Mayer equation. The two methods agreed to within 0.003 for all systems; however, the individual estimates made using the Bigeleisen-Mayer equation were found to span a range of 0.03. The DFTB2 QM/MM+DPRc method was used to map the TD-FEP isotope effect values along the phosphoryl transfer reaction coordinate, and it was found that the O2' and OS' isotope effects were largest at the early and late transition states, respectively. Taken together, the QM/MM+DPRc model and i-PI PIMD simulations provide a powerful new tool to study mechanisms and pathways of phosphoryl transfer reactions.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00151>.

Information regarding training of machine learning potential and summary of computational cost and comparison of free energy surfaces produced from different molecular dynamics software and various water models (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Darrin M. York – Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; orcid.org/0000-0002-9193-7055; Email: Darrin.York@rutgers.edu

Authors

Timothy J. Giese – Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; orcid.org/0000-0002-0653-9168

Jinzhe Zeng – Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; orcid.org/0000-0002-1515-8172

Sölen Ekesan – Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States; orcid.org/0000-0002-5598-5754

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.2c00151>

Author Contributions

T.J.G. and J.Z. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful for financial support provided by the National Institutes of Health (Nos. GM062248 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-1548562¹⁰⁵ (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) Perreault, D. M.; Anslyn, E. V. Unifying the Current Data on the Mechanism of Cleavage-Transesterification of RNA. *Angew. Chem., Int. Ed.* 1997, 36, 432–450.

- (2) Oivanen, M.; Kuusela, S.; Lönnberg, H. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brønsted Acids and Bases. *Chem. Rev.* **1998**, *98*, 961–990.
- (3) Lassila, J. K.; Zalatan, J. G.; Herschlag, D. Biological phosphoryl-transfer reactions: Understanding mechanism and catalysis. *Annu. Rev. Biochem.* **2011**, *80*, 669–702.
- (4) Kim, C. M.; Smolke, C. D. Biomedical applications of RNA-based devices. *Curr. Opin. Biomed. Eng.* **2017**, *4*, 106–115.
- (5) Hengge, A. C. Isotope effects in the study of phosphoryl and sulfuryl transfer reactions. *Acc. Chem. Res.* **2002**, *35*, 105–112.
- (6) Harris, M. E.; Piccirilli, J. A.; York, D. M. Enzyme transition states from theory and experiment. *Biochim. Biophys. Acta* **2015**, *1854*, 1727–1728.
- (7) Schramm, V. L. Enzymatic Transition States and Drug Design. *Chem. Rev.* **2018**, *118*, 11194–11258.
- (8) Gu, H.; Zhang, S.; Wong, K.-Y.; Radak, B. K.; Dissanayake, T.; Kellerman, D. L.; Dai, Q.; Miyagi, M.; Anderson, V. E.; York, D. M.; Piccirilli, J. A.; Harris, M. E. Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-O-transphosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13002–13007.
- (9) 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.
- (10) Chen, H.; Piccirilli, J. A.; Harris, M. E.; York, D. M. Effect of Zn²⁺ binding and enzyme active site on the transition state for RNA 2'-O-transphosphorylation interpreted through kinetic isotope effects. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854*, 1795–1800.
- (11) 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.
- (12) Pan, X.; Nam, K.; Epifanovsky, E.; Simmonett, A. C.; Rosta, E.; Shao, Y. A simplified charge projection scheme for long-range electrostatics in ab initio QM/MM calculations. *J. Chem. Phys.* **2021**, *154*, 024115.
- (13) Gaines, C. S.; Giese, T. J.; York, D. M. Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology. *ACS Catal.* **2019**, *9*, 5803–5815.
- (14) Ganguly, A.; Weissman, B. P.; Giese, T. J.; Li, N.-S.; Hoshika, S.; Saieesh, R.; Benner, S. A.; Piccirilli, J. A.; York, D. M. Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme. *Nat. Chem.* **2020**, *12*, 193–201.
- (15) Meuwly, M. Machine Learning for Chemical Reactions. *Chem. Rev.* **2021**, *121*, 10218–10239.
- (16) Galib, M.; Limmer, D. T. Reactive uptake of N₂O₅ by atmospheric aerosol is dominated by interfacial processes. *Science* **2021**, *371*, 921–925.
- (17) Zeng, J.; Cao, L.; Xu, M.; Zhu, T.; Zhang, J. Z. H. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat. Commun.* **2020**, *11*, 5713.
- (18) 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.
- (19) Gómez-Flores, C. L.; Maag, D.; Kansari, M.; Vuong, V.-Q.; Irle, S.; Gräter, F.; Kubař, T.; Elstner, M. Accurate Free Energies for Complex Condensed-Phase Reactions Using an Artificial Neural Network Corrected DFTB/MM Methodology. *J. Chem. Theory Comput.* **2022**, *18*, 1213–1226.
- (20) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today* **2022**, *387*, 143–149.
- (21) Zeng, J.; Zhang, L.; Wang, H.; Zhu, T. Exploring the Chemical Space of Linear Alkane Pyrolysis via Deep Potential GENERATOR. *Energy Fuels* **2021**, *35*, 762–769.
- (22) 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.
- (23) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569–585.
- (24) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Status Solidi. B* **2000**, *217*, 357–376.
- (25) 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.
- (26) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, K.; Krasny, R.; Kurtzman, T.; Lee, T.; Le-Grand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, R. M.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. *AMBER 20*; University of California, San Francisco: San Francisco, CA, 2020.
- (27) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Comput. Phys. Commun.* **2014**, *185*, 1019–1026.
- (28) 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.
- (29) 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.
- (30) 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.
- (31) Herschlag, D.; Piccirilli, J. A.; Cech, T. R. Ribozyme-Catalyzed and Nonenzymatic Reactions of Phosphate Effects upon Substitution of Sulfur for a Nonbridging Phosphoryl Diesters: Rate Oxygen Atom. *Biochemistry* **1991**, *30*, 4844–4854.
- (32) Ward, W. L.; Plakos, K.; DeRose, V. J. Nucleic acid catalysis: metals, nucleobases, and other cofactors. *Chem. Rev.* **2014**, *114*, 4318–4342.
- (33) 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.
- (34) Trippett, S. APICOPHILICITY AND RING-STRAIN IN FIVE-COORDINATE PHOSPHORANES. Phosphorus and Sulfur and the Related Elements. *J. Am. Chem. Soc.* **1976**, *1*, 89–98.
- (35) De Keijzer, A. E.; Koole, L. H.; Buck, H. M. Pseudorotation in pentacoordinated phosphorus compounds. The influence of the conformational transmission effect on the barriers to pseudorotation in cyclic alkoxyphosphoranes. *J. Am. Chem. Soc.* **1988**, *110*, 5995–6001.
- (36) López, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. Pseudorotation of Natural and Chemically Modified Biological Phosphoranes: Implications for RNA Catalysis. *Chem. Phys. Chem.* **2004**, *5*, 1045–1049.
- (37) López, C. S.; Faza, O. N.; R. de Lera, A.; York, D. M. Pseudorotation Barriers of Biological Oxyphosphoranes: A Challenge for Simulations of Ribozyme Catalysis. *Chem.—Eur. J.* **2005**, *11*, 2081–2093.

- (38) 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.
- (39) 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.
- (40) 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.
- (41) 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.
- (42) 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.
- (43) 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.
- (44) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. Description of phosphate hydrolysis reactions with the self-consistent-charge density-functional-tight-binding (SCC-DFTB) theory. I. Parameterization. *J. Chem. Theory Comput.* **2008**, *4*, 2067–2084.
- (45) Bevilacqua, P. C.; Harris, M. E.; Piccirilli, J. A.; Gaines, C.; Ganguly, A.; Kostenbader, K.; Ekesan, S.; York, D. M. An Ontology for Facilitating Discussion of Catalytic Strategies of RNA-Cleaving Enzymes. *ACS Chem. Biol.* **2019**, *14*, 1068–1076.
- (46) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (47) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (48) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform* **2011**, *3*, 33.
- (49) Horn, H. W.; Swope, W. C.; Pitera, 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.
- (50) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.
- (51) 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.
- (52) Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564–1582.
- (53) 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.
- (54) 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.
- (55) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Hsing, L.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (56) Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. Ribozyme speed limits. *RNA* **2003**, *9*, 907–918.
- (57) Zhang, L.; Han, J.; Wang, H.; Saidi, W.; Car, R.; E, W. End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. Curran Associates, Inc.: 2018; pp 4436–4446.
- (58) 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.
- (59) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Han, W.; E, W. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* **2020**, *253*, 107206.
- (60) Lu, D.; Jiang, W.; Chen, Y.; Zhang, L.; Jia, W.; Wang, H.; Chen, M. DP Train, then DP Compress: Model Compression in Deep Potential Molecular Dynamics. 2021, *arXiv:2107.02103*. <https://arxiv.org/abs/2107.02103> (accessed 2022-06-14).
- (61) Guo, Z.; Lu, D.; Yan, Y.; Hu, S.; Liu, R.; Tan, G.; Sun, N.; Jiang, W.; Liu, L.; Chen, Y.; Zhang, L.; Chen, M.; Wang, H.; Jia, W. Extending the limit of molecular dynamics with ab initio accuracy to 10 billion atoms. In *Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*; Association for Computing Machinery: New York, NY, USA, 2022; pp 205–218, DOI: 10.1145/3503221.3508425.
- (62) Masters, D.; Luschi, C. Revisiting small batch training for deep neural networks. 2018, *arXiv:1804.07612*. <https://arxiv.org/abs/1804.07612> (accessed 2022-06-14).
- (63) Ceriotti, M.; Manolopoulos, D. E. Efficient first-principles calculation of the quantum kinetic energy and momentum distribution of nuclei. *Phys. Rev. Lett.* **2012**, *109*, 100604.
- (64) Ceriotti, M.; Manolopoulos, D. E.; Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *J. Chem. Phys.* **2011**, *134*, 084104.
- (65) Ceriotti, M.; Bussi, G.; Parrinello, M. Colored-Noise Thermostat à la Carte. *J. Chem. Theory Comput.* **2010**, *6*, 1170–1180.
- (66) Ceriotti, M.; Bussi, G.; Parrinello, M. Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations. *Phys. Rev. Lett.* **2009**, *102*, 020601.
- (67) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006**, *124*, 024503.
- (68) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, I. I. L.; Voth, T. E. G. A. An accurate and simple quantum model for liquid water. *J. Chem. Phys.* **2006**, *125*, 184507.
- (69) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. PLUMED: a portable plugin for free energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (70) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (71) Gao, J.; Wong, K.-Y.; Major, D. T. Combined QM/MM and path integral simulations of kinetic isotope effects in the proton transfer reaction between nitroethane and acetate ion in water. *J. Comput. Chem.* **2008**, *29*, 514–522.
- (72) Major, D. T.; Gao, J. An integrated path integral and free-energy perturbation-umbrella sampling method for computing kinetic isotope effects of chemical reactions in solution and in enzymes. *J. Chem. Theory Comput.* **2007**, *3*, 949–960.
- (73) Voth, G. A.; Chandler, D.; Miller, W. H. Time correlation function and path integral analysis of quantum rate constants. *J. Phys. Chem.* **1989**, *93*, 7009–7015.
- (74) Kannath, S.; Adamczyk, P.; Ferro-Costas, D.; Fernández-Ramos, A.; Major, D. T.; Dybala-Defratyka, A. Role of Microsolvation and Quantum Effects in the Accurate Prediction of Kinetic Isotope Effects: The Case of Hydrogen Atom Abstraction in Ethanol by Atomic Hydrogen in Aqueous Solution. *J. Chem. Theory Comput.* **2020**, *16*, 847–859.

- (75) Truhlar, D. G.; Hase, W. L.; Hynes, J. T. Current status of transition-state theory. *J. Phys. Chem.* **1983**, *87*, 2664–2682.
- (76) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.
- (77) Bigeleisen, J.; Mayer, M. G. Calculation of Equilibrium Constants for Isotopic Exchange Reactions. *J. Chem. Phys.* **1947**, *15*, 261–267.
- (78) Bigeleisen, J.; Wolfsberg, M. Theoretical and experimental Aspects of Isotope Effects in Chemical Kinetics. *Advances in Chemical Physics*; 1958; Vol. 1, pp 15–76, DOI: 10.1002/9780470143476.ch2.
- (79) Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284–304.
- (80) Landauer, R.; Swanson, J. A. Frequency Factors in the Thermally Activated Process. *Phys. Rev.* **1961**, *121*, 1668–1674.
- (81) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
- (82) Berne, B. J.; Borkovec, M. Classical theories of reaction dynamics: Transition state theory, spatial diffusion controlled reactions, and the energy diffusion limit. *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 2717–2723.
- (83) Ceriotti, M.; Markland, T. E. Efficient methods and practical guidelines for simulating isotope effects. *J. Chem. Phys.* **2013**, *138*, 014112.
- (84) Gao, J. Chapter Fourteen - Enzymatic Kinetic Isotope Effects from Path-Integral Free Energy Perturbation Theory. *Methods Enzymol.* **2016**, *577*, 359–88.
- (85) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An open-source geometry optimizer for atomistic simulations. *J. Phys. Chem. A* **2009**, *113*, 11856–11865.
- (86) Yu, T.; Zheng, J.; Truhlar, D. G. Multi-structural variational transition state theory. Kinetics of the 1,4-hydrogen shift isomerization of the pentyl radical with torsional anharmonicity. *Chem. Sci.* **2011**, *2*, 2199.
- (87) Zheng, J.; Truhlar, D. G. Multi-path variational transition state theory for chemical reaction rates of complex polyatomic species: ethanol + OH reactions. *Faraday Discuss.* **2012**, *157*, 59–88.
- (88) Bao, J. L.; Truhlar, D. G. Variational transition state theory: theoretical framework and recent developments. *Chem. Soc. Rev.* **2017**, *46*, 7548–7596.
- (89) Yu, T.; Zheng, J.; Truhlar, D. G. Multipath Variational Transition State Theory: Rate Constant of the 1,4-Hydrogen Shift Isomerization of the 2-Cyclohexylethyl Radical. *J. Phys. Chem. A* **2012**, *116*, 297–308.
- (90) Simón-Carballido, L.; Alves, T. V.; Dybala-Defratyka, A.; Fernández-Ramos, A. Kinetic Isotope Effects in Multipath VTST: Application to a Hydrogen Abstraction Reaction. *J. Phys. Chem. B* **2016**, *120*, 1911–8.
- (91) Giese, T. J.; York, D. M. FE-ToolKit: The free energy analysis toolkit. <https://gitlab.com/RutgersLBSR/fe-toolkit> (accessed 2022-06-14).
- (92) Kellerman, D. L.; York, D. M.; Piccirilli, J. A.; Harris, M. E. Altered (transition) states: mechanisms of solution and enzyme catalyzed RNA 2'-O-transphosphorylation. *Curr. Opin. Chem. Biol.* **2014**, *21*, 96–102.
- (93) Harris, M. E.; Piccirilli, J. A.; York, D. M. Integration of kinetic isotope effect analyses to elucidate ribonuclease mechanism. *Biochim. Biophys. Acta* **2015**, *1854*, 1801–1808.
- (94) Weissman, B. P.; Li, N.-S.; York, D. M.; Harris, M.; Piccirilli, J. A. Heavy atom labeled nucleotides for measurement of kinetic isotope effects. *Biochim. Biophys. Acta* **2015**, *1854*, 1737–1745.
- (95) Wong, K.-Y.; Gu, H.; Zhang, S.; Piccirilli, J. A.; Harris, M. E.; York, D. M. Characterization of the reaction path and transition states for RNA transphosphorylation models from theory and experiment. *Angew. Chem., Int. Ed.* **2012**, *51*, 647–651.
- (96) Singh, P.; Islam, Z.; Kohen, A. Examinations of the Chemical Step in Enzyme Catalysis. *Methods in Enzymology*; 2016; Vol. 577, p 287.
- (97) Schramm, V. L. Enzymatic transition-state analysis and transition-state analogs. *Methods in Enzymology*; 1999; Vol. 308, pp 301–355.
- (98) Berti, P. J. Determining transition states from kinetic isotope effects. *Methods in Enzymology*; 1999; Vol. 308, pp 355–397.
- (99) Schramm, V. L. Enzymatic Transition States and Transition State Analog Design. *Annu. Rev. Biochem.* **1998**, *67*, 693–720.
- (100) Parkin, D. *Methods for the Determination of Competitive and Noncompetitive Kinetic Isotope Effects*; CRC Press: 1991; p 269.
- (101) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. Transesterification Thio Effects of Phosphate Diesters: Free Energy Barriers and Kinetic and Equilibrium Isotope Effects from Density-Functional Theory. *Biochemistry* **2006**, *45*, 10043–10053.
- (102) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*; 2015. <https://www.tensorflow.org/> (accessed 2022-06-14). Software available from [tensorflow.org](https://www.tensorflow.org/).
- (103) Lu, D.; Wang, H.; Chen, M.; Lin, L.; Car, R.; E, W.; Jia, W.; Zhang, L. 86 PFLOPS Deep Potential Molecular Dynamics simulation of 100 million atoms with ab initio accuracy. *Comput. Phys. Commun.* **2021**, *259*, 107624.
- (104) Jia, W.; Wang, H.; Chen, M.; Lu, D.; Lin, L.; Car, R.; E, W.; Zhang, L. Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '20*; IEEE Press: 2020; p 1.
- (105) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gathier, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62–74.