

Modeling Enzyme Reaction and Mutation by Direct Machine Learning/Molecular Mechanics Simulations

Published as part of *Journal of Chemical Theory and Computation special issue "Developments of Theoretical and Computational Chemistry Methods in Asia"*.

Xinhu Sha, Zhuo Chen, Daiqian Xie, and Yanzi Zhou*



Cite This: <https://doi.org/10.1021/acs.jctc.5c00149>



Read Online

ACCESS |



Metrics & More

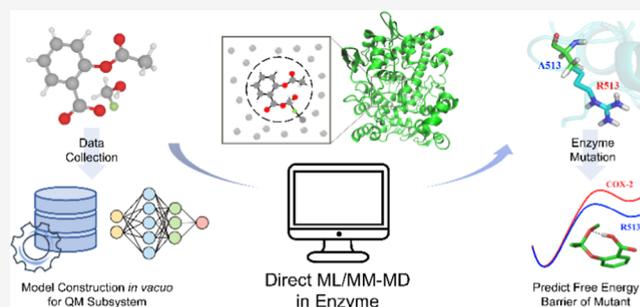


Article Recommendations



Supporting Information

ABSTRACT: Accurately modeling enzyme reactions through direct machine learning/molecular mechanics simulations remains challenging in describing the electrostatic coupling between the QM and MM subsystems. In this work, we proposed a reweighting ME (mechanic embedding) REANN (recursively embedded atom neural network) method that trains the potential and point charges of the QM subsystem in vacuo. The charge equilibration approach has been encoded into REANN to ensure conservation of the total charge of the QM subsystem. Electrostatic coupling is measured by point charges, and the polarization of the MM subsystem on the coupling can be corrected by thermodynamic perturbation after molecular dynamics simulations. We first constructed the REANN surfaces of potential energy and charges for the acylation of cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2) by aspirin. These surfaces allowed us to reproduce the free energy curves of B3LYP/MM-MD with a chemical accuracy. Subsequently, they were successfully applied to R513A of COX-2, reproducing the free energy barrier simulated by B3LYP/MM MD with a difference of less than 0.5 kcal mol⁻¹ and a speedup of 80-fold, revealing our method can predict the activity of mutants accurately and rapidly. This method is expected to be applied in virtual screening in the future.



1. INTRODUCTION

Mutant screening has always been a bottleneck in protein engineering. One effective strategy to address this issue is to transfer screening work to computers for virtual screening, which can effectively reduce the size of the library and simplify screening.^{1–4} Combining ab initio (ai) quantum mechanics/molecular mechanical (QM/MM) potential^{5–9} with molecular dynamics (MD) simulation is one of the most advanced methods to predict the free energy barrier in enzyme design,^{10–13} which considers both the electrostatic interactions within the enzyme environment and the enzyme dynamics.¹⁴ A high-level quantum chemical method is indispensable to reproduce small energy differences,^{11,15} but it is very expensive since each step of MD requires electronic structure calculations. Therefore, the practical application of ai QM/MM-MD in virtual screening is still in its early stages and requires further development before it can be widely adopted.

Many efforts have been made to develop ab initio QM/MM-MD methods aimed at reducing the need for direct, high-level QM samplings while still accurately estimating free energy changes along the reaction path. In the dual-Hamiltonian approach initially proposed by Gao,^{13,16} a higher accuracy can be achieved with an affordable computational cost by

calibrating the low-level QM model with additional functions of energy^{17–19} or reoptimizing the parameters in the low-level Hamiltonian against the high-level Hamiltonian.^{20–23} Because only a few configurations are calculated with high-level QM models, these methods are very efficient. However, the results might fluctuate, since the configurations used in ab initio calculations are too limited to describe the MM environment sufficiently. In another dual-Hamiltonian approach, samplings are performed at a low level of theory, and thermodynamic perturbation (TP) is used to obtain the high-level QM free energy.^{15,24,25} The overlap between sampling spaces at two levels determines the success of the TP correction. Nevertheless, the overlap might be small for large QM systems and lead to the failure of this reweighting method, which can be partially improved by force-matching.^{26,27}

Received: January 26, 2025

Revised: April 11, 2025

Accepted: April 14, 2025

In the past few decades, the development of machine learning (ML) has facilitated broader applications of expensive QM methods. A good training data set and machine learning algorithm can help to accurately predict electronic structural information such as energy, force, and charge of a system with a computational speed comparable to molecular mechanics (MM).^{28–32} Incorporating the machine learning^{27,31} or Δ -machine learning^{15,19,33–35} techniques with QM potential is another major family of approaches to accelerate the QM/MM-MD simulations, in which the QM/MM-MD simulations are performed on a preconstructed ML-QM/MM potential energy surface (PES). In the QM/MM method, the electrostatic embedding (EE) scheme is recommended for handling QM/MM electrostatic interactions, where the QM subsystem is polarized by the MM atoms and the MM charges are treated as one-electron terms in the effective Hamiltonian.^{10,36,37} But since the electrostatic interactions also depend on the MM environment, it is challenging to construct ML-QM/MM PES in the EE scheme. One solution is to include certain QM properties dependent on the MM environment as additional descriptors in the neural network (NN) when building the PES, for example, the QM partial charges calculated using a low-level QM model or MM electrostatic potentials at the QM grids.^{15,19,27,31,33} Alternatively, only the nearby MM atoms are considered as the input variables of the ML model.^{34,35} Nevertheless, besides the coordinates of QM atoms, there are still dozens to hundreds of additional input nodes, potentially leading to convergence and affordability problems. Moreover, the PES constructed in this way depends on the MM environment and always requires prior semiempirical QM/MM-MD simulations to calculate the QM properties or collect configurations for NN training. Very recently, the Sugita group developed a well-defined physical model to handle the electrostatic interactions between the QM and MM regions using the high-order terms from the Taylor expansion of the electrostatic operator and an equivariant model. This model achieves high accuracy and transferability among various environment media but is limited to the spherical boundary condition at present.³⁸

In our group's previous work,³⁹ we developed the reweighting ME EANN (embedded atom neural network)⁴⁰ method combining ML-QM/MM potential in the mechanic embedding (ME) scheme with weighted thermodynamic perturbation (wTP) correction to the EE scheme. This method was utilized to solution reactions and reproduced the free energy barriers of B3LYP/MM-MD simulations within 0.5 kcal mol⁻¹. Compared with B3LYP/MM-MD simulations, our method can enhance computational speed by 30–60 folds. Because the PES is constructed in the gas phase, it is independent of the environment and can be applied to various MM environments, which has been demonstrated by successful simulations of the S_N2 reaction of CH₃Cl + Cl⁻ → Cl⁻ + CH₃Cl' in water and methanol employing the same PES.

To our knowledge, only a handful of papers have employed direct machine learning potentials (MLPs) to study enzyme reactions,^{27,38} and many used Δ ML to calibrate low-level QM methods.^{35,41,42} General methods to accelerate QM/MM-MD simulations are still scarce for enzyme reactions, and it remains challenging to predict the free energy barriers with "chemical accuracy" by MLPs. In this work, we have developed the reweighting ME REANN method, in which the pseudobond approach^{43–45} is adopted to address the QM/MM boundary problem and the charge equilibration (QEq) approach⁴⁶ has

been encoded into recursively embedded atom neural network (REANN)^{47,48} to ensure conservation of the total charge of the QM subsystem. We apply it to the acylation reaction of cyclooxygenase (COX) by aspirin. Initially, we construct ML PES and charge surfaces for the QM subsystem using REANN, and subsequently, perform REANN/MM-MD on the surfaces to obtain the free energy curves of the acylation reactions for two subtypes of COX: COX-1 and COX-2. Our method can speed up the QM/MM MD simulations by 80-fold and reproduce the free energy barriers of B3LYP/MM MD with an accuracy of 0.5 kcal mol⁻¹. In addition, we successfully apply this PES to the R513A mutant of COX-2, demonstrating its robustness and sufficiency. Future work will involve further developing and applying this approach to virtual screening.

2. THEORY AND METHODS

2.1. REANN Approach. The energy and forces in the QM subsystem are calculated by a novel package with the message-passing framework based on PyTorch, namely REANN.^{47,48} The building-up of REANN starts from a series of Gaussian-type orbitals (GTOs) centered at each atom like EANN

$$\varphi_{l_x, l_y, l_z}^m(\hat{r}_{ij}) = (x_{ij})^{l_x} (y_{ij})^{l_y} (z_{ij})^{l_z} \exp[-\alpha_m (r_{ij} - r_m)^2] \quad (1)$$

where \hat{r}_{ij} represents the vector from the central atom i to its neighboring atom j and x_{ij} , y_{ij} , and z_{ij} are its Cartesian components. $l_x + l_y + l_z = L$ specifies the orbital angular momentum L . α_m and r_m are parameters used to determine the radial distribution of the m th atomic orbit. Then, the embedded atom density (EAD) vector ρ_i of the central atom i is calculated by

$$\rho_i = \sum_{l=0}^L \sum_{l_x, l_y, l_z}^{l_x+l_y+l_z=L} \frac{L!}{l_x! l_y! l_z!} \left[\sum_{j \neq i}^N c_j \sum_{m=1}^{N_\varphi} d_m \varphi_{l_x, l_y, l_z}^m(\hat{r}_{ij}) f_c(r_{ij}) \right]^2 \quad (2)$$

where c_j is the element-dependent orbital coefficient of atom j , d_m is the contraction coefficient of the m th primitive GTO, f_c is a cosine type cutoff function that makes the interaction smoothly decayed to zero at the cutoff radius, N is the number of atoms, and N_φ refers to the number of primitive GTOs for a given l . The linear combination of these primitive GTOs performs like a contracted Gaussian-type function (CGF), where N_φ can be viewed as the "n" in STO-nG. The EAD vector ρ_i is expressed by the square of the linear combination of CGFs centered at neighboring atoms within cutoff radius. Up to this point, the EADs can be considered as standard three-body descriptors.

Through the message-passing framework, REANN deepens the correlation between central atom i and its environment beyond the cutoff radius. Different from EANN,⁴⁰ the neighboring atoms also have their unique fingerprints which can be reflected on c_j based on their individual environment. In the iteration step t of message passing, the orbital coefficients of the j th atom are updated based on its neighboring environment.

$$c_j^t = g_j^{t-1} [\rho_j^{t-1} (c_j^{t-1}, r_j^{t-1})] \quad (3)$$

where c_j^{t-1} and r_j^{t-1} are the orbital coefficients and atomic positions in the neighborhood of the central atom j in the $(t-1)$ th iteration and g_j^{t-1} is the j th NN module to update c_j^t . The total energy of the QM subsystem follows the atomistic energy decomposition scheme based on ρ_i :

$E = \sum_{i=1}^N E_i = \sum_{i=1}^N NN_i(\rho_i)$. The point charge of each atom can be obtained in a similar way. In our method, two REANNs are constructed, respectively, for the QM energy and point charges. Then, the point charges are used to calculate the electrostatic interactions between the QM and MM subsystems.

The loss function of the potential energy surface is the same as that in the original REANN package

$$L = \sum_{i=1}^{N_b} \sum_{j=1}^{N_{\text{atom}}} \lambda_F \times \frac{1}{N_b \times N_{\text{atom}}} |E_{ij}^{\text{REANN}} - F_{ij}^{\text{abinitio}}|^2 + \sum_{i=1}^{N_b} \lambda_E \times \frac{1}{N_b} |E_i^{\text{REANN}} - E_i^{\text{abinitio}}|^2 \quad (4)$$

where, N_b is the size of the mini-batch. λ_F and λ_E represent the weights of the energy and force in the loss function used in the construction of PES, respectively. The superscript "ab initio" in the equation denotes the physical quantities of the QM atoms computed in vacuo at the DFT level.

2.2. Electrostatic Energy in the ME REANN. In the ME scheme, the electrostatic interactions between the QM and MM subsystems are calculated by Coulomb's law

$$E_{\text{QM/MM}}^{\text{elec,ME}} = \sum_{ij} \frac{Q_i q_j}{R_{ij}} \quad (5)$$

in which Q_i is the point charge fitted from the electrostatic potential (ESP),^{49,50} q_j is the charge of MM atom j in the force field, and R_{ij} is the distance between QM atom i and MM atom j .

The total charge of the QM atoms should be a constant. However, REANN⁴⁸ architecture cannot guarantee that the sum of all QM charges is equal to the preset value of the total charge in the DFT calculations. To ensure the conservation of total charge during simulations and address the potential future requirements for systems with different charge states, here we

have encoded the charge equilibration (QEq) approach⁴⁶ into the REANN package. Otherwise, the MD simulations are unstable, and the simulated systems will collapse on a short time scale.

It is assumed that the charge density of the system is a superposition of spherically symmetric Gaussian functions centered at the atomic positions. By minimizing the total energy with respect to atomic charge Q_i , one can get a linear system of equations. For details of the principle, please refer to CENT in Goedecker's work⁵¹ or RuNner in Behler's work,⁵² and here we applied it to fit ESP charges in vacuo. The point charges, Q_1, Q_2, \dots, Q_N (N is the number of QM atoms), are solved by a set of linear equations as follows, in which χ_i is the electronegativity for atom i , and the conservation of total charge Q_{tot} is addressed by including the constraint via the Lagrange multiplier λ . The matrix \mathbf{A} is defined as

$$\left(\begin{array}{c|c} \mathbf{A} & \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ \hline 1 & \dots & 1 & 0 \end{array} \right) \begin{pmatrix} Q_1 \\ \vdots \\ Q_N \\ \lambda \end{pmatrix} = \begin{pmatrix} -\chi_1 \\ \vdots \\ -\chi_N \\ Q_{\text{tot}} \end{pmatrix} \quad (6)$$

$$\mathbf{A}_{ij} = \begin{cases} J_i + \frac{1}{\sigma_i \sqrt{\pi}}, & \text{if } i = j \\ \frac{\text{erf}\left(\frac{r_{ij}}{\sqrt{2(\sigma_i^2 + \sigma_j^2)}}\right)}{r_{ij}}, & \text{otherwise} \end{cases} \quad (7)$$

where r_{ij} represents the distance between atoms i and j . σ_i and J_i are the width of Gaussian charge densities and element-specific hardness for atom i , respectively.

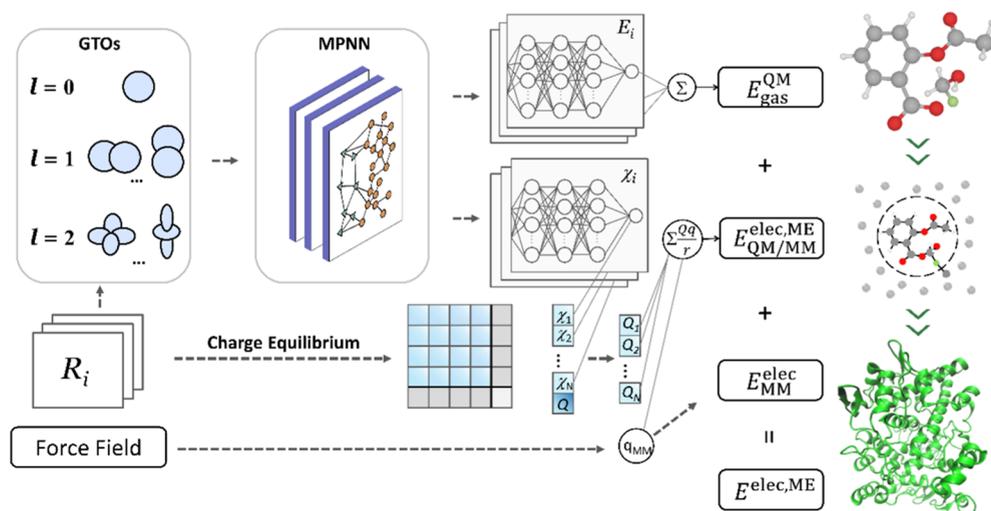


Figure 1. Schematic diagram of ME QEq-REANN framework to calculate the total electrostatic energy of a QM/MM system. Starting from the Cartesian coordinates of atoms in the QM subsystem, GTOs are constructed to represent atom-centered coordinate vectors in the local environment. After iterative message passing, atomistic neural networks output their energy E_i and χ_i , respectively. E_i are summed up to get $E_{\text{gas}}^{\text{QM}}$, and Q_i are calculated by solving the linear equations as described in eq 6. Then, the electrostatic interactions of $E_{\text{QM/MM}}^{\text{elec,ME}}$ are calculated by Coulomb's law. q_{MM} are the charges for MM atoms from the molecular force field and $E_{\text{MM}}^{\text{elec}}$ is the electrostatic interaction within the MM subsystem.

The loss function of the point charge surface is defined as

$$L = \sum_{i=1}^{N_b} \sum_{j=1}^{N_{\text{atom}}} \frac{1}{N_b \times N_{\text{atom}}} |Q_{i,j}^{\text{REANN}} - Q_{i,j}^{\text{ab initio}}|^2 \quad (8)$$

N_b is the size of the mini-batch and $Q_{i,j}^{\text{ab initio}}$ denotes the atomic charges of the QM atoms computed in vacuo at the DFT level. Since σ_i and J_i do not have explicit reference values, they are treated as trainable parameters of the NN and are jointly optimized during the minimization of the loss function. The atomic electronegativities χ_i serve as intermediate quantities, which are predicted by atomic neural networks before the charge equilibration block and vary with the QM coordinates. Using the distance matrix of QM atoms, χ_i values, and optimizable parameters σ_i and J_i , QM point charges are solved by eq 6. The total electrostatic energy is calculated as follows

$$\begin{aligned} E^{\text{elec,ME}} &= E_{\text{MM}}^{\text{elec}} + E_{\text{gas}}^{\text{QM}} + E_{\text{QM/MM}}^{\text{elec,ME}} \\ &= \sum_{i,j} \frac{q_i q_j}{r_{ij}} + \sum_i E_{i,\text{gas}}^{\text{REANN}} + \sum_{i,j} \frac{Q_i^{\text{REANN}} q_j}{R_{ij}} \end{aligned} \quad (9)$$

where $E_{i,\text{gas}}^{\text{REANN}}$ is the energy of QM atom i in the gas phase obtained by REANN and Q_i^{REANN} is the charge of QM atom fitted by QEq-REANN. The scheme to calculate the total electrostatic energy is illustrated in Figure 1.

2.3. Difference between the EE and ME Schemes.

Under the Born–Oppenheimer approximation, the total effective Hamiltonian of QM/MM systems is

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \quad (10)$$

where \hat{H}_{QM} is the Hamiltonian for the QM subsystem in the gas phase, \hat{H}_{MM} is the energies of the MM subsystem, and $\hat{H}_{\text{QM/MM}}$ is the interaction Hamiltonian between QM and MM subsystems defined as

$$\hat{H}_{\text{QM/MM}} = \hat{H}_{\text{QM/MM}}^{\text{bonded}} + \hat{H}_{\text{QM/MM}}^{\text{vdW}} + \hat{H}_{\text{QM/MM}}^{\text{elec}} \quad (11)$$

In the electrostatic embedding scheme, the MM atoms are not polarized by the QM atoms, and the energies of \hat{H}_{MM} , $\hat{H}_{\text{QM/MM}}^{\text{bonded}}$, and $\hat{H}_{\text{QM/MM}}^{\text{vdW}}$ are calculated by classical force fields. The remaining terms are related to the electrostatic interactions with QM atoms and treated together on the QM level by solving the Schrödinger equation of the following Hamiltonian

$$\hat{H}_{\text{QM-MM}}^{\text{elec}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{QM/MM}}^{\text{elec}} \quad (12)$$

Its eigenvalue consists of two parts

$$E_{\text{QM-MM}}^{\text{elec}} = E_{\text{QM}} + E_{\text{QM/MM}}^{\text{elec}} \quad (13)$$

The QM/MM electrostatics include the permanent electrostatics and relaxation terms¹⁶

$$E_{\text{QM/MM}}^{\text{elec}}[\rho] = E_{\text{QM/MM}}^{\text{perm}}[\rho^0] + \Delta E_{\text{QM/MM}}^{\text{elec}}[\rho - \rho^0] \quad (14)$$

in which ρ^0 and ρ are the electron density of the QM subsystem in the gas phase and in the presence of the MM environment, respectively. The $E_{\text{QM/MM}}^{\text{perm}}[\rho^0]$ is calculated in the gas phase as

$$E_{\text{QM/MM}}^{\text{perm}}[\rho^0] = \sum_j \phi_j^0 q_j \quad (15)$$

in which ϕ_j^0 and q_j are the electrostatic potential on the MM atomic site j in the gas phase arising from the continuous QM electron density and point charge of MM atom j , respectively.

And $\Delta E_{\text{QM/MM}}^{\text{elec}}[\rho - \rho^0]$ can be calculated in a similar way

$$\Delta E_{\text{QM/MM}}^{\text{elec}}[\rho - \rho^0] = \sum_j (\phi_j - \phi_j^0) q_j \quad (16)$$

where ϕ_j is the electrostatic potential on the MM atomic site j in the presence of the MM environment. The QM potential energy can be decomposed into the energy of the QM subsystem in the gas phase, and QM energy can be polarized by the MM subsystem

$$E_{\text{QM}}[\rho] = E_{\text{QM}}^0[\rho^0] + E_{\text{QM}}^{\text{dist}}[\rho - \rho^0] \quad (17)$$

Within the classical linear response theory^{16,53,54}

$$E_{\text{QM}}^{\text{dist}}[\rho - \rho^0] = -\frac{1}{2} \Delta E_{\text{QM/MM}}^{\text{elec}}[\rho - \rho^0] \quad (18)$$

Therefore

$$\begin{aligned} E_{\text{QM-MM}}^{\text{elec}} &= E_{\text{QM}}^0[\rho^0] + E_{\text{QM}}^{\text{dist}}[\rho - \rho^0] + E_{\text{QM/MM}}^{\text{perm}}[\rho^0] \\ &\quad + \Delta E_{\text{QM/MM}}^{\text{elec}}[\rho - \rho^0] \\ &= E_{\text{QM}}^0[\rho^0] + \frac{1}{2} \sum_j (\phi_j + \phi_j^0) q_j \end{aligned} \quad (19)$$

which means that in the electrostatic embedding (EE) scheme, $E_{\text{QM-MM}}^{\text{elec}}$ can be decomposed into the QM energy in the gas phase and the electrostatics between the charges of MM atoms and electrostatic potentials $\frac{1}{2}(\phi_j + \phi_j^0)$. While in the mechanic embedding (ME) scheme, we first obtained the QM energy and partial point charges for QM atoms by solving the Schrödinger equation in the gas phase, and then, the electrostatics are calculated by the Coulomb's law

$$E_{\text{QM-MM}}^{\text{elec,ME}} = E_{\text{QM}}^0[\rho^0] + \sum_{i,j} \frac{Q_i q_j}{R_{ij}} \quad (20)$$

in which Q_i is the point charge of QM atom i and R_{ij} is the distance between QM atom i and MM atom j . We define $E_{\text{QM-MM}}^{\text{elec,ME}^0}$ as the electrostatics calculated by the electrostatic potentials instead of point charges

$$E_{\text{QM-MM}}^{\text{elec,ME}^0} = E_{\text{QM}}^0[\rho^0] + \sum_j \phi_j^0 q_j \quad (21)$$

Now, the difference between the EE and ME schemes can be explained as

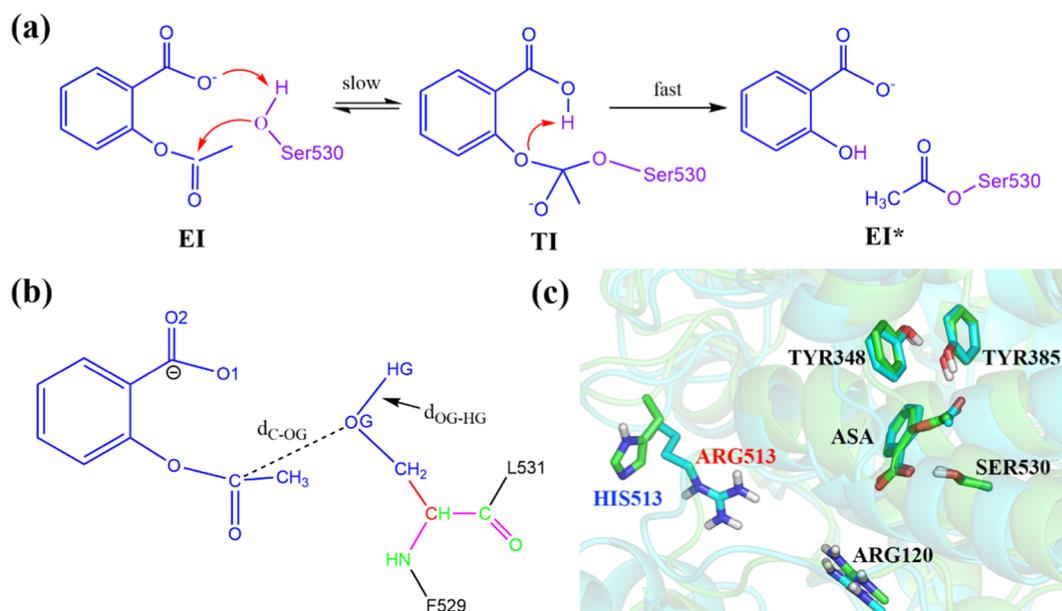


Figure 2. (a) The mechanism of aspirin acylating COX, whose first step is rate-limiting. (b) Illustration of the division of the QM/MM system for the acetylation reaction of COX. All atoms colored in blue are QM atoms, zero atoms are in green, and the pseudoatom is in red. All the left atoms in the system are MM atoms. (c) Overlap of the structures of COX-1 and COX-2 at the EI state. The carbon atoms are colored green in COX-1 and cyan in COX-2. The residue 513 is HIS in COX-1 and ARG in COX-2.

$$\begin{aligned}
 & E_{QM-MM}^{\text{elec,EE}} - E_{QM-MM}^{\text{elec,ME}} \\
 &= \Delta E^{\text{EE-ME}^0} + \Delta E^{\text{ME}^0-\text{ME}} \\
 &= \left(\frac{1}{2} \sum_j (\phi_j + \phi_j^0) q_j - \sum_j \phi_j^0 q_j \right) \\
 &+ \left(\sum_j \phi_j^0 q_j - \sum_{ij} \frac{Q_i q_j}{R_{ij}} \right) \quad (22)
 \end{aligned}$$

The first term originates from the effect of the MM environment on the electrostatic potentials, and the second term reflects the influence of the point charge approximation.

2.4. Workflow of the Reweighting ME REANN Method. In our method, a three-step procedure is employed to obtain the potential of mean force (PMF) of a reaction with REANN/MM potentials, as described below:

- (1) To address the QM/MM boundary, the pseudobond approach^{43–45} is applied, where any covalent bond formed between the QM and MM subsystems is described by a pseudobond. This pseudobond is independent of the molecular mechanical force field and offers a smooth connection between the QM and MM subsystems without introducing additional degrees of freedom into the system. Consequently, the QM subsystem remains independent of the MM environment. A series of QM configurations are selected from MD simulations, and their energies, forces, and ESP charges are calculated in the gas phase by Gaussian 16 software.⁵⁵ Then, the PES and the surface of point charges are constructed by the QEq-REANN package.
- (2) MD simulations are performed on the REANN-predicted PES and the partial charge surface. With umbrella sampling^{56,57} and MBAR analysis, we obtain the PMF along the reaction path in the ME scheme and the unbiased MBAR weight for each sample.

- (3) Finally, the PMF of the ME is reweighted to the EE scheme using wTP^{58,59} as outlined in our group's previous paper.³⁹ Whereas, the PMF curve corrected by wTP is often contaminated by statistical noise,⁵⁸ which is further smoothed by Gaussian process regression (GPR).⁶⁰

3. COMPUTATIONAL DETAILS

To test the validity of our method in enzyme reactions, we focused on the acylation of COX by aspirin. For detailed information about *ab initio* QM/MM-MD simulations and the PES training process, please refer to the [Supporting Information](#), while the details of REANN/MM-MD simulations are outlined below.

3.1. Construction of PES and Point Charge Surface.

The energies and ESP charges of the QM subsystem were calculated by Gaussian 16 software⁵⁵ on the B3LYP level. To eliminate the dependence on molecular orientation, the ESP charges were calculated in the standard orientation by Hu, Lu, and Yang charge fitting method using Gaussian's standard atomic densities (HLYGAt).⁶¹ As illustrated in the computational details in the [Supporting Information](#), 5000 and 27,021 *ab initio* points were selected from the trajectories of QM/MM-MD and REANN/MM-MD simulations by farthest point sampling (FPS)^{62,63} to fit the final PES and surface of charges, respectively, which were used to simulate the acylation of COX-1, COX-2, and R513A of COX-2. The final surface of potential energy and point charges were constructed with similar hyperparameters: the neural network for descriptors and atomistic energy or electronegativity contains two hidden layers with 64 neurons for each and one residual NN block, the maximum angular momentum is 2, and 8 radial Gaussian functions are linearly combined for any given angular momentum. The number of message-passing iterations is 2 for both the PES and the surface of charges. The L2 regularization coefficient for the ESP surface was set to

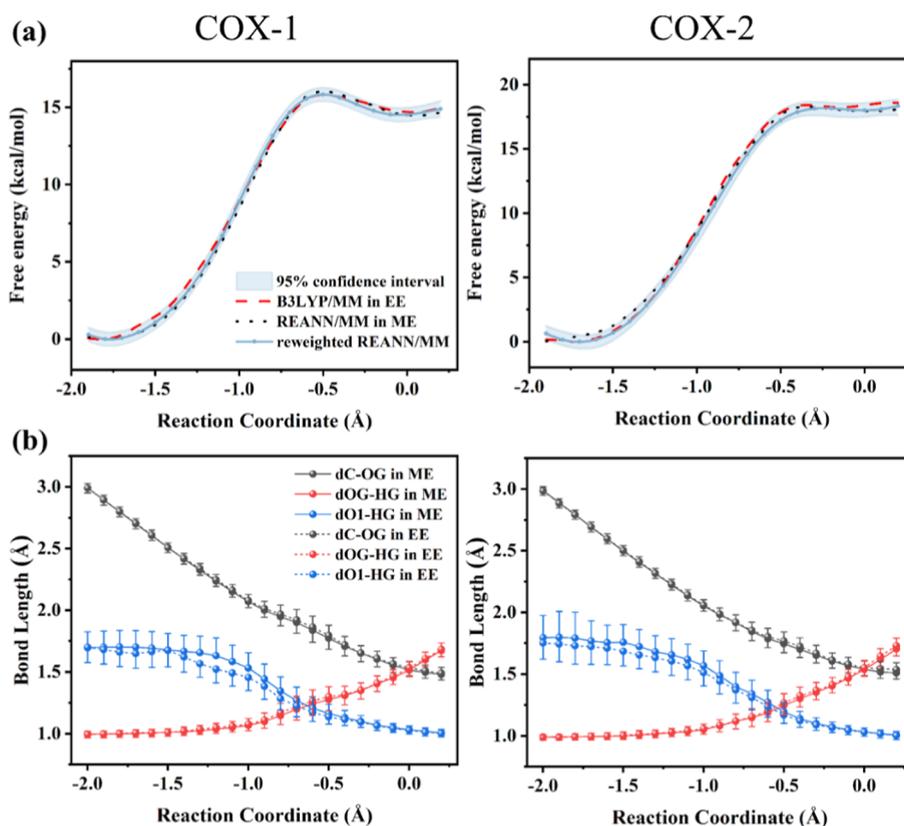


Figure 3. (a) Reweighted FE profile for the acylation of COX-1 and COX-2 calculated by the B3LYP/MM-MD, direct REANN/MM-MD, and reweighted REANN/MM-MD simulations with GPR smoothing. The 95% confidence intervals are also presented. (b) The variation of average bond distances along the reaction coordinate during MD simulations for COX-1 and COX-2.

0.005, the dropout probability was set to 0.25, and the early stopping strategy was incorporated to suppress overfitting.

3.2. REANN/MM-MD Simulations. We performed REANN/MM-MD simulations for three enzyme reactions under the same conditions as those used in B3LYP/MM-MD simulations. As illustrated in Figure 2b, the reaction coordinate is defined as $d_{\text{OG-HG}} - d_{\text{C-OG}}$. Sixteen windows centered at reaction coordinates from -2.0 to 0.2 Å were used in umbrella sampling. The force constants of the bias restraint were ranging from 30 to 70 kcal mol⁻¹ Å⁻². 500 ps REANN/MM-MD simulation under NVT conditions for each window was performed at 310 K with a time step of 1 fs through the interface developed by ourselves with Amber.⁶⁴ The configurations were saved every 0.1 ps for wTP correction and further analysis. MBAR analysis was employed to calculate the free energy curve and then it was corrected to the EE scheme by wTP followed by GPR.

4. RESULTS

4.1. QM/MM-MD Simulations. Aspirin, an ancient anti-inflammatory drug, primarily targets cyclooxygenase-2 (COX-2) in its pharmacological action.^{65,66} By acetylating Ser530 in the active site of COX-2, aspirin covalently modifies the enzyme, thereby hindering the proper binding of arachidonic acid, the native substrate of COX-2.^{67,68} However, it lacks COX-2 specificity, which will lead to side effects, such as the gastric ulceration.^{69,70} From our previous studies,⁷¹ the inhibition potency difference between the two COX subtypes mainly comes from the acylation of COX by aspirin, which proceeds in two successive stages and the initial step is rate-

limiting (see Figure 2a). In that work, a spheric boundary condition was adopted using QChem-Tinker interface.⁷² Because the Amber program mainly employs periodic boundary conditions (PBCs) to simulate the behavior of molecules in solvents, we restudied the first step by QM/MM-MD simulations using QChem-Amber interface in this work.⁷²

Figure 3a presents the free energy profiles of this step with B3LYP/MM-MD simulations for both subtypes of COX. The shapes of the two curves are similar to a high-energy intermediate. The free energy barriers are 15.9 ± 0.2 and 18.4 ± 0.4 kcal mol⁻¹ for COX-1 and COX-2, respectively, consistent with the experimental findings that aspirin is 10–100 times more potent against COX-1 than against COX-2.^{73,74} Then, the free energy curves serve as a benchmark to check the REANN/MM-MD simulations.

4.2. Model Fitting of PES and Point Charge Surface.

The QM subsystem contains 26 atoms, including aspirin and the side chain of Ser530, which directly participates in the acylation reaction (as illustrated in Figure 2b). The rest of the atoms of the enzyme are in the MM subsystem, including the main chain of Ser530. Therefore, there is a covalent bond connecting the QM and MM subsystems, which is addressed by the pseudobond approach^{43–45} in this work. In this approach, the QM subsystem is closed-shell, and no additional degrees of freedom are introduced into the system. In the ME scheme, the QM subsystem is independent of the MM environment, and its neural network surfaces for potentials and charges can be constructed independently and applied to various MM environments.

Farthest point sampling (FPS) was used to select points to mimic the original distribution, and a set of bond lengths were employed as descriptors in this work. QM calculations are unable to provide the gradients of point charges, so more data points are required for charges to achieve satisfied accuracy when the PES has already converged, as detailed in the [Supporting Information](#). 5000 and 27,021 ab initio points were finally selected to fit the PES and surface of charges, respectively.^{62,63} The root-mean-square errors (RMSE) for final surfaces are only 84 mkcal mol⁻¹, 248 mkcal mol⁻¹ Å⁻¹, and 0.016 *e* for the total energy, force, and charge on the training set, respectively (Table 1). As illustrated in Figure S1,

Table 1. Root Mean Square Errors (RMSEs) for the Final Surfaces^e

procedure	energy (mkcal mol ⁻¹)	force (mkcal mol ⁻¹ Å ⁻¹)	charge (10 ⁻³ e)
training	84.33	247.55	16.15
validation	111.03	301.70	12.48
test1 ^a	86.38	281.64	13.52
test2 ^b	99.19	295.52	13.45
test3 ^c	106.72	296.70	14.92
test4 ^d	91.29	265.07	14.66

^a8000 configurations sampled from 500 ps REANN/MM simulations for COX-1. ^b8000 configurations sampled from 500 ps REANN/MM simulations for COX-2. ^c8000 configurations sampled from 500 ps REANN/MM simulations for R513A. ^d8000 configurations sampled from 50 ps B3LYP/MM simulations for R513A. ^eThe training set consists of 5000 configurations for energy (mkcal mol⁻¹) and force (mkcal mol⁻¹ Å⁻¹) and 27,021 configurations for charges (10⁻³ e). The validation set consists of 3977 configurations randomly sampled from the trajectories of B3LYP/MM simulations for COX-1 and COX-2. Each test set consists of 8000 configurations randomly sampled from the trajectories of MD simulations.

the predicted energies, forces, and charges in vacuo exhibit high consistency with the reference values. Such a high quality of the surfaces ensures the reliability of the following MD simulations. In addition, the smoothness of the charge surfaces would influence the consistency of the QM-MM interaction energy over different configurations. Therefore, we have plotted the charge surfaces for four atoms as a function of two bond lengths related to the reaction coordinate with the rest coordinates fixed at the values of the reactant state in Figure S2, which demonstrates its smooth nature.

4.3. REANN/MM-MD Simulations of COX-1 and COX-2. We did 500 ps REANN/MM-MD simulations for COX-1 and COX-2 along the reaction path with umbrella sampling. The trajectories of the MD simulations for 16 windows are all stable, which demonstrates that our surfaces are robust enough to describe the dynamics of the reaction. Then, the PMF in the ME scheme was determined by MBAR analysis. Next, the PMF was reweighted to the EE scheme by wTP correction. To test the convergence of wTP, we calculated the PMFs with different time intervals for sampling harvesting. As illustrated in Figures S3 and S4, the PMF curves corrected by wTP are contaminated by statistical noise and become smoother with a shorter harvested time, which means that this noise can be weakened by increasing sampling frequency. Finally, the PMF was smoothed by GPR. Figure S3 compares the free energy profiles of COX-1 and COX-2 before and after GPR. The reliabilities of the wTP correction are also discussed in the [Supporting Information](#).

Figure 3a compares the PMFs of B3LYP/MM-MD, REANN/MM-MD, and REANN/MM-MD with the GPR-smoothed wTP correction for COX-1 and COX-2. The three PMFs exhibit similar shapes with the transition states (TSs) located approximately at the reaction coordinate of -0.5 Å for COX-1 and -0.4 Å for COX-2. After the wTP correction with a harvested time of 0.1 ps, the barriers are within a range of ±0.3 kcal mol⁻¹ compared with the reference values. Considering that the two subtypes of COX are 40% different in sequence, this proves the robustness and universality of our surfaces. To our surprise, even without wTP correction, direct REANN/MM-MD simulations can already reproduce the PMF of B3LYP/MM-MD simulations well. The free energy barriers of REANN/MM-MD and B3LYP/MM-MD differ by less than 0.5 kcal mol⁻¹ (see Table 2), which means that the sampling spaces of the two Hamiltonians in the ME and EE schemes might be quite similar.

Table 2. Free Energy Barriers (in kcal mol⁻¹) for the Three Enzyme Reactions

enzyme	direct REANN/MM ^a	reweighted REANN/MM ^a	B3LYP/MM ^b
COX-1	16.1 ± 0.2	15.8 ± 0.3	15.9 ± 0.2
COX-2	18.2 ± 0.1	18.1 ± 0.4	18.4 ± 0.4
R513A	17.3 ± 0.1	17.0 ± 0.3	16.9 ± 0.3

^aThe statistical error was estimated by the free energy difference between 100–300 ps and 300–500 ps. ^bThe statistical error was estimated by the free energy difference between 10–30 ps and 30–50 ps.

To further check the reliability of our surfaces, 8000 points were randomly chosen from MD simulations for COX-1 and COX-2 as test sets. The RMSEs of energy, force, and charges for the test sets are comparable to those for the training sets (see Table 1). We also examined all the configurations of the QM subsystem from the trajectories of REANN/MM-MD simulations, and less than 1% MD samples are outside the boundary of the training set for the reactions, which means that the training set sufficiently covers the possible sampling space of MD simulations for COX-1 and COX-2.

4.4. Computational Expense. The computational speed was evaluated on an Intel Xeon CPU E5-2680 2.50 GHz with 24 cores. Because the B3LYP/MM-MD simulation is expensive, we only did 50 ps B3LYP/MM-MD simulations for each window. To make a direct comparison, we estimated the wall-clock time of 500 ps B3LYP/MM-MD simulations by multiplying the time of 50 ps by a factor of 10. First, a lot of effort has been made to construct the surfaces, including single-point energy calculations and attempts to run REANN/MM-MD simulations. Fortunately, this initial setup is once and forever. As listed in Table S2, the following REANN/MM-MD simulations take 203 h for COX-1, which is 82 times faster than the 16,701 h required by B3LYP/MM-MD. In the reweighting stage, single-point calculations were performed on the B3LYP/MM level for those points sampled from REANN/MM-MD trajectories at regular intervals, so its wall-clock time depended on the harvested time. A harvested time of 0.1 ps requires an additional 181 h in this stage. Convergence tests in our group's previous work³⁹ and Figure S4 demonstrate that 0.5 ps harvested time is sufficient, allowing for an 80% reduction in reweighting time.

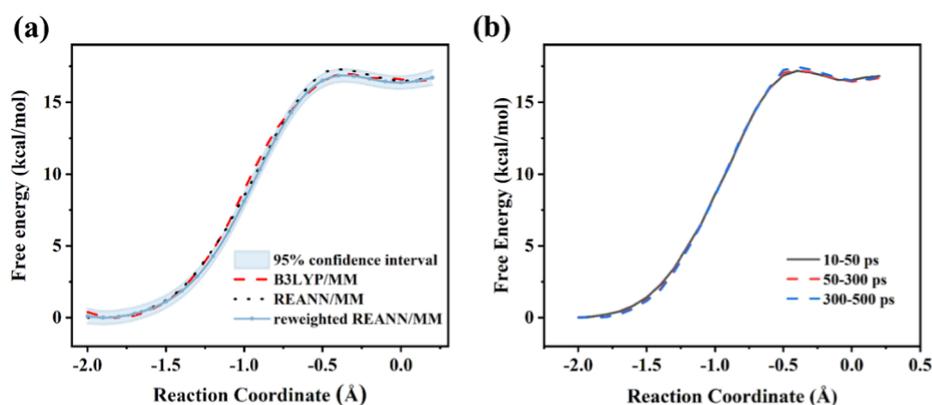


Figure 4. (a) Comparison of free energy profiles for R513A of COX-2 by the B3LYP/MM-MD, direct REANN/MM-MD, and reweighted REANN/MM-MD simulations with GPR smoothing. The 95% confidence intervals are also presented. (b) The free energy profiles of direct REANN/MM-MD simulations of 10–50, 50–300, and 300–500 ps.

4.5. Prediction of the Inhibition Potential of Aspirin on R513A of COX-2. In the ME scheme, the “offline-trained”⁷⁵ ML potentials and charge surfaces are independent of the MM environments. Ideally, if the training set is carefully prepared to comprehensively cover the whole sampling space, the REANN surfaces can be employed in various MM environments, for example, a different enzyme with the same active site. Here, we used the newly developed potential and charge surfaces to study the acetylation reaction of a mutant of COX-2 to check the robustness of these surfaces.

In our group’s previous work,⁷¹ we calculated the contribution of individual residue to the stability of transition state of acetylation. Close to the binding pocket, most residues of the two COXs are conserved and have similar effects on the TS stability as depicted in Figure 4 of previous work,⁷¹ which means that they are not responsible for the different inhibition potency of aspirin on COX-1 and COX-2. One exception is that the positive residue Arg513 in COX-2, while its counterpart in COX-1 is His513 (Figure 2c). Compared with His513 in COX-1, Arg513 in COX-2 increases the activation barrier of the acetylation reaction by about 2 kcal mol⁻¹. An analysis of the conformational changes during the acylation illustrates that the interaction between the positively charged guanidinium group of Arg513 of COX-2 and the negatively charged carboxylic group of aspirin would disfavor the protonation of the aspirin’s carboxylic group during the reaction process.⁷¹ Consequently, the mutation of Arg513 to a nonpolar residue is expected to lower the energy barrier of COX-2.

In this work, to examine the effect of the guanidinium group of Arg513 of COX-2, we calculated the free energy curve for the R513A mutant by REANN/MM-MD simulations. The mutation resulted in a decrease in the free energy barrier by about 1.1 kcal mol⁻¹, confirming our hypothesis in previous work,⁷¹ that the positively charged guanidinium group of Arg513 of COX-2 hinders the inhibition of aspirin. Unlike the wild-type COX-2, no information concerning the R513A mutant was used when we constructed surfaces. Nonetheless, the REANN/MM-MD still closely reproduced the curve of B3LYP/MM-MD well (Figure 4a) with a deviation under 0.5 kcal mol⁻¹. 8000 points were randomly chosen from MD simulations as a test set, and the RMSE was only 107 mkcal mol⁻¹ for energy as listed in Table 1. We also calculated the PMFs for various sampling durations, and the corresponding energy barriers are listed in Table S3. Evidently, the longer the

simulation time, the smaller the uncertainty of the barrier. While any MD simulations with a duration of 50 ps can reproduce the energy barrier observed in 100–500 ps simulations with an error of less than 0.6 kcal/mol. In addition, the free energy profiles of R513A with different simulation time intervals (Figure 4b) also demonstrate that 50 ps simulation is enough for the REANN/MM-MD simulations to converge. Therefore, in order to save computational costs, 50 ps MD simulations are acceptable. In summary, our surfaces can be applied to study the properties of mutants rapidly and accurately.

5. DISCUSSION

The PMFs in the ME and EE schemes are highly comparable for COX-1 and COX-2 (Figure 3a), which indicates a substantial overlap in the phase space between the Hamiltonians of the two schemes. To further prove the overlap in the phase space, a comprehensive analysis of the structures in the ME and EE schemes was conducted from various perspectives.

We examined the variation of key distances along the reaction coordinate during QM/MM-MD simulations (Figure 3b). In the first step of acylation, the oxygen atom of the hydroxyl group of Ser530 (OG) attacks the carbonyl carbon atom (C) of aspirin to form a metastable tetrahedral intermediate (TI), and meanwhile, the carboxyl group of aspirin (O1) serves as the general base to abstract the proton (HG) from the hydroxyl group of Ser530 (Figure 2a). Three distances are involved in this step: d_{C-OG} , d_{OG-HG} , and d_{O1-HG} (as labeled in Figure 2b). The three distances change monotonically, and the O–H bond does not break until the C–O distance has shortened to about 2 Å. The transfer of H to the carboxyl group of aspirin does not occur before the attack of the O atom at the carbonyl carbon of aspirin, which illustrates that they are concerted steps to form the tetrahedral intermediate. The change of the distances is almost the same in the two schemes for both COX-1 and COX-2, demonstrating that the sampling spaces of the EE and ME schemes overlap significantly during MD simulations. This may be the reason that the PMF of the ME scheme is very close to that of the EE scheme even without wTP correction.

A detailed decomposition analysis of QM-MM electrostatics is conducted to examine the similarity of Hamiltonians between the two schemes. Figures 3a and 4a reveal that the energy barriers of the two schemes do not differ much, which

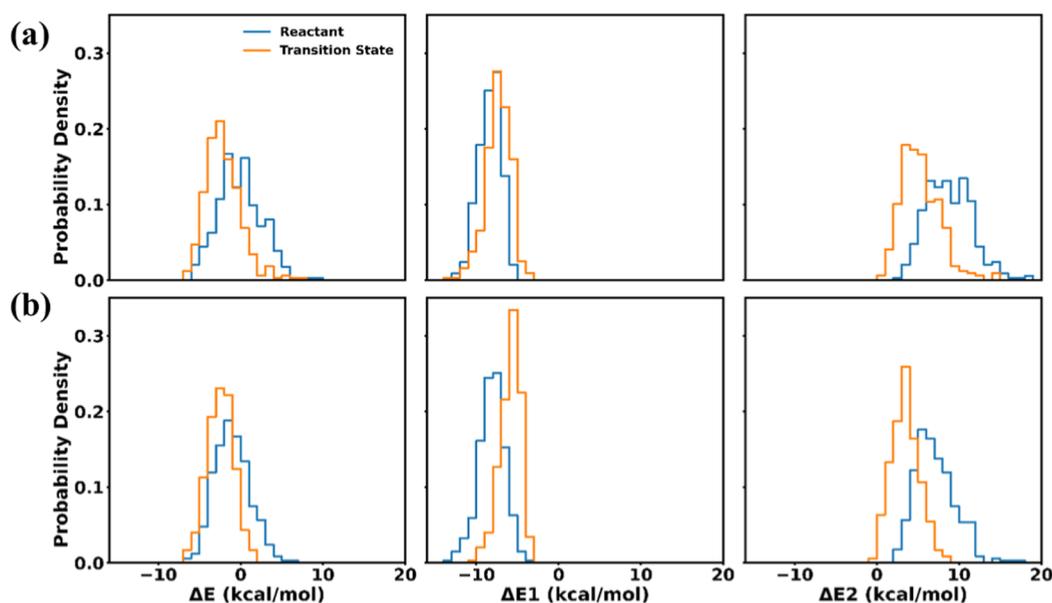


Figure 5. 8000 configurations are chosen from each window of MD simulation, and the energy differences between the ME and EE schemes are analyzed for (a) COX-1 and (b) COX-2. ΔE is the energy difference between the B3LYP/MM and REANN/MM and can be decomposed into ΔE_1 and ΔE_2 . ΔE_1 is the energy difference between the potential model of the EE and ME schemes, namely $\Delta E^{\text{EE}-\text{ME}^0}$ in eq 22. ΔE_2 is the energy difference between the potential and point charge models in the ME scheme, namely $\Delta E^{\text{ME}^0-\text{ME}}$ in eq 22.

means that the change of the environment effect along the reaction path is similar in both the ME and EE schemes. Equation 22 suggests that the ME scheme can be viewed as an approximation of the EE scheme. The discrepancies between them originate from two aspects: the influence of the MM environment on the electrostatic potentials (termed the ME approximation, labeled ΔE_1 in Figure 5), and the point charge approximation (labeled ΔE_2 in Figure 5) in the gas phase. The two effects are plotted in Figure 5 for the reactant and TS states. Clearly, the comparison of ΔE_1 illustrated that the MM environment could stabilize the whole system with a similar effect at both the TSs and reactant states, which might be due to the rigidity of enzymes. Although the point charge approximation in the gas phase is not a perfect approximation for this QM subsystem, it performs better at the TSs than at the reactant state. When considering the statistical average of the combined effects, the total difference in the free energy between ME and EE schemes remains almost constant along the reaction path. Is this phenomenon a coincidence for COX, or do all enzyme reactions possess similar characteristics? Further research on more enzyme reactions is necessary to address this question.

On the other hand, the wTP is time-consuming with additional QM calculations and always statistically contaminated. Therefore, a correction might not be necessary for such a small difference between the ME and the EE schemes. However, we cannot guarantee that the difference will always be small in all of the enzyme reactions. Therefore, it is beneficial to develop a universal correction model for arbitrary systems. In recent studies, a variant of the electrostatic embedding scheme was proposed by Kirill Zinovjev that allows training potentials of the EE scheme in vacuo, in which the responses of the QM subsystem to the MM environment were calculated by the atomic polarizabilities model with Thole damping.^{76,77} The Semelak group also proposed a polarizable ME model using minimal basis iterative stockholder (MBIS)

atomic charges in vacuo calculations.⁷⁸ These methods suggest that on-the-fly corrections may be feasible if the polarization of the QM subsystem by the environment can be accurately measured using a limited number of parameters.

6. CONCLUSIONS

Simulation of large systems through machine learning potential is still a challenge, especially for enzyme reactions. Analogous to hybrid QM/MM, directly integrating the ML potential with the MM force field is currently the most promising solution to achieve the balance between computational cost and accuracy. In this work, we have proposed a simple, efficient, and universal ML/MM model to study the dynamics of enzymes and mutants, which is capable of tracing the QM charge gradient and partially bridging the gap between the EE and ME schemes. In our model, REANN and QEq-REANN surfaces are fitted for potential energies and point charges of QM atoms, respectively. To ensure the conservation of the total charge in the close-shelled QM subsystem, the charge equilibration method is encoded into the REANN package. The validity of our model is supported by the acylation of COX by aspirin, where the free energy barriers agree well with B3LYP/MM MD simulations, with an 80-fold speed-up. One key feature of our model is that the surfaces are trained in vacuo for the QM subsystem, making them independent of the MM atomic properties. Theoretically, they can be applied to any mutant with the same active site. Ultimately, the surfaces were successfully applied to R513A of COX-2 to reproduce the energy barrier of B3LYP/MM MD simulations with a deviation of less than 0.5 kcal mol⁻¹.

In summary, our strategy offers a promising approach to the modeling of enzyme reactions. We will further develop this method to ensure that it can be applied to a general case. New strategies will be adopted to speed up the training process for charge prediction and span the relevant chemical and conformational space automatically and universally, for

example, adopting molecular descriptors, like SOAP,^{79,80} or generating configurations through uncertainty-driven active learning strategy.^{81,82} On the other hand, we have only tested the acetylation of COX by aspirin, which is well-described by the ME scheme without polarization effects in the QM region. Therefore, more complex cases are required to verify its general validity. For general reactions, the wTP correction can be employed to compensate for the difference between the ME and EE schemes. In addition, on-the-fly corrections from the ME to EE scheme can be employed to reproduce more reliable results for complex cases. We anticipate that this method will soon open up avenues for high-throughput virtual screening of enzyme evolution.

■ ASSOCIATED CONTENT

Data Availability Statement

Relevant data sets, the QEq-REANN package and its Fortran interface are open source at <https://github.com/shaxhch/QEq-REANN/>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.5c00149>.

Computational details, correlations between the predicted and reference values, REANN fitted charges, reweighted free energy profiles, histograms, and reweighting entropies in the wTP calculations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yanzi Zhou – Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China; orcid.org/0000-0002-2097-9186; Email: zhouyz@nju.edu.cn

Authors

Xinhu Sha – Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China; orcid.org/0009-0000-2672-7812

Zhuo Chen – Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China

Daiqian Xie – Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China; Hefei National Laboratory, Hefei 230088, China; orcid.org/0000-0001-7185-7085

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.5c00149>

Author Contributions

Y.Z. designed the project and X.S., Z.C., and Y.Z. discussed the neural network architecture. Y.Z. and X.S. wrote the code and X.S. performed all calculations. All authors participated in data

analyses and writing of the manuscript. Y.Z. and D.X. supervised the study and provided the funding.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 22073040 to Y.Z.), the Innovation Program for Quantum Science and Technology (2021ZD0303305 to D.X.), and the National Natural Science Foundation of China (Grant Nos. 22233003 and 22241302 to D.X.). All the calculations were carried out on the computing facilities in the High Performance Computing Center (HPCC) of Nanjing University. The authors sincerely thank Prof. Bin Jiang, Dr. Junfan Xia, and Dr. Yaolong Zhang from the University of Science and Technology of China for their help in the recursively embedded atom neural network method.

■ REFERENCES

- (1) Vaissier Welborn, V.; Head-Gordon, T. Computational design of synthetic enzymes. *Chem. Rev.* **2019**, *119*, 6613–6630.
- (2) Sheldon, R. A.; Pereira, P. C. Biocatalysis engineering: the big picture. *Chem. Soc. Rev.* **2017**, *46*, 2678–2691.
- (3) Qu, G.; Li, A.; Acevedo-Rocha, C. G.; Sun, Z.; Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem., Int. Ed.* **2020**, *59*, 13204–13231.
- (4) Zeymer, C.; Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **2018**, *87*, 131–157.
- (5) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (6) Acevedo, O.; Jorgensen, W. L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* **2010**, *43*, 142–151.
- (7) Brunk, E.; Rothlisberger, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* **2015**, *115*, 6217–6263.
- (8) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (9) Gao, J. L. Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* **1996**, *29*, 298–305.
- (10) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (11) Hu, H.; Yang, W. T. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.
- (12) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem. Commun.* **2010**, *46*, 2354–2372.
- (13) Gao, J. L. Absolute free energy of solvation from Monte Carlo simulations using combined quantum and molecular mechanical potentials. *J. Phys. Chem.* **1992**, *96*, 537–540.
- (14) Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: a tutorial review. *Int. J. Quantum Chem.* **2018**, *118*, 25558.
- (15) Shen, L.; Wu, J.; Yang, W. Multiscale quantum mechanics/molecular mechanics simulations with neural networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.
- (16) Gao, J.; Xia, X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631–635.
- (17) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. Hybrid QM/MM potentials of mean force with interpolated corrections. *J. Phys. Chem. B* **2004**, *108*, 8427–8433.

- (18) Ruiz-Pernia, J. J.; Silla, E.; Tunon, I.; Marti, S. Hybrid quantum mechanics/molecular mechanics simulations with two-dimensional interpolated corrections: application to enzymatic processes. *J. Phys. Chem. B* **2006**, *110*, 17663–17670.
- (19) Wu, J.; Shen, L.; Yang, W. Internal force corrections with machine learning for quantum mechanics/molecular mechanics simulations. *J. Chem. Phys.* **2017**, *147*, 161732.
- (20) Zhou, Y.; Pu, J. Z. Reaction path force matching: a new strategy of fitting specific reaction parameters for semiempirical methods in combined QM/MM simulations. *J. Chem. Theory Comput.* **2014**, *10*, 3038–3054.
- (21) Kuechler, E. R.; York, D. M. Quantum mechanical study of solvent effects in a prototype S_N2 reaction in solution: Cl^- attack on CH_3Cl . *J. Chem. Phys.* **2014**, *140*, 054109.
- (22) Dral, P. O.; von Lilienfeld, O. A.; Thiel, W. Machine learning of parameters for accurate semiempirical quantum chemical calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2120–2125.
- (23) Li, H.; Collins, C.; Tanha, M.; Gordon, G. J.; Yaron, D. J. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* **2018**, *14*, 5764–5776.
- (24) Rod, T. H.; Ryde, U. Quantum mechanical free energy barrier for an enzymatic reaction. *Phys. Rev. Lett.* **2005**, *94*, 138302.
- (25) Polyak, I.; Benighaus, T.; Boulanger, E.; Thiel, W. Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation. *J. Chem. Phys.* **2013**, *139*, 064105.
- (26) Pan, X. L.; Li, P. F.; Ho, J. M.; Pu, J. Z.; Mei, Y.; Shao, Y. H. 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.
- (27) 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.
- (28) Dral, P. O. Quantum chemistry in the age of machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.
- (29) Manzhos, S.; Carrington, T., Jr. Neural network potential energy surfaces for small molecules and reactions. *Chem. Rev.* **2021**, *121*, 10187–10217.
- (30) Kocer, E.; Ko, T. W.; Behler, J. Neural network potentials: a concise overview of methods. *Annu. Rev. Phys. Chem.* **2022**, *73*, 163–186.
- (31) Gastegger, M.; Schuett, K. T.; Mueller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.* **2021**, *12*, 11473–11483.
- (32) Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828–12840.
- (33) Shen, L.; Yang, W. Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks. *J. Chem. Theory Comput.* **2018**, *14*, 1442–1455.
- (34) Zeng, J.; Giese, T. J.; Ekesan, S.; 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.
- (35) Boeselt, L.; Thuerlemann, M.; Riniker, S. Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658.
- (36) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.
- (37) Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations; Topics in Current Chemistry-Series*; Springer Science & Business Media, 2007; Vol. 268, pp 173–290.
- (38) Lei, Y. K.; Yagi, K.; Sugita, Y. Learning QM/MM potential using equivariant multiscale model. *J. Chem. Phys.* **2024**, *160*, 214109.
- (39) Zhou, B.; Zhou, Y.; Xie, D. Accelerated quantum mechanics/molecular mechanics simulations via neural networks incorporated with mechanical embedding scheme. *J. Chem. Theory Comput.* **2023**, *19*, 1157–1169.
- (40) Zhang, Y.; Hu, C.; Jiang, B. Embedded atom neural network potentials: Efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.* **2019**, *10*, 4962–4967.
- (41) Giese, T. J.; Zeng, J.; Lerew, L.; McCarthy, E.; Tao, Y.; Ekesan, S.; York, D. M. Software infrastructure for next-generation QM/MM- Δ MMLP force fields. *J. Phys. Chem. B* **2024**, *128*, 6257–6271.
- (42) Gomez-Flores, C. L.; Maag, D.; Kansari, M.; Vuong, V. Q.; Irle, S.; Grater, F.; Kubar, 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.
- (43) Zhang, Y. K.; Lee, T. S.; Yang, W. T. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46–54.
- (44) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112*, 3483–3492.
- (45) Zhang, Y. K. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *J. Chem. Phys.* **2005**, *122*, 024114.
- (46) Rappe, A. K.; Goddard, W. A., III. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.
- (47) Zhang, Y.; Xia, J.; Jiang, B. Physically motivated recursively embedded atom neural networks: incorporating local completeness and nonlocality. *Phys. Rev. Lett.* **2021**, *127*, 156002.
- (48) Zhang, Y.; Xia, J.; Jiang, B. REANN: a PyTorch-based end-to-end multi-functional deep neural network package for molecular, reactive, and periodic systems. *J. Chem. Phys.* **2022**, *156*, 114801.
- (49) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (50) Sigfridsson, E.; Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* **1998**, *19*, 377–395.
- (51) Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* **2015**, *92*, 045131.
- (52) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.
- (53) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **2013**, *117*, 2328–2338.
- (54) Sodt, A. J.; Mei, Y.; Koenig, G.; Tao, P.; Steele, R. P.; Brooks, B. R.; Shao, Y. Multiple environment single system quantum mechanical/molecular mechanical (MESS-QM/MM) calculations. I. estimation of polarization energies. *J. Phys. Chem. A* **2015**, *119*, 1511–1523.
- (55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.;

- Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*. revision A. 03; Gaussian Inc.: Wallingford CT, 2016.
- (56) Boczko, E. M.; Brooks, C. L. Constant-temperature free energy surfaces for physical and chemical processes. *J. Phys. Chem.* **1993**, *97*, 4509–4513.
- (57) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (58) 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.
- (59) 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.
- (60) Seeger, M. Gaussian processes for machine learning. *Eur. J. Inf. Syst.* **2004**, *14*, 69–106.
- (61) Hu, H.; Lu, Z.; Yang, W. Fitting molecular electrostatic potentials from quantum mechanical calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1004–1013.
- (62) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W. Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels. *J. Chem. Phys.* **2017**, *146*, 244108.
- (63) Eldar, Y.; Lindenbaum, M.; Porat, M.; Zeevi, Y. Y. The farthest point strategy for progressive image sampling. *IEEE Trans. Image Process.* **1997**, *6*, 1305–1315.
- (64) Case, D. A.; Darden, T. A.; Cheatham, T. E., I; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, 2012.
- (65) Pennisi, E. Building a better aspirin. *Science* **1998**, *280*, 1191–1192.
- (66) Masferrer, J. L.; Zweifel, B. S.; Manning, P. T.; Hauser, S. D.; Leahy, K. M.; Smith, W. G.; Isakson, P. C.; Seibert, K. Selective-inhibition of inducible cyclooxygenase-2 in-vivo is antiinflammatory and nonulcerogenic. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3228–3232.
- (67) Flower, R. J. The development of COX2 inhibitors. *Nat. Rev. Drug Discovery* **2003**, *2*, 179–191.
- (68) Blobaum, A. L.; Marnett, L. J. Structural and functional basis of cyclooxygenase inhibition. *J. Med. Chem.* **2007**, *50*, 1425–1441.
- (69) Gierse, J. K.; Koboldt, C. M.; Walker, M. C.; Seibert, K.; Isakson, P. C. Kinetic basis for selective inhibition of cyclooxygenases. *Biochem. J.* **1999**, *339*, 607–614.
- (70) Meade, E. A.; Smith, W. L.; Dewitt, D. L. Differential inhibition of prostaglandin endoperoxide synthase (cyclooxygenase) isozymes by aspirin and other nonsteroidal antiinflammatory drugs. *J. Biol. Chem.* **1993**, *268*, 6610–6614.
- (71) Lei, J.; Zhou, Y.; Xie, D.; Zhang, Y. Mechanistic insights into a classic wonder drug-aspirin. *J. Am. Chem. Soc.* **2015**, *137*, 70–73.
- (72) Zhou, Y.; Wang, S.; Li, Y.; Zhang, Y. Born-Oppenheimer ab Initio QM/MM molecular dynamics simulations of enzyme reactions. *Methods Enzymol.* **2016**, *577*, 105–118.
- (73) Vane, J. R.; Bakhle, Y. S.; Botting, R. M. Cyclooxygenases 1 and 2. *Annu. Rev. Pharmacol. Toxicol.* **1998**, *38*, 97–120.
- (74) Mitchell, J. A.; Akaraseenont, P.; Thiemermann, C.; Flower, R. J.; Vane, J. R. Selectivity of nonsteroidal antiinflammatory drugs as inhibitors of constitutive and inducible cyclooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11693–11697.
- (75) Deringer, V. L.; Caro, M. A.; Csanyi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **2019**, *31*, 1902765.
- (76) Zinovjev, K.; Hedges, L.; Montagud Andreu, R.; Woods, C.; Tunon, I.; van der Kamp, M. W. emle-engine: A flexible electrostatic machine learning embedding package for multiscale molecular dynamics simulations. *J. Chem. Theory Comput.* **2024**, *20*, 4514–4522.
- (77) Zinovjev, K. Electrostatic Embedding of Machine Learning Potentials. *J. Chem. Theory Comput.* **2023**, *19*, 1888–1897.
- (78) Grassano, J. S.; Pickering, I.; Roitberg, A. E.; Gonzalez Lebrero, M. C.; Estrin, D. A.; Semelak, J. A. Assessment of embedding schemes in a hybrid machine learning/classical potentials (ML/MM) approach. *J. Chem. Inf. Model.* **2024**, *64*, 4047–4058.
- (79) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
- (80) Zhang, H.; Juraskova, V.; Duarte, F. Modelling chemical processes in explicit solvents with machine learning potentials. *Nat. Commun.* **2024**, *15*, 6114.
- (81) Lin, Q.; Zhang, Y.; Zhao, B.; Jiang, B. Automatically growing global reactive neural network potential energy surfaces: a trajectory-free active learning strategy. *J. Chem. Phys.* **2020**, *152*, 154104.
- (82) Lin, Q.; Zhang, L.; Zhang, Y.; Jiang, B. Searching configurations in uncertainty space: active learning of high-dimensional neural network reactive potentials. *J. Chem. Theory Comput.* **2021**, *17*, 2691–2701.