Machine Learning Assisted Free Energy Simulation of Solution–Phase and Enzyme Reactions

Xiaoliang Pan,[†] Junjie Yang,[†] Richard Van,[†] Evgeny Epifanovsky,[‡] Junming Ho,[¶] Jing Huang,[§] Jingzhi Pu,[*,||] Ye Mei,[*,⊥,#,@] Kwangho Nam,[*,△] and Yihan Shao[*,†]

[†]*Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Pkwy, Norman, OK 73019, United States.*
[‡]*Q-Chem, Inc. 6601 Owens Drive, Suite 105, Pleasanton, CA 94588, United States.*
[¶]*School of Chemistry, University of New South Wales, Sydney, NSW 2052, Australia.*
[§]*Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou, Zhejiang 310024, China*
[||]*Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N Blackford St, LD326, Indianapolis, IN 46202, United States*
[⊥]*State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China*
[#]*NYU–ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*
[@]*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*
[△]*Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, TX 76019, United States.*

E-mail: jpu@iupui.edu; samuel.y.mei@gmail.com; kwangho.nam@uta.edu; yihan.shao@ou.edu

Abstract

Despite recent advances in the development of machine learning potentials (MLPs) for biomolecular simulations, there has been limited effort in developing stable and accurate MLPs for enzymatic reactions. Here, we report a protocol for performing machine learning assisted free energy simulation of solution-phase and enzyme reactions at an *ab initio* quantum mechanical and molecular mechanical (*ai*-QM/MM) level of accuracy. Within our protocol, the MLP is built to reproduce the *ai*-QM/MM energy as well as forces on both QM (reactive) and MM (solvent/enzyme) atoms. As an alternative strategy, a delta machine learning potential (ΔMLP) is trained to reproduce the differences between *ai*-QM/MM and semi-empirical (*se*) QM/MM energy and forces. To account for the effect of the condensed–phase environment in both MLP and ΔMLP, the DeePMD representation of a molecular system is extended to incorporate external electrostatic potential and field on each QM atom. Using the Menshutkin and chorismate mutase reactions as examples, we show that the developed MLP and ΔMLP reproduce the *ai*-QM/MM energy and forces with an error on average less than $1.0\,\mathrm{kcal/mol}$ and $1.0\,\mathrm{kcal/mol/\mathring{A}}$ for representative configurations along the reaction pathway. For both reactions, MLP/ΔMLP-based simulations yielded free energy profiles that differed by less than $1.0\,\mathrm{kcal/mol}$ from the reference *ai*-QM/MM results, but only at a fractional computational cost.

1 Introduction

To accurately model solution-phase and enzyme reactions, it would be desirable to perform direct *ab initio* quantum mechanical and molecular mechanical (*ai*-QM/MM) free energy simulations.[1–11] In a typical *ai*-QM/MM free energy simulation, the *ai*-QM/MM potential is evaluated for each configuration of the system of interest, in which a quantum mechanical (QM) reactive region (typically containing up to 150 atoms) is embedded in thousands of or more molecular mechanical (MM) atoms (*i.e.* the rest of enzyme or solvent atoms).[12,13] In practice, the *ai*-QM/MM potential needs to be computed for 10^5 or 10^6 configurations in umbrella sampling calculations[14] or other enhanced samplings (such as metadynamics), before the mean free energy pathway and corresponding reaction free energy profile can be obtained. As such, direct *ai*-QM/MM free energy simulations are rather compute-intensive, requiring $\mathcal{O}(10^5)$ CPU hours,[15–18] and thus have yet to gain wide use.

To avoid such steep costs of direct *ai*-QM/MM free energy simulations, Gao,[19] Warshel[20] and others[21–25] developed indirect free energy simulations, where sampling is carried out using a reference semi-empirical QM/MM (*se*-QM/MM) potential and then the free energy result is corrected with a *se*-QM/MM⟶*ai*-QM/MM thermodynamic perturbation or with an interpolation between the two levels of potential, for example along the energy minimized reaction path.[26,27] Alternatively, one can adopt the multiple time-step simulation methodology from Tuckerman,[28–31] Schlick,[32,33] Nam,[34] Roux,[35] Rothslisberg,[36] and others,[37] where *se*-QM/MM trajectory propagation at regular/inner timesteps is coupled with *ai*-QM/MM trajectory corrections at outer timesteps. Needless to say, the accuracy of both indirect and multiple-timestep simulations is controlled by the quality of the *se*-QM/MM potential in use. In many cases, it is beneficial to re-optimize the *se*-QM/MM parameters[23,38] or directly modify the internal forces[39] to ensure a proper thermodynamic perturbation or interpolation correction or to maintain a stable multiple-timestep trajectory.

Recently, Yang,[40–42] Gastegger,[43] York,[44] Riniker[45] and coworkers have proposed machine learning (ML) as a new strategy to address the computational cost of direct *ai*-QM/MM free energy simulations. Specifically, for configurations of interest, artificial neural network (ANN) models were designed and trained to reproduce either the *ai*-QM/MM potential, i.e., the machine learning potential (MLP), or the difference between the *ai*-QM/MM and *se*-QM/MM potentials, hereafter, referred to as the delta machine learning potential (ΔMLP). Thus developed MLP or ΔMLP was then employed (in lieu of the *ai*-QM/MM potential) to drive the dynamical sampling of the enzyme system. These MLPs/ΔMLPs led to fairly accurate free energy barriers, with errors of around $1.0\,\mathrm{kcal/mol}$, for several solution-phase reactions.[40–42,44] In these approaches, however, the effects of solvent on the reacting system

were rather homogeneous, and their applicability to reactions in a heterogeneous solvent environment, such as in enzyme, has not been fully explored.

When compared to the construction of MLPs for gas-phase and small periodic systems, these are substantial achievements because the training of MLP for reproducing the *ai*-QM/MM potential is more challenging. Naively adding thousands of MM atoms to commonly-used ML descriptors of a gas-phase molecular system [such as symmetry functions for constructing high-dimensional neural network potentials (HDNNP) descriptors,[46–48] Coulomb matrix,[49] Faber-Christensen-Huang-Lilienfeld (FCHL) representation,,[50] tensor representation,[51] and embedding matrix[52,53]] would lead to an explosively large number of inputs for the ANN model, thus jeopardizing the convergence (in training) and affordability (for both training and production).

In the development of the above mentioned MLPs/ΔMLPs, two approaches have been proposed to efficiently include MM atoms in the ANN models. The first is an "implicit" approach, where MM atoms are implicitly accounted for through their collective perturbation to the electronic structure of the QM region. For example, Yang and coworkers[40,41] used semi-empirical QM Mulliken charges as perturbed by all the MM charges or the total MM electrostatic potential on each QM atom,[42] whereas Gastegger and coworkers[43] captured the effect of MM atoms through their net dipolar field on each QM atom. The second is an "explicit" approach, in which any MM atom within a "cutoff" distance (such as 6.0Å) from the QM region is included in the computation of ML descriptors. This approach was proposed by York and coworkers[44] for the QM/MM expansion to the embedding matrix within the DeePMD coding framework,[53,54] and also utilized by Riniker and coworkers[45] to extend the HDNNP descriptors to QM/MM calculations.

Inspired by these approaches to include MM environment in the MLP development, in this work, we aimed to develop a more robust protocol for training MLPs/ΔMLPs for free energy simulations of enzyme reactions by incorporating the effects of long-range MM electrostatic interactions, such as, under the periodic boundary conditions. Anticipating some amino acid side chains and/or solvent molecules to move in and out of the cutoff boundary with the progression of the reaction, we opted not to follow the "cutoff" approach, because the number of MM atoms retained in the cutoff might change, for example, between neighboring umbrella sampling windows and thus additional smoothing[55,56] might be needed to ensure a continuous potential energy surface. Furthermore, in enzymatic reactions, the long-range electrostatic interactions play a critical role in many catalytic reactions as well as in the stability of the molecular dynamics (MD) simulations. In addition, within the "implicit" approach, we opted not to utilize semi-empirical Mulliken charges on QM atoms, which would not be available if we aimed to directly produce the *ai*-QM/MM-quality MLP to propagate the MD trajectories. In the end, our approach would resemble that of Yang and coworkers[42] and of Gastegger *et al*,[43] but the details differed significantly as described briefly below.

Compared to previous efforts, three key features differentiate our approach from those of other groups. Firstly, long-range electrostatic effects of the MM environment is incorporated rigorously in the training of the MLP models. As will be demonstrated in Section 2, this is achieved by representing the enzyme/solvent environment as MM electrostatic potential and field (ESPF) within our recent QM/MM with augmentary charges (QM/MM-AC) scheme.[57] Secondly, a more reliable sampling/collection of training configurations is acquired by calibrating the *se*-QM/MM Hamiltonian. In the case of ΔMLP, it reduces the magnitude of the Hamiltonian differences and thus lessens the need of iterative training of the MLP model. Finally, standalone MLP and ΔMLP are developed side-by-side, thanks to the construction of a single set of ML descrip-

tors, thereby allowing a direct comparison of the two potentials for reactions of interest. With this development, we have simulated the catalytic reaction of an enzyme, chorismate mutase, which can be considered to be the first realistic and successful application of MLP/ΔMLP to model enzyme reactions under the full periodic boundary conditions.

In a separate manuscript from us,[58] a different *ai*-QM/MM-based machine learning approach was employed to improve free energy simulations of condensed-phase reactions. Specifically, ANN was trained to produce a set of chaperone polarizabilities that augment the insufficient polarizability of the AM1 semi-empirical Hamiltonian for modeling the Menshutkin reaction. With the corresponding polarization energy correction, accurate free energy barrier and reaction free energy were also obtained for the reaction.

This article is organized as follows. Section 2 introduces our overall methodology (incorporation of electrostatic embedding potential into the MLP, description of long-range QM/MM electrostatics, umbrella sampling), while more computational details for the training and free energy simulations are presented in Section 3. Results for solution-phase Menshutkin reaction and chorismate mutase-catalyzed Claisen rearrangement are presented in Section 4, with concluding remarks made in Section 5.

2 Machine Learning Potential

2.A Descriptors for the QM atoms

Our current MLP/ΔMLP implementation is based on Deep Potential - Smooth Edition (DeepPot-SE),[52,53] where each QM atom i is represented by its local environment matrix to capture the internal interactions within the QM region (with N QM atoms). Specifically, for the i-th atom in the QM region at position $\mathbf{R}_i$, its has $n = N - 1$ neighbors. The corresponding environment matrix (n-by-4) is

$$\mathscr{R}_i = \begin{pmatrix} s(R_{1i}) & s(R_{1i})\frac{x_{1i}}{R_{1i}} & s(R_{1i})\frac{y_{1i}}{R_{1i}} & s(R_{1i})\frac{z_{1i}}{R_{1i}} \\ s(R_{2i}) & s(R_{2i})\frac{x_{2i}}{R_{2i}} & s(R_{2i})\frac{y_{i1}}{R_{2i}} & s(R_{2i})\frac{z_{2i}}{R_{2i}} \\ \cdots & \cdots & \cdots & \cdots \\ s(R_{ni}) & s(R_{ni})\frac{x_{ni}}{R_{ni}} & s(R_{ni})\frac{y_{ni}}{R_{ni}} & s(R_{ni})\frac{z_{ni}}{R_{ni}} \end{pmatrix}, \quad (1)$$

where $s(R_{ji}) = 1/R_{ji} = 1/|\mathbf{R}_j - \mathbf{R}_i|$ contains the Coulomb interaction. No screening was applied to the Coulomb interaction, due to the small size of the QM region. Next, an embedding neural network, $\mathscr{G}_i$, maps a single value, $s(R_{ji})$, through multiple hidden layers into m_1 outputs

$$\mathscr{G}_i = \begin{pmatrix} (G[s(R_{1i})])_1 & (G[s(R_{1i})])_2 & \cdots & (G[s(R_{1i})])_{m_1} \\ (G[s(R_{2i})])_1 & (G[s(R_{2i})])_2 & \cdots & (G[s(R_{2i})])_{m_1} \\ \cdots & \cdots & \cdots & \cdots \\ (G[s(R_{ni})])_1 & (G[s(R_{ni})])_2 & \cdots & (G[s(R_{ni})])_{m_1} \end{pmatrix}. \quad (2)$$

The encoded feature matrix $\mathscr{D}_i$ is m_1-by-m_2

$$\mathscr{D}_i = \left(\mathscr{G}_i^1\right)^T \mathscr{R}_i \mathscr{R}_i^T \mathscr{G}_i^2, \quad (3)$$

where $\mathscr{G}_i^1$ is the same as $\mathscr{G}_i$ and $\mathscr{G}_i^2$ is a matrix that consists of the first m_2 columns of $\mathscr{G}_i$ with $m_2 \leq m_1$. Both m_1 and m_2 are hyperparameters of the model.

2.B Descriptors for the MM environment

For QM/MM interactions, our model closely resembles the regular QM/MM models, where the MM atoms are represented as point charges with no atomic identities for the QM-MM electrostatic interactions, and the QM-MM van der Waals (vdW) interactions are treated at the MM level.[4,7,59–61]

In regular QM/MM models, there are three general schemes to treat the QM-MM electrostatic interactions, namely, the continuous, the surrogate, and the hybrid schemes,[57] depending on whether the continuous electron density or its surrogate charges or a combination interact with the MM atomic charges. Since electrons are not described explicitly in most MLPs, it is natural to adopt a surrogate-like scheme in MLP where QM atoms interact with MM atoms through the electrostatic potential and field generated,

$$\phi_i = \sum_{B \in \text{MM}} \frac{q_B}{|\mathbf{R}_i - \mathbf{R}_B|}, \tag{4}$$

$$\mathscr{E}_i = \sum_{B \in \text{MM}} \frac{q_B(\mathbf{R}_i - \mathbf{R}_B)}{|\mathbf{R}_i - \mathbf{R}_B|^3}, \tag{5}$$

or higher-order Taylor expansions at the QM atomic sites. In contrast to normal QM/MM calculations, where one either evaluates the core Hamiltonian contribution associated with ϕ_i and $\mathscr{E}_i$ or derives multipole moments on each QM atom to interact with them, we choose to add the electrostatic potential ϕ_i and field $\mathscr{E}_i$ on each QM atom directly to the list of the input features of the MLP.

Specifically, the contributions of the local electrostatic potential, ϕ_i, and field, $\mathscr{E}_i$, are fed into embedding neutral networks similar to Eq. 2, leading to additional feature matrices, $\mathscr{D}_i^\phi$ and $\mathscr{D}_i^\mathscr{F}$, to capture the effect of the MM environment. For example, for the local electrostatic potential, a single value ϕ_i is mapped to m_1' outputs through multiple hidden layers, which is used as the feature matrix $\mathscr{D}_i^\phi$, i.e.,

$$\mathscr{D}_i^\phi = \begin{pmatrix} (G(\phi_i))_1 \\ (G(\phi_i))_2 \\ \cdots \\ (G(\phi_i))_{m_1'} \end{pmatrix}. \tag{6}$$

The electric field on each QM atom is projected onto the vectors pointing toward its neighbors and scaled by the distance, i.e.,

$$\mathscr{F}_i^\mathscr{F} = \begin{pmatrix} s(r_{1i}) \left(\frac{\mathbf{R}_{1i}}{R_{1i}} \right) \cdot \mathscr{E}_i \\ s(r_{2i}) \left(\frac{\mathbf{R}_{2i}}{R_{2i}} \right) \cdot \mathscr{E}_i \\ \cdots \\ s(r_{ni}) \left(\frac{\mathbf{R}_{ni}}{R_{ni}} \right) \cdot \mathscr{E}_i \end{pmatrix}. \tag{7}$$

Then, each element of the projected local field $\mathscr{F}_i^\mathscr{F}$ is mapped to m_1'' outputs through multiple hidden layers, i.e.,

$$\mathscr{G}_i^\mathscr{F} = \begin{pmatrix} (G(\mathscr{F}_{1i}))_1 & (G(\mathscr{F}_{1i}))_2 & \cdots & (G(\mathscr{F}_{1i}))_{m_1''} \\ (G(\mathscr{F}_{2i}))_1 & (G(\mathscr{F}_{2i}))_2 & \cdots & (G(\mathscr{F}_{i2}))_{m_1''} \\ \cdots & \cdots & \cdots & \cdots \\ (G(\mathscr{F}_{ni}))_1 & (G(\mathscr{F}_{ni}))_2 & \cdots & (G(\mathscr{F}_{ni}))_{m_1''} \end{pmatrix}. \tag{8}$$

The final feature matrix is a m_1''-by-m_2'' matrix, which can be calculated as

$$\mathscr{D}_i^\mathscr{F} = \left(\mathscr{G}_i^{\mathscr{F}1} \right)^T \mathscr{G}_i^{\mathscr{F}2}, \tag{9}$$

where $\mathscr{G}_i^{\mathscr{F}1}$ is the same as $\mathscr{G}_i^\mathscr{F}$ and $\mathscr{G}_i^{\mathscr{F}2}$ is a matrix that consists of the first m_2'' columns of $\mathscr{G}_i^\mathscr{F}$. In Eqs. 6 and 9, m_1', m_1'', and m_2'' are hyperparameters of the model.
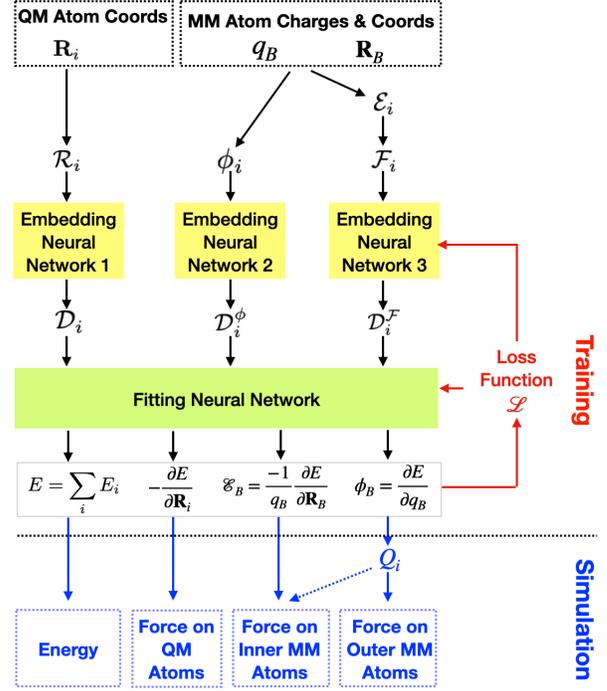
2.C Fitting Neural Network

Figure 1: Workflow for the training of MLP and its use to generate energy and forces for MD simulations.

After the three sets of feature matrices (i.e., Eqs. 3, 6 and 9) are obtained from the embedding neural networks, they are fed into a fitting neural network (NN) to obtain the total energy,

$$E = \sum_i \text{NN}(\mathscr{D}_i, \mathscr{D}_i^\phi, \mathscr{D}_i^\mathscr{F}). \tag{10}$$

With a differentiable neural network model, the analytical energy derivatives in Fig. 1 are also obtained. During the standalone MLP model training, the loss function combines the error in total energy, forces on QM atoms, as well as ESP potential/field values on MM atoms,

$$\mathscr{L} = \lambda_1 \left| E - E^{\text{ref}} \right|^2 + \lambda_2 \sum_{i \in \text{QM}} \left| -\frac{\partial E}{\partial R_i} + \frac{\partial E^{\text{ref}}}{\partial R_i} \right|^2$$
$$+ \lambda_3 \sum_{B \in \text{MM}} \left| \frac{\partial E}{\partial q_B} - \phi_B^{\text{ref}} \right|^2 + \lambda_4 \sum_{B \in \text{MM}} \left| -\frac{1}{q_B} \frac{\partial E}{\partial R_B} - \mathscr{E}_B^{\text{ref}} \right|^2, \tag{11}$$

where the learned forces on QM atoms, $-\frac{\partial E}{\partial R_i}$, and ESP potential/field values, $\frac{\partial E}{\partial q_B}$ and $-\frac{1}{q_B} \frac{\partial E}{\partial R_B}$, are obtained from the differentiable NN model in use. The same feature matrices can be applied to the development of ΔMLP model, in which the energy, force, and ESP differences between ai-QM/MM and se-QM/MM models are used as the reference values in Eq. 11. In both MLP models, λ_1, λ_2, λ_3, and λ_4 are tunable prefactors and can be adjusted during the model training.

2.D Long-Range QM/MM Interactions

For a condensed-phase reaction, the MM environment contains thousands of (or more) atoms. An explicit account of all these atoms in the MLP training becomes very inefficient, if feasible at all. A common strategy around this is to completely ignore the long-range electrostatics due to MM atoms beyond a cutoff distance.[44,45] In the context of ΔMLP-learning, this strategy means that the long-range electrostatics of ΔMLP is approximated at the low-level method. However, a systematic analysis has yet to be carried out to gauge the impact of solvent/enzyme atoms moving across the cutoff boundary during the simulation on the developed ML potentials, as it can contribute to a discontinuity on the potential energy surface.

In lieu of this approximation, we have decided to adopt our recent QM/MM with Augmented Charges (QM/MM-AC) model[57] to accurately capture long-range QM/MM electrostatics in this work. Specifically, only inner MM atoms, namely those within a cutoff distance (10Å in this work) from the QM region, are explicitly incorporated into the training of MLP and ΔMLP. This can be easily achieved by replacing the MM charges in Eqs. 4 and 5, with

$$ q_B \longrightarrow q_B^{<,C} + q_B^{AC}, \tag{12} $$

where $q_B^{<,C}$ refer to the "continuous" portion of charges on inner MM atoms and q_B^{AC} are the augmentary charges projected on the inner MM atoms from all outer MM charges and the "surrogate" portion of charges on the inner MM atoms. We refer to Ref. 57 for details of the QM/MM-AC model and how the projection is carried out.

Thus, the QM/MM-AC charges, q_B^{AC} in Eq. 12, are pre-computed for inner MM atoms of each configuration before our training of MLP and ΔMLP. After these models are constructed, the electrostatic potential on inner MM atom positions, $\phi_B = \frac{\partial E}{\partial q_B}$, is used to fit ESP charges on all QM atoms. As shown in Fig. 1, these charges are then used to provide the electrostatic potential to update the forces on outer MM atoms, through the procedure described in Ref. 57.

2.E Implementation and Training Details

Our method, which combines the Deep Potential – Smooth Edition (DeepPot-SE)[52] descriptors for QM atoms and QM/MM-AC-based descriptors for the MM environment, was implemented in PyTorch.

In this work, the hyperparameters were not fully optimized, where the recommended values from the DeePMD-kit package were adopted where applicable. For each of the local embedding networks, three hidden layers with 25, 50, and 100 (m_1) neurons, were used. For each of the electrostatic potential embedding networks and the electric field embedding networks, three hidden layers with 5, 10, and 20 (m_1' and m_1'') neurons were used. For the local embedding networks and the electric field embedding networks, the size of the axis filters were chosen to be 4 (m_2 and m_2''). For the fitting networks, three hidden layers, each of which has 240 neurons, were used.

The (standalone) MLP model was trained using the Adam optimizer. During the training, the batch size was set to 32, and the initial learning rate was set to 10^{-4}. Following DeePMD-kit,[53] the prefactor for the energy error in the loss function in Eq. 11 was much smaller than the other three at the beginning, and they gradually evolved to the same value towards the end of the training. The learning rate decayed exponentially by a factor of 0.95 every epoch. A total of 100 epochs were performed for each system. A separate ΔMLP model was also trained to reproduce the difference between the PM3*/MM and B3LYP/6-31G*/MM models using the same architecture and hyperparameters as the standalone MLP model.

For each of the two reactive systems (i.e., the solution-phase Menshutkin reaction system and the chorismate mutase reaction system), 40,000 frames were collected from the PM3*/MM simulations (see the next Section for details) and then randomly split into sets of 38,400 and 1,600 samples for training and validation, respectively. The testing set consisted of 2,000 additional configurations sampled along the reaction pathway from the *ai*-QM/MM simulations. For each sample in the training/validation sets, B3LYP/6-31G*/MM single point calculations were performed. The QM/MM-AC method[57] was used to project the MM charges that were more than 10 Å away from any QM atom and all the charges from the periodic images onto the MM atomic sites within the 10 Å cutoff. Besides the energies and the gradients of the QM atoms, the electrostatic potentials and fields on the inner MM atomic sites were also collected.

3 Simulation Details

3.A System Setup and Equilibration

In this work, we started with the Menshutkin reaction, a widely-used model system for solution chemistry, for developing the protocol of training the MLP and ΔMLP. Then the protocol is applied to the chorismate mutase reaction, which is a popular system to test new enzyme simulation methods because of its small QM region (24 QM atoms) and because the QM region is not covalently linked to the MM region.
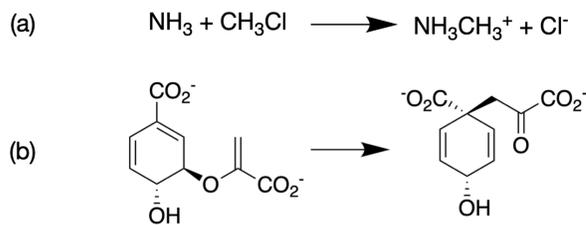

Figure 2: Schemes for (a) Menshutkin and (b) chorismate mutase reactions.

For the Menshutkin reaction (Figure 2a), the solutes (ammonia and chloromethane) were solvated in a cubic box of 723 TIP3P[62] waters, and modeled by the general Amber force field (GAFF).[63] For the chorismate mutase reaction (Figure 2b), the initial structure was prepared based on the X-ray crystal structure (PDB ID: 2CHT[64]) of *Bacillus subtilis* chorismate mutase complexed with a transition state analog, which was modified to the substrate chorismate manually, and solvated in a cubic box of 13,067 TIP3P[62] waters. 8 sodium counter ions were added to neutralize the system. The substrate and the enzyme were modeled using GAFF[63] and the AMBER ff14SB force field,[65] respectively.

Both systems were equilibrated at 300 K and 1 atm using Langevin dynamics with a friction coefficient of $5\,ps^{-1}$ and Berendsen barostat with a relaxation time of 1 ps under the periodic boundary conditions. The solutes were restrained to their initial positions by a weak harmonic potential for the Menshutkin reaction, whereas no restraints were applied for the enzymatic system. The particle mesh Ewald (PME) summation method[66,67] was used to treat the electrostatic interactions, while the van der Waals interactions were turned off to zero smoothly at a cutoff of 10 Å. The SHAKE algorithm[68] was used to constrain all bonds involving hydrogen atoms, and a time step of 2 fs was used for the MD integration using the leapfrog integrator. After equilibration, the sizes of the simulation boxes were $\sim 28\,\text{Å} \times 28\,\text{Å} \times 28\,\text{Å}$ and $\sim 76\,\text{Å} \times 76\,\text{Å} \times 76\,\text{Å}$ for the Menshutkin and chorismate mutase

reactions, respectively, which were used for the subsequent simulations in the NVT ensemble. The classical MD simulations were performed using the PMEMD program from the Amber20 package.[69]

3.B QM/MM Calculations

To simulate the bond breaking/forming process, a QM/MM description of the system was needed. To sample at the target level of theory within a moderate amount of computer time (<500,000 CPU hours), the B3LYP functional[70–72] and 6-31G* basis set[73] (in conjugation with the MM force field) were chosen as the reference *ai*-QM/MM method for the two test systems in this study. On the other hand, the PM3*/MM model, where the parameters of the standard PM3 method[74] were recalibrated through force-matching,[23] was used to simulate the reactions under study. We note that despite the same notation, the PM3* model parameters are different between the two reaction systems. For the Menshutkin reaction, the parameters were directly taken from our previous study,[23] whereas for the chorismate mutase reaction, an improved version of the parameter set were used (Table S1).

The setup of the QM/MM MD simulations was overall similar to the classical ones. During the QM/MM MD simulations, the solutes and the substrate were described by the QM method, while the rest of the system (solvent or protein) was described by the force fields used in the classical simulations. The QM/MM-AC method[57] was used to capture the long-range QM-MM electrostatic interactions. The SHAKE algorithm[68] was only applied to the MM subsystem, and the integration time step was set to 1 fs. The QM/MM MD simulations were performed using our QM/MM interface QMHub (https://github.com/panxl/qmhub) and a modified version of the SANDER program from the AmberTools20 package.[69] All DFT/MM calculations were performed using Q-Chem 5.2.[75]

3.C Umbrella Sampling

To achieve good coverage of all configurations relevant to the reactive process, enhanced sampling MD simulations are needed. Umbrella sampling[14] is a method that can enhance the sampling of the system along one or more predefined reaction coordinates by adding harmonic biasing potentials to restrain the system to the region of interest of the reaction coordinate. Typically, a series of restraint centers ξ_i^0 are determined to cover the region of interest of the reaction coordinate. Then, for each restraint center ξ_i^0 (also called window), separate simulations are conducted with the harmonic biasing potential, in the form of,

$$E_{\text{bias},i}(\xi) = k_i(\xi - \xi_i^0)^2, \qquad (13)$$

added to the Hamiltonian of the system, where k_i is the force constant of the harmonic potential at window i and ξ refers to the reaction coordinate. In practice, the number and locations of windows and the force constants are determined to ensure sufficient overlapping of the sampled configurations between neighboring windows while minimizing the computational cost. In this study, we also applied the Hamiltonian replica exchange molecular dynamics (HREMD)[76] technique to accelerate the convergence of the free energy simulation, in which the exchanges of the biasing potentials between the neighboring windows were attempted at a fixed interval of steps set to 100 fs.

To generate the training/validation sets that cover all the relevant configuration space for the reactions, the above-mentioned umbrella sampling technique was used to collect configurations along the predefined reaction coordinates, using the PM3*/MM potential. For both reaction systems, the reaction coordinates were de-

fined as $\xi = d_{\text{break}} - d_{\text{form}}$, where d_{break} and d_{form} were the bond lengths of the breaking and forming bonds, respectively. Specifically, $\xi = d_{\text{C–Cl}} - d_{\text{C–N}}$ for Menshutkin reaction and $d_{\text{C–O}} - d_{\text{C–C}}$ for chorismate mutase reaction (see Figure 2). For each reaction system, a total of 80 windows were evenly distributed with an interval of 0.05 Å to cover ξ ranged from -1.975 to 1.975 Å, and the force constants were set to be $150\,\text{kcal/mol/Å}^2$ for all the windows. For each window, 50 ps simulation was conducted, and the configurations were saved every 0.1 ps, which resulted in 500 configurations. Overall, 40,000 frames were collected for each reaction system, and for each configuration saved, the B3LYP/6-31G*/MM calculation was performed to generate the reference data.

After the ML model was trained, the same umbrella sampling simulations were performed where PM3*/MM was replaced by the MLP/MM or PM3*+ΔMLP/MM model. Multistate Bennett acceptance ratio (MBAR)[77] method as implemented in the pymbar package (https://github.com/choderalab/pymbar) was used to compute the free energy profile.

4 Results and Discussion

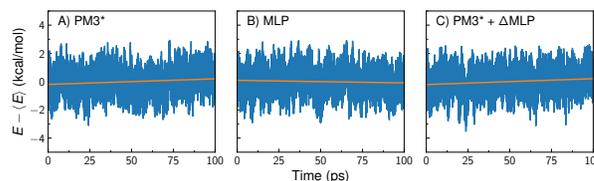## 4.A Energy Conservation During Microcanonical Ensemble Simulations

Figure 3: Conservation of the total energy in 100 ps NVE simulations of the chorismate mutase reaction using A) PM3*, B) MLP, and C) PM3*+ΔMLP models. In each figure, the line shown in orange indicates the drift of energy (see the values mentioned in the main text for each model).

As a validation of the developed MLP and ΔMLP models, especially, the conservation of the total energy, we first performed 100 ps microcanonical ensemble (NVE) MD simulations for the reactant of the chorismate mutase reaction using the PM3*/MM, MLP/MM, and PM3*+ΔMLP/MM models. The results are presented in Figure 2. Overall, a good conservation of energy was observed, with the energy drifts being $3.96 \pm 0.08 \times 10^{-3}\,\text{kcal/mol/ps}$, $-1.81 \pm 0.08 \times 10^{-3}\,\text{kcal/mol/ps}$, and $4.17 \pm 0.08 \times 10^{-3}\,\text{kcal/mol/ps}$ for the PM3*/MM, MLP/MM, and PM3*+ΔMLP/MM models, respectively. In addition, we compared the initial 500 fs trajectories with an NVE trajectory using B3LYP/6-31G*/MM under the same initial conditions (Figure S2). It showed that MLP and PM3*+ΔMLP models diverged less from the DFT/MM than the PM3* model.

4.B Overall Free Energy Results

The overall free results for the two reactive systems studied are summarized in Tables 1. For the aqueous Menshutkin reaction, the B3LYP/MM free energy barrier was calculated to be $15.5 \pm 0.1\,\text{kcal/mol}$ and the free energy of reaction was $-26.9 \pm 0.1\,\text{kcal/mol}$, respectively. Despite extensive reparameterization as outlined in Section 3.B, the PM3*/MM model still substantially overestimated the barrier height with a value of $23.3 \pm 0.1\,\text{kcal/mol}$. Similarly, the free energy of reaction was overestimated by $18.6\,\text{kcal/mol}$. In contrast, much

Table 1: Free Energy Barriers and Reaction Free Energies for Menshutkin (MEN) and Chorismate Mutase (CM) Reactions

	Free Energy Barrier (kcal/mol)				Reaction Free Energy (kcal/mol)			
	PM3*	MLP[a]	PM3* + ΔMLP	B3LYP	PM3*	MLP[a]	PM3* + ΔMLP	B3LYP
MEN	23.3 ± 0.1	16.1 ± 0.1	15.1 ± 0.1	15.5 ± 0.1	-8.3 ± 0.1	-25.9 ± 0.1	-27.5 ± 0.1	-26.9 ± 0.1
CM	14.9 ± 0.1	13.9 ± 0.1	13.9 ± 0.1	13.6 ± 0.1	-15.5 ± 0.1	-16.3 ± 0.1	-16.6 ± 0.1	-16.8 ± 0.1

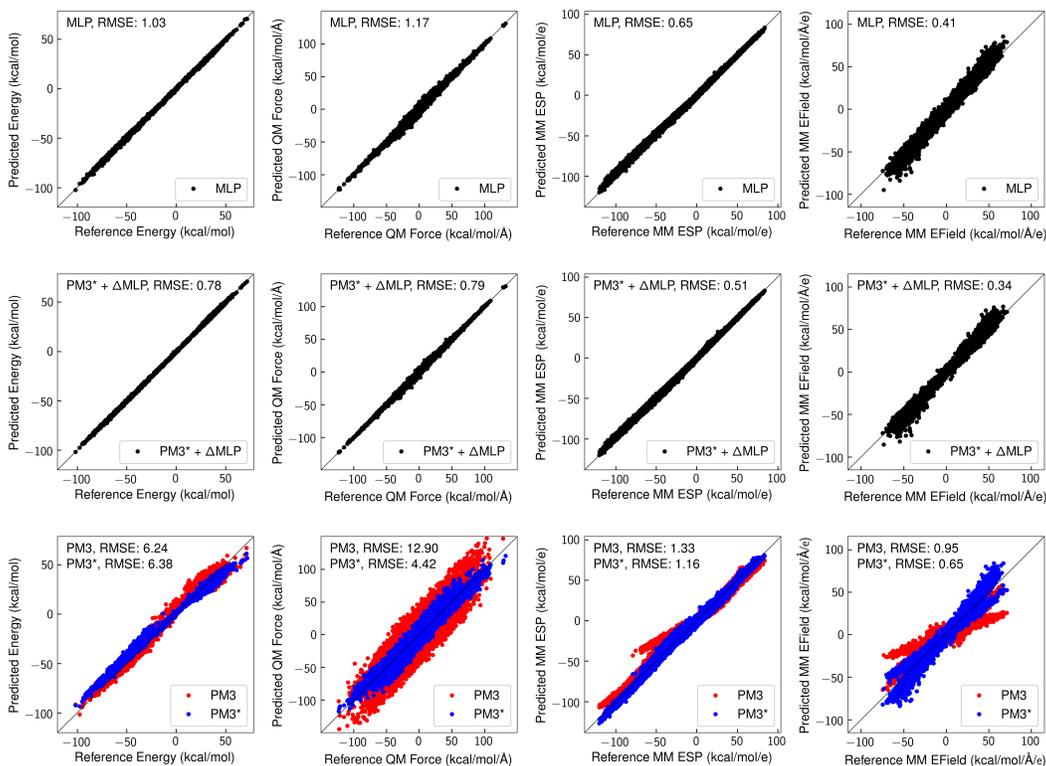[a] For the chorismate mutase reaction, the reported results are from the 2nd iteration MLP.

Figure 4: Accuracy of MLP (top), PM3*+ΔMLP (middle), PM3 and PM3* (bottom) energy, forces, electrostatic potential (ϕ) and electric field ($\mathscr{E}$) for the 2,000 testing configurations for aqueous Menshutkin reaction. In each figure, the reference values are obtained from the B3LYP/6-31G*/MM calculations. The root-mean-square error (RMSE) value is also shown for each method.

Table 2: CPU Time (in 10^3 h) for the Computation of Free Energy Profiles for Menshutkin (MEN) and Chorismate Mutase (CM) Reactions[a]

	PM3*	MLP	PM3* + ΔMLP	B3LYP
MEN	0.3	0.8	1.3	12.4
CM	1.5	11.9[b]	8.2	379.5

[a] Umbrella sampling simulations were performed for 50 ps per window for 80 windows. Thus, a total of 4 ns umbrella sampling MD simulations were performed for each QM/MM model.
[b] Including time from two iterations. Thus, a total of 100 ps for each umbrella sampling window.

improved free energy barriers and reaction free energies were obtained with MLP sampling (16.1 ± 0.1 kcal/mol for barrier and -25.9 ± 0.1 kcal/mol for the reaction free energy) and also with PM3*+ΔMLP sampling (15.1 ± 0.1 kcal/mol for barrier and -27.5 ± 0.1 kcal/mol for the reaction free energy), both within 1.0 kcal/mol from the reference B3LYP values.

Similar improvements were found for the chorismate mutase reaction. Both MLP and PM3*+ΔMLP simulations reproduced the barrier height of 13.6 ± 0.1 kcal/mol and the free energy of reaction of -16.8 ± 0.1 kcal/mol, both with errors less than 0.5 kcal/mol from the reference B3LYP values. Considering the length of the sampling for each simulation, the three results can be considered to be essentially the same. On the other hand, while remaining excellent, the PM3* model produced a barrier of 14.9 ± 0.1 kcal/mol and a reaction free energy of -15.5 ± 0.1 kcal/mol, which were only 1.3 kcal/mol higher than their corresponding B3LYP values.

We want to note that our standalone MLP and ΔMLP models have not been fully optimized in terms of their computational cost. With the current set of hyperparameters (as specified in Section 2.E), the MLP sampling of the Menshutkin reaction system took about 800 CPU hours, which was 3-fold higher than the PM3* sampling, but it was still 16 times faster than the B3LYP simu-
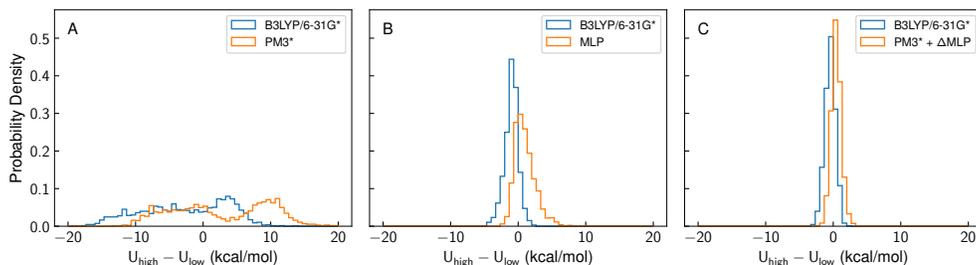
Figure 5: Distribution of high-level (B3LYP/6-31G*) and low-level (PM3*, MLP, and PM3*+ΔMLP energy differences for configuration collected from B3LYP/MM MD trajectories (blue) or low-level MD trajectories (orange) of the aqueous Menshutkin reaction.

lation (Table 2). For PM3*+ΔMLP, since both PM3* and ΔMLP energy/force evaluation were performed for each configuration, its cost increased to around 1,300 CPU hours, which was still 10 times faster than the reference B3LYP simulation. It should be noted that, our MLP and ΔMLP models were implemented using the Py-Torch framework which supports both CPU and GPU parallelization. However, the timing reported here was only based on CPU calculations.

For the chorismate mutase system, the MLP sampling took 11,900 CPU hours, which was about 8 times slower than the PM3* sampling. The higher MLP/PM3* timing ratio arose largely from two rounds of MLP training performed to obtain a stable MLP model around the reaction transition state region (see Section 4.D for details). In comparison to the 379,500 CPU hours for the B3LYP sampling, the MLP sampling offered a 32-fold speedup. With PM3*+ΔMLP, only one round of ML model training was needed, which led to a net 8,200 CPU hours for the sampling. This was 46 times faster than the reference B3LYP simulation.

Below we go over each reactive system in more detail, and show how such speedups were obtained.

4.C Menshutkin Reaction

For Menshutkin reaction, the standard PM3 model (marked red), which is shown in the bottom panels in Fig. 4, deviates substantially from the reference B3LYP values. In terms of the root-mean-square error (RMSE), PM3 energy differed by 6.24 kcal/mol, while the corresponding force error on QM atoms was 12.90 kcal/mol/Å (Table S2). Similarly, the mean-unsigned-error (MUE) for energy was 5.00 kcal/mol and for force was 9.62 kcal/mol/Å. The RMS errors in the QM electrostatic potential and field on MM atom position were substantially smaller at 1.33 kcal/mol/e and 0.95 kcal/mol/Å/e, respectively, and their corresponding MUE values were 0.67 kcal/mol/e and 0.24 kcal/mol/Å/e. The maximum errors (MAXs) were largest for PM3 among the test QM/MM models: 21.55 kcal/mol for energy, 54.32 kcal/mol/Å for force, 38.41 kcal/mol/e for electrostatic potential and 49.39 kcal/mol/Å/e for the field, respectively.

Through our automated reparameterization (which placed a larger weight on the force errors), the PM3* model (marked blue in bottom panels in Figure 4) substantially reduced the force error to 4.42 kcal/mol/Å for RMSE and 3.21 kcal/mol/Å for MUE, respectively. Meanwhile, the PM3* energy error remained at 6.38 kcal/mol for RMSD and 5.49 kcal/mol for MUE, which was evident from Fig. 8A with a broad distribution of the energy difference between the PM3* and B3LYP models.

When ΔMLP was trained, it was designed to reproduce the difference between the PM3* and B3LYP potentials. As shown in the middle panels in Fig. 4, the energy error of the combined PM3*+ΔMLP model reached 0.78 kcal/mol for RMSE and 0.62 kcal/mol for MUE, which were below 1.0 kcal/mol error and
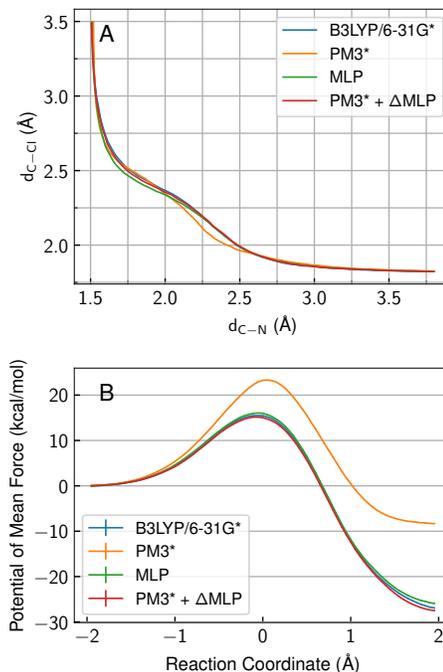

Figure 6: (A) Sampled pathway and (B) potential of mean force for the aqueous Menshutkin reaction based on umbrella sampling using PM3*+ΔMLP and MLP potentials in comparison to PM3* and B3LYP/6-31G* results.

lowest among the models tested. The reduction in the force error was more impressive, reaching a RMSE value of 0.79 kcal/mol/Å and a MUE value of 0.53 kcal/mol/Å. A moderate but more systematic improvement of the ESP and field values on MM atom position could also be noticeable. As a consequence, the distribution of PM3*+ΔMLP and B3LYP energy differences as shown in Fig. 5C became much narrower. The corresponding standard deviation of energy difference was 0.7 kcal/mol, which fell within the 1.7–2.5 kcal/mol threshold for a lower-level theory to accurately reproduce the sampling from a high-level model.[78–80]

Alternatively, MLP can be directly trained to reproduce the B3LYP energy/force as well as the QM ESP and field on MM atom positions. As shown in the top panel of Fig. 4, a standalone MLP can reproduce the B3LYP energy with 1.03 kcal/mol RMSE and 0.79 kcal/mol MUE and the forces with 1.17 kcal/mol/Å RMSE and 0.75 kcal/mol/Å MUE, respectively, which were only slightly worse than PM3*+ΔMLP. Therefore, it led to a narrow distribution of energy differences in Fig. 5B, with a slightly larger standard deviation of 1.5 kcal/mol than that of PM3*+ΔMLP.
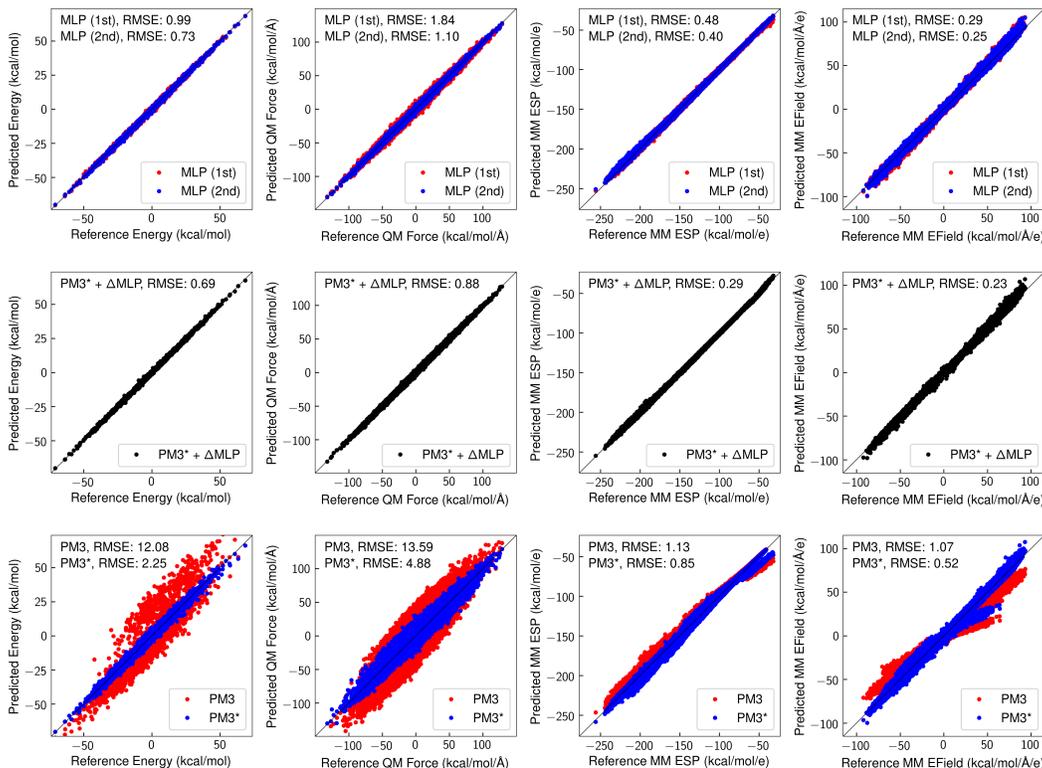
Figure 7: Accuracy of MLP (top), PM3*+ΔMLP (middle), PM3 and PM3* (bottom) energy and forces for the 2,000 testing configurations for the chorismate mutase reaction.
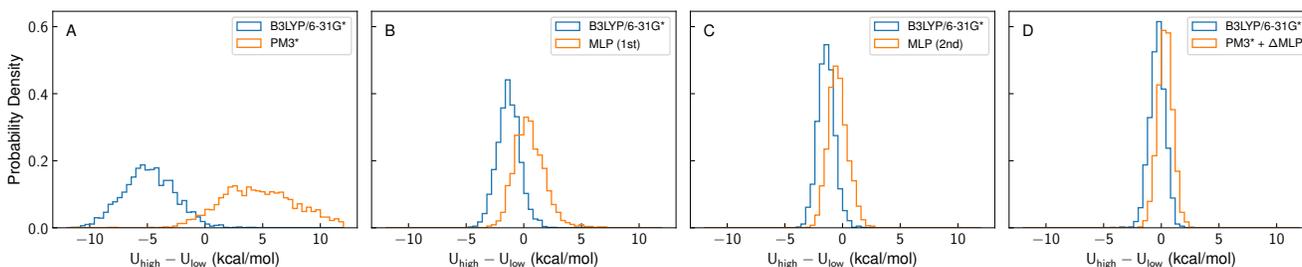

Figure 8: Distribution of high-level (B3LYP/6-31G*) and low-level (PM3*, MLP, and PM3*+ΔMLP energy differences for configuration collected from B3LYP/MM MD trajectories (blue) or low-level MD trajectories (orange) of chorismate mutase reaction.

Due to their capability to reproduce B3LYP energy and forces, both MLP and PM3*+ΔMLP models led to proper sampling of the pathway (shown in Fig. 6A) and thus accurate reaction free energy profile for the Menshutkin reaction. In addition, we expect that the conformational space sampled by both models would overlap very well with that of B3LYP.

4.D Chorismate Mutase Catalysis

Similar to the case of Menshutkin reaction in an aqueous solution, the chorismate mutase system was poorly described by the PM3 model, with an energy error of 12.08 kcal/mol and a force error of 13.59 kcal/mol/Å shown in the bottom panel of Fig. 7. A substantial improvement to the PM3 model was achieved through reparameterization, which reduced the energy error to 2.25 kcal/mol and the force error to 4.88 kcal/mol/Å. The improvement of the PM3* model is also noticeable from the comparison of MUE and MAXE values in Table S3.

The modeling of the chorismate mutase system was brought within 1.0 kcal/mol with PM3*+ΔMLP, whose RMS energy error was found to be 0.69 kcal/mol and RMS force error was 0.88 kcal/mol/Å. This was also evident in the change of the wide PM3*–B3LYP energy distribution in Fig. 8A to the narrow peaks for ΔMLP-corrected model in Fig. 8D. In addition, we point out the substantial reduction of MAXE for the forces from 33.29 kcal/mol/Å for PM3* to 7.18 kcal/mol/Å for PM3*+ΔMLP, suggesting very robust improvement of the accuracy of the PM3*+ΔMLP model.

With standalone MLP, one round of training using the configuration collected from the PM3*/MM MD trajectory did yield a good agreement with the target B3LYP/MM energy (RMSE: 0.99 kcal/mol and MUE: 0.78 kcal/mol). However, the force error was found to be 1.84 kcal/mol/Å for RMSE and 1.36 kcal/mol/Å for MUE. The MAXE of 14.78 kcal/mol/Å was also noticeably large. With such larger errors in the force, we observed some level of instability of the MD simulations using this MLP, especially near the transition state region. To address this issue, we constructed an expanded pool of configurations combining the configurations used

in the first MLP training and the configurations selected from this round of MLP sampling, and performed a second round of MLP training using the expanded configuration pool. Previously, Pu *et al.* has shown that the iterative parameterization of the *se*-QM model can systematically improve the accuracy of the (reparameterized) QM model[38] and the present second round of MLP training can be considered to be a second iteration of the MLP model training. The resultant MLP model, referred here to as MLP(2nd), further reduced the RMS force error to $1.10\,\mathrm{kcal/mol/\text{Å}}$ (and MUE to $0.81\,\mathrm{kcal/mol/\text{Å}}$), which was accompanied by a RMS energy error of $0.73\,\mathrm{kcal/mol}$ and MUE of $0.58\,\mathrm{kcal/mol}$, respectively. More impressively, the MAX force error was reduced to $8.32\,\mathrm{kcal/mol/\text{Å}}$, which was only $1.1\,\mathrm{kcal/mol/\text{Å}}$ larger than the MAX force error of PM3*+ΔMLP. The MLP(2nd) model also demonstrated a much narrower MLP(2nd)–B3LYP energy distribution than the one for MLP(1st)–B3LYP one as shown in Fig. 8.

These MLP(2nd) and PM3*+ΔMLP models yielded an accurate sampling of the reaction pathway shown in Fig. 9A. This is in contrast to the PM3* model, whose sampling was shown in Fig. 8d of Ref. 23 to deviate substantially from the B3LYP pathway between the transition and the product. As a result of the accurate energetics and sampling, both MLP(2nd) and PM3*+ΔMLP models produced free energy profiles of the chorismate mutase reaction that matched the B3LYP profile in Fig. 9B.
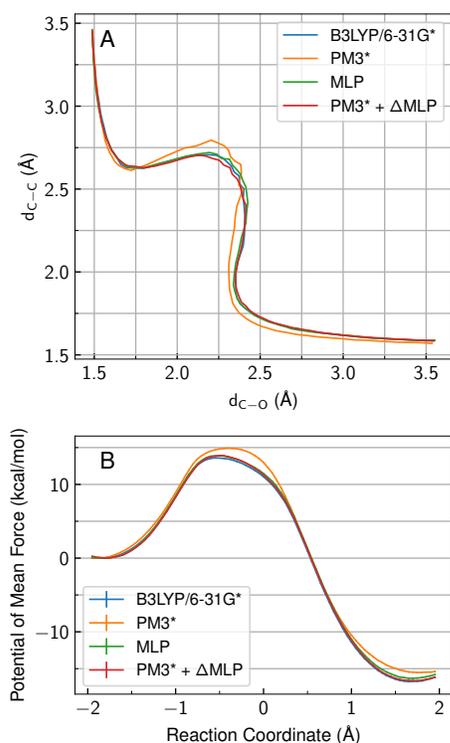

Figure 9: (A) Sampled pathway and (B) potential of mean force for the chorismate mutase reaction based on umbrella sampling using PM3*+ΔMLP and MLP (the 2nd iteration) potentials in comparison to B3LYP/6-31G* results.

5 Conclusions

Inspired by the works from Yang, York, Gastegger, Riniker and coworkers, we explored the construction of *system-specific machine-learning potentials (MLPs)* to accelerate the generation of *ab initio*-quality QM/MM free energy profiles of enzymatic reactions. A couple of advances were made in this work:

- The deep-potential of E and coworkers (i.e., DeepPot-SE) was combined with our QM/MM-AC electrostatic model to provide descriptors for a QM reactive region embedded in an enzyme or solution environment (described by MM point charges).

- A reparameterized semi-empirical quantum model, PM3*, was utilized to generate the configurations for training the atomistic machine-learning potentials. The deep-potential and QM/MM-AC descriptors of these configurations were then used to train the embedding and fitting neural networks to provide MLPs or ΔMLPs to propagate MD trajectories of condensed-phase reactive systems.

- The ΔMLP model, when combined with the PM3* model in umbrella simulations, produced accurate reaction free energies and barrier heights of two model reactions (aqueous Menshutkin reaction and chorismate mutase reaction) within $1.0\,\mathrm{kcal/mol}$. A 10–fold or 46–fold reduction in CPU time, respectively, was achieved in comparison to reference B3LYP/6-31G*/MM free energy simulations.

- The standalone MLP, which sought to directly reproduce the B3LYP energy/force on QM atoms and fitted electrostatic potentials and fields on MM atoms, could reproduce the aqueous Menshutkin free energy profile with a 16-fold CPU timing saving. For the chorismate mutase reaction, however, two rounds of MLP training was found to be necessary to ensure stable MD trajectories. With a larger QM region in chorismate mutase reaction, the timing saving was 32–fold.

On the other hand, several limitations need to be addressed in the future:

- The hyperparameters for our MLP and ΔMLP (which were fixed in this work) are needed to be systematically explored to find an optimal set of hyperparameters for each machine learning potential for efficiency and accuracy.

- In our current protocol, the samples from one or more iterations were all labelled using the high-level method. Active learning or concurrent learning[44,54] as well as a more robust data clustering algorithm can be adopted to reduce the cost of both data labelling and training.

- Our protocol for reparameterizing semi-empirical QM models and for training MLPs needs to be tested on typical enzyme reactions, where the computational savings are expected to be even more substantial with a larger QM region.

- Given the generality of our QM/MM electrostatics treatment, a MLP trained with the wild-type enzyme should be applicable to the mutants. The transferability of the trained MLPs has yet to be tested and verified.

- More accurate *ab initio* QM/MM models beyond B3LYP/6-31G*/MM should be adopted as the labelling method for our MLP and ΔMLP training.

Work in these directions are currently being pursued.

Supporting Information Available

The following files are available free of charge.

- A brief introduction of the improved semi-empirical QM recalibration protocol, recalibrated PM3 parameters for the chorismate mutase reaction (Table S1), errors with respect to B3LYP/6-31G*/MM for the Menshutkin (Table S2) and the chorismate mutase (Table S3) reactions, and the total energy during the initial 500 fs of the NVE simulations (Figure S1).

References

(1) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3188–3209.

(2) Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

(3) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109*, 15645–15650.

(4) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go From Here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

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

(6) Senn, H. M.; Thiel, W. QM/MM Studies of Enzymes. *Curr. Op. Chem. Biol.* **2007**, *11*, 182–187.

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

(8) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational Enzymology. *Chem. Commun.* **2010**, *46*, 2354.

(9) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.

(10) Masgrau, L.; Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* **2015**, *48*, 431–438.

(11) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM Simulations: What Are Some of the "Burning Issues"? *J. Phys. Chem. B* **2021**, *125*, 689–702.

(12) Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H$_2$ to Biomolecules (Nobel Lecture). *Angew. Chem. Int. Ed.* **2014**, *53*, 9992–10005.

(13) Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem. Int. Ed.* **2014**, *53*, 10020–10031.

(14) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

(15) Wang, L.; Yu, X.; Hu, P.; Broyde, S.; Zhang, Y. A Water-Mediated and Substrate-Assisted Catalytic Mechanism for *Sulfolobus solfataricus* DNA Polymerase IV. *J. Am. Chem. Soc.* **2007**, *129*, 4731–4737.

(16) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8934–8941.

(17) 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.

(18) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S. Quantum Mechanical/Molecular Mechanical Free Energy Simulations of the Self-Cleavage Reaction in the Hepatitis Delta Virus Ribozyme. *J. Am. Chem. Soc.* **2014**, *136*, 1483–1496.

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

(20) Muller, R. P.; Warshel, A. Ab Initio Calculations of Free Energy Barriers for Chemical Reactions in Solution. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(21) Heimdal, J.; Ryde, U. Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592.

(22) 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.

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

(24) 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.

(25) Wang, J.-N.; Liu, W.; Li, P.; Mo, Y.; Hu, W.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at *Ab Initio* Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 4. Adaptive QM/MM. *J. Chem. Theory Comput.* **2021**, *17*, 1318–1325.

(26) 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.

(27) Martí, S.; Moliner, V.; Tuñón, I. Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy To Explore the Potential Energy Surface in Very Large Systems. *J. Chem. Theory Comput.* **2005**, *1*, 1008–1016.

(28) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(29) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.

(30) Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, Resonance-Free Multiple Time-Step Algorithm for Molecular Dynamics with Very Large Time Steps. *Mol. Phys.* **2013**, *111*, 3579–3594.

(31) Margul, D. T.; Tuckerman, M. E. A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J. Chem. Theory Comput.* **2016**, *12*, 2170–2180.

(32) Barth, E.; Schlick, T. Overcoming Stability Limitations in Biomolecular Dynamics. I. Combining Force Splitting via Extrapolation with Langevin Dynamics in LN. *J. Chem. Phys.* **1998**, *109*, 1617–1632.

(33) Barth, E.; Schlick, T. Extrapolation Versus Impulse in Multiple-Timestepping Schemes. II. Linear Analysis and Applications to Newtonian and Langevin Dynamics. *J. Chem. Phys.* **1998**, *109*, 1633–1642.

(34) Nam, K. Acceleration of Ab Initio QM/MM Calculations under Periodic Boundary Conditions by Multiscale and Multiple Time Step Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 4175–4183.

(35) Chen, Y.; Kale, S.; Weare, J.; Dinner, A. R.; Roux, B. Multiple Time-Step Dual-Hamiltonian Hybrid Molecular Dynamics – Monte Carlo Canonical Propagation Algorithm. *J. Chem. Theory Comput.* **2016**, *12*, 1449–1458.

(36) Liberatore, E.; Meli, R.; Rothlisberger, U. A Versatile Multiple Time Step Scheme for Efficient *ab Initio* Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 2834–2842.

(37) Pan, X.; Epifanovsky, E.; Liu, J.; Pu, J.; Nam, K.; Shao, Y. Accelerating ab initio QM/MM Molecular Dynamics Simulations With Multiple TimeStep Integration and a Recalibrated Semi-empirical QM/MM Hamiltonian. *To Be Submitted*

(38) Zhou, Y.; Pu, J. 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.

(39) Kim, B.; Snyder, R.; Nagaraju, M.; Zhou, Y.; Ojeda-May, P.; Keeton, S.; Shao, Y.; Pu, J. Reaction Path-Force Matching in Collective Variables: Determining Ab Initio QM/MM Free Energy Profiles by Fitting Mean Force. *Submitted for Publication*

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

(41) Wu, J.; Shen, L.; Yang, W. Internal Force Corrections with Machine Learning for Quantum Mechanics/Molecular Mechanics Simulations. *J. Chem. Phys.* **2017**, *147*, 161732.

(42) Shen, L.; Yang, W. Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks. *J. Chem. Theory Comput.* **2018**, *14*, 1442–1455.

(43) Gastegger, M.; Schütt, K. T.; Müller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. 2020.

(44) 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*; preprint, 2021.

(45) Böselt, L.; Thürlemann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658.

(46) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(47) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(48) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

(49) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(50) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and Structural Distribution Based Representation for Universal Quantum Machine Learning. *J. Chem. Phys.* **2018**, *148*, 241717.

(51) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. *arXiv preprint* **2018**, arXiv:1704.06439.

(52) 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. Advances in Neural Information Processing Systems. 2018.

(53) 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.

(54) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; 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.

(55) York, D. M.; Karplus, M. A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *J. Phys. Chem. A* **1999**, *103*, 11060–11079.

(56) Lange, A. W.; Herbert, J. M. A Smooth, Nonsingular, and Faithful Discretization Scheme for Polarizable Continuum Models: The Switching/Gaussian Approach. *J. Chem. Phys.* **2010**, *133*, 244111.

(57) 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.

(58) Kim, B.; Shao, Y.; Pu, J. Doubly Polarized QM/MM with Machine Learning Chaperone Polarizability. *Submitted for Publication*

(59) Riccardi, D.; Li, G.; Cui, Q. Importance of van der Waals Interactions in QM/MM Simulations. *J. Phys. Chem. B* **2004**, *108*, 6467–6478.

(60) Riccardi, D.; Schaefer, P.; Yang,; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) Methods for Complex Biological Processes. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

(61) Freindorf, M.; Shao, Y.; Furlani, T. R.; Kong, J. Lennard-Jones Parameters for the Combined QM/MM Method Using the B3LYP/6-31G*/AMBER Potential. *J. Comput. Chem.* **2005**, *26*, 1270–1278.

(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(63) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.;

Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(64) Chook, Y. M.; Ke, H.; Lipscomb, W. N. Crystal Structures of the Monofunctional Chorismate Mutase from Bacillus Subtilis and Its Complex with a Transition State Analog. *Proc. Natl. Acad. Sci.* **1993**, *90*, 8600–8603.

(65) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(66) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: AnN·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(68) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(69) 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, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; 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.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. AMBER 2020, University of California, San Francisco. 2020.

(70) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(71) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

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

(73) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(74) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.

(75) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2014**, *113*, 184–215.

(76) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

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

(78) Pinnick, E. R.; Calderon, C. E.; Rusnak, A. J.; Wang, F. Achieving Fast Convergence of ab initio Free Energy Perturbation Calculations with the Adaptive Force-Matching Method. *Theor. Chem. Acc.* **2012**, *131*.

(79) Ryde, U. How Many Conformations Need To Be Sampled To Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* **2017**, *13*, 5745–5752.

(80) Boresch, S.; Woodcock, H. L. Convergence of Single-Step Free Energy Perturbation. *Mol. Phys.* **2017**, *115*, 1200–1213.

Supporting Information:

Machine Learning Assisted Free Energy Simulation of Solution–Phase and Enzyme Reactions

Xiaoliang Pan,[†] Junjie Yang,[†] Richard Van,[†] Evgeny Epifanovsky,[‡] Junming Ho,[¶]

Jing Huang,[§] Jingzhi Pu,[*,‖] Ye Mei,[*,⊥,#,@] Kwangho Nam,[*,△] and Yihan Shao[*,†]

[†]Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Pkwy, Norman, OK 73019, United States.

[‡]Q-Chem, Inc. 6601 Owens Drive, Suite 105, Pleasanton, CA 94588, United States.

[¶]School of Chemistry, University of New South Wales, Sydney, NSW 2052, Australia.

[§]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou, Zhejiang 310024, China

[‖]Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N Blackford St, LD326, Indianapolis, IN 46202, United States

[⊥]State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China

[#]NYU–ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

[@]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China

[△]Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, TX 76019, United States.

E-mail: jpu@iupui.edu; samuel.y.mei@gmail.com; kwangho.nam@uta.edu; yihan.shao@ou.edu

1 Improved Semi-Empirical QM Recalibration Protocol

In this work, the semi-empirical QM recalibration protocol was improved upon the one from our previous paper[1], which in turn is based on the Reaction Path Force Matching (RP-FM) method[2]. Inspired by the ForceBalance method[3], where the objective function is defined as the weighted sum of a set of scaled properties of the system, such as energies, forces, and dipole moments, we defined the objective function in this work to include errors in energies E, bond forces F_a, residual Cartesian forces with bond forces removed $\mathbf{F_b}$, and a quadratic penalty term with a strength of λ that restrains the parameters to their initial values. The final form of the objective function is

$$
\begin{aligned}
\mathcal{X}^2(\mathbf{x}) =\, & W_E \frac{\sum_{i \in N_s}((E_i^{\mathrm{SE}} - \bar{E}^{\mathrm{SE}}) - (E_i^{\mathrm{AI}} - \bar{E}^{\mathrm{AI}}))^2}{\sum_{i \in N_s}(E_i^{\mathrm{AI}} - \bar{E}^{\mathrm{AI}})^2} \\
& + W_F \frac{\sum_{i \in N_s} \sum_{a \in N_{\mathrm{bond}}} w_a (F_{i,a}^{\mathrm{SE}} - F_{i,a}^{\mathrm{AI}})^2}{\sum_{i \in N_s} \sum_{a \in N_{\mathrm{bond}}} w_a |F_{i,a}^{\mathrm{AI}}|^2} \\
& + W_F' \frac{\sum_{i \in N_s} \sum_{b \in N_{\mathrm{atom}}} |\mathbf{F}_{i,b}^{\mathrm{SE}} - \mathbf{F}_{i,b}^{\mathrm{AI}}|^2}{\sum_{i \in N_s} \sum_{b \in N_{\mathrm{atom}}} |\mathbf{F}_{i,b}^{\mathrm{AI}}|^2} \\
& + \lambda |\mathbf{x} - \mathbf{x_0}|^2,
\end{aligned}
$$

where W_E, W_F, and W_F' are the weights for the errors in energies, bond forces, and residual Cartesian forces. w_a is the weight of force error for bond a. Instead of optimizing the semi-empirical QM parameters directly, we defined $\mathbf{x}$ as the relative deviations from the standard semi-empirical QM parameter set. In this way, the initial parameters $\mathbf{x_0}$ is a vector containing only zeros. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm as implemented in the SciPy package was used to perform the minimization, without any bounds on the parameters. The first derivatives with respect to the parameters were calculated using the finite difference method.

[1] Pan, X.; Li, P.; Ho, J.; Pu, J.; Mei, Y.; Shao, Y. *Phys. Chem. Chem. Phys.* **2019**, *21*, 20595–20605.

[2] Zhou, Y.; Pu, J. *J. Chem. Theory Comput.* **2014**, *10*, 3038–3054.

[3] Wang, L.-P.; Chen, J.; Voorhis, T. V. *J. Chem. Theory Comput.* **2013**, *9*, 452–460.

Table S1: Recalibrated PM3 Parameters for Chorismate Mutase Reaction

parameters	H		C		O	
	PM3	PM3*	PM3	PM3*	PM3	PM3*
U_{ss} (eV)	−13.073321	−14.104461	−47.270320	−35.771165	−86.993002	−89.017515
U_{pp} (eV)			−36.266918	−31.777119	−71.879580	−77.254225
ζ_{ss} (au)	0.967807	1.016984	1.565085	1.790595	3.796544	3.777840
ζ_{pp} (au)			1.842345	1.658069	2.389402	2.009813
β_{ss} (eV)	−5.626512	−5.288635	−11.910015	−12.887428	−45.202651	−45.244602
β_{pp} (eV)			−9.802755	−5.801079	−24.752515	−14.749516
G_{ss} (eV)	14.794208	19.086036	11.200708	10.120272	15.755760	9.484059
G_{pp} (eV)			10.796292	14.662809	13.654016	17.429415
G_{sp} (eV)			10.265027	6.633904	10.621160	11.894290
G_{p2} (eV)			9.042566	10.024770	12.406095	13.438287
H_{sp} (eV)			2.290980	2.762577	0.593883	0.601093
a_1 (unitless)	1.128750	1.151601	0.050107	0.036006	−1.131128	−1.123957
b_1 (1/Å^2)	5.096282	5.899212	6.003165	4.411918	6.002477	6.150366
c_1 (Å)	1.537465	1.525685	1.642214	1.667922	1.607311	1.571011
a_2 (unitless)	−1.060329	−1.030224	0.050733	0.048980	1.137891	1.146880
b_2 (1/Å^2)	6.003788	6.516917	6.002979	5.815149	5.950512	6.103272
c_2 (Å)	1.570189	1.561958	0.892488	0.876823	1.598395	1.569670

Table S2: Errors with Respect to B3LYP/6-31G*/MM for Menshutkin Reaction

	MLP				PM3* + ΔMLP				PM3				PM3*			
	E[a]	F_{QM}[b]	ϕ_{MM}[c]	ε_{MM}[d]	E	F_{QM}	ϕ_{MM}	ε_{MM}	E	F_{QM}	ϕ_{MM}	ε_{MM}	E	F_{QM}	ϕ_{MM}	ε_{MM}
MSE		−0.01	0.18	0.00		0.00	0.15	0.00		0.02	−0.02	0.00		0.01	−0.11	0.00
MUE	0.79	0.75	0.37	0.10	0.62	0.53	0.31	0.09	5.00	9.62	0.67	0.24	5.49	3.21	0.70	0.22
MAX	3.79	15.49	10.57	33.13	2.57	12.04	8.26	33.33	21.55	54.32	38.41	49.39	16.55	32.64	16.22	35.51
RMSE	1.03	1.17	0.65	0.41	0.78	0.79	0.51	0.34	6.24	12.90	1.33	0.95	6.38	4.42	1.16	0.65

[a] Total energy (kcal mol⁻¹). Because the energies were shifted to have zero mean for each model, MSE of energy was zero by definition.
[b] Forces of QM atoms (kcal mol⁻¹ Å⁻¹).
[c] Electrostatic potential at MM atomic sites (kcal mol⁻¹ e^{-1}).
[d] Electric field at MM atomic sites (kcal mol⁻¹ Å⁻¹ e^{-1}).

Table S3: Errors with Respect to B3LYP/6-31G*/MM for Chorismate Mutase Reaction

	MLP (1st)				MLP (2nd)				PM3* + ΔMLP				PM3				PM3*			
	E[a]	F_{QM}[b]	ϕ_{MM}[c]	ε_{MM}[d]	E	F_{QM}	ϕ_{MM}	ε_{MM}	E	F_{QM}	ϕ_{MM}	ε_{MM}	E	F_{QM}	ϕ_{MM}	ε_{MM}	E	F_{QM}	ϕ_{MM}	ε_{MM}
MSE		0.07	0.07	0.01		0.08	0.02	0.01		−0.03	0.10	0.00		0.17	0.15	−0.03		0.18	−0.11	0.00
MUE	0.78	1.36	0.31	0.08	0.58	0.81	0.23	0.07	0.54	0.65	0.21	0.06	9.76	10.07	0.51	0.22	1.79	3.68	0.42	0.15
MAX	3.76	14.78	9.29	15.71	3.08	8.32	9.40	16.03	3.97	7.18	7.13	17.42	34.55	58.03	23.87	46.64	9.20	33.29	16.20	32.77
RMSE	0.99	1.84	0.48	0.29	0.73	1.10	0.40	0.25	0.69	0.88	0.29	0.23	12.08	13.59	1.13	1.07	2.25	4.88	0.85	0.52

[a] Total energy (kcal mol⁻¹). Because the energies were shifted to have zero mean for each model, MSE of energy was zero by definition.
[b] Forces of QM atoms (kcal mol⁻¹ Å⁻¹).
[c] Electrostatic potential at MM atomic sites (kcal mol⁻¹ e^{-1}).
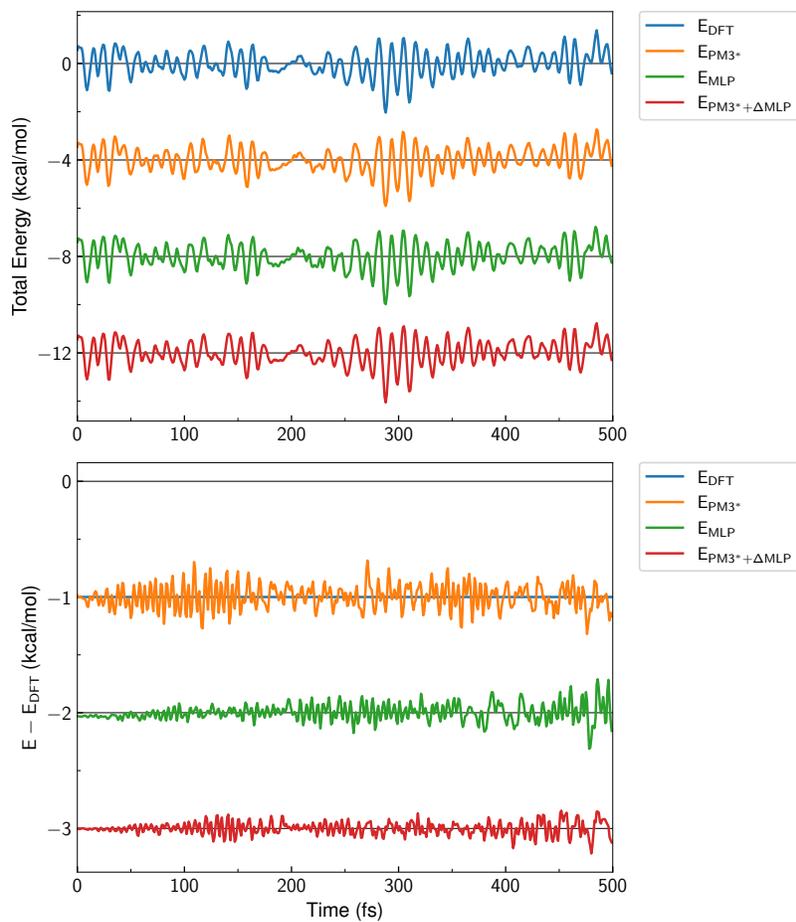[d] Electric field at MM atomic sites (kcal mol⁻¹ Å⁻¹ e^{-1}).

Figure S1: The total energy (top) and the energy difference with respective to DFT/MM (bottom) during the initial 500 fs of the NVE simulations. For each model, the energies were shifted for clarity.