doi.org/10.26434/chemrxiv.11819268.v1

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

Christian Devereux, Justin Smith, Kate Davis, Kipton Barros, Roman Zubatyuk, Olexandr Isayev, Adrian Roitberg

Submitted date: 06/02/2020 • Posted date: 07/02/2020

Citation information: Devereux, Christian; Smith, Justin; Davis, Kate; Barros, Kipton; Zubatyuk, Roman; Isayev, Olexandr; et al. (2020): Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11819268.v1

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, 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 non-bonded interactions. ANI-2x is shown to accurately predict molecular energies compared to DFT with a ~10^6 factor speedup and a negligible slowdown compared to ANI-1x. The resulting model is a valuable tool for drug development that can potentially replace both quantum calculations and classical force fields for myriad applications.

File list (2)

Extending the applicability of the ANI deep learning molecular potential to Sulfur and Halogens

Christian Devereux[1], Justin S. Smith[2,3,], Kate K. Davis [1], Kipton Barros[3], Roman Zubatyuk[4], Olexandr Isayev[4,*], Adrian E. Roitberg[1,*]*

[1] Department of Chemistry, University of Florida, Gainesville, FL 32611, USA
[2] Center for Non-Linear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[3] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[4] Department of Chemistry, Carnegie Mellon University, Pittsburgh PA, 15213

* Corresponding authors; email: JSS (just@lanl.gov), OI (olexandr@olexandrisayev.com) and AER (roitberg@ufl.edu)

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, 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 non-bonded interactions. ANI-2x is shown to accurately predict molecular energies compared to DFT with a ~10^6 factor speedup and a negligible slowdown compared to ANI-1x. The resulting model is a valuable tool for drug development that can potentially replace both quantum calculations and classical force fields for myriad applications.

KEYWORDS. Active learning, machine learning, molecular potentials, force field, neural network, deep learning

Introduction

The application of machine learning (ML) methods in chemistry is rising in popularity due to success in areas such as robotic chemical synthesis[1], drug and materials 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[6,11–20], atomic partial charges[21,22], dipoles, and quadrupoles[20,23] with accuracies greatly surpassing classical physics-based techniques. Some researchers have shown that models can be trained to multiple properties simultaneously.[24] The speed, accuracy, and transferability of ML property predictors promises to revolutionize the computational design of drugs and materials by bridging the speed vs. 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, e.g. geometry optimization or molecular dynamics simulations. ML potentials for both materials[19,25–27] and biological[11,20,28] (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 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, e.g. proteins and large drug molecules, intractable. 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 dataset of molecular conformations from small molecules as a training dataset. 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, we can often achieve both linear scaling in system size, and extensible potentials. 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. While 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 in 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.[29–32]

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 dataset and Behler and Parrinello type descriptors[33] with a modified symmetry function.[11] The ANI-1 dataset[34] (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 lead to sparse coverage of chemical space, e.g. torsion space. In response, an active learning algorithm using query by committee (QBC) for selecting new data was employed to automatically diversify the dataset[23]. 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 is 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. Due to the vastness of chemical space, the early proof of concept ANI-1x dataset only sampled molecules with a limited number of atomic elements: H, C, N, and O. While these four elements cover a large swath of interesting organic chemistry, further expansion and diversification of this dataset 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[35] (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 [https://github.com/isayev/ASE_ANI].

Methods

Building a database of S, F, Cl containing molecules

The database of molecules used to build the active learning-based ANI-2x training dataset is composed of molecules from a variety of sources, including the GDB-11[36,37] database, the CheMBL[38] database, and the s66x8[39] benchmark. The GDB-11 dataset contains an enumeration of chemically feasible organic molecules containing the heavy elements: C, N, O, and F. From this database, we combinatorically replaced the chemical symbols O with S and F with Cl for all molecules containing up to 8 non-hydrogen atoms. From the ChEMBL[38] database, molecules

containing S, F, and Cl were sampled. Also, conformers of amino acids and di-peptides containing S were randomly generated using the Rdkit[40] chem informatics package. These sampling techniques mirror those used in building the ANI-1x dataset[35].
In the results section below, we commonly show the mean absolute error (MAE) and root mean squared error (RMSE) as 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 atomic simulation environment (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.[29,35] 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 non-bonded interaction sampling, were required to contain S, F, or Cl. To generate non-equilibrium 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.[29,35] As with the ANI-1x active learning process, only small molecules are searched in early iterations with molecule size increasing as the process proceeds. In the end, more than 50 active learning cycles were carried out yielding a dataset of 4,695,707 molecular conformations from 13,405 chemical isomers. Combined with the original ANI-1x dataset and torsion refinement dataset from the ANI-1ccx work, the final ANI-2x dataset consists of 8.9 million molecular conformations.

Active-learning torsion refinement

As previously presented in our work developing the ANI-1ccx potential[35], we carried out an active learning torsion refinement on a randomly selected subset of molecules from the ChEMBL drug molecule database[38,41]. 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 degrees, 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 then added to the training dataset. This process is referred to as "torsional

refinement". A torsional refinement was carried out during each iteration of the active learning process.

Active-learning non-bonded interaction refinement

To improve sampling of non-bonded interactions we use the s66x8[42] 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 non-bonded interactions since these dimers represent the smallest systems containing their respective interaction. This active learning cycle was bootstrapped from the ANI-1x dataset 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 Atomic Simulation Environment[43] package. NVT molecular dynamics simulations were then carried out using the latest ANI potential. Every 5 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 were generated by taking all waters within 6 Å from the center of the N_{th} water. A random selection of waters was then deleted from each cluster until between 2 and 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. This termination criteria helps 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 due to 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. ANI-2x was trained using the loss function:

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

Where $\hat{E}_i$ and $\hat{f}_{ij}$ are the energies and forces predicted by any for a given molecule, E_i and f_{ij} are the QM energy and forces. $l_0$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. 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 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.

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 non-bonded 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[44]. It contains all atomic elements (C, H, N, O, S, F, Cl) considered in this work. For the 2D potential energy scans we selected four molecules (bendamustine, cysteine-dipeptide, 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[45]. This collection includes drug-like ligands used to access performance of conformer generating methods like OMEGA or ETKDG.

Torsion Profiles

The Genentech torsion benchmark[46] 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-degree 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[46]. 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 slightly higher error than its reference DFT.

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.

Method	MAE	RMSE
DFT	0.35	0.50
ANI-2x	0.42	0.59
OPLS3	0.66	1.02

Table 1. MAE and RMSE between ANI-2x, ωB97X/6-31G*, and OPLS3 against CCSD(T)/CBS on the Genentech torsion benchmark[46].

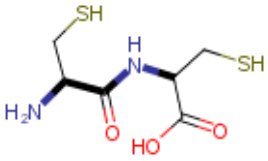
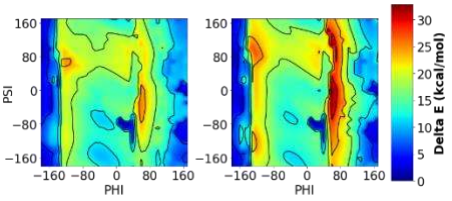
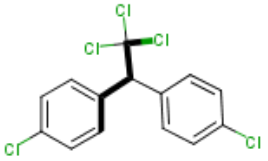
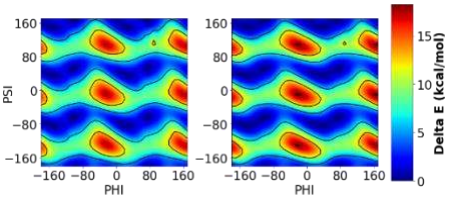
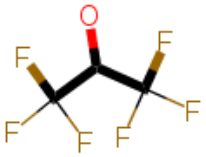
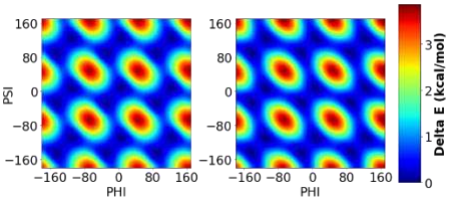
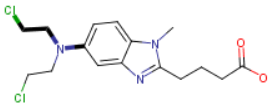
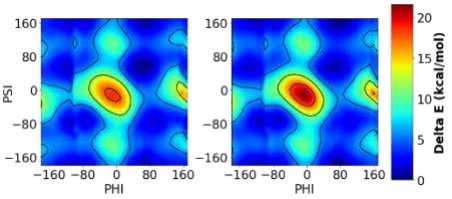
Name	Molecule	MAE	RMSE	Scan (Left:ANI Right:DFT)
Cysteine-Dipeptide (25 atoms)		1.75	2.55	
DDT (28 atoms)		0.53	0.70	
Hexafluoroacetone (10 atoms)		0.08	0.11	
Bendamustine (44 atoms)		0.50	0.66	

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 those 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 of the relative energies in kcal/mol between ANI and DFT.

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[47] software package while DDT, hexafluoroacetone, and bendamustine DFT optimizations were performed using Gaussian 16[48]. ANI-2x accurately predicts the location of the minima and maxima for all four molecules. For DDT, hexafluoroacetone, and bendamustine, ANI shows sub-chemical accuracy. ANI shows 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. In the SI, 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 since 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, 0.08 kcal/mol/atom^{-1/2}. Since many publications on ML potentials present results as uncorrected per atom MAE we also provide these results: 0.07, 0.02, 0.01, 0.01 kcal/mol/atom. However, we stress that the latter metric of comparison is unreliable since 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 due to 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 due to its relatively accurate forces and efficiency. However, even MP2 is too costly to perform high throughput conformer searches across large datasets. 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 and missed candidates as well as 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 druglike molecules. These molecules were taken from the test set used in recent work on validating force fields by Kanal, Keith, and Hutchinson[45]. 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 was embedded in 3D space using the Rdkit[40] 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 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.

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[49], and all MMFF94 calculations were performed using the Open Babel software package[50]. Mean absolute error and root mean square error 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 lower error compared to MP2/cc-PVTZ for both 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 seconds compared to 552 seconds with ωB97X/6-31G*. While it's important to note that these calculations were done using different hardware (ANI-2x on a GPU, DFT on a CPU), we believe 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.

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

Table 2. Geometry comparison of the optimized structures predicted by DFT, ANI, PM6, and MMFF94 with the optimized structures predicted by MP2.

	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)	1.23	1.91	2.96	3.83	1.68	1.90	1.96
RMSE (kcal/mol)	2.01	2.67	3.78	5.17	2.20	2.51	2.62

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. Mean absolute error (MAE) and root mean square error (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.

COMP6v2

The COMP6 benchmark from our recent work on applying active learning techniques to build general-purpose ML potentials[35] 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 numbers of molecules and conformers in COMP6v2 is provided in the supplemental information TS8.

The errors shown in Figure 2a, Figure 2b, and Figure 2c 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 dataset corresponds to the conformer space visited in near ambient temperature dynamical simulation. In Figure 2a, within the 200kcal/mol energy range, most of the benchmarks achieve sub-chemical accuracy (1 kcal/mol) errors, while total energy errors (Figure 2b) tend to be larger due to 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

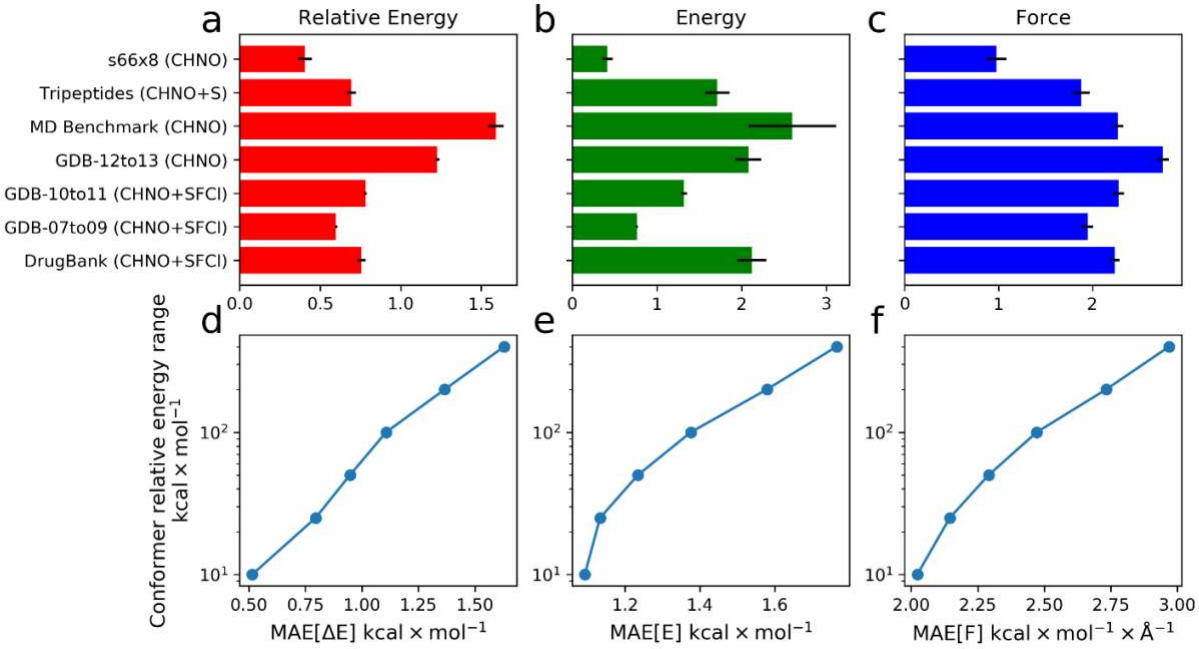


Figure 2. a,b,c) Errors for molecules within 200 kcal/mol from the minima compared to DFT reference data from the updated COMP6 benchmark for the chemical elements C, H, N, O, S, F, and Cl. d,e,f.) Relative conformer energy range considered vs. mean absolute error (MAE) of relative energy, total energy,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

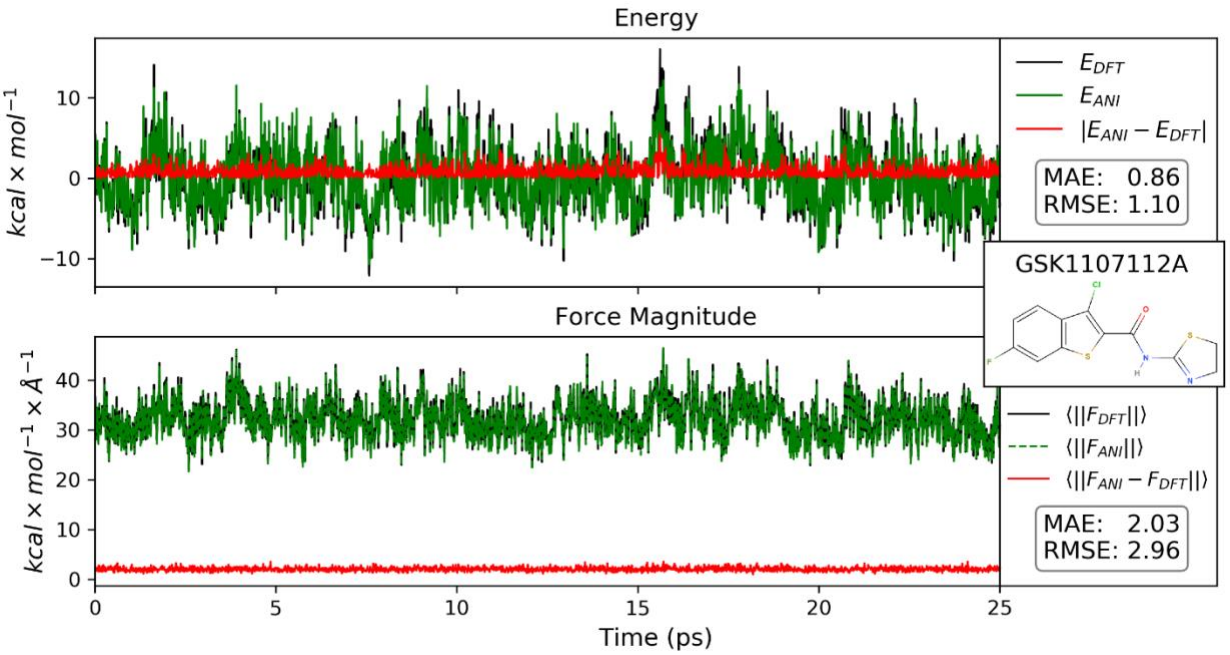


Figure 3. Energies (shifted to the mean) and force magnitudes along with corresponding errors for a 25ps NVT molecular dynamics (MD) trajectory using the Langevin thermostat at 300K. This figure represents the final 25ps of a 1.5ns MD simulation in vacuum. The drug ligand GSK1107112A was chosen as an example because it contains all atomic elements (C, H, N, O, S, F, 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.

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 since each atomic energy prediction has an error associated with it.

Figures 2d-2f show 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 dataset due to the sampling techniques used, i.e. 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 doesn't necessarily represent the error obtained in room temperature simulation.[51]

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 a 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 300K. This trajectory shows the applicability of ANI-2x to MD simulations for systems containing the chemical elements C, H, N, and O as well as 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 dataset. 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 timescale simulations, ANI-2x is still sampling structures that agree well with reference DFT. The ANI potential took approximately 12.0 GPU hours to run the 3.75 million steps require 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.

Non-bonded interactions

Two datasets were chosen to show that ANI-2x accurately predicts non-bonded interaction energies. The X40 dataset was obtained from the Benchmark Energy and Geometry Database[52]. 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 dataset was taken from work done by Thomas A. Halgren in 1996 to measure the performance of MMFF94 for intermolecular interactions[53], primarily hydrogen bonds. Avogadro[54] 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 dataset: C, H, N, O, S, and F.

Each dataset 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. 1). 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.

$$IE = E_{AB} - (E_A + E_B) \quad (1)$$

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 SI TS5. SI TS6 provides a deeper look into the X40 dataset, 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, the Halgren data set is larger in size and contains a more diverse set of systems with only hydrogen bonds, where the X40 dataset is smaller and contains a large range of interaction types, with only 8 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 dataset 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.

Error Metric	ANI vs. DFT Halgren	ANI vs. DFT X40	ANI vs. CCSD(T) (X40)	DFT vs. CCSD(T) (X40)
MAE	1.24	1.51	1.7	1.9
RMSE	1.77	2.44	2.4	2.7

Table 4. MAE and RMSE comparing ANI-2x to DFT interaction energies for the X40 dataset and the dataset from Halgren, as well as the MAE and RMSE of the interaction energies calculated by ANI-2x and DFT compared to CCSD(T)/CBS calculations from the X40 dataset. 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, Cl) within the thermal applicability range of interest to bio-chemists 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[55]. Although ANI-2x shows slightly higher error on the Genentech torsion benchmark than Schrodinger-ANI, we believe the inclusion of force training and the diverse sampling techniques used when training ANI-2x makes 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. Further, 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 dataset.

Looking forward, the addition of long range interactions by combining ANI-2x and ML-based charge models, such as the Affordable Charge Assignment (ACA)[56] model, can provide corrections to missing long range interactions. Further studies need to be carried out with models such as HIP-NN[6], AIMNet[24], and Schnet[19] to determine if iterative long-range information transfer scheme provides advantages in the realm of general-purpose potentials, and to quantify those advantages vs. overall computational cost. To further increase the applicability of general-purpose ML potentials, techniques and datasets 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 as well as other ML potentials provide an appealing alternative approach to traditional methods. The ANI methodology, coupled with active learning data sampling, provides a systematic approach to generating such methods. It drastically reduces the human effort required for fitting a force field, it 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

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

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 NSF grant number ACI-1053575. This research in part was done using resources provided by the Open Science Grid[57,58] 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. https://doi.org/10.1038/s41586-018-0307-8.

(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* (2), 268–276. https://doi.org/10.1021/acscentsci.7b00572.

(3) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* **2018**, *4* (7), eaap7885. https://doi.org/10.1126/sciadv.aap7885.

(4) Behler, J. Representing Potential Energy Surfaces by High-Dimensional Neural Network Potentials. *Journal of Physics Condensed Matter*. 2014, p 183001. https://doi.org/10.1088/0953-8984/26/18/183001.

(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* (11), 5255–5264. https://doi.org/10.1021/acs.jctc.7b00577.

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

(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* (24), 241709. https://doi.org/10.1063/1.5019667.

(8) Glielmo, A.; Sollich, P.; De Vita, A. Accurate Interatomic Force Fields via Machine Learning with Covariant Kernels. *Phys. Rev. B* **2017**, *95* (21), 214302. https://doi.org/10.1103/PhysRevB.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* (1), 511–522. https://doi.org/10.1021/acs.jpcc.6b10908.

(10) Kruglov, I.; Sergeev, O.; Yanilkin, A.; Oganov, A. R. Energy-Free Machine Learning Force Field for Aluminum. *Sci. Rep.* **2017**, *7*, 8512. https://doi.org/10.1038/s41598-017-08455-3.

(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* (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.

(12) Han, J.; Zhang, L.; Car, R.; E, W. Deep Potential: A General Representation of a Many-Body Potential Energy Surface. *Commun. Comput. Phys.* **2017**, *23*, 629–639.

(13) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115* (16), 1032–1050. https://doi.org/10.1002/qua.24890.

(14) 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* (3), 479–506. https://doi.org/10.1080/0144235X.2016.1200347.

(15) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. Representation of Intermolecular Potential Functions by Neural Networks. *J. Phys.* **1998**, *102* (24), 4596–4605.

(16) 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* (6), 064103. https://doi.org/10.1063/1.3682557.

(17) 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* (22), 224103. https://doi.org/10.1063/1.4953560.

(18) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114* (10), 3371–3383. https://doi.org/10.1021/jp9105585.

(19) 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* (24), 241722. https://doi.org/10.1063/1.5019779.
(20) 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.* **2017**, *9* (8), 2261–2269. https://doi.org/10.1039/C7SC04934J.
(21) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58* (3), 579–590. https://doi.org/10.1021/acs.jcim.7b00663.
(22) 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* (9), 4687–4698. https://doi.org/10.1021/acs.jctc.8b00524.
(23) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8* (10), 6924–6935. https://doi.org/10.1039/C7SC02267K.
(24) 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*, eaav6490. https://doi.org/10.1126/sciadv.aav6490.
(25) 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* (C). https://doi.org/10.1016/j.jcp.2014.12.018.
(26) 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. https://doi.org/10.1038/srep34256.
(27) 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* (15), 6220–6227. https://doi.org/10.1021/acs.chemmater.7b00789.
(28) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3* (5), e1603015. https://doi.org/10.1126/sciadv.1603015.
(29) 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. https://doi.org/10.1038/s41467-019-10827-4.
(30) 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* (13), 135502. https://doi.org/10.1103/PhysRevLett.117.135502.
(31) 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* (1), 57–64. https://doi.org/10.1021/acscentsci.8b00551.
(32) Chmiela, S.; Sauceda, H. E.; Müller, K. R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9* (1), 3887. https://doi.org/10.1038/s41467-018-06169-2.
(33) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401.
(34) 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. https://doi.org/10.1038/sdata.2017.193.

(35) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148* (24). https://doi.org/10.1063/1.5023802.
(36) Fink, T.; Raymond, 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 Discove. *J. Chem. Inf. Model.* **2007**, *47* (2), 342–353. https://doi.org/10.1021/ci600423u.
(37) Fink, T.; Bruggesser, H.; Reymond, J. L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. *Angew. Chemie - Int. Ed.* **2005**, *44* (10), 1504–1508. https://doi.org/10.1002/anie.200462457.
(38) 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* (D1), D1083–D1090. https://doi.org/10.1093/nar/gkt1031.
(39) 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* (31), 20905–20925. https://doi.org/10.1039/C6CP00688D.
(40) Landrum, G. RDkit: Open-source Cheminformatics.
(41) 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* (2), 334–337. https://doi.org/10.3390/challe5020334.
(42) 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* (4), 238–248. https://doi.org/10.1071/CH17588.
(43) 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. *Journal of Physics Condensed Matter*. IOP Publishing July 2017, p 273002. https://doi.org/10.1088/1361-648X/aa680e.
(44) 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* (2), 313–321. https://doi.org/10.1002/cmdc.201200428.
(45) 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* (5), e25512. https://doi.org/10.1002/qua.25512.
(46) 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* (6), 1265–1275. https://doi.org/10.1021/acs.jcim.6b00614.
(47) 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. Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT. *Gaussian 09 Revision C.01*. 2010.
(48) 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**.
(49) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**. https://doi.org/10.1002/wcms.81.
(50) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**. https://doi.org/10.1186/1758-2946-3-33.
(51) 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* (11), 114102. https://doi.org/10.1063/1.5078687.

(52) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8* (11), 4285–4292. https://doi.org/10.1021/ct300647k.

(53) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17* (5-6), 520–552. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<520::AID-JCC2>3.0.CO;2-W.

(54) 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. Cheminform.* **2012**. https://doi.org/10.1186/1758-2946-4-17.

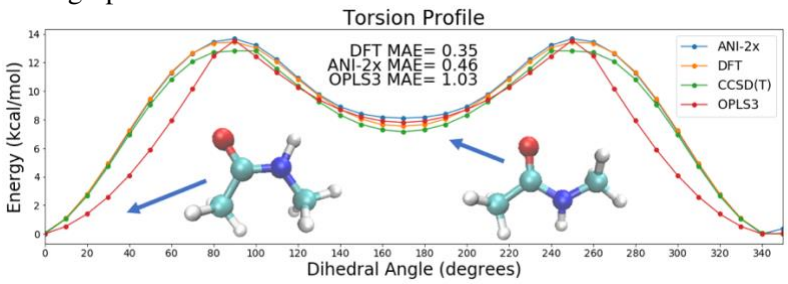
(55) Stevenson, J.; Jacobson, L. D.; Zhao, Y.; Wu, C.; Maple, J.; Leswing, K.; Harder, E.; Abel, R. Schrodinger-ANI: An Eight-Element Neural Network Interaction Potential with Greatly Expanded Coverage of Druglike Chemical Space. December 12, 2019. https://doi.org/10.26434/chemrxiv.11319860.v1.

(56) 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. https://doi.org/10.1021/acs.jpclett.8b01939.

(57) 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. In *Journal of Physics: Conference Series*; IOP Publishing, 2007; Vol. 78, p 012057. https://doi.org/10.1088/1742-6596/78/1/012057.

(58) Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Würthwein, F. The Pilot Way to Grid Resources Using GlideinWMS. In *2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009*; IEEE, 2009; Vol. 2, pp 428–432. https://doi.org/10.1109/CSIE.2009.950.

TOC graphic


Torsion Profile
DFT MAE= 0.35
ANI-2x MAE= 0.46
OPLS3 MAE= 1.03
ANI-2x
DFT
CCSD(T)
OPLS3
Energy (kcal/mol)
Dihedral Angle (degrees)

Supplementary Information

TS1: Mean absolute error, root mean square error, R-squared correlation, and spearman rank between relative conformer energies predicted by ANI and MP2 for each molecule in the conformer test set. Energy range is the difference between the highest and lowest energy conformer as predicted by MP2.

Molecule	MAE (kcal/mol)	RMSE (kcal/mol)	R_2	Spearman Rank	Energy Range (kcal/mol)
omegacsd_FABSOQ10	1.092	1.329	0.365	0.467	1.262
omegacsd_DIAVER	1.272	1.527	0.754	0.527	4.762
omegacsd_CDBMPI10	1.556	1.981	0.589	0.571	6.125
omegacsd_GALSEM	1.918	2.535	0.422	0.318	7.416
astex_1n2j	1.428	1.922	0.781	0.922	8.088
omegapdb_1m5f	1.433	1.753	0.739	0.854	8.798
omegapdb_1v2k	1.213	1.583	0.853	0.901	8.958
astex_1of1	1.870	1.870	0.869	0.911	10.252
omegacsd_SEYMIS	2.700	4.341	0.235	0.484	10.491
omegacsd_FIBREN	2.304	2.843	0.785	0.871	10.613
omegapdb_1qxw	2.376	3.296	0.466	0.675	10.745
omegacsd_SPIRIL	1.581	1.960	0.820	0.837	11.107
omegacsd_CUHNEY10	1.542	2.121	0.785	0.794	11.107
omegapdb_2f7p	1.537	1.879	0.837	0.910	11.689
omegacsd_CEJTIU	3.239	4.264	0.705	0.901	12.056
omegapdb_2byh	3.100	4.156	0.421	0.650	12.216
omegapdb_1z1r	3.342	4.155	0.629	0.808	13.824
omegapdb_1gs5	1.978	3.354	0.694	0.827	14.936
astex_1meh	1.551	1.914	0.947	0.958	15.013
omegapdb_1xom	1.414	1.712	0.917	0.896	15.102

TS2: DFT@ANI

Mean absolute error, root mean square error, R-squared correlation, and spearman rank between relative conformer energies predicted by ANI and DFT for each molecule in the conformer test set. The structures were optimized using ANI-2x and single point calculation on those structures were performed with ωb97x/6-31g*. Energy range is the difference between the highest and lowest energy conformer as predicted by DFT.

Molecule	MAE (kcal/mol)	RMSE (kcal/mol)	R_2	Spearman Rank	Energy Range (kcal/mol)
omegacsd_FABSOQ10	0.973	1.167	0.471	0.692	2.596
omegacsd_DIAVER	1.218	1.442	0.674	0.855	4.557
omegacsd_CDBMPI10	1.332	1.623	0.707	0.462	4.756
Omegacsd_GALSEM	1.044	1.328	0.797	0.768	6.662
omegapdb_1v2k	0.909	1.099	0.926	0.946	8.732
omegacsd_SEYMIS	0.838	1.062	0.936	0.967	9.120
astex_1n2j	1.368	1.708	0.842	0.928	10.574
omegapdb_1m5f	1.399	1.745	0.795	0.854	11.020
astex_1of1	0.537	0.700	0.982	0.982	11.618
omegapdb_2f7p	1.183	1.531	0.894	0.940	12.207
omegacsd_SPIRIL	2.219	2.716	0.818	0.809	12.567
omegacsd_FIBREN	2.015	2.462	0.813	0.938	13.063
omegapdb_1qxw	2.838	3.490	0.649	0.777	13.212
omegacsd_CEJTIU	0.705	0.873	0.988	0.926	14.155
omegapdb_1xom	0.836	1.053	0.972	0.931	14.582
omegacsd_CUHNEY10	1.397	1.792	0.957	0.877	15.657
omegapdb_1gs5	1.014	1.247	0.969	0.979	15.989
omegapdb_2byh	2.624	3.234	0.675	0.824	16.724
omegapdb_1z1r	1.888	2.363	0.900	0.947	18.865
astex_1meh	2.017	3.009	0.873	0.916	20.207

TS3: ANI@DFT

Mean absolute error, root mean square error, R-squared correlation, and spearman rank between relative conformer energies predicted by ANI and DFT for each molecule in the conformer test set. The structures were optimized using ωb97x/6-31g* and single point calculation on those structures were performed with ANI-2x. Energy range is the difference between the highest and lowest energy conformer as predicted by DFT.

Molecule	MAE (kcal/mol)	RMSE (kcal/mol)	R_2	Spearman Rank	Energy Range (kcal/mol)
omegacsd_FABSOQ10	1.076	1.305	0.141	0.250	2.205
omegacsd_DIAVER	0.518	0.648	0.835	0.960	3.394
omegacsd_CDBMPI10	0.916	1.104	0.823	0.724	5.036
omegacsd_SEYMIS	1.561	1.899	0.927	0.962	7.652
omegacsd_GALSEM	1.153	1.553	0.795	0.790	8.165
omegapdb_1v2k	1.030	1.271	0.891	0.911	8.190
omegapdb_1m5f	1.406	1.715	0.775	0.897	9.466
omegacsd_SPIRIL	1.553	1.980	0.837	0.861	9.808
astex_1n2j	1.113	1.361	0.895	0.934	10.655
omegapdb_1qxw	1.636	2.075	0.788	0.852	11.424
astex_1of1	0.596	0.749	0.982	0.971	11.594
omegacsd_FIBREN	2.067	2.675	0.836	0.953	12.298
omegapdb_2f7p	1.228	1.505	0.905	0.927	12.373
omegacsd_CEJTIU	0.933	1.151	0.980	0.864	12.907
omegacsd_CUHNEY10	1.020	1.268	0.966	0.889	13.376
omegapdb_2byh	2.466	3.131	0.619	0.833	13.980
omegapdb_1xom	1.010	1.262	0.961	0.897	15.533
omegapdb_1gs5	1.021	1.306	0.961	0.962	15.945
omegapdb_1z1r	2.396	2.912	0.816	0.885	16.596
astex_1meh	1.218	1.542	0.966	0.970	18.899

TS4: ANI@ANI_DFT@DFT

Mean absolute error, root mean square error, R-squared correlation, and spearman rank between relative conformer energies predicted by ANI and DFT for each molecule in the conformer test set. The conformers were optimized with each method and the correlation between the methods is shown below.

Molecule	MAE (kcal/mol)	RMSE (kcal/mol)	R_2	Spearman Rank
omegacsd_FABSOQ10	1.251	1.521	0.135	0.179
omegacsd_DIAVER	1.433	1.702	0.174	0.322
omegacsd_CDBMPI10	1.209	1.499	0.750	0.514
omegacsd_SEYMIS	0.828	1.004	0.963	0.967
omegacsd_GALSEM	1.965	2.526	0.400	0.385
omegapdb_1v2k	1.341	1.658	0.830	0.911
omegapdb_1m5f	1.590	1.955	0.689	0.847
omegacsd_SPIRIL	1.595	1.942	0.838	0.795
astex_1n2j	1.209	1.491	0.875	0.931
omegapdb_1qxw	2.124	2.662	0.651	0.752
astex_1of1	1.320	1.978	0.873	0.931
omegacsd_FIBREN	2.024	2.526	0.803	0.923
omegapdb_2f7p	1.198	1.511	0.903	0.933
omegacsd_CEJTIU	0.916	1.132	0.981	0.887
omegacsd_CUHNEY10	0.669	0.841	0.976	0.872
omegapdb_2byh	2.932	3.871	0.468	0.672
omegapdb_1xom	1.109	1.412	0.954	0.895
omegapdb_1gs5	1.114	1.413	0.953	0.981
omegapdb_1z1r	3.673	4.757	0.599	0.761
astex_1meh	1.454	1.911	0.947	0.945

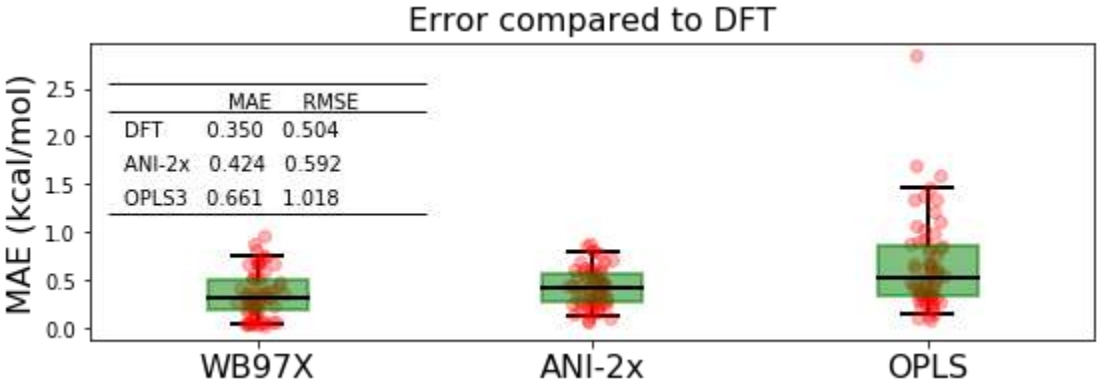
TS5: MAE and RMSE comparing ANI-2x to DFT interaction energies including deformation energy for the X40 dataset and the dataset from Halgren.

Error Metric	**Halgren**	**X40**
MAE	1.25	1.30
RMSE	1.81	1.89

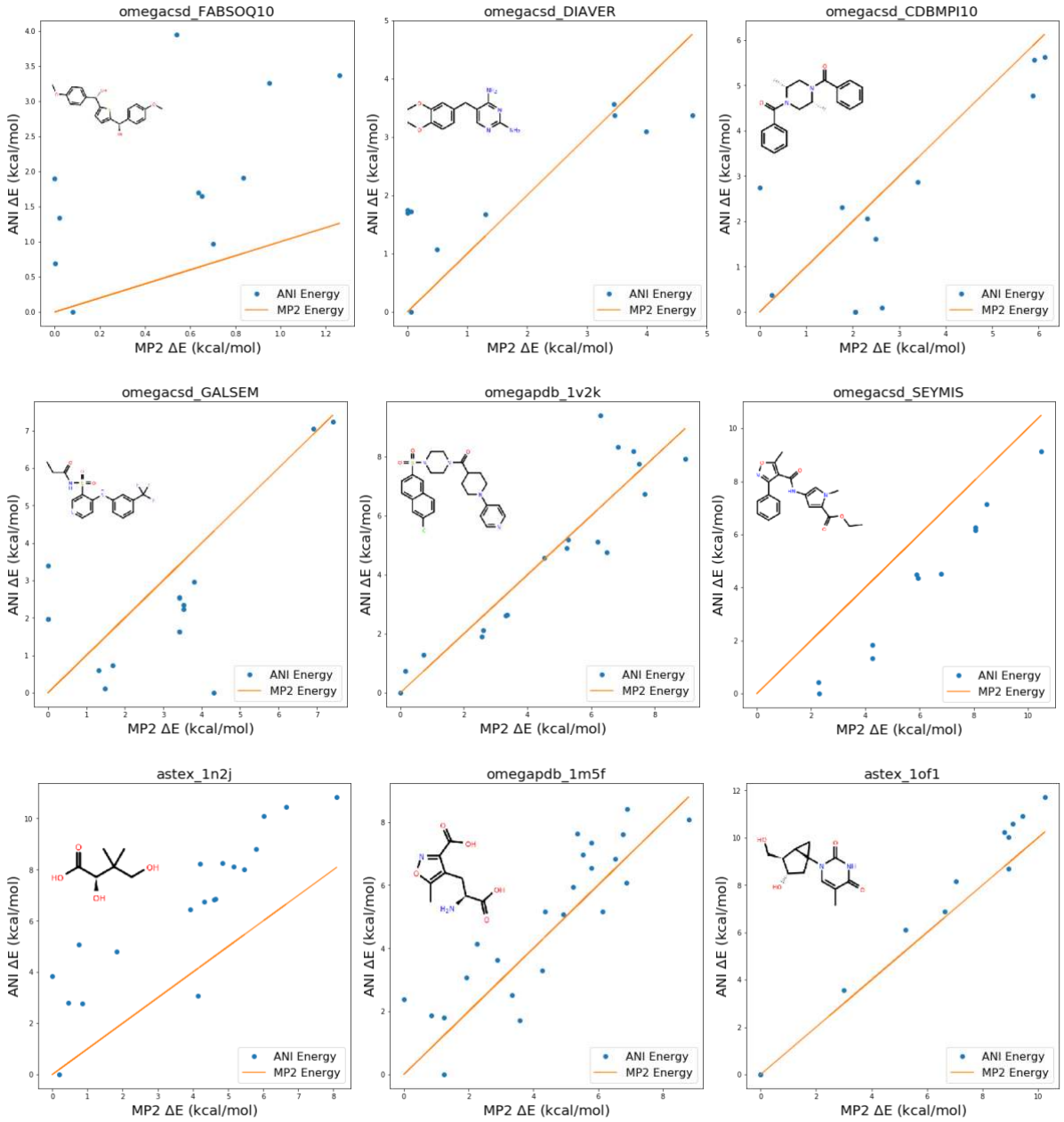
TS6: MAE and RMSE comparing ANI-2x to DFT interaction energies for the X40 dataset, separated by interaction type and whether or not deformation energy was included in the interaction energy values. The number of systems from the X40 dataset for each interaction type is also reported.

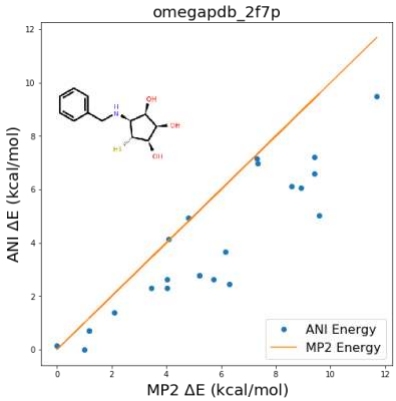
	Error Metric	**London Dispersion**	**Induction**	**Dipole-dipole Interaction**	**Stacking**	**Halogen Bonds**	**Hydrogen Bonds**
Number of Systems		2	4	2	2	4	8
No Deformation Energy	**MAE**	0.60	0.90	0.32	1.43	1.07	2.59
	RMSE	0.66	1.06	0.33	1.90	1.24	3.73
Including Deformation Energy	**MAE**	0.58	0.89	0.32	1.03	1.06	2.10
	RMSE	0.64	1.06	0.32	1.34	1.23	2.81

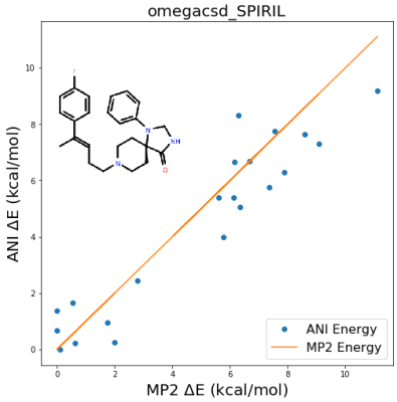
FS1: Each red point represents the MAE across all 36 conformers compared to benchmark CCSD(T)/CBS data for one molecule in the genentech benchmark. The boxes extend to include 50% of the points and the whiskers extend to include 90% of the points. The middle line represents the median value. The table shows the MAE and RMSE of each method across all conformers of all molecules.

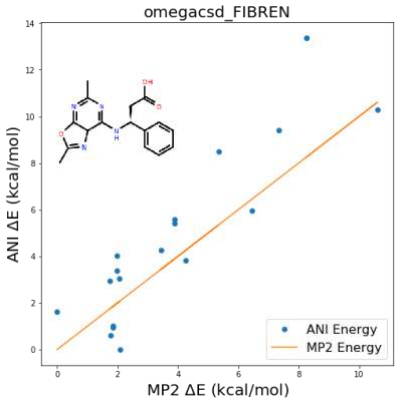


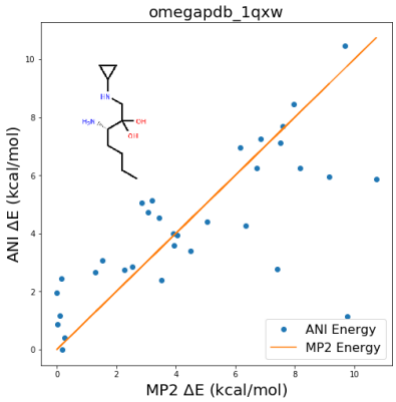
FS2: ANI-2x energies vs. MP2 energies for each of the molecules in the conformer test case. One conformation of each molecule is shown to help demonstrate the types of molecules studied.

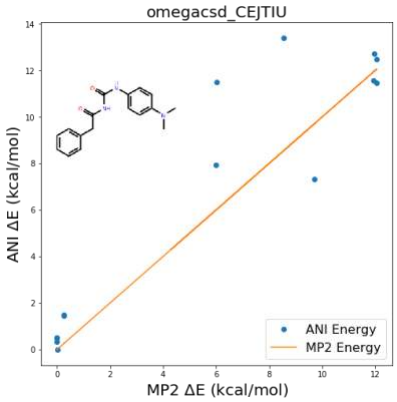


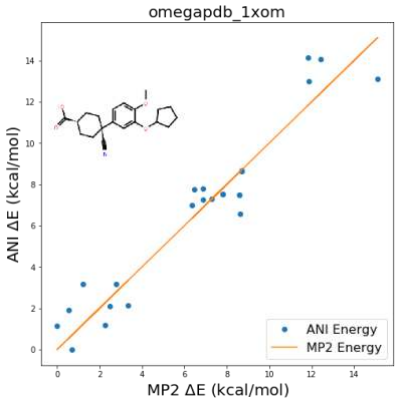

omegapdb_2f7p
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

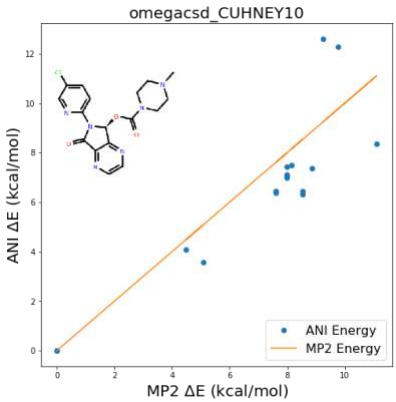

omegacsd_SPIRIL
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

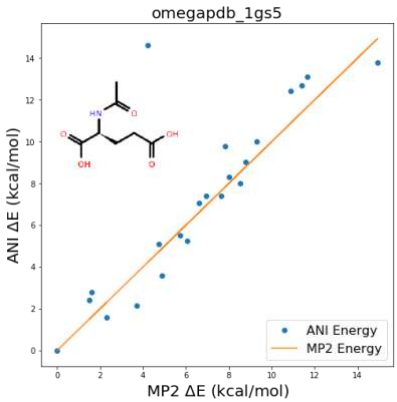

omegacsd_FIBREN
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

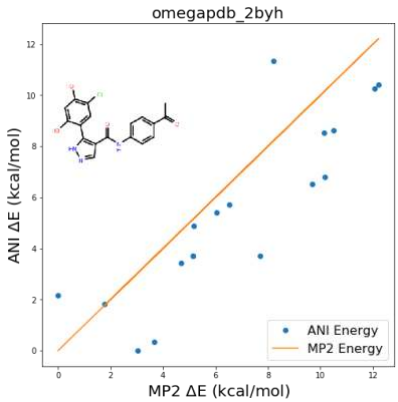

omegapdb_1qxw
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

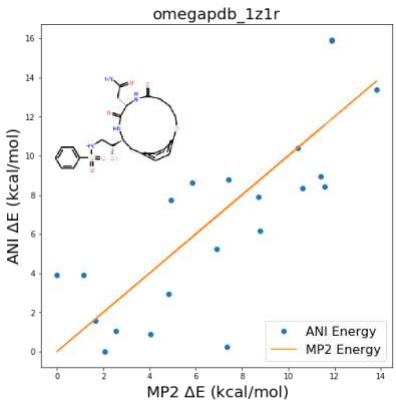

omegacsd_CEJTIU
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

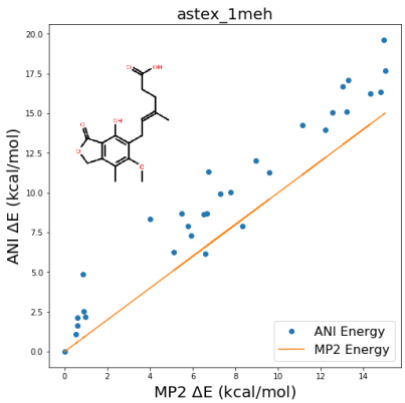

omegapdb_1xom
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy


omegacsd_CUHNEY10
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

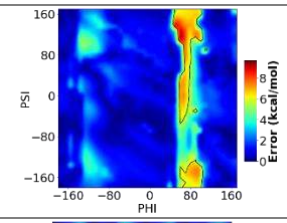
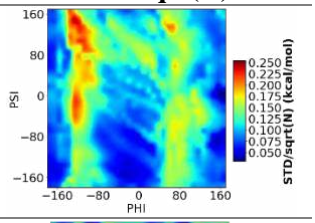
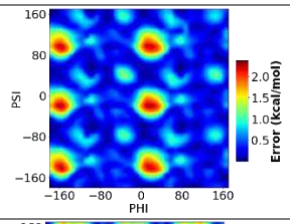
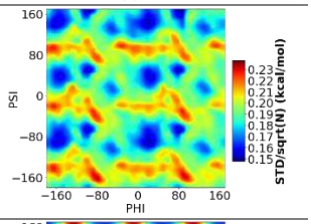
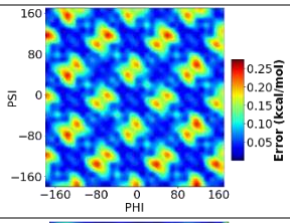
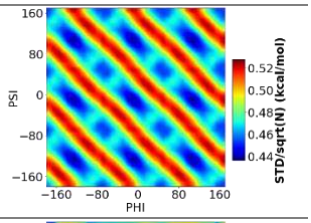
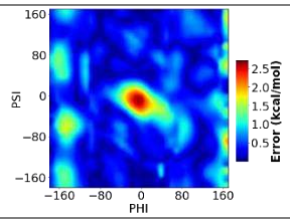
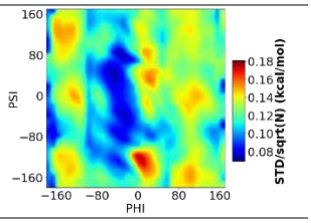

omegapdb_1gs5
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy


omegapdb_2byh
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy


omegapdb_1z1r
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy


astex_1meh
ANI ΔE (kcal/mol)
MP2 ΔE (kcal/mol)
ANI Energy
MP2 Energy

FS3: ANI-2x error and ensemble standard deviation divided by the square root of number of atoms in the system for each of the molecules from the 2D torsion scan benchmark.

Name	Error	STD/sqrt(N)
Cysteine-Dipeptide		
DDT		
Hexafluoroacetone		
Bendamustine		

Methods:

The bond lengths, angles, and dihedrals analyzed during the conformer search and score test were attained by iterating over all bonds using the RDKit software package. RMSD calculations were performed using open babel's python implementation pybel and exclude hydrogens.

ANI-2x contains an ensemble of 8 models. Each model contains a separate atomic network for each species. Table S4 shows details for these networks.

TS7:

		Hydrogen		
	Layer 1	Layer 2	Layer 3	Layer 4
Nodes	256	192	160	1
Activation	CELU	CELU	CELU	Linear
Regularization	L2 (5.0E-3)	L2(1.0E-6)	L2(1.0E-6)	None
		Carbon		
Nodes	224	192	160	1
Activation	CELU	CELU	CELU	Linear
Regularization	L2(5.0E-3)	L2(1.0E-6)	L2(1.0E-6)	None
		Nitrogen		
Nodes	192	160	128	1
Activation	CELU	CELU	CELU	Linear
Regularization	L2(5.0E-3)	L2(1.0E-6)	L2(1.0E-6)	None
		Oxygen		
Nodes	192	160	128	1
Activation	CELU	CELU	CELU	Linear
Regularization	L2(5.0E-3)	L2(1.0E-6)	L2(1.0E-6)	None
		Sulfur/Fluorine/Chlorine		
Nodes	160	128	96	1
Activation	CELU	CELU	CELU	Linear
Regularization	L2(5.0E-3)	L2(1.0E-6)	L2(1.0E-6)	None

TS8: Comprehensive Machine-learning Potential version 2 (COMP6v2) benchmark.

BENCHMARK	# OF MOLECULES	# OF CONFORMATIONS	AVERAGE ATOMS/MOLECULE (STDV.)
S66X8	66	528	20(7)
TRIPEPTIDES	345	3536	51(7)
MD	14	1791	75(73)
BENCHMARK			
GDB-12TO13	2000	24000	26(3)
GDB-10TO11	1746	41670	21(3)
GDB-07TO09	2625	63000	16(3)
DRUGBANK	1451	23203	44(18)

The ANI model atomic environment vector (AEV) is computed using the in-house NeuroChem software suite. These AEVs are computed identically to those published in the ANI-1 work.7 In this work the atomic elements C, H, N, and O are described by the AEVs (using the parameters below) yielding a total of 384 AEV elements per atom. The AEV parameters used to train each model are supplied below.

Radial Parameters:

- Radial Cutoff= 5.1 $\AA$
- η_{Radial}= 19.70000
- $R_{SRadial}$=[8.0000000e+01, 1.0687500e+00, 1.3375000e+00, 1.6062500e+00, 1.8750000e+00, 2.1437500e+00, 2.4125000e+00, 2.6812500e+00, 2.9500000e+00, 3.2187500e+00, 3.4875000e+00, 3.7562500e+00, 4.0250000e+00, 4.2937500e+00, 4.5625000e+00, 4.8312500e+00] $\AA$

Angular Parameters:

- Angular Cutoff= 3.5 $\AA$
- $\eta_{Angular}$= 12.50000
- $R_{sRadial}$=[8.0000000e-01, 1.1375000e+00, 1.4750000e+00, 1.8125000e+00, 2.1500000e+00, 2.4875000e+00, 2.8250000e+00, 3.1625000e+00]
- ζ= 14.10000
- θ_s=[3.9269908e-01, 1.1780972e+00, 1.9634954e+00, 2.7488936e+00]