ANI Neural Networks Meet Electrostatics: A ML/MM Implementation in Amber

Jonathan A.Semelak,[†,‡,#] Ignacio Pickering,[¶,#] Kate Huddleston,[¶] Justo Olmos,[§] Juan Santiago Grassano,[†,‡] Camila Clemente,[§] Salvador I. Drusin,[‖,⊥] Marcelo Martí,[§] Mariano Camilo Gonzalez Lebrero,[†,‡] Adrian E. Roitberg,[*,¶] and Dario A. Estrin[*,†,‡]

†*Facultad de Ciencias Exactas y Naturales, Departamento de Química Inorgánica, Analítica y Química Física, Universidad de Buenos Aires, Intendente Güiraldes 2160, C1428EHA Buenos Aires, Argentina*

‡*CONICET - Universidad de Buenos Aires, Instituto de Química-Física de los Materiales, Medio Ambiente y Energía (INQUIMAE), Ciudad Universitaria, Pabellón 2, C1428EHA Buenos Aires, Argentina*

¶*Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States*

§*Facultad de Ciencias Exactas y Naturales, Departamento de Química Biológica e IQUIBICEN-CONICET, Universidad de Buenos Aires, Intendente Güiraldes 2160, C1428EHA Buenos Aires, Argentina.*

‖*Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina.*

⊥*Instituto de Química Rosario (IQUIR-CONICET), Rosario, Argentina*

#*These authors contributed equally to this work.*

E-mail: roitberg@ufl.edu; dario@qi.fcen.uba.ar

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Abstract

We present a novel integration of the ANI neural networks into the Amber software suite, offering a sophisticated machine learning/molecular mechanics (ML/MM) framework. The implementation is designed as a general-purpose tool for the simulation of neutral organic molecules, requiring no additional training for its use beyond the initial setup. The framework leverages a new ANI potential that accurately predicts geometry-dependent atomic partial charges at the Minimal Basis Iterative Stockholder (MBIS) level, enhancing the modeling of electrostatic interactions within ML/MM systems. Additionally, we incorporate a polarization correction to address the distortion effects on the ML subsystem from MM point charges. Our approach is validated through simulations of solvation profiles, vibrational spectra, and torsion free energy profiles of small molecules in aqueous environments, as well as protein-ligand interactions. Our findings demonstrate that this ML/MM framework can approximate QM/MM electrostatic embedding with significantly reduced computational demands, paving the way for more efficient and accurate simulations in computational chemistry.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

1 Introduction

The 2013 Nobel Prize in Chemistry was awarded to Karplus, Levitt, and Warshel for their pioneering development of a method to model interactions between a quantum mechanics (QM) subsystem and a molecular mechanics (MM) subsystem, giving rise to the innovative hybrid QM/MM theory.[1–5] At the heart of QM/MM coupling is electrostatics, which could be the critical component in the development of new machine learning[6–12] (ML) molecular mechanics frameworks (ML/MM).

Hybrid ML/MM models have emerged as a promising solution to integrate the high accuracy of ML in predicting energy and forces[13–22] of isolated systems at the QM level with the computational efficiency of MM force fields adept at describing larger systems. Despite being faster than QM methods, ML models are still slower than classical force fields and often employ local descriptors[23,24] with a cutoff radius, which neglects long-range interactions and challenges their scalability. This limitation has catalyzed extensive research into ML/MM frameworks, which can be divided into three main categories. The first[25–27] employs a mechanical embedding strategy similar to early QM/MM methods,[28] in which the coupling energy is computed assigning a set of fixed partial charges to the ML atoms (or any other way to approximate the electrostatic potential) and calculating the coulombic interaction between those charges and the MM subsystem. The second category[29–34] address the influence of external potentials in the ML predictions by incorporating information of the MM subsystem as an extra input of the ML model. These methods have proven to be highly reliable in comparisons with QM/MM methods, although they require training on data specifically derived from QM/MM level simulations, which restricts their broader application. Finally, the third category of ML/MM methods[35–37] proposes an intermediate approach, using the ML model to predict the *in vacuo* energy of a system while adjusting for environmental effects through *a posteriori* corrections to both ML/MM and pure ML energy (and forces) calculations. These methods seek to achieve the sophistication of electrostatic embedding,[28] representing the most widely used QM/MM implementations, while

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

maintaining transferability.

In a previous work[37] we examined various approaches for integrating ML models with classical molecular mechanics, specifically focusing on the use of *in vacuo* predicted atomic charges[38–42] within ML/MM frameworks. Our findings demonstrated that employing minimal basis iterative stockholder[43] (MBIS) atomic charges derived *in vacuo*, coupled with a simple polarization correction using fixed atomic polarizabilities, achieves excellent agreement with the reference coupling energies observed in QM/MM simulations utilizing electrostatic embedding.

The straightforward method to calculate MBIS charges *in vacuo* necessitates a fully converged ground-state electronic density associated with the geometry of interest. Using an *ab initio* method such as density functional theory (DFT) to obtain this electronic density would, however, defeat the purpose of our proposed ML/MM scheme, since this step would constitute the bottleneck of the method. In view of this, we decided to leverage the known capability of the ANI[13–16] architecture to accurately reproduce atomic charges,[40] and trained a novel neural network model, which we will refer to as ANI-MBIS-q, that directly predicts MBIS charges from the local chemical environment (coordinates and atomic numbers) of each atom, bypassing the need for the electronic density.[44]

The ANI-MBIS-q model was trained on precalculated MBIS charges from structures in the ANI-2x dataset,[16,45] which correspond to electronic densities converged at the ωB97X/6-31G* level of theory[46] (we note however that MBIS charges have been shown to be largely independent of the basis set used).[43] Furthermore, since ANI-MBIS-q is implemented in PyTorch,[47] it can leverage its *autograd* mechanism to produce also the gradients of the atomic charges with respect to the coordinates, which is fundamental for enabling the implementation of our ML/MM scheme.

In this work, we present and showcase the implementation of a ML/MM scheme that takes advantage of the capability of the ANI-MBIS-q model to predict accurate atomic charges and charge-gradients, together with that of the ANI-2x model of producing fast and accurate

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

energies and forces at the ωB97X/6-31G* level of theory, by combining both together into a single model, ANI-MBIS, and embedding it in the context of a previously thoroughly validated ML/MM coupling approach. In addition, we introduce an interface that allows us to use ANI-style neural networks implemented in the TorchANI[15] framework, as calculators for the QM region in QM/MM simulations performed in SANDER, the molecular dynamics (MD) engine provided in AmberTools,[48] one of the leading software-suites for biomolecular simulations. This implementation, which is planned for inclusion in the next AmberTools release, is efficient, in that it avoids the cost that would incur from the naive approach of calling into Python code for every model evaluation. It also requires no familiarity with the internals of PyTorch, or any other machine learning framework from prospective users, and can be seamless called through the SANDER engine. We finally note that, the ANI models being general-purpose models that perform well for a wide range of systems, our scheme can be used without the need to train models from scratch.

This article is organized as follows: We first discuss the theory and implementation details of our ML/MM scheme. Next, we evaluate the performance and accuracy of our implementation for obtaining accurate MD simulations by the calculation of solvation profiles, vibrational spectra, and torsional free-energy profiles of small molecules in aqueous solution. Finally, we assess our ML/MM approach's capability to generate a physically accurate description of protein-ligand interactions.

2 Theory

Similar to the additive QM/MM scheme,[1,2,28] the total energy in our ML/MM framework, where the ML method predicts properties *in vacuo*, is expressed as:

$$E = E^{\mathrm{vac}}_{\mathrm{ML}}(\mathbf{R}^{\mathrm{ML}}) + E^{\mathrm{D}}_{\mathrm{ML}}(\mathbf{R}^{\mathrm{ML}}, \mathbf{R}^{\mathrm{MM}}) + E_{\mathrm{MM}}(\mathbf{R}^{\mathrm{MM}}) + E_{\mathrm{ML/MM}}(\mathbf{R}^{\mathrm{ML}}, \mathbf{R}^{\mathrm{MM}}) \tag{1}$$

In this expression, $\mathbf{R}^{\mathrm{ML}}$ and $\mathbf{R}^{\mathrm{MM}}$ represent the nuclear coordinates of atoms in the ML

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

and MM region. $E^{\mathrm{vac}}_{\mathrm{ML}}$ is the *in vacuo* energy for the ML region, $E^{\mathrm{D}}_{\mathrm{ML}}$ represents the distortion energy that accounts for the cost of polarizing the ML region, E_{MM} is the energy of the MM region, and $E_{\mathrm{ML/MM}}$ denotes the interaction energy between the ML and MM subsystems. This interaction energy includes a non-electrostatic contribution utilizing predetermined Lennard-Jones parameters for ML atoms, alongside an electrostatic component, $E^{\mathrm{elec}}_{\mathrm{ML/MM}}$, calculated as follows:

$$E^{\mathrm{elec}}_{\mathrm{ML/MM}} = \sum_{i\in\mathrm{MM}} \sum_{j\in\mathrm{ML}} \frac{q_i q_j^{\mathrm{vac}}(\mathbf{R}^{\mathrm{ML}})}{|\mathbf{R}_i^{\mathrm{MM}} - \mathbf{R}_j^{\mathrm{ML}}|} + E^{\mathrm{P}}_{\mathrm{ML/MM}}(\mathbf{R}^{\mathrm{ML}}, \mathbf{R}^{\mathrm{MM}}) \tag{2}$$

Where q_i and q_j^{vac} are the atomic partial charges of the i^{th} MM atom and the j^{th} ML atom, respectively, while $\mathbf{R}_i^{\mathrm{MM}}$ and $\mathbf{R}_j^{\mathrm{ML}}$ are their coordinates. The dependency of q_j^{vac} on the coordinates of the ML atoms is indicated explicitly, as well as their *in vacuo* nature. The MM q_i charges, on the other hand, are fixed parameters. $E^{\mathrm{P}}_{\mathrm{ML/MM}}$ is the polarization energy, adjusted for the influence of surrounding dipoles using an effective dielectric constant ε and the polarizability of isolated atoms α_j^{free}:

$$E^{\mathrm{P}}_{\mathrm{ML/MM}}(\mathbf{R}^{\mathrm{ML}}, \mathbf{R}^{\mathrm{MM}}) = -\frac{1}{\varepsilon} \sum_{j\in\mathrm{ML}} \alpha_j^{\mathrm{free}} |\mathbf{E}_{\mathrm{MM}}(\mathbf{R}_j^{\mathrm{ML}}, \mathbf{R}^{\mathrm{MM}})|^2 \tag{3}$$

Here, we assume that each ML atom functions as an independent isotropic polarizable center, with its associated induced dipoles being proportional to the electric field evaluated on its coordinates, $\mathbf{E}_{\mathrm{MM}}$, which is generated exclusively by the MM point charges.[1,37] The distortion energy, $E^{\mathrm{D}}_{\mathrm{ML}}$, is then calculated as half the negative value of the polarization energy.[49–52]

The only parameters that need to be assigned beforehand to the ML atoms are the atomic polarizabilities (and Lennard-Jones ones, but this is a common practice in QM/MM simulations, and are usually taken from MM force fields). Based on a previous study,[37] the *in vacuo* atomic charges for the ML region are MBIS charges, the atomic polarizabilities

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

are taken from the recommended[53] values of static scalar dipole polarizabilities for neutral atoms and ε is set to 2.

3 Methodology

3.1 ANI Neural Networks

The ANI family of atomistic neural network potentials are based off the Behler-Parrinello[23] architecture, with modified element-specific descriptors named Atomic Environment Vectors (AEVs). Three previously published and validated models in this family (each consisting of an 8 network ensemble) are available out of the box in the TorchANI library. In this article we focus on the ANI-2x model, which can predict energies and forces for neutral molecules with elements in the set H, C, N, O, F, S, Cl.

Directly interleaving calls to the TorchANI[15] library with SANDER's Fortran code would be inefficient and prohibitively expensive. This is due to each call to the ML calculators needing to either load the models from scratch, or communicate with a concurrent Python process. Additionally, the Python interpreter itself introduces overhead in between each PyTorch operator call, which is an extra cost on top of the usual PyTorch dispatch mechanism. To get around this issues we decided to implement the interface as a dynamic shared object that consists of a C wrapper over the underlying LibTorch intermediate representation executor.

Previous work[40] has shown that the ANI architecture is can be extended to the prediction the of atomic properties. Here, we use the ANI-MBIS-q model for the prediction of atomic charges of the ML subsystem on-the-fly. This model was trained on MBIS atomic charges corresponding to ωB97X/6-31G* electron densities, using the loss function

$$\mathcal{J}(\theta) = \frac{1}{M} \sum_{m \in \mathcal{D}} \frac{1}{N_m} \sum_i \left(q_i(\mathbf{r}_i, Z_i, \ldots, \theta) - q_i^{\text{ref}} \right)^2 ,$$

where M is the number of conformations in the data set $\mathcal{D}$, N_m the number of atoms in

https://doi.org/10.26434/chemrxiv-2024-m9xgc ORCID: https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

each conformation, θ is the full vector of parameters of the model, $\mathbf{r}_i, Z_i, \ldots$ represents the dependence of the predicted atomic charge on the coordinates and atomic number of atom i and *all atoms surrounding it* within a given cutoff distance (5.2 Å for the model considered in this article), and q_i^{ref} is the corresponding reference atomic charge. This loss function is minimized using the AdamW[54,55] optimizer, with default parameters, and a per-layer weight-decay analogous to the one used in the training procedure for the ANI-2x model.[16]

The network architecture and descriptor parameters are the same as those used in ANI-2x,.[16,40] The output value output from the final layer of the network, which we will call q_i^*, is not used directly in the loss function; it is instead normalized by multiplying it with a factor $f_i(q_m^*)$ that depends on all values output by the atomic networks for a given molecule, and guarantees that the sum of the atomic charges is equal to the charge of the QM region (neutral in all the results we present in this article). We found a normalizing factor given by

$$f_i = \frac{(w_i q_i^*)^2}{\sum_i (w_i q_i^*)^2} \,,$$

where $w_i = \chi_Z/\eta_Z$, and χ, η are the atomic number dependent electronegativity and hardness of atom i, taken from reference data,[56] ensured good stability when training.

To further improve the efficiency and modularity of our the implementation, we took advantage of the fact that both models that run in parallel, namely ANI-2x and ANI-MBIS-q, use AEVs with the same parameters, and we calculate them a single time and reuse them. The calculation of the AEVs is the bottleneck both in terms of memory usage and computational cost, so this optimization helps the model perform with a similar cost as a single ensemble.

3.2 Implementation of the ML/MM interface in Amber

The ML/MM framework is integrated into the extensible QM/MM SANDER interface,[57] which allows the user to invoke an ML/MM simulation with ANI serving as the ML calcu-

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

lator. If the ANI-MBIS model is selected, the ML/MM coupling is performed within the embedding theory discussed above. The other ANI models (ANI-1x, ANI-2x, ANI-1cxx) only work at the mechanical embedding level, in which the atomic charges for the ML region are read from the topology file, and nor distortion nor polarization corrections are included. We will focus here in the use of the ANI-MBIS model.

Figure 1 delineates the implementation pipeline and its application to perform MD simulations. The parameters for the simulation correspond to those of the topology file plus the polarizabilities of each atom from the ML region. SANDER constructs the neighbor list for the ML region based on the ML/MM cutoff specified by the user. The interface receives the element and coordinates of all the ML atoms, as well as the charges and coordinates of all MM atoms within the cutoff radius. The selected ANI model is initialized at the start of the simulation.

At each simulation step, the ANI-MBIS model computes the energy and MBIS atomic charges, predicting these at the ωB97X/6-31G* level of theory. The corresponding forces and gradients are obtained by back-propagation. The ML/MM electrostatic coupling term and its associated forces are computed by the interface, ensuring that the gradients of the predicted MBIS charges are considered to maintain energy conservation. The interface gathers this data and communicates with SANDER's main code for the propagation of the nuclear coordinates.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

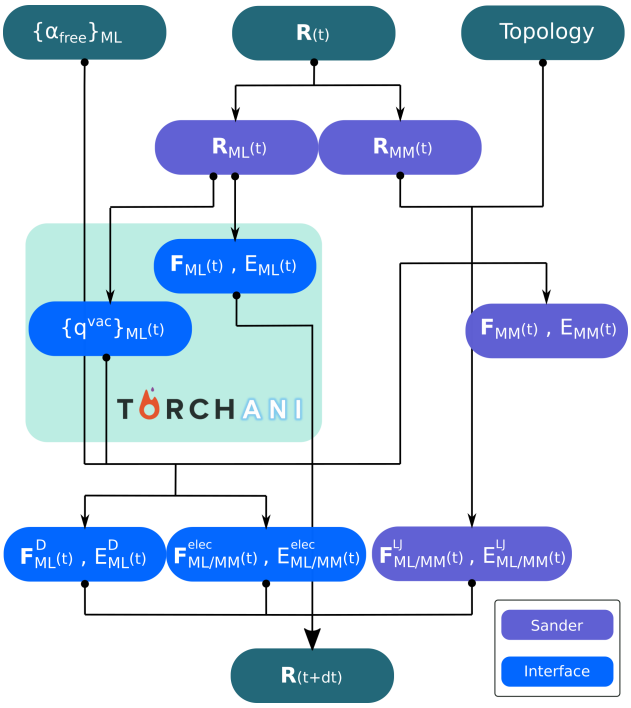

Figure 1: Implementation Pipeline of the ML/MM Framework in Amber's SANDER Interface. This diagram illustrates the workflow from input parameters through the molecular dynamics simulation process, highlighting the integration of the ANI-MBIS model. Key steps include the initialization of the ANI model, computation of energies and MBIS charges, and the subsequent calculation of forces.

3.3 Simulation Details

3.3.1 Case studies

The selection of problems was made based on popular applications of QM/MM simulations, covering for example studies in aqueous solution,[58,59] spectroscopy[60,61] and drug discovery.[62,63] Specifically, we tested our ML/MM model in the following systems in aqueous solution: (i) The solvation structure and vibrational spectra of phenol; (ii) torsion free energy profile of N-methylacetamide; (iii) The 2-dimensional free energy surface of the central Ramachandran angles of the capped alanine tetrapeptide; and (iv) two protein-ligand complexes: enzyme FKBP12 bound to an FK506 analogue[64] and avidin[65] bound to a biotin analogue; the existence of both complexes is supported by experimental data,[64,65] and we chose these analogues instead of the natural ligands for being neutral compounds.

The ML subsystem was integrated by the solute in cases (i) to (iii), and by ligands in

https://doi.org/10.26434/chemrxiv-2024-m9xgc ORCID: https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

the case of the protein-ligand complexes.

3.3.2 General Procedures

ML/MM simulations were performed employing different embedding schemes: the polarizable mechanical embedding using the ANI-MBIS model (equation (2)), a mechanical embedding scheme using the ANI-MBIS model without polarization and distortion correction (but MBIS charges predicted on-the-fly), and the actual mechanical embedding with ANI-2x, using fixed charges for the ML region, obtained from the corresponding force field (as specified below). QM/MM reference simulations were performed using the ωB97X/6-31G* level of theory to match the level of theory that the ANI models were trained on.

Each simulation was performed in a truncated octahedral box with about 3000 TIP3P[66] water molecules for non-protein systems, and about 13000 for protein-ligand systems. The cutoff for the MM interactions was 8.0 Å, in every case. The cutoff for ML/MM and QM/MM interactions was 18.0 Å, for the simulations of phenol and 15 Å, for all the other systems. We employed GAFF[67] parameters when modeling phenol, N-methylacetamide and both ligands at the MM level, using RESP[68] charges from HF/6-31G* electron densities. The alanine tetrapeptide and the two proteins where modeled using the ff19SB[69] force field. Lennard-Jones parameters used were the ones corresponding to each MM force field. In the case of the avidin complex the starting structure corresponded to the Protein Data Bank Identifier 1AVD,[70] for the FKBP complex it was 1FKH.[64] The initial ligand structures were achieved in both cases by simply removing atoms from the original ligands, this approach is validated by the high structural similarities between the ligand-analogue pairs (both analogues being smaller than the ligands) and has been used before for these same complexes.[71] Missing hydrogen atoms were added in Modeller,[72] ensuring the ligand remained neutral. All simulations were performed with our ML/MM and ANI codes integrated in AmberTools24, and the QM/MM simulations employed the existing interface with Gaussian09.[73] The system preparations were made with LEaP,[74] antechamber, and parmchk2[75] tools from AmberTools24.

https://doi.org/10.26434/chemrxiv-2024-m9xgc ORCID: https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

Dynamics visualizations and molecular drawings were performed with VMD,[76] version 1.9.3.

3.3.3 Systems preparation

For the non-protein systems, a step-wise protocols was employed to relax the system, consisting on a short minimization followed a 20 ps long thermalisation to 300 K in the NVT ensemble (constant amount of substance, volume and temperature), and a 50 ps long MD simulation in the NPT ensemble (same as NVT but constant pressure instead of volume, set to 1 bar) to reach a stationary density value. In the case of the protein complexes, the heating was performed in 2 ns, and the NPT MD simulation was 4 ns long. These relaxation simulations were performed using the Berendsen thermostat and barostat.[77] All simulations were run using the SHAKE algorithm[78] for keeping the bonds involving hydrogen still, with the exception of the full MM phenol simulations. In all the cases, the heating and equilibration protocols were performed treating the whole system at the MM level. Production runs were performed in the NVT ensemble using the Langevin thermostat,[79] with a collision frequency of 5 ps^{-1}. The ML/MM production runs were performed starting from the last frame of the equilibration protocol. In the case of phenol, the time step was 0.5 fs and the system was also simulated *in vacuo*, for further comparison. The ML/MM simulations of the N-methyl-acetamide, the alanine tetrapeptide and the protein complexes used a time step of 1 fs, and the corresponding MM simulations a time step of 2 fs. In every case in which QM/MM simulations were performed for comparison, the time step was the same as in the ML/MM case.

3.3.4 Radial distribution functions and vibrational density of states

Solvation profiles were obtained as radial distribution functions (RDF) from 50 ps long NVT simulations of phenol in water, at the MM, ML/MM and QM/MM levels, using CPPTRAJ.[80] The temporal trace of the O-H distance was also obtained with CPPTRAJ for further analysis. Vibrational density of states (VDOS) were computed to determine the O-H

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

stretching wavenumber by computing the Fourier transform of the derivative of the C-O distance temporal trace, using https://github.com/JonathanSemelak/VDOS.

3.3.5 Umbrella sampling simulations

The torsion free energy profile of N-Methylacetamide C-C-N-C dihedral angle was obtained by performing umbrella simulations at the MM, ML/MM and QM/MM levels, using 50 ps long NVT simulations (300 K) for every case but the QM/MM simulations (for which 5 ps long ones were used). Starting structures for each window were obtained from a steered molecular dynamics simulations. A force constant of 200 kcal/(mol rad^2) was used in each case and windows were spaced by 5 degree each.

For the alanine tetrapeptide, the 2-dimensional free energy surface corresponding to its central Ramachandran angles was obtained. Windows were separated by 6 degrees. The starting structures for each window were obtained starting from a previously equilibrated structure by performing a short MD simulation with a force constant of 50 kcal/(mol rad^2). Sampling on each window used a 125 kcal/(mol rad^2) force bias. Simulations were performed both at the MM and ML/MM levels using 50 ps long NVT simulations. No QM/MM simulations were performed in this case due to the prohibitive amount of QM simulations needed in order to obtain the free energy surface.

The final free energy profile and free energy surfaces were obtained with the umbrella integration method (UI).[81,82]

3.3.6 Analysis of protein-ligand interactions

The performance of our ML/MM implementation for modeling protein-ligand interactions was evaluated by comparing 2 ns long ML/MM and MM trajectories for the FKBP12 complex, in terms of complex stability and the presence of relevant protein-ligand interactions. For the avidin complex, given the smaller size of the ligand in comparison with the FKBP12 complex, we evaluated the accuracy of the predicted protein-ligand interaction energy by

https://doi.org/10.26434/chemrxiv-2024-m9xgc ORCID: https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

comparing it to actual QM/MM calculations. Single point QM/MM calculations were performed on frames took from a 10 ns long ML/MM production run, and the electrostatic coupling energy (i.e. the electrostatic protein-ligand interaction), was compared to that obtained at the ML/MM level. For comparison, the interaction at the MM level was computed for the same structures. The non-electrostatic component of the interaction was omitted in the analysis since its values are equal for all methods.

4 Results and discussion

4.1 Solvation and vibrational spectrum

We begin by presenting the simulation of a single phenol molecule in aqueous solution (Figure 2, panel A). Phenol was chosen because its -OH group participates in hydrogen bonding, offering a simple yet informative solvation structure, along with a vibrational spectrum that is particularly sensitive to its environment.

Panel B illustrates the solvation structure through radial distribution functions. The first solvation shell peak is accurately reproduced at the QM/MM level only when the full ML/MM embedding is used, incorporating on-the-fly predicted MBIS charges with distortion and polarization corrections. In contrast, the MM/MM simulation (i.e. fully MM) shows a lower and broader first solvation shell peak, centered at larger distances. The radial distribution function from the ML/MM simulation using force field charges for the ML/MM region (labeled ML/MM (ME) in the figure) closely resembles that of the MM/MM case. These results underscore the importance of precisely predicting the electrostatic potential of the ML region for capturing solute-solvent interactions. Since solvation patterns beyond the first solvation shell may require longer simulations, our comparison is limited to the first shell.

Panel C presents the vibrational spectrum obtained from the vibrational density of states of the O-H stretching mode. Results for an isolated, *in vacuo*, molecule are also shown to

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

assess the ML/MM implementation's ability to capture the characteristic red shift of phenol in an aqueous environment. This diagnostic feature was chosen because the peak positions in vibrational spectra reveal subtle properties of the potential energy surface, particularly its curvature around local minima.

For the O-H stretching peak, the QM and ML simulations (*in vacuo*) place it around $\sim$3800 cm^{-1}, consistent with prior reports of the ANI-2x model's accuracy in reproducing organic molecule vibrational modes. The introduction of water shifts the peak to lower energy (red shift) by $\sim$400 cm^{-1} at the QM/MM level, indicating a decrease in the curvature of potential energy surface along the O-H mode. While subtle, this shift is still captured by our ML/MM implementation. The agreement is semi-quantitative, showing a $\sim$300 cm^{-1} red shift, but it surpasses the performance of the ML/MM (ME) level. In contrast, the full MM simulations exhibit the O-H stretching peak around $\sim$3000 cm^{-1} for both the isolated and solvated cases.

4.2 Free energy

Thermodynamic information is often derived from MD simulations based on the ergodic hypothesis, which assumes that a sufficiently long simulation will sample a representative population of states. From these, free energy differences can be inferred by analyzing the probability of the system occupying each state. To demonstrate that our ML/MM implementation can accurately reproduce realistic phase spaces (and thus realistic free energy landscapes) we computed the free energy profile for the C-N bond rotation in N-Methylacetamide (Figure 3, Panels A and B) and generated Ramachandran plots for the alanine tetrapeptide in water (Figure 3, Panels C and D).

For N-Methylacetamide, both the ML/MM implementation relying on MBIS charges and the simpler ML/MM (ME) approach showed good agreement with reference QM/MM simulations (Figure 3, Panel B). Even though longer simulations may be necessary for the QM/MM case for a more accurate comparison, the differences in barrier heights and free

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

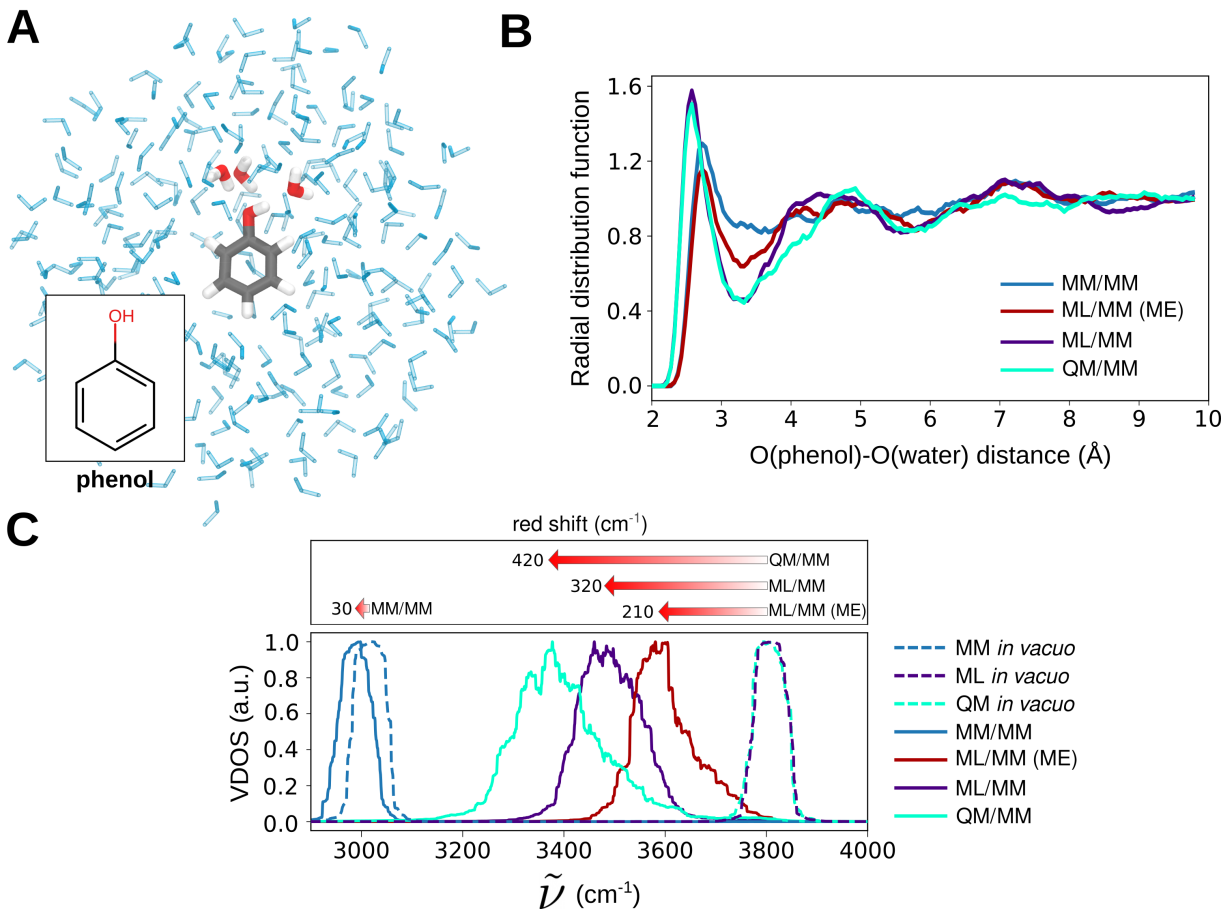

Figure 2: Solvation analysis and vibrational spectrum of phenol in water. (A) Snapshot of a phenol molecule in aqueous solution. (B) Radial distribution functions illustrating the solvation structure around the -OH group. Results are shown for QM/MM, full MM, and ML/MM simulations with different embedding schemes. (C) Vibrational density of states (VDOS) for the O-H stretching mode, comparing simulations of phenol in solution to the isolated molecule (*in vacuo*). The red shift observed in the aqueous environment is captured at the QM/MM and ML/MM levels, with differences in accuracy between embedding methods.

energy changes between the methods were within ∼2 kcal/mol. We hypothesize that the absence of significant deviations at the ML/MM (ME) level may be due to minimal changes in solvation patterns throughout the process. However, the free energy profile obtained from the MM/MM simulations was notably different.

Ramachandran plots depict the free energy landscape associated with the φ and ψ dihedral angles of peptides, which are critical in evaluating conformational changes in proteins. Since these conformational shifts are largely governed by dihedral angle changes, accurately modeling the free energies of these angles is essential for reliable protein simulations. In our study, we explored the free energy surface corresponding to the central Ramachandran angles

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

of the alanine tetrapeptide in water (Figure 3, Panel D), comparing ML/MM and MM/MM results. We selected alanine tetrapeptide as it was one of the peptides used in parametrizing the Amber ff19SB force field, which was also applied here. Remarkably, the key secondary structure minima were consistent across the ML/MM and MM/MM surfaces. These minima include: right-handed α-helix, left-handed α-helix, polyproline II and β-strand, as showed on Figure 3.

Due to the relatively large size of alanine tetrapeptide, the extensive sampling required for a 2-dimensional umbrella sampling at the QM/MM level was not feasible. To verify whether the identified minima align with those from QM/MM, we conducted 1 ps long QM/MM simulations, treating the tetrapeptide at the QM level and starting from each identified minimum. These simulations confirmed that the structures remained stable near their respective minima, indicating that these minima are also present in the QM energy landscape. Furthermore, similar profiles obtained for alanine dipeptide at the same QM/MM level[36] show reasonable agreement with our ML/MM profiles.

4.3 Protein-ligand interactions

The use of MD-based methods for predicting protein-ligand interactions is a common practice in computational biology. A key challenge in these approaches is balancing accuracy and computational efficiency, especially since drug screening is one of the primary applications.

Binding energy is largely influenced by changes in the ligand's conformation due to its interaction with the protein environment, as well as the interaction energy between the protein and ligand. To assess the performance of our ML/MM implementation for modeling protein-ligand interactions, we conducted both MM and ML/MM simulations of the FKBP12 enzyme (a cis-trans peptidyl-prolyl isomerase) in complex with a FK506 related ligand (Figure 4, panels A and B).

Both ML/MM and MM/MM simulations exhibited stable root mean square deviation (RMSD) values throughout the simulation, with the ML/MM system showing slightly greater

https://doi.org/10.26434/chemrxiv-2024-m9xgc ORCID: https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

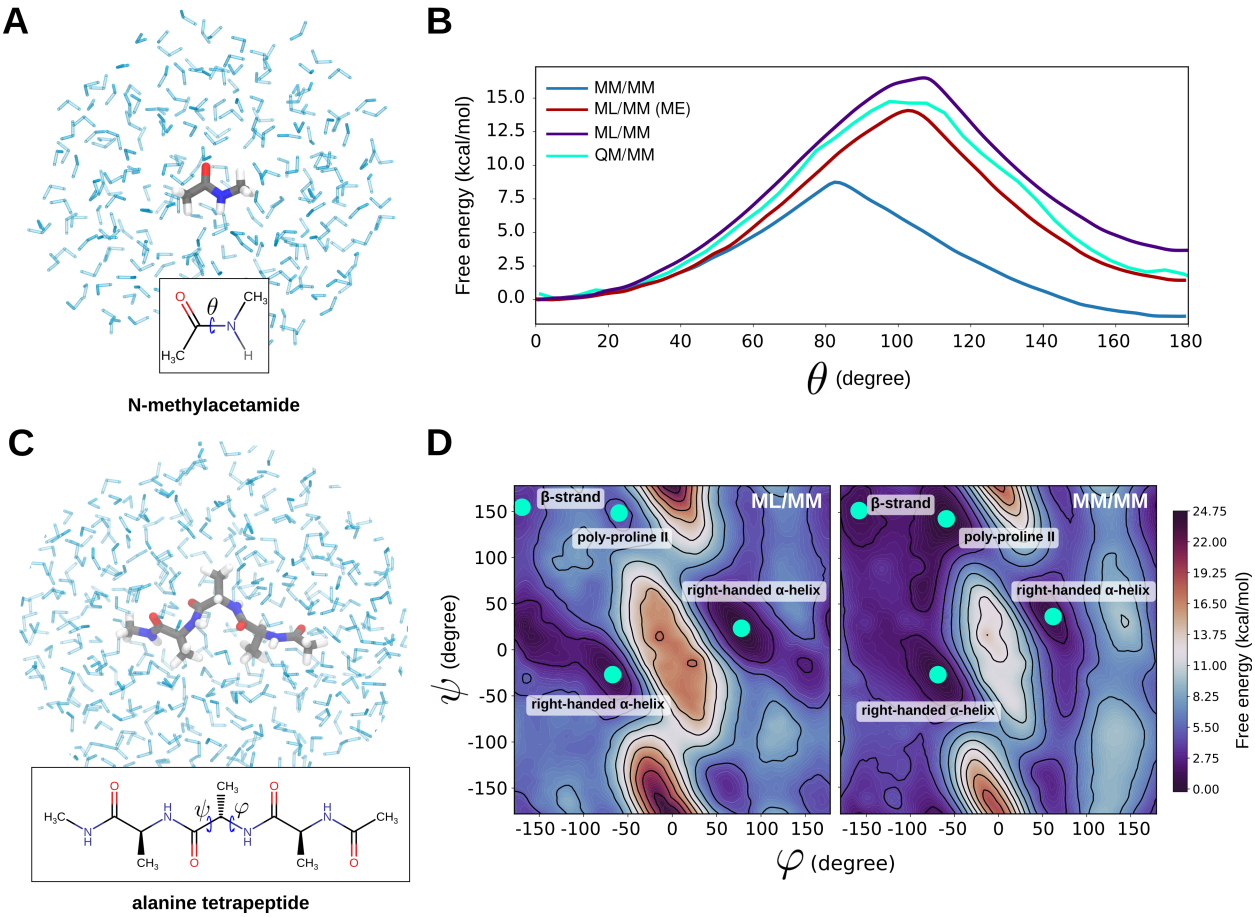

Figure 3: Free energy profiles and conformational analysis of N-Methylacetamide and alanine tetrapeptide in water. (A) Snapshot of a N-Methylacetamide molecule in water. (B) Free energy profile for the rotation of the C-N bond in N-Methylacetamide at the QM/MM, ML/MM, and MM/MM levels. The ML/MM and ML/MM (ME) methods show good agreement with the reference QM/MM profile, with differences in barrier heights and free energy changes within ∼2 kcal/mol. (C) Snapshot of the alanine tetrapeptide in water. (D) Ramachandran plot displaying the free energy surface for the φ and ψ dihedral angles of the alanine tetrapeptide, with ML/MM and MM/MM simulations showing consistent locations of key secondary structure minima, including right-handed α-helix, left-handed α-helix, polyproline II, and β-strand.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

flexibility (not shown). We focused on 13 key protein-ligand contacts identified from previous studies.[71] Hydrophobic contacts were defined for three regions of the ligand: the pipecolyl ring, the phenyl ring, and the tertiary pentyl fragment (Figure 4, panel A). A contact was considered present if the distance between these regions and the side chain of the corresponding amino acid was 4 Å or less. For hydrogen bonds, a distance of 3 Å or less between the acceptor and the donor hydrogen was used as a cutoff.

We analyzed uncorrelated frames from both ML/MM and MM simulations to identify hydrophobic contacts between the tertiary pentyl fragment and Ile56, Leu97, Tyr82, and Ile90; between the phenyl ring and His87, Ile56, and Tyr82; and between the pipecolyl ring and Phe46, Tyr26, Trp59, and Phe99. Additionally, we identified hydrogen bond interactions between O1 and O4 with Ile56 and Ile91, respectively. Hydrogen bonds were observed in nearly all frames, and the number of hydrophobic contacts was similar across both trajectories (Figure 4, panel B).

In the case of the avidin complex (Figure 4, panel C), we quantitatively evaluated the accuracy of our ML/MM implementation for protein-ligand interaction energies. From the 10 ns ML/MM simulation, 200 uncorrelated frames were extracted for single-point energy calculations at the MM/MM and QM/MM levels. The electrostatic interaction energy between the protein and ligand, $E_{\mathrm{PL}}^{\mathrm{elec}}$, was calculated at the ML/MM level as $E_{\mathrm{ML/MM}}^{\mathrm{elec}} + E_{\mathrm{ML}}^{\mathrm{D}}$. For the QM/MM level, this was obtained by subtracting the energy of a mechanical embedding calculation from that of an electrostatic embedding calculation, accounting for interactions between the electronic density and MM point charges as well as distortion effects. At the MM/MM level, a simple coulombic interaction was used. As shown in Figure 4, panel D, the ML/MM electrostatic interaction energies aligned more closely with QM/MM results than with MM/MM calculations.

These findings indicate that our ML/MM approach can produce stable simulations while offering a physically accurate representation of protein-ligand interactions, highlighting its potential for use in in-silico drug design experiments.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

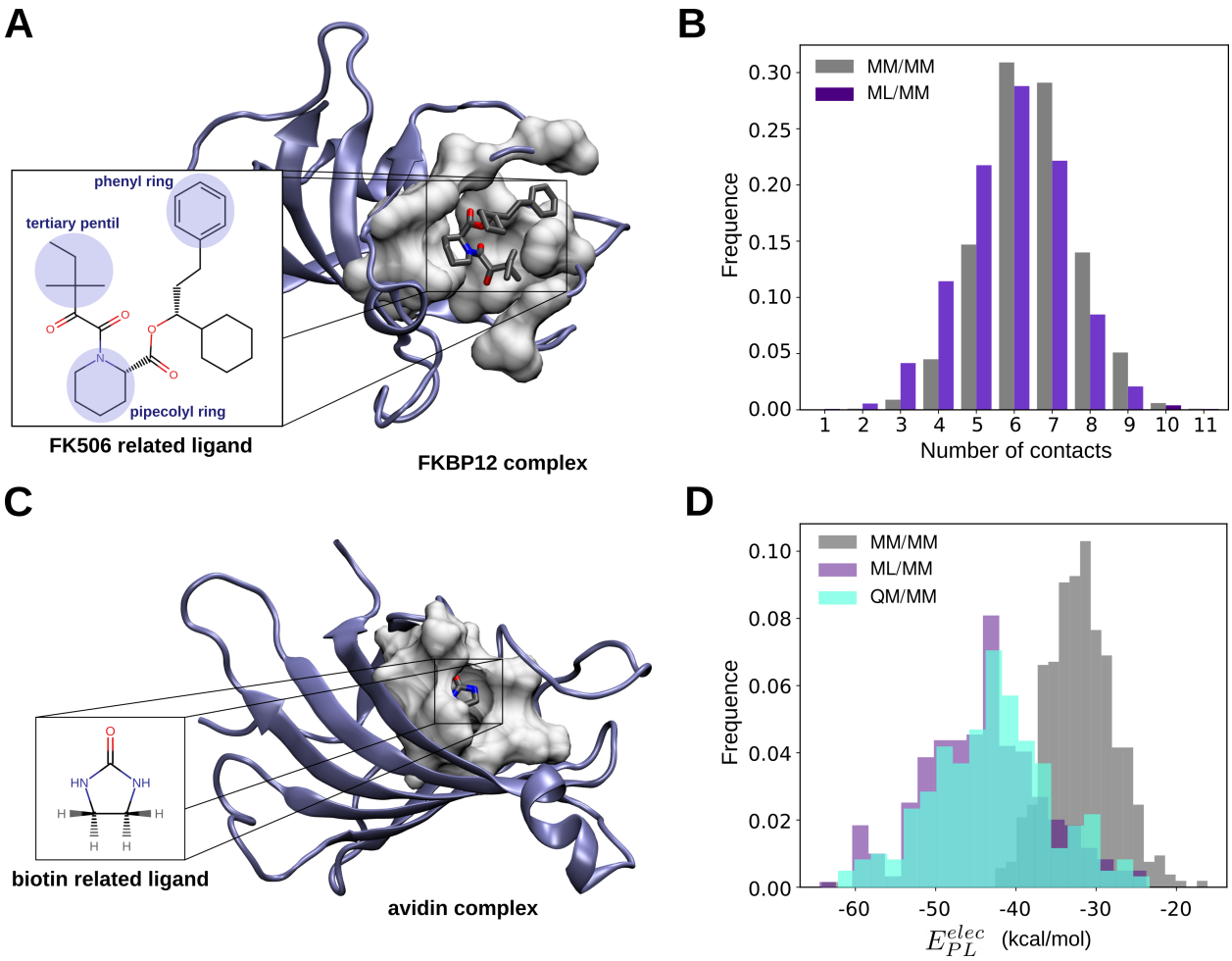

Figure 4: Protein-ligand interaction analysis for FKBP12 and avidin complexes. (A) A snapshot of the FKBP12 complex (excluding water molecules for clarity) in complex with the FK506-related ligand, highlighting the ligand fragments (pipecolyl ring, phenyl ring, and tertiary pentyl fragment) involved in hydrophobic interactions. (B) Histogram of hydrophobic contacts between key ligand fragments and surrounding amino acids from the ML/MM and MM simulations. (C) Snapshot of the avidin complex. (D) Electrostatic interaction energies between the protein and ligand in the avidin complex, calculated at the ML/MM, QM/MM, and MM/MM levels from 200 uncorrelated frames extracted from the 10 ns ML/MM MD simulation.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

5 Conclusions

In this work, we introduced a novel ML/MM framework that integrates ANI neural networks with the AmberTools software suite, enabling accurate simulations of molecular systems with significantly reduced computational cost compared to traditional QM/MM approaches. Our method exploits the ANI-MBIS model for the on-the-fly prediction of MBIS atomic charges and incorporates polarization and distortion corrections, allowing for a more accurate representation of electrostatic interactions within the ML/MM systems. The implementation within the AmberTools software suite ensures seamless integration into existing workflows.

Our results demonstrate that this framework can effectively reproduce free energy profiles, solvation structures, and vibrational spectra in aqueous environments, showing good agreement with QM/MM simulations. Additionally, the method successfully captures key features in protein-ligand interactions, with ML/MM predictions aligning closely with QM/MM calculations for electrostatic energies, particularly in the avidin complex.

The ML/MM framework presented herein is designed as a general-purpose tool for the simulation of neutral organic molecules, requiring no additional training for its use beyond the initial setup. This feature makes it highly accessible and versatile for a wide range of applications in computational chemistry. Furthermore, we anticipate that this implementation will serve as a foundation for future developments in ML models, with potential extensions to simulate charged species and chemical reactions in ML/MM frameworks.

Data Availability Statement

The software developed for this work will be available in future AmberTools releases.

Acknowledgement

J.A.S gratefully acknowledge CONICET for its fellowships. M.G.L. and D.A.E. are University of Buenos Aires and CONICET staff. AER acknowledges supports from a USA National

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Science Foundation Award OAC-2311632.

This research was supported by grants of the Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico,y la Innovación (PICT 2020-01828), Universidad de Buenos Aires (UBACYT 20020220200002BA), and Consejo Nacional de Investigaciones Científicas y Técnicas (11220200102757CO).

References

(1) 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.

(2) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. Frontier bonds in QM/MM methods: A comparison of different approaches. *The Journal of Physical Chemistry A* **2000**, *104*, 1720–1735.

(3) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2007**, *117*, 185–199.

(4) Groenhof, G. Introduction to QM/MM simulations. *Biomolecular simulations: methods and protocols* **2013**, 43–66.

(5) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.

(6) Hagg, A.; Kirschner, K. N. Open-source machine learning in computational chemistry. *Journal of Chemical Information and Modeling* **2023**, *63*, 4505–4532.

(7) Meuwly, M. Machine learning for chemical reactions. *Chemical Reviews* **2021**, *121*, 10218–10239.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(8) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical reviews* **2021**, *121*, 9816–9872.

(9) Dral, P. O. Quantum chemistry in the age of machine learning. *The journal of physical chemistry letters* **2020**, *11*, 2336–2347.

(10) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep learning for computational chemistry. *Journal of computational chemistry* **2017**, *38*, 1291–1307.

(11) Goldman, B. B.; Walters, W. P. Machine learning in computational chemistry. *Annual Reports in Computational Chemistry* **2006**, *2*, 127–140.

(12) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine learning for molecular simulation. *Annual review of physical chemistry* **2020**, *71*, 361–390.

(13) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical science* **2017**, *8*, 3192–3203.

(14) 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. *Nature communications* **2019**, *10*, 2903.

(15) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: A free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *Journal of chemical information and modeling* **2020**, *60*, 3408–3415.

(16) 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

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

to sulfur and halogens. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202.

(17) Unke, O. T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation* **2019**, *15*, 3678–3693.

(18) Schütt, K.; Kessel, P.; Gastegger, M.; Nicoli, K.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A deep learning toolbox for atomistic systems. *Journal of chemical theory and computation* **2018**, *15*, 448–455.

(19) Manzhos, S.; Carrington Jr, T. Neural network potential energy surfaces for small molecules and reactions. *Chemical Reviews* **2020**, *121*, 10187–10217.

(20) Shao, Y.; Hellström, M.; Mitev, P. D.; Knijff, L.; Zhang, C. PiNN: A python library for building atomic neural networks of molecules and materials. *Journal of chemical information and modeling* **2020**, *60*, 1184–1193.

(21) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical science* **2018**, *9*, 2261–2269.

(22) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science advances* **2019**, *5*, eaav6490.

(23) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **2007**, *98*, 146401.

(24) Gallegos, M.; Isamura, B. K.; Popelier, P. L.; Martín Pendás, A. An Unsupervised Machine Learning Approach for the Automatic Construction of Local Chemical Descriptors. *Journal of Chemical Information and Modeling* **2024**, *64*, 3059–3079.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(25) Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials. *BioRxiv* **2020**, 2020–07.

(26) Eastman, P.; Galvelis, R.; Peláez, R. P.; Abreu, C. R.; Farr, S. E.; Gallicchio, E.; Gorenko, A.; Henry, M. M.; Hu, F.; Huang, J.; others OpenMM 8: molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B* **2023**, *128*, 109–116.

(27) Galvelis, R.; Varela-Rial, A.; Doerr, S.; Fino, R.; Eastman, P.; Markland, T. E.; Chodera, J. D.; De Fabritiis, G. NNP/MM: accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics. *Journal of chemical information and modeling* **2023**, *63*, 5701–5708.

(28) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *The Journal of Physical Chemistry* **1996**, *100*, 10580–10594.

(29) 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. *Journal of chemical theory and computation* **2021**, *17*, 6993–7009.

(30) Böselt, L.; Thürlemann, M.; Riniker, S. Machine learning in QM/MM molecular dynamics simulations of condensed-phase systems. *Journal of Chemical Theory and Computation* **2021**, *17*, 2641–2658.

(31) 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. *Journal of Chemical Theory and Computation* **2021**, *17*, 5745–5758.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(32) Gastegger, M.; Schütt, K. T.; Müller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chemical science* **2021**, *12*, 11473–11483.

(33) Hofstetter, A.; Böselt, L.; Riniker, S. Graph-convolutional neural networks for (QM) ML/MM molecular dynamics simulations. *Physical Chemistry Chemical Physics* **2022**, *24*, 22497–22512.

(34) Lei, Y.-K.; Yagi, K.; Sugita, Y. Learning QM/MM potential using equivariant multi-scale model. *The Journal of Chemical Physics* **2024**, *160*.

(35) Zinovjev, K. Electrostatic Embedding of Machine Learning Potentials. *Journal of Chemical Theory and Computation* **2023**, *19*, 1888–1897.

(36) Zinovjev, K.; Hedges, L.; Montagud Andreu, R.; Woods, C.; Tuñón, I.; van der Kamp, M. W. emle-engine: a flexible electrostatic machine learning embedding package for multiscale molecular dynamics simulations. *Journal of Chemical Theory and Computation* **2024**,

(37) Grassano, J. S.; Pickering, I.; Roitberg, A. E.; González Lebrero, M. C.; Estrin, D. A.; Semelak, J. A. Assessment of Embedding Schemes in a Hybrid Machine Learning/Classical Potentials (ML/MM) Approach. *Journal of Chemical Information and Modeling* **2024**, *64*, 4047–4058.

(38) Kancharlapalli, S.; Gopalan, A.; Haranczyk, M.; Snurr, R. Q. Fast and accurate machine learning strategy for calculating partial atomic charges in metal–organic frameworks. *Journal of Chemical Theory and Computation* **2021**, *17*, 3052–3064.

(39) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine learning of partial charges derived from high-quality quantum-mechanical calculations. *Journal of chemical information and modeling* **2018**, *58*, 579–590.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(40) Sifain, A. E.; Lubbers, N.; Nebgen, B. T.; Smith, J. S.; Lokhov, A. Y.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Discovering a transferable charge assignment model using machine learning. *The journal of physical chemistry letters* **2018**, *9*, 4495–4501.

(41) Bereau, T.; Andrienko, D.; Von Lilienfeld, O. A. Transferable atomic multipole machine learning models for small organic molecules. *Journal of chemical theory and computation* **2015**, *11*, 3225–3233.

(42) Korolev, V. V.; Mitrofanov, A.; Marchenko, E. I.; Eremin, N. N.; Tkachenko, V.; Kalmykov, S. N. Transferable and extensible machine learning-derived atomic charges for modeling hybrid nanoporous materials. *Chemistry of Materials* **2020**, *32*, 7822–7831.

(43) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal basis iterative stockholder: atoms in molecules for force-field development. *Journal of Chemical Theory and Computation* **2016**, *12*, 3894–3912.

(44) Huddleston, K. K. Exploring the Atom in a Molecule through Atomic Energies and Partial Charges Using the ANAKIN-ME Methodology. PhD Dissertation, University of Florida, 2024.

(45) Huddleston, K.; Zubatyuk, R.; Smith, J.; Roitberg, A.; Isayev, O.; Pickering, I.; Devereux, C.; Barros, K. ANI-2x Release. `https://doi.org/10.5281/zenodo.10108942`, 2023; Data set.

(46) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *The Journal of Chemical Physics* **2008**, *128*, 084106.

(47) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; others Pytorch: An imperative style, high-

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

performance deep learning library. *Advances in neural information processing systems* **2019**, *32*.

(48) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E. I.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. J. AmberTools. *Journal of Chemical Information and Modeling* **2023**, *63*, 6183–6191.

(49) Berendsen, H. J.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(50) Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47*, 2795–2803.

(51) Milne, A. W.; Jorge, M. Polarization Corrections and the Hydration Free Energy of Water. *J. Chem. Theory Comput.* **2018**, *15*, 1065–1078.

(52) Thole, B. T. Molecular Polarizabilities Calculated With a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(53) Schwerdtfeger, P.; Nagle, K. Table of Static Dipole Polarizabilities of the Neutral Elements in the Periodic Table. *Mol. Phys.* **2019**, *117*, 1200–1225.

(54) Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. International Conference on Learning Representations. 2018.

(55) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2017.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(56) Cárdenas, C.; Heidar-Zadeh, F.; Ayers, P. W. Benchmark Values of Chemical Potential and Chemical Hardness for Atoms and Atomic Ions (Including Unstable Anions) from the Energies of Isoelectronic Series. *Physical Chemistry Chemical Physics* **2016**, *18*, 25721–25734.

(57) Götz, A. W.; Clark, M. A.; Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *Journal of computational chemistry* **2014**, *35*, 95–108.

(58) Hu, H.; Yang, W. Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes. *Journal of Molecular Structure: THEOCHEM* **2009**, *898*, 17–30.

(59) Gao, J. Combined QM/MM simulation study of the Claisen rearrangement of allyl vinyl ether in aqueous solution. *Journal of the American Chemical Society* **1994**, *116*, 1563–1564.

(60) Callis, P. R.; Vivian, J. T. Understanding the variable fluorescence quantum yield of tryptophan in proteins using QM-MM simulations. Quenching by charge transfer to the peptide backbone. *Chemical physics letters* **2003**, *369*, 409–414.

(61) Morzan, U. N.; Alonso de Armino, D. J.; Foglia, N. O.; Ramirez, F.; Gonzalez Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in complex environments from QM-MM simulations. *Chemical reviews* **2018**, *118*, 4071–4113.

(62) Gleeson, M. P.; Gleeson, D. QM/MM calculations in drug discovery: a useful method for studying binding phenomena? *Journal of chemical information and modeling* **2009**, *49*, 670–677.

(63) Lodola, A.; De Vivo, M. The increasing role of QM/MM in drug discovery. *Advances in protein chemistry and structural biology* **2012**, *87*, 337–362.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(64) Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H. J.; Konialian, A. L.; Yen, H. K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J.; Levy, M. A.; others Design, synthesis, and kinetic evaluation of high-affinity FKBP ligands and the X-ray crystal structures of their complexes with FKBP12. *Journal of the American Chemical Society* **2002**, *115*, 9925–9938.

(65) Green, N. Thermodynamics of the binding of biotin and some analogues by avidin. *Biochemical Journal* **1966**, *101*, 774–780.

(66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(67) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.

(68) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.

(69) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, 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. *Journal of Chemical Theory and Computation* **2020**, *16*, 528–552.

(70) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. Three-dimensional Structure of the Tetragonal Crystal Form of Egg-white Avidin in its functional Complex with Biotin at 2· 7 Å Resolution. *Journal of molecular biology* **1993**, *231*, 698–710.

(71) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

dynamics simulations with restraining potentials. *Biophysical journal* **2006**, *91*, 2798–2814.

(72) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54*, 5.6.1–5.6.37.

(73) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 09 Revision A.2. Gaussian, Inc., 2009.

(74) Schafmeister, C.; Ross, W.; Romanovski, V.; David, R.; Kollman, P. A. LEaP. University of California, San Francisco, 1995.

(75) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(76) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(77) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(78) 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. *Journal of computational physics* **1977**, *23*, 327–341.

(79) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **1978**, *17*, 1302.

(80) Roe, D. R.; Cheatham, T. E. I. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.

(81) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration". *The Journal of Chemical Physics* **2005**, *123*, 144104.

(82) Kästner, J.; Thiel, W. Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration. *The Journal of Chemical Physics* **2006**, *124*, 234106.

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

For Table of Contents Use Only

ANI Neural Networks Meet Electrostatics for a ML/MM Implementation in Amber

Jonathan Alexis Semelak, Ignacio Pickering, Kate Huddleston, Justo Olmos, Camila Clemente, Juan Santiago Grassano, Marcelo Martí, Mariano Camilo Gonzalez Lebrero, Adrian Roitberg*, and Dario Ariel Estrin*

https://doi.org/10.26434/chemrxiv-2024-m9xgc **ORCID:** https://orcid.org/0000-0002-9615-8405 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0