

Hybrid ML/MM Simulations for Computational Enzymology: Recent Advances and Challenges

Published as part of *The Journal of Physical Chemistry Letters* special issue "80th Anniversary of Jilin University".

Xinhu Sha,[#] Chenyu Wu,[#] Daiqian Xie,^{*} and Yanzi Zhou^{*}



Cite This: *J. Phys. Chem. Lett.* 2026, 17, 8521–8534



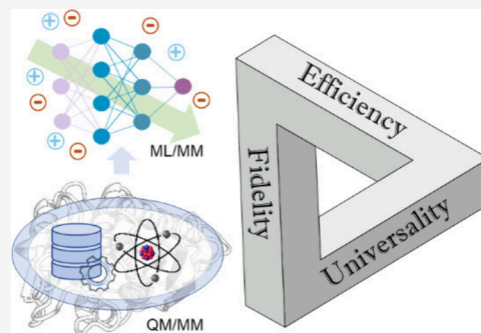
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The integration of QM/MM methods with molecular dynamics (MD) simulations has become a powerful tool for elucidating enzymatic reaction mechanisms at atomistic resolution, providing valuable insights for biocatalyst design and drug development. However, the high computational cost of QM methods, amplified by the extensive configurational sampling inherent in MD, remains a key bottleneck in studying complex enzymatic processes. Hybrid machine learning/molecular mechanics (ML/MM) methods, which replace the quantum calculations with machine-learned interatomic potentials, offer a promising alternative toward near-QM/MM accuracy at substantially reduced computational cost. This Perspective surveys the central methodological challenges in developing ML/MM frameworks, including the generation of high-quality reference data and the treatment of multiscale coupling. Emerging applications in biosystems are highlighted, and future directions are outlined toward accurate, efficient, and broadly transferable ML/MM models to support next-generation workflows for biomolecular simulations.



1. INTRODUCTION

Enzymes can accelerate biochemical reactions by several orders of magnitude under mild conditions, and their catalytic mechanisms have long been the focus of scientific research due to their high efficiency and specificity.^{1–4} A key goal in the field of enzyme catalysis is to develop highly effective catalysts for reactions.^{5–7} However, the limited understanding of structure–function relationships at the atomic level remains a major obstacle to rational enzyme design. Although traditional experimental methods, such as X-ray crystallography,^{8,9} nuclear magnetic resonance (NMR) spectroscopy,^{10,11} and kinetic analysis,^{12–14} can reveal static structures or apparent kinetics, they are difficult to capture transient reaction processes such as transition states or dynamic conformational changes. Molecular simulations and modeling offer a complementary approach, providing detailed, atomic-level insight into the fundamental processes of enzymatic catalysis.⁴

The reaction mechanism of an enzyme is not only determined by its active site, but also influenced by the environment.^{15,16} Furthermore, enzymes are inherently dynamic, exhibiting motions spanning multiple time and length scales that are pivotal to their biological function.^{17–20} Therefore, a thorough sampling of the conformational space is essential for precisely capturing the impact of the environment on enzymatic reactions.^{15,16,21} One of the most widely adopted approaches is combining quantum mechanical/molecular mechanical

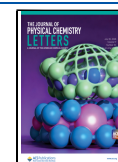
(QM/MM) potential,^{22–26} striking a balance between QM accuracy and MM sampling efficiency, and then employing molecular dynamics simulation (MD) or Monte Carlo simulation (MC) to sample the configuration space.^{15,16,21,27–29} This framework has enabled mechanistic investigations of increasingly complex reactive biosystems, such as ribozymes and riboswitches in RNA catalysis.^{30,31} Nevertheless, the accuracy of QM/MM methods depend on the QM level and the size of the QM region. Though high-level QM methods provide more accurate results, their prohibitive cost typically confines the size of the QM region to dozens of atoms and the MD simulation duration to a few tens of picoseconds,^{32–34} thereby limiting conformational sampling to obtain statistically reliable free energies. As a result, QM/MM-MD simulations, characterized by extensive and repetitive computations, have now reached a developmental plateau, highlighting the urgent need for a linear-scaling framework that

Received: May 16, 2026

Revised: June 24, 2026

Accepted: July 8, 2026

Published: July 16, 2026



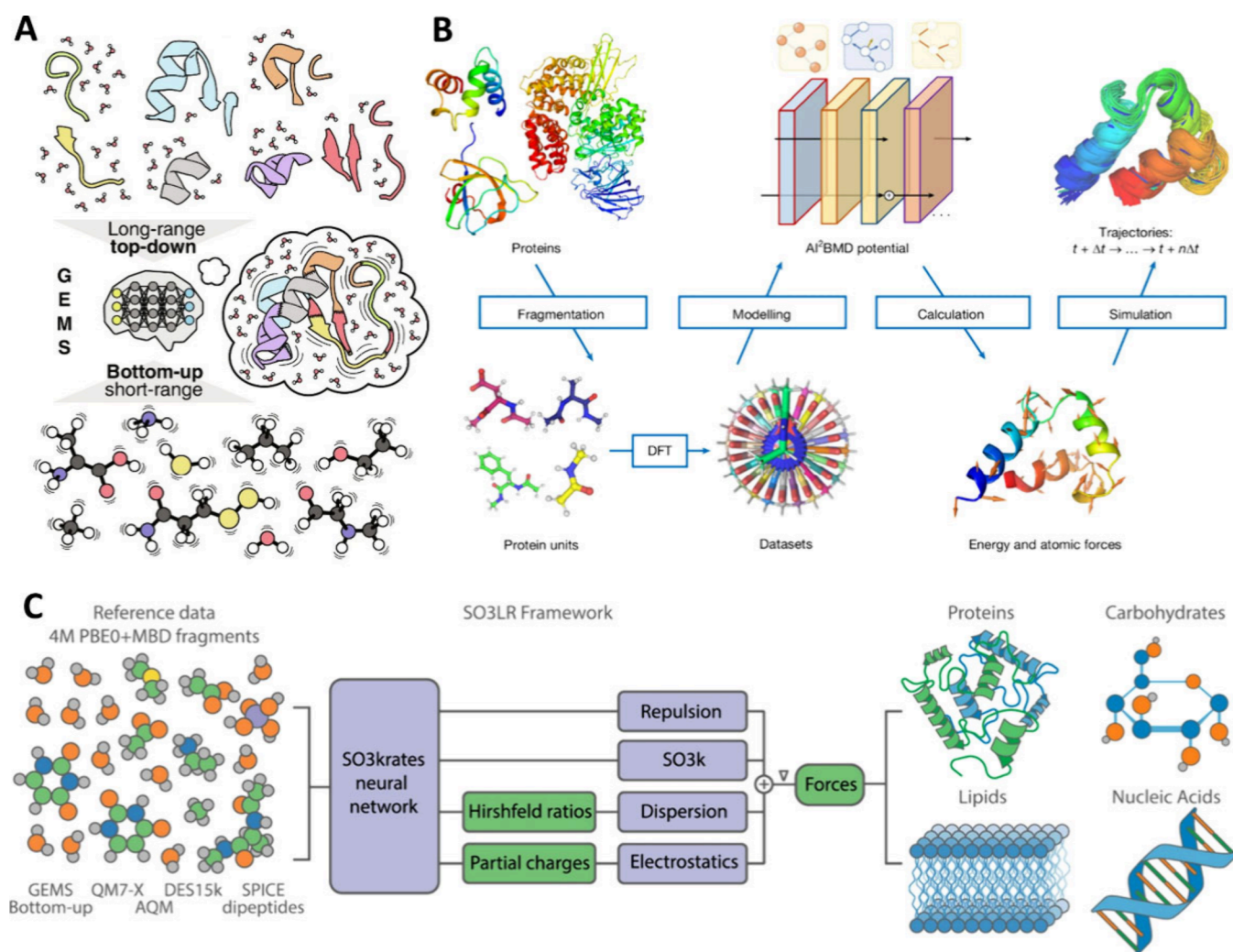


Figure 1. Representative examples of universal full-QM machine learning potentials pretrained for biomolecular systems. (A) Overview of the GEMS method, which follows a ‘bottom-up’ strategy to train small molecular pieces and a ‘top-down’ strategy in chemical fragmentation. [Image adapted from ref 44. Copyright 2024, American Association for the Advancement of Science.] (B) The overall pipeline of AI²BMD, where proteins are divided into protein units by a fragmentation process. [Reproduced from ref 45. Copyright 2024, Springer Nature.] (C) Overview of the SO3LR model pretrained on multiple data sets, which utilizes explicit empirical formulas to describe long-range dispersion and electrostatic interactions. [Reproduced from ref 46. Copyright 2025, American Chemical Society.]

can dramatically reduce computational cost while retaining ab initio accuracy.

The past two decades have witnessed a paradigm shift in molecular modeling.^{35,36} Supervised machine learning has catalyzed the development of MLPs as efficient alternatives to expensive QM calculations, which essentially employ parametrized functions to fit target physical properties with near-ground-truth accuracy.³⁷ The rapid evolution of machine learning algorithms and the growing availability of high-quality QM data sets have unlocked generic full-QM MLPs,^{37,38} which are pretrained and can be directly applied across various domains of chemical and materials science.^{39–41} These models are typically based on gas-phase or periodic data with only up to hundreds of atoms, and often rely on a locality assumption,⁴² limiting their ability to capture long-range interactions in large heterogeneous systems without special strategies.⁴³ As illustrated in Figure 1, for biosystems, pioneering studies have demonstrated that chemical fragmentation strategies can train a neural network and characterize protein molecular dynamics with ab initio accuracy, which enables precise free-energy

calculations for protein folding and thermodynamic property estimation consistent with experimental observations.^{44–46} Although the inference speed of these MLPs is orders of magnitude faster than ab initio calculations, they are not comparable to conventional force fields yet, motivating the development of the hybrid machine learning/molecular mechanics (ML/MM) approach.^{46–48} The ML/MM approach extends the QM/MM paradigm by greatly reducing computational costs while maintaining QM accuracy through machine-learned potentials for active centers, and leverages the high efficiency of the MM force field in describing nonreactive regions and long-range interactions. Therefore, the ML/MM method provides an ideal framework for describing complex biochemical reactions in solution, but several practical challenges remain to be addressed before its full potential can be realized.

A major challenge in developing robust MLPs for QM/MM systems lies in generating high-quality reference data sets, as training typically demands extensive QM/MM calculations to achieve adequate coverage of chemical space.^{49,50} In some cases,

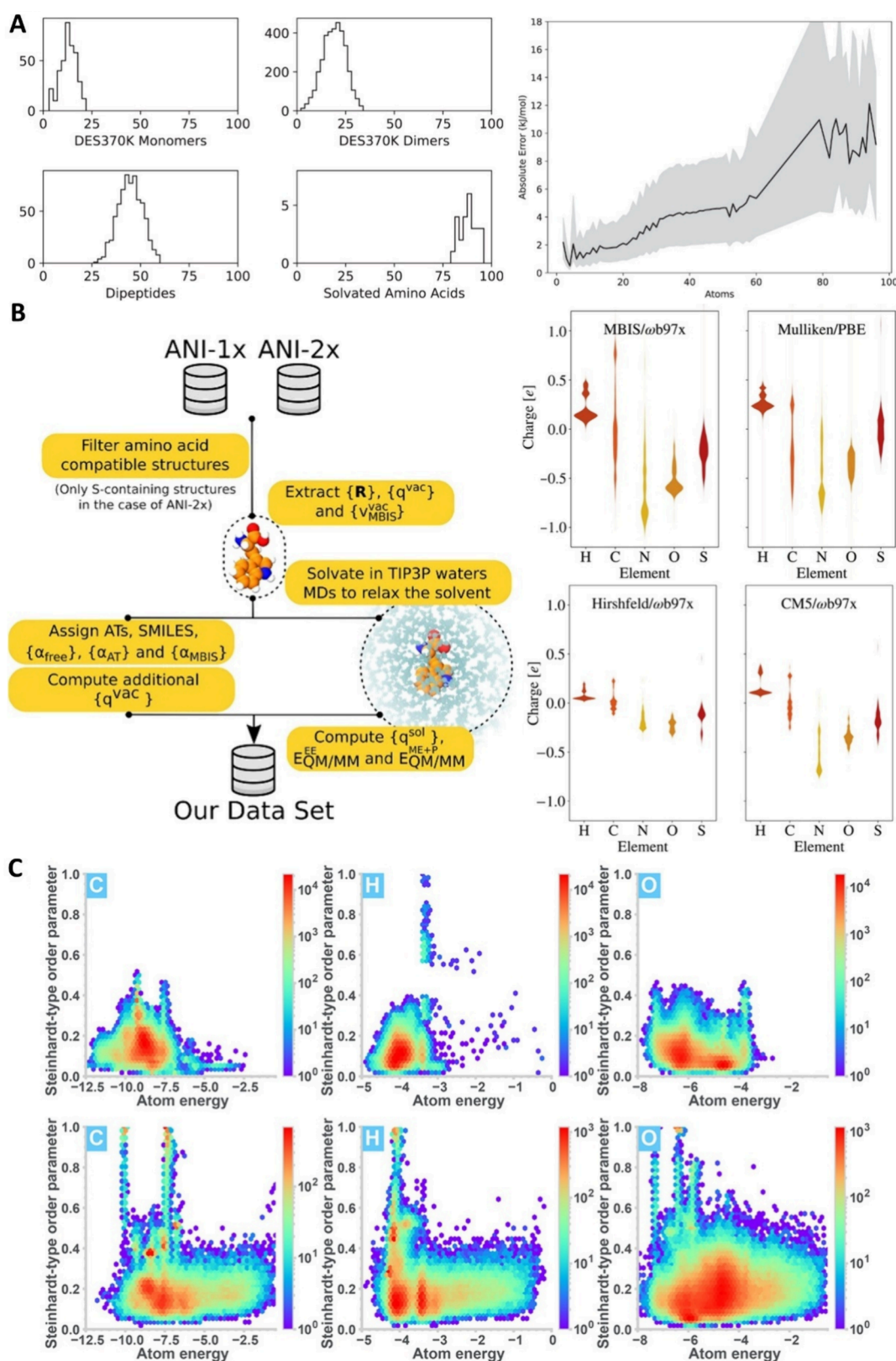


Figure 2. (A) Histogram of the size of molecules in each subset of SPICE and absolute error as a function of the total number of atoms in a molecule. [Reproduced from ref 66. Copyright 2025, Springer Nature.] (B) The flowchart used to construct QM/MM reference data set from public data sets ANI-1x and ANI-2x for assessing different ML/MM coupling schemes. [Reproduced from ref 57. Copyright 2024, American Chemical Society.] (C) Coverage analysis of the whole data set generated by MD with structures from a public database (the first row) compared to REICO (the second row). [Reproduced from ref 67. Copyright 2026, American Chemical Society.]

the cost of generating training data is comparable to that of the original QM/MM-MD simulations themselves. Fortunately,

homologous enzymes and engineered variants often share a conserved active site and analogous reaction mechanism.⁵¹

Therefore, formulating a coherent physical picture in the ML/MM framework and enhancing the transferability of ML/MM models across heterogeneous enzyme environments is critical to save the cost of the training process for new enzyme variants.

The fundamentally different treatments of electronic structure in QM and MM methods give rise to a critical challenge for the ML/MM framework: how to accurately describe the electrostatics, including electrostatic anisotropy and polarization effects, between different subsystems. The QM/MM electrostatic interactions can be handled at different levels of sophistication through mechanical, electrostatic, or polarized embedding schemes, with electrostatic embedding (EE) generally recommended.^{15,52} However, training an MLP for the QM region in the EE scheme is practically challenging, as the electrostatics of the QM region are coupled with the MM environment, introducing thousands of additional degrees of freedom from the surrounding MM atoms.⁵³

This Perspective overviews the key challenges in developing next-generation ML/MM models that demand higher computational efficiency, quantum-level accuracy, and broader transferability. Future progress hinges on advances in three closely connected areas: constructing high-quality training data sets, developing robust methodologies for multiscale coupling between ML and MM regions, and expanding applications to increasingly complex enzymatic systems. Among these, multiscale coupling represents both the central challenge and a defining feature that distinguishes ML/MM methods from gas-phase MLPs. In particular, the accurate and efficient treatment of electrostatic interactions arising from MM environmental polarization remains a major obstacle for physically grounded neural-network models, as discussed in detail below.

2. TRAINING SET CONSTRUCTION AND CHEMICAL SPACE COVERAGE

The applicability of MLPs is typically confined to the domain of their training data, with extrapolations to configurations far beyond this domain proving inherently unreliable.^{54,55} As shown in Figure 2A, fitting errors inevitably increase with system size, posing a great challenge for constructing high-quality data sets for large biosystems. In most ML/MM studies, researchers employ system-specific training data sets tailored to a given enzyme or solution system. While this strategy reduces the cost of QM reference calculations and improves fitting accuracy for tested systems, it inherently limits the coverage of chemical space and compromises transferability. Multiscale data sets with high diversity in QM chemical reaction space and enrichment in MM environments have become a key foundation for the development of transferable ML/MM methods. However, conventional MD sampling remains confined to near minimum-energy conformations and is insufficient to meet this demand.⁴⁹ To date, only a limited number of studies have reported multiscale data sets designed for universally applicable ML/MM models. Several ML/MM studies^{56–59} relabeled existing QM data sets with small organic molecules in the gas phase, such as QM7,⁶⁰ MD17,⁶¹ and ANI series,⁶² for application across diverse solution or even enzyme systems, as illustrated in Figure 2B. Collecting data sets from chemically diverse small molecules has proven effective in delivering accurate predictions for QM subsystems and MM environments not encountered during training.⁶³ For example, Zeng and co-workers trained a universal Δ -learning MLP based on multiple public data sets, capable of correcting DFTB3/3OB results to ω B97X/6-31G* level.⁶⁴ More recently, the Riniker group

reported the BMS25 data set solvated with TIP4P_FB water point charges, which contains configurations with up to 487 atoms in the QM region and nearly 8000 atoms in the MM region.⁶⁵ This data set is poised to guide the development of generic ML/MM models for biomolecular systems.

Significant progress in the exploration of chemical space has been achieved across chemical and materials science. Several data sets provide specific subsets of configurations for target applications. For example, SPICE⁶⁶ aggregates drug-like molecules, DORTS⁶⁸ samples reaction trajectories via normal mode sampling⁶⁹ (NMS), and QCML spans diverse molecular geometries and various electronic states.⁷⁰ Increasing attention has recently shifted toward the systematic inclusion of nonequilibrium configurations, which remain largely underexplored in ML/MM data set construction.⁷¹ An alternative strategy seeks to enhance neural network transferability by guiding models toward fundamental intermolecular interactions through random exploration of chemical space. ANI-1xnr employs a nanoreactor-based sampling strategy, introducing a random schedule function for oscillating temperature and density to accelerate the reaction process and explore the reaction path.⁷² Similarly, Figure 2C illustrates a distinct random exploration via imaginary chemicals optimization (REICO) sampling strategy developed by Hu and co-workers.⁴¹ Rather than relying on conventional samplers such as MD in other works, they perform geometry optimizations on randomly generated element-based structures. Coverage analysis in Figure 2C confirms that this strategy not only outperforms MD sampling but also extends exploration into high-energy regions critical for ensuring the robustness and generalizability of the resulting MLPs.⁶⁷ Although the data sets of these works were originally developed for gaseous or periodic material systems, and are not yet directly applicable to biological systems, they nevertheless offer promising directions for developing more transferable ML/MM potentials in biosystems.⁴⁴

3. MULTISCALE COUPLING AND PHYSICS-BASED EMBEDDING

Transferability is particularly important for practical applications in biosystems. However, without physically interpretable representations, neural network potentials may only learn spurious correlations within limited training data sets, leading to overfitting and unstable energy predictions under even minor structural perturbations.^{51,73} The no-free-lunch theorem underscores that inductive bias constitutes one of the most effective strategies for achieving generalization with limited labeled data.⁷⁴ Specifically, for MLPs in the gas phase, the mapping from atomic coordinates and species to molecular properties⁴⁸ typically embeds several physics-based inductive biases to enhance predictive performance. These include the formula-consistent derivation of physical quantities, such as forces and dipole moments;^{59,75} the preservation of rotational, permutational, and inversional symmetries through global descriptor-based models,⁷⁶ local descriptor-based models,⁷⁷ and message passing networks;^{78,79} and the equivariance of higher-order tensors enabled by equivariant graph neural networks (EGNN), such as NequIP,⁷³ MACE,⁸⁰ and Allegro.⁸¹

When extending MLPs to ML/MM frameworks, additional inductive biases become necessary to account for interactions with the surrounding environment. In particular, models should preserve global subsystem properties (such as the charge state of QM regions),⁸² provide a unified and computationally efficient representation of external charges from the MM environment,

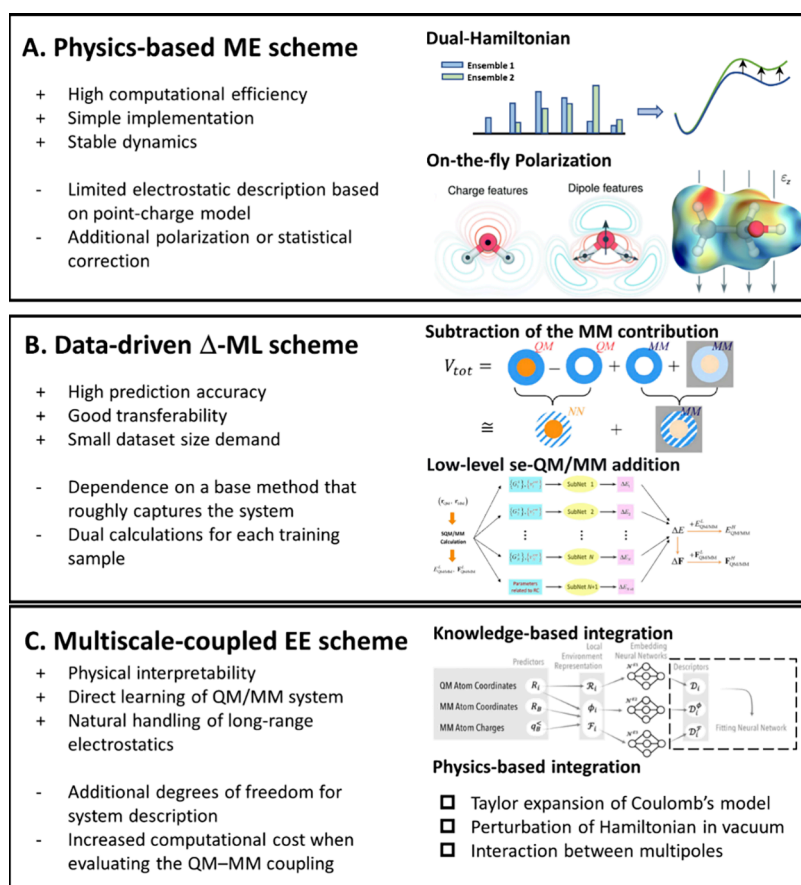


Figure 3. Summary of embedding schemes in ML/MM multiscale frameworks. (A) Physics-based ME scheme. Beyond point-charge interactions, the deviation from the EE scheme is bridged by the dual-Hamiltonian method and the on-the-fly polarization methods. [Image reproduced from ref 59. Copyright 2021, The Royal Society of Chemistry.] (B) Data-driven Δ -ML scheme. The electrostatic difference between high-level and low-level QM/MM results is learned by a neural network. The low-level reference can be either a molecular force field (reproduced from ref 83, Copyright 2022, American Chemical Society) or semiempirical QM/MM calculations (reproduced from ref 84, Copyright 2018, American Chemical Society). (C) Multiscale-coupled EE scheme, including knowledge-based integration and physics-based integration methods. [Image adapted from ref 51. Copyright 2025, American Chemical Society.]

and ensure physically consistent electrostatic interactions whose long-range asymptotic behavior correctly recovers Coulombic interactions.

Incorporating physics-based representations of multiscale coupling with embedding schemes of QM/MM methods has become a longstanding issue in the field. Before detailing ML/MM frameworks built upon these inductive biases, we first review the mainstream EE and ME schemes in QM/MM simulations. In the EE scheme, MM point charges are treated as a one-electron perturbation to the QM Hamiltonian in vacuo $\hat{H}_{\text{QM}}^0$, yielding an effective Hamiltonian $\hat{H}^{\text{eff}}$:

$$\begin{aligned}\hat{H}^{\text{eff}} &= \hat{H}_{\text{QM}}^0 + \hat{H}_{\text{QM/MM}}^{\text{elec}} \\ &= \hat{H}_{\text{QM}}^0 + \sum_{i,j} \left(\frac{Z_i q_j}{|R_i - R_j|} - \int \frac{q \rho(r')}{|r' - R_j|} dr' \right)\end{aligned}\quad (1)$$

where Z_i and q_j are the charges of QM nucleus i and MM point charge j , respectively, with R_i and R_j being their atomic coordinates. ρ is the density of electron and r' is its coordinate.

For brevity, we denote the eigenvalue of the effective Hamiltonian as $E_{\text{QM-MM}}^{\text{elec}}$, which decomposes into the energy of the QM subsystem in vacuo and the electrostatic interactions between the QM and MM subsystems:

$$\begin{aligned}E_{\text{QM-MM}}^{\text{elec,EE}} &= E_{\text{QM}}^0 + E_{\text{QM/MM}}^{\text{elec}} = \langle \Phi | \hat{H}^{\text{eff}} | \Phi \rangle \\ &= \langle \Phi | \hat{H}_{\text{QM}}^0 + \hat{H}_{\text{QM/MM}}^{\text{elec}} | \Phi \rangle\end{aligned}\quad (2)$$

where Φ is the eigenfunction of $\hat{H}^{\text{eff}}$.

In contrast, the ME scheme decouples these two contributions, and the electrostatic interaction between QM subsystems and MM subsystems simply follows Coulomb's law:

$$\begin{aligned}E_{\text{QM-MM}}^{\text{elec,ME}} &= E_{\text{QM}}^0 + E_{\text{QM/MM}}^{\text{elec}} \\ &= \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle + \sum_{j \in \text{MM}} \left(\sum_{i \in \text{QM}} \frac{Q_i}{|R_i - R_j|} \right) q_j\end{aligned}\quad (3)$$

where Φ_0 is the eigenfunction of $\hat{H}_{\text{QM}}^0$, and Q_i is the point charge of QM atom i .

Within ML/MM frameworks, electrostatic coupling can be described by three distinct embedding schemes as illustrated in Figure 3.

3.1. Physics-Based ME Scheme

In the traditional ME scheme, the QM/MM electrostatic interactions are computed under Coulomb's law with fixed charges. That is, the partial charges of QM atoms are

predetermined and kept constant throughout the simulation, which is unreasonable. Consequently, numerous studies have focused on updating partial charges of QM atoms in response to configurational changes during simulation.^{56,58,59,85–89} These works constructed an additional interface that provides approximate electrostatic interactions for host MLP⁹⁰ trained in vacuo or semiempirical QM/MM calculations. In fact, selecting an appropriate charge partitioning model as training labels, such as MBIS charges, and updating the QM atomic charges as a function of conformation has long been recognized as an important strategy in improving simulation accuracy.^{57,58} In our previous work, we developed an optimized ME scheme integrated with the embedded atom neural network (EANN),⁹¹ in which the partial charges of QM atoms in vacuo are predicted by the neural network, and the QM/MM electrostatic interactions are calculated by Coulomb's law.^{85,86} To enforce charge conservation and describe different charge states of the QM system, either the charge equilibration (Qeq) methods,^{82,92} or computing charge one by one via the formula $\langle q \rangle = \frac{Q_{\text{tot}}}{N} - \frac{1}{N} \sum_i^N q_i$ for simplicity, is widely adopted in ML/MM models.^{86,87,93}

In practice, simple Coulomb's interaction cannot faithfully reproduce the QM/MM results in the EE scheme, prompting the development of several approaches that perform low-level QM/MM simulations as a foundation and statistically calibrate the results to reference values. One such approach, the dual-Hamiltonian approach, samples configurations from the low-level configuration space and computes free energy in targeted ensembles via thermodynamic perturbation (TP) or its variant methods.⁹⁴ For example, several studies^{33,95,96} reweight the ML/MM sampling back to the QM/MM level and verify that their developed methods successfully reproduce the distribution of ensembles in QM/MM methods and thus obtain similar free energy results. Our group has also corrected the polarization term of the MM subsystem on the electrostatic coupling by thermodynamic perturbation after molecular dynamics simulations, and successfully applied this approach in organic reactions and enzyme mutations.^{85,86} However, this statistical correction approach suffers from notable drawbacks: the conformational spaces of the two Hamiltonians must overlap sufficiently; otherwise, reweighted free energies exhibit significant statistical noise or even fail to converge.

In the EE scheme, the description of QM/MM electrostatic interactions extends far beyond simple Coulombic terms—the QM electron density must be polarized by the electrostatic environment, and the point charge model cannot faithfully reproduce the true charge distribution near QM subsystems.¹⁵ Accordingly, numerous studies employ alternative physical quantities, such as dipoles, to approximate real charge distribution in the EE scheme, thereby enhancing the transferability of the trained model. To account for the polarization effects in the ME scheme, Gastegger developed the FieldSchNet framework.⁵⁹ In the method, solvent molecules are represented as an electrostatic field, and the polarized charges of QM subsystems are treated as a response property and derived from energies and electrostatic fields. An alternative strategy explicitly introduces a polarization term. In the ANI-MBIS package, beyond Coulomb interactions, the polarization term is expressed as a function of the surrounding dipole moment and the dielectric constant of QM atoms.⁵⁸ Similarly, in the EMLE framework developed by Zinovjev, the QM/MM electrostatic energy is divided into static and induced components.^{56,63,87}

The static term derives the electrostatic potential from MBIS partitioning to replace the point-charge model in a damped form, while the induced term is treated as a correction to atomic polarizabilities, employing a modified dipole–dipole interaction tensor to compute the induced electrostatic potential.⁹⁷

3.2. Data-Driven Δ -ML Scheme

In QM/MM methods, several studies correct low-level simulations on the fly through sparse high-level calculations or parametrized models, including multiple time step (MTS)⁹⁸ and force-matching methods.^{61,99,100} For ML/MM frameworks, direct simulation lacks the capacity to capture long-range interactions and suffers from prohibitive computational requirements coupled with slow convergence, potentially compromising the stability of long-term simulations and the reliability of dynamic results.^{71,101} Consequently, a major branch of Δ ML/MM works has been regarded as a substitute for direct ML/MM simulations, which can be considered as “data-driven calibration”.^{33,53,84,101–104} For example, Yang et al. introduce the MM electrostatic potential and the reaction coordinate as additional descriptors to train the potential energy difference between ab initio and semiempirical QM/MM methods^{33,84} or MM force fields.⁷¹ BuRNN defines a buffer region encompassing QM atoms and nearby MM atoms, employing SchNet to capture the difference in both energy and atomic forces.^{83,105}

3.3. Multiscale-Coupled EE Scheme

The pursuit of a transferable machine learning potential has long been regarded as the holy grail of computational chemistry, demanding that the “offline-trained” PESs apply to various unlabeled configurations in order to substantially reduce the computational cost associated with retraining.^{106,107} Advancing multiscale coupling toward physics-based integration of QM and MM subsystems is one of the most effective and economical strategies, and numerous ML/MM works adopt the EE scheme to calculate the QM energy in vacuo and QM/MM electrostatic energy in an end-to-end fashion.^{51,96,101,108–111} A distinct advantage of these EE scheme-based frameworks is that forces on all related atoms can be directly obtained via automatic differentiation of the energy with respect to atomic coordinates in PyTorch.¹¹² Unlike earlier schemes, these approaches seek to integrate the influence of MM atoms in more sophisticated ways, rather than through a simple additive combination. The central challenge in training these models lies in representing the MM environment at an affordable computational cost, despite its thousands of degrees of freedom.

Several works distinguish QM–QM, QM–MM, and MM–MM atomistic pairs interactions using separate descriptor sets and interaction blocks,^{51,101} which may be termed as “knowledge-based framework”.³⁶ Nam and co-workers constructed MM-environment descriptors analogous to the well-known DeePot-SE descriptor and assigned the three atomistic pair types to distinct embedding neural networks within the DeepMD architecture,¹¹³ where direct ML/MM and Δ ML/MM results are systematically compared. Another series of work incorporated physics-based descriptors³⁶ to describe the perturbation from MM polarization, focused primarily on the external potential^{71,96} or field^{51,101} generated by the environment. Gao and co-workers treated the polarization energy from the MM environment as a higher-order perturbation to the QM energy in vacuo, with the energy contribution from this term averaged over all neighboring MM charges for the QM regions.⁹⁶ Similarly, Sugita and co-workers utilized a Taylor expansion of Coulomb's law to represent the external potential,

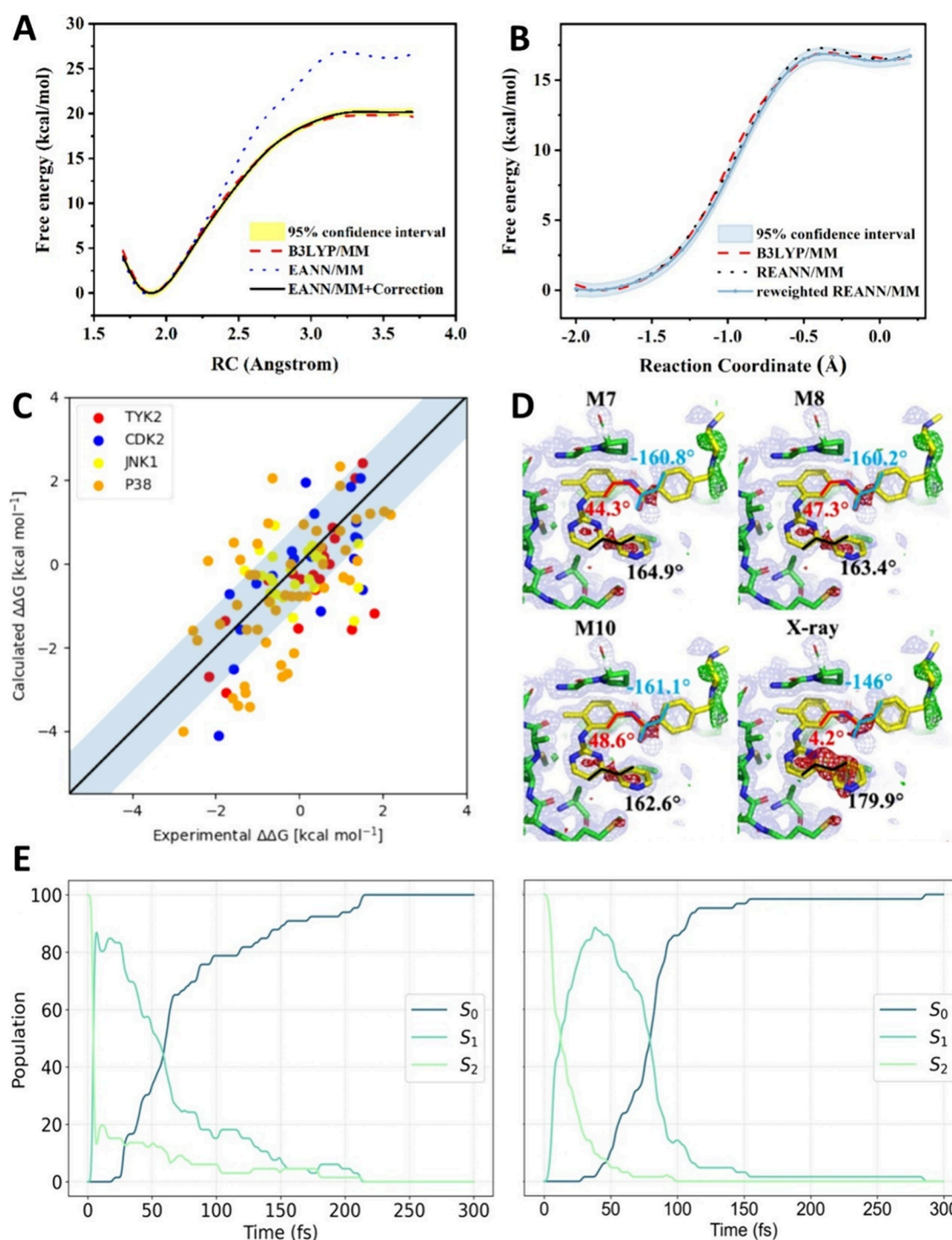


Figure 4. Examples of ML/MM applications. (A) Solution systems. Reweighted free energy profile for the S_N1 fragmentation of *tert*-butyl chloride calculated from wTP calculations, direct QM/MM-MD, and ML/MM-MD simulations. [Reproduced from ref 85. Copyright 2023, American Chemical Society.] (B) Enzyme systems. Free energy profiles for the R513A mutant of COX-2 from the B3LYP/MM-MD, direct REANN/MM-MD, and reweighted REANN/MM-MD simulations smoothed by Gaussian process regression. [Reproduced from ref 86. Copyright 2025, American Chemical Society.] (C) Protein–ligand binding free energies. Results for 108 relative binding free energies $\Delta\Delta G$ in the work from Torpenter with AIMNet2. [Reproduced from ref 119. Copyright 2025, American Chemical Society.] (D) Quantum refinements. Structures for the binding sites of imatinib in spleen tyrosine kinase from various quantum refinements using different machine-learning-integrated multiscale schemes, compared with the X-ray crystal structure. [Reproduced from ref 120. Copyright 2024, Springer Nature.] (E) Photodynamics simulation. Population curves from photodynamics simulations using the Field-MACE model (left) and quantum chemistry (right) for the ring-opening reaction of furan following excitation to the second excited singlet state (S_2). [Image adapted from ref 115. Copyright 2026, Springer Nature.]

thereby accounting for long-range interactions and polarization within EGNN.¹⁰⁹ Notably, Riniker and co-workers recently proposed the anisotropic message-passing (AMP) formalism to

describe long-range interactions in ML/MM frameworks.^{93,108,114} The electrostatic energy, expressed as multipole–multipole interactions, appears to be a promising

direction; however, accurately describing short-range interactions near the QM/MM boundary remains challenging.^{108,115}

3.4. Approximate EE Scheme with ME Simplicity

According to eqs 1 and 2, the QM/MM electrostatics can be expressed as

$$\begin{aligned} E_{\text{QM/MM}}^{\text{elec,EE}} &= \langle \Phi | \hat{H}^{\text{eff}} | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle \\ &= \langle \Phi | \hat{H}_{\text{QM}}^0 | \Phi \rangle + \langle \Phi | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle \\ &= \langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle + \langle \Phi | \hat{H}_{\text{QM}}^0 | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle \\ &\quad + \langle \Phi | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle \end{aligned} \quad (4)$$

with

$$\begin{aligned} E_{\text{dist}} &= \langle \Phi | \hat{H}_{\text{QM}}^0 | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle \\ E_{\text{pol}} &= \langle \Phi | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi \rangle - \langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle \end{aligned} \quad (5)$$

Based on linear response theory, Gao and Xia demonstrated as early as 1992 that nearly half of the QM–MM electron relaxation energy is used to distort the QM subsystem,¹¹⁶ i.e.,

$$E^{\text{dist}} \approx -\frac{1}{2} E^{\text{pol}} \quad (6)$$

The QM/MM electrostatics can be approximately calculated by

$$\begin{aligned} E_{\text{QM/MM}}^{\text{elec,EE}} &\approx \langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle + \frac{1}{2} (\langle \Phi | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi \rangle \\ &\quad - \langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle) \\ &= \frac{1}{2} (\langle \Phi_0 | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi_0 \rangle + \langle \Phi | \hat{H}_{\text{QM/MM}}^{\text{ele}} | \Phi \rangle) \end{aligned} \quad (7)$$

The expectation values of $\hat{H}_{\text{QM/MM}}^{\text{ele}}$ in eq 7 can be calculated via the electrostatic potential of QM atoms at the MM atom sites. Substituting eq 7 into eq 3 yields

$$\begin{aligned} E_{\text{QM-MM}}^{\text{elec}} &= E_{\text{QM}}^0 + E_{\text{QM/MM}}^{\text{elec}} \\ &\approx \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle + \frac{1}{2} \sum_{j \in \text{MM}} (V_j^0 + V_j) q_j \\ &= \langle \Phi_0 | \hat{H}_{\text{QM}}^0 | \Phi_0 \rangle + \sum_{j \in \text{MM}} V_j' q_j \end{aligned} \quad (8)$$

where V_j^0 and V_j denote the electrostatic potentials generated by the QM subsystem at the MM atomic site j in vacuo and in the MM environment, respectively. q_j is the point charge of the MM atom j , assigned by the MM force field and held fixed. Relative to eqs 2 and 3, eq 8 constitutes an approximation to the EE scheme and preserves the simplicity of the ME scheme. The total system energy is elegantly decoupled, and the QM/MM electrostatic energy is reformulated in terms of the electrostatic potential of polarized and unpolarized QM subsystems, which is predicted by the MM atom electrostatic field-induced recursively embedded atom neural network (MIREANN) in our work.¹¹⁷ With a more precise representation of electrostatic potential (ESP) in neural networks, such as AMP,⁹³ this methodology is expected to achieve higher accuracy and extend to more complex enzyme systems.

4. EMERGING APPLICATIONS OF ML/MM FRAMEWORKS

Explicit and accurate description of solute–solvent interactions serves as one of the most fundamental ‘touchstones’ for evaluating the performance of ML/MM frameworks, and many early studies have focused on organic reactions or the configuration of biomolecules in solvents. For example, Figure 4A compares the free energies from the ML/MM and QM/MM simulations for the S_N1 fragmentation of *tert*-butyl chloride from our group.⁸⁵ Other representative test systems span the S_N2 substitution of halomethanes,^{85,109} the torsion of peptides,^{64,71,108} and the Claisen rearrangement of allyl vinyl ether.^{33,96} Quantitative metrics, including fitting errors, radial distribution functions (RDFs), free energy surfaces, and average bond lengths, provide rigorous assessments of framework performance.¹¹⁸ Moreover, field-responsive MLPs enable the prediction of dipole moments under external electric fields,^{59,75} thereby facilitating the investigation of solvent effects on infrared spectra via ML/MM simulations.^{59,83,105} Nevertheless, applying ML/MM solely to describe solvent effects on the QM region leaves much of its potential for biomolecular systems untapped. Below, we highlight several emerging and promising research directions that extend ML/MM frameworks beyond conventional solvent benchmark systems.

4.1. Enzyme Reaction Mechanism

When applied to reactions in protein environments, only a few studies have reported long-time ML/MM-MD simulations to date, and these generally exhibit significantly larger fitting errors than those observed in liquid-phase systems.^{53,63,101,102,109,121} Many of these works employ Δ -learning strategies calibrated against MM force fields or semiempirical QM calculations to mitigate collapse issues, which may introduce either additional computational cost or numerical instability, particularly for bond breaking and formation.⁷¹ Consequently, there is an urgent need for universal and robust ML/MM frameworks capable of directly simulating enzyme catalysis.⁴⁷ In this regard, recently we have developed direct ML/MM models that have been successfully applied to cyclooxygenase (COX)⁸⁶ (Figure 4B) and to PETase.¹¹⁷ Furthermore, metalloenzyme catalysis is a common chemical process in biological systems, in which both spin-multiplicity transitions and radical formation challenge the development of robust MLPs.¹⁰⁷ Several metalloenzyme systems have been investigated using ML/MM simulations,^{122,123} but as these works are largely system-specific and focus primarily on the construction of MLPs with multiple spin states rather than on multiscale coupling, we do not elaborate further here. The applications of representative ML/MM implementations in enzymatic catalysis or organic reactions in solutions are summarized in Table 1.

4.2. Ligand Binding

Protein–ligand interactions are ubiquitous in biology, driving extensive efforts to predict binding affinity via binding free energies (Figure 4C) to guide drug design.^{119,124–126} These studies require neural networks to accurately explore more intricate stereochemical space, and consequently, they often rely on either system-specific active learning strategies or transferable pretrained MLPs. Beyond free energy calculations, several groups^{120,127,128} have extended ML/MM frameworks to the geometry optimization of biomacromolecules, aiming to reproduce the results of quantum refinement (QR) methods. These applications impose even more stringent demands on the

Table 1. Representative ML/MM Applications in Enzymatic Catalysis and Organic Reactions

Physics-based ME schemes			
Reference	ML size	Representative applications	Charge model
Zinovyev et al. ^{56,87}	22, 66	Alanine dipeptide; SARS-CoV-2 Mpro	MBIS (Qeq) + Thole model
Gradisteanu et al. ⁶³	24, 49	Diels–Alderase reaction; Chorismate mutase	MBIS (Qeq) + Thole model
Zhou et al. ⁸⁵	6, 14	S _N 1, S _N 2 reactions	ESP (Qeq) + wTP correction
Sha et al. ⁸⁶	26	Cyclooxygenase	ESP (Qeq) + wTP correction
Gastegger et al. ^{59,88}	6–23	Claisen rearrangement	Atomic polar tensor charges
Semelak et al. ⁵⁸	10–73	Alanine dipeptide; Avidin/FKBP12	MBIS + Polarization model
Data-driven Δ -ML schemes			
Reference	ML size	Representative applications	Calibration level
Gomez-Flores et al. ¹⁰²	15	Thiol–disulfide exchange	DFTB $\rightarrow$ CCSD(T)
Böselt, Hofstetter et al. ⁵³	7–63	S _N 2 reactions; S-adenosylmethionine/cytosine	DFTB $\rightarrow$ DFT
Shen et al. ^{33,84}	6–14	S _N 2 reaction; Proton transfer; Claisen rearrangement	AM1, DFTB $\rightarrow$ DFT
Song et al. ⁷¹	22	Alanine dipeptide	MM $\rightarrow$ DFT/MM
Zeng, Giese et al. ^{95,104}	~20–30	Phosphoryl transfer reaction	DFTB2 $\rightarrow$ MNDO/d
Multiscale-coupled EE schemes			
Reference	ML size	Representative applications	Coupling method
Pan, Arattu et al. ^{51,101}	9–69	Menshutkin reaction; Chorismate mutase; Dihydrofolate reductase	Fitting neural network
Lei et al. ^{50,109}	6–35	S _N 2 reaction; Proton transfer; Michael addition in chalcone isomerase; Kemp elimination	Taylor expansion
Xie et al. ⁹⁶	14, 22	Alanine dipeptide; Claisen rearrangement	Energy perturbation
Sha et al. ¹¹⁷	6–63	S _N 2 reaction; PETase; Cyclooxygenase	Linear response theory
Pultar et al. ¹⁰⁸	~22–350	Alanine dipeptide; Ni–P complex; Pyridine and quinoline dimers	Multipole interactions

models, as accurate geometry optimization requires high-fidelity PESs across broad configurational spaces. Figure 4D provides a representative example by comparing structures for the binding sites of imatinib in spleen tyrosine kinase from various quantum refinements using different machine-learning-integrated multiscale schemes and the X-ray crystal structure. Accurate structural characterization of multiscale systems containing cofactors, ligands, or reactive intermediates remains challenging for current AI-based structure prediction methods, underscoring the potential value of physics-based ML/MM approaches in this context.¹²⁹

4.3. Excited-State Dynamics

Several studies have applied ML/MM frameworks to non-adiabatic effects and excited-state dynamics. Zinovyev and co-workers combined their EMLE framework with MM force fields for conformational sampling via ML/MM simulations. High-level QM/MMPol¹³⁰ calculations were subsequently carried out

on selected snapshots to evaluate properties of excited states, analogous to the dual-Hamiltonian approach.¹³¹ Another study involved excited-state nonadiabatic dynamics through ML/MM-based trajectory surface hopping¹³² simulations, as shown in Figure 4E, which utilizes a neural network to predict energies and forces across multiple electronic states, enabling faithful reproduction of the electronic dynamics and structural rearrangements observed in QM/MM surface hopping reference simulations. These pioneering efforts demonstrate the feasibility of excited-state multiscale simulations of complex biosystems, which were computationally prohibitive before.^{115,133}

5. CHALLENGES AND OUTLOOK

As an emerging paradigm that integrates the QM/MM approach with machine learning, the development of reliable ML/MM potentials remains at an early stage. The major challenges impeding progress in ML/MM frameworks can be broadly categorized into three interconnected aspects: efficiency, fidelity, and universality.

5.1. Efficiency

Recent studies suggest that increasingly complex physically motivated constraints may limit MLP scalability, and many current implementations remain insufficiently optimized for parallelization on GPUs.¹³⁴ In addition, for ML/MM frameworks, the increasingly sophisticated representations of MM atoms, together with tighter coupling between QM and MM atoms within neural architectures, impose additional computational burdens on the inference efficiency.⁵³ Another subtle yet critical issue concerns the compatibility between the neural network inference interface and the molecular dynamics program. In many cases, the Python-based ML models and MD software based on compiled languages (C or Fortran) require specialized interfaces for data exchange, for example, the socket communication of Sugita's work¹⁰⁹ to leverage PyTorch. Direct ML/MM simulations frequently suffer from numerical instability in practice, largely owing to limited model transferability and insufficient sampling coverage in chemical reaction space.^{51,71} Iteratively expanding the training data sets through on-the-fly sampling for each new system is computationally unsustainable and incompatible with the long-term goal of universal ML/MM potentials.

5.2. Fidelity

Theoretical studies of enzyme mutants impose extremely high demands on the accuracy of ML/MM frameworks in predicting atomistic behavior. For modern full-QM MLPs, the prediction of energies for small molecules in vacuo can easily achieve “chemical accuracy”.⁷³ However, unlike in periodic materials science, the simple application of point-charge models within Ewald summation¹³⁵ to compute long-range electrostatics for biosystems is insufficient; the central challenge of ML/MM research instead lies in the physically consistent description of multiscale QM/MM coupling. When confronted with complex electrostatic coupling in QM/MM simulations, many contemporary ML/MM models routinely achieve single-point energy errors within 1–2 kcal/mol for solution-phase systems.^{56,108} Such fitting errors can be further amplified in heterogeneous enzymatic environments, limiting the ability of current ML/MM models to distinguish subtle energetic differences arising from local structural perturbations or enzyme mutations. In addition, as Jaakkola and co-workers have pointed out, practical validation of ML/MM models should extend beyond conventional energy

and force errors.¹¹⁸ Future benchmarking efforts should also assess the stability of long-time scale molecular dynamics simulations, the accurate reproduction of ensemble properties (e.g., radial distribution functions and free-energy profiles), and transferability across different enzyme variants. Collectively, these criteria provide a more comprehensive assessment of model reliability and predictive capability in realistic enzymatic environments.

5.3. Universality

Constructing universal reactive data sets for biomolecules is undoubtedly challenging. Even in vacuo, locating transition states and sufficiently exploring reactive pathways across the vast chemical reaction space is already a formidable task. This challenge becomes substantially greater in heterogeneous biomolecular environments.⁶⁸ The lack of interpretability hinders direct transferability of neural network potentials trained on small- to medium-sized systems to the application of complex reactive biological systems, such as protein post-translational modifications (PTMs) and full-length ribozymes in RNA molecules.^{30,43} Robust and transferable MLPs for realistic biological systems remain largely unexplored, except for a few emerging “top-down” approaches.^{73,108} Furthermore, charge and spin states lack a systematic, physics-based description in ML/MM frameworks, despite the crucial roles of metalloenzyme catalysis in life sciences.^{136,137}

In summary, ML/MM represents a promising evolution of conventional QM/MM methods through the integration of machine learning techniques. With continued advances in data set construction and multiscale electrostatic modeling, it is poised to become a routine tool for large-scale simulations in computational enzymology.

AUTHOR INFORMATION

Corresponding Authors

Daiqian Xie – *Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry, Nanjing University, Nanjing 210023, China; Hefei National Laboratory, Hefei 230088, China; orcid.org/0000-0001-7185-7085; Email: dqxie@nju.edu.cn*

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

Authors

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

Chenyu Wu – *Institute of Theoretical and Computational Chemistry, State Key Laboratory of Coordination Chemistry, Key Laboratory of Mesoscopic Chemistry, School of Chemistry, Nanjing University, Nanjing 210023, China*

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpclett.6c01601>

Author Contributions

#Authors Xinhu Sha and Chenyu Wu contributed equally to this Perspective. All authors reviewed and revised the manuscript.

Notes

The authors declare no competing financial interest.

Biographies



Xinhu Sha is a Ph.D. student at Nanjing University, China. His scientific interest lies in incorporating machine learning potentials with multiscale simulations in biomolecular systems.



Chenyu Wu is a Ph.D. student at Nanjing University, China. His research interests focus on the development of universal machine learning interatomic potentials and the generation scheme of novel reference datasets.



Daiqian Xie is a distinguished Professor at Nanjing University, China. He received his B.S. from Sichuan University in 1983 and Ph.D. in Physical Chemistry from Jilin University in 1988. After a postdoctoral appointment at Jilin University from 1988 to 1991, he became a

professor at Sichuan University from 1991 to 2001. He has been a visiting professor at the University of New Mexico, Duke University, National University of Singapore, and Nanyang Technological University. His research focuses on polyatomic potential energy surfaces and the quantum dynamics of molecular inelastic scattering, photodissociation, and chemical reactions.



Yanzi Zhou is an Associate Professor at Nanjing University, China. She received her Ph.D. from Sichuan University in 2007. From 2007 to 2011, she worked as a postdoctoral fellow at the Department of Chemistry in New York University. Her research focuses on the dynamic simulation of biomolecules, theoretical research on enzyme reaction mechanisms, development of QM/MM methods, construction of high-precision potential energy surfaces for van der Waals complexes, and prediction of the vibrational and rotational spectra.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 22073040 to Y.Z., 22233003 and 22241302 to D.X.) and the Innovation Program for Quantum Science and Technology (2021ZD0303305 to D.X.).

REFERENCES

- (1) Vaissier Welborn, V.; Head-Gordon, T. Computational design of synthetic enzymes. *Chem. Rev.* **2019**, *119* (11), 6613–6630.
- (2) Wolfenden, R.; Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **2001**, *34* (12), 938–945.
- (3) Mak, W. S.; Siegel, J. B. Computational enzyme design: Transitioning from catalytic proteins to enzymes. *Curr. Opin. Struct. Biol.* **2014**, *27*, 87–94.
- (4) Vaissier Welborn, V. Structural dynamics and computational design of synthetic enzymes. *Chem. Catal.* **2022**, *2* (1), 19–28.
- (5) Carrea, G.; Riva, S. Properties and synthetic applications of enzymes in organic solvents. *Angew. Chem., Int. Ed.* **2000**, *39* (13), 2226–2254.
- (6) Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440* (7086), 940–943.
- (7) Dürrenberger, M.; Ward, T. R. Recent achievements in the design and engineering of artificial metalloenzymes. *Curr. Opin. Chem. Biol.* **2014**, *19*, 99–106.
- (8) Gao, Y.; Yang, W. Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science* **2016**, *352* (6291), 1334–1337.
- (9) Xu, D.; Meisburger, S. P.; Ando, N. Correlated motions in structural biology. *Biochemistry* **2021**, *60* (30), 2331–2340.
- (10) Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **2016**, *49*, e6.
- (11) Stiller, J. B.; Jordan Kerns, S.; Hoemberger, M.; Cho, Y.-J.; Otten, R.; Hagan, M. F.; Kern, D. Probing the transition state in enzyme catalysis by high-pressure NMR dynamics. *Nat. Catal.* **2019**, *2* (8), 726–734.
- (12) Yoon, S.; Ollie, E.; York, D. M.; Piccirilli, J. A.; Harris, M. E. Rapid kinetics of pistol ribozyme: Insights into limits to RNA catalysis. *Biochemistry* **2023**, *62* (13), 2079–2092.
- (13) Markin, C. J.; Mokhtari, D. A.; Sundén, F.; Appel, M. J.; Akiva, E.; Longwell, S. A.; Sabatti, C.; Herschlag, D.; Fordyce, P. M. Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. *Science* **2021**, *373* (6553), No. eabf8761.
- (14) Ganguly, A.; Weissman, B. P.; Giese, T. J.; Li, N.-S.; Hoshika, S.; Rao, S.; Benner, S. A.; Piccirilli, J. A.; York, D. M. Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme. *Nat. Chem.* **2020**, *12* (2), 193–201.
- (15) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198–1229.
- (16) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem. Commun.* **2010**, *46* (14), 2354–2372.
- (17) Kohen, A. Role of dynamics in enzyme catalysis: Substantial versus semantic controversies. *Acc. Chem. Res.* **2015**, *48* (2), 466–473.
- (18) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450* (7171), 913–916.
- (19) Masgrau, L.; Truhlar, D. G. The importance of ensemble averaging in enzyme kinetics. *Acc. Chem. Res.* **2015**, *48* (2), 431–438.
- (20) Nagel, Z. D.; Klinman, J. P. A 21st century revisionist's view at a turning point in enzymology. *Nat. Chem. Biol.* **2009**, *5* (8), 543–550.
- (21) Hu, H.; Yang, W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59* (1), 573–601.
- (22) 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* (2), 227–249.
- (23) Acevedo, O.; Jorgensen, W. L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* **2010**, *43* (1), 142–151.
- (24) Brunk, E.; Rothlisberger, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* **2015**, *115* (12), 6217–6263.
- (25) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733.
- (26) Gao, J. Hybrid quantum and molecular mechanical simulations: An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* **1996**, *29* (6), 298–305.
- (27) Hu, P.; Wang, S.; Zhang, Y. How do set-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by ab initio QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.* **2008**, *130* (12), 3806–3813.
- (28) Hu, P.; Wang, S.; Zhang, Y. Highly dissociative and concerted mechanism for the nicotinamide cleavage reaction in Sir2Tm enzyme suggested by ab initio QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.* **2008**, *130* (49), 16721–16728.
- (29) Gao, J. Absolute free energy of solvation from Monte Carlo simulations using combined quantum and molecular mechanical potentials. *J. Phys. Chem.* **1992**, *96* (2), 537–540.
- (30) Gaines, C. S.; Giese, T. J.; York, D. M. Cleaning up mechanistic debris generated by twister ribozymes using computational RNA enzymology. *ACS Catal.* **2019**, *9* (7), 5803–5815.
- (31) McCarthy, E.; Ekesan, S.; Giese, T. J.; Wilson, T. J.; Deng, J.; Huang, L.; Lilley, D. M. J.; York, D. M. Catalytic mechanism and pH dependence of a methyltransferase ribozyme (MTR1) from computational enzymology. *Nucleic Acids Res.* **2023**, *51* (9), 4508–4518.
- (32) Li, P.; Liu, F.; Shao, Y.; Mei, Y. Computational insights into endo/exo selectivity of the Diels–Alder reaction in explicit solvent at ab initio quantum mechanical/molecular mechanical level. *J. Phys. Chem. B* **2019**, *123* (24), 5131–5138.

- (33) Shen, L.; Wu, J.; Yang, W. Multiscale quantum mechanics/molecular mechanics simulations with neural networks. *J. Chem. Theory Comput.* **2016**, *12* (10), 4934–4946.
- (34) Yan, W.; Lai, X.; Chen, Y.; Zhang, W.; Wu, J.; Xu, X. General framework for geometric deep learning on tensorial properties of molecules and crystals. *J. Am. Chem. Soc.* **2025**, *147* (51), 47044–47056.
- (35) Jurich, C.; Shao, Q.; Ran, X.; Yang, Z. J. Physics-based modeling in the new era of enzyme engineering. *Nat. Comput. Sci.* **2025**, *5* (4), 279–291.
- (36) Schlick, T.; Portillo-Ledesma, S. Biomolecular modeling thrives in the age of technology. *Nat. Comput. Sci.* **2021**, *1* (5), 321–331.
- (37) Dral, P. O. Quantum chemistry in the age of machine Learning. *J. Phys. Chem. Lett.* **2020**, *11* (6), 2336–2347.
- (38) Han, Y.; Wang, Z.; Wei, Z.; Liu, J.; Li, J. Machine learning builds full-QM precision protein force fields in seconds. *Brief. Bioinform.* **2021**, *22* (6), No. bbab158.
- (39) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **2023**, *5* (9), 1031–1041.
- (40) Liao, Y.-L.; Wood, B. M.; Das, A.; Smidt, T. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. Presented at the *Twelfth International Conference on Learning Representations (ICLR)*, 2024.
- (41) Yang, C.; Wu, C.; Xie, W.; Xie, D.; Hu, P. General reactive element-based machine learning potentials for heterogeneous catalysis. *Nat. Catal.* **2025**, *8* (9), 891–904.
- (42) Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072.
- (43) Bonollo, G.; Trèves, G.; Komarov, D.; Mansoor, S.; Moroni, E.; Colombo, G. Advancing molecular simulations: Merging physical models, experiments, and AI to tackle multiscale complexity. *J. Phys. Chem. Lett.* **2025**, *16* (15), 3606–3615.
- (44) Unke, O. T.; Stohr, M.; Gansch, S.; Unterthiner, T.; Maennel, H.; Kashubin, S.; Ahlin, D.; Gastegger, M.; Medrano Sandonas, L.; Berryman, J. T.; Tkatchenko, A.; Müller, K.-R.; et al. Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Sci. Adv.* **2024**, *10* (14), No. eadn4397.
- (45) Wang, T.; He, X.; Li, M.; Li, Y.; Bi, R.; Wang, Y.; Cheng, C.; Shen, X.; Meng, J.; Zhang, H.; et al. Ab initio characterization of protein molecular dynamics with AI²BMD. *Nature* **2024**, *635* (8040), 1019–1027.
- (46) Kabylda, A.; Frank, J. T.; Suárez-Dou, S.; Khabibrakhmanov, A.; Medrano Sandonas, L.; Unke, O. T.; Chmiela, S.; Müller, K.-R.; Tkatchenko, A. Molecular simulations with a pretrained neural network and universal pairwise force fields. *J. Am. Chem. Soc.* **2025**, *147* (37), 33723–33734.
- (47) Yang, Y.; Zhang, S.; Ranasinghe, K. D.; Isayev, O.; Roitberg, A. E. Machine learning of reactive potentials. *Annu. Rev. Phys. Chem.* **2024**, *75*, 371–395.
- (48) Grassano, J. S.; Pickering, I.; Roitberg, A. E.; Estrin, D. A.; Semelak, J. A. From QM/MM to ML/MM: A new era in multiscale modeling. *Chem. Phys. Rev.* **2025**, *6* (4), No. 041304.
- (49) Kulichenko, M.; Barros, K.; Lubbers, N.; Li, Y. W.; Messerly, R.; Tretiak, S.; Smith, J. S.; Nebgen, B. Uncertainty-driven dynamics for active learning of interatomic potentials. *Nat. Comput. Sci.* **2023**, *3* (3), 230–239.
- (50) Lei, Y.; Yagi, K.; Sugita, Y. Efficient training of neural network potentials for chemical and enzymatic reactions by continual learning. *J. Chem. Theory Comput.* **2025**, *21* (5), 2695–2711.
- (51) Arattu Thodika, A. R.; Pan, X.; Shao, Y.; Nam, K. Machine learning quantum mechanical/molecular mechanical potentials: Evaluating transferability in dihydrofolate reductase-catalyzed reactions. *J. Chem. Theory Comput.* **2025**, *21* (2), 817–832.
- (52) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100* (25), 10580–10594.
- (53) Bösel, L.; Thuerlemann, M.; Riniker, S. Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems. *J. Chem. Theory Comput.* **2021**, *17* (5), 2641–2658.
- (54) Eastman, P.; Pritchard, B. P.; Chodera, J. D.; Markland, T. E. Nutmeg and SPICE: Models and data for biomolecular machine learning. *J. Chem. Theory Comput.* **2024**, *20* (19), 8583–8593.
- (55) Morrow, J. D.; Gardner, J. L. A.; Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **2023**, *158* (12), No. 121501.
- (56) Zinoviev, K. Electrostatic embedding of machine learning potentials. *J. Chem. Theory Comput.* **2023**, *19* (6), 1888–1897.
- (57) Grassano, J. S.; Pickering, I.; Roitberg, A. E.; Gonzalez Lebrero, M. C.; Estrin, D. A.; Semelak, J. A. Assessment of embedding schemes in a hybrid machine learning/classical potentials (ML/MM) approach. *J. Chem. Inf. Model.* **2024**, *64* (10), 4047–4058.
- (58) Semelak, J. A.; Pickering, I.; Huddleston, K.; Olmos, J.; Grassano, J. S.; Clemente, C. M.; Drusin, S. I.; Marti, M.; Gonzalez Lebrero, M. C.; Roitberg, A. E.; et al. Advancing multiscale molecular modeling with machine learning-derived electrostatics. *J. Chem. Theory Comput.* **2025**, *21* (10), 5194–5207.
- (59) Gastegger, M.; Schuett, K. T.; Mueller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.* **2021**, *12* (34), 11473–11483.
- (60) Blum, L. C.; Raymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131* (25), 8732–8733.
- (61) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3* (5), No. e1603015.
- (62) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **2020**, *7* (1), 134.
- (63) Gradisteanu, V.; Chan, E. W.; Hedges, L.; Malagarriga, M.; David, R.; de la Puente, M.; Laage, D.; Tuñón, I.; van der Kamp, M. W.; Zinoviev, K. Simulating enzyme catalysis with electrostatically embedded machine learning potentials. *Chem. Sci.* **2026**, *17*, 8542–8556.
- (64) Zeng, J.; Tao, Y.; Giese, T. J.; York, D. M. QD π : A quantum deep potential interaction model for drug discovery. *J. Chem. Theory Comput.* **2023**, *19* (4), 1261–1275.
- (65) Thürlemann, M.; Pultar, F.; Gordiy, I.; Riniker, S. Biomolecular multiscale simulation (BMS25) dataset to train neural network potentials for QM/MM settings with electrostatic embedding. *ChemRxiv*, December 15, 2025. [10.26434/chemrxiv-2025-xzp6k](https://doi.org/10.26434/chemrxiv-2025-xzp6k) (accessed June 4, 2026).
- (66) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y.; De Fabritiis, G.; Markland, T. E.; et al. SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Sci. Data* **2023**, *10* (1), 11.
- (67) Wu, C.; Yang, C.; Fang, Z.; Xie, W.; Xie, D.; Hu, P. Hitchhikers guide to training more general machine learning potentials in heterogeneous catalysis. *J. Chem. Theory Comput.* **2026**, *22* (4), 1970–1981.
- (68) Ren, K.; Tang, K.; Zhao, Y.; Zhang, L.; Du, J.; Meng, Q.; Liu, Q. Generic machine learning potential for accelerating transition state search in organic synthesis. *ChemRxiv*, September 22, 2025, DOI: [10.26434/chemrxiv-2025-mzz6w-v3](https://doi.org/10.26434/chemrxiv-2025-mzz6w-v3) (accessed June 4, 2026).
- (69) Brezina, K.; Beck, H.; Marsalek, O. Reducing the cost of neural network potential generation for reactive molecular systems. *J. Chem. Theory Comput.* **2023**, *19* (19), 6589–6604.
- (70) Gansch, S.; Unke, O. T.; Ahlin, D.; Maennel, H.; Kashubin, S.; Müller, K.-R. The QCML dataset, Quantum chemistry reference data from 33.5M DFT and 14.7B semi-empirical calculations. *Sci. Data* **2025**, *12* (1), 406.
- (71) Song, G.; Yang, W. NepoIP/MM: Toward accurate biomolecular simulation with a machine learning/molecular mechanics model

incorporating polarization effects. *J. Chem. Theory Comput.* **2025**, *21* (11), 5588–5598.

(72) Zhang, S.; Makoś, M. Z.; Jadrlich, R. B.; Kraka, E.; Barros, K.; Nebgen, B. T.; Tretiak, S.; Isayev, O.; Lubbers, N.; Messerly, R. A.; et al. Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential. *Nat. Chem.* **2024**, *16* (5), 727–734.

(73) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13* (1), 2453.

(74) Goyal, A.; Bengio, Y. Inductive biases for deep learning of higher-level cognition. *Proc. R. Soc. A* **2022**, *478* (2266), No. 20210068.

(75) Zhang, Y.; Jiang, B. Universal machine learning for the response of atomistic systems to external fields. *Nat. Commun.* **2023**, *14*, 6424.

(76) Jin, Z.; Braams, B. J.; Bowman, J. M. An ab initio based global potential energy surface describing $\text{CH}_5^+ \rightarrow \text{CH}_3^+ + \text{H}_2$. *J. Phys. Chem. A* **2006**, *110* (4), 1569–1574.

(77) Yu, Q.; Ma, R.; Qu, C.; Conte, R.; Nandi, A.; Pandey, P.; Houston, P. L.; Zhang, D. H.; Bowman, J. M. Extending atomic decomposition and many-body representation with a chemistry-motivated approach to machine learning potentials. *Nat. Comput. Sci.* **2025**, *5* (5), 418–426.

(78) Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional networks on graphs for learning molecular fingerprints. In *NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems*, Vol. 2, 2015; pp 2224–2232.

(79) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

(80) Batatia, I.; Kovács, D. P.; Simm, G. N. C.; Ortner, C.; Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, 2022.

(81) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14* (1), 579.

(82) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12* (1), 398.

(83) Lier, B.; Poliak, P.; Marquetand, P.; Westermayr, J.; Oostenbrink, C. BuRNN: Buffer region neural network approach for polarizable-embedding neural network/molecular mechanics simulations. *J. Phys. Chem. Lett.* **2022**, *13* (17), 3812–3818.

(84) Shen, L.; Yang, W. Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks. *J. Chem. Theory Comput.* **2018**, *14* (3), 1442–1455.

(85) Zhou, B.; Zhou, Y.; Xie, D. Accelerated quantum mechanics/molecular mechanics simulations via neural networks incorporated with mechanical embedding scheme. *J. Chem. Theory Comput.* **2023**, *19*, 1157–1169.

(86) Sha, X.; Chen, Z.; Xie, D.; Zhou, Y. Modeling enzyme reaction and mutation by direct machine learning/molecular mechanics simulations. *J. Chem. Theory Comput.* **2025**, *21* (9), 4335–4346.

(87) Zinovjev, K.; Hedges, L.; Montagud Andreu, R.; Woods, C.; Tunon, I.; van der Kamp, M. W. emle-engine: A flexible electrostatic machine learning embedding package for multiscale molecular dynamics simulations. *J. Chem. Theory Comput.* **2024**, *20* (11), 4514–4522.

(88) Szczepaniak, F.; Suh, D.; Im, W. CHARMM-GUI hybrid ML/MM builder for hybrid machine learning and molecular mechanical modeling and simulations. *J. Chem. Inf. Model.* **2026**, *66* (6), 2960–2966.

(89) Wang, X.; Wu, X.; Brooks, B. R.; Wang, J. Accurate free energy calculation via multiscale simulations driven by hybrid machine learning and molecular mechanics potentials. *J. Chem. Theory Comput.* **2025**, *21* (14), 6979–6987.

(90) Kim, D.; Wang, X.; Vargas, S.; Zhong, P.; King, D. S.; Inizan, T. J.; Cheng, B. A universal augmentation framework for long-range electrostatics in machine learning interatomic potentials. *J. Chem. Theory Comput.* **2025**, *21* (24), 12709–12724.

(91) Zhang, Y.; Hu, C.; Jiang, B. Embedded atom neural network potentials: Efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.* **2019**, *10* (17), 4962–4967.

(92) Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B* **2015**, *92* (4), No. 045131.

(93) Thürlmann, M.; Bösel, L.; Riniker, S. Learning atomic multipoles: Prediction of the electrostatic potential with equivariant graph neural networks. *J. Chem. Theory Comput.* **2022**, *18* (3), 1701–1710.

(94) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426.

(95) Zeng, J.; Giese, T. J.; Ekesan, S.; York, D. M. Development of range-corrected deep learning potentials for fast, accurate quantum mechanical/molecular mechanical simulations of chemical reactions in solution. *J. Chem. Theory Comput.* **2021**, *17* (11), 6993–7009.

(96) Xie, Z.; Li, Y.; Xia, Y.; Zhang, J.; Yuan, S.; Fan, C.; Yang, Y. I.; Gao, Y. Q. Multiscale force field model based on a graph neural network for complex chemical systems. *J. Chem. Theory Comput.* **2025**, *21* (5), 2501–2514.

(97) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59* (3), 341–350.

(98) 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* (10), 4175–4183.

(99) 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* (37), 20595–20605.

(100) Kim, B.; Snyder, R.; Nagaraju, M.; Zhou, Y.; Ojeda-May, P.; Keeton, S.; Hege, M.; Shao, Y.; Pu, J. Reaction path-force matching in collective variables: Determining ab initio QM/MM free energy profiles by fitting mean force. *J. Chem. Theory Comput.* **2021**, *17* (8), 4961–4980.

(101) Pan, X.; Yang, J.; Van, R.; Epifanovsky, E.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-learning-assisted free energy simulation of solution-phase and enzyme reactions. *J. Chem. Theory Comput.* **2021**, *17* (9), 5745–5758.

(102) Gomez-Flores, C. L.; Maag, D.; Kansari, M.; Vuong, V. Q.; Irle, S.; Grater, F.; Kubar, T.; Elstner, M. Accurate free energies for complex condensed-phase reactions using an artificial neural network corrected DFTB/MM methodology. *J. Chem. Theory Comput.* **2022**, *18* (2), 1213–1226.

(103) Hofstetter, A.; Bösel, L.; Riniker, S. Graph-convolutional neural networks for (QM)ML/MM molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2022**, *24* (37), 22497–22512.

(104) Giese, T. J.; Zeng, J.; York, D. M. Transferability of MACE graph neural network for range corrected Δ -machine learning potential QM/MM applications. *J. Phys. Chem. B* **2025**, *129* (22), 5477–5490.

(105) Crha, R.; Poliak, P.; Gillhofer, M.; Oostenbrink, C. Alchemical free-energy calculations at quantum-chemical precision. *J. Phys. Chem. Lett.* **2025**, *16* (4), 863–869.

(106) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **2019**, *31* (46), No. 1902765.

(107) Wang, X.; Wang, J.; Li, W.-L. Machine learning/molecular mechanics enzymology for the next generation of computational enzymatic catalysis. *Chem. Catal.* **2026**, *6* (3), No. 101658.

(108) Pultar, F.; Thürlmann, M.; Gordiy, I.; Doloszeski, E.; Riniker, S. Neural network potential with multiresolution approach enables accurate prediction of reaction free energies in solution. *J. Am. Chem. Soc.* **2025**, *147* (8), 6835–6856.

- (109) Lei, Y.; Yagi, K.; Sugita, Y. Learning QM/MM potential using equivariant multiscale model. *J. Chem. Phys.* **2024**, *160* (21), No. 214109.
- (110) Haghighi, S.; Viquez Rojas, C.; Bhat, S.; Isayev, O.; Slipchenko, L. ANI/EFP: Modeling long-range interactions in ANI neural network with effective fragment potentials. *J. Chem. Theory Comput.* **2024**, *20* (20), 9138–9147.
- (111) Haghighi, S.; Urbina, A. S.; Slipchenko, L. V. Beyond the cutoff: Hybrid ML/MM electrostatics for neural network potentials. *J. Chem. Phys.* **2026**, *164* (10), No. 104102.
- (112) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, 2019.
- (113) 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.
- (114) Thürlmann, M.; Riniker, S. Anisotropic message passing: Graph neural networks with directional and long-range interactions. Presented at the *Eleventh International Conference on Learning Representations (ICLR)*, 2023.
- (115) Barrett, R.; Dietschreit, J. C. B.; Westermayr, J. Incorporating long-range interactions via the multipole expansion into ground and excited-state molecular simulations. *npj Comput. Mater.* **2026**, *12* (1), 135.
- (116) Gao, J.; Xia, X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258* (5082), 631–635.
- (117) Sha, X.; Guo, X.; Chen, Z.; Xie, D.; Zhou, Y. Efficient deep learning framework for multiscale simulations in enzymes via linear response approximation. *ChemRxiv*, February 2, 2026, DOI: 10.26434/chemrxiv.10001828/v1 (accessed June 4, 2026).
- (118) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Ketten, S.; Gomez-Bombarelli, R.; Jaakkola, T. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv*, October 13, 2022, DOI: 10.48550/arXiv.2210.07237 (accessed June 4, 2026).
- (119) Karwounopoulos, J.; Bieniek, M.; Wu, Z.; Baskerville, A. L.; König, G.; Cossins, B. P.; Wood, G. P. F. Evaluation of machine learning/molecular mechanics end-state corrections with mechanical embedding to calculate relative protein–ligand binding free energies. *J. Chem. Theory Comput.* **2025**, *21* (2), 967–977.
- (120) Yan, Z.; Wei, D.; Li, X.; Chung, L. W. Accelerating reliable multiscale quantum refinement of protein–drug systems enabled by machine learning. *Nat. Commun.* **2024**, *15* (1), 4181.
- (121) Giese, T. J.; Zeng, J.; Lerew, L.; McCarthy, E.; Tao, Y.; Ekesan, S.; York, D. M. Software infrastructure for next-generation QM/MM- Δ MLP force fields. *J. Phys. Chem. B* **2024**, *128* (26), 6257–6271.
- (122) Xu, M.; Zhu, T.; Zhang, J. Z. H. Automatically constructed neural network potentials for molecular dynamics simulation of zinc proteins. *Front. Chem.* **2021**, *9*, No. 692200.
- (123) Ding, Y.; Huang, J. DP/MM: A hybrid model for zinc–protein interactions in molecular dynamics. *J. Phys. Chem. Lett.* **2024**, *15* (2), 616–627.
- (124) Bensberg, M.; Eckhoff, M.; Thomasen, F. E.; Bro-Jørgensen, W.; Teynor, M. S.; Sora, V.; Weymuth, T.; Husstein, R. T.; Knudsen, F. E.; Krogh, A.; et al. Machine learning-enhanced calculation of quantum-classical binding free energies. *J. Chem. Theory Comput.* **2025**, *21* (16), 8182–8198.
- (125) Sabanés Zariquiey, F.; Farr, S. E.; Doerr, S.; De Fabritiis, G. QuantumBind-RBFE: Accurate relative binding free energy calculations using neural network potentials. *J. Chem. Inf. Model.* **2025**, *65* (8), 4081–4089.
- (126) Lahey, S.-L. J.; Rowley, C. N. Simulating protein–ligand binding with neural network potentials. *Chem. Sci.* **2020**, *11* (9), 2362–2368.
- (127) Zubatyuk, R.; Biczysko, M.; Ranasinghe, K.; Moriarty, N. W.; Gokcan, H.; Kruse, H.; Poon, B. K.; Adams, P. D.; Waller, M. P.; Roitberg, A. E.; Isayev, O.; Afonine, P. V.; et al. AQuaRef: Machine learning accelerated quantum refinement of protein structures. *Nat. Commun.* **2025**, *16* (1), 9224.
- (128) Vant, J. W.; Lahey, S.-L. J.; Jana, K.; Shekhar, M.; Sarkar, D.; Munk, B. H.; Kleinekathöfer, U.; Mittal, S.; Rowley, C.; Singharoy, A. Flexible fitting of small molecules into electron microscopy maps using molecular dynamics simulations with neural network potentials. *J. Chem. Inf. Model.* **2020**, *60* (5), 2591–2604.
- (129) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500.
- (130) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic energy transfer in condensed phase studied by a polarizable QM/MM model. *J. Chem. Theory Comput.* **2009**, *5* (7), 1838–1848.
- (131) Zinovjev, K.; Curutchet, C. Improved description of environment and vibronic effects with electrostatically embedded ML potentials. *J. Phys. Chem. Lett.* **2025**, *16* (3), 774–781.
- (132) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93* (2), 1061–1071.
- (133) Tiefenbacher, M. X.; Bachmair, B.; Chen, C. G.; Westermayr, J.; Marquetand, P.; Dietschreit, J. C. B.; González, L. Excited-state nonadiabatic dynamics in explicit solvent using machine learned interatomic potentials. *Digital Discovery* **2025**, *4* (6), 1478–1491.
- (134) Qu, E.; Krishnapriyan, A. S. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. In *Advances in Neural Information Processing Systems 37 (NeurIPS)*, 2024.
- (135) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (136) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM simulations: What are some of the “burning issues”? *J. Phys. Chem. B* **2021**, *125* (3), 689–702.
- (137) Tzeliou, C. E.; Mermigki, M. A.; Tzeli, D. Review on the QM/MM methodologies and their application to metalloproteins. *Molecules* **2022**, *27* (9), 2660.