

# Accurate Free Energy Calculation via Multiscale Simulations Driven by Hybrid Machine Learning and Molecular Mechanics Potentials

Xujian Wang,<sup>†</sup> Xiongwu Wu,<sup>\*,‡</sup> Bernard R. Brooks,<sup>‡</sup> and Junmei Wang<sup>\*,†</sup>

<sup>†</sup>*Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening  
Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, Pennsylvania 15261,  
United States*

<sup>‡</sup>*Laboratory of Cell Biology, National Heart, Lung and Blood Institute, National Institutes  
of Health, Bethesda, MD, USA*

E-mail: wuxw@nhlbi.nih.gov; junmei.wang@pitt.edu

## Abstract

Our study focused on the implementation and testing of machine learning inter-atomic potentials (MLIPs) into the AMBER software suite. This implementation enables us to perform a novel type of molecular dynamics simulation utilizing the hybrid machine learning/molecular mechanics (ML/MM) potentials. To underpin the capabilities of ML/MM simulations, we first validated our implementation at a fundamental physical level by confirming energy and momentum conservation laws. The successful validation indicates that our implementation is able to produce adequate and physically interpretable samplings. Building upon this, for the first time to the best of our knowledge, we proposed an ML/MM-compatible thermodynamic integration (TI)

protocol to tackle real-world challenges, such as solvation free energy calculation. Our results demonstrate that this computational protocol can predict hydration free energies with an accuracy of less than 1.00 kcal/mol compared to experimental data, paving the way for the use of ML/MM in multiscale simulations to addressing future drug design problems. Moreover, by applying ML/MM in molecular dynamics simulations of protein-ligand complexes, we demonstrated that the adequate samplings enable us to accurately reproduce experimental binding free energies. Thus, our implementation can offer new insights into biomolecular systems using the ML/MM "microscope". Last, we demonstrated that our implementation can achieve nanosecond timescale simulations daily after significant effort being put to improve the code performance. In a conclusion, we have successfully implemented ML/MM potential to AMBER software package after overcoming limitations in current multi-scale simulations including low computational efficiency. We have advanced TI theory allowing us to accurately predict free energies with ML/MM potentials.

## 1 Introduction

In the field of molecular dynamics (MD) simulations, much work has contributed to improving molecular mechanics force fields (MMFF) to achieve higher accuracy in reproducing experimental metrics. Efforts include extending general small molecule force fields,<sup>1,2</sup> developing new protein force fields,<sup>3-6</sup> and creating force fields for other biomolecules, such as DNA<sup>7,8</sup> and lipids.<sup>9,10</sup> Classical force fields aim to reproduce quantum mechanics (QM) results, however, it remains a challenge to accurately matching these QM results, especially when chemical reactions are involved. To overcome the challenge, a feasible solution is to combine the computational efficient MMFF method with the accurate QM methods. In the 1970s, Warshel Arieh and Levitt Michael<sup>11</sup> proposed quantum mechanics/molecular mechanics (QM/MM) molecular dynamics, which applies a QM model which includes density functional theory (DFT)<sup>12,13</sup> to describe the essential part of the system (such as atoms

involved in a chemical reaction), and MMFF to describe the rest of the system.<sup>14,15</sup> This multi-scale simulation technology make the investigation of electronic structures and chemical reactions in a large system come true.

In recent years, many efforts have been made to improve QM/MM MD simulations. For instance, electrostatic embedding approach was proposed to accurately calculate the electrostatic interaction,<sup>16,17</sup> and polarizable force fields have been introduced to account for interatomic polarization effects.<sup>18–20</sup> Additionally, improvements in long-range interaction processing have enhanced the accuracy of simulations.<sup>21,22</sup> Most of the above measures focus on improving simulation accuracy. However, the primary limitation to the broad application of multi-scale simulation techniques remains the simulation performance. That’s to say the true bottleneck in QM/MM MD is the QM calculation, which is difficult to be accelerated significantly.

However, the rise of various artificial intelligence (AI) techniques has sparked a wave of developing machine learning interatomic potentials (MLIPs).<sup>23–33</sup> MLIPs are trained using machine learning algorithms to reproduce ab initio quantities, such as potential energy and atomic forces. For example, ANI-2x,<sup>27</sup> which was trained on data from  $\omega$ B97x-D/6-31G(d)<sup>34,35</sup> calculations, achieves near-DFT levels of accuracy while maintaining computational efficiency comparable to molecular mechanics. Given by their accuracy and performance, MLIPs could potentially serve the role of ab initio models in simulating biomolecular systems with QM/MM. Incorporation of MLIPs into a molecular dynamics engine to develop a brand-new multi-scale simulation technique is appealing given their near-QM level of accuracy and near-MM level of efficiency. Thus, machine learning/molecular mechanics molecular dynamics (ML/MM MD), represents a promising opportunity for biomolecular simulations.<sup>36–38</sup> Many MLIP models represented by the ANI series<sup>23–27</sup> and MACE series<sup>28–30</sup> have demonstrated robustness, transferability, and high performance. Therefore, they are ideal candidates for incorporation into MD simulation platforms.

Most of the time, free energy calculations are performed using molecular mechanics (MM)

which is based upon the applied MMFFs. However, MM-based free energy methods sometimes cannot produce accurate energy values due to a lack of adequate description of some subparts of the simulation system with MMFFs, for which a QM-level description is often necessary. It is appealing to apply ML/MM MD approach to free energy calculations, given the advantage of ML over QM. In this framework, critical regions of a system, such as those being involved in chemical reactions and ligand and protein binding, are modeled with MLIPs, while the surrounding environment is modeled by MM. This approach enables high-quality free energy calculations as it strikes a balance between computational accuracy and efficiency.

In this work, we incorporated MLIPs into the AMBER simulation platform.<sup>39,40</sup> We then used the modified code to validate conservation laws within the ML/MM approach, thereby ensuring a thermodynamically consistent system. Next, we advanced the theory of free energy calculations using the thermodynamic integration (TI) method with the core part being described by MLIPs. We tested the new technique by performing hydration free energy calculations via multi-scale ML/MM molecular dynamics simulations. To further explore potential applications of ML/MM MD to macromolecular systems, we applied the ML/MM MD engine to sample conformations and MM-PBSA method to predict binding free energy for a set of protein-ligand complexes. Finally, we performed performance tests on our code. By leveraging advancements in MLIPs and integrating them with multi-scale simulation techniques, we expect that our ML/MM approach can help address many current challenges in molecular simulation.

## 2 Theory

### 2.1 Fundamental theory for ML/MM approach

The ML/MM approach is highly similar to the QM/MM approach.<sup>11,41</sup> We divide the total energy into three components: the ML part, the MM part, and the ML-MM interaction part:

$$E_{\text{total}} = E_{\text{ML}} + E_{\text{MM}} + E_{\text{ML-MM}} \quad (1)$$

The ML part is computed using MLIPs:

$$E_{\text{ML}} = \text{MLIP}(\mathbf{R}^{\text{ML}}, \mathbf{N}^{\text{ML}}) \quad (2)$$

Notably, the functional form of MLIP varies from model to model. Here,  $\mathbf{R}^{\text{ML}}$  represents the atomic coordinates in the ML region, and  $\mathbf{N}^{\text{ML}}$  denotes the corresponding atomic numbers in the ML region. Similar to QM, the potential energy depends solely on the system's atomic coordinates and atomic numbers.

The MM part is governed by classical force field equations,<sup>42</sup> consisting of bond, angle, and dihedral bonded terms, as well as electrostatic and van der Waals nonbonded terms:

$$\begin{aligned} E_{\text{MM}} = & \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} V_n [1 + \cos(n\phi - \gamma)] \\ & + \sum_{i,j} \left( \frac{q_i q_j}{r_{ij}} + \epsilon \left[ \left( \frac{A}{r_{ij}} \right)^{12} - \left( \frac{B}{r_{ij}} \right)^6 \right] \right) \end{aligned} \quad (3)$$

For the ML-MM interaction, although several methods have been proposed to improve the accuracy of electrostatic<sup>41</sup> and van der Waals<sup>43</sup> interactions, we use the most widely used functional forms. This involves using Coulomb's law and employing Lennard-Jones (LJ) potentials to describe the interactions. The atomic partial charges and atomic LJ parameters for the ML region are also obtained from the employed MMFF.

$$E_{\text{ML-MM}} = \sum_{i \in \text{MM}} \sum_{j \in \text{ML}} \frac{q_i q_j}{|\mathbf{R}_i^{\text{MM}} - \mathbf{R}_j^{\text{ML}}|} + \sum_{i \in \text{MM}} \sum_{j \in \text{ML}} \epsilon \left[ \left( \frac{A}{|\mathbf{R}_i^{\text{MM}} - \mathbf{R}_j^{\text{ML}}|} \right)^{12} - \left( \frac{B}{|\mathbf{R}_i^{\text{MM}} - \mathbf{R}_j^{\text{ML}}|} \right)^6 \right] \quad (4)$$

In the equation,  $\mathbf{R}_i^{\text{MM}}$  and  $\mathbf{R}_j^{\text{ML}}$  represent the atomic coordinates in the MM and ML regions, respectively. In van der Waals interactions, the  $A$  and  $B$  parameters, which are calculated using mixing rules, are atom type dependent.

## 2.2 Virials calculation for accurate thermodynamic system simulation

Accurate virials calculation can guarantee the correct thermodynamic behavior, especially in NPT ensemble. The fundamental expression virial calculation equation is as follows:

$$\delta W = \sum_i \mathbf{F}_i \cdot \delta \mathbf{u}_i \quad (5)$$

Where  $\delta W$  denotes the complete virials of the system.  $\mathbf{F}_i$  and  $\delta \mathbf{u}_i$  denote the atomic force and atomic virtual displacement, respectively. For atom in MM region and ML region, their forces are calculated differently:

$$F_{\text{total}}^i = F_{\text{MM}}^i + F_{\text{ML-MM}}^i, \quad i \in \text{MM region} \quad (6)$$

For an atom in MM region, its force ( $F_{\text{total}}^i$ ) consists of MM force ( $F_{\text{MM}}^i$ ) and the part due to its interaction with the ML region ( $F_{\text{ML-MM}}^i$ ). According to the eqs.3 and 4, it can be calculated analytically.

$$F_{\text{total}}^j = F_{\text{ML}}^j + F_{\text{ML-MM}}^j, \quad j \in \text{ML region} \quad (7)$$

While for the atom in ML region, its force consists of ML force ( $F_{\text{ML}}^i$ ) itself and the part due to its interaction with the MM region ( $F_{\text{ML-MM}}^i$ ). The latter ( $F_{\text{ML-MM}}^j$ ) can be calculated analytically, while the former is derived from a complicated neural network. But owing to an auto-gradient technique, which makes it possible for us to estimate the gradient of energy using the energy from eq.2 .

$$F_{\text{ML}} = -\nabla \cdot \mathbf{E}_{\text{ML}} \quad (8)$$

By accurately calculating virials, we are able to generate real thermodynamics for the simulation system.

### 2.3 Thermodynamic integration (TI) for ML/MM approach

TI is a robust method for estimating free energy changes during phase transitions.<sup>44</sup> Traditionally, TI calculations follow the equation below:

$$\Delta G = G_{\lambda=1} - G_{\lambda=0} = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (9)$$

When calculating solvation free energy using TI, it is necessary to determine the free energy in the solvent phase ( $\langle \frac{\partial V}{\partial \lambda} \rangle_{\text{solvent}}$ ) and the gas phase ( $\langle \frac{\partial V}{\partial \lambda} \rangle_{\text{gas}}$ ) separately. Considering the bonded and non-bonded terms in ML are not separable, we need to develop a new protocol to enable TI calculations using ML/MM approach. In our protocol, we assumed the contributions from solute itself in the gas and solvent phases are the same. Under this assumption, we only need to do TI calculations in the solvent phase and the values  $\langle \frac{\partial V}{\partial \lambda} \rangle$  are calculated without considering the intramolecular energy of the solute. We then estimate the reorganization energy, which accounts for the free energy difference associated with the different conformational ensembles in the gas and solvent phases. We expect that the reorganization energy can mitigate the prediction error due to the above assumption. Typically, for hydration free energy calculation:

$$\Delta G_{\text{solvation}} = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{wat}} d\lambda + \Delta G_{\text{reorg}} \quad (10)$$

Here, we define the reorganization energy ( $\Delta G_{\text{reorg}}$ ) as the difference between the mean values of the potential energies for conformations sampled in the in aqueous and gas phases:

$$\Delta G_{\text{reorg}} = \langle E_{\text{ML}} \rangle_{\text{wat}} - \langle E_{\text{ML}} \rangle_{\text{gas}} \quad (11)$$

At this stage,  $\langle \frac{\partial V}{\partial \lambda} \rangle_{\text{wat}}$  and  $\Delta G_{\text{reorg}}$  were calculated using a MLIP which has a near-DFT level accuracy. For more detailed theoretical hypothesis and formula derivation, please refer to the supporting information.

## 3 Results and discussion

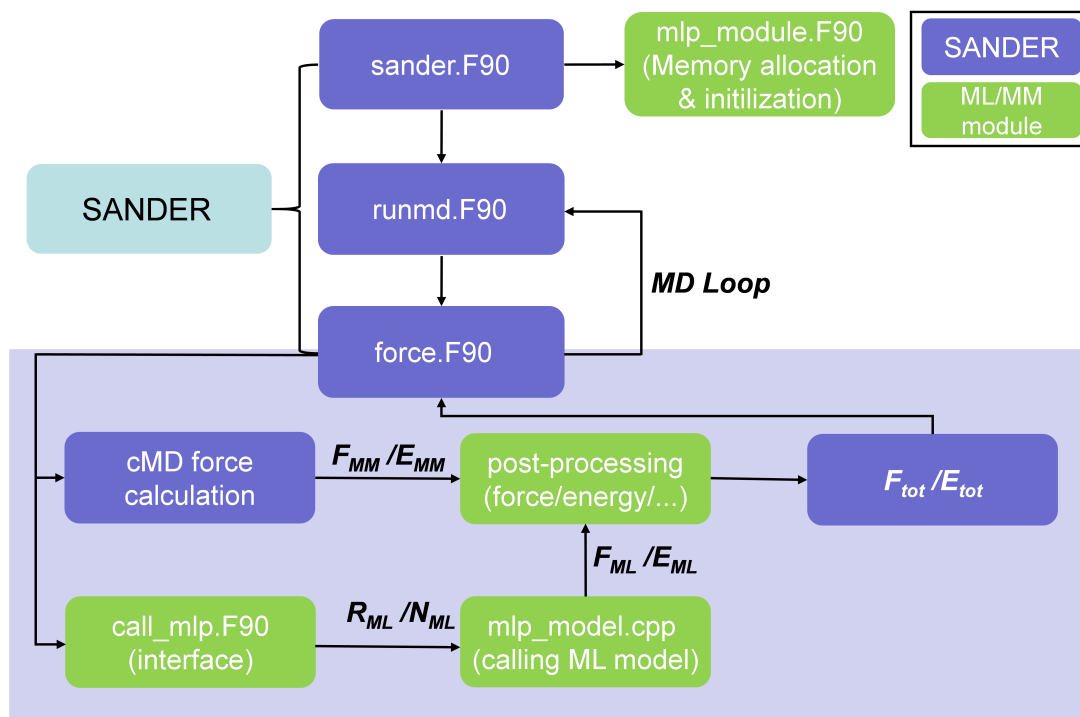
### 3.1 Implementation detail

The development of our code is based on the Amber 2023 software suite, mainly focused on SANDER, a highly efficient MD engine that utilizes CPU resources for MD simulations. To enhance simulation performance, we considered two strategies: built-in implementation and parallel computing.

Notably, most MLIPs are trained and used within the PyTorch framework,<sup>45</sup> which is based on Python. In contrast, SANDER is primarily written in Fortran. Directly calling MLIPs by SANDER could result in high latency due to the frequent initialization of MLIPs in each MD step. Furthermore, Being a high-level language, Python makes direct memory manipulation challenging. To address these issues, we implemented the MLIPs in a built-in manner (see Scheme 1), meaning that the MLIPs are initialized alongside SANDER and the memory is retained until the MD simulation finishes. This approach avoids frequent memory allocation by storing the ML model in system cache for rapid access. We implemented the MLIP model in C++ to enable both inference and auto-gradient, which are



essential for atomic force calculations. To integrate SANDER and MLIPs, we developed a Fortran/C++ interface, allowing MLIPs to communicate with SANDER directly through memory addresses, thereby reducing communication latency.



Scheme 1: The implementation details of ML/MM module in SANDER MD engine in Amber. This module is initialized alongside SANDER. Once the MD loop begins, the ML model is invoked through an intrinsic Fortran/C++ interface, facilitating direct communication between the ML algorithms and the core MD routines. The outputs generated by the ML model are then meticulously processed, then they are incorporated into the SANDER mainstream.

To further enhance performance, we adapted the code for both the standard SANDER and the Message Passing Interface (MPI) enabled SANDER.MPI. In this configuration, SANDER manages conventional MD (cMD) calculations, accelerated through MPI, while the ML/MM module is responsible for the MLIP inference. The MLIP inference is conducted separately on graphics processing units (GPU), which significantly accelerates overall processing. During SANDER's calls to the ML/MM interface, inference jobs are dispatched to the GPU, allowing asynchronous computing: cMD calculations occur on CPU cores, while

MLIP inference takes place on GPU (see Scheme 1).

Currently, we have integrated the ANI series (ANI-1x,<sup>25</sup> ANI-1ccx,<sup>24</sup> and ANI-2x<sup>27</sup>) and the MACE series<sup>29</sup> (MACE-OFF23(S), MACE-OFF23(M), MACE-OFF23(L)) MLIPs into SANDER. Our ML/MM implementation mirrors the mechanical embedding approach<sup>41</sup> used in QM/MM frameworks. Users specify the atomic ML region, with the remaining atoms being in the MM region.

This integrated approach enables the ML/MM module to augment traditional MD simulations with machine learning capabilities while maintaining compatibility with the Amber 2023 SANDER framework.

### 3.2 Validation of conservation laws

MD is a robust simulation technology used to model molecular behavior at the atomic level. Leveraging this technology enables researchers to investigate the mechanisms underlying biological processes. For example, MD can be used to unravel biocatalysts behavior<sup>46,47</sup> and understand protein-drug interactions.<sup>48,49</sup> However, the accurate reproduction of real molecular behaviors in MD simulations relies on the MD engine can strictly adhere to fundamental physical laws, such as Newton's equations of motion. By following these laws, we can assess whether our ML/MM approach upholds the two most fundamental conservation laws—energy and momentum—thereby ensuring the reliability of our simulation results.

In this study, we tested the MD system by simulating erlotinib,<sup>50</sup> an EGFR inhibitor, in water under the microcanonical ensemble (NVE, i.e. constant number of particles, volume and total energy of the system). The system contains 151 atoms (Figure 1A), with 52 belonging to erlotinib, defined as the ML region, while the remaining 99 atoms belonging to 33 water molecules. In this configuration, the ML region constitutes approximately 34.4% of the system's atoms, allowing us to observe its influence on the total energy contribution as fully as possible.

We employed ANI-2x, which showed its accurate reproduction of DFT results and high

reliability,<sup>51–53</sup> as the MLIP model in the ML/MM MD approach and used cMD as a reference to conduct a 1 nanosecond (ns) simulation with a timestep of 0.1 femtosecond (fs). For the ML/MM simulation, the average energy was -255.28 kcal/mol (Figure 1B), with a standard deviation of 0.03 kcal/mol, while the cMD simulation showed an average energy of -153.75 kcal/mol (Figure 1C), with a standard deviation of 0.12 kcal/mol. The difference in the absolute values of the total energy arose because, in the ML/MM approach, the atoms in the ML region are calculated using the MLIP rather than molecular mechanics. Notably, the ML/MM approach demonstrates a lower deviation in energy throughout the simulation, indicating MLIP can reduce the fluctuation of potential energy. By describing a portion of these atoms with MLIP, the computational error from the MM region is minimized. Specifically, in the ML region, the ANI-2x model's potential surface is smooth and nearly consistent, resulting in a balanced net force that helps the ML/MM simulation adhere strictly to the conservation of energy law.

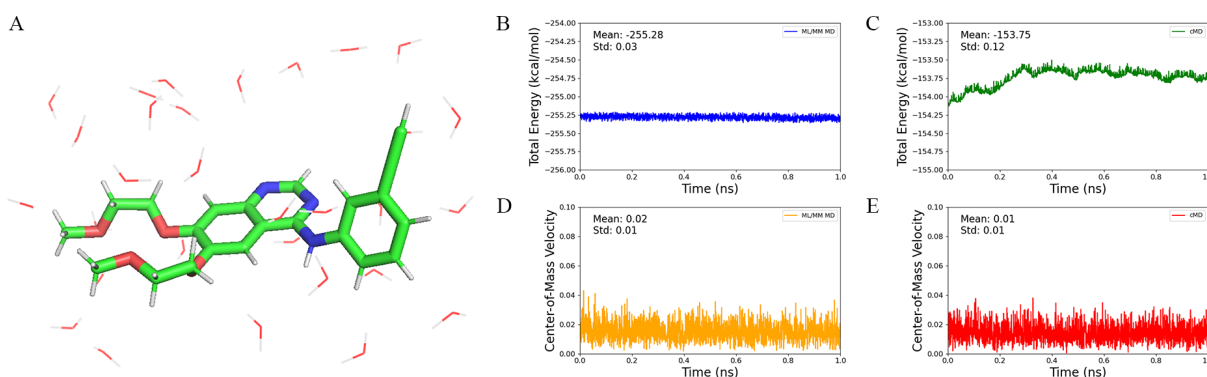


Figure 1: Validation of Energy and Momentum Conservation. (A) and (B) show the energy curves throughout the simulation for the ML/MM MD approach and the cMD approach, respectively. Figures (C) and (D) show the velocity of the center of mass during the simulation for the two approaches.

Regarding momentum conservation, two simulations were initiated from the same minimized structure, with the initial momentum being set to zero. Throughout the simulations, the velocity of the center of mass remained close to zero: 0.02 for ML/MM (Figure 1D) and 0.01 for cMD (Figure 1E), respectively. Neither system was subjected to any external forces,

as evidenced by the small fluctuations observed. To further support this observation, we also calculated the translational and rotational energies of both systems. The translational energy remained below 0.15 kcal/mol for both systems (Figure S1), while the rotational energy was much lower, at 0.02 kcal/mol for ML/MM and 0.07 kcal/mol for cMD. These findings indicate that the ML/MM approach can consistently simulate the thermodynamics of a system and produce molecular behaviors without obeying laws of thermodynamics.

By validating these two fundamental rules, we believe we have successfully integrated MLIPs into the MD framework. This integration allows MLIPs to adequately respond to MM interactions, adjust their conformations according to environmental changes, thereby enable the exchange of energy and force between these components in a physically interpretable way.

### 3.3 Thermodynamic integration for hydration free energy estimation

TI is a useful technique in which a perturbation is applied to facilitate the system's transition from  $V_0$  to  $V_1$ ,<sup>44</sup> allowing for an estimation of the resulting free energy difference. This method is widely used in solvation energy prediction<sup>54–57</sup> and binding free energy between a ligand and a receptor.<sup>58,59</sup> In this study, we develop a ML/MM-based TI calculation protocol for hydration free energy calculation.

Mobley and Guthrie reported hundreds of molecules with experimental hydration free energy data.<sup>60</sup> Notably, for these molecules, the energy estimated using the MMFF method with the traditional TI protocol demonstrated a deviation of  $\pm 1.5$  kcal/mol.<sup>55–57</sup> We randomly selected 20 compounds containing the elements C, H, O, N, F, and Cl from this dataset, applied our newly-proposed protocol to predict hydration free energies using ANI-2x in combination with GAFF2.<sup>2</sup> We used ABCG2 charge model<sup>61</sup> to calculate the electrostatic interactions, considering it outperforms AM1-BCC in many molecular property calculations. Figure 2 illustrates the prediction accuracy of different models. Note that the results for CGenFF<sup>1</sup>

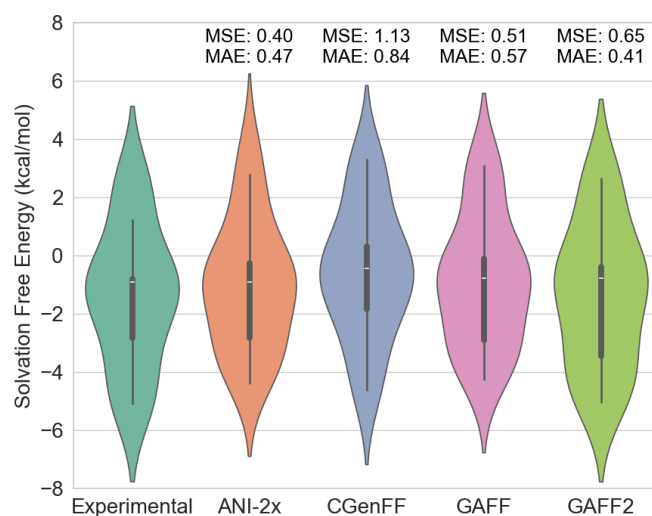


Figure 2: Prediction of hydration free energy using classical force field and Machine Learning/Molecular Mechanics Approaches

and GAFF<sup>2</sup> were directly obtained from previous publications.<sup>56,60</sup> In summary, the overall data distributions of ANI-2x and GAFF2 are relatively similar, with nearly the same mean absolute error (MAE) of 0.47 kcal/mol and 0.41 kcal/mol (Figure 2), respectively, which are significantly lower than those obtained using either CGenFF (0.84 kcal/mol) or GAFF (0.57 kcal/mol). We suspect that the slightly lower accuracy of ANI-2x compared to GAFF2 is due to some cooperation or consistency issue between them. After all, ANI-2x was trained to reproduce high-accuracy DFT energetics and forces ( $\omega$ B97x-D/6-31G(d)), whereas GAFF<sup>2</sup> and TIP3P water<sup>62</sup> were developed to reproduce both quantum mechanics and experimental data. This difference might account for the observed discrepancy. However, the quartile line distribution and mean squared error (MSE) indicate that the hydration free energies estimated by ANI-2x are closer to the experimental data. All these results indicate that our postulated theory regarding ML/MM demonstrates its comparability to the traditional TI approach in a novel way. Traditional TI, however, employs a gradual scaling-down method to reduce intramolecular interactions, which may also affect interactions between water and the molecule. This creates a highly coupled system; while our approach aims to reasonably

decouple these interactions, further efforts are needed to estimate the coupling effects in TI calculations, thereby enhancing the accuracy of ML/MM TI calculations.

### 3.4 Protein ligand complex simulation

Biological macromolecules are essential molecules that carry out complex functions in organisms. Applying ML/MM to the simulation of these macromolecules can enhance our understanding of their mechanisms with near-DFT accuracy at atomistic level.

We selected six well-studied protein-ligand complexes<sup>63,64</sup> as candidates for our analysis and then conducted ML/MM MD simulations. Figures 3A and 3B illustrate that these proteins exhibit small fluctuations (less than 1 Å) during the 5 ns simulations. The ligands described by MLIPs also exhibit small fluctuations of less than 2 Å (Figures 3C and 3D). These results demonstrate that our ML module is successfully implemented and suitable for real-world simulations. Thus our ML/MM approach is appropriate for investigating protein-ligand interactions in drug design.

To quantitatively assess the performance of the ML/MM approach, we calculated the B-factors for each structure and compared them with experimental data. Most proteins exhibited a high correlation, with Pearson correlation coefficients exceeding 0.5 (Figure S2). Among the structures, the Myeloid cell leukemia 1 protein (PDB ID: 4HW3) showed the lowest correlation coefficient of 0.18. However, the dynamics of the protein may be still reasonable due to the following reason. The original protein from the PDB entry is a multimer, but only the biological unit of the protein, which is a monomer, was chosen to do the simulations. (Figure S3). In this altered environment, transitioning from protein-protein interactions to a solvent-based context, the protein's dynamic behavior may be altered.

Additionally, visualizations provide a more intuitive understanding. We plotted the protein (PDB ID: 4GIH) with B-factors derived from both experimental data and our calculations (Figures 3E, 3F, and 3G). In our approach, most structures results in similar B-factor values as observed. Additional structure-based B-factor comparisons are presented in Figure

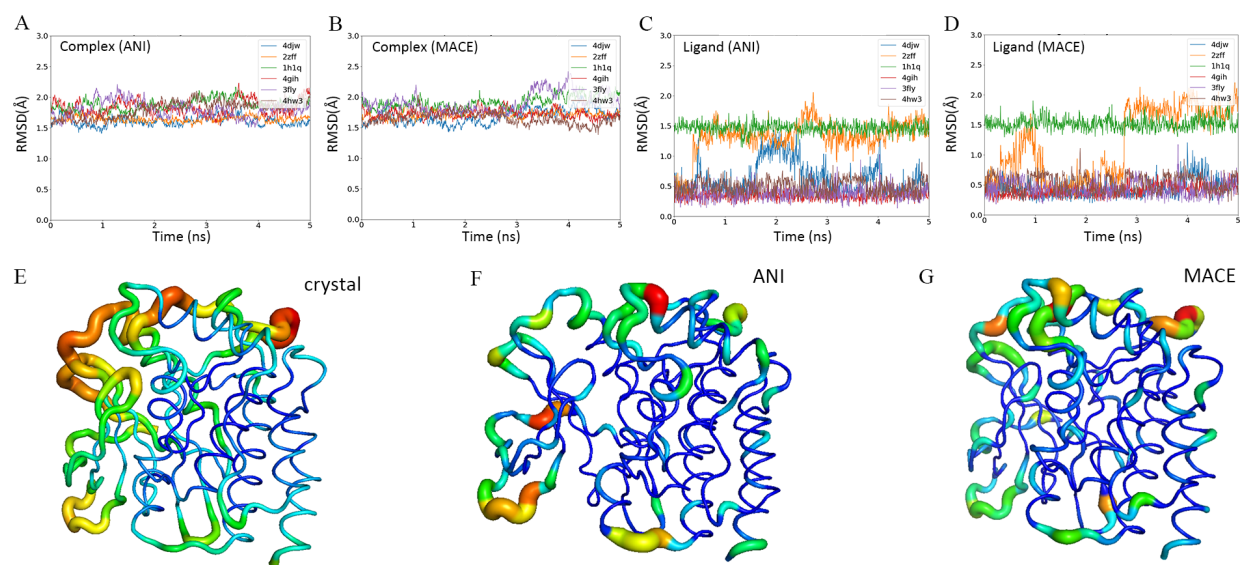


Figure 3: The protein-ligand simulation using an ML/MM approach. (A) and (B) illustrate the root mean square deviation (RMSD) of the entire protein-ligand complex calculated using ANI-2x and MACE-OFF23(S), respectively. (C) and (D) display the ligand RMSD based on results from the two MLIPs. (E), (F), and (G) show the B-factor color-mapped structures, where the B-factors are derived from the crystal structure and our simulation, respectively.

S4.

MM-PBSA, a popular endpoint free energy calculation method, is widely used in predicting protein-ligand binding free energy.<sup>65,66</sup> The performance of this method strongly depends on the quality of the samplings. We evaluated a protocol which applies ML/MM MD to sample the complex structure and then calculates MM-PBSA binding free energy with the same MMFFs utilized in the ML/MM mix potential. We evaluated this protocol using CDK2 receptor in complex with 19 ligands.<sup>63,64</sup> We found that sampling by MACE inference achieved the best performance with RMSE of 0.65 kcal/mol and  $R^2$  of 0.58, better than those of cMD, which were 0.68 kcal/mol and 0.54, respectively. The performance of ANI-2x inference is slightly worse than cMD (RMSE = 0.77 kcal/mol and  $R^2$  = 0.41). Overall, the quality of samplings by ML/MM MD is acceptable, and more test is going on for the endpoint free energy calculation protocol. The detailed MD sampling protocol and the result of free energy decomposition into different energetic terms were presented in the

Table S1-3.

### 3.5 Performance Test

Simulation performance is the main factor that limits the broad applications of QM/MM MD, whereas ML/MM MD provides a promising alternative. Therefore, for further ML/MM MD development, performance is a primary concern. We collected data from protein-ligand systems to evaluate the effectiveness of our built-in methods and parallel processing approaches in accelerating computations. All computation results were obtained using the NVIDIA L40s and the Intel Xeon Platinum 8462Y+.

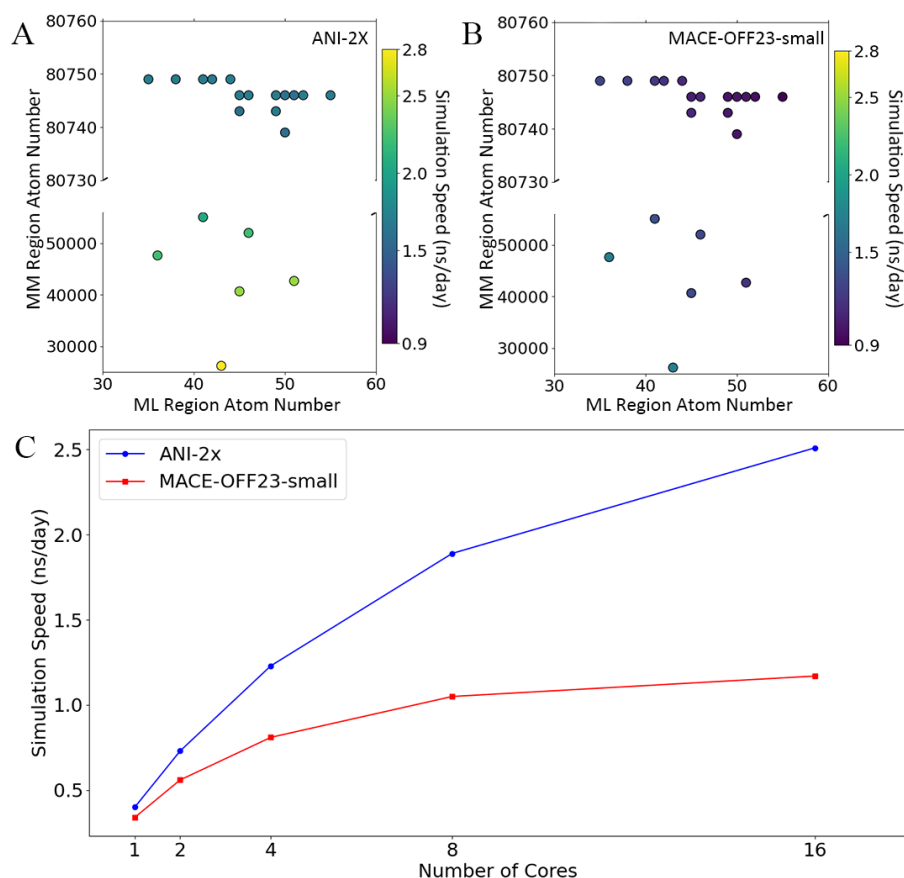


Figure 4: The performance evaluation of the ML/MM implementation in the Amber software suite. (A) and (B) illustrate the performance differences when the ANI-2x and MACE-OFF23(S) were used in ML/MM simulations of various protein-ligand complex scenarios, specifically focusing on the variation in atom numbers between the ML and MM regions. (C) shows the influence of the number of cores on simulation performance.



Figures 4A and 4B illustrate the performance of the ANI-2x and MACE-OFF23(S) models within the ML/MM framework. In most cases using ANI-2x, simulations achieve over 2 ns/day, while MACE operates at approximately 1.5 ns/day. Notably, all simulations used a timestep of 1 fs. When bonds involving hydrogen were constrained using the SHAKE algorithm,<sup>67</sup> the timestep can be extended to 2 fs. This approach is suitable for large systems where detailed hydrogen behavior is not the focus, effectively doubling the simulation performance.

Our implementation is designed to perform MD simulations on CPU and ML inference on GPU, with these two processes running in parallel. To achieve optimal performance, we need to carefully balance the workloads on CPU and GPU. For testing, we selected a protein-ligand complex with nearly 42,000 atoms (Figure 4C). We found that as the number of CPU cores increased up to 16 CPU cores, the simulation performance of both ML models improved. Especially, ANI-2x shows a steeper increase than MACE, indicating the more CPUs are desirable for ANI-2x than MACE. Before the simulation speed reaching the plateau of the curves, the bottleneck of ML/MM MD is the limited number of CPU cores, however, the bottleneck becomes the inference itself when the simulation speed reaches the plateau.

As shown in Figure 4, the ANI-2x model consistently outperforms MACE-OFF23(S) in terms of speed. Compared to ANI-2x, MACE is a larger model, capable of processing additional inputs such as environmental forces and cell size, and predicting properties like dipole moments and stress. Even though these features are not utilized in our current study, they do impact computational performance. Conversely, the ANI-2x model is designed for efficiency, focusing on the rapid prediction of molecular energies and atomic forces. Although MACE currently runs slower than ANI-2x, various methods, such as reducing model parameters<sup>29</sup> and implementing the JAX MD framework,<sup>68</sup> have been proposed to enhance its speed. Overall, compared to traditional QM/MM MD simulations, ML/MM enables nanosecond-timescale simulations with near-DFT level of accuracy.

## 4 Conclusion

In this study, we implemented the ML/MM approach within the Amber software suite and used it to validate the conservation of energy and momentum. This demonstrates that the ML region and MM region can exchange energy harmoniously, underscoring the robustness of our implementation. Notably, we proposed a new thermodynamic integration protocol for applying ML/MM mixed potentials in alchemical free energy calculations. This development offers new insights in biomolecular systems via high accurate and efficient free energy simulations enabled by multi-scale ML/MM potentials. Additionally, the high quality samplings by multi-scale ML/MM MD is achieved as the crystal structure properties can be very well predicted. Combining ML/MM MD sampling with the MM-PBSA endpoint free energy method, we are able to very well predict the binding free energies of a series of compounds binding to a receptor. Thus our approach can integrate seamlessly with other advanced methodologies to expand its functionality. Last, performance testing indicates that the ML/MM approach is a promising method for simulating molecular systems which can achieve a nanosecond per day sampling speed for a typical biological system in drug design.

In summary, ML/MM represents a promising direction for the future of molecular simulations. By leveraging the Amber platform, we anticipate that ML/MM can be extended beyond its current capabilities, and be easily combined with other advanced simulation technologies. We envision our future work will focus on integrating more MLIPs, further expanding the method's functionality, and enabling ML/MM to handle increasingly complex scenarios.

## 5 Supporting Information

Detailed theoretical hypothesis and formula derivation for thermodynamic integration in the ML/MM approach; computational details; validation of conservation laws using translational and rotational energy; B-factors derived from ML/MM MD simulations; structure of myeloid

cell leukemia 1 protein and B-factor color-mapped structures.

## Acknowledgement

This work was supported by the following funds from the National Institutes of Health (NIH) and the National Science Foundation (NSF): NIH R01GM147673 and NSF 1955260. The authors would like to thank the computing resources provided by the Center for Research Computing (facility RRID: SCR\_022735) at the University of Pittsburgh (NSF award number OAC-2117681), and the Pittsburgh Supercomputer Center (grant number BIO210185).

## References

- (1) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, J., Alexander D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (2) 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.
- (3) 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.
- (4) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguetta, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific

Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552.

- (5) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, J., Alexander D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (6) Huang, J.; MacKerell, J., Alexander D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.
- (7) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, J., Alexander D. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348–362.
- (8) Ivani, I. et al. ParmBSC1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55–58.
- (9) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, J., Alexander D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
- (10) Dickson, C. J.; Madej, B. D.; Skjevik, A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879.
- (11) 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.

- (12) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (13) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (14) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.
- (15) Brás, N. F.; Fernandes, P. A.; Ramos, M. J. QM/MM Study and MD Simulations on the Hypertension Regulator Angiotensin-Converting Enzyme. *J. Chem. Inf. Model.* **2022**, *62*, 3638–3650.
- (16) Ryde, U. The Coordination of the Catalytic Zinc Ion in Alcohol Dehydrogenase Studied by Combined Quantum-Chemical and Molecular Mechanics Calculations. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153–164.
- (17) Rod, T. H.; Ryde, U. Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol O-Methyltransferase. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.
- (18) Song, C.; Wang, L.-P. A Polarizable QM/MM Model That Combines the State-Averaged CASSCF and AMOEBA Force Field for Photoreactions in Proteins. *J. Chem. Theory Comput.* **2024**, *20*, 6632–6651.
- (19) Liu, C.; Jiang, H.; Li, Y.; Xue, B.; Yao, Y.-Y.; Yang, Z.-Z. Development of a QM/MM(ABEEM) Method Combined with a Polarizable Force Field to Investigate the Excision Reaction Mechanism of Damaged Thymine. *Phys. Chem. Chem. Phys.* **2023**, *25*, To be updated.

- (20) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.
- (21) Nam, K.; Gao, J.; York, D. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.
- (22) Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comput. Chem.* **2008**, *29*, 1019–1031.
- (23) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (24) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.
- (25) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148*, 241733.
- (26) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.
- (27) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.

- (28) Batatia, I. et al. A Foundation Model for Atomistic Materials Chemistry. *arXiv* **2023**, *2401.00096*, arXiv preprint.
- (29) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Kapil, V.; Witt, W. C.; Magdău, I.-B.; Cole, D. J.; Csányi, G. MACE-OFF23: Transferable Machine Learning Force Fields for Organic Molecules. *arXiv* **2023**, *2312.15211v2*, arXiv preprint.
- (30) Kovács, D. P.; Batatia, I.; Arany, E. S.; Csányi, G. Evaluation of the MACE Force Field Architecture: From Medicinal Chemistry to Materials Science. *J. Chem. Phys.* **2023**, *159*, 044118.
- (31) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, eaav6490.
- (32) Anstine, D.; Zubatyuk, R.; Isayev, O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *ChemRxiv* **2023**, Preprint.
- (33) Zubatyuk, R.; Smith, J. S.; Nebgen, B. T.; Tretiak, S.; Isayev, O. Teaching a Neural Network to Attach and Detach Electrons from Molecules. *Nat. Commun.* **2021**, *12*, 4870.
- (34) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (35) Ditchfield, R.; Hehre, W.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

- (36) Eastman, P. et al. OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials. *J. Phys. Chem. B* **2024**, *128*, 109–116.
- (37) Dickel, D.; Nitol, M.; Barrett, C. D. LAMMPS Implementation of Rapid Artificial Neural Network Derived Interatomic Potentials. *Comput. Mater. Sci.* **2021**, *196*, 110481.
- (38) Wang, Y.; Jaffrelo Inizan, T.; Liu, C.; Piquemal, J.-P.; Ren, P. Incorporating Neural Networks into the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2024**, *128*, 2381–2388.
- (39) Case, D. A. et al. Amber 2024. 2024; University of California, San Francisco.
- (40) Case, D. A. et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191.
- (41) 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.
- (42) González, M. Force Fields and Molecular Dynamics Simulations. *Coll. SFN* **2011**, *12*, 169–200.
- (43) Giese, T. J.; York, D. M. Charge-Dependent Model for Many-Body Polarization, Exchange, and Dispersion Interactions in Hybrid Quantum Mechanical/Molecular Mechanical Calculations. *J. Chem. Phys.* **2007**, *127*, 194101.
- (44) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (45) Paszke, A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada, 2019.
- (46) Wang, X.; Liu, H.; Wang, J.; Chang, L.; Cai, J.; Wei, Z.; Pan, J.; Gu, X.; Li, W.-L.; Li, J. Enzyme Tunnel Dynamics and Catalytic Mechanism of Norcoclaurine Synthase:



- Insights from a Combined LiGaMD and DFT Study. *J. Phys. Chem. B* **2024**, *128*, 9385–9395.
- (47) Wang, J.; Xu, Y.; Wang, X.; Li, J.; Hua, Z. Mechanism of Mutation-Induced Effects on the Catalytic Function of TEV Protease: A Molecular Dynamics Study. *Molecules* **2024**, *29*, 1071.
- (48) Cai, L.; Han, F.; Ji, B.; He, X.; Wang, L.; Niu, T.; Zhai, J.; Wang, J. In Silico Screening of Natural Flavonoids against 3-Chymotrypsin-like Protease of SARS-CoV-2 Using Machine Learning and Molecular Modeling. *Molecules* **2023**, *28*, 8034.
- (49) Wang, J.; Blount, P. Feeling the Tension: The Bacterial Mechanosensitive Channel of Large Conductance as a Model System and Drug Target. *Curr. Opin. Physiol.* **2023**, *31*, 100627.
- (50) Wu, Y.-L.; Zhou, C.; Liam, C.-K.; Zhang, Y.; Xia, F.; Zuo, Y. First-line Erlotinib versus Gemcitabine/Cisplatin in Patients with Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer: Analyses from the Phase III, Randomized, Open-Label, ENSURE Study. *Lung Cancer* **2015**, *89*, 1883–1889.
- (51) Wang, L.; He, X.; Ji, B.; Han, F.; Niu, T.; Cai, L.; Zhai, J.; Hao, D.; Wang, J. Geometry Optimization Algorithms in Conjunction with the Machine Learning Potential ANI-2x Facilitate the Structure-Based Virtual Screening and Binding Mode Prediction. *Biomolecules* **2024**, *14*, 648.
- (52) Han, F.; Hao, D.; He, X.; Wang, L.; Niu, T.; Wang, J. Distribution of Bound Conformations in Conformational Ensembles for X-ray Ligands Predicted by the ANI-2X Machine Learning Potential. *J. Chem. Inf. Model.* **2023**, *63*, 6608–6618.
- (53) Wang, L.; He, X.; Ji, B.; Han, F.; Niu, T.; Cai, L.; Zhai, J.; Hao, D.; Wang, J. Geometry Optimization Algorithms in Conjunction with the Machine Learning Potential

- ANI-2x Facilitate the Structure-Based Virtual Screening and Binding Mode Prediction. *Biomolecules* **2024**, *14*, 648.
- (54) Martins, S. A.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Prediction of Solvation Free Energies with Thermodynamic Integration Using the General Amber Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 3570–3577.
- (55) Chakravorty, A.; Hussain, A.; Cervantes, L. F.; Lai, T. T.; III, C. L. B. Exploring the Limits of the Generalized CHARMM and AMBER Force Fields through Predictions of Hydration Free Energy of Small Molecules. *J. Chem. Inf. Model.* **2024**, *64*, 4089–4101.
- (56) Karwounopoulos, J.; Åsmund Kaupang; Wieder, M.; Boresch, S. Calculations of Absolute Solvation Free Energies with Transformato—Application to the FreeSolv Database Using the CGenFF Force Field. *J. Chem. Theory Comput.* **2023**, *19*, 5988–5998.
- (57) Vassetti, D.; Pagliai, M.; Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1983–1995.
- (58) Garbett, N. C.; Chaires, J. B. Thermodynamic Studies for Drug Design and Screening. *Expert Opin. Drug Discovery* **2012**, *7*, 299–314.
- (59) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5*, 4611–4619.
- (60) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711–720.

- (61) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A Fast and High-Quality Charge Model for the Next Generation General AMBER Force Field. *J. Chem. Phys.* **2020**, *153*, 114502.
- (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, L. et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (64) Goel, H.; Hazel, A.; Ustach, V. D.; Jo, S.; Yu, W.; Alexander D. MacKerell, J. Rapid and Accurate Estimation of Protein–Ligand Relative Binding Affinities Using Site-Identification by Ligand Competitive Saturation. *Chem. Sci.* **2021**, *12*, Add specific pages if available.
- (65) Wang, J. M.; Hou, T. J.; Xu, X. J. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 287–306.
- (66) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478–9508.
- (67) 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.
- (68) Schoenholz, S. S.; Cubuk, E. D. JAX M.D. A Framework for Differentiable Physics. Advances in Neural Information Processing Systems. 2020.