JCTC Journal of Chemical Theory and Computation

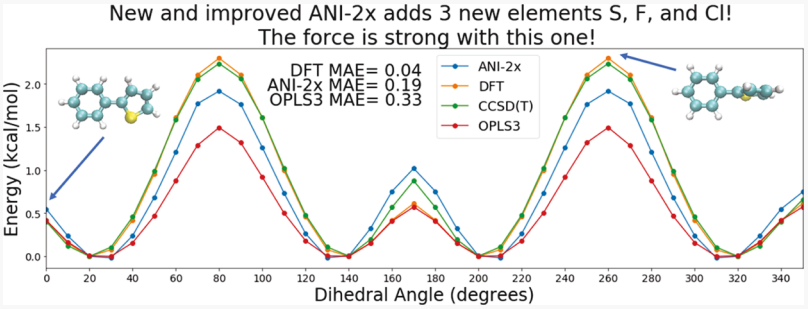
pubs.acs.org/JCTC Article

Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens

Christian Devereux, Justin S. Smith,* Kate K. Davis, Kipton Barros, Roman Zubatyuk, Olexandr Isayev,* and Adrian E. Roitberg*

Cite This: J. Chem. Theory Comput. 2020, 16, 4192−4202 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Machine learning (ML) methods have become powerful, predictive tools in a wide range of applications, such as facial recognition and autonomous vehicles. In the sciences, computational chemists and physicists have been using ML for the prediction of physical phenomena, such as atomistic potential energy surfaces and reaction pathways. Transferable ML potentials, such as ANI-1x, have been developed with the goal of accurately simulating organic molecules containing the chemical elements H, C, N, and O. Here, we provide an extension of the ANI-1x model. The new model, dubbed ANI-2x, is trained to three additional chemical elements: S, F, and Cl. Additionally, ANI-2x underwent torsional refinement training to better predict molecular torsion profiles. These new features open a wide range of new applications within organic chemistry and drug development. These seven elements (H, C, N, O, F, Cl, and S) make up ∼90% of drug-like molecules. To show that these additions do not sacrifice accuracy, we have tested this model across a range of organic molecules and applications, including the COMP6 benchmark, dihedral rotations, conformer scoring, and nonbonded interactions. ANI-2x is shown to accurately predict molecular energies compared to density functional theory with a $\sim 10^6$ factor speedup and a negligible slowdown compared to ANI-1x and shows subchemical accuracy across most of the COMP6 benchmark. The resulting model is a valuable tool for drug development which can potentially replace both quantum calculations and classical force fields for a myriad of applications.

INTRODUCTION

The application of machine learning (ML) methods in chemistry is rising in popularity because of success in areas such as robotic chemical synthesis,[1] drug and material prediction,[2,3] and quantum mechanical property prediction.[4,5] The latter area of research aims to provide high accuracy predictions of quantum mechanical (QM) reference calculations, while maintaining a computational cost comparable to classical force fields. ML-based property predictors have been employed to predict molecular atomization energies,[5−7] forces,[8−10] potential energy surfaces,[6,11−21] atomic partial charges,[22−24] dipoles, and quadrupoles[21,25−27] with accuracies greatly surpassing classical physics-based techniques. Some researchers have shown that models can be trained to multiple properties simultaneously.[28] The speed, accuracy, and transferability of ML property predictors promise to revolutionize the computational design of drugs and materials by bridging the speed versus accuracy gap between quantum mechanics and classical methods.

Many ML methods have been developed with the aim of predicting an atomistic potential energy surface for a variety of applications, for example, geometry optimization or molecular dynamics simulations. ML potentials for both materials[20,29−31] and biological[11,21,32] (organic) systems have been published. Two classes of methods have been proposed for learning the potential energy of organic molecules: dedicated ML potentials and transferable ML potentials. Dedicated ML potentials are designed to describe the potential energy surface of a single

Received: February 6, 2020
Published: June 16, 2020

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

system or a small class of systems using as little QM reference data as possible. These models tend to provide highly accurate energies and forces for molecules with a relatively low number of degrees of freedom. For dedicated ML potentials, QM calculations are required prior to any application. Therefore, the effective computational scaling of dedicated models is that of the underlying QM method, making their use in applications to large biological systems, for example, proteins and large drug molecules, intractable. As the number of degrees of freedom within a system increases, the amount of QM data needed to fully cover configurational space increases greatly. For large enough systems, this method could become impractical. The added time for QM data generation also makes such models unfeasible for high throughput studies on databases of small molecules.

Transferable ML potentials aim to accurately simulate an entire class of molecules, with the objective of avoiding direct QM calculations prior to an application. That is to say, once the model has been trained, it can be applied to a multitude of systems with no new QM calculations being needed. This yields a linear scaling method in most cases. Current transferable methods for organic molecules work by generating a very large and highly diverse data set of molecular conformations from small molecules as a training data set. Model locality is employed to ensure that training to small systems yields extensibility to larger systems. Much of chemistry admits a nearsightedness principle. As a consequence, one can often achieve both linear scaling in system size and extensible potentials by computing atomic energies as a function of the atom's local environment, as pioneered by Behler and Parinello.[33] The overall philosophy behind transferable models is to provide enough data diversity during the learning process that the model is trained only once and is forced to learn the local atomic physics at play, rather than only describing the potential energy surface in a limited subset of atomic configuration space. Transferable models can generalize features learned from different systems. Interaction types that are learned from some molecules in the training set do not need to be sampled for all molecules, decreasing the amount of data needed for each system when training. Although transferable ML potentials are vastly more general than dedicated ML potentials, they tend to be somewhat less accurate, on the order of chemical accuracy (1 kcal/mol error), or better under near ambient conditions. However, for many practical applications, the level of accuracy achieved by transferable ML potentials is more than sufficient to provide quantitative results.[34−37]

The ANAKIN-ME (ANI) method[11] is one example of a technique for building transferable neural network-based molecular potentials. The key components of ANI models are the data set and Behler and Parrinello-type descriptors[33] with a modified symmetry function.[11] The ANI-1 data set[38] (used to train the ANI-1 potential) was built from 57,000 small CHNO-containing molecules perturbed into 22 million randomly selected molecular conformations. Test cases showed ANI-1 to be chemically accurate compared to reference density functional theory (DFT) ωB97X/6-31G* calculations. However, random normal-mode sampling is based on a harmonic approximation, which leads to sparse coverage of chemical space, for example, torsion space. In response, an active learning algorithm using query by committee (QBC) for selecting new data was employed to automatically diversify the data set.[25,39] This QBC method uses the disagreement of an ensemble of models to accurately predict energies for a given molecule. When poorly described structures are identified, QM data are generated for these conformations. A massive search of conformational space was carried out using this QBC active learning algorithm, resulting in the ANI-1x potential. Because of the vastness of chemical space, the early proof-of-concept ANI-1x data set only sampled molecules with a limited number of atomic elements H, C, N, and O. Although these four elements cover a large swath of interesting organic chemistry, further expansion and diversification of this data set will lead to even greater applicability.

In this study, we extended the previously developed ANI model using our automated active learning algorithm to include the elements S, F, and Cl. These specific elements are chosen because of their ubiquitous applicability, for example, in protein simulation and small-molecule drug design. The resulting potential, ANI-2x, is chemically accurate compared to reference DFT calculations in multiple test cases. These test cases include the original COMP6 benchmark[40] (with C-, H-, N-, and O-containing molecules) combined with a new sister version, COMP6v2, which contains all seven chemical elements (C, H, N, O, S, F, and Cl). The applicability of the ANI-2x potential on relaxed torsion scans involving the new chemical elements and on a small drug-like molecule conformer search are shown. Interaction energies were also predicted by ANI-2x and compared to reference DFT calculations. The ANI-2x potential is available for free on our GitHub repository package integrated with the atomic simulation environment (ASE) library [github.com/isayev/ASE_ANI and github.com/aiqm/torchani/tree/master/torchani/resources].

All the codes and data needed to reproduce all figures and tables in this manuscript can be found in github.com/aiqm/Scripts_for_ani2x_jtct_paper.

METHODS

Atomic Environment Vectors. ANI-2x uses the same method for the construction of its atomic environment vectors (AEVs) as previous ANI models,[11] but new terms have been added to account for the additional chemical elements. These AEVs are a form of Behler and Parrinello-type descriptors[33] with a modified symmetry function. An AEV, $\vec{G}_i^{\,X} = \{G_1, G_2, G_3, ..., G_M\}$, composed of elements $G_{M,}$ is computed using original Behler and Parrinello symmetry functions (BPSFs)[33] for each ith atom of the molecule with atomic number X. The AEV examines the atom's radial and angular chemical environment and is a numerical representation of its local chemical environment. The following piecewise cutoff function executes the local atomic environment approximation

$$f_c(R_{ij}) = \begin{cases} 0.5 \times \cos\left(\frac{\pi R_{ij}}{R_c}\right) + 0.5 & \text{for } R_{ij} \le R_c \\ 0.0 & \text{for } R_{ij} > R_c \end{cases} \quad (1)$$

where R_{ij} is the distance between atom i and atom j and R_c is the cutoff radius. To examine the local radial environment for atom i, the following original radial BPSF is used, which produces radial elements G_m^R of $\vec{G}_i^{\,X}$.

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

$$G_m^{\mathrm{R}} = \sum_{j\neq i}^{\text{all atoms}} e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij}) \tag{2}$$

Each element is over a set of parameters, η and R_s, where η changes the width of the Gaussian distribution and R_s shifts the center of the peaks.

The ANI method[11] uses a modified version of the angular BPSF,[33] where arbitrary number values are allowed for the parameter θ_s and an exponential factor has been added that allows an R_s parameter. When examining atoms i, j, and k, angle θ_{ijk} is computed with two distances, R_{ij} and R_{ik}. The local angular environment for atom i is examined through the single element $G_m^{A_{\mathrm{mod}}}$ of $\vec{G}_i^{\,X}$, which is the product of angular and radial factors

$$G_m^{A_{\mathrm{mod}}} = 2^{1-\zeta} \sum_{j,k\neq i}^{\text{all atoms}} (1+\cos(\theta_{ijk}-\theta_s))^{\zeta} \exp\left[-\eta \left(\frac{R_{ij}+R_{ik}}{2} - R_s\right)^2\right] f_c(R_{ij}) f_c(R_{ik}) \tag{3}$$

Now, each element is over a set of our parameters: ζ, θ_s, η, and R_s, where η and R_s are the radial terms. ζ changes the width of the peaks in the angular environment. θ_s serves a similar purpose to R_s in radial function, by probing specific regions of the angular environment.[11]

Building a Database of S-, F-, and Cl-Containing Molecules. The database of molecules used to build the active learning-based ANI-2x training data set is composed of molecules from a variety of sources, including the GDB-11[41,42] database, the CheMBL[43] database, and the s66x8[44] benchmark. The GDB-11 data set contains an enumeration of chemically feasible organic molecules containing the heavy elements C, N, O, and F. From this database, we combinatorially replaced the chemical symbols O with S and F with Cl for all molecules containing up to eight non-hydrogen atoms. From the ChEMBL[43] database, molecules containing S, F, and Cl were sampled. Also, conformers of amino acids and dipeptides containing S were randomly generated using the Rdkit[45] chem informatics package. These sampling techniques mirror those used in building the ANI-1x data set.[40]

In the Results and Discussion section below, we commonly show the mean absolute error (MAE) and root mean square error (RMSE) as a measure of accuracy of various properties. The properties we measure are the relative energies of molecular conformers, the force components acting on atoms, the absolute potential energies, and errors for various geometric features such as bonds, angles, and torsions. For each test set, all calculations were performed using the DFT functional ωB97X with the 6-31G* basis set, using Gaussian 09 and Gaussian 16. All ANI-2x optimizations were performed using the LBFGS algorithm, as implemented in the ASE Python package.

Active Learning in Chemical Space. The active learning process used in this work directly mirrors that of the active learning process published in the development of the ANI-1x and ANI-1ccx potentials.[34,40] Therefore, we point all interested parties to this work for a detailed description of the active learning processes. The primary difference in this work's process is that all molecules sampled during active learning, except for the s66x8 nonbonded interaction sampling, were required to contain S, F, or Cl. To generate nonequilibrium conformations, we employed dimer sampling, normal-mode sampling, N trajectory molecular dynamics sampling, and ML-driven torsion sampling in all iterations of active learning. A detailed description of these methods is provided in our earlier work.[34,40] As with the ANI-1x active learning process, only small molecules are searched in early iterations with molecule size increasing as the process proceeds. We began by adding molecules with two heavy atoms and iterated at this number until the number of molecules added per iteration began to decrease; then, the process was repeated by adding one heavy atom at a time. In the end, more than 50 active learning cycles were carried out, yielding a data set of 4,695,707 molecular conformations from 13,405 chemical isomers. Combined with the original ANI-1x data set and torsion refinement data set from the ANI-1ccx work, the final ANI-2x data set consists of 8.9 million molecular conformations.

Active Learning Torsion Refinement. As previously presented in our work on developing the ANI-1ccx potential,[40] we carried out an active learning torsion refinement on a randomly selected subset of molecules from the ChEMBL drug molecule database.[43,46] In the development of ANI-2x, 250 SMILES strings were selected at random and then embedded into 3D space using the RDKit cheminformatics package. A rotatable torsion was selected at random for each molecule. If a torsion contained hydrogens or was a member of a ring, then it was not selected. During the active learning cycles, the latest version of the ANI potential ensemble was used to relax the selected torsion every 10°, resulting in 36 conformations. All conformations that have an ensemble disagreement over a set threshold were selected, and normal modes were computed using the ANI ensemble. Four data points were then generated using normal-mode sampling (as presented in our previous work[11]), and QM calculations were performed for each and then added to the training data set. This process is referred to as "torsional refinement". A torsional refinement was carried out during each iteration of the active learning process.

Active Learning Nonbonded Interaction Refinement. To improve sampling of nonbonded interactions, we use the s66x8[47] benchmark to generate training data. s66x8 contains eight structures along the dissociation path of 66 C-, H-, N-, and O-containing dimer systems. Such systems must be sampled to improve the accuracy of nonbonded interactions because these dimers represent the smallest systems containing their respective interaction. This active learning cycle was bootstrapped from the ANI-1x data set and potential. We first generate normal modes using DFT for each of the eight structures along the path. We then use normal-mode sampling to generate random structures along the path of dissociation for each dimer. We carried out 26 active learning cycles to generate 195,291 conformations of random dimers.

Active Learning for Improved Bulk Water. A novel type of sampling was employed to improve the ANI-2x description of bulk water. Water molecules with random position and orientation were placed within a bounding box with random edge lengths. The density of these systems was restricted to be between 0.8 and 1.20 g/cm^3. The resulting box of water molecules was optimized using the current ANI potential and LBFGS, as implemented in the ASE[48] package. NVT molecular dynamics simulations were then carried out using the latest ANI potential. Every five time steps, the simulation was paused, and the box was broken into N small clusters, where N is the number of water molecules in the box. The N clusters

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

were generated by taking all water molecules within 6 Å from the center of the *N*th water. A random selection of water molecules was then deleted from each cluster until 2 to 15 molecules remained. Finally, the active learning selection process was carried out, and QM calculations were performed for any selected clusters. The ensemble disagreement used for selection was ρ = 0.32 kcal/mol/atom$^{1/2}$. When any single cluster has an ensemble disagreement larger than 3ρ, the simulation is terminated. These termination criteria help prevent highly unphysical configurations from forming during the simulation. During each iteration of the ANI-2x active learning process, 10 random boxes of water were sampled, as described above.

Force Training. ANI-2x was trained to molecular energies and forces. The forces predicted by the model are the analytical derivatives of the molecular energies, assuring that energy is conserved when running simulations. Force training was not used during the active learning process because of the increased computational costs associated with training the model to forces. However, a force-trained model predicts at the same computational speed as a model trained to just energies when predicting energies and forces. This force training was done with the intent to improve model accuracy during molecular dynamics. This is accomplished by providing additional information about the molecule's potential energy surface to the model. ANI-2x was trained using the loss function

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[(\hat{E}_i - E_i)^2 + \frac{l_0}{M_i}\sum_{j=1}^{M_i}(\hat{f}_{ij} - f_{ij})^2\right] \quad (4)$$

where $\hat{E}_i$ and $\hat{f}_{ij}$ are the energies and forces predicted for a given molecule, respectively, and E_i and f_{ij} are the QM energy and forces, respectively. l_0 is chosen to balance the force and energy terms during training, and a value of 0.1 is used for ANI-2x. N is the number of systems, and M is the atoms per system. The balancing term between the energy and force components of the loss function is important because the molecular forces do not contain information about the molecule's absolute energy. Without a balance between these two terms, the model may neglect to learn about energies all together in favor of forces. When training a model, the derivative of the loss function must be taken with respect to all weights in the model. The loss function involves forces and thus involves the derivative of the energy with respect to atomic positions. To train the model, we require the gradient of the loss function with respect to the model parameters. In other words, we require second derivatives of the energy. Frameworks such as Tensorflow or Pytorch can perform iterated back propagation and thus automate the procedure of calculating the gradient of the loss function. Such a code transformation would be extremely challenging to perform on our CUDA/C++ implementation of ANI. For this reason, a finite differentiation is used to approximate the model's forces during training. This numerical method of force training has been detailed in other work.[49]

RESULTS AND DISCUSSION

To illustrate the utility of the ANI-2x potential, we have conducted case studies mimicking typical molecular modeling applications: (a) molecular dynamics simulations, (b) potential energy scans, (c) conformer search and ranking, (d) challenging benchmark COMP6v2 database developed in this work, and (e) accuracy of nonbonded interactions from existing benchmarks. For MD simulations, we selected the GSK1107112A (CHEMBL1527187) compound as a real-life industrial compound open sourced with the GSK Tuberculosis Screening campaign.[50] It is much larger than any molecules contained in the training set and contains all atomic elements (C, H, N, O, S, F, and Cl) considered in this work, something not true for most of the model's training set. For the 2D potential energy scans, we selected four molecules (bendamustine, cysteine-dipeptide, dichlorodiphenyltrichloroethane (DDT), and hexafluoroacetone) that provide diverse structures containing both sulfur and halogens. For the conformer search program, we selected 20 molecules from the recent benchmark set of low-energy conformer evaluations.[51] This collection includes drug-like ligands used to assess performance of conformer-generating methods such as OMEGA or ETKDG.

Torsion Profiles. The Genentech torsion benchmark[52] was used to assess the ability of the ANI-2x model to predict torsion profiles. This benchmark consists of 62 molecules containing H, C, N, O, F, Cl, and S. For each molecule in the test set, 36 conformations were generated by rotating one of the bonds in 10° increments. Each structure was then optimized, and energies were computed to produce relaxed torsion profiles at various levels of theory. The reference data are the CCSD(T)/CBS calculations provided by Sellers et al[52] The error between ANI-2x, our reference DFT, and OPLS3 against CCSD(T)/CBS are shown in Table 1. ANI outperforms OPLS3 and has only a slightly higher error than its reference DFT.

Table 1. MAE and RMSE between ANI-2x, ωB97X/6-31G*, and OPLS3 against CCSD(T)/CBS on the Genentech Torsion Benchmark[52]

method	MAE (kcal/mol)	RMSE (kcal/mol)
DFT	0.36	0.51
ANI-2x	0.42	0.59
OPLS3	0.67	1.02

To validate the predictive performance of ANI-2x on real-world molecules, 2D torsion profiles were computed for four different systems containing some combination of the chemical elements C, H, N, O, S, F, and Cl. Two dihedrals were chosen for each molecule and rotated in ten-degree increments in turn to create a 36 by 36 dihedral profile of the molecules, 1296 structures in total per molecule. The resulting structures were then optimized with our reference level of theory, freezing the appropriate dihedrals along the rotation path. Each structure optimized with DFT was then reoptimized with ANI-2x.

This was done so that the time-consuming QM optimizations could be performed in parallel and to assure that the structures generated from each method, DFT and ANI-2x, were in the same local minimum. Very minor differences in potential energy surfaces can lead to large differences in final structures, especially when using two different optimizers and starting from conformations far from the desired minima. The results of these scans are shown in Figure 1. DFT optimizations of the cysteine dipeptide were performed using the Gaussian 09[53] software package, while DDT, hexafluoroacetone, and bendamustine DFT optimizations were performed using Gaussian 16.[54] ANI-2x accurately predicts the location of the minima and maxima for all four molecules. For DDT, hexafluoroacetone, and bendamustine,

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

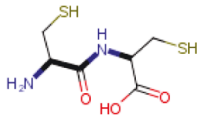
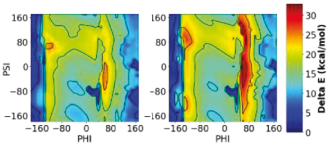
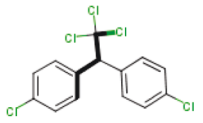
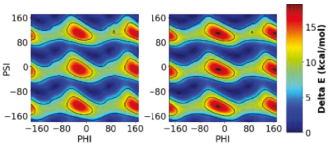
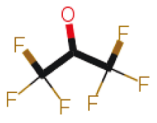
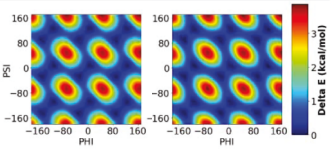
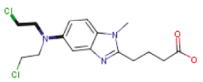
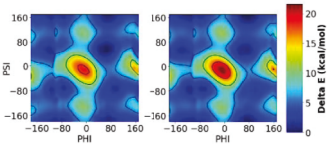
Name	Molecule	MAE (kcal/mol)	RMSE (kcal/mol)	Scan (Left:ANI Right:DFT)
Cysteine-Dipeptide (25 atoms)		1.75	2.55	
DDT (28 atoms)		0.53	0.71	
Hexafluoroacetone (10 atoms)		0.09	0.11	
Bendamustine (44 atoms)		0.50	0.66	

Figure 1. Relaxed 2D torsion profiles for ANI-2x (left) and DFT (right). Two dihedrals (shown in bold) were rotated in ten-degree increments about one another to generate the confirmations to be optimized with DFT. Each confirmation was optimized with the appropriate dihedrals frozen; then, these structures where again optimized with ANI-2x. The bonds composing the scanned dihedrals are bolded in the second column, and the third and fourth columns show the MAE and RMSE, respectively, of the relative energies in kcal/mol between ANI and DFT.

ANI shows subchemical accuracy. ANI shows a greater error on the cysteine dipeptide, but the energy range covered by the scan is much greater for the dipeptide than for the other three systems. For the cysteine dipeptide, ANI-2x has the correct placement for the minima and maxima but does not get the barrier height of the maxima correct, most likely because high-energy configurations are more sparse in the training data. In the Supporting Information, FS3 shows a 2D plot of ANI-2x's error on each molecule compared to the ensemble standard deviation of the 2x model. The standard deviation is divided by the square root of the number of atoms in the system to account for error cancelation between the individual atomic networks. This standard deviation is used as a measure of uncertainty for the ANI-2x model, where high standard deviation/sqrt(N) means the model is less reliable. It is important to note that this uncertainty metric is not the same as a conventional error bar and does not say how far or close to the true answer ANI-2x is, but rather how familiar the network is with the type of system it is being applied to.

For the four 2D torsions shown, ANI-2x performs remarkably well, especially considering the number of chemical elements and total number of atoms in these systems. Larger systems such as the cysteine dipeptide can experience a higher error overall because the error grows with the number of atoms. This produces the following per atom error cancellation-corrected MAE for each of the four torsion scans: 0.35, 0.10, 0.03, and 0.08 kcal/mol/atom$^{-1/2}$. Because many publications on ML potentials present results as the uncorrected per atom MAE, we also provide these results: 0.07, 0.02, 0.01, and 0.01 kcal/mol/atom. However, we stress that the latter metric of comparison is unreliable because larger systems will experience more error cancellation and will thus appear to have a significantly lower error.

Conformer Search and Ranking. Optimizing molecules using higher levels of theory, such as coupled cluster, is often impractical because of the high cost of force calculations with these methods. For this reason, MP2 is often used to optimize molecules before using other methods for energy calculations because of its relatively accurate forces and efficiency. However, even MP2 is too costly to perform high-throughput conformer searches across large data sets. Classical force fields and semi-empirical methods are often employed to siphon through molecules to find the best candidates for drug development before performing costly QM calculations. Unfortunately, because of the limited accuracy of such methods, this often leads to several false negatives, missed candidates, and wasted computational time on false positives.

A test set of 20 molecules was used to determine how well ANI-2x predicts the relative energies of different local minima for drug-like molecules. These molecules were taken from the test set used in recent work on validating force fields by Kanal, Keith, and Hutchinson.[51] We choose the 7 molecules with the highest and 13 molecules with the lowest conformer relative energy correlation between empirical methods and DFT B3LYP ground truth. The SMILES strings for each of these molecules were embedded in 3D space using the Rdkit[45] software package, and a conformational search was performed to generate between 10 and 35 conformations of each molecule. We used an earlier version of the ANI model during the conformational search to optimize each conformer to

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

ensure that each conformation was near a different, unique local minimum. These conformations were then optimized with ANI-2x, PM6, MMFF94, ωB97X/6-31G*, and MP2/cc-PVTZ. We compared the optimized geometries predicted by each method with the MP2-optimized structures. These results are shown in Table 2.

Table 2. Geometry Comparison of the Optimized Structures Predicted by DFT, ANI, PM6, and MMFF94 with the Optimized Structures Predicted by MP2

property	DFT	ANI-2x	PM6	MMFF94
bond length MAE (A)	0.0050	0.0053	0.015	0.0076
angles MAE (degree)	0.19	0.28	0.66	0.48
torsion MAE (degree)	3.36	5.41	11.15	5.28
rmsd (A)	0.30	0.43	0.69	0.44

Single-point calculations were then carried out on each of the ANI-2x-optimized structures using DFT and on the DFT-optimized structures using ANI-2x. We then determined the R^2 correlation and Spearman rank of relative conformer energies predicted/computed by ANI-2x, PM6, MMFF94, and ωB97X/6-31G* and compared with those obtained with MP2. The same comparison was also carried out between ANI-2x and ωB97X. These results are shown in Table 3. In Table 3, ANI@DFT refers to the single-point energies of ANI-2x and DFT on the structures optimized by DFT, DFT@ANI refers to the single-point energies of ANI-2x and DFT on the structures optimized by ANI, and ANI_DFT refers to the ANI single-point energies on the ANI-optimized structures with the DFT single-point energies on the DFT-optimized structures. All MP2 calculations were performed using the Orca software package,[55] and all MMFF94 calculations were performed using the Open Babel software package.[56] MAE and RMSE are reported for each method compared to MP2 and between ANI-2x and DFT. Errors reported represent the errors in the relative energies between all conformations of each molecule.

ANI-2x shows a lower error compared to MP2/cc-PVTZ for both the relative energy and optimized geometry than both PM6 and MMFF94, without needing to rely on any specific atom typing or connectivity information. Not surprisingly, ANI-2x shows better correlation compared to ωB97X in the ANI@DFT and DFT@ANI comparison than in the ANI_DFT because in the latter, the energies being compared come from different structures. The speed and high correlation of ANI-2x with QM methods make it a useful tool for conformer scoring and for generating structures to be further optimized with levels of theory beyond DFT. As an example of the speed of the ANI model, the average time for ANI-2x to perform a single-point calculation on any of the conformations was ∼0.02 s compared to 552 s with ωB97X/6-31G*. While it is important to note that these calculations were done using different hardware (ANI-2x on a GPU and DFT on a CPU), we believe that this illustrates the potential of ANI-2x for large-scale and high-throughput studies. We have shown that ANI-2x is capable of outperforming both semi-empirical and classical methods, while still operating several orders of magnitude faster than QM methods.

COMP6v2. The COMP6 benchmark from our recent work on applying active learning techniques to build general-purpose ML potentials[40] has been extended in this work to include the chemical elements S, F, and Cl. This benchmark is now referred to as COMP6v2. Figure 2 shows the relative energy between all conformers (ΔE), the absolute potential energy (E), and the force component (F) accuracy of ANI-2x on the combined COMP6v2 benchmark. In COMP6v2, the GDB10to13 set has been split into GBD10to11 and GDB12to13. All GDB sets, excluding GDB12to13, have been augmented to contain S, F, and Cl. The DrugBank test set has also been augmented to include these new atomic elements. The tripeptide test set now includes cysteine and methionine, both of which contain sulfur. The number of molecules and conformers in COMP6v2 is provided in the Supporting Information, TS8.

The errors shown in Figure 2a−c are for conformations restricted to within 200 kcal/mol from the nearest energy minima for a given molecule. This energy range is significantly higher than the energy range of conformers visited in room temperature MD simulations. Varying this range allows us to gauge the performance generality of a potential for simulations at a specific temperature. For example, if we restrict this range to 30 kcal/mol, then the data set corresponds to the conformer space visited in near ambient temperature dynamical simulation. In Figure 2a, within the 200 kcal/mol energy range, most of the benchmarks achieve subchemical accuracy (1 kcal/mol) errors, while total energy errors (Figure 2b) tend to be larger because of the bias error for different molecules. However, for many applications which depend on torsional energy barriers and relative populations of conformers, only accurate relative energies are required. Another trend to note here is that the error grows as the number of atoms per molecule grows, for a given benchmark. This trend is expected with atomistic ML potentials because each atomic energy prediction has an error associated with it.

Figure 2d−f shows that as the relative energy range of conformers in the test set is reduced, the overall error drops. This phenomenon can be explained by the fact that near equilibrium, conformers represent the average over the entire

Table 3. Mean Spearman Rank, R^2 Correlation, MAE, and RMSE of Relative Conformer Energies Predicted by DFT, ANI, PM6, and MMFF94 Compared with Those Obtained with MP2[a]

		comparison to MP2			comparison to DFT		
metric	DFT	ANI-2x	PM6	MMFF94	ANI@DFT	DFT@ANI	ANI_DFT
mean R^2	0.79	0.68	0.35	0.52	0.83	0.83	0.74
mean spearman	0.86	0.75	0.45	0.56	0.86	0.87	0.77
MAE (kcal/mol)	1.24	1.93	3.00	3.84	1.45	1.69	1.77
RMSE (kcal/mol)	2.03	2.71	3.83	5.19	1.92	2.32	2.50

[a]MAE and RMSE are computed across all conformers of all molecules. ANI@DFT compares the single-point energies of ANI-2x and DFT on the structures optimized by DFT. DFT@ANI compares the single-point energies of ANI-2x and DFT on the structures optimized by ANI. ANI_DFT compares the ANI single-point energies on the ANI-optimized structures with the DFT single-point energies on the DFT-optimized structures.

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

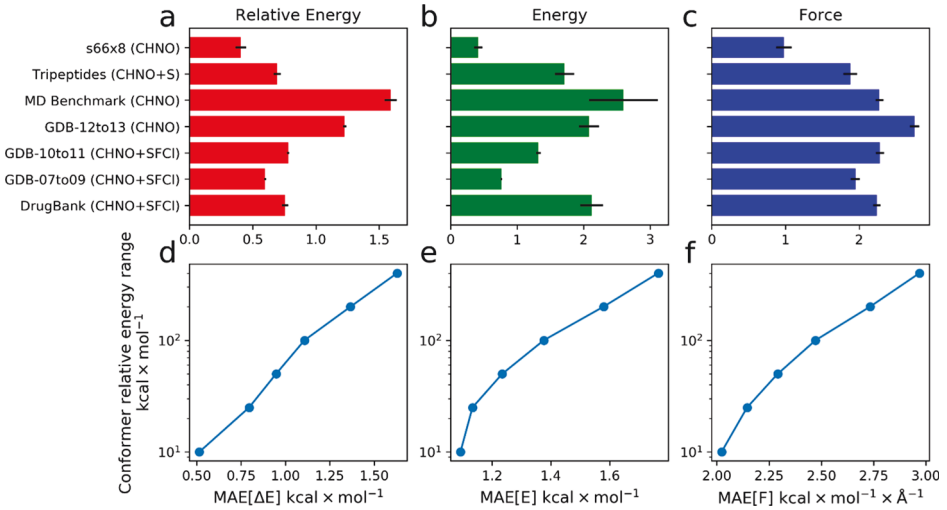


Figure 2. (a–c) Errors for molecules within 200 kcal/mol from the minima compared to DFT reference data from the updated COMP6 benchmark for the chemical elements C, H, N, O, S, F, and Cl. (d–f.) Relative conformer energy range considered vs MAE of relative energy, total energy, and forces over the entire updated COMP6 benchmark.

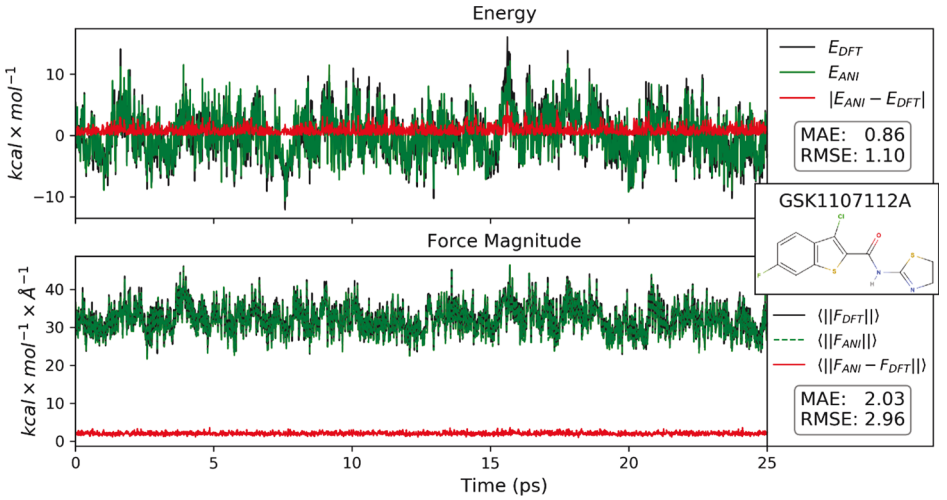


Figure 3. Energies (shifted to the mean) and force magnitudes along with the corresponding errors for a 25 ps *NVT* molecular dynamics (MD) trajectory using the Langevin thermostat at 300 K. This figure represents the final 25 ps of a 1.5 ns MD simulation in vacuum. The drug ligand GSK1107112A was chosen as an example because it contains all atomic elements (C, H, N, O, S, F, and Cl) considered in this work. Energies were shifted to the mean energy over the trajectory. Black shows the DFT-computed properties, green is the ANI-2x-computed properties, and red is the absolute difference between the values.

data set because of the sampling techniques used, that is, MD sampling and normal-mode sampling. This fact is important to remember as it can lead to misconceptions that the errors shown on full benchmarks considering a high energy range does not necessarily represent the error obtained in room temperature simulation.[57]

Molecular Dynamics Trajectory. For a machine learning-based potential to be applicable in molecular dynamics (MD) simulations, it must represent a mathematical potential and ensure conservation of energy and momentum. By construction, ANI models are guaranteed to be conservative to numerical precision. Furthermore, to achieve meaningful sampling time scales, the potential must be computationally efficient. As the first step, we investigate the feasibility of applying the ANI-2x model in MD simulations by generating an MD trajectory of the GSK1107112A compound. This is a much more challenging task than traditional error evaluation as it requires a sampling of the vast configuration space and computing dynamical observables. We then ran single-point DFT calculations on the final 25 ps of the simulation to calculate the energy and forces according to our reference level of theory, ωB97X-6-31G*.

Figure 3 provides energy and force magnitudes along with errors for the final 25 ps of a 1.5 ns NVT MD simulation. The simulation used a time step of 0.4 fs, and the thermostat was set to a temperature of 300 K. This trajectory shows the applicability of ANI-2x to MD simulations for systems containing the chemical elements C, H, N, O, S, F, and Cl. Energy errors compared to DFT reference calculations for the provided portion of the trajectory are 0.86/1.10 MAE/RMSE in kcal/mol. This error represents chemical accuracy for a molecule that was not explicitly added to the training data set. Force magnitude errors are 2.03/2.96 MAE/RMSE in kcal/mol/Å. The simulation ran for 1.5 ns, and the final 25 ps were chosen to show that even after long time scale simulations, ANI-2x is still sampling structures that agree well with

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192–4202

reference DFT. The ANI potential took approximately 12.0 GPU hours to run the 3.75 million steps required for the 1.5 ns simulation in the NeuroChem package (https://github.com/isayev/ASE_ANI). At 27 atoms, this system is too small to saturate the GPU for peak efficiency; therefore, efficiency will grow with larger system sizes. The DFT calculations for the final 25 ps (2500 frames of the trajectory) took 192 CPU core hours.

Nonbonded Interactions. Two data sets were chosen to show that ANI-2x accurately predicts nonbonded interaction energies. The X40 data set was obtained from the Benchmark Energy and Geometry Database.[58] It consists of noncovalent complexes that participate in a variety of interaction types, such as London dispersion, dipole−dipole interactions, and hydrogen bonds.[2] Only the systems containing C, H, N, O, F, and Cl were used in this study; those containing I and Br were omitted. The second data set was taken from work done by Thomas A. Halgren in 1996 to measure the performance of MMFF94 for intermolecular interactions,[59] primarily hydrogen bonds. Avogadro[60] was used to create the systems, using the same bond distances as the literature. Structures containing charged species were excluded in these tests. The following elements were used in the Halgren data set: C, H, N, O, S, and F.

Each data set was optimized using ANI-2x, and the energy was calculated using the same potential. The same was done for DFT. The interaction energy is defined as the difference between the energy of the complex (E_{AB}) and the sum of the energies of the individual molecules ($E_A + E_B$) at the same geometries, as in the dimer complex (eq 5). This is the common approach in the field (not including deformation energy in the interaction energy) because it allows for the contribution to the total energy from nonbonded interaction to be studied independently of the molecules' other properties.

$$\mathrm{IE} = E_{AB} - (E_A + E_B) \tag{5}$$

Table 4 shows the MAE and RMSE of the interaction energies calculated with ANI-2x and DFT. The results in this table do not include deformation energy. Error metrics of interaction energies with deformation energy included are shown in the Supporting Information, TS5. The Supporting Information, TS6, provides a deeper look into the X40 data set, showing the error metrics for each interaction type and how many systems were provided for each. It was found that the interaction type with the highest error is hydrogen bonding. However, when comparing these values, it is important to note that the Halgren data set is larger in size and contains a more diverse set of systems with only hydrogen bonds, where the X40 data set is smaller and contains a large range of interaction types, with only eight systems representing the hydrogen bond. To reduce the errors across separate interaction types, more strategic dimer sampling is necessary. Table 4 also shows the same error metrics for the X40 data set comparing ANI and DFT to CCSD(T)/CBS energy values. ANI shows a lower error than DFT compared to CCSD; however, the values are comparable. This shows that ANI-2x can be substituted for DFT when studying these types of systems.

Table 4. MAE and RMSE Comparing ANI-2x To DFT Interaction Energies for the X40 Data Set and the Data Set from Halgren and the MAE and RMSE of the Interaction Energies Calculated by ANI-2x and DFT Compared to CCSD(T)/CBS Calculations from the X40 Data Set[a]

error metric (kcal/mol)	ANI vs DFT Halgren	ANI vs DFT X40	ANI vs CCSD(T) (X40)	DFT vs CCSD(T) (X40)
MAE	1.24	1.51	1.71	1.93
RMSE	1.77	2.44	2.43	2.71

[a]All in kcal/mol.

Discussions and Concluding Remarks. Continued development of new and improved deep learning molecular potentials promises to change the way molecular simulation is conducted for years to come. As these potentials improve, their range of applicability grows. The presented ANI-2x potential provides chemically accurate energy predictions for molecules containing seven atomic elements (H, C, N, O, S, F, and Cl) within the thermal applicability range of interest to biochemists and computational drug designers. It has been tested across a wide range of applications relevant to drug development on diverse test sets. When compared to trusted QM methods, ANI-2x shows similar accuracy to DFT and outperforms MMFF94 and PM6 for conformer scoring. Another model has been developed by Stevenson et al. using a similar methodology (Schrodinger-ANI) that incorporates S, F, Cl, and P.[61] Although ANI-2x shows a slightly higher error on the Genentech torsion benchmark than Schrodinger-ANI, we believe that the inclusion of force training and the diverse sampling techniques used when training ANI-2x makes it better suited for applications such as molecular dynamics. Still, Schrodinger-ANI is further evidence that general-purpose machine learning models can be extended to new chemical elements without sacrificing accuracy on previously sampled systems.

The ANI-2x potential retains the same computational scaling as classical force fields, providing a 10^6 speedup over the DFT level it has been trained against. Furthermore, the addition of more atomic species has a negligible impact on the overall numerical speed of ANI potentials, despite $O(N^2)$ growth in the size of the atomic environment descriptors. Parameterization to new chemical elements has been shown to have no noticeable negative impact on the accuracy of ANI-2x. In fact, the addition of molecules containing new chemical elements to the training set can improve the model's accuracy by increasing the diversity of chemistry in the training data set.

Looking forward, the addition of long-range interactions by combining ANI-2x and ML-based charge models, such as the affordable charge assignment (ACA)[62] model, can provide corrections to missing long-range interactions. Further studies need to be carried out with models such as HIP-NN,[6] AIMNet,[28] and Schnet[20] to determine if iterative long-range information transfer scheme provides advantages in the realm of general-purpose potentials and to quantify these advantages versus overall computational cost. To further increase the applicability of general-purpose ML potentials, techniques and data sets need to be developed to allow the models to describe more than just singlet spin and neutral charge states.

Small-molecule force field development is a challenging, labor-intensive effort and cannot be easily automated. Force fields are usually developed by large consortia of academic and industrial groups working together over an extended period of time to parametrize a model addressing a particular class of problems. The ANI-2x potential developed in this work and other ML potentials provide an appealing alternative approach to traditional methods. The ANI methodology, coupled with

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

active learning data sampling, provides a systematic approach for generating such methods. It drastically reduces the human effort required for fitting a force field, automates the method development, and provides systematic improvement. Using a neural network, as universal approximators, does not require one to choose a functional form. These capabilities will dramatically accelerate development of new models, while also producing more accurate force fields with clear dependencies on reference QM data and tools for uncertainty quantification.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00121.

Breakdown of the conformer scoring results for each molecule; MAE and RMSE for nonbonded interaction energies including deformation energies for the X40 data set and a breakdown by the interaction type for the Halgren data set; box-and-whisker plot comparing DFT, ANI-2x, and OPLS on the Genetech torsion benchmark; comparison of ANI-2x's error and ensemble standard deviation; hyperparameters and network architecture used to train ANI-2x; and description of the COMP6v2 benchmark (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Justin S. Smith − *Center for Non-Linear Studies and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*; Email: just@lanl.gov

Olexandr Isayev − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*; orcid.org/0000-0001-7581-8497; Email: olexandr@olexandrisayev.com

Adrian E. Roitberg − *Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States*; orcid.org/0000-0003-3963-8784; Email: roitberg@ufl.edu

Authors

Christian Devereux − *Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States*

Kate K. Davis − *Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States*

Kipton Barros − *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States*

Roman Zubatyuk − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00121

Funding

A.E.R. thanks NSF CHE-1802831, and O.I. thanks NSF CHE-1802789. This work was partially supported by the LANL Laboratory Directed Research and Development (LDRD) and the Advanced Simulation and Computing Program (ASC) programs. We acknowledge computer time on the CCS-7 Darwin cluster at LANL. JSS was partial supported by the Center for Nonlinear Studies (CNLS) and the Nicholas C. Metropolis Postdoctoral Fellowship. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. The authors acknowledge Extreme Science and Engineering Discovery Environment (XSEDE) award DMR110088, which is supported by the NSF grant number ACI-1053575. This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by the National Science Foundation award OAC-1818253. This research, in part, was done using resources provided by the Open Science Grid[63,64] which is supported by the award 1148698 and the U.S. DOE Office of Science. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. DOE Office of Science. We gratefully acknowledge the support and hardware donation from NVIDIA Corporation and express our special gratitude to Jonathan Lefman.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377−381.

(2) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268−276.

(3) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* **2018**, *4*, No. eaap7885.

(4) Behler, J. Representing Potential Energy Surfaces by High-Dimensional Neural Network Potentials. *J. Phys.: Condens. Matter* **2014**, *26*, 183001.

(5) 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.

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

(7) Gastegger, M.; Schwiedrzik, L.; Bittermann, M.; Berzsenyi, F.; Marquetand, P. wACSF-Weighted atom-centered symmetry functions as descriptors in machine learning potentials. *J. Chem. Phys.* **2018**, *148*, 241709.

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

(9) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511−522.

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

(11) 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.

(12) Han, J.; Zhang, L.; Car, R.; Weinan, E. Deep Potential: A General Representation of a Many-Body Potential Energy Surface. *Commun. Comput. Phys.* **2018**, *23*, 629−639.

(13) 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.

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

(15) Jiang, B.; Li, J.; Guo, H. Potential Energy Surfaces from High Fidelity Fitting of Ab Initio Points: The Permutation Invariant

Polynomial - Neural Network Approach. *Int. Rev. Phys. Chem.* **2016**, *35*, 479−506.
(16) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. Representation of Intermolecular Potential Functions by Neural Networks. *J. Phys.* **1998**, *102*, 4596−4605.
(17) Morawietz, T.; Sharma, V.; Behler, J. A Neural Network Potential-Energy Surface for the Water Dimer Based on Environment-Dependent Atomic Energies and Charges. *J. Chem. Phys.* **2012**, *136*, 064103.
(18) Kolb, B.; Zhao, B.; Li, J.; Jiang, B.; Guo, H. Permutation Invariant Potential Energy Surfaces for Polyatomic Reactions Using Atomistic Neural Networks. *J. Chem. Phys.* **2016**, *144*, 224103.
(19) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114*, 3371−3383.
(20) 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.
(21) Yao, K.; Herr, J. E.; Toth, D. W.; Mcintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261−2269.
(22) 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.
(23) 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.
(24) Artrith, N.; Morawietz, T.; Behler, J. High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *83*, 153101.
(25) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.
(26) Fletcher, T. L.; Popelier, P. L. A. Multipolar Electrostatic Energy Prediction for All 20 Natural Amino Acids Using Kriging Machine Learning. *J. Chem. Theory Comput.* **2016**, *12*, 2742−2751.
(27) Fletcher, T. L.; Davie, S. J.; Popelier, P. L. A. Prediction of Intramolecular Polarization of Aromatic Amino Acids Using Kriging Machine Learning. *J. Chem. Theory Comput.* **2014**, *10*, 3708−3719.
(28) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, No. eaav6490.
(29) 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.
(30) De Jong, M.; Chen, W.; Notestine, R.; Persson, K.; Ceder, G.; Jain, A.; Asta, M.; Gamst, A. A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-Nary Inorganic Polycrystalline Compounds. *Sci. Rep.* **2016**, *6*, 34256.
(31) Legrain, F.; Carrete, J.; Van Roekeghem, A.; Curtarolo, S.; Mingo, N. How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids. *Chem. Mater.* **2017**, *29*, 6220−6227.
(32) 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*, No. e1603015.
(33) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(34) 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.
(35) Faber, F. A.; Lindmaa, A.; Von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC2D6) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.
(36) Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. *ACS Cent. Sci.* **2019**, *5*, 57−64.
(37) 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.
(38) Smith, J. S.; Isayev, O.; Roitberg, A. E. Data Descriptor: ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017**, *4*, 170193.
(39) Artrith, N.; Behler, J. High-Dimensional Neural Network Potentials for Metal Surfaces: A Prototype Study for Copper. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 045439.
(40) 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.
(41) Fink, T.; Reymond, J.-L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342−353.
(42) Fink, T.; Bruggesser, H.; Reymond, J.-L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. *Angew. Chem., Int. Ed.* **2005**, *44*, 1504−1508.
(43) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.
(44) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66x8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.
(45) Landrum, G. *RDkit: Open-Source Cheminformatics*, 2016.
(46) Davies, M.; Nowotka, M.; Papadatos, G.; Atkinson, F.; van Westen, G.; Dedman, N.; Ochoa, R.; Overington, J. MyChEMBL: A Virtual Platform for Distributing Cheminformatics Tools and Open Data. *Challenges* **2014**, *5*, 334−337.
(47) Kesharwani, M. K.; Karton, A.; Sylvetsky, N.; Martin, J. M. L. The S66 Non-Covalent Interactions Benchmark Reconsidered Using Explicitly Correlated Methods Near the Basis Set Limit. *Aust. J. Chem.* **2018**, *71*, 238−248.
(48) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The atomic simulation environment-a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
(49) Smith, J. S.; Lubbers, N.; Thompson, A. P.; Barros, K. Simple and Efficient Algorithms for Training Machine Learning Potentials to Force Data. **2020**, arXiv:2006.05475, ChemRxiv, 1−7.
(50) Ballell, L.; Bates, R. H.; Young, R. J.; Alvarez-Gomez, D.; Alvarez-Ruiz, E.; Barroso, V.; Blanco, D.; Crespo, B.; Escribano, J.; González, R.; et al. Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. *ChemMedChem* **2013**, *8*, 313−321.
(51) Kanal, I. Y.; Keith, J. A.; Hutchison, G. R. A Sobering Assessment of Small-Molecule Force Field Methods for Low Energy Conformer Predictions. *Int. J. Quantum Chem.* **2018**, *118*, No. e25512.
(52) 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.
(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09* Revision D.01; Gaussian Inc.: Wallingford CT. Gaussian 09 Revision C.01, 2010.

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202

(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian, Inc.: Wallingford CT, 2016.
(55) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *1*, 73.
(56) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.
(57) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular Force Fields with Gradient-Domain Machine Learning: Construction and Application to Dynamics of Small Molecules with Coupled Cluster Forces. *J. Chem. Phys.* **2019**, *150*, 114102.
(58) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(59) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(60) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeerschd, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 17.
(61) Stevenson, J.; Jacobson, L. D.; Zhao, Y.; Wu, C.; Maple, J.; Leswing, K.; Harder, E.; Abel, R. Schrodinger-ANI: An Eight-Element Neural Network Interaction Potential with Greatly Expanded Coverage of Druglike Chemical Space. **2019**, arXiv:1912.05079.
(62) 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.
(63) Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Würthwein, F.; Foster, I.; Gardner, R.; Wilde, M.; Blatecky, A.; McGee, J.; Quick, R. The Open Science Grid. *J. Phys.: Conf. Ser.* **2007**, *78*, 012057.
(64) Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Würthwein, F. The Pilot Way to Grid Resources Using GlideinWMS. *2009 WRI World Congress on Computer Science and Information Engineering*, 2009; Vol. *2*, pp 428−432.

https://dx.doi.org/10.1021/acs.jctc.0c00121
J. Chem. Theory Comput. 2020, 16, 4192−4202