

# Tell Machine Learning Potentials What They are Needed for: Simulation-Oriented Training Exemplified for Glycine

Fuchun Ge,<sup>1†</sup> Ran Wang,<sup>1†</sup> Chen Qu,<sup>2</sup> Peikun Zheng,<sup>1,3</sup> Apurba Nandi,<sup>4,5</sup>  
Riccardo Conte,<sup>6</sup> Paul L. Houston,<sup>7</sup> Joel M. Bowman,<sup>4\*</sup> Pavlo O. Dral<sup>1\*</sup>

<sup>1</sup>*State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, and Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), Xiamen University, Xiamen, Fujian 361005, China*

<sup>2</sup>*Independent Researcher, Toronto, Ontario M9B0E3, Canada*

<sup>3</sup>*Present address: Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*

<sup>4</sup>*Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia, 30322, United States*

<sup>5</sup>*Department of Physics and Materials Science, University of Luxembourg, Luxembourg City L-1511, Luxembourg*

<sup>6</sup>*Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*

<sup>7</sup>*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States; Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

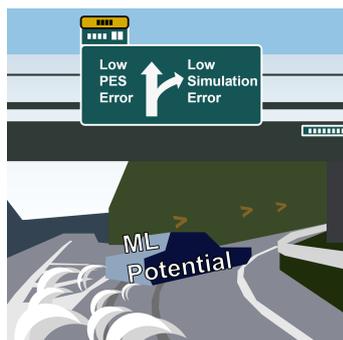
*E-mails: [jmbowma@emory.edu](mailto:jmbowma@emory.edu); [dral@xmu.edu.cn](mailto:dral@xmu.edu.cn)*

*†Equal contribution*

## Abstract

Machine learning potentials (MLPs) are widely applied as an efficient alternative way to represent potential energy surfaces (PES) in many chemical simulations. The MLPs are often evaluated with the root-mean-square errors on the test set drawn from the same distribution as the training data. Here, we systematically investigate the relationship between such test errors and the simulation accuracy with MLPs on an example of a full-dimensional, global PES for the glycine amino acid. Our results show that the errors in the test set do not unambiguously reflect the MLP performance in different simulation tasks such as relative conformer energies, barriers, vibrational levels, and zero-point vibrational energies. We also offer an easily accessible solution for improving the MLP quality in a simulation-oriented manner, yielding the most precise relative conformer energies and barriers. This solution also passed the stringent test by the diffusion Monte Carlo simulations.

## TOC Graphic



Potential energy surface (PES)<sup>1-5</sup> is one of the most essential concepts in computational chemistry, which can be abstracted as a multivariable function of the positions of nuclei with the output indicating the potential energy profile<sup>6, 7</sup>. The negative gradients of this function are the forces that atoms experience. PES is a potent instrument that allows us to understand molecular structures<sup>8</sup>, their stability<sup>9-11</sup>, and reactivity governed by nuclear dynamics<sup>12-18</sup>. PES is extensively utilized for the exploration of the conformational landscape and reaction mechanisms<sup>19-34</sup>.

Despite PES being an essential prerequisite to perform many chemical simulations, the construction often becomes a bottleneck in the application: highly accurate *ab initio* electronic structure methods offer reliable potential energies but have high computational cost and scaling, while low-level empirical or semiempirical methods are fast but less robust. Fortunately, this conflict can be tackled by applying machine learning (ML) techniques to return an analytic functional form that mimics the PES, resulting in what is called machine learning potentials (MLPs).<sup>35-51</sup> By serving as an efficient replacement for quantum mechanically generated potential energies (and gradients), MLPs have developed rapidly over the past decade and various algorithms emerged. MLPs based on linear regression,<sup>49, 52</sup> using permutationally invariant polynomials,<sup>50, 51</sup> neural networks (NN),<sup>44, 47, 53</sup> and kernel methods (KM)<sup>48, 54-58</sup> have been broadly applied for representations of correct *ab initio* PES for atomistic simulations in chemistry, physics, and materials science.<sup>35-40, 43-52</sup>

However, the convenience of MLPs, a black-box statistical approximator, comes with an inborn drawback of lacking the physical model foundation, making it difficult to get an intuitive sense of its reliability. The usual practice is to chase numerical metrics such as RMSEs (root-mean-square error) and MAEs (mean absolute error) on the test set drawn from the same distribution as the training data to evaluate MLPs' performance in PES. The hope is that these metrics would reflect the MLP performance in the actual simulation. Nevertheless, these numerical metrics, on the one hand, are *not representative*: most of the metrics are single numbers that only give an overall description of the error, this "average accuracy" cannot regard all the nuances in individual PES regions. On the other hand, they can be *biased* or even deceptive since the values are highly dependent on the composition of the test data, unable to reflect the poor prediction of the ML model for those data not drawn from the same distribution as the training set.

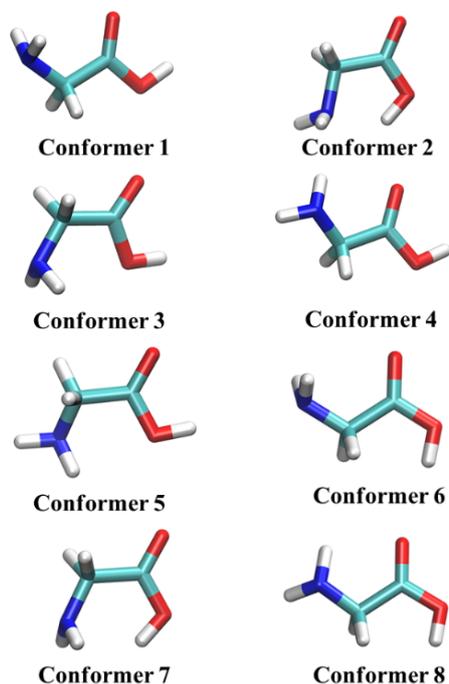
To overcome this limitation and obtain more reliable MLPs on chemical simulations, the obvious way is to directly validate MLPs by performing and examining the correctness of simulations with MLPs. An increasing number of studies attempted to evaluate MLPs beyond the statistical errors like MAEs and RMSEs on the test sets. Ong *et al.* presented an evaluation of MLP models based on local environment descriptors (GAP, MTP, NNP, SNAP, and qSNAP) and made a noteworthy observation of the absence of a consistent relationship between the accuracy in energies or forces and the accuracy in chemical simulations.<sup>59</sup> For instance, the SNAP model demonstrates superior performance in accurately reproducing polymorphic energy differences across various systems, even though it exhibits significantly higher RMSEs in energies and forces. Similarly, despite the GAP model exhibiting relatively low RMSE in predicting energies and forces, it demonstrates the least accuracy in polymorphic energy differences. Unfortunately, they did not delve into the underlying causation. Tkatchenko *et al.* assessed the performance of four MLPs including BPNN, SchNet, GAP/SOAP, and sGDML on the PES of glycine and azobenzene molecules and to check the reliability of MLPs, MD simulations were employed to reproduce the transition processes of azobenzene.<sup>60</sup> Their analysis revealed the qualitative problems with MLPs in trajectories. Csanyi *et al.* carried out error evaluations for simulation tasks beyond test RMSE, such as normal-mode prediction and dihedral torsional profile prediction.<sup>61</sup> More than that, MLPs whose test errors are not that bad may lead to catastrophic, non-physical events in the simulations: the most common one is molecular dissociation which should not happen.<sup>61-64</sup> Bowman and co-workers have introduced the term “DMC-certified” to indicate that the MLP can be used in a rigorous, diffusion Monte Carlo (DMC) calculation of the zero-point vibrational energy and wavefunction.<sup>65</sup> These studies provided valuable insight that the evaluation of MLPs only on the test RMSE is insufficient for judging their quality in actual simulations.

One common feature of the challenging cases is the proper description of the rich conformational space. The situation should be even more complicated when the global PES is needed,<sup>39, 55, 66</sup> i.e., for accurate vibrational spectra and zero-point vibrational energy calculations, e.g., with DMC, where the potential is needed that can be applied for very different geometries, including highly distorted ones. Training such potentials is a challenge in itself as the energy distribution is much broader and the quality of the reference quantum mechanical data can be problematic too. These challenges were the focus of a recent Perspective article that specifically examined the limitations of the MD17 dataset compared

to much more extensive datasets for a number of molecules including malonaldehyde and glycine.<sup>65</sup> The latter molecule is the subject of this paper. The suggested solutions to fitting global PES try to improve the MLPs accuracy for the more important low-energy regions with, e.g., weighting schemes (where higher weights are assigned to lower energy during the fitting) or by even ignoring higher-energy points above some cutoff during the hyperparameter tuning stage and further refinement with the self-correction procedures. Such solutions are still quite cumbersome and usually require manual experimentations with the choice of suitable weighting functions<sup>67</sup> or cutoffs<sup>55, 67</sup> to obtain reasonable results.

Inspired by the above challenges and proposed solutions, we investigate how to construct robust, ‘DMC-certified’, MLP for the global PES featuring multiple conformers and transitions between them. We take a challenging example glycine which has eight different conformers (Figure 1).<sup>67</sup> Even non-global glycine PES was highlighted as a challenge to different classes of the state-of-the-art MLPs, including the neural network (NN) and kernel methods with local and global descriptors.<sup>60</sup> No clear solution was suggested. Glycine global PES was successfully fit with the permutationally invariant polynomial (PIP) MLP only after the use of a simple weighting function and using high-energy cutoff to ignore energy gradient information for highly distorted structures during the fitting.<sup>67</sup> Hence, the question also remains, whether such complicated cases can even be described with the other off-the-shelf MLPs and, more importantly, whether a general solution to make it possible and easier is feasible. In this study, we show for ANI-type NNs that while off-the-shelf implementation indeed cannot be used, it is possible to systematically improve their performance to the state-of-the-art level (i.e., achieving accuracy on par with PIP). ANI is an abbreviation for ANAKIN-ME, which in turn stands for Accurate Neural network engINe for Molecular Energies.<sup>68</sup> It employs a modified version of Behler and Parrinello symmetry functions<sup>35</sup> to generate single-atom atomic environment vectors for the representation of molecules. Despite ANI’s known problems of using the non-global descriptor<sup>69</sup> and being less accurate than some of the other MLP types based on the test RMSE evaluations.<sup>46, 70</sup> We develop a protocol to fit such an NN potential semi-automatically by focusing on getting the simulation result right instead of the lower test RMSE. The key idea behind the protocol is telling the NN what parts of PESs are more important than others by adjusting the weights for the errors in training points so that the resulting simulation error is the smallest. In our case, we use the flexible energy weighting function which gives higher weights for the more important low-

energy regions of PES. This approach should be general enough to be applied in similar cases (e.g., different NN potentials and applications).



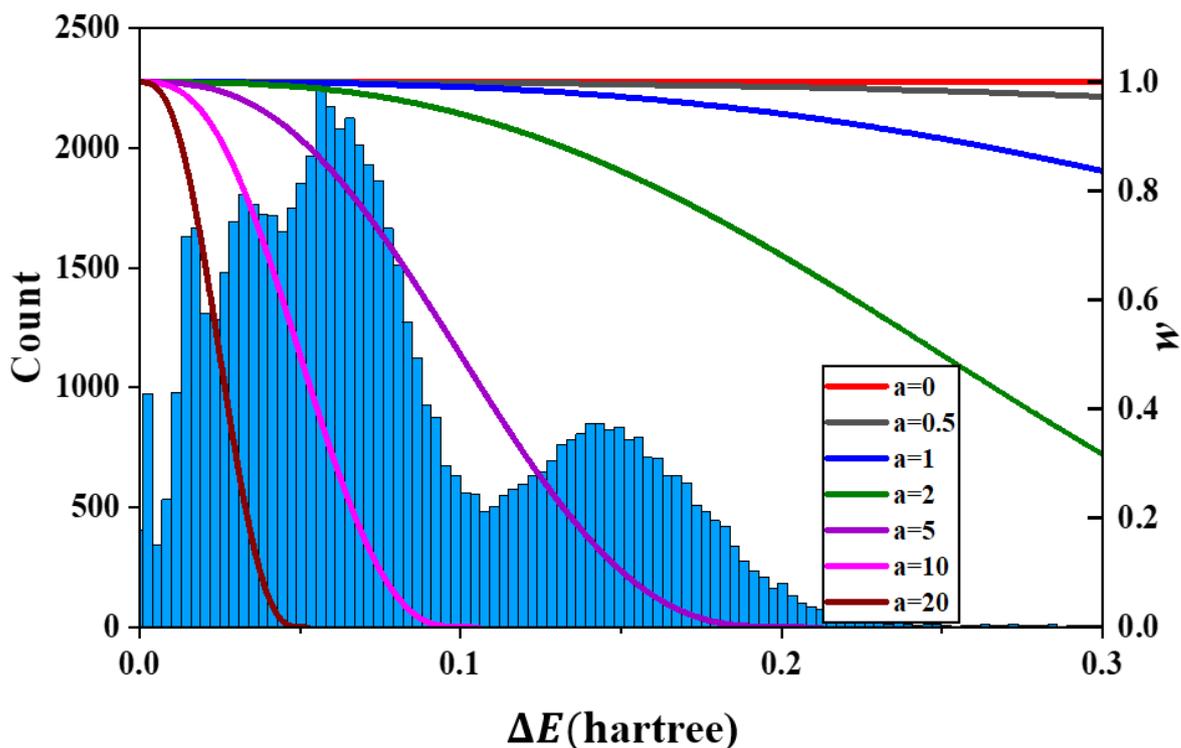
**Figure 1.** Equilibrium structures of eight conformers of glycine.

*How well does the test RMSE reflect the simulation performance?* We first investigate how well the test RMSE of an MLP reflects its performance for a glycine PES that aims to describe these conformers. The data set used is taken from the literature and its construction was not a focus of this study;<sup>67</sup> it contains 70099 geometries with B3LYP/aug-cc-pVDZ energies and gradients; these were shown to have excellent agreement with the gold-standard coupled-cluster level for the relative energies of the 8 conformers<sup>71</sup>. The dataset was obtained from every 10th step of AIMD trajectories initiated at the minima of conformers 1, 2, 3, 4 as well as several saddle points separating them for total energies between 100 and 80,000  $\text{cm}^{-1}$ . Additional energies were added to fill in holes found in DMC calculations (details are given in Ref. <sup>67</sup>). To estimate the test RMSE we employ the standard approach of splitting the data set into the training and test set with a ratio of 9:1 and calculate the RMSE for energy predictions on the test set. We fit the ANI-type<sup>53</sup> NN on the training data set by using 10% of it for validation. The fitting is performed with the off-the-shelf setting as implemented in MLatom<sup>72</sup> interfaced to TorchANI<sup>73</sup> (see Methods). As the regions on the PES with very high energy are less important in actual simulations (see below), the general solution in the

literature is to downweight the contributions from this region. Thus, here we also implemented the training of ANI-type NNs with a weighting function. We choose the weighting function which has the desired form of providing larger weights  $w$  to points with lower energies while also having significant weights for higher-energy regions as defined by the energy  $\Delta E$  (in hartree) relative to the global minimum:

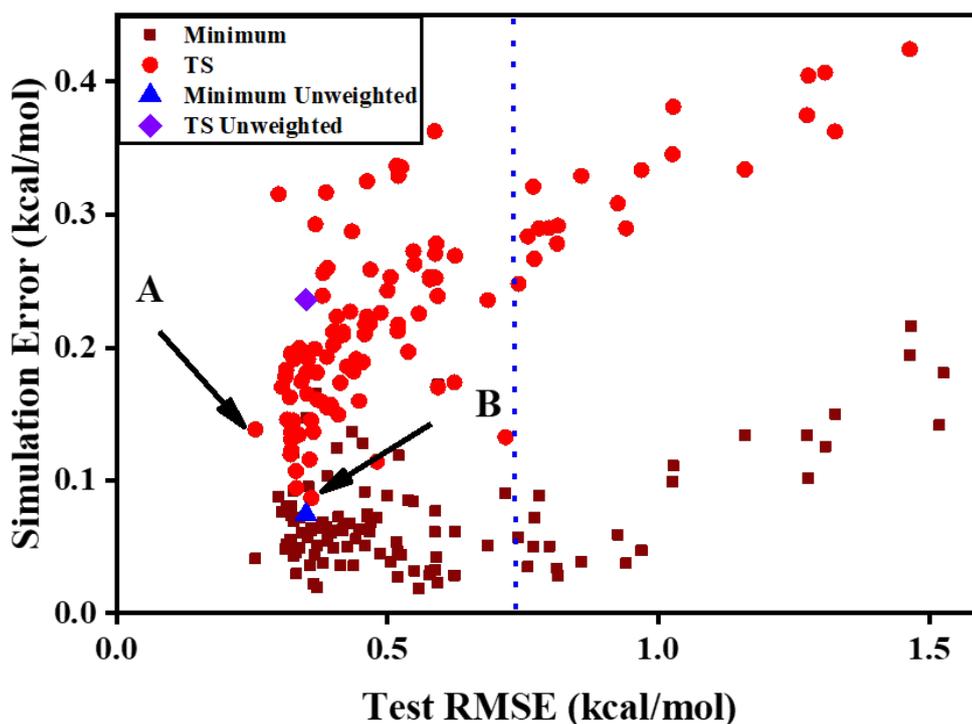
$$w(\Delta E) = \max_i \frac{1}{1 + a(\Delta E - E_i)^2}$$

This function was the best among many possible choices we tried. The original form without parameter  $a$  was introduced for a different problem in the literature.<sup>74</sup> The parameter  $a$  gives us the additional flexibility of controlling how quickly the weighting function drops to 0 as the relative energy  $\Delta E$  increases (Figure 2). The larger  $a$  is, the more quickly the weighting function drops, and the less significant the high-energy data points are. Note that for  $a=0$ , all weights  $w=1$ , i.e., the training is equivalent to the off-the-shelf training of ANI without any weighting function. This is the only difference between the off-the-shelf ANI and other weighted variants we discuss below.



**Figure 2.** The shape of the weighting function is defined by Eq. 1. The dependence of the weighting function shape on the parameter  $a$  is shown by the curves, and the histogram of the distribution of  $\Delta E$  of the data set is also shown.

We train multiple MLPs for a set of parameters  $a$  (including  $a=0$ ) by restarting the training with different randomly initiated NN weights, yielding different test RMSEs even for the same parameter  $a$  (see Methods). For an illustration of the MLP's performance in simulations, we perform geometry optimizations of minima and transition states and evaluate the simulation error defined as the mean absolute error (MAE) in the energies of these stationary points relative to the global minimum. The general trend is that the lower the test RMSE the lower the simulation error (Figure 3). However, the clear correlation is only up to some point (denoted by the blue dotted line in Figure 3). For the region with smaller test RMSEs, no correlation is observed and the spread is quite broad. Two highlighted examples marked **A** and **B** in Figure 3 show extreme cases: for a relatively small test RMSE of 0.26 kcal/mol (point **A**) we can have a rather large simulation MAE of 0.14 kcal/mol for transition states while for a larger test RMSE of 0.36 kcal/mol (point **B**), we obtain two times smaller MAE of 0.08 kcal/mol. This has big implications if highly accurate results are needed as we cannot judge the quality of the MLP for the simulations solely based on the test RMSE. The simulation errors for the off-the-shelf ANI trained without the weighting function (equivalent to using  $a=0$  in the weighting function, marked 'unweighted' in Figure 3) are higher than the errors of both **A** and **B** models trained with the weighting function, while the test RMSE of off-the-shelf ANI is in between the test errors of the **A** and **B** models. All of these observations clearly illustrate the insufficient reliability of judging the model quality by only checking the test RMSE.



**Figure 3.** The relationship between the test RMSE and the simulation error (mean absolute error, MAE) performance of ANI-type NN in simulations (relative energies of conformers and transition states). Test RMSE is the root-mean-square error in energies, TS represents the MAE of MLPs for predicting fifteen main transition states connecting the local minima, Minima represents the MAE in predicting relative energies of all the eight minima compared with the B3LYP/aug-cc-pVDZ method.

*Simulation-oriented training strategy.* Above we saw that the simulation result is very sensitive to the MLP training and just looking at the test RMSE does not provide a confident estimate of the quality of the simulations. Thus, here we report the implementation of the automatable, simulation-oriented training protocol aiming at obtaining the robust MLP with good performance in actual simulations rather than only checking the test RMSE.

The key to successful simulation-oriented training lies in a) choosing an appropriate, computationally affordable, property-based metric describing the quality of the simulations and b) designing an effective model-optimization strategy based on the chosen metric. We fine-tune the weighting function to optimize the model so that it produces better performance in simulations judged property-based metric. Here, we choose the error in the relative energies of eight glycine conformers. To independently validate the model we use the related but different property-based metrics such as errors in transition state energies, normal mode frequencies, and zero-point vibrational energies.

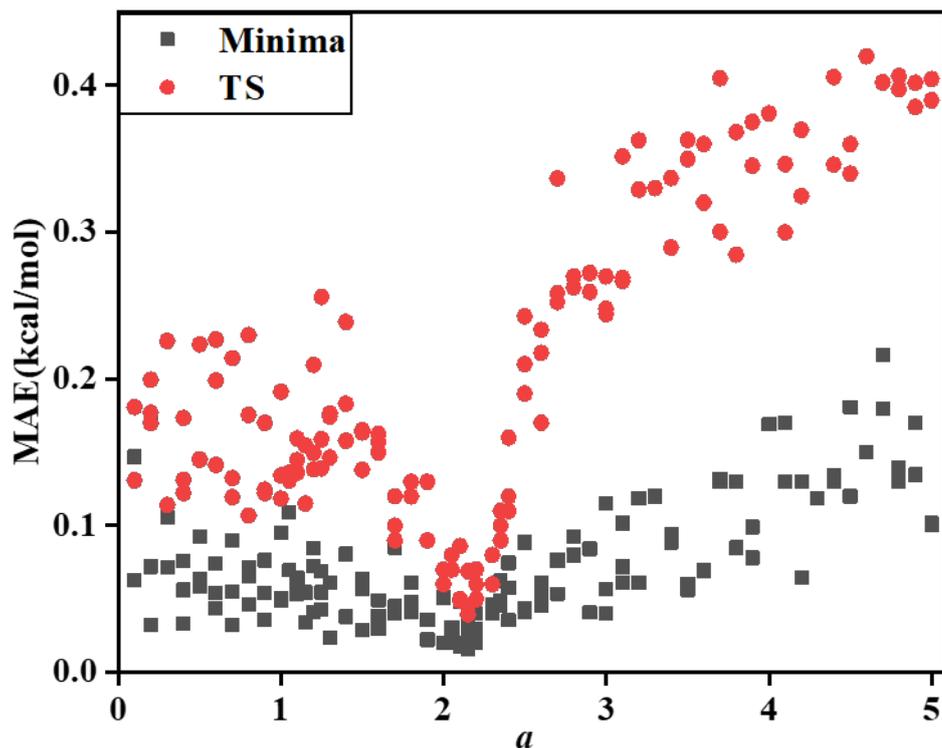
Of course, the reference values for the property-based metric used for model optimization should be known. It means that the benefit of using MLP should be then in

applying it to related but much more computationally expensive simulation tasks. Later we will show that the resulting MLP models can be successfully used for such related tasks: very computationally expensive DMC calculations.

To design an effective model-optimization strategy, we optimize the parameter  $a$  in the weighting function defined by Eq. 1 that controls the regional importance of the training points based on relative energies. We treat this parameter as the hyperparameter to be tuned for the model selection, i.e., we optimize it to obtain the lowest simulation MAE in relative energies of conformers with the MLP being fitted.

Model selection can be easily done given the simplicity of our approach: we just need to scan different values of the parameter  $a$  to obtain the lowest simulation MAE for the minima. We see that such an optimal value of  $a$  is around 2 for the global glycine PES (Figure 4).

To ensure that this strategy does not overfit, we also investigate whether lower errors for one type of simulation lead to lower errors in another, related, type of simulation. Here, we check the MAE for the relative energy values of fifteen primary transition states predicted by the ANI (denoted as TS, red circles in Figure 4). These MAEs were not used for the model selection, i.e., they represent a completely independent test. The trend observed in the TS convincingly aligns with the trend observed for the minima. This indicates a close correlation between the accuracy of transition states predicted by the MLP and the accuracy of the energy minima predicted by the same MLP trained using our strategy. Furthermore, it is noteworthy that as the parameter increases (exceeding 2.5), the TS energies show a more rapid increase in errors, which implies that while different simulations (such as identifying energy minima and transition states on the PES) exhibit an overall similar dependence on the weighting function, they do not share the same level of sensitivity to different regions in the training PES data set.



**Figure 4.** Optimization of the parameter  $a$  in the weighting function. MAE represents the mean absolute error of MLPs for predicting relative energies of all the eight minima compared with B3LYP/aug-cc-pVDZ method (labeled Minima, black) and of the fifteen main transition states (labeled TS, red) connecting the local minima.

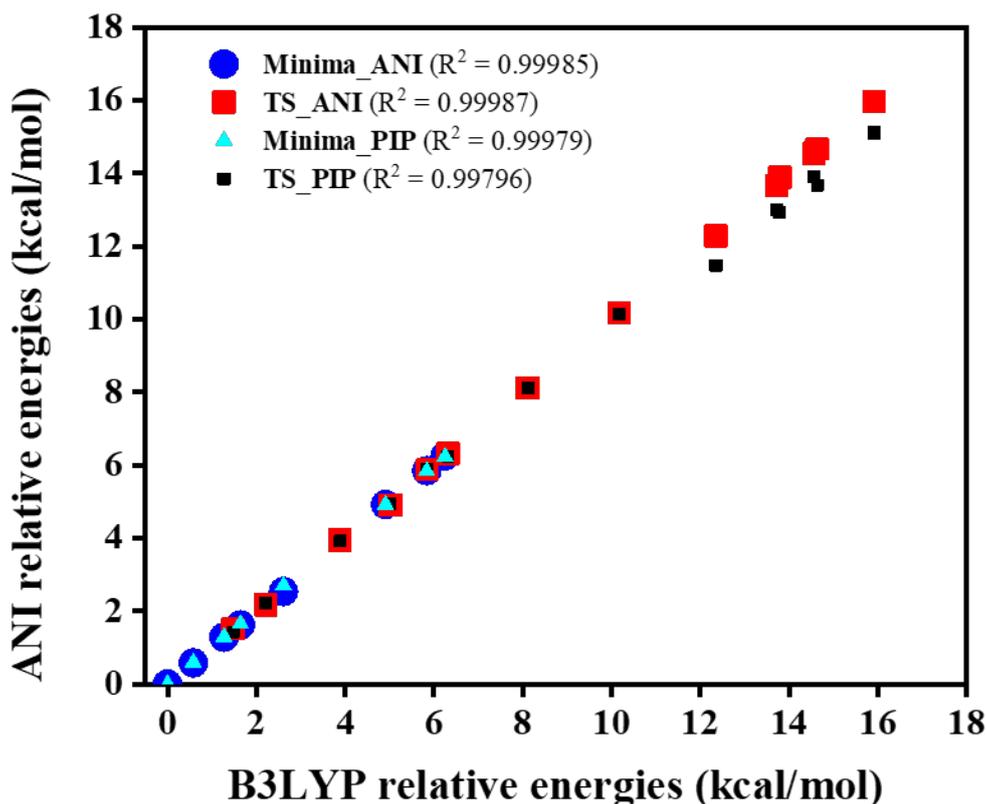
*Performance for the minima and transition states.* Having identified the optimal range of  $a$  around 2.0 that corresponds to both small MAE in relative energies of minima and TSs, we subsequently fine-tuned  $a$  around 2.0 based on MAE in minima and ultimately obtained the best model with  $a$  set to 2.15. We will refer to this MLP as ‘the best ANI model’. The best ANI model has an MAE value of 0.016 kcal/mol for the relative energies of the eight minima and an MAE value of 0.039 kcal/mol for the relative energies of the 15 primary saddle points (Table 1). The simulation errors of energy-weighted ANI models with  $a$  set to 2.15 are consistently much lower than those of off-the-shelf ANI trained without using the weighting function (as can be seen from results of repeated experiments, Table 1). This is despite off-

the-shelf ANI model’s lower test RMSEs indicating that the lower test RMSEs do not guarantee better performance in simulations.

The energies of each conformer relative to the global minimum given by the best ANI model and the reference B3LYP/aug-cc-pVDZ method are shown in Figure 5. Since all the points are essentially located on the diagonal of Figure 5, this implies the best ANI model we trained under the energy-weighted scheme is capable of accurately identifying all the eight energy minima and fifteen main saddle points on the PES of glycine. Compared to the B3LYP/aug-cc-pVDZ method which generates the dataset, our training can significantly reduce the computational cost for simulations without sacrificing accuracy.

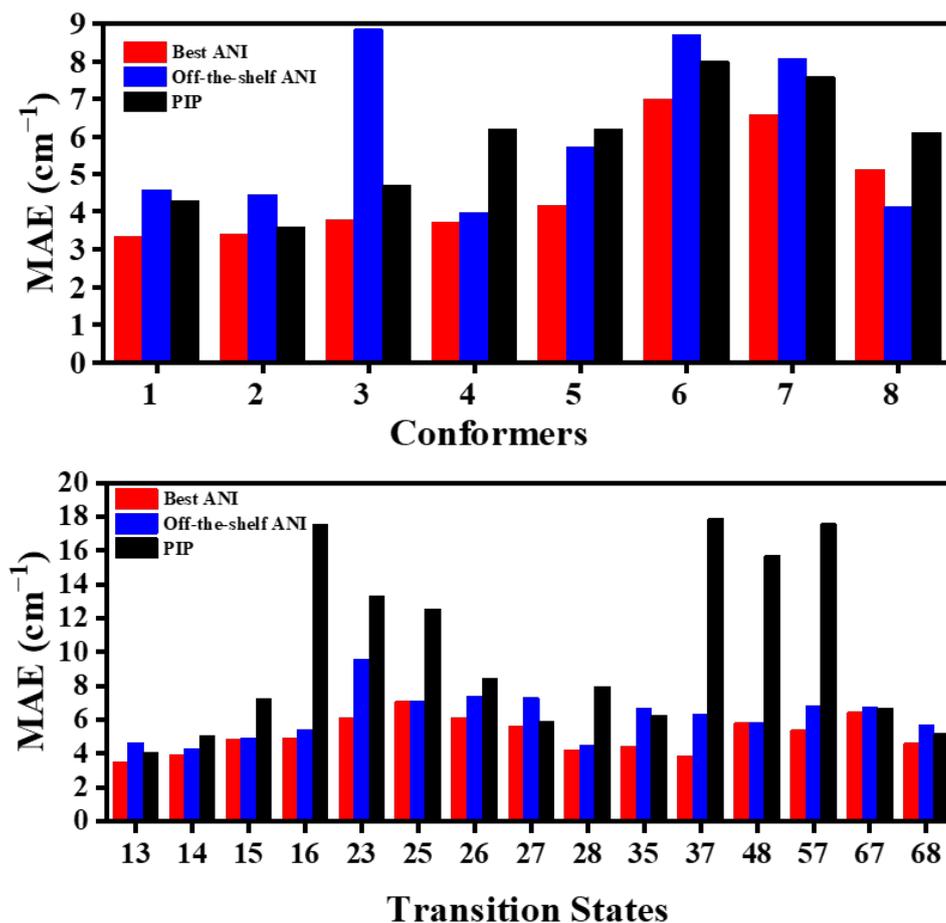
**Table 1.** Statistics of simulation errors and test RMSE in kcal/mol with the off-the-shelf ANI and the ANI trained with the optimized weighting parameter  $\alpha= 2.15$  (energy-weighted ANI). Each set of models is trained five times. Best ANI and PIP are single models, where the best ANI is the best performing energy-weighted ANI (with the lowest MAE in minima).

|           |           | off-the-shelf ANI | energy-weighted ANI | best ANI | PIP <sup>67</sup> |
|-----------|-----------|-------------------|---------------------|----------|-------------------|
| MAE       | in minima | 0.079±0.024       | 0.023±0.0062        | 0.016    | 0.021             |
| MAE       | in TS     | 0.19±0.040        | 0.055±0.0088        | 0.039    | 0.35              |
| test RMSE |           | 0.38±0.077        | 0.41±0.036          | 0.45     | –                 |



**Figure 5.** Energies relative to the global minimum of conformers and saddle points obtained by the best ANI model, PIP potential and the B3LYP/aug-cc-pVDZ.

*Performance for the vibrational levels.* To further ensure the reliability of the best ANI model we obtained, we subsequently performed normal-mode calculations for the eight conformers and the same fifteen saddle points that were previously investigated. For each optimized geometry, the twenty-four harmonic frequencies were calculated using the reference DFT method (B3LYP/aug-cc-pVDZ) and the best ANI model, and MAE was calculated based on the deviation between ANI and DFT results. The MAEs of harmonic frequencies of the minima and transition states are shown in Figure 6. The best ANI model is evidently the most robust MLP with the MAEs of only a few wavenumbers for the energy minima. Impressively, ANI potential can achieve good agreement not only for transition states in the low-energy region but also for those in the high-energy region (16, 23, 25, 37, 57, and 48 in Figure 6), whose energies are higher than 11 kcal/mol. The PIP has an overall similar performance for low-energy TSs but is markedly worse for the higher-energy TS which might be because its training used a different training approach (another weighting function, removed energy gradients for higher-energy regions, and no optimization of the weighting function).



**Figure 6.** Mean absolute errors (MAE, in  $\text{cm}^{-1}$ ) for the harmonic frequencies of energy minima and transition states (their selection and numbering taken from literature<sup>67</sup>) obtained from the ANI potential (red), the off-the-shelf ANI potential (blue) and PIP potential (black) compared to those at B3LYP/aug-cc-pVDZ level.

*Performance for the harmonic zero-point vibrational energies (ZPVE).* Based on the harmonic frequencies we calculated for energy minima and transition states, we further calculated their approximate harmonic zero-point vibrational energies (ZPVEs); the results are shown in

Table 2. For the off-the-shelf ANI model, having a smaller test RMSE cannot guarantee its accuracy in predicting better harmonic ZPVEs. For the transition states, this is especially evident when compared to the performance of the best ANI model, which achieved nearly half of the error. This result highlights that solely relying on the test error may be deceptive.

**Table 2.** Statistics of simulation errors (MAE in harmonic zero-point vibrational energies for minima and transition states in  $\text{cm}^{-1}$ ) and test RMSE in kcal/mol with one of the off-the-shelf ANI models, the best ANI trained with the energy-weighting function and optimized parameter  $a=2.15$ , and PIP.

|               | off-the-shelf ANI     | best ANI              | PIP <sup>67</sup>     |
|---------------|-----------------------|-----------------------|-----------------------|
| MAE in minima | 33.9 $\text{cm}^{-1}$ | 24.8 $\text{cm}^{-1}$ | 27.6 $\text{cm}^{-1}$ |
| MAE in TS     | 42.0 $\text{cm}^{-1}$ | 15.0 $\text{cm}^{-1}$ | 30.4 $\text{cm}^{-1}$ |
| test RMSE     | 0.29 kcal/mol         | 0.45 kcal/mol         | –                     |

*Performance for the DMC-level zero-point vibrational energies (ZPVE).* To give a more comprehensive description of the morphology around the local minima of the MLP, we calculate the zero-point vibrational energies (ZPVE) of each conformer employing rigorous diffusion Monte Carlo (DMC) simulations. The results are presented in

. The agreement between the best ANI model and previously reported state-of-the-art PIP potential is remarkable taking into account that DMC samples have a much larger space than harmonic-oscillator frequencies (we do not have reference B3LYP calculations due to their prohibitive cost for such simulations). Most importantly, ANI can correctly describe the key feature of the double-well pairs on PES (conformers 1 and 4, 2 and 7, 3 and 5, and 6 and 8) as the ZPVEs for the conformers in each pair are very close. The differences between ZPVEs for each conformer in the double-well pair range from  $1 \text{ cm}^{-1}$  to  $14 \text{ cm}^{-1}$  (0.003 to 0.04 kcal/mol). They are somewhat larger than the differences with PIP ( $1\text{--}3 \text{ cm}^{-1}$ ), however, the latter were averaged over several repeated DMC calculations while ANI results are bootstrapped from a single repeat which overall leads to higher standard deviations in ANI's ZPVEs and may partly explain the overall better performance of PIP. The results discussed above ensure the reliability of the MLP constructed by the best ANI as they are 'DMC-certified'. It also demonstrates the high quality of the PES data set.

Unfortunately, we cannot compare ANI or PIP results with the reference calculations as they would require the DMC simulations with the B3LYP/aug-cc-pVDZ method for all eight conformers. One DMC simulation for a single conformer requires 30000 walkers and 55000 steps comprising roughly  $1.6 \cdot 10^9$  potential evaluations with B3LYP. This is the reason why the MLPs (PIP or ANI) are fitted in the first place – to enable expensive calculations such as DMC or *ab initio* MD.

**Table 3.** Zero-point vibrational energies (in  $\text{cm}^{-1}$ ) of each conformer, obtained from DMC by ANI and PIP. Note that PIP results are averages of several simulation repeats while ANI results are bootstrapped from a single repeat leading to higher standard deviations.

| Conformer | Best ANI       | PIP <sup>67</sup> |
|-----------|----------------|-------------------|
| 1         | 17154 $\pm$ 38 | 17151 $\pm$ 5     |
| 4         | 17140 $\pm$ 20 | 17150 $\pm$ 6     |
| 2         | 17442 $\pm$ 40 | 17430 $\pm$ 3     |
| 7         | 17434 $\pm$ 34 | 17430 $\pm$ 4     |
| 3         | 17701 $\pm$ 27 | 17717 $\pm$ 4     |
| 5         | 17696 $\pm$ 34 | 17720 $\pm$ 6     |
| 6         | 18819 $\pm$ 24 | 18843 $\pm$ 6     |
| 8         | 18818 $\pm$ 31 | 18841 $\pm$ 4     |

*How does the simulation-oriented training of ANI compare to the state-of-the-art PIP?*

The focus of this paper was to investigate how to properly train popular MLPs on challenging data sets. We demonstrated that special care is needed for a balanced description of different PES regions which we achieve by introducing the protocol based on the energy-weighted function whose shape is adjusted to produce the lowest error in the computationally cheap simulation. This MLP can then be used for much more computationally expensive calculations such as DMC.

Although comparison with other MLPs is not the focus, it is important to discuss the results in the context of the previous state-of-the-art PIP calculations. In this study, we knew the PIP results and hence had the opportunity to strive to achieve better performance than PIP. However, it is important to emphasize the published PIP results used different weighting functions which were not optimized using the protocol introduced here. Hence, our results with ANI can be viewed as additional, independent validation of the PIP fit on glycine: our best ANI results agree very well with PIP for most of the examined simulation errors (Table

1–

and Figure 5–Figure 6). The only exception is high-energy TSs, which are described better by ANI. These TSs are less important for simulations with DMC and the higher MAE with PIP can be explained by using another weighting function which goes much steeper and faster to zero in the high-energy region and also the neglect of the gradients for high-energy geometries. In the future, PIP will likely adopt a similar simulation-oriented protocol for finding the best weighting function shape.

Finally, given the near-perfect agreement between the DMC zero-point vibrational energies as well as the excellent agreement for the relative energies and harmonic frequencies of the minima it is worth commenting on the speed of evaluation. This is especially relevant for DMC calculations, which are highly compute intensive. Previous comparisons of ANI and PIP for similarly precise fits for ethanol indicate that the PIP PES runs roughly 10–100 times faster than the ANI PES based on the evaluation with the old version of MLatom.<sup>75</sup> This factor is consistent with a more recent assessment of speed and precision for ANI and PIP PESs for aspirin.<sup>76</sup>

For glycine, running 30000 walkers for DMC simulations (with 55000 steps in each walker) with ANI took ca. 60 hours on 1 NVIDIA 3090 GPU, and would take ca. 107 CPU-hours on Intel(R) Xeon(R) Gold 6240 CPU @2.60GHz processors with the new MLatom version with improved performance. The published timings for PIP evaluation for the same number of walkers and steps on a similar Xeon processor (2.4 GHz Intel) took roughly 30 CPU-hours;<sup>67</sup> a factor of ca. 3.5 faster (based on CPU-time comparison). The training speed is also a factor. Training of ANI on 1 NVIDIA 3090 GPU took ca. 48 h, while PIP took only 12.5 hours (3.5 hours to obtain the PIP basis and 9 hours for the fit) on a single CPUs of the above hardware types.

Performance improvements of both ANI and PIP methods are certainly possible. ANI evaluations can be sped up by, e.g., converting the network in the C++ code and recompiling it into a more efficient executable and by reducing the size of the network which is probably unnecessarily large for 70k training points. (The ANI network was designed for training on millions of points but we did not optimize its architecture for this specific application to keep the modifications to the default settings of ANI to the minimum.) Performance enhancements of PIPs are also possible. One that has already been reported for the PIP PES for aspirin<sup>76</sup> is the reduction of the PIP basis with very little loss in precision but a significant speed-up in both training and prediction.

*Conclusions and outlook.* We demonstrate the performance of the off-the-shelf machine learning potentials cannot be solely judged by the test RMSEs but rather should be evaluated on the simulation errors.

To systematically improve the quality of MLPs, we propose simulation-oriented training which utilizes the energy-weighted training scheme where a single hyperparameter is tuned to obtain the better performance of MLPs in simulations. We demonstrate the underlying relationship between various simulations, wherein the precision in predicting relative energies for energy minima influences the accuracy in predicting transition states. Importantly, we can identify a region in the parameter space where both the simulation accuracy for the energy minima and transition states are high and obtain the best MLP model around that region.

The best MLP model's reliability is further confirmed through normal-mode calculations for energy minima and transition states. Additionally, the advanced DMC method is utilized to obtain zero-point vibrational energies (ZPVEs) for energy minima. All the simulation results obtained ensure the reliability of the MLP obtained through a simulation-oriented approach.

To facilitate the usage of the method, we provide a convenient tool implemented in MLatom, an open-source package that can run both locally and on the online XACS cloud computing.<sup>77</sup>

While the focus of this study was how we can properly train the model given already sampled points, a promising direction for future investigation is combining different sampling procedures such as active learning<sup>78-82</sup> (and its often-used variant adaptive sampling<sup>83</sup>) with our simulation-oriented training strategy.

## **Methods**

*ANI model.* A development version of MLatom<sup>72</sup> was used in this work. Its interface to TorchANI<sup>73</sup> (version 2.2.3) was used for generating the AEV descriptor used in ANI models. The parameters of the AEV descriptor were chosen from the ANI-1ccx<sup>84</sup> model. Each ANI model consists of four networks for the H/C/N/O atoms, respectively. While each network implements 3 hidden layers of 160, 128, and 96 CELU-activated neurons.

In the training process, both losses on energy ( $L_E$ ) and gradients ( $L_{\text{grad}}$ ) are calculated with mean squared error weighted by relative energy reference and accounted for in the final loss  $L$ :

$$L = L_E + 0.1 \cdot L_{\text{grad}} = \frac{\sum_i^{N_{\text{tr}}} (w_i (\hat{E}_i - E_i))^2}{N_{\text{tr}}} + 0.1 \cdot \frac{\sum_i^{N_{\text{tr}}} \sum_j^{3N_{\text{at}}} (w_i (\hat{F}_{i,j} - F_{i,j}))^2}{3N_{\text{at}} N_{\text{tr}}}, \quad (\text{SEQ Equation 1 ARABIC 2})$$

where  $w_i$  is the weight calculated with Eq. 1,  $E_i$  is the centered energy of the training point  $i$ , and the  $F_{i,j}$  is the force component of the training point  $i$  (negative energy gradient).

The learning rate was scheduled by the *ReduceLRonPlateau* scheduler provided by PyTorch<sup>85</sup>, monitoring  $L_E$  with patience of 32 epochs and a reduce factor of 0.5 to reduce it from the initial value of  $1 \times 10^{-3}$ . Early stopping triggered by a minimum learning rate of  $1 \times 10^{-5}$  was also used to prevent overfitting.

The tuning process of parameter  $a$  starts from 0.1 and was performed at intervals of 0.1, continuing until it reached 5, with 3 repeats at each point. While the unweighted, off-the-shelf ANI was trained for 5 times.

*Geometry optimization.* Geometry optimization on ANI potentials was performed with the same version of MLatom as above. The optimization algorithm was performed via the interface to Gaussian 16<sup>86</sup>. Optimization settings used in Gaussian were *opt(nomicro,calcall,noeigen)* for energy minima, and *opt(ts,calcfc,noeigen,nomicro)* for transition states. The optimized geometries at B3LYP/ aug-cc-pVDZ reported in the same literature<sup>67</sup> of the glycine data set were used as the initial guesses.

*Diffusion Monte Carlo.* Diffusion Monte Carlo calculations for ANI potentials were performed by *PyVibDMC* module (<https://github.com/rjdirisio/pyvibdmc>, version 1.3.7) interfaced to MLatom. For faster DMC calculations, we implemented in MLatom an efficient batch parallelization of the potential surface evaluations from many walkers. For each energy minimum, the pre-optimized geometry was scaled by a factor of 1.01 then was used as the starting geometry for 30000 random walkers. Every simulation consists of 55000 steps of 1 atomic unit of time each, and the zero-point vibrational energy was bootstrapped by averaging the mean energies from eight segments with 5000 steps after the 15000<sup>th</sup> step, and a

standard deviation of the mean values was calculated from these eight segments. The convergence of each simulation was manually checked.

### **Data availability**

No data was generated in this study.

### **Code availability**

The code for all simulations is publicly available in the open-source package MLatom (<https://github.com/dralgroup/mlatom>). Best, off-the-shelf, and **A** and **B** examples of trained ANI potentials are available at [https://github.com/dralgroup/ani\\_glycine24](https://github.com/dralgroup/ani_glycine24). This repository also describes how to train the models with and without the weighting function and use them with MLatom for the simulations. The PIP PES is available online as supplementary material at <https://pubs.aip.org/aip/jcp/article/153/24/244301/200351/Full-dimensional-ab-initio-potential-energy>.

### **Author contributions**

J.M.B. and P.O.D. initiated the study and outlined the initial study plan. F.G. made implementations and R.W. performed a selection of energy-weighting functions, their parameters, and calculations, both analyzed the calculations and wrote the manuscript under the supervision of P.O.D. P.Z. contributed to the original implementation of the geometry optimizations. A.N. and C.Q. provided additional explanations about the PIP calculations. All authors discussed the project and contributed to the final version of the manuscript.

### **Acknowledgments**

P.O.D. acknowledges funding by the National Natural Science Foundation of China (No. 22003051 and funding via the Outstanding Youth Scholars (Overseas, 2021) project), the Fundamental Research Funds for the Central Universities (No. 20720210092). This project is supported by Science and Technology Projects of Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM) (No: RD2022070103). R.C. thanks Università degli Studi di Milano for funding under grant PSR2022\_DIP\_005\_PI\_RCONT. J.B. acknowledges support from NASA grant (80NSSC22K1167) The authors also thank Adrian Roitberg for the initial discussions on off-the-shelf ANI. R.W. thanks Shengheng Yan for the discussions and his help with technical issues.

## References

- (1) Born, M.; Huang, K., *Dynamical Theory of Crystal Lattices*. The University Press: Oxford, 1954.
- (2) Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, *389* (20), 457–484.
- (3) Murrell, J. N. In *Molecular Potential Energy Functions*, 1985.
- (4) Dawes, R.; Ndengué, S. A., Single- and multireference electronic structure calculations for constructing potential energy surfaces. *Int. Rev. Phys. Chem.* **2016**, *35* (3), 441–478.
- (5) Werner, H.-J., Matrix-Formulated Direct Multiconfiguration Self-Consistent Field and Multiconfiguration Reference Configuration-Interaction Methods. In *Adv. Chem. Phys.*, 1987; pp 1–62.
- (6) Child, M. S., Molecular Potential Energy Functions. *Physics Bulletin* **1985**, *36* (12), 507.
- (7) Schatz, G. C., The analytical representation of electronic potential-energy surfaces. *Rev. Mod. Phys.* **1989**, *61* (3), 669–688.
- (8) Dawes, R.; Quintas-Sánchez, E., The Construction of Ab Initio-Based Potential Energy Surfaces. In *Reviews in Computational Chemistry*, 2018; Vol. 31, pp 199–263.
- (9) Shostak, S. L.; Ebenstein, W. L.; Muentner, J. S., The dipole moment of water. I. Dipole moments and hyperfine properties of H<sub>2</sub>O and HDO in the ground and excited vibrational states. *J. Chem. Phys.* **1991**, *94* (9), 5875–5882.
- (10) Xie, D.; Guo, H.; Peterson, K. A., Accurate ab initio near-equilibrium potential energy and dipole moment functions of the ground electronic state of ozone. *J. Chem. Phys.* **2000**, *112* (19), 8378–8386.
- (11) Harding, L. B.; Klippenstein, S. J.; Lischka, H.; Shepard, R., Comparison of multireference configuration interaction potential energy surfaces for H + O<sub>2</sub> → HO<sub>2</sub>: the effect of internal contraction. *Theor. Chem. Acc.* **2013**, *133* (2), 1429.
- (12) Chen, W.; Hase, W. L.; Schlegel, H. B., Ab initio classical trajectory study of H<sub>2</sub>CO → H<sub>2</sub>+CO dissociation. *Chem. Phys. Lett.* **1994**, *228* (4), 436–442.
- (13) Lourderaj, U.; Park, K.; Hase, W. L., Classical trajectory simulations of post-transition state dynamics. *Int. Rev. Phys. Chem.* **2008**, *27* (3), 361–403.
- (14) Lourderaj, U.; Song, K.; Windus, T. L.; Zhuang, Y.; Hase, W. L., Direct dynamics simulations using Hessian-based predictor-corrector integration algorithms. *J. Chem. Phys.* **2007**, *126* (4), 044105.
- (15) Hu, W.; Lendvay, G.; Maiti, B.; Schatz, G. C., Trajectory Surface Hopping Study of the O(<sup>3</sup>P) + Ethylene Reaction Dynamics. *J. Phys. Chem. A* **2008**, *112* (10), 2093–2103.
- (16) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y., Accelerated Computation of Free Energy Profile at ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14* (11), 5583–5596.
- (17) Collins, M. A., Molecular potential-energy surfaces for chemical reaction dynamics. *Theor. Chem. Acc.* **2002**, *108* (6), 313–324.
- (18) Bowman, J. M.; Czakó, G.; Fu, B., High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* **2011**, *13* (18), 8094–8111.
- (19) Nguyen, T. L.; Li, J.; Dawes, R.; Stanton, J. F.; Guo, H., Accurate Determination of Barrier Height and Kinetics for the F + H<sub>2</sub>O → HF + OH Reaction. *J. Phys. Chem. A* **2013**, *117* (36), 8864–8872.
- (20) Li, J.; Jiang, B.; Song, H.; Ma, J.; Zhao, B.; Dawes, R.; Guo, H., From ab Initio Potential Energy Surfaces to State-Resolved Reactivities: X + H<sub>2</sub>O ↔ HX + OH [X = F, Cl, and O(<sup>3</sup>P)] Reactions. *J. Phys. Chem. A* **2015**, *119* (20), 4667–4687.
- (21) Klippenstein, S. J.; Pande, V. S.; Truhlar, D. G., Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J. Am. Chem. Soc.* **2014**, *136* (2), 528–546.
- (22) Powell, A. D.; Dattani, N. S.; Spada, R. F. K.; Machado, F. B. C.; Lischka, H.; Dawes, R., Investigation of the ozone formation reaction pathway: Comparisons of full configuration interaction quantum Monte Carlo

and fixed-node diffusion Monte Carlo with contracted and uncontracted MRCI. *J. Chem. Phys.* **2017**, *147* (9), 094306.

(23) Agaoglou, M.; García-Garrido, V. J.; Katsanikas, M.; Wiggins, S., The phase space mechanism for selectivity in a symmetric potential energy surface with a post-transition-state bifurcation. *Chem. Phys. Lett.* **2020**, *754*, 137610.

(24) Ziegler, T.; Autschbach, J., Theoretical Methods of Potential Use for Studies of Inorganic Reaction Mechanisms. *Chem. Rev.* **2005**, *105* (6), 2695-2722.

(25) Andrade, M. F. C.; Ko, H.-Y.; Zhang, L.; Car, R.; Selloni, A., Free energy of proton transfer at the water–TiO<sub>2</sub> interface from ab initio deep potential molecular dynamics. *Chem. Sci.* **2020**, *11* (9), 2335-2341.

(26) Bernardi, F.; Olivucci, M.; Robb, M. A., Potential energy surface crossings in organic photochemistry. *Chem. Soc. Rev.* **1996**, *25* (5), 321-328.

(27) Neelamraju, S.; Wales, D. J.; Gosavi, S., Protein energy landscape exploration with structure-based models. *Curr. Opin. Struct. Biol.* **2020**, *64*, 145-151.

(28) Barone, V.; Fusè, M.; Lazzari, F.; Mancini, G., Benchmark Structures and Conformational Landscapes of Amino Acids in the Gas Phase: A Joint Venture of Machine Learning, Quantum Chemistry, and Rotational Spectroscopy. *J. Chem. Theory Comput.* **2023**, *19* (4), 1243-1260.

(29) Szori, M.; Fittschen, C.; Csizmadia, I. G.; Viskolcz, B., Allylic H-Abstraction Mechanism: The Potential Energy Surface of the Reaction of Propene with OH Radical. *J. Chem. Theory Comput.* **2006**, *2* (6), 1575-1586.

(30) Jiang, B.; Li, J.; Guo, H., High-Fidelity Potential Energy Surfaces for Gas-Phase and Gas–Surface Scattering Processes from Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11* (13), 5120-5131.

(31) Tishchenko, O.; Truhlar, D. G.; Ceulemans, A.; Nguyen, M. T., A Unified Perspective on the Hydrogen Atom Transfer and Proton-Coupled Electron Transfer Mechanisms in Terms of Topographic Features of the Ground and Excited Potential Energy Surfaces As Exemplified by the Reaction between Phenol and Radicals. *J. Am. Chem. Soc.* **2008**, *130* (22), 7000-7010.

(32) Zou, Y.; Houk, K. N., Mechanisms and Dynamics of Synthetic and Biosynthetic Formation of Delitschiapyrones: Solvent Control of Ambimodal Periselectivity. *J. Am. Chem. Soc.* **2021**, *143* (30), 11734-11740.

(33) Zeng, J.; Cao, L.; Xu, M.; Zhu, T.; Zhang, J. Z. H., Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat. Comm.* **2020**, *11* (1), 5713.

(34) Wales, D. J., Exploring energy landscapes. *Annu. Rev. Phys. Chem.* **2018**, *69*, 401-425.

(35) Behler, J.; Parrinello, M., Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401.

(36) Bartók, A. P.; Kondor, R.; Csányi, G., On representing chemical environments. *Phys. Rev. B* **2013**, *87* (18), 187115.

(37) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schutt, K. T.; Tkatchenko, A.; Müller, K. R., Machine Learning Force Fields. *Chem. Rev.* **2021**, *121* (16), 10142–10186.

(38) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M., Physics-inspired structural representations for molecules and materials. *Chem. Rev.* **2021**, *121* (16), 9759–9815.

(39) Manzhos, S.; Carrington, T., Jr., Neural Network Potential Energy Surfaces for Small Molecules and Reactions. *Chem. Rev.* **2021**, *121* (16), 10187–10217.

(40) Gokcan, H.; Isayev, O., Learning molecular potentials with neural networks. *WIREs Comput. Mol. Sci.* **2021**, *12* (2), e1564.

(41) Braams, B. J.; Bowman, J. M., Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28* (4), 577–606.

(42) Qu, C.; Yu, Q.; Bowman, J. M., Permutationally Invariant Potential Energy Surfaces. *Annu. Rev. Phys. Chem.* **2018**, *69*, 151–175.

(43) Friederich, P.; Hase, F.; Proppe, J.; Aspuru-Guzik, A., Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20* (6), 750–761.

- (44) Behler, J., Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072.
- (45) Ceriotti, M.; Willatt, M. J.; Csányi, G., Machine Learning of Atomic-Scale Properties Based on Physical Principles. In *Handbook of Materials Modeling*, Andreoni, W.; Yip, S., Eds. Springer: Cham, 2018.
- (46) Pinheiro Jr, M.; Ge, F.; Ferré, N.; Dral, P. O.; Barbatti, M., Choosing the right molecular machine learning potential. *Chem. Sci.* **2021**, *12* (43), 14396–14413.
- (47) Zeng, J.; Cao, L.; Zhu, T., Neural network potentials. In *Quantum Chemistry in the Age of Machine Learning*, Dral, P. O., Ed. Elsevier: Amsterdam, Netherlands, 2023; pp 279–294.
- (48) Hou, Y.-F.; Dral, P. O., Kernel method potentials. In *Quantum Chemistry in the Age of Machine Learning*, Dral, P. O., Ed. Elsevier: Amsterdam, Netherlands, 2023; pp 295–312.
- (49) Tallec, G.; Laurens, G.; Fresse-Colson, O.; Lam, J., Potentials based on linear models. In *Quantum Chemistry in the Age of Machine Learning*, Dral, P. O., Ed. Elsevier: Amsterdam, Netherlands, 2023; pp 253–277.
- (50) Xie, Z.; Bowman, J. M., Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting via Monomial Symmetrization. *J. Chem. Theory Comput.* **2010**, *6* (1), 26–34.
- (51) Houston, P. L.; Qu, C.; Yu, Q.; Conte, R.; Nandi, A.; Li, J. K.; Bowman, J. M., PESPIP: Software to fit complex molecular and many-body potential energy surfaces with permutationally invariant polynomials. *J. Chem. Phys.* **2023**, *158* (4), 044109.
- (52) Novikov, I. S.; Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V., The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn.: Sci. Technol.* **2021**, *2* (2), 025002.
- (53) Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203.
- (54) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G., Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403.
- (55) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W., Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels. *J. Chem. Phys.* **2017**, *146* (24), 244108.
- (56) 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), e1603015.
- (57) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A., sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38–45.
- (58) Hou, Y.-F.; Ge, F.; Dral, P. O., Explicit Learning of Derivatives with the KREG and pKREG Models on the Example of Accurate Representation of Molecular Potential Energy Surfaces. *J. Chem. Theory Comput.* **2023**, *19* (8), 2369–2379.
- (59) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P., Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A* **2020**, *124* (4), 731–745.
- (60) Vassilev-Galindo, V.; Fonseca, G.; Poltavsky, I.; Tkatchenko, A., Challenges for machine learning force fields in reproducing potential energy surfaces of flexible molecules. *J. Chem. Phys.* **2021**, *154* (9), 094119.
- (61) Kovacs, D. P.; Oord, C. V.; Kucera, J.; Allen, A. E. A.; Cole, D. J.; Ortner, C.; Csanyi, G., Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. *J. Chem. Theory Comput.* **2021**, *17* (12), 7696–7711.
- (62) Schwalbe-Koda, D.; Tan, A. R.; Gomez-Bombarelli, R., Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. *Nat. Commun.* **2021**, *12* (1), 5104.
- (63) Miksch, A. M.; Morawietz, T.; Kästner, J.; Urban, A.; Artrith, N., Strategies for the construction of machine-learning potentials for accurate and efficient atomic-scale simulations. *Mach. Learn. Sci. Technol.* **2021**, *2* (3), 031001.

- (64) Zhang, L.; Hou, Y.-F.; Ge, F.; Dral, P. O., Energy-conserving molecular dynamics is not energy conserving. *Phys. Chem. Chem. Phys.* **2023**, *25* (35), 23467–23476.
- (65) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q., The MD17 datasets from the perspective of datasets for gas-phase "small" molecule potentials. *J. Chem. Phys.* **2022**, *156* (24), 240901.
- (66) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G., Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152* (20), 204110.
- (67) Conte, R.; Houston, P. L.; Qu, C.; Li, J.; Bowman, J. M., Full-dimensional, ab initio potential energy surface for glycine with characterization of stationary points and zero-point energy calculations by means of diffusion Monte Carlo and semiclassical dynamics. *J. Chem. Phys.* **2020**, *153* (24), 244301.
- (68) Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203.
- (69) Pozdnyakov, S. N.; Willatt, M. J.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M., Incompleteness of Atomic Structure Representations. *Phys. Rev. Lett.* **2020**, *125* (16), 166001.
- (70) Batatia, I.; Kovács, D. P.; Simm, G. N. C.; Ortner, C.; Csányi, G. In *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*, Advances in Neural Information Processing Systems, <https://openreview.net/forum?id=YPPSngE-ZU>, 2022 (accessed 2024-04-07).
- (71) Orján, E. M.; Nacsa, A. B.; Czakó, G., Conformers of dehydrogenated glycine isomers. *J. Comput. Chem.* **2020**, *41* (22), 2001–2014.
- (72) Dral, P. O.; Ge, F.; Hou, Y.-F.; Zheng, P.; Chen, Y.; Barbatti, M.; Isayev, O.; Wang, C.; Xue, B.-X.; Pinheiro Jr, M.; Su, Y.; Dai, Y.; Chen, Y.; Zhang, S.; Zhang, L.; Ullah, A.; Zhang, Q.; Ou, Y., MLatom 3: A Platform for Machine Learning-Enhanced Computational Chemistry Simulations and Workflows. *J. Chem. Theory Comput.* **2024**, *20* (3), 1193–1213.
- (73) 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. *J. Chem. Inf. Model.* **2020**, *60* (7), 3408–3415.
- (74) Yan, S.; Wang, B.; Lin, H., Tracking the Delocalized Proton in Concerted Proton Transfer in Bulk Water. *J. Chem. Theory Comput.* **2023**, *19* (2), 448–459.
- (75) Houston, P. L.; Qu, C.; Nandi, A.; Conte, R.; Yu, Q.; Bowman, J. M., Permutationally invariant polynomial regression for energies and gradients, using reverse differentiation, achieves orders of magnitude speed-up with high precision compared to other machine learning methods. *J. Chem. Phys.* **2022**, *156* (4), 044120.
- (76) Houston, P. L.; Qu, C.; Yu, Q.; Pandey, P.; Conte, R.; Nandi, A.; Bowman, J. M., No Headache for PIPs: A PIP Potential for Aspirin Outperforms Other Machine-Learned Potentials. *arXiv:2401.09316v1 [physics.chem-ph]* **2024**.
- (77) *Xiamen Atomistic Computing Suite (XACS)*, <https://XACScLOUD.com>, Xiamen University: 2022–2024 (accessed 2024-04-07).
- (78) 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.
- (79) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E., Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148* (24), 241733.
- (80) Botu, V.; Ramprasad, R., Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quantum Chem.* **2015**, *115* (16), 1074–1083.
- (81) Li, Z.; Kermode, J. R.; De Vita, A., Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **2015**, *114* (9), 096405.
- (82) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W., Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Materials* **2019**, *3* (2), 023804.
- (83) Gastegger, M.; Behler, J.; Marquetand, P., Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017**, *8* (10), 6924–6935.

(84) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E., Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10* (1), 2903.

(85) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S., PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Adv. Neural Inf. Process. Syst.*, Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; Garnett, R., Eds. Curran Associates, Inc.: 2019; Vol. 32, pp 8026–8037.

(86) 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 Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.01*, Wallingford, CT, 2016.