THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

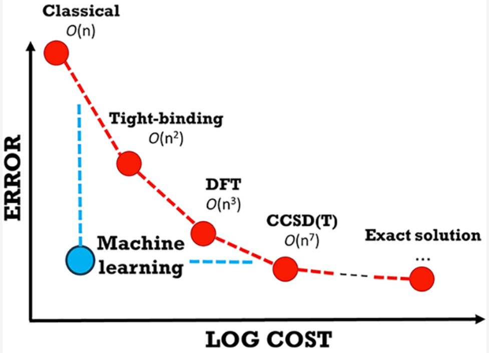
The Rise of Neural Networks for Materials and Chemical Dynamics

Maksim Kulichenko, Justin S. Smith, Benjamin Nebgen, Ying Wai Li, Nikita Fedik, Alexander I. Boldyrev, Nicholas Lubbers, Kipton Barros, and Sergei Tretiak*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 6227−6243 Read Online

ACCESS | Metrics & More | Article Recommendations

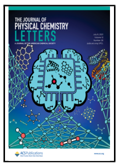
ABSTRACT: Machine learning (ML) is quickly becoming a premier tool for modeling chemical processes and materials. ML-based force fields, trained on large data sets of high-quality electron structure calculations, are particularly attractive due their unique combination of computational efficiency and physical accuracy. This Perspective summarizes some recent advances in the development of neural network-based interatomic potentials. Designing high-quality training data sets is crucial to overall model accuracy. One strategy is active learning, in which new data are automatically collected for atomic configurations that produce large ML uncertainties. Another strategy is to use the highest levels of quantum theory possible. Transfer learning allows training to a data set of mixed fidelity. A model initially trained to a large data set of density functional theory calculations can be significantly improved by retraining to a relatively small data set of expensive coupled cluster theory calculations. These advances are exemplified by applications to molecules and materials.



Computational atomistic modeling has become a crucial element in our discovery and understanding of the fundamental properties of natural and human-made materials. In the early decades of computing, atomistic computation studied simple model systems and small quantum mechanical (QM) systems. Future advances in science and technology will require accurate modeling capabilities for ever larger and more complex molecules and materials. However, modeling of even medium-sized molecules within sophisticated quantum mechanical methods such as *ab initio* techniques[1,2] involves significant computational effort, and large systems of practical interest are out of reach. As a rule of thumb (see the abstract graphic), computational chemistry methods that approach the exact quantum mechanical solution for an electronic system have computational costs that grow very rapidly with system size. An important benchmark for QM methods is to achieve so-called chemical accuracy, errors of approximately 1 kcal/mol, which is the energy scale associated with thermal fluctuations at ambient temperatures. Benchmark *ab initio* methods such as coupled cluster single−double with perturbative triples[2] [CCSD(T)] provide such accuracy without empirical fitting parameters and directly capture physical interactions such as electrostatic interactions and dynamic electron correlation. Approximating the Schrödinger equation at the CCSD(T) level of accuracy requires computing a massive number of many-center electron integrals and repeated diagonalization of the very large self-consistent Hamiltonian matrix. The computational cost of CCSD(T) scales as $O(N^7)$, which limits practical system sizes to perhaps $N \sim 10^2$ atoms. Density functional theory (DFT) models[3] can reduce this scaling to $O(N) - O(N^3)$, underpinning the practical success of this approach. However, relative to CCSD(T), the inexact functionals of DFT limit its accuracy, and the computational costs of DFT can still be considerable for systems with >10^3 atoms. To further reduce the computational cost, approaches utilizing effective Hamiltonian models such as semiempirical methods, e.g., Austin model 1 (AM1)[4] or density functional tight binding (DFTB),[5] neglect three- and four-atom center integrals at the cost of accuracy. The remaining two-electron integrals are replaced by empirical parameters to account for the neglected behavior. The significant gain in computational efficiency, however, may require compromises with respect to accuracy. Quite often, properly adjusted parameters work well within a certain range of chemical systems, but the transferability of such methods is not always obvious and should be studied for every specific case.

Modeling of dynamical processes (chemical reactions, shocks, protein folding, and phase transitions in materials, to name a few) requires large-scale molecular dynamics (MD) simulations. Here finite-temperature dynamical trajectories sample potential energy surfaces (PESs) (defined by the energy of a system as a function of nuclear coordinates or geometry) and are generated

Received: April 26, 2021
Accepted: June 9, 2021
Published: July 1, 2021



https://doi.org/10.1021/acs.jpclett.1c01357

by using forces (i.e., gradients of the energy) calculated "on the fly". *Ab initio* MD (AIMD) uses forces computed from electronic structure calculations. While accurate, AIMD is numerically expensive as outlined above, which severely limits its applications in terms of system size and time scales. A more dramatic computational simplification is to neglect quantum mechanics entirely with the use of classical force fields, which approximate the system as a classical "beads and springs" model with additional terms for Coulomb and dispersion interactions. These models typically exhibit $O(N)$ scaling with a low prefactor facilitating MD simulations of systems with millions or even billions of atoms, from which thermodynamic properties as well as non-equilibrium processes can be directly computed. Classical force fields traditionally assume a fixed, physically motivated functional form. The total energy of a system is split into bonded terms for covalently bonded atoms and nonbonded terms. The bonded terms, in turn, consist of three components (bond lengths, bond angles, and dihedral angles), and each term has empirical parameters. The parameters for each term depend not only on the chemical elements in the bonds but often on the local chemical environment. Different carbon−carbon bonds, for example, will be assigned different parameters to represent single or double bonds, or ring versus chain topologies. Determining which parameters can be shared or must be different can be a laborious process and is often arbitrary. This problem persists for nonbonded terms, due to the necessity of determining a reasonable atomic charge for ionic compounds or polar molecules. One strong disadvantage of force fields is that the bonding-oriented modeling approach limits their applicability to nonreactive conditions. Thus, they are not reliable for investigations of, for example, reaction pathways and transition states or generally dynamics far from equilibrium. There are more expensive force field models, such as ReaxFF,[6] that are capable of handling transition states and chemical reactions, when explicitly parametrized to a specific set of reactions. While the scalability of force fields is excellent, their accuracy and transferability are severely constrained. This necessitates large amounts of human time and effort spent by researchers to adapt their parameters directly to the system under study prior to scientific simulations of complex or novel chemicals and materials.

Machine learning (ML)-based potentials are bridging the gap between highly accurate quantum mechanics simulations and the affordable, but less transferable, classical force field approaches.[7−9] ML workflows establish, by definition, an empirical model of a data set and are capable emulating the underlying electron structure calculations to very high levels of accuracy. A broad variety of ML approaches have been developed, and the remarkable success across a variety of chemistry and materials fields has been remarkable.[10−24] At the highest level, all ML-based potentials can be understood as intelligent schemes for interpolating on the training data. To maximize the transferability of the ML model, this interpolation happens in an abstract space, in which the representation of the training data point has undergone a transformation, usually a nonlinear transformation to a high-dimensional space. One broad class of ML algorithms consists of so-called kernel methods; here, the ML practitioner specifies a kernel function that serves as a measure of similarity between two inputs (e.g., two local atomic environments).[25−28] The greater this similarity measure, the closer the model outputs are expected to be. In kernel ridge regression (KRR) or the closely related Gaussian process regression (GPR), model complexity naturally grows with data set size. A disadvantage of such nonparametric modeling, however, is that the ML models can become very computationally expensive for large data sets, both in the training and in the application phases. The technique of sparsification can significantly reduce the computational costs for kernel-based modeling. Alternatively, it can be advantageous to assume some fixed (parametric) functional form for the model, typically with a large number of fitting parameters. Standard examples of parametric models include, e.g., linear regression, and its generalization to neural networks. In this context, a large area of research is the design of good descriptors (also known as feature vectors) that highlight the important chemical features in an input data point (i.e., a local atomic environment). With very carefully designed descriptors,[25,29−31] relatively simple ML models, such as linear regression, can achieve very high accuracies. Alternatively, with neural networks, the input training data can be supplied in a relatively simple representation, and the burden of good descriptor design may be shifted into the automated ML training procedure.

Machine learning (ML)-based potentials are bridging the gap between highly accurate quantum mechanics simulations and the affordable, but less transferable, classical force field approaches.

In this Perspective, we focus on the development and application of neural network (NN)-based interatomic potentials as surrogate energy models to *ab initio* methods, replacing the conventional classical force fields and enabling large-scale MD simulations.[6,32,33] This development promises to provide MD simulations with much improved accuracy and transferability, while preserving $O(N)$ numerical scaling. NNs are highly flexible, nonlinear functions with thousands to millions of parameters that are optimized to fit a data set. A large number of parameters provides many degrees of freedom, allowing an optimization algorithm to decide the best mapping that transforms the input into an estimation of a desired property. The optimization algorithm requires a training set, which is a set of inputs (e.g., molecular geometries) and corresponding labels (e.g., reference energies and/or forces). An important concept in machine learning is generalization error. Machine learning models of all forms can suffer from overfitting, a condition in which the model learns to fit the training data but does not make similarly accurate predictions about new data. In

In this Perspective, we focus on the development and application of neural network (NN)-based interatomic potentials as surrogate energy models to *ab initio* methods, replacing the conventional classical force fields and enabling large-scale MD simulations.

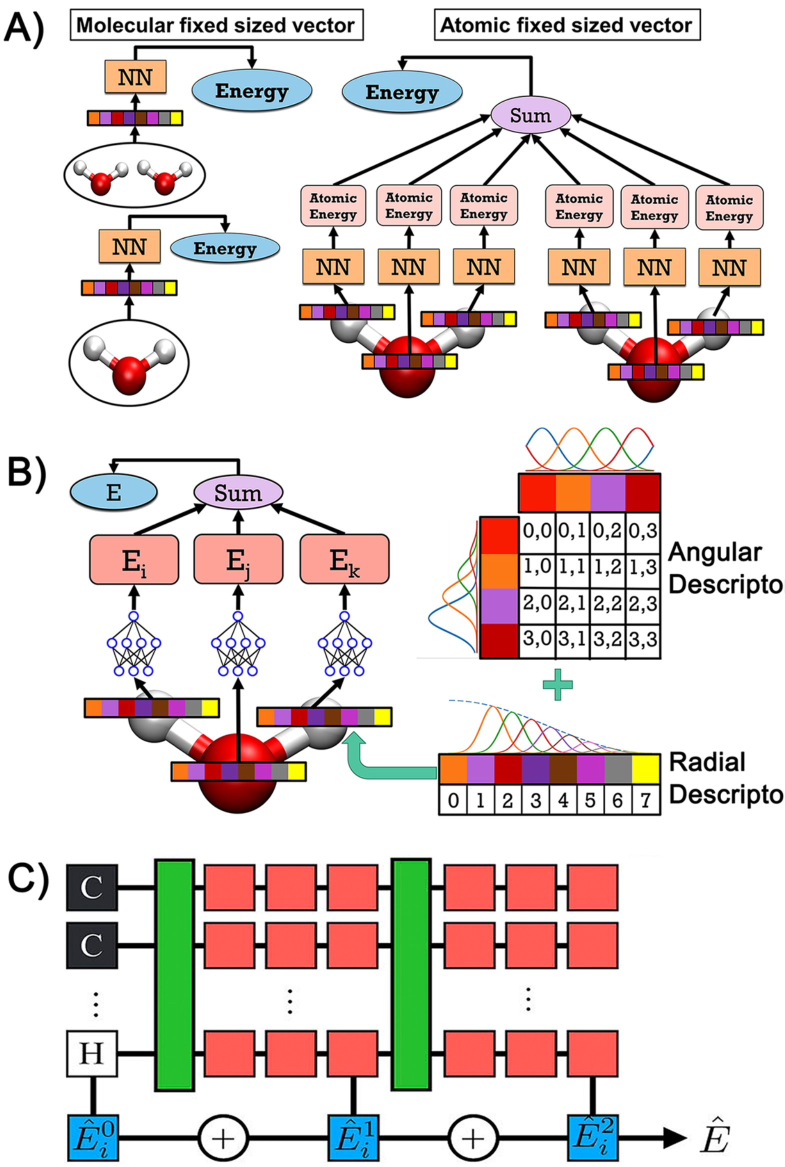


Figure 1. (A) Non-extensible fixed input size NN with the entire system serving as input into one network (left) and extensible Behler−Parrinello type NN with a distinct network for each chemical element environment (right). (B) **ANI** NN architecture. There is a distinct NN for each chemical element, and the total energy is a sum of NN outputs. The input is a vector of radial symmetry functions and a matrix of angular symmetry functions. Descriptors differentiate between different elements within a cutoff. (C) **HIP-NN** architecture. The input is atomic species reflecting dressed atom approximation in $\hat{E}_i^0$. Interaction layers (green boxes) transmit information between atoms within a local neighborhood. The energy terms decrease hierarchically as $\hat{E}_i^0 < \hat{E}_i^1 < \hat{E}_i^2$. Reproduced with permission from ref 73. Published 2018 American Institute of Physics.

the context of ML-based potentials, we can measure model accuracy in different ways. One important measure is size extensibility, the ability of the model to make accurate predictions for systems much larger than those observed in the training process. Another important measure is chemical transferability, the ability of the model to make accurate predictions for systems that are structurally distinct from those observed during training.

Neural Networks in Brief. In their barest form, NNs can be treated as black boxes that are trained to learn the mapping between inputs (here, geometry of a local atomic environment) and outputs (here, molecular energy). At the core of neural networks are artificial neurons. Each neuron forms a cheap, parametrizable nonlinear function from several inputs, to a single output, or activation, using a set of weights and bias parameters and an activation function. The activations are organized into layers. The input layer accepts feature vectors (for our purposes, a set of numbers containing information describing the molecular composition and geometry). In the first layer, a set of neurons take this feature vector and compute a vector of activations, giving rise to a hidden layer (hidden, because its values are not directly constrained by the input or outputs in the data set) whose activations constitute a new, processed feature vector. This new set of information is then passed to another layer of neurons, and so on, until the final output layer, which constitutes the final output of the network (e.g., predicting molecular energy) .[34]

When first initialized with random parameters, a neural network will not compute any particularly meaningful function. The parameters of the network need to be optimized to perform the task at hand. This is accomplished via a loss function. This loss function is a scalar quantifying the overall difference between the predicted and reference values. A typical loss function is the mean squared error between the predictions and true reference values (e.g., the result of *ab initio* simulation). To train the network, batches of examples from the data set are presented, and the prediction and associated loss function are computed. The key step to updating the network is the gradient of the loss function of the parameters; by taking the gradient of the loss function and stepping the weights and biases of all neurons subtly in the direction that reduces the loss, one can improve the network. This process is repeated over the data set over many iterations, sometimes millions, until a satisfactory match to the reference data set is constructed. To avoid overfitting to the training data, a common strategy is early stopping, whereby the neural network training procedure is stopped when model performance is no longer improving on a held-out validation data set.

A key algorithm called backpropagation (i.e., reverse mode automatic differentiation)[35−37] is used to make this process efficient. Naïve gradient algorithms such as finite differencing scale very poorly with respect to the number of parameters with which the gradient must be taken. In contrast, backpropagation allows for the simultaneous calculation of all parameter gradients to floating-point accuracy. This operation is computationally efficient, as the time scales in precisely the same way as the computation of the network output. The key of the method is to harness the recursive formulation of the multivariate chain rule along with dynamic programming to iteratively calculate gradients of all variables backward, starting from the loss, proceeding to the final layers of the network, through the hidden layers, and eventually back to the network inputs. Automatic differentiation plays another key role for neural network potentials. The algorithm can be used to compute the gradient of a network trained to produce energy with respect to the input positions, yielding the force corresponding to the energy prediction. In a suitable programming framework, automatic differentiation thus does not require a separate set of codes to predict force and energy; the force can be cleanly expressed as the gradient of the energy. Similarly, training to forces requires gradients of a loss function that includes these forces, and successive calls to automatic differentiation can accomplish this. As such, in many cases there is very little programming needed to promote a model from energy prediction and training to force prediction and training, and it is not difficult to define models whose forces conserve energy (see recent reviews[38,39] for a more detailed discussion of ML potential architectures).

Beyond these generalities, neural networks come in many flavors. Perhaps the most salient is the architecture of the network, the specification of the overall structure of the input features, neuron functional forms, and interconnections between neurons, rather than the individual parameter values.

Neural Network Architectures for Potentials. The first successful attempts to apply NNs in chemistry and physics were made in the late 1980s and early 1990s. These include the analysis of nuclear magnetic resonance (NMR),[40] mass spectra,[41] and the predictions of protein structures.[42,43] Since then, ML has been actively developed to predict total[44−48] and atomization energies,[49,50] forces,[51−53] dipole moments,[47,54,55] assignment of atomic charges,[56,57] chemical reactions,[58,59] new materials,[60−62] etc. In 1995, Blank et al., for the first time, implemented NNs to establish a structure−energy relationship.[63] The input geometry representation of a system was the first challenge that early research efforts faced when dealing with NN potentials. Pioneering works were devoted to simple molecular systems of a fixed size like diatomic molecules adsorbed on crystalline surfaces[63,64] and water dimers.[65] Direct geometric parameters of a system such as bond lengths and angles or simply Cartesian coordinates served as fixed-size inputs for the NN (Figure 1A, left). This implies explicit limitation to the extensibility of earlier ML potentials: NNs can work only with inputs of a fixed size, because the addition of more atoms increases the number of input neurons, introducing new, unfitted parameters into the system.

Efforts by Hobday et al. in 1999 introduced a neural network architecture that offered size-extensible predictions for hydrocarbons by providing a bond-centered network; local energy predictions are made for each pair of atoms within a fixed cutoff radius from each other.[66] In 2007, Behler and Parrinello proposed an extensible NN representation for high-dimensional PESs.[34] The main idea is that the total energy can be represented by a sum of effective individual atomic contributions (Figure 1A, right), that is, $E = \sum_i^N E_i$. Here, N is the number of atoms in a system. This atom-centered notion of system energy remains the leading technique for building size-extensible neural networks, although bond-centered and other approaches have not been abandoned.[67,68] While intuitive and largely successful, the general strategy of decomposing energy as a sum of local contributions may not be fully consistent with the underlying quantum mechanics. Behler also articulated three criteria for molecular descriptors that machine learning potentials should satisfy: rotational and translational invariance; the exchange of two identical atoms should yield the same result; and the representation (input vector) should describe a molecule's geometry in a unique way given a set of atomic positions and types.[69] Combined with a local energy decomposition approach, these principles have led to modern NN potentials, which provide physical guarantees on the functional form of the energy without eradicating the flexibility of the original black-box approaches and the many approximations of classical force fields.

These principles were used to construct the high-dimensional symmetry function (SF) approach to featurizing atomic environments. Here, the Cartesian coordinates are transformed into a set of SF values $\{G_i^\mu\}$, where μ indexes the various symmetry functions and i indexes the atoms. The set of $\{G_i^\mu\}$ hence describes the atomic environment of each atom in the system. There are two types of SFs, the radial, or two-body, SF, which describes the distances between atom i and all neighboring atoms j; and the angular, or three-body, SF, which provides information about angles between atom i and pairs of neighboring atoms j and k. In both types, spatial locality is ensured by a smooth radial cutoff (R_c) function:

$$f_c = \begin{cases} 0.5 \times \cos\left(\dfrac{\pi R_{ij}}{R_c}\right) + 0.5, \text{ for } R_{ij} \le R_c \\ 0, \text{ for} R_{ij} > R_c \end{cases}$$

which reflects the decrease in the level of interaction as the distance R_{ij} between two atoms increases. The radial atomic environment of atom i is probed by a vector $G_m^R = \sum_{j \neq i}^{\text{all atoms}} \exp[-\eta_m (R_{ij} - R_m^S)^2] f_c(R_{ij})$, where m indexes a set of hyperparameters η and R_s. Altogether, these quantities define an atomic environment vector (AEV) for every atom with sizes ranging from tens to thousands of elements depending on the particular implementation.[70,71] Originally, the R_m^S parameters were set to a constant 0, varying only η_m to change the width of the Gaussian for different SF probes. Later, varying R_m^S enabled the probing of specific radial shells with a constant η set such that the Gaussian probes overlap. This original formulation limits the number of distinct chemical elements [such as hydrogen (H), carbon (C), or oxygen (O)] that NNs can learn, because this NN architecture implies a separate NN for each chemical element. Additionally, for N chemical elements, the complexity of angular descriptors scales as $N(N + 1)/2$.

Although the Behler−Parrinello SFs solved the problem of size extensibility in many cases, it did not lead to chemical transferability of the learned potentials. This problem might have two causes. First, it is possible that the original SFs are not descriptive enough to recognize the common spatial patterns of atoms (e.g., rings, heteroatomic bonds, functional groups, etc.) in the molecular representation, a reason that hinders learning interactions in one molecule and then applying this knowledge for another molecule. Second, the original SFs do not distinguish between different chemical elements in the summation over the neighboring atoms. Therefore, the individual chemical element specific NN potentials are unable to differentiate between atom types within a given cutoff distance. These limitations restrict the original applications of SFs to systems with few atom types or small single-molecule data sets.

The *ANI-1* NN model introduced in 2017 and constructed using the ANI architecture[71] (Figure 1B) modified the Behler−Parrinello SFs to address chemical transferability and build, arguably, the first extensible potential. The modified angular symmetry functions were designed to spatially localize the description of the angular environment of each atom within the cutoff radius in a manner similar to varying the R_m^S parameters, which spatially localize the radial environment within the local cutoff. The ANI architecture also employs a specific set of symmetry functions to provide better transferability on large and diverse data sets by opting for maximum spatial locality for each symmetry function in the AEV. Maximizing the symmetry function spatial locality makes distinct features in a chemical environment more recognizable, thus improving transferability. ANI also provides radial and angular descriptors for each distinct chemical element (and pairs of elements) that might be present within a given cutoff radius. In other words, it is able to differentiate atomic numbers in the local environment of atom *i*. Numerous tests have shown that ***ANI-1*** is an extensible potential for organic molecular compounds containing the elements C, H, N, and O that reaches the accuracy of DFT used as a reference method. Herein we will use bold italic notation for NN potentials (such as ***ANI-1***) to emphasize specific ML models capable of describing PESs. ANI-1 then stands for a respective data set used to train the ***ANI-1*** potential.

Further developments in atomistic NN architectures, pioneered by Schütt et al. in the deep tensor neural network (*DTNN*),[44] have focused on a notion of end-to-end learning, also adopted prominently by *SchNet*[72] and *HIP-NN*[73] (Figure 1C). In loose terms, it corresponds to the replacement of the SF approach with the concept of an interaction layer that plays a role similar role to that of the SFs, but with two important differences. First, the parameters of these layers are fully learnable, allowing the characterization of the atomic environment itself to be dynamically adjusted during training. Second, these layers are designed to be stacked, allowing later layers to characterize the atomic environment of an atom by adding information about neighbors. This cascading approach allows the energy function to remain local but accounts for the larger environment in terms of a multiple of the cutoff length used for a single interaction layer. ***HIP-NN***, the hierarchically interacting particle neural network (Figure 1C), further partitions atomic energy contributions for each interaction layer, such that the total energy $E = \sum_i^N E_i$ is further decomposed as $E_i = \sum_{n=0}^{\text{Ninteraction}} E_i^n$, with n indexing each level of interaction. The network can be regularized to statistically reflect the notion that $E_i^0 > E_i^1 > ... > E_i^N$, as in a series approximation such as the many-body expansion. The zeroth energy term corresponds to the dressed atom approximation[67] because the input to the zeroth layer is simply atomic species. Further terms refine early predictions within higher-order functions, and the hierarchical assumption corresponds to the notion that simpler functions should be able to account for much of the remaining energy to fit; the last and therefore highest-complexity neurons correspond to large many-body effects, which should not account for a large fraction of the system energy.

The NN architectures schematically presented in Figure 1 emphasize some commonly used principles for constructing ML interatomic potentials. One key point that is currently a subject of intense research is recapturing long-range interactions such as Coulombic interactions, van der Waals forces, and dispersive interactions in these inherently local models.[74,75] Much work has focused specifically on including Coulombic interactions in ML potential models.[47,76,77] Other recent work has focused on developing a NN potential for modeling the dynamics of electronically excited states.[78] These models are required to predict not only energies and forces for all excited states of interest but also all couplings to model transitions between states. The current state-of-the-art in this area is the SchNarc model, which was shown to very accurately model the exited state dynamics on the methylenimmonium cation and thioformaldehyde, independently. While excited state ML potentials are very promising, similar to the early days of ground state ML potential development, they are currently limited to modeling a single system at a time requiring retraining from application to application. This also limits the size of molecular systems that can be modeled using such an approach, because building an extensible potential is currently feasible. As the field evolves (see a recent review[79]), we hope to see workflows extended in such a way as to capture general excited state potential energy surfaces covering large swaths of chemical space simultaneously.

> One key point that is currently a subject of intense research is recapturing long-range interactions such as Coulombic, Van der Waals forces and dispersive interactions in these inherently local models.

Data Set Diversity. Having a good NN architecture is important, but the largest factor in ML model accuracy is the quality and diversity of the training data. As the use of ML potentials in chemistry has grown, there has been an increasing focus on methodologies for generating training data sets. Many early ML potential development efforts were focused on specific applications. For example, a small data set of hundreds to thousands of conformations can be generated for the same

> Having a good NN architecture is important, but the largest factor in ML model accuracy is the quality and diversity of the training data.

molecule and used to train a machine learning potential to study that specific system.[47,80] An example from materials science is the generation of data from selected crystal structures of interest.[34,70,81] This specialization approach often leads to very accurate fits because of the narrow scope of the data. Once trained, the ML model provides a very fast prediction of energies and forces that can be applied in MD. However, this approach to data essentially scales the same as the underlying QM method used for reference data generation because new QM data must be calculated for any system of interest. Researchers also require expertise in fitting ML potentials to use specialized ML potentials.

A recent focus has been the development of ML potentials that can be applied to broad classes of molecules (e.g., organic molecules)[47,82] and materials.[83,84] A general-use ML potential should be applicable to systems larger than those in the training data set (size extensibility) and accurately describe a wide variety of configurations and conformations of the elements in the original training set (transferability). Thus, an adequate coverage of the problem space of interest defining the diversity of training sets is critical. To this end, we delineate the chemical and conformational diversities of the data sets. The former refers to the coverage of molecular space, chemical elements, and chemical bonding patterns, whereas the latter alludes to how well the PES of a system is represented via non-equilibrium geometry samples. Both diversities are critical for obtaining a truly general-use NN potential valid for non-equilibrium situations and for generally addressing reactive chemistry processes: the model should be trained to a data set that covers the broadest possible range of local interactions and spatial patterns.

> A general-use ML potential should be applicable to systems larger than those in the training data set (size extensibility) and accurately describe a wide variety of configurations and conformations of the elements in the original training set (transferability).

To gain a sense of combinatorial difficulty and chemical diversity, one may consider the magnitude of the GDB-17 database, created by the Chemical Space Project.[85] This database contains 166.4 billion organic molecules of ≤17 atoms of C, N, O, S, and halogens (plus the necessary hydrogens to saturate unfilled bonds). Inspired by GDB-17, smaller data sets of organic molecules have become popular for benchmarking ML potentials. For example, the QM-9[86] data set has ∼134K stable small organic molecules (C, H, O, N, and F elements) with up to nine non-hydrogen atoms. QM-9 contains relaxed equilibrium geometries and includes molecular properties like energy and dipole moment, calculated at the DFT level. Performance in predicting QM-9 molecular energies is a particular popular benchmark. Soon after this data set was introduced, models reached accuracies well below 1 kcal/mol, for example, as reported in refs 44, 87, and 88. The scatter plot in Figure 2A shows HIP-NN performance on this QM-9 benchmark. Here, HIP-NN molecular energy predictions are plotted against reference DFT calculations for ∼20K test molecules that were withheld from the training process. The scatter plot is barely distinguishable from a straight line, and the mean absolute error (MAE) of HIP-NN predictions is a remarkable 0.26 kcal/mol.[73]

Although QM-9 is a very popular benchmark, and a great first test for new ML model architectures, it lacks conformational diversity. A more recent alternative, the ANI-1 benchmark data set,[71] contains non-equilibrium conformations for ∼57K molecules that have up to eight non-hydrogen atoms (C, N, and O). Through sampling of the normal modes, a total of ∼20 million non-equilibrium molecular conformations are included in the ANI-1 data set, along with DFT molecular energy calculations. Trained to this data set, the ANI-1 neural network potential demonstrated a strong ability to be generalized to new molecules. Figure 2B illustrates the performance of the ANI-1 potential when tested on a data set of molecules containing up to 10 non-hydrogen atoms. Note that the molecules in the testing data were all larger than those in the training data. Nonetheless, the ANI-1 potential achieved near chemical accuracy, with a root-mean-square error (RMSE) of 1.8 kcal/mol, and the resulting potential provides a smooth PES for conformational changes, such a dihedral rotation in larger drug molecules (Figure 2C). The ANI-1 potential significantly outperforms popular semiempirical methods such as PM6 in reproducing DFT energies. In Figure 3C, we compare the diversity of chemical environments around H and C atoms in equilibrium QM-9 and non-equilibrium ANI-1 data sets. This visualization is generated using the t-distributed stochastic neighbor embedding (t-SNE)[89] technique for portraying multidimensional data in a reduced two-dimensional (2D) representation. Thus, each point in Figure 3C corresponds to H or C atomic environments in each distinct molecule in the data sets. As expected, ANI-1 is much more diverse than the QM-9 data set, which contains only equilibrium geometries.

NN architectures constructed using the framework described above have already delivered a wealth of important chemical information.[57,70,90,91] For example, another popular NN potential, ***TensorMol***, which was trained on ∼370K water clusters, perfectly describes the PES of breaking a hydrogen bond in a water dimer (Figure 3D).[47] Furthermore, the predicted dipoles of a system are in excellent agreement with DFT results (see Figure 4 in the original work).[47] Additionally, besides predicting total energies, ML potentials provide access to interesting atom-wise properties such as local chemical potentials. For example, the local chemical potential is defined as the energy of atom A at position **r** in molecule M. The local chemical potential isosurfaces (generated via the DTNN[44]) of the hydrogen test charge for some organic molecules are shown in Figure 2E. These isosurfaces enable the estimation of bond saturation and degrees of aromaticity. For instance, the DTNN predicts the relative aromaticity by comparing benzene and toluene. Furthermore, it is possible to estimate the stability of different fragments of a molecule because DTNN naturally provides the respective atomic energies. Indeed, $C_6O_3H_6$ has the

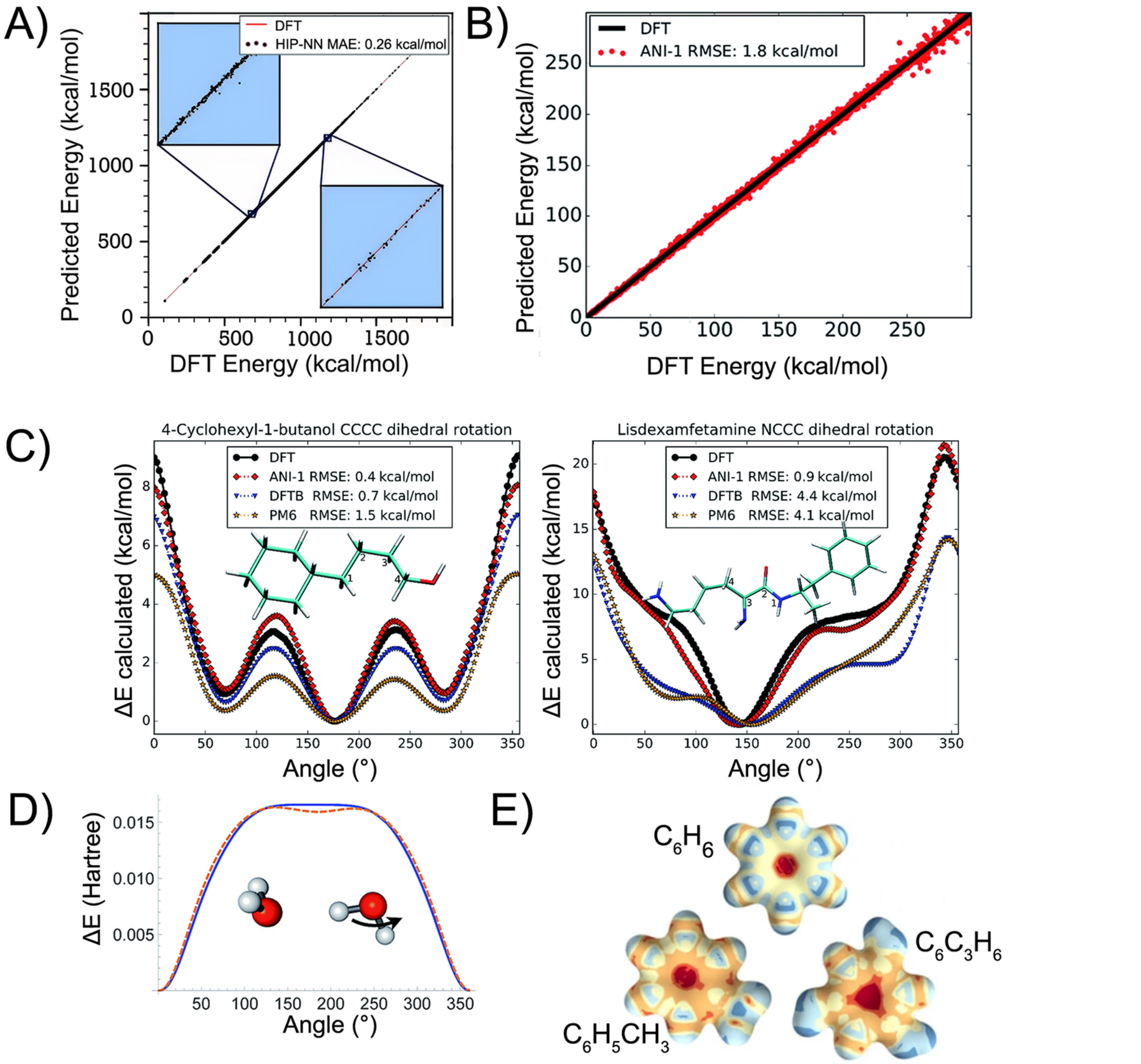


Figure 2. (A) Correlation plot between DFT energies and **HIPP-NN** predictions for the QM-9 database for equilibrium structures with an MAE of 0.26 kcal/mol. (B) Correlation plot between DFT energies and **ANI-1** NN predictions. Relative energy comparisons from random conformations of a random sampling of 134 molecules from GDB-11 all with 10 heavy atoms. There is an average of 62 conformations, and therefore energies, per molecule. None of the molecules from this set are included in any of the ANI training sets. Reproduced with permission from ref 71. Copyright 2017 The Royal Society of Chemistry. (C) One-dimensional potential surface scan generated from DFT, the **ANI-1** potential, and two popular semiempirical methods, DFTB and PM6. The atoms used to produce the scan coordinate are labeled in the images of the molecules in subplots. Reproduced with permission from ref 71. Copyright 2017 The Royal Society of Chemistry. (D) **TensorMol** PES of breaking a hydrogen bond between two waters by rotating one water around the O−H bond. The DFT (ωB97X-D/6-311G**) results are shown as a dashed orange line, and the **TensorMol** force field results are plotted as a solid blue line. Reproduced with permission from ref 47. Copyright 2018 The Royal Society of Chemistry. (E) DTNN local chemical potentials $\Omega C(\mathbf{r})$ of some organic molecules (using a hydrogen test charge on $\sum_i \|\mathbf{r} - \mathbf{r}_i\| = 3.7$ Å). Reproduced with permission from ref 44. Copyright 2017 Springer Nature Ltd.

most stable carbon ring in the GBD-9 database according to DTNN predictions.

Automated Data Set Construction. Importantly, generation of large training data sets for NNs, which are exemplified in Figure 2, requires a significant investment of computational time. This raises an important question of automatic generation of new data points, particularly addressing the underrepresented regions in the existing data set. Beyond direct reduction of the numerical efforts, this would also help to reduce human labor by minimizing the involvement of the researcher in the data generation process. Active learning (AL) aims to expand data sets based on iterative applications of sampling with uncertainty quantification (UQ)-driven selection of poorly represented atomic systems. Training labels (energies and forces) are generated for the UQ-selected atomic systems using *ab initio* simulations and then added to the training data set. The ML potential is retrained, and the sampling and UQ selection is repeated iteratively. AL enables a significant reduction in the number of *ab initio* simulations needed while ensuring the maximal diversity of a data set. In essence, AL helps to automate the development of ML potentials while also removing human bias in data selection. A simple but effective ensemble UQ strategy called query by committee[92] relies on the comparison of several NNs trained independently, allowing the ML model to

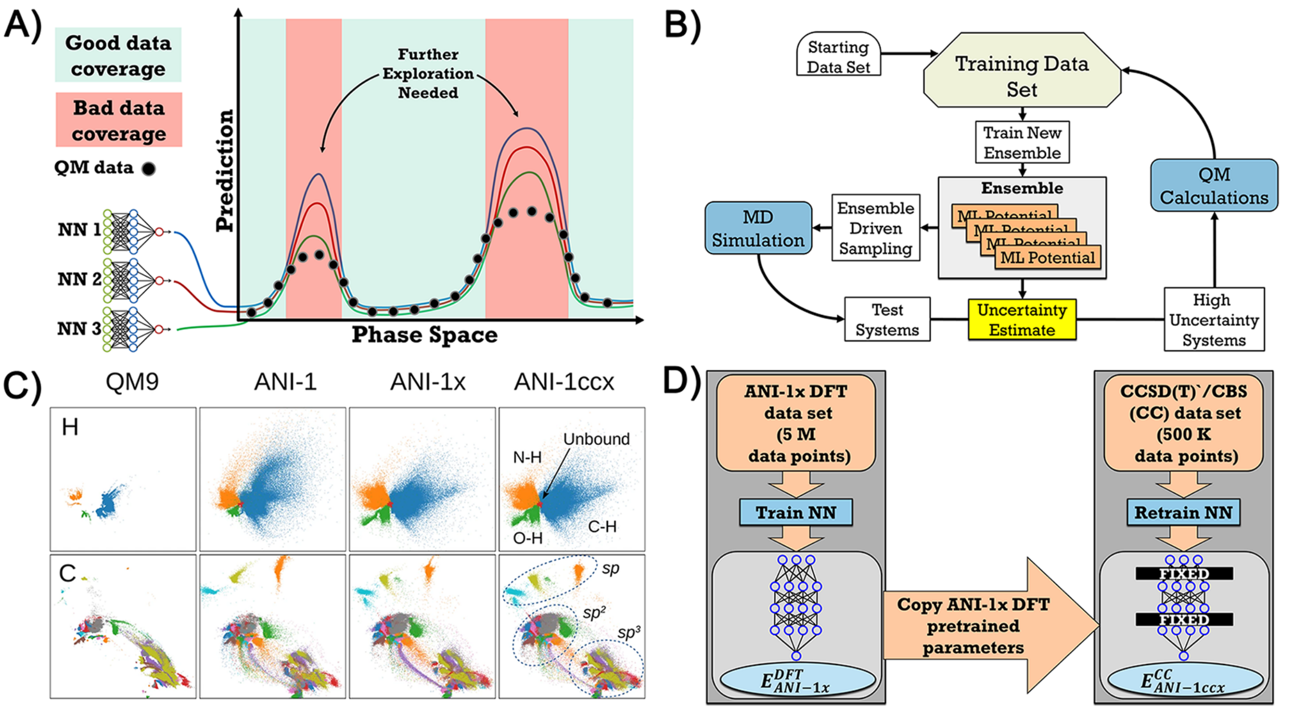


Figure 3. (A) Schematic representation of the ensemble deviation of NNs in poorly covered regions of phase space. (B) AL workflow for the MD sampler. (C) t-SNE representation of hydrogen and carbon environments in QM-9 (134K equilibrium structures), ANI-1 (20 million non-equilibrium structures), ANI-1x (5 million non-equilibrium structures), and ANI-1ccx (500K non-equilibrium structures) data sets. Reproduced with permission from ref 93. Copyright 2020 Springer Nature Ltd. (D) Example of a DFT → CCSD(T)/CBS transfer learning workflow.

> Active learning enables a significant reduction in the number of *ab initio* simulations needed while ensuring the maximal diversity of a dataset.

select needed data before running reference QM calculations (Figure 3A). The schematic workflow of AL is shown in Figure 3B. It is performed using an ensemble of NNs within, for example, an MD simulation. While MD is running with the existing pretrained ML potential, each member of the NN ensemble is making a prediction. If the NNs' predictions deviate too much from each other, the uncertainty of the model is deemed high for this part of the phase space (i.e., conformations) and the training set should be augmented near this data point. ANI-1x, a data set of properties for 5 million non-equilibrium DFT calculations, is created in this fashion.[93] This strategy is further applied to the development of the ANI-2x data set,[82] which extends ANI-1x to new elements S, F, and Cl. As shown in the t-SNE plot (Figure 3C), ANI-1x preserves most of the chemical diversity of the ANI-1 data set while being 4 times smaller. Moreover, some conformations with O−H interactions are even better sampled in ANI-1x, resulting in a slightly more accurate ***ANI-1x*** potential compared to the original ***ANI-1*** model.

Careful AL-based data set design becomes especially important for numerically expensive, highly accurate QM methods such as the CCSD(T) combined with the complete basis set (CBS) level of theory, considered a gold standard in quantum chemistry. The ANI-1ccx data set is generated by iteratively applying an active learning "filter" to ANI-1x, which results in properties of ∼500K molecular structures evaluated at a highly accurate level of theory near CCSD(T)/CBS quality [dubbed CCSD(T)*/CBS].[93] The 2D t-SNE representation of this data set shown in Figure 3C indicates that ANI-1ccx essentially preserves the ANI-1x chemical diversity.

Notably, the extreme computational cost of constructing such data sets usually limits the database expansion; the resulting NNs thus lack reliability due to overfitting and other issues. A subsequent question about how to properly combine and take advantage of multiple data sets computed with different fidelities, such as numerically cheap DFT results (that may encompass, e.g., 5 million structures in the ANI-1x data set[93]) and expensive CCSD(T)/CBS simulations (that may contain, e.g., only 500K data points in the ANI-1ccx data set[93]), then arises. Transfer,[94,95] delta,[96,97] and joint[98,99] learning techniques address these issues. For example, transfer learning starts with a "cheap" but very diverse DFT data set to obtain an extensible and transferable NN model with DFT accuracy. Then this model is retrained, or refined, on a smaller, higher-accuracy data set, after fixing the majority of the NN parameters. A schematic workflow of transfer learning applied to the ANI-1x and ANI-1ccx data set[93] is shown in Figure 3D.

A concerted application of active and transfer learning techniques can describe non-equilibrium processes with quantitative chemical accuracy. For example, Figure 4A shows the potential performance of ***ANI-1ccx*** on the ISOL6 benchmark that contains chemical reactions and isomerization energies. The graph shows the differences among the reference energies at the CCSD(T)-F12a/aug-cc-pVDZ level, energies computed with the DFT model (*ω*B97X/6-31g*), ***ANI-1x*** and ***ANI-1ccx*** potentials, and the CCSD(T)*/CBS approximation scheme. As is frequently the case for chemical reactions, the *ω*B97X with the modest basis set 6-31g* is not particularly

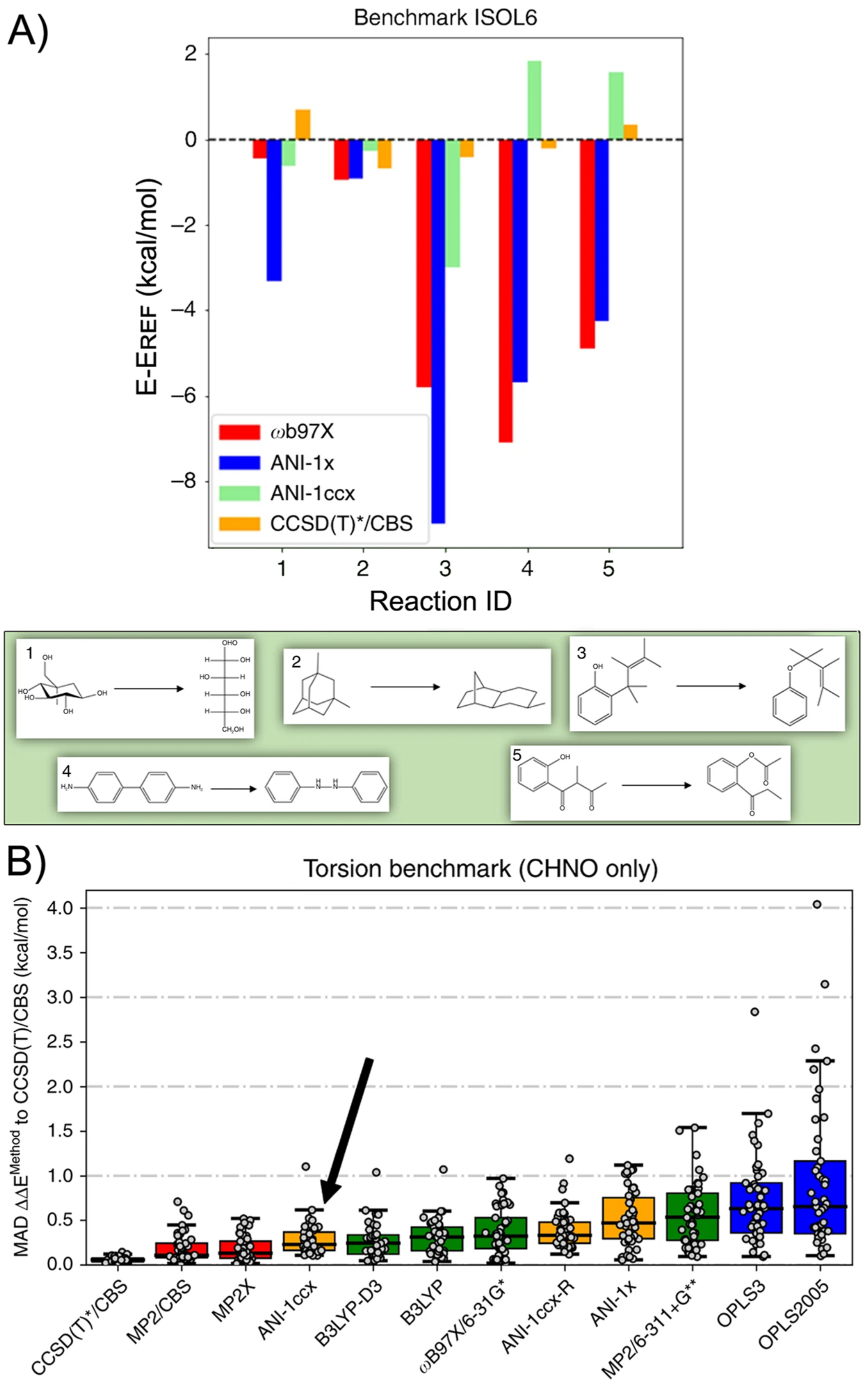


Figure 4. (A) Accuracy in predicting reaction and isomerization energy. **ANI-1ccx** reaction and isomerization energy difference prediction on the ISOL6 benchmark, relative to the reference CCSD(T)-F12a/aug-cc-pVDZ data. Methods compared are the **ANI-1ccx** transfer learning potential, **ANI-1x** trained only on DFT data, the DFT reference (ωB97X), and coupled-cluster extrapolation scheme CCSD(T)*/CBS. Reproduced with permission from ref 103. Copyright 2019 Springer Nature Ltd. (B) Accuracy in predicting torsional energies relevant to drug discovery. Methods compared are QM (red and green), molecular mechanics (blue), and **ANI** (orange) performance on 45 torsion profiles containing C, H, N, and O atomic elements. The gray dots represent the MAD of a given torsion scan vs the gold standard CCSD(T)/CBS approach. Reproduced with permission from ref 103. Copyright 2019 Springer Nature Ltd.

accurate. As expected, ***ANI-1x***, which was trained to data from the same level of theory, mimics the performance of DFT. The ***ANI-1ccx*** potential derived via transfer learning is much more accurate and closely resembles the behavior of the CCSD(T)*/CBS method.[93] Therefore, the transfer learning approach significantly improves the accuracy of the pretrained ***ANI-1x*** model, while preserving its original transferability and extensibility.

For an example of torsional profiles, Figure 4B compares the behavior of the ***ANI-1ccx*** model with ***ANI-1ccx-R***, which is trained solely to the ANI-1ccx data set with only 500K data points. Additionally, it examines the accuracy of various ANI models against other QM and molecular mechanics methods.

The transfer learning approach significantly improves the accuracy of the pre-trained model while preserving its original transferability and extensibility.

Torsional energy profiles play an important role in modeling soft materials and in drug discovery. These are key quantities in classical force fields that require careful parametrization. We use the torsion benchmark of small organic molecules containing C, H, N, and O atoms reported by Sellers et al.[100] The whisker plots in Figure 4B compare ***ANI*** potentials (yellow), three expensive wave function-based QM methods (red boxes), four less expensive QM methods (green boxes), and two versions of OPLS (optimized potentials for liquid simulations) force fields,[101,102] a classical force field designed for accuracy on diverse small organic molecules (blue boxes).[103] The ***ANI-1x*** potential, trained on DFT data, achieves an MAE of 0.47 kcal/mol. Its performance mimics MP2/6-311+G** and lags just behind the ANI-1ccx-R potential. The ***ANI-1x*** potential outperforms the two force fields. At the same time, the ***ANI-1ccx*** potential has a median MAE of 0.23 kcal/mol, 2 times smaller than that of ***ANI-1x***, which outperforms all utilized DFT methods and approaches the accuracy of much more expensive QM methods. It is important to emphasize that there is no increase in numerical expense for ***ANI-1ccx*** over ***ANI-1x*** at prediction time. Both ***ANI*** potentials scale linearly compared to $O(N^3)$, which is typical for a hybrid DFT approach. Another important insight from torsional benchmarks is that the transfer learning approach results in accuracy that is better than that of a model trained only on a small, high-quality data set.

In addition to molecular gas phase simulations, ML potentials have been successfully applied in various condensed phase simulations.[49,70,104,105] For example, ***ANI-Al*** is a ML potential recently proposed for aluminum solid state simulations.[84] As a continuation of the ANI-1x strategy, here active learning was utilized to obtain a diverse training set with minimal human intervention. The active learning loop includes three main steps: (1) MD simulation at varying temperatures using the best available ***ANI-Al*** model to sample new configurations, (2) ensemble uncertainty estimation and DFT calculations for new configurations that meet the uncertainty threshold, and (3) training a new ***ANI-Al*** model to the augmented training data. Although each AL MD simulation is initialized to a random disordered system (melts), after several iterations, the AL starts capturing ordered configurations like FCC, HCP, BCC, etc. (Figure 5A). This is demonstrated in Figure 5B, which is a 2D t-SNE representation colored with respect to the AL iteration

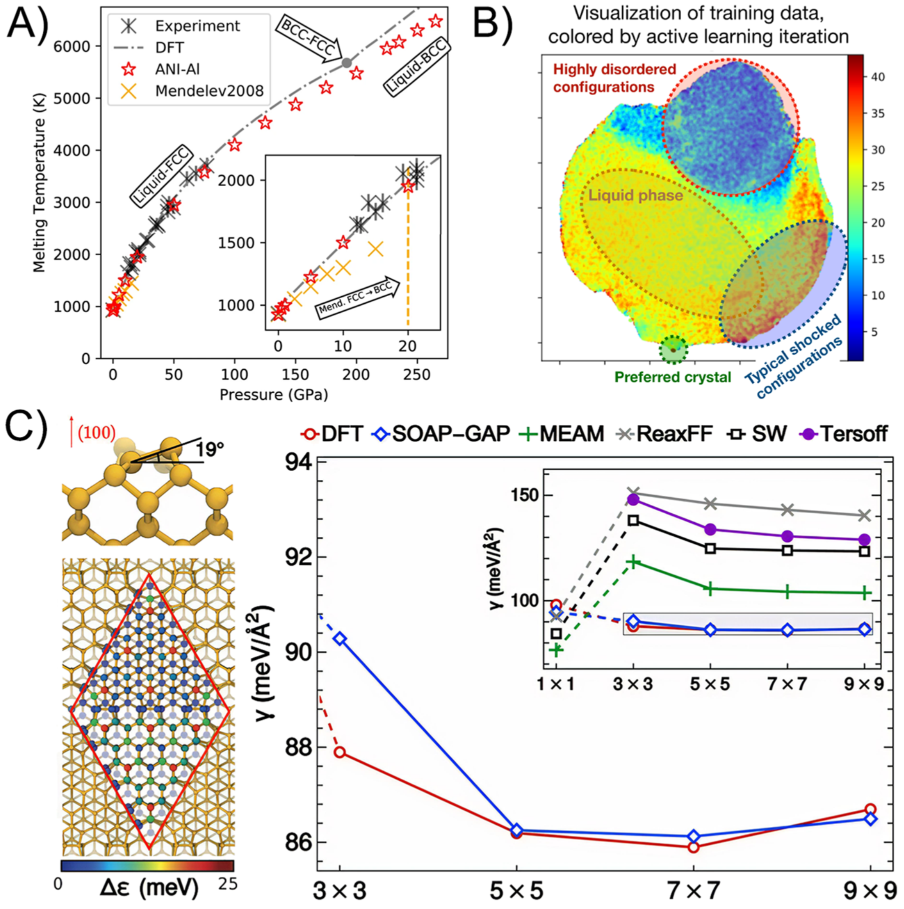


Figure 5. (A) Melt curve as a function of pressure for DFT, **ANI-Al**, and the EAM potential of Mendelev et al., compared with experimental data. Reproduced with permission from ref 84. Copyright 2021 Springer Nature Ltd. (B) t-SNE representation of training data for **ANI-Al** colored with respect to an active learning iteration at which a sample was taken. (C) **SOAP-GAP** ML model that correctly predicts surface Si dimer tilt and that 7 × 7 is the ground state structure. Reproduced with permission from ref 81. Copyright 2017 American Association for the Advancement of Science.

number at which that region of space was sampled. Thus, the uncertainty estimation during the MD run gradually accounts for essential crystal structures. Importantly, the construction of the training set is fully automated. Although crystal structures are eventually sampled through the AL procedure, these formed only through nucleation in the MD and were never directly seeded by hand. The required human input was primarily the range of temperatures for MD sampling and densities for the random system construction and hyperparameters for the ML model.

Figure 5A shows the performance of ***ANI-Al*** in predicting the liquid−solid coexistence line in pressure versus temperature coordinates. The ML model is compared to experimental data, DFT, and the classical potential of Mendelev et al.[106] that is parametrized to model the melting point of aluminum ($T = 933$ K) at atmospheric pressure. Both ***ANI-Al*** and Mendelev potentials predict the melting point of 925 K at this pressure, which is in good agreement with the experimental data. Then, the Mendelev potential starts to underestimate the melting temperature beyond 5 GPa, while ***ANI-Al*** provides quantitatively accurate predictions up to ∼50 GPa, which corresponds to the range of densities sampled in the AL process. Interestingly, the qualitative behavior of ***ANI-Al*** remains reasonable up to 250 GPa, which is well beyond what was included in the training data.[106]

There are alternative ML strategies for addressing the structural properties of solids. For example, recently, Ceriotti and co-workers[81] successfully applied a kernel-based ML approach to model crystalline Si. Although the bulk Si can be easily handled by many popular classical force fields, its surface exhibits nontrivial structural and electrical properties. Si dimers on the (100) surface are subject to Jahn−Teller distortion, which results in their being tilted relative to the crystal surface plane (Figure 5C). Empirical force fields fail to model such behavior, while the ***SOAP*** (smooth overlap of atomic positions)−***GAP*** (Gaussian approximation potential) correctly predicts the 19° tilt in agreement with DFT (Figure 5C).[81] Another puzzling feature of the Si crystal is the 7 × 7 reconstruction of the (111) surface described by a dimer-adatom-stacking-fault (DAS) model.[107] Empirical potentials do not recognize the 7 × 7 cell as the lowest-energy configuration and mistakenly predict unreconstructed 1 × 1 to be the lowest-energy structure (Figure 5C). At the same time, the ***SOAP-GAP*** potential correctly predicts the 7 × 7 configuration to be an energy minimum. This clearly reflects the extensibility of this model as training data included reconstruction unit cells up to 3 × 3 in size.[81]

In conclusion, machine learning has become a ubiquitous tool for modeling molecular and material potential energy surfaces across diverse chemical spaces. These ML potentials enable MD simulations at scales approaching those that can be achieved with classical force fields, and with accuracies approaching those of *ab initio* techniques. Recent trends show that a vast range of ML techniques can be applied to an inexhaustible variety of applications in chemistry and materials science. As the chemical systems become larger and more complex, the classical methods are becoming either inaccurate or too expensive for handling such complexity, because bridging between quantum mechanical and classical models is getting more tedious and difficult. Thus, ML frameworks are a logical extension of chemical science, which can learn from a vast amount of data. Overall, modern ML algorithms provide the best ratio between speed and accuracy, being significantly cheaper than *ab initio* and more accurate and transferable than classical force fields.

ML frameworks are a logical extension of chemical science, which are able to learn from a vast amount of data.

In this Perspective, we discussed the history of and recent advances in NN-based interatomic potentials applied to molecular systems and solids. Provided with sufficiently diverse training data, neural networks easily reproduce geometry-dependent properties, providing accurate energy profiles of large molecules and phase diagrams of solids. However, there is still room for improvement. Development of general-use ML potentials that can be applied to a wide variety of systems is an active area of research. These universal potentials enable routine high-throughput experimentation. Also, researchers who perform *in silico* experimentation frequently lack massive computational resources or expertise to develop a new model for each system or material.

There are several opportunities to improve ML model architectures by designing them to incorporate more physics. ML-based potentials typically assume spatial locality. The force on an atom is assumed to depend only on neighboring atoms within some radius, typically around 5−10 Å. To accurately model some materials, it will be important to capture long-range effects such as Coulomb interactions. Recent work has proposed ML-based potentials that employ self-consistent charge equilibration schemes, inspired by their use in classical polarizable force fields.[77,108] Going beyond Coulomb interactions, there may be other types of long-range interactions that an ML model should ideally capture, e.g., dispersion and effects due to delocalized electron wave functions. Finding general frameworks for introducing such long-range interactions into ML models seems to be an outstanding challenge. If one naively increases the cutoff distance in the symmetry functions (i.e., in the representation of the atomic geometry provided to the neural network), then ML accuracy usually worsens due rapid growth in the number of tunable parameters. Finding new ways to introduce flexible long-range interactions, while imposing a strong (and physics-inspired) regularization to control model capacity, seems to be an important area for future development.

Finding new ways to introduce flexible long-range interactions, while imposing a strong (and physics-inspired) regularization to control model capacity, seems an important area for future development.

Along these lines, another challenge facing most existing ML models is in the prediction of electronic properties like spectroscopic states and excitation energies.[109−112] A promising direction here is to train ML models that predict matrix elements for an effective quantum Hamiltonian, as a function of the atomic geometry.[113−115] The effective Hamiltonian will typically have a very small basis set but must still be solved using traditional approaches from quantum chemistry. This hybrid ML/QM approach introduces great modeling flexibility, but also significant computational cost; training such ML models can become especially expensive. Future algorithms that retain the ML/QM modeling power, while reducing the computational costs, will be crucial to making this modeling approach practical in more contexts.

A concomitant effort in data generation is a challenging task by itself. The extensibility and transferability of a model strongly depend on the training set; thus, new techniques for potential energy surface exploration, such as that inspired by an active learning strategy, are essential for ML development in chemistry and materials. Here fully automated generation of the data set (as exemplified by the development of a general potential for elemental aluminum[84]) is a very promising route. Arguably, a common source of inaccuracy of general ML potentials, whether in the context of chemistry or materials, comes from biasing the data set toward the regions of chemical space that a researcher perceives as important. This bias can cause an ML potential to incorrectly favor some configurations over others in MD simulations. In contrast, an automated, active learning approach can lead to highly diverse data sets that perform well even beyond what the ML modeler might have originally anticipated. In the case of the ***ANI-Al*** potential,[84] the model could be used to perform highly accurate shock simulations, even though no shocked data were included in the training data set. Although narrowly focused models may have many practical uses, they run the risk of entering an untrained region of chemical space in MD simulations. Moreover, uncertainty quantification and sampling of a relevant phase space become particularly important for the deeper study of the reactivity and kinetics that require fast sampling of rare events such as transition states. These challenges have yet to be fully addressed by ML and active learning and will need to be solved before tackling more difficult problems, such as training ML potentials for condensed phase reactive chemistry.

In retrospect, the flourishing ML applications in chemistry are reminiscent of quantum chemistry and molecular dynamics developments during the past century. Initially, application of software based on quantum mechanical models was an arduous affair for the end user (such as a synthetic chemist or spectroscopist), an exercise reserved mostly for experienced theoreticians, but eventually, the market was flooded with user-friendly quantum chemistry software along with copious guidelines on how to select proper methods. Subsequently, for example, a broad choice of DFT functionals or classical force fields is available to the community of practitioners, making their use routine. We observe a similar situation with ML methods when more and more ready-to-use software packages (e.g., TorchANI,[116] AIMNet code,[117] ASE_ANI repo,[71] SchNet,[72] AENet,[70] DeePMD,[118] AMP,[119] PROPhet,[120] TensorMol,[47] and others) and data sets (e.g., ANI-1[46] and ANI-1ccx[93] sets, ChemSpider,[121] ISO17,[88] etc.[122,123]) for different purposes are released, with more released every year, which potentially may revolutionize our chemical science discovery process and have a major impact on academic and industrial research and development.

■ AUTHOR INFORMATION

Corresponding Author

Sergei Tretiak − *Theoretical Division, Center for Nonlinear Studies, and Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States;* orcid.org/0000-0001-5547-3647; Email: serg@lanl.gov

Authors

Maksim Kulichenko − *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States; Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322, United States*

Justin S. Smith − *Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*

Benjamin Nebgen − *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States;* orcid.org/0000-0001-5310-3263

Ying Wai Li − *Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*

Nikita Fedik − *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States; Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322, United States*

Alexander I. Boldyrev − *Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322, United States;* orcid.org/0000-0002-8277-3669

Nicholas Lubbers − *Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*

Kipton Barros − *Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpclett.1c01357

Notes

The authors declare no competing financial interest.

Biographies

Maksim Kulichenko studied Applied Mathematics and Physics during his undergraduate work at Moscow State University. He is currently finishing his Ph.D. studies under the supervision of Prof. Alexander I. Boldyrev at Utah State University. His Ph.D. studies are focused on chemical bonding and nonlinear optical materials. In close collaboration with Los Alamos National Laboratory, his recent research is mostly focused on machine learning in chemistry, application of active learning and biased potentials for constructing datasets.

Dr. Justin S. Smith is a staff scientist at Los Alamos National Laboratory who specializes in the application and development of machine learning methods in chemistry and materials sciences. He was awarded the Nicholas C. Metropolis postdoctoral fellowship at Los Alamos National Laboratory after completing his Ph.D. in chemistry at the University of Florida prior to becoming a staff scientist. His work focuses on developing methods for constructing data sets through active learning and the design of novel architectures for building accurate and general-purpose machine learning-based potential energy predictors of atomic systems. He has developed and applied models to atomic systems relevant to organic chemistry and materials science.

Due to his training as a chemist, Ben Nebgen has interest in applying recent advances in data science and machine learning to the field of theoretical chemistry. Development of machine learned molecular potentials as well as machine learning assisted effective Hamiltonian methods have been a primary focus in this area. Additionally, he has applied recent advances in theoretical chemistry to various practical applications such as the simulation of ionic liquids. Recently, his research interests have expanded to the field of tensor methods, non-negative matrix factorization, and other data driven methodologies. Extensive experience with high performance computing has made this transition possible.

Dr. Ying Wai Li is a staff scientist at Los Alamos National Laboratory with research interests spanning statistical and condensed matter physics, algorithm design, and high-performance computing. She studied Physics during her undergraduate and M.Phil. work at The Chinese University of Hong Kong, and obtained a Ph.D. from The University of Georgia, U.S. Her expertise is in the state-of-the-art classical and parallel Monte Carlo methods for the study of thermodynamics and phase transitions, first principles methods (density functional theory and quantum Monte Carlo) for the study of material properties, and recently the application of machine learning techniques to computer simulations and data analytics.

Nikita Fedik is a Ph.D. student in Prof. Alexander I. Boldyrev research group at Utah State University. His research interests span different areas of computational chemistry, from design of new clusters and materials to data science and machine learning for chemical discovery. His current collaborative projects with Los Alamos National Laboratory are focused on the development of efficient dataset generation protocols and empirical and semiempirical methods dynamically parametrized by machine learning. Additionally, Nikita has long-lasting passion for computers, and he is responsible for advancement of supercomputer infrastructure in his research group.

Prof. Alexander I. Boldyrev received his B.Sc./M.Sc.(1974) in chemistry from Novosibirsk University, his Ph.D. in physical chemistry from Moscow State University, and his Dr. Sci. in chemical physics from Moscow Physico-Chemical Institute (1984). He is currently a R. Gaurth Hansen Professor at the Department of Chemistry and Biochemistry at Utah State University. His current scientific interest is the development of new chemical bonding models for clusters, molecules, solid-state materials, novel two-dimensional materials and other chemical species, where conventional chemical bonding models are not applicable.

Dr. Nicholas Lubbers is a staff scientist at Los Alamos National Laboratory whose current work lies at the intersection of machine learning and the physical sciences. He studied Engineering Physics during his undergraduate work at the Colorado School of Mines, and continued studying Physics, earning a Ph.D. From Boston University. His work has applied and developed machine learning methods for materials science, seismology, fluid mechanics, and porous media, and has focused in particular on the modeling of atomistic systems using neural network approaches.

Dr. Kipton Barros is a staff scientist at Los Alamos National Laboratory who works in the areas of physics and chemistry of materials, computational science, and machine learning. He studied computer science as an undergraduate at Carnegie Mellon University and physics during his Ph.D. at Boston University. Barros' work spans many areas of statistical physics, including the kinetics of phase transformations, collective motion in granular matter, dielectric effects in soft matter systems, multi-scale simulation methods, and algorithms for quantum Monte Carlo codes. A recent focus is the development of machine learning methods to extract insight from scientific data, and to accelerate simulation workflows.

Sergei Tretiak received his M.S. (1994) from Moscow Institute of Physics and Technology (Russia) and his Ph.D. (1998) from the University of Rochester (US) where he worked with Prof. Shaul Mukamel. He was then a Director-funded Postdoctoral Fellow in Theoretical Division at Los Alamos National Laboratory (LANL) and became LANL staff scientist in 2001. He is currently a deputy group leader at T-1 (Theoretical Division, LANL), Adjunct Professor at the University of California (Santa Barbara, CA) and Skolkovo Institute of Science & Technology (Russia). The overarching theme of his research is to develop a theoretical framework for electronic properties and dynamics in complex molecular and semiconductor structures as well as machine learning techniques for chemical dynamics.

■ ACKNOWLEDGMENTS

The work at Los Alamos National Laboratory (LANL) was supported by the LANL Directed Research and Development Funds (LDRD) and performed in part at the Center for Nonlinear Studies (CNLS) and the Center for Integrated Nanotechnologies (CINT), a U.S. Department of Energy Office of Science user facility at LANL. This research used resources provided by the Los Alamos National Laboratory Institutional Computing (IC) Program. A.I.B. acknowledges the R. Gaurth Hansen Professorship.

■ REFERENCES

(1) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503−506.

(2) Purvis, G. D.; Bartlett, R. J. A Full Coupled-cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910−1918.

(3) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(4) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(5) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-

Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.

(6) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.

(7) Zubatiuk, T.; Isayev, O. Development of Multimodal Machine Learning Potentials: Toward a Physics-Aware Artificial Intelligence. *Acc. Chem. Res.* **2021**, *54*, 1575−1585.

(8) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. General-Purpose Machine Learning Potentials Capturing Nonlocal Charge Transfer. *Acc. Chem. Res.* **2021**, *54*, 808−817.

(9) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336−2347.

(10) Koch, W.; Bonfanti, M.; Eisenbrandt, P.; Nandi, A.; Fu, B.; Bowman, J.; Tannor, D.; Burghardt, I. Two-Layer Gaussian-Based MCTDH Study of the S1 ← S0 Vibronic Absorption Spectrum of Formaldehyde Using Multiplicative Neural Network Potentials. *J. Chem. Phys.* **2019**, *151*, 064121.

(11) Westermayr, J.; Faber, F. A.; Christensen, A. S.; von Lilienfeld, O. A.; Marquetand, P. Neural Networks and Kernel Ridge Regression for Excited States Dynamics of $CH_2NH_2^+$: From Single-State to Multi-State Representations and Multi-Property Machine Learning Models. *Machine Learning: Science and Technology* **2020**, *1*, 025009.

(12) Mazhnik, E.; Oganov, A. R. Application of Machine Learning Methods for Predicting New Superhard Materials. *J. Appl. Phys.* **2020**, *128*, 075102.

(13) Jha, D.; Ward, L.; Paul, A.; Liao, W.; Choudhary, A.; Wolverton, C.; Agrawal, A. ElemNet : Deep Learning the Chemistry of Materials from Only Elemental Composition. *Sci. Rep.* **2018**, *8*, 17593.

(14) Jørgensen, P. B.; Schmidt, M. N.; Winther, O. Deep Generative Models for Molecular Science. *Mol. Inf.* **2018**, *37*, 1700133.

(15) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* **2018**, *4*, eaap7885.

(16) Degiacomi, M. T. Coupling Molecular Dynamics and Deep Learning to Mine Protein Conformational Space. *Structure* **2019**, *27*, 1034−1040.e3.

(17) Ghosh, K.; Stuke, A.; Todorović, M.; Jørgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke, P. Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Adv. Sci.* **2019**, *6*, 1801367.

(18) Matsuzaka, Y.; Uesawa, Y. Optimization of a Deep-Learning Method Based on the Classification of Images Generated by Parameterized Deep Snap a Novel Molecular-Image-Input Technique for Quantitative Structure−Activity Relationship (QSAR) Analysis. *Front. Bioeng. Biotechnol.* **2019**, *7*, 65.

(19) Cova, T. F. G. G.; Pais, A. A. C. C. Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. *Front. Chem.* **2019**, *7*, 809.

(20) Wood, M. A.; Thompson, A. P. Extending the Accuracy of the SNAP Interatomic Potential Form. *J. Chem. Phys.* **2018**, *148*, 241721.

(21) 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*, 025002.

(22) Zhai, H.; Alexandrova, A. N. Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization. *J. Chem. Theory Comput.* **2016**, *12*, 6213−6226.

(23) Zhou, G.; Chu, W.; Prezhdo, O. V. Structural Deformation Controls Charge Losses in $MAPbI_3$: Unsupervised Machine Learning of Nonadiabatic Molecular Dynamics. *ACS Energy Lett.* **2020**, *5*, 1930−1938.

(24) Prezhdo, O. V. Advancing Physical Chemistry with Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 9656−9658.

(25) 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*, 136403.

(26) Rosenbrock, C. W.; Homer, E. R.; Csányi, G.; Hart, G. L. W. Discovering the Building Blocks of Atomic Systems Using Machine Learning: Application to Grain Boundaries. *npj Comput. Mater.* **2017**, *3*, 29.

(27) Ferré, G.; Haut, T.; Barros, K. Learning Molecular Energies Using Localized Graph Kernels. *J. Chem. Phys.* **2017**, *146*, 114107.

(28) Grisafi, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M. Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.

(29) Bartók, A. P.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051−1057.

(30) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14*, 1153−1173.

(31) 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.

(32) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(33) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(34) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(35) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, 1912.01703.

(36) Griewank, A. On Automatic Differentiation. In *Mathematical Programming: Recent Developments and Applications*; Kluwer Academic Publishers: Boston, 1989; pp 83−108.

(37) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, 2016.

(38) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, DOI: 10.1021/acs.chemrev.0c01111.

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

(40) Thomsen, J. U.; Meyer, B. Pattern Recognition of the 1H NMR Spectra of Sugar Alditols Using a Neural Network. *J. Magn. Reson.* **1989**, *84*, 212−217.

(41) Curry, B.; Rumelhart, D. E. MSnet: A Neural Network Which Classifies Mass Spectra. *Tetrahedron Comput. Methodol.* **1990**, *3*, 213−237.

(42) Holley, L. H.; Karplus, M. Protein Secondary Structure Prediction with a Neural Network. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 152−156.

(43) Rabow, A. A.; Scheraga, H. A. Lattice Neural Network Minimization Application of Neural Network Optimization for Locating the Global-Minimum Conformations of Proteins. *J. Mol. Biol.* **1993**, *232*, 1157−1168.

(44) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.

(45) 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*, e1603015.

(46) 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*, 241733.

(47) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261−2269.

(48) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.

(49) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(50) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.

(51) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(52) Glielmo, A.; Sollich, P.; De Vita, A. Accurate Interatomic Force Fields via Machine Learning with Covariant Kernels. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 214302.

(53) Kruglov, I.; Sergeev, O.; Yanilkin, A.; Oganov, A. R. Energy-Free Machine Learning Force Field for Aluminum. *Sci. Rep.* **2017**, *7*, 8512.

(54) Sifain, A. E.; Lubbers, N.; Nebgen, B. T.; Smith, J. S.; Lokhov, A. Y.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Discovering a Transferable Charge Assignment Model Using Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 4495−4501.

(55) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.

(56) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579−590.

(57) Nebgen, B.; Lubbers, N.; Smith, J. S.; Sifain, A. E.; Lokhov, A.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Transferable Dynamic Molecular Charge Assignment Using Deep Neural Networks. *J. Chem. Theory Comput.* **2018**, *14*, 4687−4698.

(58) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C−N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186−190.

(59) Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; et al. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem.* **2018**, *4*, 522−532.

(60) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput. Mater.* **2017**, *3*, 1−13.

(61) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73−76.

(62) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal Fragment Descriptors for Predicting Properties of Inorganic Crystals. *Nat. Commun.* **2017**, *8*, 15679.

(63) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural Network Models of Potential Energy Surfaces. *J. Chem. Phys.* **1995**, *103*, 4129−4137.

(64) Lorenz, S.; Groß, A.; Scheffler, M. Representing High-Dimensional Potential-Energy Surfaces for Reactions at Surfaces by Neural Networks. *Chem. Phys. Lett.* **2004**, *395*, 210−215.

(65) Tai No, K.; Ha Chang, B.; Yeon Kim, S.; Shik Jhon, M.; Scheraga, H. A. Description of the Potential Energy Surface of the Water Dimer with an Artificial Neural Network. *Chem. Phys. Lett.* **1997**, *271*, 152−156.

(66) Hobday, S.; Smith, R.; Belbruno, J. Applications of Neural Networks to Fitting Interatomic Potential Functions. *Modell. Simul. Mater. Sci. Eng.* **1999**, *7*, 397−412.

(67) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.

(68) Pun, G. P. P.; Batra, R.; Ramprasad, R.; Mishin, Y. Physically Informed Artificial Neural Networks for Atomistic Modeling of Materials. *Nat. Commun.* **2019**, *10*, 2339.

(69) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032−1050.

(70) Artrith, N.; Urban, A. An Implementation of Artificial Neural-Network Potentials for Atomistic Materials Simulations: Performance for TiO_2. *Comput. Mater. Sci.* **2016**, *114*, 135−150.

(71) 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*, 3192−3203.

(72) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet − A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(73) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical Modeling of Molecular Energies Using a Deep Neural Network. *J. Chem. Phys.* **2018**, *148*, 241715.

(74) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-Covalent Interactions across Organic and Biological Subsets of Chemical Space: Physics-Based Potentials Parametrized from Machine Learning. *J. Chem. Phys.* **2018**, *148*, 241706.

(75) 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*, 094119.

(76) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

(77) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A Fourth-Generation High-Dimensional Neural Network Potential with Accurate Electrostatics Including Non-Local Charge Transfer. *Nat. Commun.* **2021**, *12*, 398.

(78) Westermayr, j.; Gastegger, M.; Marquetand, P. Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 3828−3834.

(79) Westermayr, J.; Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **2020**, DOI: 10.1021/acs.chemrev.0c00749.

(80) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9*, 3887.

(81) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, e1701816.

(82) Devereux, C.; Smith, J. S.; Davis, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192−4202.

(83) Deringer, V. L.; Pickard, C. J.; Csányi, G. Data-Driven Learning of Total and Local Energies in Elemental Boron. *Phys. Rev. Lett.* **2018**, *120*, 156001.

(84) 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*, 1257.

(85) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48*, 722−730.

(86) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022.

(87) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. In *International Conference on Machine Learning*; PMLR, 2017; pp 1263−1272.

(88) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *arXiv* **2017**, 1706.08566.

(89) van der Maaten, L. Learning a Parametric Embedding by Preserving Local Structure. In *Artificial Intelligence and Statistics*; PMLR, 2009; pp 384−391.

(90) Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152*, 050902.

(91) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

(92) Seung, H. S.; Opper, M.; Sompolinsky, H. Query by Committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*; COLT '92; Association for Computing Machinery: New York, NY, 1992; pp 287−294.

(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*, 134.

(94) Taylor, M. E.; Stone, P. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research* **2009**, *10*, 1633−1685.

(95) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345−1359.

(96) Hu, L.; Wang, X.; Wong, L.; Chen, G. Combined First-Principles Calculation and Neural-Network Correction Approach for Heat of Formation. *J. Chem. Phys.* **2003**, *119*, 11501−11507.

(97) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.

(98) Zheng, S.; Hao, Y.; Lu, D.; Bao, H.; Xu, J.; Hao, H.; Xu, B. Joint Entity and Relation Extraction Based on a Hybrid Neural Network. *Neurocomputing* **2017**, *257*, 59−66.

(99) Shen, C.; Luo, J.; Lai, Z.; Ding, P. Multiview Joint Learning-Based Method for Identifying Small-Molecule-Associated MiRNAs by Integrating Pharmacological, Genomics, and Network Knowledge. *J. Chem. Inf. Model.* **2020**, *60*, 4085−4097.

(100) Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. Model.* **2017**, *57*, 1265−1275.

(101) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.

(102) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509−1519.

(103) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.

(104) Cubuk, E. D.; Schoenholz, S. S.; Rieser, J. M.; Malone, B. D.; Rottler, J.; Durian, D. J.; Kaxiras, E.; Liu, A. J. Identifying Structural Flow Defects in Disordered Solids Using Machine-Learning Methods. *Phys. Rev. Lett.* **2015**, *114*, 108001.

(105) Kruglov, I. A.; Yanilkin, A.; Oganov, A. R.; Korotaev, P. Phase Diagram of Uranium from Ab Initio Calculations and Machine Learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2019**, *100*, 174104.

(106) Mendelev, M. I.; Kramer, M. J.; Becker, C. A.; Asta, M. Analysis of Semi-Empirical Interatomic Potentials Appropriate for Simulation of Crystalline and Liquid Al and Cu. *Philos. Mag.* **2008**, *88*, 1723−1750.

(107) Takayanagi, K.; Tanishiro, Y.; Takahashi, M.; Takahashi, S. Structural Analysis of Si(111)-7 × 7 by UHV-transmission Electron Diffraction and Microscopy. *J. Vac. Sci. Technol., A* **1985**, *3*, 1502−1506.

(108) Rappe, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(109) Scher, J. A.; Govind, N.; Chakraborty, A. Evidence of Skewness and Sub-Gaussian Character in Temperature-Dependent Distributions of One Million Electronic Excitation Energies in PbS Quantum Dots. *J. Phys. Chem. Lett.* **2020**, *11*, 986−992.

(110) Dimitrov, S. D.; Azzouzi, M.; Wu, J.; Yao, J.; Dong, Y.; Tuladhar, P. S.; Schroeder, B. C.; Bittner, E. R.; McCulloch, I.; Nelson, J.; et al. Spectroscopic Investigation of the Effect of Microstructure and Energetic Offset on the Nature of Interfacial Charge Transfer States in Polymer: Fullerene Blends. *J. Am. Chem. Soc.* **2019**, *141*, 4634−4643.

(111) Dral, P. O.; Barbatti, M.; Thiel, W. Nonadiabatic Excited-State Dynamics with Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 5660−5663.

(112) Xue, B.-X.; Barbatti, M.; Dral, P. O. Machine Learning for Absorption Cross Sections. *J. Phys. Chem. A* **2020**, *124*, 7199−7210.

(113) Li, H.; Collins, C.; Tanha, M.; Gordon, G. J.; Yaron, D. J. A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians. *arXiv* **2018**, 1808.04526v2.

(114) Zubatyuk, T.; Nebgen, B.; Lubbers, N.; Smith, J. S.; Zubatyuk, R.; Zhou, G.; Koh, C.; Barros, K.; Isayev, O.; Tretiak, S. Machine Learned Hückel Theory: Interfacing Physics and Deep Neural Networks. *arXiv* **2019**, 1909.12963v1.

(115) Zhou, G.; Nebgen, B.; Lubbers, N.; Malone, W.; Niklasson, A. M. N.; Tretiak, S. Graphics Processing Unit-Accelerated Semiempirical Born Oppenheimer Molecular Dynamics Using PyTorch. *J. Chem. Theory Comput.* **2020**, *16*, 4951−4962.

(116) 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*, 3408−3415.

(117) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Science Advances* **2019**, *5*, eaav6490.

(118) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-Kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178−184.

(119) Khorshidi, A.; Peterson, A. A. Amp: A Modular Approach to Machine Learning in Atomistic Simulations. *Comput. Phys. Commun.* **2016**, *207*, 310−324.

(120) Kolb, B.; Lentz, L. C.; Kolpak, A. M. Discovering Charge Density Functionals and Structure-Property Relationships with PROPhet: A General Framework for Coupling Machine Learning and First-Principles Methods. *Sci. Rep.* **2017**, *7*, 1192.

(121) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123−1124.

(122) Balcells, D.; Skjelstad, B. B. TmQM Dataset—Quantum Geometries and Properties of 86k Transition Metal Complexes. *J. Chem. Inf. Model.* **2020**, *60*, 6135−6146.

(123) Ropo, M.; Schneider, M.; Baldauf, C.; Blum, V. First-Principles Data Set of 45,892 Isolated and Cation-Coordinated Conformers of 20 Proteinogenic Amino Acids. *Sci. Data* **2016**, *3*, 160009.