

# A transferable active-learning strategy for reactive molecular force fields

Tom A. Young,<sup>a</sup> Tristan Johnston-Wood,<sup>a</sup> Volker L. Deringer,<sup>b</sup> Fernanda Duarte<sup>a</sup>

<sup>a</sup> Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, United Kingdom

<sup>b</sup> Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, United Kingdom

email: [volker.deringer@chem.ox.ac.uk](mailto:volker.deringer@chem.ox.ac.uk); [fernanda.duarte@chem.ox.ac.uk](mailto:fernanda.duarte@chem.ox.ac.uk)

## Abstract

Predictive simulations of dynamic processes in molecular systems require fast, accurate and reactive interatomic potentials. Machine learning offers a promising approach to construct force-field models for large-scale molecular simulation by fitting to high-level quantum-mechanical data. However, machine-learned force fields generally require considerable human intervention and data volume. Here we show that, by leveraging hierarchical and active learning, accurate Gaussian Approximation Potential (GAP) models for diverse chemical systems can be developed in an autonomous way, requiring only hundreds to a few thousand energy and gradient evaluations on the reference potential-energy surface. Our approach relies on a decomposition of the condensed-phase molecular system into intra- and inter-molecular terms, and on the definition of a prospective error metric to quantify accuracy. We demonstrate applications to a range of molecular systems: from bulk water, organic solvents, and a solvated ion onwards to the description of chemical reactivity, including, a bifurcating Diels–Alder reaction in the gas phase and non-equilibrium dynamics ( $S_N2$  reaction) in explicit solvent. The method provides a route to routinely generating machine-learned force fields for complex and/or reactive molecular systems.

# 1. Introduction

Molecular simulation is a cornerstone in computational chemistry, providing dynamical insights beyond experimental resolution and quantitative predictions in chemical and biochemical settings.<sup>1</sup> Realistic simulations of reactions need to include solvents and other aspects of the chemical environment, and often must be run over extended timescales. In this context, empirical interatomic potentials (force fields) can generally sample the configuration space of interest well, but are limited in accuracy and transferability.<sup>2</sup> Moreover, most of these force fields are parameterised for isolated entities with fixed connectivity, and thus unable to describe bond breaking/forming processes. In contrast, *ab initio* methods provide an accurate description of the potential-energy surface (PES) – which is particularly critical for reactions in solution – but, because of their high computational cost and unfavourable scaling behaviour, are limited to a few hundred atoms and simulation times of picoseconds in *ab initio* molecular dynamics (AIMD).<sup>3</sup>

Machine learning (ML) approaches have the potential to revolutionize force-field based simulations, aiming to provide the best of both worlds,<sup>4–6</sup> and have indeed begun to provide new insights into a range of challenging research problems.<sup>7–15</sup> The development of a truly general ML potential mapping nuclear coordinates to total energies and forces is, however, precluded by the curse of dimensionality. Within small chemical subspaces, models can be achieved using neural networks (NNs),<sup>6,16–20</sup> kernel-based methods such as Gaussian processes (GP)<sup>21,22</sup> and gradient-domain machine learning (GDML),<sup>23</sup> and linear fitting with properly chosen basis functions,<sup>24,25</sup> each with different data requirements and transferability.<sup>26</sup> In the present work, we employ the Gaussian Approximation Potential (GAP) framework,<sup>21</sup> which has been used to generate force fields for a range of elemental,<sup>27–29</sup> multicomponent inorganic,<sup>30,31</sup> and recently gas-phase organic systems.<sup>14,32</sup> Initial studies of condensed phase molecular systems with GAPs include fluid methane<sup>33</sup> and phosphorus.<sup>34</sup> These potentials, while accurate, have required considerable computational effort and human oversight. Indeed, condensed-phase NN<sup>35,36</sup> and GAP fitting approaches typically require several thousand reference (“ground truth”) evaluations.

Active learning (AL) has started to emerge as one of the key strategies for generating reference databases and accelerating the fitting process.<sup>30,37–41</sup> Notable examples in the modelling of materials include an early demonstration of a “query-by-committee” approach in fitting a high-dimensional NN potential for elemental copper,<sup>38</sup> the fitting of Moment Tensor Potential (MTP)<sup>25</sup> models based on the D-optimality criterion<sup>42</sup> which has been applied, e.g., to the prediction of elemental crystal structures<sup>37</sup> and multicomponent alloys,<sup>39</sup> and the deep potential generator (DP-GEN)<sup>43,44</sup> that provides an interface to deep neural network potential models.<sup>45</sup> More recently, AL approaches have also been combined with GP based force fields including GAP,<sup>46</sup> and included within a first-principles MD implementation such that it allows the “on the fly” fitting of force fields for a specific simulation system.<sup>47,48</sup>

Efficient approaches to generate training data become even more important in computational organic chemistry, in particular when exploring chemical reactions where accuracy beyond DFT is required.

Reactive AIMD simulations to describe bond breaking/forming processes in organic reactions in explicit solvents are demanding even for small systems at a DFT level, and often out of reach for high-level wavefunction-based methods, for which only single point (energy/force) evaluations are possible. Very recently, AL approaches have started to be adopted for fitting reactive potentials for organic molecules based on single point evaluations at quantum-chemical levels of theory. Notable examples include the modelling of gas-phase pericyclic reactions,<sup>13</sup> the exploration of reactivity during methane combustion,<sup>49</sup> and the decomposition of urea in water.<sup>40</sup>

Here, we show how *ab initio* quality GAPs can be trained with hundreds of *total* ground truth evaluations for liquid water, requiring no *a priori* knowledge of the system. We then show how this methodology can be directly transferred to other chemical systems in the gas phase, and in implicit and explicit solvent, focusing on the applicability to a range of scenarios that are relevant in computational chemistry. This work therefore outlines an efficient and general approach for constructing reactive GAP models that enable condensed-phase molecular simulations.

## 2. Results and Discussion

Despite GAP fitting being increasingly used for inorganic systems, we found that the same fitting strategies did not easily transfer to the description of complex molecular systems. Indeed, our initial attempts to generate GAPs for  $\text{H}_2\text{O}_{(l)}$ ,  $\text{Pd(II)}_{(aq)}$  and  $\text{Zn(II)}_{(aq)}$  employing reported strategies without adaptation were unsuccessful. Even with a high correlation and low error on energies in unseen test data, potentials were not stable for more than a few femtoseconds. In the following section, we therefore introduce a new strategy along with a prospective error metric, to develop robust models for gas-phase and condensed-phase molecular systems.

### 2.1 A prospective error metric

The initial step in validating supervised machine learning (ML) tends to follow the splitting of a dataset into training and test sets, training the model, then evaluating its performance on the test set with a squared error (RMSE/MSE) or a correlation ( $R^2$ ) metric. As with model overfitting, this ‘retrospective’ validation strategy ultimately limits the applicability of these models.<sup>50–56</sup> In an ML potential, the minimum required domain of applicability is the region of configuration space likely to be sampled during a simulation with the potential. However, this region is not known *a priori*, making the choice of test data problematic if not impossible for use in a standard train/test data split approach. In addition, one would also like to ensure high accuracy in regions sampled on the ground truth surface (especially for early versions of an evolving potential), but being able to quantify this accuracy requires dynamics at the ground truth method level in the first place, which is much more challenging than sampling with an efficient potential.

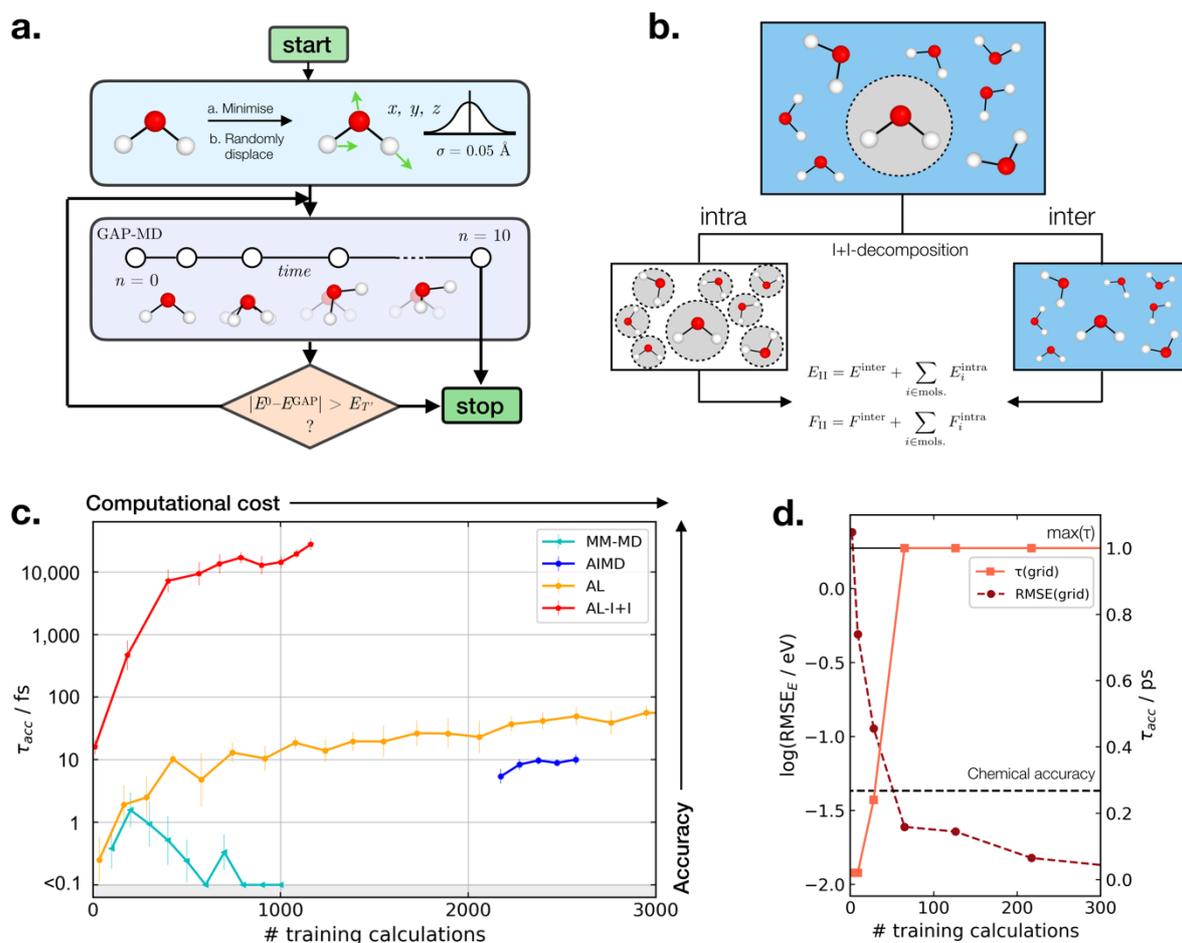
Using a train/test set split with high structural similarity between the two sets can lead to highly misleadingly accuracy whenever the potential is to be taken outside the training region in computational practice. For example, splitting an AIMD trajectory of water into a training and test set with an odd/even frame split (50:50) and training a simple GAP model yields an energy error on the order of 1–2 kcal mol<sup>-1</sup> (**Figure S1a**). However, simulations with this potential in the same configuration space sample unphysical configurations within 10 fs (**Figure S1b**), making RMSE over the test data an insufficient metric in quantifying the quality of a potential.

Considering that single-point reference energy evaluations are reasonably cheap, a ‘prospective’ validation scheme is possible, where the error metric operates in the configuration space sampled in a simulation. With this in mind, we propose a temporal cumulative error metric ( $\tau_{\text{acc}}$ , Eqn. 1), defined as the time required for the cumulative error (absolute difference between true ( $E^0$ ) and predicted ( $E^{\text{GAP}}$ )) to exceed a given threshold ( $E_T$ ); the larger  $\tau_{\text{acc}}$ , the more robust the potential. Note that a potential’s stability can far exceed  $\tau_{\text{acc}}$ , as shown in the following. Here only errors above a lower-bound threshold value ( $E_l$ ) contribute to the cumulative error. The lower threshold is required to account for the residual error that is due to the finite

radial cut-off of the model. In the following we take  $E_T$  to be 10 times  $E_l$ , but it may be adjusted dependent on the simulation.

$$\tau_{\text{acc}} = \text{time} \quad : \quad E_T < \sum_{i \in \text{frames}} \max(|E_i^0 - E_i^{\text{GAP}}| - E_l, 0) \quad (1)$$

This metric has several advantages in that (a) it ensures that a potential with high accuracy will result in stable dynamics; (b) it allows the user to specify the level of accepted error according to the quality of the training method, thus not penalising where the error is within the difference between the ground truth and the true PES (i.e. a larger threshold would be more suitable for a less accurate reference method); (c) it penalises large errors, even if they only occur for single configurations, which is important as they may lead to instabilities in the ML-driven MD trajectory and (d) it enables a quoted accuracy to include regions that may not be accessible to the ground truth (e.g. long-time behaviour). Overall, this metric depends on the lower bound and total error, interval between evaluations, and the simulation on which it is evaluated, so while not unique it is – crucially – prospective. As explained in the next paragraphs, we found this metric to be essential in developing an efficient training strategy and accurate potentials for bulk water (**Figure 1**).



**Figure 1.** (a) Schematic of the active learning loop implemented here for fitting GAP models, where the GAP-MD exploration is run for  $n^3 + 2$  femtoseconds, where  $n$  (the number of evaluations) is incremented after each the error is evaluated. (b) Schematic illustrating the separation into inter- intra-molecular terms (I+I) for a bulk water system; these are described by separate GAP models and then added to give the combined prediction for energies,  $E$ , and forces,  $F$ . (c) Learning curves for a bulk water GAP model using different training strategies.  $\tau_{\text{acc}}$  with  $E_l = 0.1$  eV,  $E_t = 1$  eV, 10 fs interval, 300 K, from the same random minimised configuration of 10 waters in a  $7 \text{ \AA}$  cubic box. Error bars quoted as the standard errors in the mean from 5 independent repeats. The horizontal axis denotes the number of evaluations in training data generation. See **Tables S1-2** for detailed methods. DFTB(3ob) ground truth. Minimum  $\tau_{\text{acc}}$  is shown as 0.1 fs for plotting on a log scale. (d) Water monomer training in  $\tau_{\text{acc}}$  and RMSE over the full 3D PES, see **Figure S2** for details.

## 2.2 Water models

For bespoke ML potentials to be routinely developed for molecular systems, one would hope to be able to complete the data generation and model training within a matter of hours to days, and to quantify the accuracy of the resulting potential. With this in mind, here we train a GAP to simulate bulk water, aiming to minimize the number of required ground truth evaluations as well as the required human intervention in the setup, while maximizing stability (measured by  $\tau_{\text{acc}}$ ). A selection of training strategies is discussed in the following paragraphs and their results are outlined in **Figure 1**.

We initially employed training strategies found to work well in elemental materials by, for example, fitting a combined potential with two- and three-body GAPs; this approach was, however, found to be detrimental to the potential's stability. This can be understood considering a water dimer ( $\text{HO}-\text{H}_c\cdots\text{OH}_2$ ); here, a two-body description that treats the two O-H<sub>c</sub> interactions on the same footing is a poor approximation, in view of the different order of magnitude between the interactions (**Figure S3**). The following GAPs therefore only employ a smooth overlap of atomic positions<sup>57</sup> (SOAP) descriptor for an exclusively many-body description of atomic environments (except for AL-I+I where a different approach is used, *vide infra*).

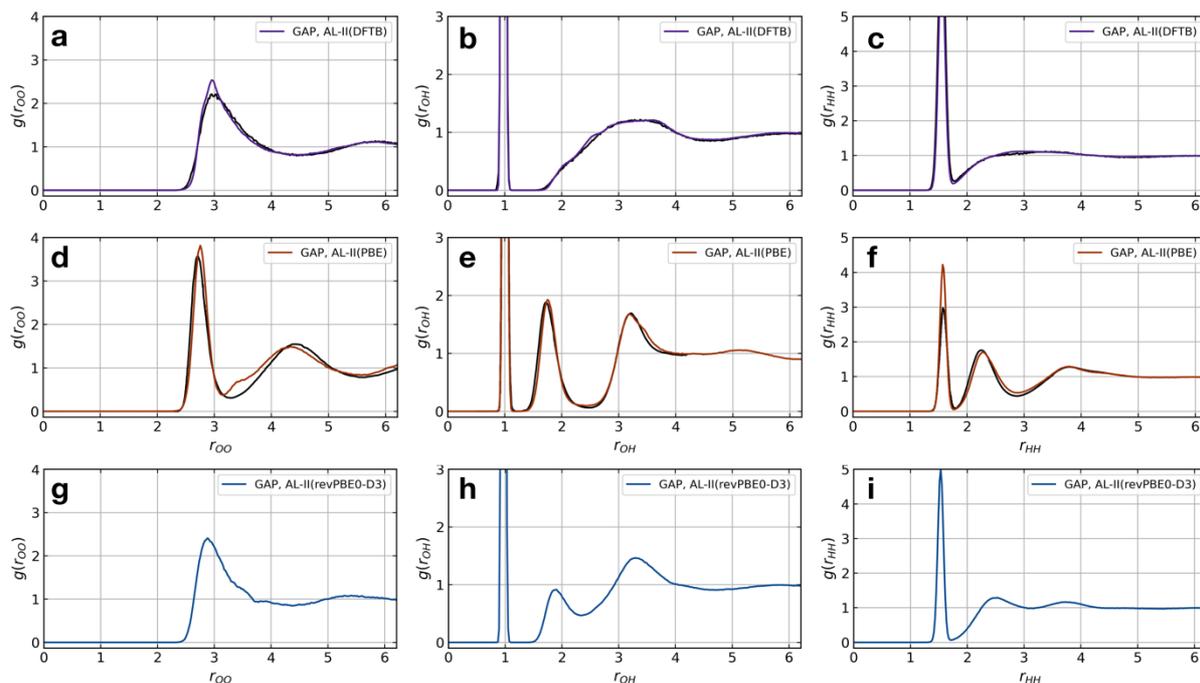
Similarly, an emerging approach in generating training data for elemental GAPs is to initialize the database with randomized configurations (with reasonable constraints, as in *ab initio* random structure searching<sup>58</sup>), evaluate reference energies and forces, then train the model. However, a similar approach of randomly placing water molecules does not afford a stable potential. A similar result is observed when the most diverse configurations are selected using the CUR algorithm<sup>59</sup>, as in ref. <sup>60</sup> (**Figure S4**) or when applying intramolecular displacements, following minimisation (**Figure S4**). Classical molecular mechanics (MM) at temperatures of 100–1000 K were also found to be an ineffective strategy for generating reference configurations (**Figure S5**), reaching  $\tau_{\text{acc}}$  of only a few fs (“MM-MD”, **Figure 1c**). This is in line with the results reported in ref. 14. Note that this is not because the GAP cannot fit reference energies and forces from MM configurations (**Figure S6**), but because of a poor configuration space overlap with the ground truth PES (**Figure S7**). Selecting configurations from an AIMD simulation at 300 K (AIMD, **Figure 1c**) was an improvement over training on random and MM-generated configurations, with  $\tau_{\text{acc}} \sim 10$  fs. Adding AIMD configurations, however, saturates in accuracy quickly even if those are obtained at higher temperatures (**Figure S8**). Using AIMD configurations also requires significant cost ( $O(10^3)$  evaluations). Finally, active learning from only a few randomly generated configurations provides a modest uplift in accuracy (AL, **Figure 1c**), with accuracy on-par with GAP trained on AIMD configurations at a third of the required reference data.

Only when the relevant energy scales of the system are decomposed by treating intra- and inter-molecular components separately (**Figure 1b**) does our approach generate a potential that is stable for  $O(\text{ps})$  (AL-I+I, **Figure 1c**). We note that this idea is related to, but different in detail from, the hierarchical fitting of

GAPs<sup>33,61</sup> and related ML models<sup>32,62</sup> using different levels of computational approaches. Here, we employ the same level of input data throughout, but describe the stronger (e.g., covalent) and weaker intermolecular terms with separate fits that are afterwards combined to give the final model. The intramolecular GAP for water contains only 2- and 3-body terms and the training data are chosen using an evenly spaced grid over the full 3-atom PES ( $8\times 8\times 8$  grid points in  $r_{\text{OH}}$  and  $r_{\text{HH}}$ ,  $\sim 0.1$  Å spacing, **Figure S9**). Energy and force evaluations of this potential are a simple sum of intra- and inter-molecular terms, but require the former to be evaluated in an expanded simulation box to ensure no non-bonded hydrogens are present within the radius of the 2- and 3-body descriptors on oxygen (**Figure S10**). Here the PES is fairly low-dimensional, so a full and reasonably dense grid is available, which in turn allows us to define a rigid error measure, where we find the error to be inversely correlated with  $\tau_{\text{acc}}$  (**Figure 1d**).

Using an acceptable error of 10 meV/H<sub>2</sub>O molecule for a description of bulk water, which is similar to that achieved in a recent NN fit of water,<sup>35</sup> we find that this potential affords  $\tau_{\text{acc}} > 10$  ps with just a few hundred ground truth evaluations (AL-I+I, **Figure 1c**). This accuracy is on-par with a bespoke NN<sup>35</sup> ( $\tau_{\text{acc}} = 7.6\pm 0.7$  ps) when trained on DFTB energy and forces from a partially manually curated dataset of  $\sim 7000$  training configurations. For comparison, an ANI<sup>18</sup> potential trained using  $\sim 6000$  (H<sub>2</sub>O)<sub>5</sub> clusters from the ANI-1x dataset gives  $\tau_{\text{acc}}$  of  $60\pm 9$  fs, implying that with these training data ANI does not afford an accurate bulk water model; somewhat expected considering the training data does not include any periodic configurations.

The model fitted using our approach (AL-I+I) yields a good description of the ground-truth radial distribution functions (RDFs), both considering the location and intensities of the peaks corresponding to the first and second coordination shells (**Figure 2a-c**). This is despite the relatively short-range atomic cut-offs (3 Å, O only) used. Interestingly, for a DFT-quality GAP simply re-evaluating energies and forces on DFTB-derived active-learned configurations is insufficient, with the DFTB configurations being high in energy at the DFT level ( $\sim 5$  eV, **Figure S11**). However, applying an active learning strategy with a PBE reference method and a slightly larger 3.5 Å cut-off generates excellent agreement with the true AIMD simulation from ref. <sup>63</sup>, in only a few hours of total training time (**Figure 2d-f**). The real significance, of course, is in moving to more accurate ground-truth methods, for which a full MD would not be straightforward: indeed, using the same method, a hybrid DFT-quality water model can be generated within a few days, which would be inaccessible with other methods (the generation of the GAP model required  $\sim 5$  days on 20 CPU cores, **Figure 2g-i**). These results suggest that the training strategy (and hyperparameter selection) presented here is suitable independent of the reference method.



**Figure 2.** Oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen RDFs from NVT MD simulations of 64 water molecules in a 12.42 Å cubic box, with ground truth (black) and GAP (purple) simulations. (a–c) DFTB(3ob params) ground truth, 100 ps, 300 K,  $r_c^{\text{SOAP}}(\text{O}) = 3$  Å. (d–f) DFT(PBE) reference RDF data extracted from ref.<sup>63</sup>, 30 ps, 330 K,  $r_c^{\text{SOAP}}(\text{O}) = 3.5$  Å. (g–i) DFT(revPBE0-D3) GAP, 30 ps, 330 K,  $r_c^{\text{SOAP}}(\text{O}) = 4.0$  Å.

### 2.3 Other solvent systems

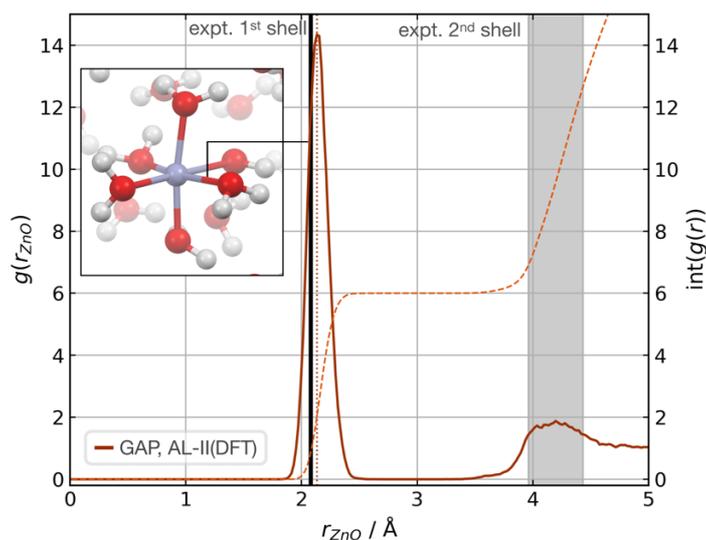
Using an identical training strategy to the one described for water, we then trained GAPs for a selection of organic solvents with varying interactions (acetonitrile, MeCN; methanol, MeOH; acetone,  $\text{CH}_3\text{C}(\text{O})\text{CH}_3$ ; pyridine,  $\text{C}_5\text{H}_5\text{N}$ ; ammonia,  $\text{NH}_3$ ). To generate the reference simulation data for this proof-of-concept reasonably quickly, a DFTB ground truth is employed, with chlorinated solvents not selected due to a large discrepancy between the DFTB-generated and experimental C–Cl bond dissociation energy (**Figure S12**). A uniform grid over the intramolecular PESs is now no longer possible, thus AL is used to develop an initial intramolecular potential trained using GAP-MD at 1600 K (see **Figure S13**). This temperature is used to sample higher-energy configurations more efficiently. In all cases, only hundreds of ground truth evaluations were necessary to generate GAPs affording stable dynamics, with  $\tau_{\text{acc}}$  values on the order of picoseconds (**Figure S14**, **Table 1**, SI §S1). For a representative example, the computed RDFs for acetonitrile (MeCN) compare well with the ground truth (**Figure S15**). As with the water models above, to quantitatively evaluate bulk properties training an accurate reference method and the inclusion of nuclear quantum effects would be necessary.<sup>33,35</sup> Nevertheless, this example already demonstrates that our training method is applicable to a range of chemical systems beyond water or aqueous solutions.

**Table 1.** Average number of *total* ground truth evaluations (over 5 repeats quoted with a standard error in the mean) required to obtain a potential with  $\tau_{\text{acc}} > 3$  ps, where  $E_T = 1$  eV,  $E_I = 0.1$  eV, 300 K. All SOAPs used 3.0 Å cut-offs; they are centred on the stated atomic species, and include all atoms within the neighbourhood of those atoms (including hydrogen). See **Table S3** for more detailed parameters.

| Solvent                             | SOAP descriptors centred on | $N_{\text{intra}}$ | $N_{\text{inter}}$ |
|-------------------------------------|-----------------------------|--------------------|--------------------|
| MeCN                                | C, N                        | 269±12             | 120±6              |
| MeOH                                | C, O                        | 221±13             | 292±49             |
| CH <sub>3</sub> C(O)CH <sub>3</sub> | C, O                        | 566±8              | 359±29             |
| C <sub>5</sub> H <sub>5</sub> N     | C, N                        | 249±36             | 243±11             |
| NH <sub>3</sub>                     | N                           | 38±4               | 109±24             |

## 2.4 Zn(II) in aqueous solution

Modelling metal ions in solution remains one of the main challenges for general purpose force fields.<sup>64</sup> Historically, they have been developed by fitting van der Waals parameters to reproduce RDFs and hydration free energy of aquo complexes, which are expected to be transferrable to more complex environments. However, while simple, these models have been often found to lead to unstable simulations or poorly describe structural properties.<sup>64</sup> Considering these challenges and their relevance in biomolecular modelling, we decided to use our strategy to generate a GAP for aqueous Zn(II) ion as a representative system. Here the system was decomposed into a  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$  cluster and the remaining water molecules and training the intermolecular interactions separately, using an identical strategy with slightly larger 4.0 Å intermolecular cut-offs on oxygen. This model reproduces the experimental coordination number, and Zn–O distances of both the first and second hydration shells without further optimization (**Figure 3**). From random points in the configuration space of  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$  and 20 water molecules (intermolecular distances  $> 1.7$  Å, 10 Å cubic box)  $\tau_{\text{acc}}$  reached 0.5 ps under a 300 K simulation ( $E_I = 0.037$  eV / H<sub>2</sub>O, 20 fs interval). Note this value is far short of the 100 ps simulations performed to generate the RDF and illustrates that a potential may be stable for  $t \gg \tau_{\text{acc}}$ .



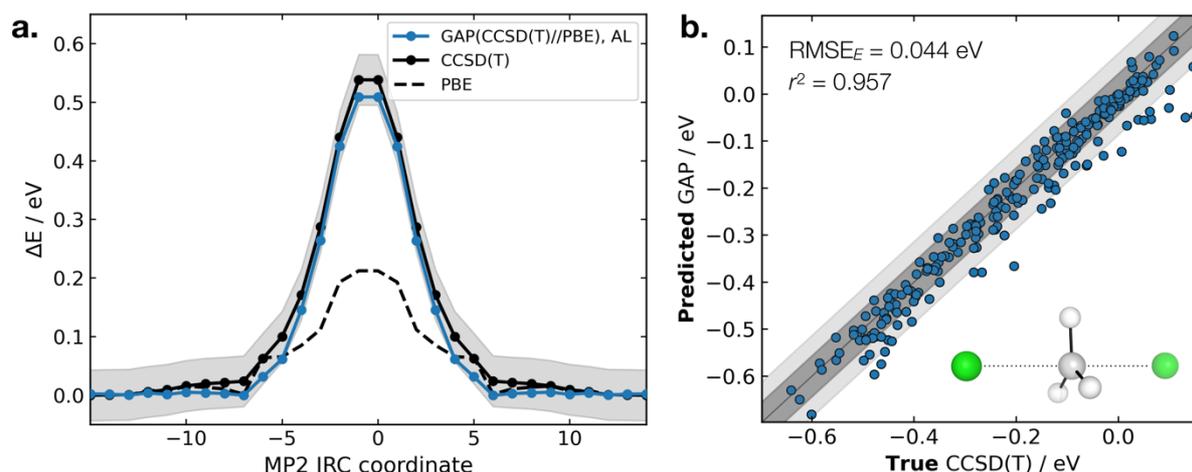
**Figure 3.** Zn–O radial distribution function averaged from 1 ns of cumulative ( $10 \times 100$  ps) NVT MD simulations of Zn(II) in aqueous solution at 300 K, with the experimental modal Zn–O distance shown in black. Experimental (X-ray diffraction) Zn–O distances from ref.<sup>65</sup>, octahedral first hydration shell. Shaded area denotes the range of experimental second hydration shell (ref.<sup>65</sup> and cited within). GAP trained as those in **Table 1** using a PBE/400 eV ground truth, intra-Zn(H<sub>2</sub>O)<sub>6</sub> used a O SOAP  $r_c = 3.0$  Å and inter  $r_c = 4.0$  Å.

## 2.5 Reaction dynamics in gas and solvent phase

The high dimensionality and ensuing flexibility of machine-learned force fields makes them highly suitable in principle to study reaction dynamics – the latter usually require many costly electronic structure calculations to obtain atomic level descriptions of reaction mechanisms, solvent effects or post-transition state (TS) dynamics.<sup>66,67</sup> In the following section, we show that our data-efficient strategy enables accurate reactive potentials ( $\tau_{\text{acc}} > 100$  fs) with only a few hundreds of DFT evaluations for set of prototypical organic reactions.

### 2.5.1 Bimolecular nucleophilic substitution (S<sub>N</sub>2) in the gas phase

The S<sub>N</sub>2 nucleophilic substitution reaction is fundamental in organic chemistry and has been extensively studied using AIMD and analytically fit PESs.<sup>66,68</sup> However, even with efficient approaches to fitting PES, AIMD methods still require tens of thousands of energy evaluations.<sup>69</sup> Here, we generated a reactive GAP to study the reaction between chloride and methyl chloride. By initialising active learning from the TS the true intrinsic reaction coordinate is reproduced to within 1 kcal mol<sup>-1</sup> (**Figure S16**). Interestingly, and unlike the previous attempt to generate a DFT-quality GAP by evaluating energies and forces on DFTB active-learned configurations, here uplifting a DFT-GAP to an accurate wavefunction GAP is possible. This method allows coupled cluster [CCSD(T)]-quality IRC (**Figure 4a**) and dynamics to be propagated from the TS with just 55 energy and (numerical) force evaluations (**Figure 4b/S16**). The resultant GAP is considerably more accurate than the underlying DFT over the IRC (dashed, **Figure 4a**).

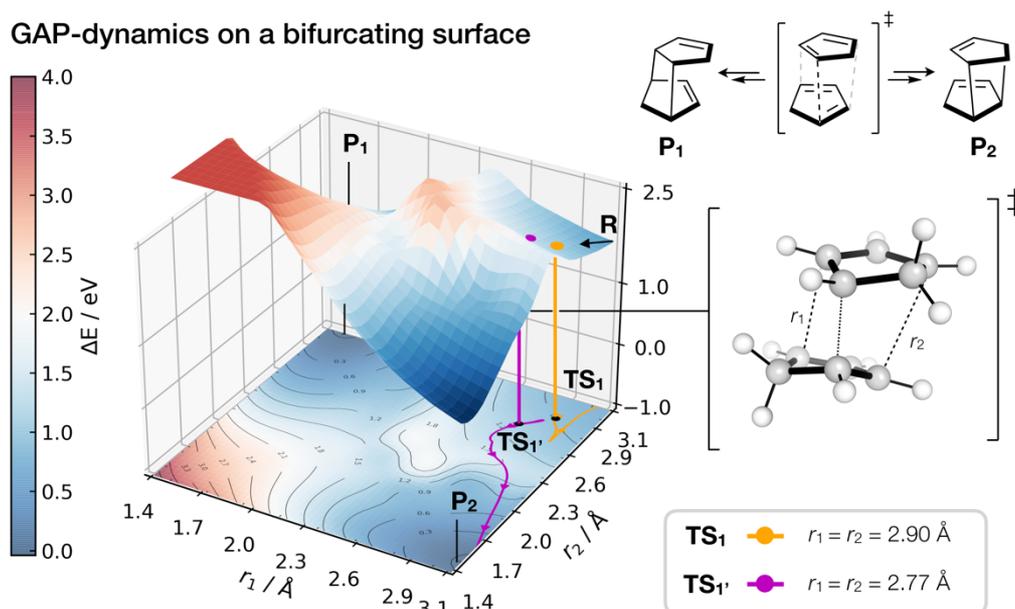


**Figure 4.** (a) Predictions (GAP) and true (DLPNO-CCSD(T)/ma-def2-TZVPP) energy values on the MP2/ma-def2-TZVPP intrinsic reaction coordinate (IRC). Shaded region bounds the ‘chemically accurate’ ( $\pm 1$  kcal mol<sup>-1</sup>) region. (b) Errors between GAP predictions and true energies from ten 100 fs GAP-MD trajectories initialised from the TS (300 K). Dark and light shaded area bound the  $\pm 1$  kcal mol<sup>-1</sup> and  $\pm 2$  kcal mol<sup>-1</sup> regions, respectively.

## 2.5.2 Post-TS Bifurcating pathway in a Diels–Alder reaction

GAPs for more complex reactions involving reactions that proceed on a bifurcating PES can also be trained. These reactions typically require AIMD simulations, where selectivity is determined from the average behaviour of many trajectories leading to either product. Other approaches have also been developed.<sup>70</sup>

We explored the dimerization between cyclopentadiene, for which *endo* selectivity has been rationalised on the basis of bifurcating reaction pathways.<sup>71</sup> Once again, initiating active learning from the literature TS (**TS<sub>1</sub>**, **Figure 5**) using a DFT method analogous to the one used in the original work by Caramella and co-workers affords a reactive potential from which 500 fs trajectories were propagated. Interestingly, we found that propagating the system from this TS did not afford any products (**P<sub>1</sub>** or **P<sub>2</sub>**, **Figure 5**), with all trajectories leading to the reactant state (**Figure S19**). Further investigation and generation of the relaxed 2D potential energy surface over the two possible forming C–C bonds ( $r_1$ ,  $r_2$ ) leading to products provided a rather different surface to the one suggested in ref.<sup>71</sup>, with a flat portion then an incline as  $r_1$ ,  $r_2$  shorten below 2.9 Å, with a steeply exergonic reverse reaction (IRC, **Figure S19**). As noted by Caramella, following the IRC forwards from **TS<sub>1</sub>** the reaction proceeds to another **TS<sub>1'</sub>**, which is similar in energy ( $\Delta E = 2$  kcal mol<sup>-1</sup>). By training a GAP at 500 K and propagating GAP-MD from **TS<sub>1'</sub>**, sampling the area of the PES around a valley-ridge inflection point (VRI), trajectories did lead to the expected two products (e.g., purple line, **Figure 5** and **Figure S21**). With no *a-priori* knowledge, apart from the structure of **TS<sub>1</sub>**, the topology of the bifurcating surface can therefore be revealed efficiently using GAP dynamics. Given the strategy is completely automated, and with training taking a few hours to days, this enables the routine examination of dynamical effects at the transition state following location of a TS.



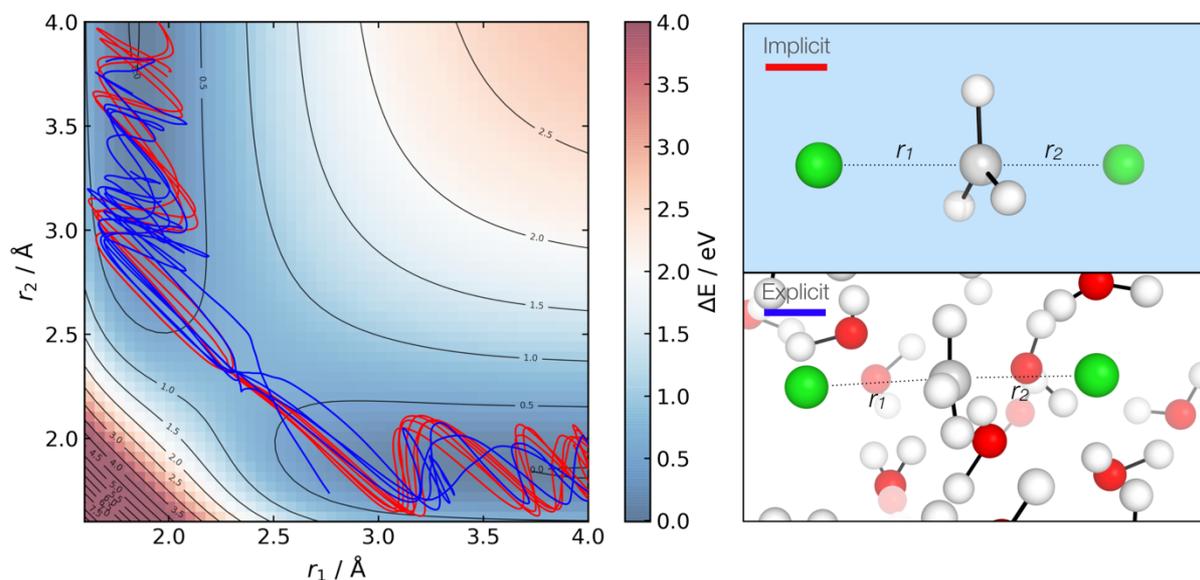
**Figure 5.** 2D PES (B3LYP/def2-SVP) along the forming bond distances ( $r_1, r_2$ ) in the dimerization of cyclopentadiene. An example of GAP-propagated reactive dynamics (300 K) is shown from  $TS_1$  (7N in ref. <sup>71</sup>), which leads to reactants, (representative trajectory in orange), and from  $TS_{1'}$ , which leads to products (a representative trajectory is shown in purple). 3D projection is truncated at 2.5 eV above the minimum for plotting. Interpolated surface used a cubic spline using *scipy.interp2d* with the raw surfaces shown in **Figure S22**. All trajectories shown in **Figure S20-21**.

## 2.6 Solvated reaction dynamics

The ability to accurately describe bond-breaking/forming paths in the condensed-phase is crucial if this strategy could be applied to increasingly complex processes, such as enzymatic reactions. Towards this goal, and having accurately generated potentials for condensed phase molecular systems and gas-phase reactions, we decided to extend our active learning strategy to explicitly solvated reactions; once again, using the  $S_N2$  reaction between chloride and methyl chloride as a test case.  $S_N2$  reactions have been used as a test case for a recent ML potential, but those studies have been limited to the gas phase and used thousands of training points.<sup>20</sup> Only recently, a NN has been trained by Parrinello and coworkers to study urea decomposition in water.<sup>40</sup>

Adopting an identical strategy to the one employed for the systems described above, the intra and intermolecular PESs dynamics can be propagated from the TS and the effect of explicit solvation interrogated (**Figure 6**). Using only the gas-phase TS as an *a-priori* point on the PES the training is complete in a day ( $\sim 100$ s CPUh) if only short time ( $< 1$  ps) dynamics are of interest. Interestingly, the behaviour in explicit water (blue, **Figure 6**) solvent differs from the implicit counterpart (red, **Figure 6**). Solvent reorganisation is instantaneous in the implicit case which results in oscillations in the C-Cl bond

characteristic of a gas phase reaction. In contrast, in explicit solvent the dynamics are more complex, with a slower transition from the product channel. Additionally, one of the 10 trajectories recrosses the barrier after 170 fs of simulation (**Figure S23**), where the solvent has not reorganised to accommodate the anionic chloride yet, making the path to products shallower in energy.



**Figure 6.** Reactive GAP MD trajectories (lines) propagated from the TS, trained on implicitly solvated configurations (CPCM(Water)-PBE/def2-SVP, red) and explicitly solvated configurations (PBE/400 eV, blue). Ground truth implicitly solvated surface (cubic interpolated, 5:1) underneath with the error to the GAP prediction in **Figure S24**. Intramolecular GAP used C-only ( $r_c^{\text{SOAP}} = 6 \text{ \AA}$ ) descriptor with the intermolecular explicit solvent using a SOAP on O ( $r_c^{\text{SOAP}} = 3.5 \text{ \AA}$ ) and Cl ( $r_c^{\text{SOAP}} = 4.5 \text{ \AA}$ , **Figure S25**).

The component-wise separation of the system also leads to the possibility of training to a more accurate *ab initio* surface for the gas phase reaction, in a similar way to QM/MM, but here a ML(A)/ML(B) partition is available where A and B are two different ground truth methods. Application of this kind of hierarchical ML potential fitting will be the subject of further work.

### 3. Conclusion

Studying dynamic processes and the effect of explicit solvation on chemical reactions demands a rapid method to develop bespoke force-field models. Here, we demonstrated that within the Gaussian Approximation Potential (GAP) machine learning framework, accurate and robust models can be developed efficiently for gas-phase and condensed-phase molecular reactions. Our strategy employs only a few randomly selected points in the configuration space, from which active learning training of intra and intermolecular components to the energy and forces is carried out. The developed method is completely automated and publicly available (<https://github.com/t-young31/gap-train>). We also define a prospective error metric, which is found to be crucial in developing robust potentials, whereas correlation on a predefined test set being insufficient. We illustrate the generality of this approach by modelling bulk water, Zn(II) in aqueous solution, and chemical reactions in the gas phase and explicit solvent, including post-TS cyclisation and S<sub>N</sub>2 reactions. The diversity of the examples presented here demonstrates the general applicability of the strategy and encourages applying this approach in the modelling of more complex reactions in homogeneous and heterogenous environments.

## 4. Methods

### 4.1 Technical Details

All Gaussian Approximation Potentials (GAPs) were trained using the GAP and QUIP codes (Singularity distribution, commit #66c553f) and a Smooth Overlap of Atomic Positions (SOAP)<sup>57</sup> kernel with hyperparameters as defined in **Table S1**. GAP-MD simulations were performed with ASE<sup>72</sup> interfaced to QUIP with the quippy wrapper using the Langevin integrator with 0.5 fs timesteps unless otherwise specified. Initial configurations, CUR<sup>59</sup> selection and all learning curves were generated with the *gap-train* module, which was used to run the automated fitting.<sup>73</sup> CUR selection used SOAPs averaged over atoms in a configuration using the Dscribe<sup>74</sup> package ('inner' averaging, over entries of the coefficient vector). Intra+Inter (I+I) energy and force evaluations used an expansion factor of 10 to ensure no intermolecular atoms were within the intra GAP cut-off. The NumPy<sup>75</sup> implementation introduces a negligible computational overhead for expanding the box ( $\sim 0.1$  ms step<sup>-1</sup> real time) but requires two GAP calculations on the inter and intra components, currently carried out in serial mode. All GAPs with the exception of the revPBE0-D3 water potential were trained in less than a day on 10 CPU cores. The revPBE0-D3 water potential was constructed from scratch in 5 days (1 intra + 4 inter) and used 20 CPU cores.

Periodic DFTB calculations performed with DFTB+<sup>76</sup> using 3ob<sup>77</sup> parameters, and molecular equivalents using GFN2-XTB<sup>78</sup> in XTB v. 6.2.3. Periodic pure GGA DFT calculations performed with GPAW<sup>79,80</sup> v. 19.8.1 with the PBE<sup>81</sup> functional and a 400 eV plane-wave cut-off from a *dzp* LCAO initial guess at the gamma point. Hybrid periodic DFT calculations with the revPBE0<sup>82,83</sup> functional combined with the D3<sup>84</sup> dispersion correction were performed with CP2K.<sup>85</sup>

Molecular DFT, MP2 and coupled cluster [DLPNO-CCSD(T)] calculations performed with ORCA<sup>86,87</sup> v. 4.2.1 wrapped with autode<sup>88</sup> using PBE<sup>81</sup> and PBE0<sup>83</sup> functionals, (ma)-def2-SVP, def2-TZVP, ma-def2-TZVPP basis sets.<sup>89</sup> AIMD calculations at the DFTB level performed with DFTB+ with 3OB parameters<sup>77</sup> and MM simulations performed with GROMACS<sup>90,91</sup> 2019.2 with TIP3P parameters.<sup>92</sup>

ANI-1x NN trained using on a subset (6276 configurations) of the full ANI-1x dataset,<sup>93</sup> containing only water clusters with energies and forces re-evaluated at DFTB(3ob) using the TorchANI<sup>94</sup> implementation with only O and H networks. All default parameters and training with forces using a batch size of 128, appropriate for fewer data. Best validation RMSE = 1.18 kcal mol<sup>-1</sup> after 218 epochs. HDNNP<sup>16</sup> trained on 7258 RPBE configurations from ref.<sup>95</sup> re-evaluated at DFTB(3ob) with n2p2<sup>96</sup> using the default parameters and symmetry functions.

```
1 import gaptrain as gt
2 gt.GTConfig.n_cores = 4
3
4 h2o = gt.System(box_size=[10, 10, 10])
5 h2o.add_solvent('h2o', n=20)
6
7 data, gap = gt.active.train_ii(h2o, method_name='dftb')
```

---

**Figure 7.** Input script required to train a bulk water model from scratch at the DFTB level using four CPU cores.

## 5. Acknowledgements

TAY and TJW acknowledge the EPSRC Centre for Doctoral Theory and Modelling in Chemical Sciences (EP/L015722/1). TAY studentship is generously supported by AWE. Ministry of Defence ©British Crown Copyright 2020/AWE. The authors acknowledge the use of the Cirrus UK National Tier-2 HPC Service at EPCC (<http://www.cirrus.ac.uk>) funded by the University of Edinburgh and EPSRC (EP/P020267/1). We thank Gábor Csányi for initial discussions.

## 6. References

- (1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation : From Algorithms to Applications*, 2nd ed.; Academic Press: Cambridge, Massachusetts, 2002.
- (2) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7* (2), e32131.
- (3) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab Initio Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc. Natl. Acad. Sci.* **2005**, *102* (19), 6654–6659.
- (4) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71* (1), 361–390.
- (5) Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152* (5), 050902.
- (6) Unke, O. T.; Koner, D.; Patra, S.; Käser, S.; Meuwly, M. High-Dimensional Potential Energy Surfaces for Molecular Simulations: From Empiricism to Machine Learning. *Mach. Learn. Sci. Technol.* **2020**, *1* (1), 013001.
- (7) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation Mechanism for the Direct Graphite-to-Diamond Phase Transition. *Nat. Mater.* **2011**, *10* (9), 693–697.
- (8) Sosso, G. C.; Miceli, G.; Caravati, S.; Giberti, F.; Behler, J.; Bernasconi, M. Fast Crystallization of the Phase Change Compound GeTe by Large-Scale Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2013**, *4* (24), 4241–4246.
- (9) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High Sp<sup>3</sup> Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, *120* (16), 166101.
- (10) Niu, H.; Bonati, L.; Piaggi, P. M.; Parrinello, M. Ab Initio Phase Diagram and Nucleation of Gallium. *Nat. Commun.* **2020**, *11* (1), 2654.
- (11) Cheng, B.; Mazzola, G.; Pickard, C. J.; Ceriotti, M. Evidence for Supercritical Behaviour of High-Pressure Liquid Hydrogen. *Nature* **2020**, *585* (7824), 217–220.
- (12) Deringer, V. L.; Bernstein, N.; Csányi, G.; Ben Mahmoud, C.; Ceriotti, M.; Wilson, M.; Drabold, D. A.; Elliott, S. R. Origins of Structural and Electronic Transitions in Disordered Silicon. *Nature* **2021**, *589* (7840), 59–64.
- (13) Ang, S. J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Gómez-Bombarelli, R. Active Learning Accelerates Ab Initio Molecular Dynamics on Reactive Energy Surfaces. *Chem* **2021**.
- (14) Cole, D. J.; Mones, L.; Csányi, G. A Machine Learning Based Intramolecular Potential for a Flexible Organic Molecule. *Faraday Discuss.* **2020**.
- (15) Ruffa, D. A.; Bruce Macdonald, H.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards Chemical Accuracy for Alchemical Free Energy Calculations with Hybrid Physics-Based Machine Learning / Molecular Mechanics Potentials. *bioRxiv* **2020**.
- (16) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401.
- (17) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56* (42), 12828–12840.
- (18) 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.
- (19) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8* (1), 13890.
- (20) 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.
- (21) 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.
- (22) Bartók, A. P.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *Int. J. Quantum Chem.* **2015**, *115* (16), 1051–1057.
- (23) 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.
- (24) 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. *J. Comput. Phys.* **2015**, *285*, 316–330.
- (25) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14* (3), 1153–1173.
- (26) 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.
- (27) Szlachta, W. J.; Bartók, A. P.; Csányi, G. Accuracy and Transferability of Gaussian Approximation Potential Models for Tungsten. *Phys. Rev. B* **2014**, *90* (10), 104108.
- (28) Deringer, V. L.; Csányi, G. Machine Learning Based Interatomic Potential for Amorphous Carbon. *Phys. Rev. B* **2017**, *95* (9), 094203.
- (29) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8* (4), 041048.
- (30) 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), 104.
- (31) Mocanu, F. C.; Konstantinou, K.; Lee, T. H.; Bernstein, N.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Modeling the Phase-Change Memory Material, Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub>, with a Machine-Learned Interatomic Potential. *J. Phys. Chem. B* **2018**, *122* (38), 8998–9006.
- (32) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152* (20), 204110.
- (33) Veit, M.; Jain, S. K.; Bonakala, S.; Rudra, I.; Hohl, D.; Csányi, G. Equation of State of Fluid Methane from First Principles with Machine Learning Potentials. *J. Chem. Theory Comput.* **2019**, *15* (4), 2574–2586.
- (34) Deringer, V. L.; Caro, M. A.; Csányi, G. A General-Purpose Machine-Learning Force Field for Bulk and Nanostructured Phosphorus. *Nat. Commun.* **2020**, *11* (1), 5461.
- (35) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. Ab Initio Thermodynamics of Liquid and Solid Water. *Proc. Natl. Acad. Sci.* **2019**, *116* (4), 1110–1115.
- (36) Schran, C.; Brezina, K.; Marsalek, O. Committee Neural Network Potentials Control Generalization Errors and Enable Active Learning. *J. Chem. Phys.* **2020**, *153* (10), 104105.
- (37) Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R. Accelerating Crystal Structure Prediction by Machine-Learning Interatomic Potentials with Active Learning. *Phys. Rev. B* **2019**, *99* (6), 064114.
- (38) Artrith, N.; Behler, J. High-Dimensional Neural Network Potentials for Metal Surfaces: A Prototype Study for Copper. *Phys. Rev. B* **2012**, *85* (4), 045439.
- (39) Gubaev, K.; Podryabinkin, E. V.; Hart, G. L. W.; Shapeev, A. V. Accelerating High-Throughput Searches for New Alloys with Active Learning of Interatomic Potentials. *Comput. Mater. Sci.* **2019**, *156*, 148–156.
- (40) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using Metadynamics to Build Neural Network Potentials for Reactive Events: The Case of Urea Decomposition in Water. **2020**.
- (41) 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.
- (42) Podryabinkin, E. V.; Shapeev, A. V. Active Learning of Linearly Parametrized Interatomic Potentials. *Comput. Mater. Sci.* **2017**, *140*, 171–180.
- (43) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active Learning of Uniformly Accurate Interatomic Potentials for Materials Simulation. *Phys. Rev. Mater.* **2019**, *3* (2), 023804.
- (44) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; E, W. DP-GEN: A Concurrent Learning Platform for the Generation of Reliable Deep Learning Based Potential Energy Models. *Comput. Phys. Commun.* **2020**, *253*, 107206.
- (45) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001.
- (46) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-Fly Active Learning of Interpretable Bayesian Force Fields for Atomistic Rare Events. *npj Comput. Mater.* **2020**, *6* (1), 20.
- (47) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Phys. Rev. Lett.* **2019**, *122* (22), 225701.
- (48) 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.
- (49) 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. Commun.* **2020**, *11* (1), 5713.
- (50) Kearnes, S. Pursuing a Prospective Perspective. **2020**.
- (51) Cawley, G. C.; Talbot, N. L. C. On Over-Fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

- (52) Kramer, C.; Gedeck, P. Leave-Cluster-Out Cross-Validation Is Appropriate for Scoring Functions Derived from Diverse Protein Data Sets. *J. Chem. Inf. Model.* **2010**, *50* (11), 1961–1969.
- (53) Li, Y.; Yang, J. Structural and Sequence Similarity Makes a Significant Impact on Machine-Learning-Based Scoring Functions for Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2017**, *57* (4), 1007–1012.
- (54) Chen, L.; Cruz, A.; Ramsey, S.; Dickson, C. J.; Duca, J. S.; Hornak, V.; Koes, D. R.; Kurtzman, T. Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening. *PLoS One* **2019**, *14* (8), e0220113.
- (55) Kovacs, D. P.; Mccorkindale, W.; Lee, A. A. Quantitative Interpretation Explains Machine Learning Models for Chemical Reaction Prediction and Uncovers Bias. *ChemRxiv* **2020**.
- (56) Sheridan, R. P. Time-Split Cross-Validation as a Method for Estimating the Goodness of Prospective Prediction. *J. Chem. Inf. Model.* **2013**, *53* (4), 783–790.
- (57) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87* (18), 184115.
- (58) Pickard, C. J.; Needs, R. J. Ab Initio Random Structure Searching. *J. Phys. Condens. Matter* **2011**, *23* (5), 053201.
- (59) Mahoney, M. W.; Drineas, P. CUR Matrix Decompositions for Improved Data Analysis. *Proc. Natl. Acad. Sci.* **2009**, *106* (3), 697–702.
- (60) Bernstein, N.; Csányi, G.; Deringer, V. L. De Novo Exploration and Self-Guided Learning of Potential-Energy Surfaces. *npj Comput. Mater.* **2019**, *5* (1), 99.
- (61) Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. Machine-Learning Approach for One- and Two-Body Corrections to Density Functional Theory: Applications to Molecular and Condensed Water. *Phys. Rev. B* **2013**, *88* (5), 054104.
- (62) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The  $\Delta$ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2087–2096.
- (63) Zheng, L.; Chen, M.; Sun, Z.; Ko, H.-Y.; Santra, B.; Dhuvad, P.; Wu, X. Structural, Electronic, and Dynamical Properties of Liquid Water by Ab Initio Molecular Dynamics Based on SCAN Functional within the Canonical Ensemble. *J. Chem. Phys.* **2018**, *148* (16), 164505.
- (64) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117* (3), 1564–1686.
- (65) Ohtaki, H.; Radnai, T. Structure and Dynamics of Hydrated Ions. *Chem. Rev.* **1993**, *93* (3), 1157–1204.
- (66) Pratihari, S.; Ma, X.; Homayoon, Z.; Barnes, G. L.; Hase, W. L. Direct Chemical Dynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139* (10), 3570–3590.
- (67) Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angew. Chem. Int. Ed.* **2008**, *47* (40), 7592–7601.
- (68) Xie, J.; Hase, W. L. Rethinking the SN2 Reaction. *Science (80-. )*. **2016**, *352* (6281), 32–33.
- (69) Szabó, I.; Czakó, G. Dynamics and Novel Mechanisms of S N 2 Reactions on Ab Initio Analytical Potential Energy Surfaces. *J. Phys. Chem. A* **2017**, *121* (47), 9005–9019.
- (70) Lee, S.; Goodman, J. M. Rapid Route-Finding for Bifurcating Organic Reactions. *J. Am. Chem. Soc.* **2020**, *142* (20), 9210–9219.
- (71) Caramella, P.; Quadrelli, P.; Toma, L. An Unexpected Bispericyclic Transition Structure Leading to 4+2 and 2+4 Cycloadducts in the E Ndo Dimerization of Cyclopentadiene. *J. Am. Chem. Soc.* **2002**, *124* (7), 1130–1131.
- (72) Hjørth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermod, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; 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.
- (73) Young, T.; Johnston-Wood, T. Gap-Train. <https://github.com/t-young31/gap-train> 2020.
- (74) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. Dscribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
- (75) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362.

- (76) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshayé, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irlé, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Rezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W. -z.; Frauenheim, T. DFTB+, a Software Package for Efficient Approximate Density Functional Theory Based Atomistic Simulations. *J. Chem. Phys.* **2020**, *152* (12), 124101.
- (77) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9* (1), 338–354.
- (78) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652–1671.
- (79) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-Space Grid Implementation of the Projector Augmented Wave Method. *Phys. Rev. B* **2005**, *71* (3), 035109.
- (80) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Duřak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsarlis, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic Structure Calculations with GPAW: A Real-Space Implementation of the Projector Augmented-Wave Method. *J. Phys. Condens. Matter* **2010**, *22* (25), 253202.
- (81) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.
- (82) Zhang, Y.; Yang, W. Comment on “Generalized Gradient Approximation Made Simple.” *Phys. Rev. Lett.* **1998**, *80* (4), 890–890.
- (83) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110* (13), 6158–6170.
- (84) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (85) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; TAILLEFUMIER, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöb, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An Electronic Structure and Molecular Dynamics Software Package - Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.* **2020**, *152* (19), 194103.
- (86) Neese, F. The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2* (1), 73–78.
- (87) Neese, F. Software Update: The ORCA Program System, Version 4.0. *WIREs Comput. Mol. Sci.* **2018**, *8* (1), 1–6.
- (88) Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. AutodE: Automated Calculation of Reaction Energy Profiles—Application to Organic and Organometallic Reactions. *Angew. Chemie Int. Ed.* **2020**, anie.202011941.
- (89) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297.
- (90) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
- (91) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (92) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (93) 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.
- (94) 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.

- (95) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van Der Waals Interactions Determine the Unique Properties of Water. *Proc. Natl. Acad. Sci.* **2016**, *113* (30), 8368–8373.
- (96) Singraber, A.; Morawietz, T.; Behler, J.; Dellago, C. Parallel Multistream Training of High-Dimensional Neural Network Potentials. *J. Chem. Theory Comput.* **2019**, *15* (5), 3075–3092.