Challenges and Opportunities for Machine Learning Potentials in Transition Path Sampling: Alanine Dipeptide and Azobenzene Studies

Nikita Fedik[1,2], Wei Li[3*], Nicholas Lubbers[3], Benjamin Nebgen[1], Sergei Tretiak[1,2,4], Ying Wai Li[3]

1. Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
2. Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
3. Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
4. Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract

The growing interest in machine learning (ML) tools within chemistry and material science stems from their novelty and ability to predict properties almost as accurately as underlying electronic structure calculations or experiments. Transition path sampling (TPS) offers a practical way to explore transition routes between metastable minima such as conformers and isomers on the multidimensional potential energy surface. However, TPS has historically suffered from the computational cost vs. accuracy trade-off between affordable force-field simulations and expensive high-fidelity quantum mechanical calculations. ML interatomic potentials combined with TPS offer a new approach for the exploration of transition pathways at near-quantum mechanical accuracy, while keeping the computational cost comparable to classical force fields. In this study, we employ the HIP-NN-TS and ANI-1x neural network-based ML potentials, both trained on the ANI-1x dataset of 5 million HCNO structures. We first verify the correctness of our approach by applying it to alanine dipeptide and compare the resulting energy surface and transition paths to the literature. Our findings suggest that proposed approach holds promise for conformational searches, as evidenced by the chemical accuracy (errors $\lesssim$ 1 kcal/mol) for thermal molecular dynamics trajectories of alanine dipeptide. While we were able to successfully reconstruct alanine dipeptide's potential landscape using both HIP-NN-TS and ANI-1x frameworks, we observed that ML models with lower accuracy may still locate additional important conformations. We also find that active learning, augmenting the training data by structures taken from TPS trajectories, improved the accuracy by ~30% with small amounts of additional data. Finally, we evaluated a more intricate case, azobenzene, and reiterated that seemingly simple torsions may bear a challenge for ML potentials and limit their applications in TPS. Inability of HIP-NN-TS to correctly describe the energetics of major rotational pathway in azobenzene isomerization highlights deficiencies of the reference method in describing the electronic degrees of freedom. Our study underscores the importance of domain expertise in selecting physically meaningful pathways for benchmarking ML potentials, especially considering the intricacies of electronic structure in chemical dynamics and non-equilibrium processes.

Introduction

The Potential Energy Surface (PES) is a central concept for modeling chemical and physical processes at the atomistic scale[1]. PES describes the energy of a system with respect to nuclear coordinates showing not only the critical minimum energy structures of a system but also highlighting the relevant passages connecting them, which may play a role in various chemical transformations such as chemical reactions or conformational changes. Despite the importance of PES[1] and its maturity with research occurring for more than a century[2], theoretical chemistry still does not have a single recommended tool or method for computing and exploring multidimensional PESs accurately and effortlessly.

Specifically, given the 3N-6 degrees of freedom with respect to the number of atoms N, exhaustive exploration of PES comes at a significant computational expense even at standard quantum mechanical (QM) calculations at Density Functional Theory (DFT) level, limiting dynamic simulations typically to only 10-10^3 of atoms and picosecond timescales. Alternatively, classical interatomic potentials provide many orders of magnitude speedup compared to DFT calculations, achieving linear scaling with a small prefactor. However, classical force-fields (FF)[3] are usually parameterized using a combination of top-down and bottom-up feedback in order to address specific processes or systems of interest[4–7], for instance, protein simulations[6,8]. Moreover, in many cases, potentials are parameterized explicitly for given molecular topologies. This leads to a regime where classical FF are derived to match our intuition, leading to inherent human biases in their predictions.

Over the last decade, machine learning (ML) techniques have been increasingly and successfully used for the construction of interatomic potentials[9–17]. For instance, neural networks, one of the backbones of ML, became capable of automated extraction of structural features from the atomic configuration of a many-body molecular system and establishing a relationship between the features and fundamental properties such as the energies and interatomic forces [18–22]. A sufficiently large (~10^5-10^6) and diverse dataset of atomic configurations with corresponding energies and forces from QM calculations, either DFT or higher-fidelity electronic structure methods, are used in the training of the ML potentials based on one or several target metrics[23–25]. Machine learning interatomic potentials (MLIPs) have been shown to achieve near QM accuracy in predicting energies and forces across diverse atomic configurations, while scaling linearly with the number of atoms with a larger prefactor compared to classical potentials. Examples of such techniques include the high-dimensional neural network potential[26,27,21,28], Gaussian approximation potential (GAP)[29,30], the spectral neighbor analysis potential (SNAP)[9,31], and moment tensor potentials (MTP)[32,33], and end-to-end neural network architectures[34–37].

ML potentials have undoubtedly proven their capability to provide accurate insights into complex atomistic systems near *equilibrium*. Roughly, ML performance on near-equilibrium systems can be assessed by straightforward metrics such as root-mean square error (RMSE) for selected test sets. However, simulations of chemical dynamics and reactivity rely on accurate treatment of *non-equilibrium* structures and phenomena. ML architectures, training procedures and dataset collection techniques are evolving to account for *non-equilibrium* processes[38–42]. This shift also implies that simple averaging

accuracy metrics should not be used alone to gauge the accuracy of the ML models for complex PESs. This concern has been raised several times[43–45] recently, underpinning the importance of evaluating ML potential methods beyond the error in energies/forces alone, such as the radial distribution function[46], dissociation and torsion energy curves[47]. Volker *et al*[48] suggested evaluating the "correctness" of MLIP as the ability to correctly describe minima and maxima on PES for systems of interest.

Presumably, for a dynamic process, one can rather evaluate a RMSE metric or similar on a trajectory (Fig. 1A, left and middle panels) instead of on a random near-equilibrium test set. Still, it would only provide a fractional knowledge of accuracy since most of the chemical transitions proceed through multiple channels (Fig. 1A, right panel). Computing reference QM data for validation on multiple trajectories obtained by molecular dynamics (MD) is rarely feasible. Besides, MD simulations can only adequately sample the stable reactant and product states, however, rare events at long time scales such as high-energy barriers of transition states are not easily accessible with direct MD. To access such states at a higher frequency by using MD simulations, several enhanced sampling techniques have been developed in the past. These include umbrella sampling[49], uncertainty-driven dynamics[50], metadynamics[51–53], steered MD[54], and transition path sampling (TPS)[55–60]. TPS is an attractive method since it samples reactive trajectories in which the rare barrier-crossing events are guaranteed to occur, and it does not require a prior knowledge of the reaction coordinate which can be complex. TPS creates an ensemble of dynamical trajectories between the reactant and product states; its computational cost is primarily limited by the underlying energy/forces calculator of MD simulations.

As a result, path sampling can immediately benefit from improved accuracy compared to classical FFs in MD, and improved scaling compared to QM energy evaluations by leveraging the advances of ML potentials. Additionally, TPS can help detect performance discrepancies between ML potentials for rare and important configurations involved in transition processes, beyond those found using ordinary MD algorithms.

To date TPS has mostly been implemented using classical force fields[61–63], tight-binding methods[64], or pure electronic structure computations[65–67]. Only a few studies have explored the use of machine learning for enhanced sampling[68–73], , MLIPs performance for MD of flexible molecules[74,75] and TPS's performance in condensed phase through the lenses of MLIPs[76,77]. While recent pioneering studies[76,77] clearly demonstrated MLIPs can reproduce ensembles of TPS trajectories and relevant regions of PESs, significant knowledge gaps persist. First, there is limited understanding of whether different MLIPs trained on the same data will produce comparable TPS pathways. Second, previous works generated data specifically for targeted reactions. A key advantage of MLIPs is their potential to recycle existing data with minimal or no modification, yet it remains unclear whether general-purpose datasets such as ANI-1x are suitable for this task.

In this work we evaluate ML potentials beyond structural equilibrium and simple error metrics. In particular, we examine TPS dynamics associated with transitions between metastable states, a general non-equilibrium problem. We do not aim at establishing a completely new metric for evaluating accuracy

of dynamic ML predictions, rather we highlight what factors are important for validating a PES for rare phenomena. Specifically, we evaluate the accuracy and suitability of two ML potentials, ANI-1x[23,24] and HIP-NN-TS[34,78], among other tools in MD and TPS simulations for two seemingly simple test cases, alanine dipeptide (AD, Fig. 1B) and azobenzene (AZ, Fig. 1D). We show that conscious choice of test systems, physics-guided trajectories, and underlying QM level are essential to approach the problem at the right angle; otherwise, traditional RMSE metrics may give a false perception of accuracy for unphysical processes. In addition, we explore possible benefits of active learning for improving MLIP performance for *non-equilibrium* structures generated by TPS.

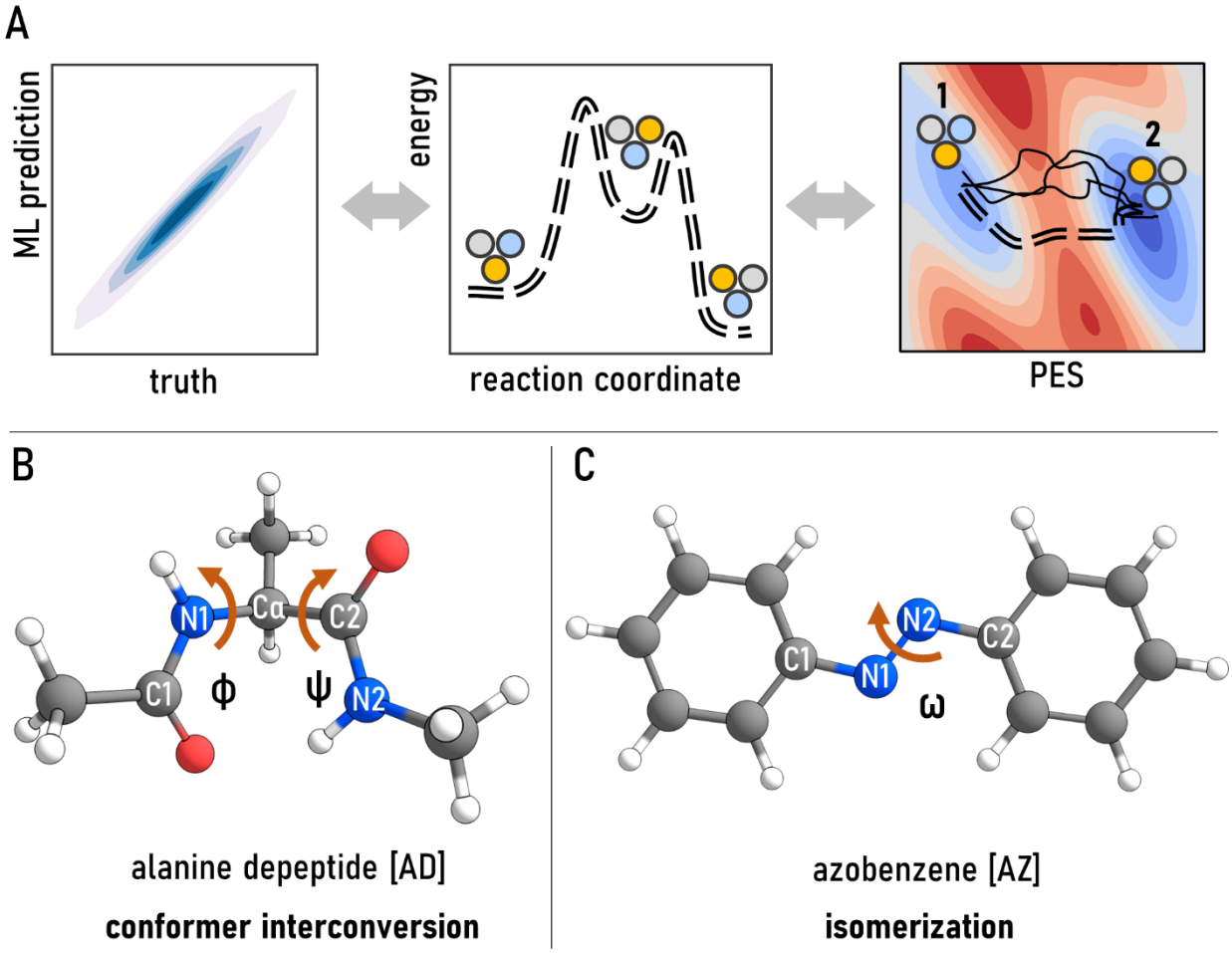

Figure 1. A. Scatter correlation plot (left) is a standard way to evaluate a MLIP emphasizing the accuracy of predictions for a test set. However, the benchmark often spans a narrow set of test structures like a specific trajectory (middle). In reality, transition from state 1 to state 2 is a collection of different paths (right) which renders single-trajectory benchmarks inadequate for assessing accuracy of the MLIP. Qualitatively, discovery of known trajectories (solid and dashed lines) and relevant basins (1 and 2) by ML-TPS can serve as an indicator of good performance. **B**. Alanine dipeptide (AD) structure and relevant dihedral angles ϕ and ψ used to differentiate between states. Figure 1 A, right panel abstractly exemplifies multitude of possible transitions when phase space of both angles is considered. AD is a prototypical example of conformational interconversion. **C**. Azobenzene (AZ) structure

with relevant dihedral angle ω highlighted. Note that in contrast to conformer switching, isomerization requires rearrangement of atoms or chemical bonds.

Molecules for test cases

For the benchmarking of MLIPs using TPS simulations we focused on two torsional transformations, which come in two distinct flavors: conformational interconversion and isomerization. We selected two well-studied cases: alanine dipeptide (AD) (Fig. 1B) and azobenzene (AZ) (Fig. 1C). AD, "a hydrogen atom of biomolecules", which was extensively studied both theoretically[79–83,70] and experimentally[84,85], and is regarded as a simple and insightful model for parametrizing protein backbone torsion potential[86]. The various reported[70,79,87,88] PESs of AD exhibit different energy minima and barrier heights, depending on the effective potential model. Note that stable and metastable states of AD are conformers as they require rotation along single bonds.

Projection of the PES onto ϕ (C1-N1-Cα-C2) and ψ (N1-Cα-C2-N2) backbone dihedral angles suggests many possible transition paths (see Fig. 1A, right panel for abstract example) by which the molecule might interconvert among the basins of the PES . We use TPS to explore transition paths which connect the two main metastable conformers, denoted as C5 and $C7_{eq}$ (Fig. 3), using different ML potentials. This probes the capability of the ML potentials in regions corresponding to very rare events undersampled by ordinary MD simulations.

AZ is a molecule composed of two phenyl rings linked by two nitrogen atoms, presenting a geometrically simple configuration. It is recognized as a molecular switch sensitive to external stimuli[89–91], triggering *cis-trans* isomerization. While AZ initially appeared less intricate compared to AD due to fewer structural degrees of freedom (it has only one relevant dihedral angle (ω)), this system was chosen because of its complex electronic structure[90,92,89,93,94]. Isomerization, in general, is a more intricate process requiring higher activation energy and rearrangement of chemical bonds, in contrast to conformer interconversion seen in AD. Our conscious choice of AZ is to challenge the MLIP-aided TPS on evaluating relevant trajectories highly dependent on electronic degrees of freedom.

MLIP models and data

As a reference electronic structure method for all calculations in this study we used ωB97X/6-31G(d)[95,96] DFT level (see Computational Methods) consistent with ANI-1x[24] dataset. The latter[23,24] contains 5M energies and atomic gradients of equilibrium and nonequilibrium structures (but not transition states) computed using this method and was constructed by active learning to cover diverse regions of chemical space. Furtherrly, we trained the Hierarchically Interacting Particle Neural Network with Tensor Sensitivity, HIP-NN-TS[78], which generalizes interaction layers to pass tensor information between atomic sites. We would like to emphasize that TS suffix is not related to the common abbreviation of transition state, and this variant of HIP-NN is still a general atomistic network, not tuned for locating transition states. Details of HIP-NN architecture are summarized elsewhere[34,37] All HIP-NN-TS potentials were trained using the same protocol summarized in Table 1. HIP-NN and HIP-NN-TS along with

training examples and documentation are publicly available at github.com[97]. A second MLIP architecture used is TorchANI, a Pytorch implementation of classical ANI potential based on modified Behler-Parrinello symmetry functions (Table 1).

Even though in this work we compare results produced by HIP-NN-TS and ANI-1x potentials, we cannot directly attribute performance discrepancies to architecture differences due to the low interpretability. To learn more, we point curious readers to a comprehensive review[37] comparing these architectures and their descriptors.

	HIP-NN-TS	TorchANI
#N trainable parameters	652k	2613k
Descriptor	learnable 20 sensitivity functions	fixed vector of symmetry functions see[98] for parameters
Architecture	26 hidden layers 2 interaction blocks 4 atomic layers with 128 features	128 hidden layers 128 features each
Max epochs	150	unlimited
Loss function	energy + gradient	energy + gradient
Batch size	512	2560
Learning rate	$0.5x10^{-3}$	$1x10^{-3}$
Optimizer	ADAM	ADAM

Table 1. Hyperparameters and architecture details of MLIPs used in this work.

For TPS simulations based on MLIP, we employed OpenPathSampling (OPS)[105,106] software package. Furtherly, we interfaced OPS with the Atomic Simulation Environment (ASE) package[99] to leverage its MD engines.

Results and Discussion

I. Regular MD trajectories

To better understand the advantage of MLIPs over other classical force fields widely employed for MD and TPS, we first compared their energies and forces over regular MD trajectories under NVT condition, where the number of atoms (N) and volume (V) are kept fixed at temperature T = 300K. Initiated from equilibrium, the MD trajectories for AD and AZ span 1.5 ps timescale, employing the HIP-NN-TS potential. 10k snapshots were randomly sampled along each MD trajectory, their energies and forces then are recalculated using different various levels of theory: Amber14[100], Amoeba2013[8], ANI-1x[23,101], and ωB97x/6-31G(d)[95,96] (DFT) - see Fig. 2 and Table 2.

	Alanine Dipeptide (AD)				Azobenzene (AZ)			
	Energy		Forces		Energy		Forces	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DFT/Amoeba2013	1.920	2.431	6.032	8.598	-	-	-	-
DFT/Amber14	1.843	2.343	7.378	11.664	1.768	2.262	7.226	11.579
DFT/ANI-1x	0.668	0.833	1.863	2.758	0.658	0.838	3.329	5.774
DFT/HIP-NN-TS	0.426	0.543	0.799	1.137	0.405	0.524	1.171	1.791
ANI-1x/HIP-NN-TS	0.568	0.715	1.867	2.812	0.642	0.799	2.899	5.247
Amoeba2013/Amber14	1.627	2.076	4.976	7.222	-	-	-	-

Table 2. Prediction accuracy of total energy (kcal/mol) and forces (kcal/mol/Å) for 10k snapshots selected from regular NVT MD at 300K for the AD and AZ test systems. Correlation plots for AD are plotted in Fig. 2. Compared to DFT references, HIP-NN-TS provides best accuracy for both AD and AZ molecules. Section V on AZ elaborates that low error in ML-MD may not always guarantee accuracy in applications. Amoeba2013 does not have parameters for double nitrogen bonds, hence it was not used for AZ calculations.

Both MLIPs achieve remarkable accuracy in energy predictions with RMSEs below 1 kcal/mol compared with reference DFT calculations, under the desired threshold of "chemical accuracy"[102] (Table 2, Fig. 2A). HIP-NN-TS clearly outperforms ANI-1x, which was among the most accurate symmetry-function based architectures of the previous generation. Superior accuracy of HIP-NN-TS is a benefit of the recently incorporated Tensor Sensitivity,[78] which captures higher order many-body information. Both ML models significantly outperform traditional FFs whose errors are at least 300% higher (Table 2, Fig. 2B). Previous benchmarks are in agreement with the observed improvements[78]. We noticed that FF models show a much larger disagreement not only with DFT (Fig. 2B) but also with each other (Table 1 and Fig. 2C right), even though they are both parameterized from experimental data. This verifies that the different fitting methods used in different models give rise to significant discrepancies even on simple test cases[46]. These results seemingly suggest that HIP-NN-TS is the most accurate model for our test cases. However, in the next Section we will discuss that better predictive accuracy may not be the ultimate indicator for TPS performance.

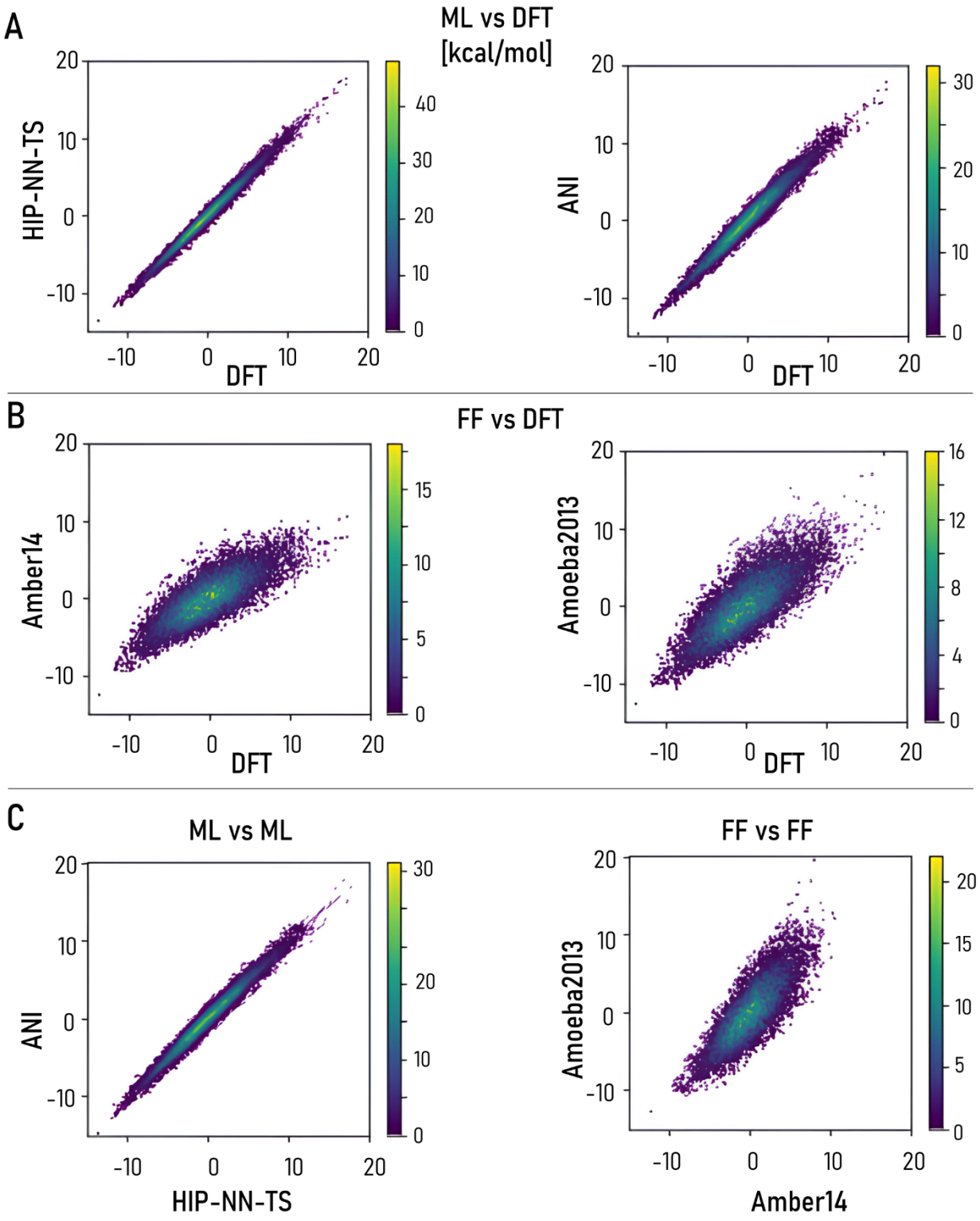

Figure 2. Density correlation plots for energy predictions by various models (MLIPs and FFs) against each other. **A**. MLIPs vs reference DFT. **B**. FFs vs DFT. **C**. MLIPs and FFs against each other. HIP-NN-TS demonstrates better correlation for energy predictions than ANI-1x in line with the lower RMSE (Table 1). However, this is not a guarantee of better recovery of relevant basins by TPS as discussed in the next section. FFs exhibit significant spread with RMSEs

closer to 2 kcal/mol placing them below "chemical accuracy" for test systems. Noticeably, HIP-NN vs. ANI plot shows much better correlation than Amoeba13 vs. Amber14 FF. It underpins that despite discrepancies in MLIPs architectures, neural network-based approaches come to a better agreement with each other and reference theory than empirical counterparts, which should be meticulously tuned for the task at hand.

II. MLIP PES of Alanine Dipeptide

The AD system has a multitude of metastable states, with $C7_{eq}$ recognized as the global minimum[103,61,82]. C5, α_R and $C7_{ax}$ are other important low-lying conformations[104,103,80,82] (Fig. 3A) . The backbone dihedral angles ϕ and ψ act as differentiating order parameters for these conformers (Fig. 3A). Alanine dipeptide PES (Fig. 3B,C) phase space was sampled using brute force approach where structures with fixed incremental combination of ϕ and ψ (-180° to 180°, step size of 5°) were relaxed using FIRE optimizer alongside the HIP-NN-TS and ANI-1x results used as energy calculators.

Both models exhibit noteworthy agreement in predicting the PES, particularly in delineating the locations of the four metastable state basins (C5, $C7_{ax}$, α_R and $C7_{eq}$, see Fig. 3A) and the associated energy barriers. ANI-1x simulations reveal a division of the $C7_{ax}$ state to yield another known state, α_L (Fig. 3A), located at the higher side of the basin. Both the highest energy barrier on the PES (the red regions) and the barrier separating the α_R and $C7_{ax}$ states appear to be higher by around 1.5 kcal/mol, when compared to HIP-NN-TS. Comparatively, ANI-1x's PES (Fig. 3C) closely aligns with those constructed by empirical FFs such as Charm22 in vacuum[105], OPLS/AGBNP with implicit solvation[80], as well as Hartree-Fock and MP2 simulations[82]. Remarkably, ANI-1x demonstrates an upward shift (Fig. 3B) in the valley between C5 and $C7_{eq}$ states, while maintaining close conformity with the basins and energy barrier locations of $C7_{ax}$ and α_L. Notably, the energy minima of the $C7_{ax}$ and α_L states exhibit greater stability in the ANI-1x model. The HIP-NN-TS model's PES aligns well with the latest ϕ-ψ diagram produced the DeepPot-SE model[70], a deep learning potential trained specifically for AD.

While both HIP-NN-TS and ANI-1x models yield comparable PESs (Fig. 3B,C), it is crucial to highlight that ANI-1x identifies an additional major basin, α_L, whereas HIP-NN integrates it with the larger basin of the $C7_{ax}$ conformer. Note that α_L is a well-known minima at MP2[82] and coupled-cluster levels[106]. It is also a true minima at ωb97x DFT level (the reference functional used in ANI-1x dataset) in conjunction with various Pople basis sets[106]. While we emphasize that ANI-1x MLIP discovered α_L as a distinct PES region, we did not confirm whether it is a true local minimum.

Intriguingly, HIP-NN demonstrates quantitative superiority for regular MD trajectories (Fig. 2 and Table 2), prompting a fundamental research question regarding the preference between a model describing more basins with less precision versus a model providing more accurate descriptions overall but differentiating less conformers. This is particularly interesting given that the same dataset was used to train both models. Although our analysis is based on a single case, we posit that a less accurate MLIP, as determined by the RMSE benchmark, may still be practical or even preferable for TPS conformational switching, subject to careful case-specific testing. Additionally, we acknowledge the non-deterministic

nature of ML approaches, advocating for the comparison of ensembles of trained MLIPs for a more comprehensive understanding.

III. Transition Path Sampling of AD: going from C5 to $C7_{eq}$ conformer

Multiple known minima[82,87] for AD signify multitude of possible transitions between conformers. Previous simulations indicate that the time scale for the transition from $C7_{eq}$ to α_R in solution is approximately 0.25-1.0 nanosecond[61,80]. For reference, $\alpha_R \rightarrow C7_{ax}$ interconversion is rarer and spans about 10 nanosecond timeframe[80], while a transition between $C7_{ax}$ and $C7_{eq}$ conformers might require microsecond simulations[107]. Consequently, capturing these transitions sufficiently necessitates either very long trajectories or some form of importance sampling. Given the extensive duration of these transitions and a rather proof-of-concept nature of the current study, we opted to focus on sampling the transition paths from the C5 to the $C7_{eq}$ states, whose interconversion trajectories typically span 3-5 picoseconds[108,109].

To ensure adequate TPS sampling, we defined relatively large regions for these two basins, as depicted by the red rectangles in Fig. 3D,E. Given the consistent basin locations produced by the HIP-NN-TS and ANI-1x models, we adopted the same boundary definitions for the collective variables in the TPS simulations for both models. The stable state C5 region was defined with boundaries $-180 \leq \varphi \leq -150$ and $150 \leq \psi \leq 180$, while the stable state $C7_{eq}$ region was defined with boundaries $-120 \leq \varphi \leq -60$ and $30 \leq \psi \leq 80$[82].

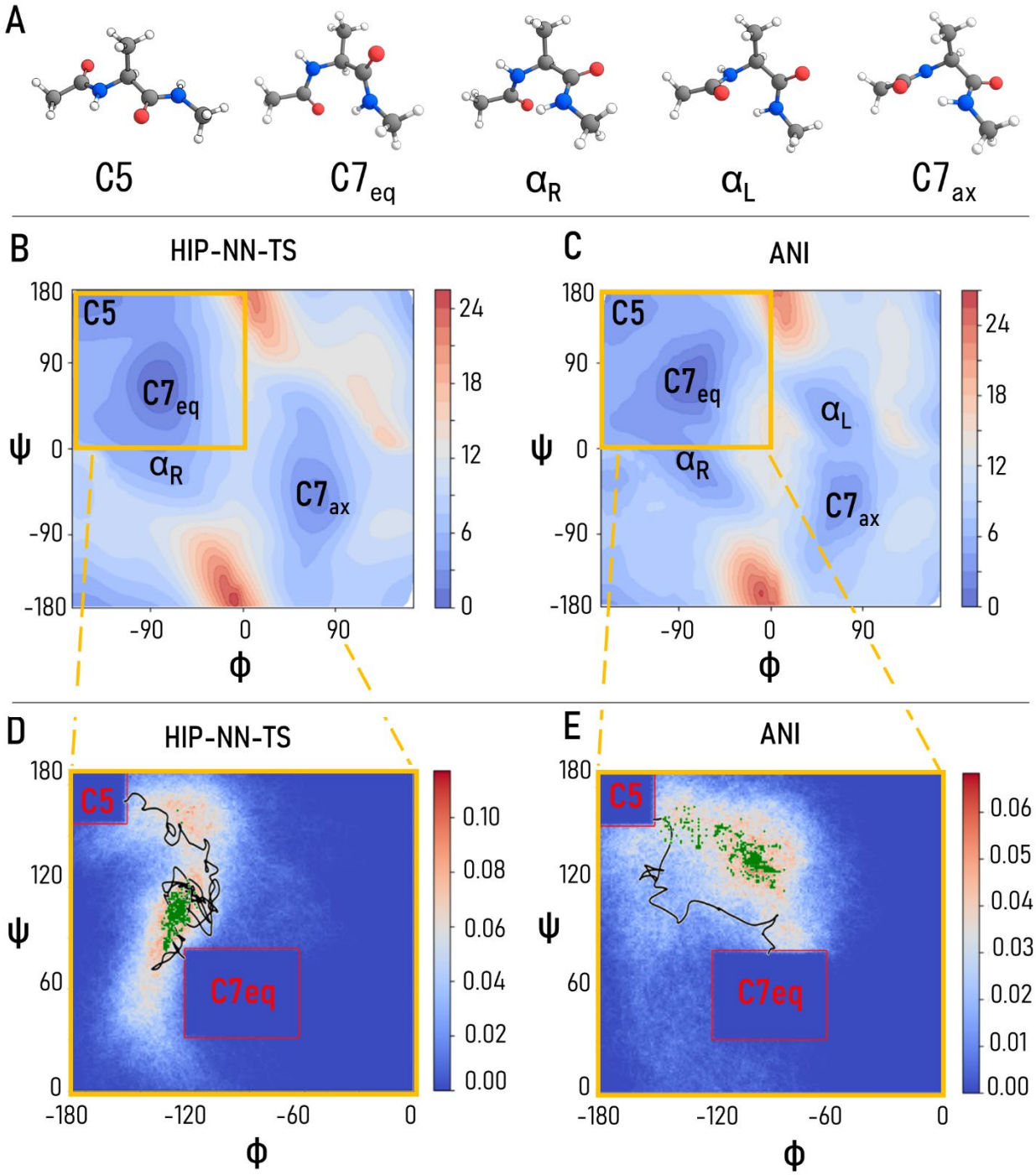

Figure 3. **A**. Major conformations of AD discovered by MLIPs. Structures reproduced from comprehensive theoretical analysis[82]. **B** and **C**. PES's of AD reconstructed by HIP-NN-TS and ANI-1x potentials, respectively. Note that HIP-NN-TS does not differentiate between $C7_{ax}$ and α_L states while ANI-1x treats them as separate basins in line with other works[82]. Noticeably, HIP-NN-TS is more accurate for regular MD trajectories but does not discover the α_L basin on AD PES. It exemplifies that RMSE may not be indicative enough to prefer one potential over another. **D** and **E.** Path density obtained by ML-TPS for the C5 to $C7_{eq}$ transition based on 10k trajectories. Despite both HIP-NN-TS and ANI-1x being trained to the same data, they produce strikingly different TPS ensembles. It emphasizes strong TPS

dependency on the underlying MLIP. Black lines denote the initial MD trajectories at 300K; green dots mark the top 100 most frequently sampled configurations. Red boxes draw boundaries of the $C7_{ax}$ and α_R states.

TPS was initiated from the C5 state to the $C7_{eq}$ state by relaxing the initial trajectory at 500K to 300K using the ASE[99] engine with HIP-NN-TS and ANI-1x potentials. By conducting this step at a relatively high temperature, the trajectories were unhindered by barriers associated with variables other than ϕ and ψ, allowing for the quick generation of a trajectory at 500 K connecting the two stable states. During the subsequent random walk through trajectory space, momenta gradually adjusted to reach ensembles with kinetic energy corresponding to 300 K. The resulting trajectory connecting the two regions was then considered as the initial trajectory for transition path sampling.

In this work, a stochastic "one-way shooting[60,110,111] approach was adopted as implemented in OpenPathSampling[60,110,111], wherein shooting points for generating trial paths were randomly selected from the previously accepted path via so called Monte Carlo move[112,55,60]. A new velocity was assigned randomly for each atom at that point, along with randomly perturbed momenta. The dynamics were integrated forward and backward in time, and the new pathway was accepted if it ended up in C5 and $C7_{eq}$ on both ends, otherwise, it was rejected.

A total of 10k pathways were collected as a transition path ensemble, in which every trajectory was saved after 14 trial moves. In total, 1,438 uncorrelated paths were obtained, with a reasonable acceptance rate for newly generated trajectories (68%), similar to OPS benchmark example results[110,111] using classical FFs. Path density plots for the flexible path length ensemble in the (ϕ, ψ) plane depict the occurrence of pathways for both the HIP-NN-TS model (Fig. 3D) and the ANI-1x model (Fig. 3E). The initial trajectory is represented by black lines. Despite the similar PES descriptions (Fig. 3B,C) with only minor differences, as discussed in previous sections, the transition path ensembles reveal notable distinctions.

With the HIP-NN-TS model, two channels emerge, providing favorable pathways for interconversion trajectories. Conversely, the ANI-1x model reveals a single channel, extending across a broad region between the C5 and $C7_{eq}$ states, as depicted in Fig. 4E. HIP-NN-TS model exhibits similar behavior departing from C5 state; however, it rapidly diverges to the left in the middle, forming a second, more visited channel upon reaching the $C7_{eq}$ state.

Most frequently visited configurations from TPS trajectories produced by HIP-NN-TS (green dots in Fig. 3D), demonstrate clustering in the vicinity of $C7_{eq}$ state, even with the presence of the two-branched channel. In contrast, most frequently visited configurations by the ANI-1x model are dispersed along the single channel (Fig. 3E). Within the HIP-NN-TS model, the most frequently visited sets involve dihedral angles ranges, such as ϕ ≈ (-105, -130), ψ ≈ (80, 115), with a minor clustering around angles ϕ ≈ (-115, -120), ψ ≈ (140, 160), which are in proximity to the $C7_{eq}$ state.

For the ANI-1x model, most visited states occur with dihedral angle pairs ϕ, ψ ≈ (-130, -70), ψ ≈ (115, 145). This observation aligns with the PES plot discussed earlier, indicating similar locations of the C5 and $C7_{eq}$ states for both models. However, these states are separated by a more uniform barrier in the ANI-1x model. In the HIP-NN model, however, the lower energy barrier is more confined, prompting the molecule transition to follow the narrower path and veering toward the left side of the $C7_{eq}$ state, as illustrated in the density plot (Fig. 3D).

These observations support an interesting conclusion - two MLIPs trained to the same data reveal very distinct TPS pathways. Their PESs (Fig. 3B,C) differ only slightly, but the dynamic simulations revealed sufficiently different channels (Fig. 3D,E). It suggests that TPS might be very sensitive to the choice of a particular MLIP even when the same training data is used. Furthermore, it corroborates our previous assertion that similarity in accuracy in static benchmarks based purely on RMSE is not indicative of the similarity in TPS simulations or in the transition trajectories.

IV. Probing Active Learning for TPS

Even though we find RMSE is not an all-in-one metric for comparing two MLIPs in the context of TPS, in general, RMSE can be used to systematically track quantitative improvement of a single model. On average, more accurate MLIP should reproduce the underlying QM PES with greater accuracy. Beware that qualitatively it does not always reflect the discovery of relevant minima and conformers (see Section II and a recent tutorial[48]). Importantly, that RMSE evaluated over equilibrium data is insufficient to estimate performance for non-equilibrium transition pertaining to TPS.

Keeping in mind that most MLIPs are trained to equilibrium structures, here we propose a simple probe whether we can systematically improve the accuracy of predictions for non-equilibrium TPS-relevant structures leveraging simplified active learning[23,113,114,39,115] (AL) approach. We augment the dataset with structures directly taken from the TPS trajectories. The new HIP-NN-TS-AL model was then tested on the exact trajectory frames which do not appear in the train set. Though not a true AL automated loop, this simple pipeline provides insight into whether collecting structures from transition trajectories will be beneficial for further ML-TPS improvement. As a proof of principle, we carried out a single iteration of AL by manually injecting 12k such structures from the most visited regions with a 10800:1200 train:test split, and retraining HIP-NN-TS on the combined dataset (original ANI-1x + 10800 "TPS structures", Fig. 4A) to yield HIP-NN-TS-AL.

HIP-NN-TS-AL demonstrates ~35% accuracy improvement when tested on the subset of 1200 TPS structures (Fig. 4A,B). To quantify non-deterministic errors, we trained three HIP-NN–TS-AL models with different random seeds for weight initialization following the same protocol as for the original HIP-NN-TS. Improvements of ~38% energy RMSE and ~24% force MAE are observed in all three models (Fig. 4C). This is a remarkable improvement given that only 10k new configurations (0.2% increase in total data) were inserted into the original 5M data points of ANI-1x. We conclude that AL holds great promise in improving MLIP performance for *non-equilibrium* structures occurring during TPS conformational search. It is

possible that further automated and systematic model refinement covering previously unexplored *non-equilibrium* regions can make ML-TPS a viable competitor to widely used electronic structure calculations and FFs for TPS.

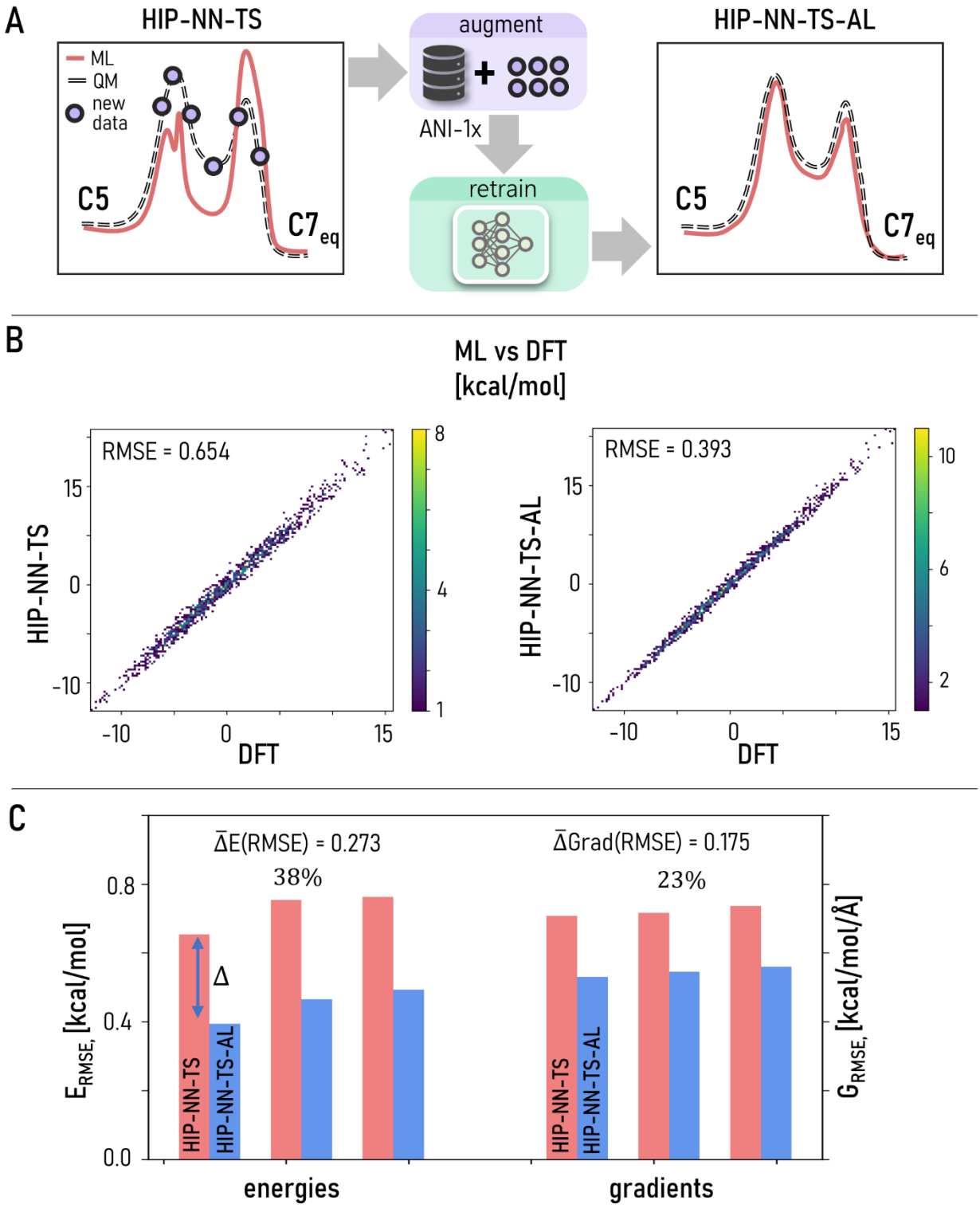

Figure 4. **A**. Conceptually, AL targets structures (violet dots) from undersampled regions. By augmenting training data with new relevant structures and retraining the MLIP, we can improve inference accuracy. In particular, we explore AL for improving prediction of energy and forces of structures encountered during $C7_{eq}$ to α_R transition. **B**. [left] Accuracy of HIP-NN-TS on the 1200 TPS test structures. [right] Accuracy of HIP-NN-TS after one iteration of AL (HIP-NN-TS-AL) on the same test set. 0.2% data augmentation leads to ~30% accuracy improvement, holding promises for AL-refinement of MLIPs for TPS. **C**. Accuracy of HIP-NN-TS and HIP-NN-TS-AL for 1200 test structures averaged across 3 different random number seeds to account for non-deterministic factors. AL-models show unanimous improvement for energy and gradient inference compared to original HIP-NN-TS models trained to ANI-1x data only.

V. Azobenzene: a cautionary tale

In this subsection we emphasize a common wisdom from the ML community: real-life ML model accuracy should be evaluated in the specific application domain[45,48]. In our particular case, MLIP for TPS exploration should correctly assess major pathways and identify the lowest-energy one, which is a kinetically dominant transition channel for the system disjoint from the training set. Analysis of thermal MD simulations (Table 2) demonstrates that HIP-NN-TS predictions for AZ should be accurate within ~0.5 kcal/mol, suggesting that reliable information about the system can be derived from MLIP simulations. The AZ molecule is a geometrically simple arrangement with two benzene rings bridged by two N atoms (Fig. 1C). Both conjugated rings should exhibit aromaticity to some extent, preventing their distortion during MD. Therefore, AZ is an excellent test case for TPS simulations having only one easy-activating degree of freedom associated with the movement around the N-N bridge. Ultimately, it leads to isomerization from *cis*- to *trans*-states or vice versa. However, the complexity lies in the finer details.

The controllable isomerization of azobenzene (AZ) has enabled a wide array of emerging applications[116,117], from light-responsive supramolecular self-assembly to synthetic vision restoration. In-depth studies[118] demonstrated that the interplay between isomerization mechanisms depend on numerous factors including solvent, steric hindrance of functional group, heat and pressure among others. Previous studies have already established multiple isomerization mechanisms among which two are widely recognized - inversion and rotation[119–122] (Fig. 5). While the community initially believed that inversion is the more energetically-favorable route of thermal isomerization, further studies found that the rotational mechanism is prefered[93,121]. Capturing two of these major pathways is the key to describing AZ transformations.

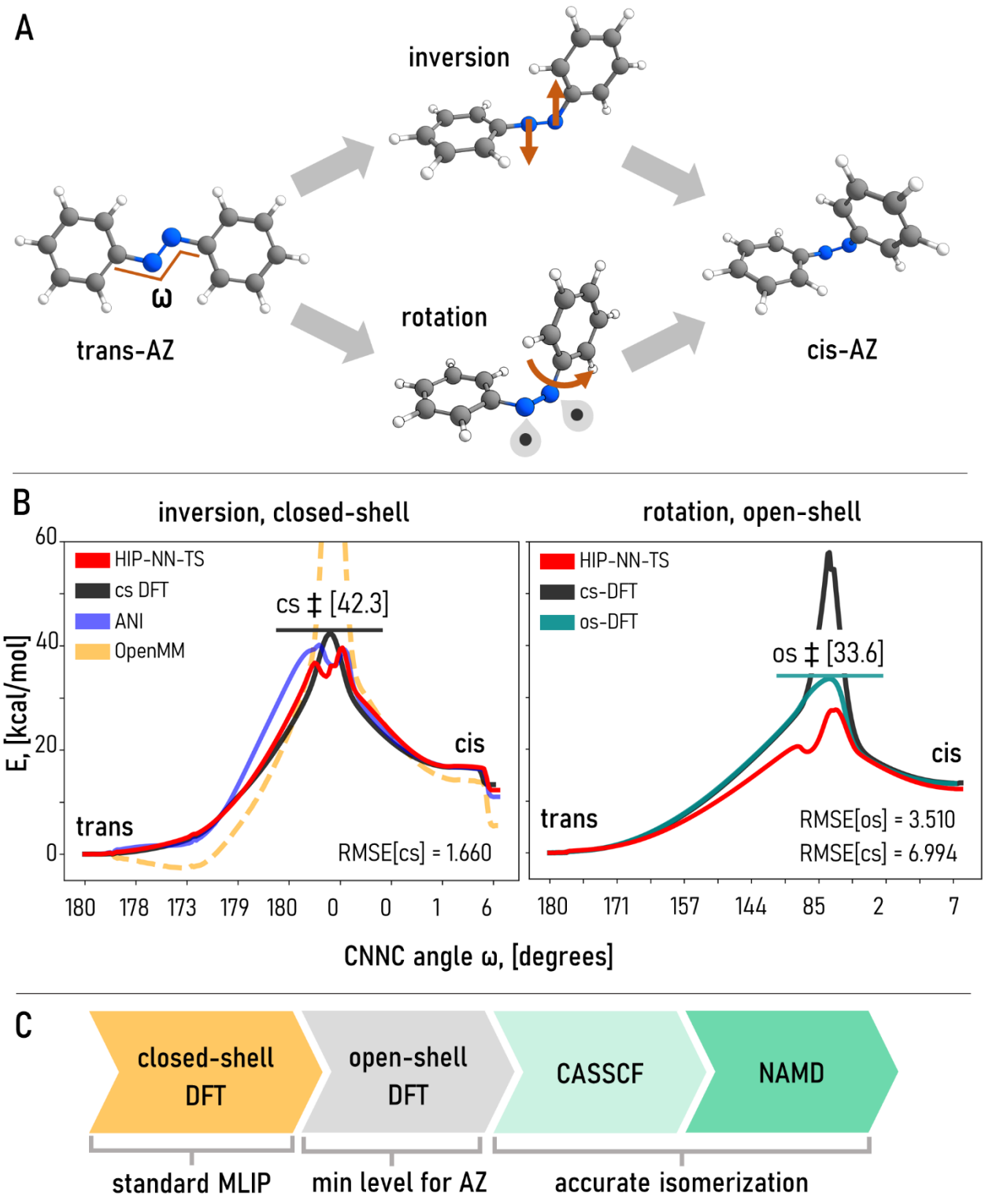

Figure 5. Isomerization of azobenzene. **A**. Two pathways of thermal isomerization. Inversion involves the concerted bending of two bridging N atoms without significant changes in ω dihedral angle. The rotation mechanism is based

on the second ring rotation relative to the first one with gradual changes in ω angle. Rotation requires double bond breaking and leads to the configuration with unpaired electrons on two N atoms. Red arrows indicate the movement of N atoms in imaginary mode. **B**. [left] Energy profile of the inversion mechanism of *trans-cis* AZ isomerization based on closed-shell DFT (cs-DFT) calculations. The transition state (‡) of inversion is 42.3 kcal/mol higher than trans-AZ. HIP-NN-TS yields the best agreement (RMSE=1.660 kcal/mol) compared to cs-DFT while ANI-1x provides a reasonable approximation. The OpenMM barrier exceeds 120 kcal/mol. [Right]. The rotational mechanism is based on unrestricted broken-symmetry open-shell -DFT (os-DFT) calculations. At this DFT level, this path (33.6 kcal/mol) is energetically preferred over inversion. The HIP-NN-TS RMSE for os-DFT trajectory is 3.510 kcal/mol, noticeably higher than for the inversion path. cs-DFT here was calculated on top of the os-trajectory since TS could not be located at cs-PES. Overall, cs-DFT gives an unrealistically elevated barrier for the rotational path. **C.** Hierarchy of theoretical levels for exploring AZ isomerization: cs-DFT (common for MLIP training) cannot describe a preferable rotational path through bond breaking while os-DFT and CASSCF can capture rotation. AZ isomerization involves ground state crossing with excited states, described by non-adiabatic molecular dynamics (NAMD). This highlights the importance of being aware of possible electronic structure changes in test cases.

We calculated the inversion isomerization pathway using a closed-shell DFT (cs-DFT) approach (ωB97X/6-31G(d) model chemistry) starting with a hand-crafted and optimized transition state (see Computational Methods for details and Fig. 5B, left). We recomputed energies for all 179 points on a smooth isomerization curve and found that HIP-NN-TS was the most accurate model in line with previous observations. Even though the energy in the vicinity of TS dips unphysically for both HIP-NN-TS and ANI-1x (possibly due to absence of transition states in training data), overall, HIP-NN-TS curve is fairly accurate (RMSE = 1.660 kcal/mol), especially for the initial and final states. ANI-1x is slightly worse in terms of accuracy, while the empirical OpenMM force-field exhibits an energy spike for the barrier exceeding 120 kcal/mol, beyond any practicality (Fig. 5B, left). These results seem to support the conclusion that HIP-NN-TS, or any other MLIP trained on the ANI-1x database, could be a viable tool for exploring isomerization in azobenzene or similar double-bonded switches.

An alternative and important route of thermal isomerization proceeds through a bond-breaking mechanism[93] because the bridging nitrogen atoms are connected by a double bond that blocks any low-barrier rotation, in contrast to the freely-rotating single bonds in AD conformers. Roughly, such rotational isomerization proceeds through a biradical state when uncoupled electrons occupy spatially different orbitals on different N atoms. Such biradicals (often referred to as open-shell singlets[123,124]) are not properly described by cs-DFT which inherently assumes double occupancy for each orbital. Unless expensive high-fidelity calculations with careful manual selection of active orbitals are conducted, e.g. the complete active space self-consistent field method **(**CASSCF)[90,125] (Fig. 5C), the more affordable approximation would be to manually break the symmetry of orbitals in unrestricted open-shell DFT (os-DFT) calculations[93,126]. Despite the *ad-hoc* nature of such a treatment, this solution typically approximates states with broken bonds quite reasonably[127,128].

To get an insight into a lower-energy rotational pathway, we recalculated the trajectory from a newly estimated TS optimized with os-DFT. The os-DFT TS structure is very different from its cs-DFT counterpart. It exhibits an imaginary mode of rotation leading to bond breaking, in contrast to the

inversion mechanism for cs-DFT. Rotation is accompanied by substantial changes in dihedral angle ω, in agreement with previous studies[122,64,119]. Most importantly, the os-DFT pathway has a lower barrier by almost 9 kcal/mol, rendering it far more energetically favorable. Simply speaking, isomerization will much more likely proceed through an open-shell rotational trajectory under thermal stimuli, at least at the given DFT level[93]. Therefore, the closed-shell inversion trajectory for which the MLIPs yield great accuracy, is an incomplete test case giving false perception on reliability of ML prediction that is based on inaccurate reference QM methodology.

Once more, we recalculated the new open-shell rotational pathway (Fig. 5B right) with HIP-NN-TS which yielded a significantly worse match with an unphysical energy dip near the TS and higher RMSE of 3.510 kcal/mol. Though noticeably worse, ~3.5 kcal discrepancy might be deemed acceptable for some reactive applications, leading us to believe that MLIP still might be a fair choice. However, we should keep in mind that MLIP was trained solely on closed-shell data, which completely ignores structures relevant to the rotational pathway. Hence, aside from the quantitative discrepancy of predictions outlined above, it is very unlikely for MLIP to discover open-shell pathways which proceed through very different transitions, both geometrically and electronically if training data are lacking physically justifiable and reliable QM reference results. To prove this point, we recalculated the os-DFT trajectory by single-point cs-DFT. Closed-shell treatment incorrectly describes the barrier energetics with the TS reaching 50 kcal/mol above the os-DFT maxima (Fig. 5B right). Moreover, cs-DFT reoptimization starting from an os-TS geometry with broken symmetry does not lead to a saddle point with one imaginary frequency at all, ignoring the most favorable path altogether.

To reiterate, cs-DFT cannot discover the favorable rotation mechanism. In the context of dynamic simulations carried out by MLIP trained to cs-DFT data, it would likely prompt the simulations to avoid common isomerization channels. This limitation renders TPS exploration incomplete or simply incorrect (Fig. 5C). While beyond the scope of the current study, we believe it would be essential in the future to include broken-symmetry and open-shell paths in the dataset and probe the TPS performance for azobenzene as a model system (Fig. 5C).

An unresolved issue is that even an open-shell singlet is a poor description for the AZ transition state, which is often photo-activated and isomerized by state crossing[90–92,119,121,125]. Such effects are usually captured by non-adiabatic dynamics[129,90] and are beyond the scope of this article (Fig. 5C). Therefore, non-excited dynamics, even with an open-shell treatment, is already a rough approximation to more realistic light-induced molecular switching (Fig. 5C). Unless data sets account for these challenges, production-quality ML-TPS will not be achievable for paths involving formation and cleavage of bonds.

Conclusions

In this article we explore the usability and pitfalls of machine learning interatomic potentials for transition path sampling. In general, we conclude that MLIPs can be promising tools for TPS explorations of PESs as both HIP-NN-TS and ANI-1x were able to reproduce AD PES with sufficient details. As expected,

both models performed significantly more accurately than classical FFs on regular MD trajectories for AD and AZ. However, our results indicate that significant caution should be exercised in simulating isomerization when proper description and energetics of bond breaking and formation is required.

With that in mind, we summarize our research lessons learned from using MLIPs for TPS:

1. **Lower error metrics do not guarantee correct discovery of basins on PESs**

   Although HIP-NN-TS demonstrated superior accuracy in predicting energy and forces compared to ANI-1x for the test molecules, an interesting observation emerged during PES exploration. Seemingly less accurate, ANI-1x revealed an additional basin, α_L, adjacent to another major basin. This counterintuitive finding underscores the limitation of relying solely on root mean square error (RMSE) calculated for random samples as the sole indicator of the correctness of dynamic explorations like TPS.

2. **Different MLIP architectures trained on the same data may produce different trajectories in TPS**

   Despite the same training data, HIP-NN-TS and ANI-1x revealed different channels for the transition from C5 to $C7_{eq}$ basin. It is particularly interesting given the relatively close agreement of the two PESs, especially near C5 and $C7_{eq}$ basins. Therefore, we conclude that different MLIPs with similar reasonable RMSE-based accuracy, can produce sufficiently different TPS trajectories.

3. **Active learning can substantially improve MLIP accuracy on TPS trajectories**

   Despite the potential vagueness of RMSE as a definitive indicator for comparing different MLIPs in TPS, a lower RMSE still signifies a closer alignment of the predicted PES with the reference one. By employing a manual data augmentation procedure, we emulated a single iteration of an active learning loop. The original ANI-1x (5M structures) dataset was augmented by 10800 structures selected from TPS trajectories. With a minute data augmentation of ~0.2%, a substantial improvement of ~ 30% in energy and gradient predictions was achieved for TPS-relevant data points, underscoring the tangible benefits of active learning. New class of foundational ML models, a very broad non-system specific potentials[130,131], opens exciting venues for ML-TPS with minor active learning fine-tuning.

4. **Choice of test cases and domain of applicability of MLIPs should account for electronic structure features**

   Illustrated with the azobenzene example, a system comprising two benzene rings bridged by two nitrogen atoms, we emphasize that seemingly simple transformations like torsion and isomerization may demand more sophisticated theoretical approaches than conventional low-cost closed-shell DFT. Failure to adequately describe possible bond breaking and formation effects in electronic structure methods could lead to the generation of incorrect or artificially high-energy trajectories, which are easy to overlook without prior knowledge. Therefore, we caution against reliance on incomplete or unphysical transformations for MLIP evaluation. For instance, MLIP

performance is often evaluated on torsional profiles[132], which is only valid for conformer-like interconversion around single bonds unless MLIP was trained on reactive data. Therefore, *a priori* intuition of bond type is advised. Altogether, it underpins that all test cases for ML-TPS and associated training data should be carefully inspected for electronic degrees of freedom. When considering AD and AZ molecules through the lens of MLIPs and TPS, we conjecture that general databases such as ANI-1x might be adequately representative to describe conformational interconversion, but will be deficient for isomerization involving chemical bonding rearrangement. The need for a community effort in generating high-accuracy reference data for bonding breaking processes at CASSCF[133] or truncated configuration interaction (CI)[134] levels is apparent. So far, such endeavors were limited to small molecules only[133,134]. Advent of reactivity-focused data such as Transition-1x[41] and RGD1[135] and architectures such as EquiReact[136] can shape the future of ML-TPS when accurate description of bond-changing events is required.

Beyond data, lessons learned urge for robust metrics quantifying how well minima and saddle points are recovered by MLIPs. Sophisticated training and testing of MLIPs for TPS[76,77] and dynamics simulations will continue relying on human expertise in computational chemistry underscoring an importance of physics-informed ML frameworks[14,137,48].

Data Availability

MLIPs, datasets, ASE and OpenPathSampling package used in this study are publicly available and free of charge. The supporting data, including comprehensive compilations of links to all utilized tools, pretrained HIP-NN-TS models, an additional 12k alanine dipeptide configurations from section IV, and azobenzene isomerization pathways, are accessible at github.com/nikitafedik/ml_tps_si.

Computational Methods

All DFT calculations in the article were done at ωB97X/6-31G(d)[95,96] level employing tight SCF convergence for consistency with ANI-1x[23,24] database. All electronic structure calculations were performed in the Gaussian 16 package[138]. For AD, closed-shell DFT trajectory (Fig 5B), transition state (TS) was computed based on manual geometry guess through Berny algorithm[139] with suppressed curvature tes (opt=TS, calcfc, noeigentest). Smooth reaction coordinate in the vicinity of TS was built as reverse and forward internal reaction coordinate (IRC) integration without recorrection step, default step size of 0.1 Bohr and recomputation of Hessian every 10 steps with total trajectory length up to 300 steps. Left and right shoulders of trajectory, descents to *cis*- and *trans*-isomers, respectively, were determined by further optimization of final IRC states with Hessian calculations at each step for smoothness. AZ open-shell trajectory was built following an almost identical protocol employing unrestricted broken-symmetry DFT through default orbital shift (VShift) by 0.1 mHartree and HOMO-LUMO mixing. Built-in wave function analysis confirmed TS instability (stable=opt).

References

(1) The Concept of the Potential Energy Surface. In *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Lewars, E., Ed.; Springer US: Boston, MA, 2003; pp 9–41. https://doi.org/10.1007/0-306-48391-2_2.

(2) Marcelin, M. R. Contribution à l'étude de la cinétique physico-chimique. *Ann. Phys.* **1915**, *9* (3), 120–231. https://doi.org/10.1051/anphys/191509030120.

(3) MacKerell, A. D. Empirical Force Fields. In *Computational Methods for Protein Structure Prediction and Modeling: Volume 1: Basic Characterization*; Xu, Y., Xu, D., Liang, J., Eds.; BIOLOGICAL AND MEDICAL PHYSICS BIOMEDICAL ENGINEERING; Springer: New York, NY, 2007; pp 45–69. https://doi.org/10.1007/978-0-387-68372-0_2.

(4) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17* (5–6), 490–519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

(5) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65* (3), 712–725. https://doi.org/10.1002/prot.21123.

(6) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121* (16), 4023–4039. https://doi.org/10.1021/acs.jpcb.7b02320.

(7) Wildman, J.; Repiščák, P.; Paterson, M. J.; Galbraith, I. General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution. *J. Chem. Theory Comput.* **2016**, *12* (8), 3813–3824. https://doi.org/10.1021/acs.jctc.5b01195.

(8) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063. https://doi.org/10.1021/ct4003702.

(9) Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials. *Journal of Computational Physics* **2015**, *285*, 316–330. https://doi.org/10.1016/j.jcp.2014.12.018.

(10) Deringer, V. L.; Csányi, G. Machine Learning Based Interatomic Potential for Amorphous Carbon. *Phys. Rev. B* **2017**, *95* (9), 094203. https://doi.org/10.1103/PhysRevB.95.094203.

(11) 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. https://doi.org/10.1103/PhysRevMaterials.3.023804.

(12) 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. https://doi.org/10.1021/acs.jpca.9b08723.

(13) Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152* (5), 050902. https://doi.org/10.1063/1.5126336.

(14) Anstine, D. M.; Isayev, O. Machine Learning Interatomic Potentials and Long-Range Physics. *J. Phys. Chem. A* **2023**, *127* (11), 2417–2431. https://doi.org/10.1021/acs.jpca.2c06778.

(15) Ko, T. W.; Ong, S. P. Recent Advances and Outstanding Challenges for Machine Learning

Interatomic Potentials. *Nat Comput Sci* **2023**, *3* (12), 998–1000. https://doi.org/10.1038/s43588-023-00561-9.

(16) Liu, Y.; He, X.; Mo, Y. Discrepancies and Error Evaluation Metrics for Machine Learning Interatomic Potentials. *npj Comput Mater* **2023**, *9* (1), 1–13. https://doi.org/10.1038/s41524-023-01123-3.

(17) Mishin, Y. Machine-Learning Interatomic Potentials for Materials Science. *Acta Materialia* **2021**, *214*, 116980. https://doi.org/10.1016/j.actamat.2021.116980.

(18) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145* (17), 170901. https://doi.org/10.1063/1.4966192.

(19) Artrith, N.; Urban, A. An Implementation of Artificial Neural-Network Potentials for Atomistic Materials Simulations: Performance for TiO2. *Computational Materials Science* **2016**, *114*, 135–150. https://doi.org/10.1016/j.commatsci.2015.11.047.

(20) Kulichenko, M.; Smith, J. S.; Nebgen, B.; Li, Y. W.; Fedik, N.; Boldyrev, A. I.; Lubbers, N.; Barros, K.; Tretiak, S. The Rise of Neural Networks for Materials and Chemical Dynamics. *J. Phys. Chem. Lett.* **2021**, *12* (26), 6227–6243. https://doi.org/10.1021/acs.jpclett.1c01357.

(21) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072. https://doi.org/10.1021/acs.chemrev.0c00868.

(22) Tokita, A. M.; Behler, J. How to Train a Neural Network Potential. *The Journal of Chemical Physics* **2023**, *159* (12), 121501. https://doi.org/10.1063/5.0160326.

(23) 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. https://doi.org/10.1063/1.5023802.

(24) 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. *Scientific Data* **2020**, *7* (1), 134. https://doi.org/10.1038/s41597-020-0473-z.

(25) Smith, J. S.; Nebgen, B.; Mathew, N.; Chen, J.; Lubbers, N.; Burakovsky, L.; Tretiak, S.; Nam, H. A.; Germann, T.; Fensin, S.; Barros, K. Automated Discovery of a Robust Interatomic Potential for Aluminum. *Nat Commun* **2021**, *12* (1), 1257. https://doi.org/10.1038/s41467-021-21376-0.

(26) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401.

(27) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134* (7), 074106. https://doi.org/10.1063/1.3553717.

(28) Eckhoff, M.; Behler, J. *High-Dimensional Neural Network Potentials for Magnetic Systems Using Spin-Dependent Atom-Centered Symmetry Functions*; 2021.

(29) 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. https://doi.org/10.1103/PhysRevLett.104.136403.

(30) Klawohn, S.; Darby, J. P.; Kermode, J. R.; Csányi, G.; Caro, M. A.; Bartók, A. P. Gaussian Approximation Potentials: Theory, Software Implementation and Application Examples. *The Journal of Chemical Physics* **2023**, *159* (17), 174108. https://doi.org/10.1063/5.0160898.

(31) Wood, M. A.; Thompson, A. P. Extending the Accuracy of the SNAP Interatomic Potential Form. *The Journal of Chemical Physics* **2018**, *148* (24), 241721. https://doi.org/10.1063/1.5017641.

(32) 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. https://doi.org/10.1088/2632-2153/abc9fe.

(33) Podryabinkin, E.; Garifullin, K.; Shapeev, A.; Novikov, I. MLIP-3: Active Learning on Atomic Environments with Moment Tensor Potentials. *The Journal of Chemical Physics* **2023**, *159* (8), 084112. https://doi.org/10.1063/5.0155887.

(34) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical Modeling of Molecular Energies Using a Deep Neural Network. *J. Chem. Phys.* **2018**, *148* (24), 241715. https://doi.org/10.1063/1.5011181.

(35) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15* (1), 448–455. https://doi.org/10.1021/acs.jctc.8b00908.

(36) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15* (6), 3678–3693. https://doi.org/10.1021/acs.jctc.9b00181.

(37) Fedik, N.; Zubatyuk, R.; Kulichenko, M.; Lubbers, N.; Smith, J. S.; Nebgen, B.; Messerly, R.; Li, Y. W.; Boldyrev, A. I.; Barros, K.; Isayev, O.; Tretiak, S. Extending Machine Learning beyond Interatomic Potentials for Predicting Molecular Properties. *Nat Rev Chem* **2022**, *6* (9), 653–672. https://doi.org/10.1038/s41570-022-00416-3.

(38) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angewandte Chemie International Edition* **2017**, *56* (42), 12828–12840. https://doi.org/10.1002/anie.201703114.

(39) Zhang, S.; Makoś, M.; Jadrich, R.; Kraka, E.; Barros, K.; Nebgen, B.; Tretiak, S.; Isayev, O.; Lubbers, N.; Messerly, R.; Smith, J. Exploring the Frontiers of Chemistry with a General Reactive Machine Learning Potential. ChemRxiv November 3, 2022. https://doi.org/10.26434/chemrxiv-2022-15ct6.

(40) Lindsey, R. K.; Huy Pham, C.; Goldman, N.; Bastea, S.; Fried, L. E. Machine-Learning a Solution for Reactive Atomistic Simulations of Energetic Materials. *Propellants, Explosives, Pyrotechnics* **2022**, *47* (8), e202200001. https://doi.org/10.1002/prep.202200001.

(41) Schreiner, M.; Bhowmik, A.; Vegge, T.; Busk, J.; Winther, O. Transition1x - a Dataset for Building Generalizable Reactive Machine Learning Potentials. *Sci Data* **2022**, *9* (1), 779. https://doi.org/10.1038/s41597-022-01870-w.

(42) Stark, W. G.; Westermayr, J.; Douglas-Gallardo, O. A.; Gardner, J.; Habershon, S.; Maurer, R. J. Machine Learning Interatomic Potentials for Reactive Hydrogen Dynamics at Metal Surfaces Based on Iterative Refinement of Reaction Probabilities. *J. Phys. Chem. C* **2023**, *127* (50), 24168–24182. https://doi.org/10.1021/acs.jpcc.3c06648.

(43) Behler, J.; Csányi, G. Machine Learning Potentials for Extended Systems: A Perspective. *Eur. Phys. J. B* **2021**, *94* (7), 142. https://doi.org/10.1140/epjb/s10051-021-00156-1.

(44) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121* (16), 10073–10141. https://doi.org/10.1021/acs.chemrev.1c00022.

(45) Carbone, M. R. When Not to Use Machine Learning: A Perspective on Potential and Limitations.

MRS Bulletin **2022**, *47* (9), 968–974. https://doi.org/10.1557/s43577-022-00417-z.

(46) Rosenberger, D.; Smith, J. S.; Garcia, A. E. Modeling of Peptides with Classical and Novel Machine Learning Force Fields: A Comparison. *J. Phys. Chem. B* **2021**, *125* (14), 3598–3612. https://doi.org/10.1021/acs.jpcb.0c10401.

(47) Kovács, D. P.; Oord, C. van der; Kucera, J.; Allen, A. E. A.; Cole, D. J.; Ortner, C.; Csányi, G. Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. *J. Chem. Theory Comput.* **2021**, *17* (12), 7696–7711. https://doi.org/10.1021/acs.jctc.1c00647.

(48) Morrow, J. D.; Gardner, J. L. A.; Deringer, V. L. How to Validate Machine-Learned Interatomic Potentials. *The Journal of Chemical Physics* **2023**, *158* (12), 121501. https://doi.org/10.1063/5.0139611.

(49) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal of Computational Physics* **1977**, *23* (2), 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.

(50) Kulichenko, M.; Barros, K.; Lubbers, N.; Li, Y. W.; Messerly, R.; Tretiak, S.; Smith, J. S.; Nebgen, B. Uncertainty-Driven Dynamics for Active Learning of Interatomic Potentials. *Nat Comput Sci* **2023**, *3* (3), 230–239. https://doi.org/10.1038/s43588-023-00406-5.

(51) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proceedings of the National Academy of Sciences* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(52) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603.

(53) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11* (7), 2731–2736. https://doi.org/10.1021/acs.jpclett.0c00497.

(54) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A. E., Reich, S., Skeel, R. D., Eds.; Lecture Notes in Computational Science and Engineering; Springer: Berlin, Heidelberg, 1999; pp 39–65. https://doi.org/10.1007/978-3-642-58360-5_2.

(55) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry* **2002**, *53* (1), 291–318. https://doi.org/10.1146/annurev.physchem.53.082301.113146.

(56) Dellago, C.; Bolhuis, P. G. Activation Energies from Transition Path Sampling Simulations. *Molecular Simulation* **2004**, *30* (11–12), 795–799. https://doi.org/10.1080/08927020412331294869.

(57) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling Methods. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Lecture Notes in Physics; Springer: Berlin, Heidelberg, 2006; pp 349–391. https://doi.org/10.1007/3-540-35273-2_10.

(58) Dellago, C.; Bolhuis, P. G. Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*; Holm, C., Kremer, K., Eds.; Advances in Polymer Science; Springer: Berlin, Heidelberg, 2009; pp 167–233.

https://doi.org/10.1007/978-3-540-87706-6_3.

(59) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-Sampling Strategies for Simulating Rare Events in Biomolecular Systems. *Current Opinion in Structural Biology* **2017**, *43*, 88–94. https://doi.org/10.1016/j.sbi.2016.11.019.

(60) Bolhuis, P. G.; Swenson, D. W. H. Transition Path Sampling as Markov Chain Monte Carlo of Trajectories: Recent Algorithms, Software, Applications, and Future Outlook. *Advanced Theory and Simulations* **2021**, *4* (4), 2000237. https://doi.org/10.1002/adts.202000237.

(61) Bolhuis, P. G.; Dellago, C.; Chandler, D. Reaction Coordinates of Biomolecular Isomerization. *Proceedings of the National Academy of Sciences* **2000**, *97* (11), 5877–5882. https://doi.org/10.1073/pnas.100127697.

(62) Bolhuis, P. G. Transition-Path Sampling of β-Hairpin Folding. *Proceedings of the National Academy of Sciences* **2003**, *100* (21), 12129–12134. https://doi.org/10.1073/pnas.1534924100.

(63) Zhou, H.; Tao, P. Dynamics Sampling in Transition Pathway Space. *J. Chem. Theory Comput.* **2018**, *14* (1), 14–29. https://doi.org/10.1021/acs.jctc.7b00606.

(64) Muždalo, A.; Saalfrank, P.; Vreede, J.; Santer, M. Cis-to-Trans Isomerization of Azobenzene Derivatives Studied with Transition Path Sampling and Quantum Mechanical/Molecular Mechanical Molecular Dynamics. *J. Chem. Theory Comput.* **2018**, *14* (4), 2042–2051. https://doi.org/10.1021/acs.jctc.7b01120.

(65) Rowley, C. N.; Woo, T. K. Generation of Initial Trajectories for Transition Path Sampling of Chemical Reactions with Ab Initio Molecular Dynamics. *The Journal of Chemical Physics* **2007**, *126* (2), 024110. https://doi.org/10.1063/1.2424712.

(66) Bučko, T.; Benco, L.; Hafner, J.; Ángyán, J. G. Monomolecular Cracking of Propane over Acidic Chabazite: An Ab Initio Molecular Dynamics and Transition Path Sampling Study. *Journal of Catalysis* **2011**, *279* (1), 220–228. https://doi.org/10.1016/j.jcat.2011.01.022.

(67) Sun, G.; Jiang, H. Ab Initio Molecular Dynamics with Enhanced Sampling for Surface Reaction Kinetics at Finite Temperatures: CH2 ⇌ CH + H on Ni(111) as a Case Study. *The Journal of Chemical Physics* **2015**, *143* (23), 234706. https://doi.org/10.1063/1.4937483.

(68) Bonati, L.; Zhang, Y.-Y.; Parrinello, M. Neural Networks-Based Variationally Enhanced Sampling. *Proceedings of the National Academy of Sciences* **2019**, *116* (36), 17641–17647. https://doi.org/10.1073/pnas.1907975116.

(69) Bonati, L.; Piccini, G.; Parrinello, M. Deep Learning the Slow Modes for Rare Events Sampling. *Proceedings of the National Academy of Sciences* **2021**, *118* (44), e2113533118. https://doi.org/10.1073/pnas.2113533118.

(70) Yao, S.; Van, R.; Pan, X.; Hwan Park, J.; Mao, Y.; Pu, J.; Mei, Y.; Shao, Y. Machine Learning Based Implicit Solvent Model for Aqueous-Solution Alanine Dipeptide Molecular Dynamics Simulations. *RSC Advances* **2023**, *13* (7), 4565–4577. https://doi.org/10.1039/D2RA08180F.

(71) Lelièvre, T.; Robin, G.; Sekkat, I.; Stoltz, G.; Cardoso, G. V. Generative Methods for Sampling Transition Paths in Molecular Dynamics. arXiv January 31, 2023. http://arxiv.org/abs/2205.02818 (accessed 2024-03-19).

(72) Ojha, A. A.; Thakur, S.; Ahn, S.-H.; Amaro, R. E. DeepWEST: Deep Learning of Kinetic Models with the Weighted Ensemble Simulation Toolkit for Enhanced Sampling. *J. Chem. Theory Comput.* **2023**, *19* (4), 1342–1359. https://doi.org/10.1021/acs.jctc.2c00282.

(73) Kikutsuji, T.; Mori, Y.; Okazaki, K.; Mori, T.; Kim, K.; Matubayasi, N. Explaining Reaction Coordinates of Alanine Dipeptide Isomerization Obtained from Deep Neural Networks Using Explainable Artificial Intelligence (XAI). *The Journal of Chemical Physics* **2022**, *156* (15), 154108. https://doi.org/10.1063/5.0087310.

(74) 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. https://doi.org/10.1063/5.0038516.

(75) Williams, C. D.; Kalayan, J.; Burton, N. A.; Bryce, R. A. Stable and Accurate Atomistic Simulations of Flexible Molecules Using Conformationally Generalisable Machine Learned Potentials. ChemRxiv February 16, 2024. https://doi.org/10.26434/chemrxiv-2024-r75jz.

(76) Benayad, Z.; David, R.; Stirnemann, G. Prebiotic Chemical Reactivity in Solution with Quantum Accuracy and Microsecond Sampling Using Neural Network Potentials. *Proceedings of the National Academy of Sciences* **2024**, *121* (23), e2322040121. https://doi.org/10.1073/pnas.2322040121.

(77) David, R.; Tuñón, I.; Laage, D. Competing Reaction Mechanisms of Peptide Bond Formation in Water Revealed by Deep Potential Molecular Dynamics and Path Sampling. *J. Am. Chem. Soc.* **2024**, *146* (20), 14213–14224. https://doi.org/10.1021/jacs.4c03445.

(78) Chigaev, M.; Smith, J. S.; Anaya, S.; Nebgen, B.; Bettencourt, M.; Barros, K.; Lubbers, N. Lightweight and Effective Tensor Sensitivity for Atomistic Neural Networks. *The Journal of Chemical Physics* **2023**, *158* (18), 184108. https://doi.org/10.1063/5.0142127.

(79) Smith, P. E.; Pettitt, B. M.; Karplus, M. Stochastic Dynamics Simulations of the Alanine Dipeptide Using a Solvent-Modified Potential Energy Surface. *J. Phys. Chem.* **1993**, *97* (26), 6907–6913. https://doi.org/10.1021/j100128a027.

(80) Chekmarev, D. S.; Ishida, T.; Levy, R. M. Long-Time Conformational Transitions of Alanine Dipeptide in Aqueous Solution: Continuous and Discrete-State Kinetic Models. *J. Phys. Chem. B* **2004**, *108* (50), 19487–19495. https://doi.org/10.1021/jp048540w.

(81) Vymětal, J.; Vondrášek, J. Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides — The Alanine Dipeptide Case Study. *J. Phys. Chem. B* **2010**, *114* (16), 5632–5642. https://doi.org/10.1021/jp100950w.

(82) Mironov, V.; Alexeev, Y.; Mulligan, V. K.; Fedorov, D. G. A Systematic Study of Minima in Alanine Dipeptide. *Journal of Computational Chemistry* **2019**, *40* (2), 297–309. https://doi.org/10.1002/jcc.25589.

(83) Schwalbe-Koda, D.; Tan, A. R.; Gómez-Bombarelli, R. Differentiable Sampling of Molecular Geometries with Uncertainty-Based Adversarial Attacks. *Nat Commun* **2021**, *12* (1), 5104. https://doi.org/10.1038/s41467-021-25342-8.

(84) Mehta, M. A.; Fry, E. A.; Eddy, M. T.; Dedeo, M. T.; Anagnost, A. E.; Long, J. R. Structure of the Alanine Dipeptide in Condensed Phases Determined by 13C NMR. *J. Phys. Chem. B* **2004**, *108* (9), 2777–2780. https://doi.org/10.1021/jp037871q.

(85) Bhate, M. P.; Woodard, J. C.; Mehta, M. A. Solvation and Hydrogen Bonding in Alanine- and Glycine-Containing Dipeptides Probed Using Solution- and Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **2009**, *131* (27), 9579–9589. https://doi.org/10.1021/ja902917s.

(86) Feig, M. Is Alanine Dipeptide a Good Model for Representing the Torsional Preferences of Protein

Backbones? *J. Chem. Theory Comput.* **2008**, *4* (9), 1555–1564. https://doi.org/10.1021/ct800153n.

(87) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. Conformational Study of the Alanine Dipeptide at the MP2 and DFT Levels. *J. Phys. Chem. A* **2002**, *106* (13), 3213–3218. https://doi.org/10.1021/jp013952f.

(88) Strodel, B.; Wales, D. J. Free Energy Surfaces from an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide. *Chemical Physics Letters* **2008**, *466* (4), 105–115. https://doi.org/10.1016/j.cplett.2008.10.085.

(89) Bandara, H. M. D.; Burdette, S. C. Photoisomerization in Different Classes of Azobenzene. *Chem. Soc. Rev.* **2012**, *41* (5), 1809–1825. https://doi.org/10.1039/C1CS15179G.

(90) Merritt, I. C. D.; Jacquemin, D.; Vacher, M. Cis → Trans Photoisomerisation of Azobenzene: A Fresh Theoretical Look. *Phys. Chem. Chem. Phys.* **2021**, *23* (35), 19155–19165. https://doi.org/10.1039/D1CP01873F.

(91) Jerca, F. A.; Jerca, V. V.; Hoogenboom, R. Advances and Opportunities in the Exciting World of Azobenzenes. *Nat Rev Chem* **2022**, *6* (1), 51–69. https://doi.org/10.1038/s41570-021-00334-w.

(92) Böckmann, M.; Doltsinis, N. L.; Marx, D. Nonadiabatic Hybrid Quantum and Molecular Mechanic Simulations of Azobenzene Photoswitching in Bulk Liquid Environment. *J. Phys. Chem. A* **2010**, *114* (2), 745–754. https://doi.org/10.1021/jp910103b.

(93) Gagliardi, L.; Orlandi, G.; Bernardi, F.; Cembran, A.; Garavelli, M. A Theoretical Study of the Lowest Electronic States of Azobenzene: The Role of Torsion Coordinate in the Cis–Trans Photoisomerization. *Theor Chem Acc* **2004**, *111* (2), 363–372. https://doi.org/10.1007/s00214-003-0528-1.

(94) Conti, I.; Garavelli, M.; Orlandi, G. The Different Photoisomerization Efficiency of Azobenzene in the Lowest Nπ* and Ππ* Singlets: The Role of a Phantom State. *J. Am. Chem. Soc.* **2008**, *130* (15), 5216–5230. https://doi.org/10.1021/ja710275e.

(95) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54* (2), 724–728. https://doi.org/10.1063/1.1674902.

(96) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *The Journal of Chemical Physics* **2008**, *128* (8), 084106. https://doi.org/10.1063/1.2834918.

(97) Lanl/Hippynn, 2024. https://github.com/lanl/hippynn (accessed 2024-07-22).

(98) Aiqm/Torchani, 2021. https://github.com/aiqm/torchani (accessed 2021-04-22).

(99) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Du\lak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29* (27), 273002. https://doi.org/10.1088/1361-648X/aa680e.

(100) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *WIREs Computational Molecular Science* **2013**, *3* (2), 198–210.

https://doi.org/10.1002/wcms.1121.

(101) 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. https://doi.org/10.1021/acs.jcim.0c00451.

(102) Pople, J. A. Quantum Chemical Models (Nobel Lecture). *Angewandte Chemie International Edition* **1999**, *38* (13–14), 1894–1902. https://doi.org/10.1002/(SICI)1521-3773(19990712)38:13/14<1894::AID-ANIE1894>3.0.CO;2-H.

(103) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks III, C.; Pople, J. A Theoretical Study of Alanine Dipeptide and Analogs. *International Journal of Quantum Chemistry* **1989**, *36* (S16), 311–322. https://doi.org/10.1002/qua.560360725.

(104) Apostolakis, J.; Ferrara, P.; Caflisch, A. Calculation of Conformational Transitions and Barriers in Solvated Systems: Application to the Alanine Dipeptide in Water. *The Journal of Chemical Physics* **1999**, *110* (4), 2099–2108. https://doi.org/10.1063/1.477819.

(105) Chen, M.; Cuendet, M. A.; Tuckerman, M. E. Heating and Flooding: A Unified Approach for Rapid Generation of Free Energy Surfaces. *The Journal of Chemical Physics* **2012**, *137* (2), 024102. https://doi.org/10.1063/1.4733389.

(106) Kang, Y. K. Assessment of CCSD(T), MP2, DFT-D, CBS-QB3, and G4(MP2) Methods for Conformational Study of Alanine and Proline Dipeptides. *Chemical Physics Letters* **2014**.

(107) Chakraborty, D.; Banerjee, A.; Wales, D. J. Side-Chain Polarity Modulates the Intrinsic Conformational Landscape of Model Dipeptides. *J. Phys. Chem. B* **2021**, *125* (22), 5809–5822. https://doi.org/10.1021/acs.jpcb.1c02412.

(108) Wei, D.; Guo, H.; Salahub, D. R. Conformational Dynamics of an Alanine Dipeptide Analog: An *Ab Initio* Molecular Dynamics Study. *Phys. Rev. E* **2001**, *64* (1), 011907. https://doi.org/10.1103/PhysRevE.64.011907.

(109) Velez-Vega, C.; Borrero, E. E.; Escobedo, F. A. Kinetics and Reaction Coordinate for the Isomerization of Alanine Dipeptide by a Forward Flux Sampling Protocol. *The Journal of Chemical Physics* **2009**, *130* (22), 225101. https://doi.org/10.1063/1.3147465.

(110) Swenson, D. W. H.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. OpenPathSampling: A Python Framework for Path Sampling Simulations. 1. Basics. *J. Chem. Theory Comput.* **2019**, *15* (2), 813–836. https://doi.org/10.1021/acs.jctc.8b00626.

(111) Swenson, D. W. H.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. OpenPathSampling: A Python Framework for Path Sampling Simulations. 2. Building and Customizing Path Ensembles and Sample Schemes. *J. Chem. Theory Comput.* **2019**, *15* (2), 837–856. https://doi.org/10.1021/acs.jctc.8b00627.

(112) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient Transition Path Sampling: Application to Lennard-Jones Cluster Rearrangements. *The Journal of Chemical Physics* **1998**, *108* (22), 9236–9245. https://doi.org/10.1063/1.476378.

(113) Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. On-the-Fly Active Learning of Interatomic Potentials for Large-Scale Atomistic Simulations. *J. Phys. Chem. Lett.* **2020**, *11* (17), 6946–6955. https://doi.org/10.1021/acs.jpclett.0c01061.

(114) Sivaraman, G.; Krishnamoorthy, A. N.; Baur, M.; Holm, C.; Stan, M.; Csányi, G.; Benmore, C.; Vázquez-Mayagoitia, Á. Machine-Learned Interatomic Potentials by Active Learning: Amorphous

and Liquid Hafnium Dioxide. *npj Comput Mater* **2020**, *6* (1), 1–8. https://doi.org/10.1038/s41524-020-00367-7.

(115) Wilson, N.; Willhelm, D.; Qian, X.; Arróyave, R.; Qian, X. Batch Active Learning for Accelerating the Development of Interatomic Potentials. *Computational Materials Science* **2022**, *208*, 111330. https://doi.org/10.1016/j.commatsci.2022.111330.

(116) Fedele, C.; Ruoko, T.-P.; Kuntze, K.; Virkki, M.; Priimagi, A. New Tricks and Emerging Applications from Contemporary Azobenzene Research. *Photochem Photobiol Sci* **2022**, *21* (10), 1719–1734. https://doi.org/10.1007/s43630-022-00262-8.

(117) Poutanen, M.; Ikkala, O.; Priimagi, A. Structurally Controlled Dynamics in Azobenzene-Based Supramolecular Self-Assemblies in Solid State. *Macromolecules* **2016**, *49* (11), 4095–4101. https://doi.org/10.1021/acs.macromol.6b00562.

(118) W. Giles, L.; J. Faul, C. F.; F. Tabor, R. Azobenzene Isomerization in Condensed Matter: Lessons for the Design of Efficient Light-Responsive Soft-Matter Systems. *Materials Advances* **2021**, *2* (13), 4152–4164. https://doi.org/10.1039/D1MA00340B.

(119) Cembran, A.; Bernardi, F.; Garavelli, M.; Gagliardi, L.; Orlandi, G. On the Mechanism of the Cis−trans Isomerization in the Lowest Electronic States of Azobenzene: S0, S1, and T1. *J. Am. Chem. Soc.* **2004**, *126* (10), 3234–3243. https://doi.org/10.1021/ja038327y.

(120) Crecca, C. R.; Roitberg, A. E. Theoretical Study of the Isomerization Mechanism of Azobenzene and Disubstituted Azobenzene Derivatives. *J. Phys. Chem. A* **2006**, *110* (26), 8188–8203. https://doi.org/10.1021/jp057413c.

(121) Axelrod, S.; Shakhnovich, E.; Gómez-Bombarelli, R. Thermal Half-Lives of Azobenzene Derivatives: Virtual Screening Based on Intersystem Crossing Using a Machine Learning Potential. *ACS Cent. Sci.* **2023**, *9* (2), 166–176. https://doi.org/10.1021/acscentsci.2c00897.

(122) Maurer, R. J.; Reuter, K. Assessing Computationally Efficient Isomerization Dynamics: ΔSCF Density-Functional Theory Study of Azobenzene Molecular Switching. *The Journal of Chemical Physics* **2011**, *135* (22), 224303. https://doi.org/10.1063/1.3664305.

(123) Gryn'ova, G.; Coote, M. L.; Corminboeuf, C. Theory and Practice of Uncommon Molecular Electronic Configurations: Uncommon Molecular Electronic Configurations. *WIREs Comput Mol Sci* **2015**, *5* (6), 440–459. https://doi.org/10.1002/wcms.1233.

(124) Krylov, A. I. The Quantum Chemistry of Open-Shell Species. In *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2017; pp 151–224. https://doi.org/10.1002/9781119356059.ch4.

(125) Wang, L.; Xu, W.; Yi, C.; Wang, X. Isomerization and Electronic Relaxation of Azobenzene after Being Excited to Higher Electronic States. *Journal of Molecular Graphics and Modelling* **2009**, *27* (7), 792–796. https://doi.org/10.1016/j.jmgm.2008.11.011.

(126) Ess, D. H.; Cook, T. C. Unrestricted Prescriptions for Open-Shell Singlet Diradicals: Using Economical Ab Initio and Density Functional Theory to Calculate Singlet–Triplet Gaps and Bond Dissociation Curves. *J. Phys. Chem. A* **2012**, *116* (20), 4922–4929. https://doi.org/10.1021/jp300633j.

(127) Gräfenstein, J.; Kraka, E.; Filatov, M.; Cremer, D. Can Unrestricted Density-Functional Theory Describe Open Shell Singlet Biradicals? *International Journal of Molecular Sciences* **2002**, *3* (4), 360–394. https://doi.org/10.3390/i3040360.

(128) Fedik, N.; Kulichenko, M.; Steglenko, D.; Boldyrev, A. I. Can Aromaticity Be a Kinetic Trap? Example of Mechanically Interlocked Aromatic [2-5]Catenanes Built from Cyclo[18]Carbon. *Chem. Commun.* **2020**, *56* (18), 2711–2714. https://doi.org/10.1039/C9CC09483K.

(129) Magee, J. L.; Shand, W. Jr.; Eyring, H. Non-Adiabatic Reactions. Rotation about the Double Bond*. *J. Am. Chem. Soc.* **1941**, *63* (3), 677–688. https://doi.org/10.1021/ja01848a012.

(130) Allen, A. E. A.; Lubbers, N.; Matin, S.; Smith, J.; Messerly, R.; Tretiak, S.; Barros, K. Learning Together: Towards Foundational Models for Machine Learning Interatomic Potentials with Meta-Learning. arXiv July 8, 2023. https://doi.org/10.48550/arXiv.2307.04012.

(131) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J.; Berger, F.; Bernstein, N.; Bhowmik, A.; Blau, S. M.; Cărare, V.; Darby, J. P.; De, S.; Della Pia, F.; Deringer, V. L.; Elijošius, R.; El-Machachi, Z.; Falcioni, F.; Fako, E.; Ferrari, A. C.; Genreith-Schriever, A.; George, J.; Goodall, R. E. A.; Grey, C. P.; Grigorev, P.; Han, S.; Handley, W.; Heenen, H. H.; Hermansson, K.; Holm, C.; Jaafar, J.; Hofmann, S.; Jakob, K. S.; Jung, H.; Kapil, V.; Kaplan, A. D.; Karimitari, N.; Kermode, J. R.; Kroupa, N.; Kullgren, J.; Kuner, M. C.; Kuryla, D.; Liepuoniute, G.; Margraf, J. T.; Magdău, I.-B.; Michaelides, A.; Moore, J. H.; Naik, A. A.; Niblett, S. P.; Norwood, S. W.; O'Neill, N.; Ortner, C.; Persson, K. A.; Reuter, K.; Rosen, A. S.; Schaaf, L. L.; Schran, C.; Shi, B. X.; Sivonxay, E.; Stenczel, T. K.; Svahn, V.; Sutton, C.; Swinburne, T. D.; Tilly, J.; van der Oord, C.; Varga-Umbrich, E.; Vegge, T.; Vondrák, M.; Wang, Y.; Witt, W. C.; Zills, F.; Csányi, G. A Foundation Model for Atomistic Materials Chemistry. arXiv March 1, 2024. https://doi.org/10.48550/arXiv.2401.00096.

(132) 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. https://doi.org/10.1039/C6SC05720A.

(133) Hu, Q. H.; Johannesen, A. M.; Graham, D. S.; Goodpaster, J. D. Neural Network Potentials for Reactive Chemistry: CASPT2 Quality Potential Energy Surfaces for Bond Breaking. *Digital Discovery* **2023**, *2* (4), 1058–1069. https://doi.org/10.1039/D3DD00051F.

(134) Rano, M.; Ghosh, D. Efficient Machine Learning Configuration Interaction for Bond Breaking Problems. *J. Phys. Chem. A* **2023**, *127* (16), 3705–3713. https://doi.org/10.1021/acs.jpca.2c09103.

(135) Zhao, Q.; Vaddadi, S. M.; Woulfe, M.; Ogunfowora, L. A.; Garimella, S. S.; Isayev, O.; Savoie, B. M. Comprehensive Exploration of Graphically Defined Reaction Spaces. *Sci Data* **2023**, *10* (1), 145. https://doi.org/10.1038/s41597-023-02043-z.

(136) van Gerwen, P.; Briling, K. R.; Bunne, C.; Somnath, V. R.; Laplaza, R.; Krause, A.; Corminboeuf, C. EquiReact: An Equivariant Neural Network for Chemical Reactions. arXiv December 13, 2023. https://doi.org/10.48550/arXiv.2312.08307.

(137) Fedik, N.; Nebgen, B.; Lubbers, N.; Barros, K.; Kulichenko, M.; Li, Y. W.; Zubatyuk, R.; Messerly, R.; Isayev, O.; Tretiak, S. Synergy of Semiempirical Models and Machine Learning in Computational Chemistry. *The Journal of Chemical Physics* **2023**, *159* (11).

(138) Gaussian 16, Revision C.01, 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, Inc., Wallingford CT, 2016.

(139) Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures. *Journal of Computational Chemistry* **1982**, *3* (2), 214–218. https://doi.org/10.1002/jcc.540030212.