

# Evaluation of Thermochemical Machine Learning for Potential Energy Curves and Geometry Optimization

Dakota L. Folmsbee,<sup>†</sup> David R. Koes,<sup>‡</sup> and Geoffrey R. Hutchison<sup>\*,†,¶</sup>

<sup>†</sup>*Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, PA  
15260*

<sup>‡</sup>*Department of Department of Computational & Systems Biology, School of Medicine,  
University of Pittsburgh, 3420 Forbes Avenue, Pittsburgh, PA 15260*

<sup>¶</sup>*Department of Chemical and Petroleum Engineering, University of Pittsburgh, 3700  
O'Hara Street, Pittsburgh, PA 15261*

E-mail: geoffh@pitt.edu

## Abstract

While many machine learning methods, particularly deep neural networks, have been trained for density functional and quantum chemical energies and properties, the vast majority of these methods focus on single-point energies. In principle, such ML methods, once trained, offer thermochemical accuracy on par with density functional and wave function methods but at speeds comparable to traditional force fields or approximate semiempirical methods. So far, most efforts have focused on optimized equilibrium single-point energies and properties. In this work, we evaluate the accuracy of several leading ML methods across a range of bond potential energy curves and torsional potentials. Methods were trained on the existing ANI-1 training set, calculated using the  $\omega$ B97X / 6-31G(d) single points at non-equilibrium geometries. We find that across a range of small molecules, several methods offer both qualitative accuracy (e.g., correct minima, both repulsive and attractive bond regions, anharmonic shape, and single minima) and quantitative accuracy in terms of the mean absolute percent error near the minima. At the moment, ANI-2x, FCHL, and a new libmolgrid-based convolutional neural net show good performance.

## Introduction

Machine learning (ML) methods have been proposed as surrogates for time-consuming quantum mechanical calculations, such as density functional and first-principles methods, for their rapid prediction potential once trained.<sup>1-11</sup> For ML to be a successful surrogate, the methods need to be able to perform property predictions adequately for optimized geometries, capture not just the well of the potential energy curve but also the anharmonicity that force field methods fail to capture, and appropriately handle multiple conformations of the same molecule.

Numerous studies have shown the proficiency of ML methods to predict thermochemical parameters at already optimized geometries utilizing various types of representations and

neural network structures.<sup>2,12</sup> Early representations, such as Coulomb Matrix<sup>13</sup> and bag-of-features,<sup>14,15</sup> demonstrated success in property predictions with further iterations of representations such as FCHL<sup>16,17</sup> continuing to improve the property prediction at optimized geometries. These ML methods are typically trained on the QM7<sup>13,18</sup> or QM9<sup>19,20</sup> data sets consisting of optimized molecules with up to 7 or 9 heavy atoms respectively and help to demonstrate ML’s potential as a surrogate.

Additional deep neural network (DNN) methods, like ANI<sup>3-5,21</sup> and BAND NN,<sup>22</sup> used training data beyond optimized single points to better evaluate the potential surface for dynamics and geometry optimizations. These methods utilize the ANI-1 data set,<sup>23</sup> or ANI-2 data set in the case of ANI-2x, for training as they contain both equilibrium and non-equilibrium structures of up to eight heavy atoms containing H, C, N, and O with the non-equilibrium structures being generated from normal-mode sampling. The training set for ANI-2x adds the additional elements of F, Cl, and S while providing additional torsion sampling data.<sup>5</sup> The BAND NN model uses a subset of the ANI-1 data set with only non-equilibrium geometries with energies within 30 kcal/mol of the equilibrium energy. Although these methods have been shown to perform adequately in their respective papers, the range for bond stretch applications has been limited to the harmonic portion of the potential energy curve, rarely examining the potential energy curves further from equilibrium.

Recent work has expanded the knowledge on ML performance for predicting and ranking thermally accessible conformations.<sup>24</sup> Though ML was not tasked with large bond stretches as in this work, the ability of ML methods to rank conformational energy was only comparable to that of semiempirical methods. While this is not equivalent to the accuracy of density functional (DFT) or *ab initio* electronic structure methods, for ML methods to be an accurate surrogate for quantum chemical methods, continued advancements in ML models and training sets are needed to provide further performance improvements.

For ML to become a viable replacement for current methods, ML needs to achieve optimized

geometries and predict properties without relying on force field (FF) methods. Most FFs have been refined for small molecules and biomolecules and can struggle with non-covalent and steric interactions for applications such as conjugated polymers. While these issues can be lessened with specific parameterization,<sup>25,26</sup> geometries of FFs generally can be less than ideal.<sup>27</sup> ML trained on higher levels of theory ideally captures these non-covalent interactions and provides better initial optimized geometries.

With the rapid adoption of ML, there has been a growing desire to use ML in molecular dynamics (MD) applications to provide more accurate simulations than FFs at a much lower cost than time-consuming quantum mechanical calculations.<sup>24</sup> For ML to be reliable, it needs to properly predict geometric changes that occur in MD simulations from non-equilibrium bond stretching to torsional barriers. This work seeks to examine how well the current state of ML performs at these tasks, as well as to display the methods’ understanding of chemical physics to help decide key needs for ML to improve as a surrogate for computationally expensive quantum calculations.

## Methods

### Molecules

A mixture of small and large molecules was chosen to evaluate ML performance on potential energy surfaces for a total of 17 bond stretches and 5 dihedral scans. The molecules examined were benzene (C-C and C-H stretching), methanol, methane, CO, H<sub>2</sub>, ethylene, water, acetylene, hydrogen cyanide, N<sub>2</sub>, ammonia, biphenyl, aspartame, sucrose, dialanine, and diglycine. Bond stretches were evaluated every 0.1Å while dihedrals were evaluated every 20° with the exception of biphenyl which was every 15°.

## Computational Methods

The reference method,  $\omega$ B97X,<sup>28</sup> was performed using Orca 4.0.1<sup>29</sup> while the force field calculations, MMFF94<sup>30-34</sup> and GAFF,<sup>35</sup> were performed using Open Babel version 3.0.<sup>36</sup>

Machine learning methods and representations included the pre-trained models ANI-1x,<sup>3,4</sup> ANI-2x,<sup>5</sup> BAND-NN,<sup>22</sup> as well as FCHL,<sup>37</sup> Bag of Bonds (BOB),<sup>38</sup> and Extended Connectivity Fingerprints (ECFP).<sup>39,40</sup> Scikit-learn<sup>41</sup> was used for kernel ridge regression (KRR) and bayesian ridge regression (BRR) for BOB and random forest regression (RFR) with BOB and ECFP representations while FCHL used the custom KRR in QML.

We also trained a deep convolutional neural network (Colorful CNN), an approach that has been successfully used in protein-ligand binding affinity prediction.<sup>42,43</sup> The input molecule is represented as a voxelized grid of atomic densities as generated by the libmolgrid library.<sup>44</sup> Our network has six modules separated by pooling operations each with seven convolutional layers and was trained on the ANI-1x data set.<sup>45</sup> The trained Colorful CNN model can be found at <https://github.com/hutchisonlab/ml-benchmark>.

Due to method scaling efficiency for memory usage, a subset of the ANI-1 data set was taken for training representations using BOB/KRR and BOB/BRR. For consistency, ECFP/RFR and BOB/RFR were additionally trained on this subset. The subset consists of 5 non-equilibrium geometries for every molecule with up to 7 heavy atoms, as well as 5 non-equilibrium geometries for half of the molecules with 8 heavy atoms, to create a training set consisting of 33,496 molecules and 167,480 non-equilibrium geometries. All molecules from the test set were removed from the training set. This training set was additionally used for BOB/RFR and ECFP/RFR. An additional subset of the first 5000 non-equilibrium geometries was used for FCHL/KRR. Increasing the training set for FCHL/KRR had a negative impact on prediction performance so our results are with the model trained on 1000 different molecules for a total of 5000 non-equilibrium geometries.

# Results and discussion

To illustrate the qualitative performance of potential energy surface predictions, we analyzed both small and larger molecules outside of the ANI-1 data set used for training for each ML method. We wish to focus on how the methods perform not only around the bond length at the energy minima,  $r_0$ , but also in the attractive and repulsive regimes to gain a better understanding of how ML methods would behave if given less ideal starting geometries for a task such as geometry optimization.

**Table 1: Overview of machine learning performance sorted by median mean absolute percent error (MAPE).**

Methods	Median MAPE <sup>1</sup>	$r_0$ <sup>2</sup>	Repulsive Wall <sup>3</sup>	Attractive Forces <sup>4</sup>	Minima after 2Å <sup>5</sup>
$\omega$ B97X 6-31G(d)	0	17	17	17	0
ANI-2x	0.002	17	13	17	12
BOB/BRR	0.227	0	5	5	9
FCHL/KRR	0.255	10	16	15	13
Colorful CNN	0.2555	16	17	17	13
ANI-1x	0.265	16	11	17	5
BOB/KRR	0.313	1	9	11	13
BOB/RFR	43.881	2	3	0	8
BAND-NN	99.310	11	9	5 <sup>6</sup>	5 <sup>6</sup>
MMFF94	100.050	14	17	0	0
GAFF	100.133	13	17	0	0
ECFP/RFR	193.370	0	0	0	0

<sup>1</sup> Median mean absolute percent error over all 17 molecules from  $r_0 \pm 0.25\text{\AA}$ .

<sup>2</sup> The number of molecules in which the lowest predicted energy point matches DFT.

<sup>3</sup> The number of times the method predicted a repulsive wall as the bond was compressed.

<sup>4</sup> The number of times the method predicted anharmonic attractive forces after  $r_0$ .

<sup>5</sup> The number of molecules predicted to have a local or global minima after 2Å.

<sup>6</sup> BAND-NN regularly would not predict energies for geometries with a bond stretch of 2Å or greater.

Each ML method was evaluated on the criteria demonstrated in Table 1 for bond stretches. The median mean absolute percent error (MAPE) was calculated from the energy values ranging from  $r_0 \pm 0.25\text{\AA}$  for the molecules to determine how accurate and precise the ML predicted energies are. Since the ANI-1 training set samples harmonic displacements around the  $r_0$  (e.g., Figure S1) this range corresponds mostly to interpolation. Comparisons for repulsive short-range and attractive long-range interactions – extrapolations outside the training range are compiled in Table S1. The  $r_0$  evaluation criteria considered whether the

method correctly predicted the DFT equilibrium bond length to be the lowest energy bond length. Additional evaluation criteria included the qualitative prediction of a repulsive wall, anharmonic long range interactions, and if there were incidences of additional minima past 2Å.

While methods like BOB/BRR and BOB/KRR had the second and fifth-lowest median MAPE, their ability to predict the geometry with the lowest energy, a repulsive wall, and attractive forces was quite poor compared to the other top methods based on MAPE. Other methods utilizing RFR also performed poorly, often predicting stepwise energy surfaces seen in Figure 1, thus being incapable of consistently predicting  $r_0$ , attractive, or repulsive forces. This is seen in Figure 1b when the bond breaking causes the only change in the ECFP representation and leads to the higher energy. Other ML methods such as ANI-1x, ANI-2x, FCHL, and Colorful CNN were able to accurately predict energies while also predicting the repulsive and attractive forces of the molecule. In short, while random forest methods may have accuracy at single-point properties, they prove inherently inaccurate for potential energy and should be avoided.

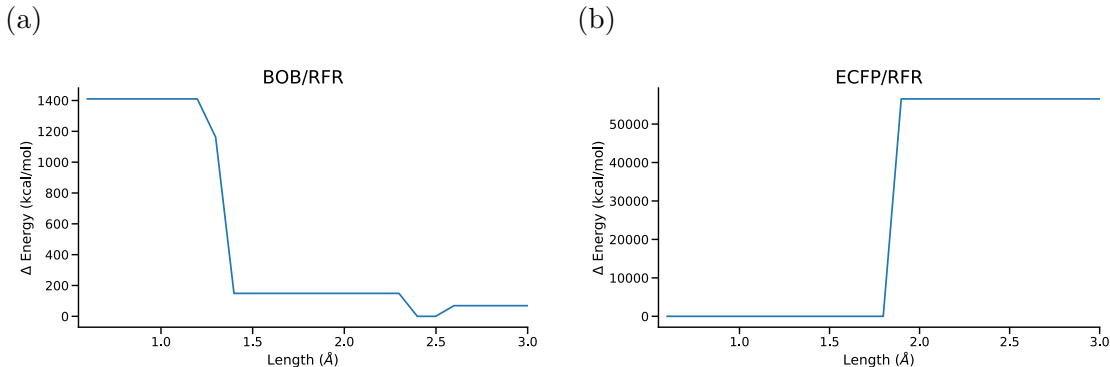


Figure 1: N<sub>2</sub> potential energy curves for ML methods utilizing random forest regression for predictions using (a) BOB and (b) ECFP for the ML descriptors.

A possible advantage for the ANI-1x and ANI-2x models is that some molecules in our test evaluation are found in the ANI-1x training set. In the training of the other methods, molecules in our test set were purposefully left out of the training set but may be present in

the ANI-1x and ANI-2x model. For that reason, we will focus the remainder of our discussion on molecules outside of the ANI-1 training set, examining the best overall performers, ANI-1x, ANI-2x, FCHL, and Colorful CNN from Table 1. The performance of all methods is included in the supplemental information.

Figure 2a displays the performance of ANI-1x, ANI-2x, Colorful CNN, and FCHL on the N-N bond stretch of  $\text{N}_2$ . While each of these ML methods predicts the correct  $r_0$ , there are issues in the prediction of the potential energy curve. ANI-1x, ANI-2x, and Colorful CNN fail to accurately depict the repulsive region with ANI-2x lowering in energy as the bond was compressed to  $0.6\text{\AA}$ . FCHL depicts the repulsive wall but inaccurately predicts the energy as the bond is compressed. All four methods accurately determined the attractive forces to about  $2\text{\AA}$  with ANI-2x matching  $\omega\text{B97X}$  to  $2.25\text{\AA}$ .

The H-H stretch of  $\text{H}_2$  in Figure 2b indicates one possible issue for ML. All four methods performed poorly with ANI-2x being the only method to obtain the correct  $r_0$ . This performance is likely due to the absence of H-H bonding data within the training set.  $\text{H}_2$ , while a unique bond, demonstrates the need to be careful when applying ML to molecules or chemistry completely outside the scope of the training set.

Figure 2c and 2d demonstrate the prediction capability of these ML methods on bond stretches for molecules larger than the training set. FCHL was only able to accurately capture the shape of the potential energy curve for dialanine, failing to capture the well of the potential energy curve for aspartame, perhaps from the difficulties training the entire ANI-1 set. ANI-1x, ANI-2x, and Colorful CNN retain both repulsive and attractive information while having accurate energies to that of  $\omega\text{B97X}$  for both aspartame and dialanine. These methods do continue to exhibit difficulty in accurately predicting bond compression under  $1\text{\AA}$  as well as bond stretching after  $2\text{\AA}$ .



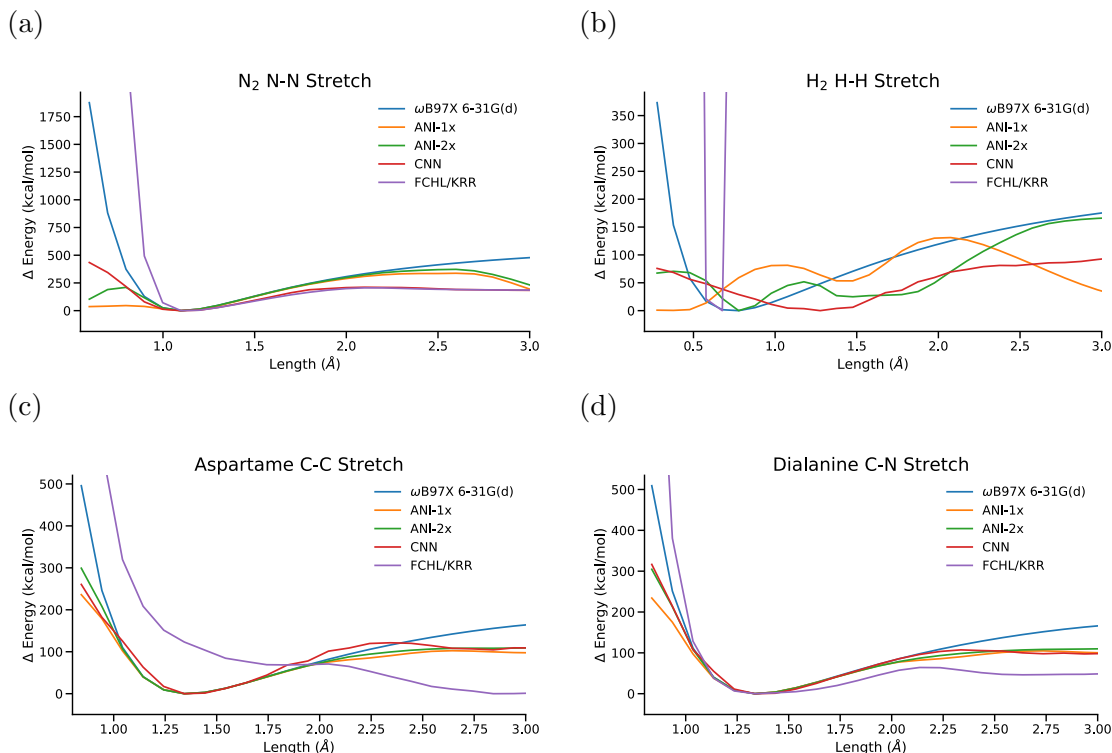


Figure 2: Bond stretch potential energy curves for (a) N<sub>2</sub>, (b) H<sub>2</sub>, (c) aspartame, (d) dialanine using total SCF energies in kcal/mol.

For bond stretches, ANI-1x, ANI-2x, Colorful CNN, and FCHL models show promise with initial training indicating these methods can accurately predict the bottom of the potential energy well. While force fields such as MMFF94 or GAFF can be used to obtain optimized geometries near this regime, ultimately ML methods should exhibit accuracy not only at single-point energy evaluation tasks, but at qualitatively and quantitatively accurate potential energy curves. Further training on long-range attractive forces might enable ML models to evaluate non-covalent interactions.

As an example, further evaluations were carried out on energy predictions from frozen-rotor dihedral angle scans performed with  $\omega$ B97X 6-31G(d) for biphenyl and sucrose. Table 2 compiles the predicted lowest energy angle for these molecules as well as the barrier energies from  $-45^\circ$  to  $0^\circ$  for biphenyl and  $0^\circ$  to  $-60^\circ$  for sucrose.

ANI-1x and ANI-2x properly predict the lowest energy angle for biphenyl while Colorful CNN

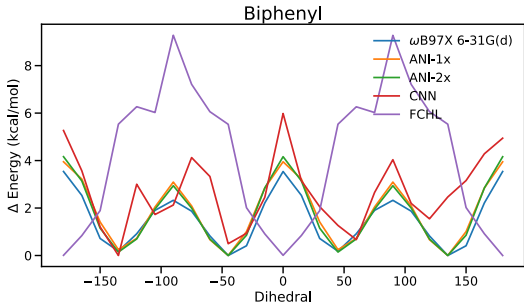
predicts  $-45^\circ$  to be a local, but not global, minima. FCHL improperly predicts rotation energies as seen in Figure 3a, predicting  $0^\circ$ ,  $180^\circ$ , and  $-180^\circ$  to be the lowest energy dihedrals. All of the methods over-predicted the height of the energy barrier for biphenyl.

For sucrose, all four methods correctly predicted the lowest energy angle. ANI-1x best captures the energy of the dihedral angles, seen in Figure 3b, with ANI-2x and Colorful CNN under-predicting the energy for most angles. Unlike with biphenyl, FCHL captures the shape of the torsion scan for sucrose but vastly over predicts the energies at each angle.

**Table 2: The ML prediction of  $\theta_0$  and the barrier energy between the lowest and highest energy dihedrals for biphenyl and sucrose compared to the reference  $\omega$ B97X 6-31G(d) method.**

Methods	Biphenyl		Sucrose	
	$\theta_0$ ( $^\circ$ )	Barrier Energy (kcal/mol)	$\theta_0$ ( $^\circ$ )	Barrier Energy (kcal/mol)
$\omega$ B97X 6-31G(d)	-45	3.54	0	$2.45 \times 10^3$
ANI-1x	-45	3.95	0	$2.50 \times 10^3$
ANI-2x	-45	4.16	0	$1.93 \times 10^3$
Colorful CNN	-135	5.49	0	$9.46 \times 10^2$
FCHL/KRR	180	5.52	0	$9.73 \times 10^4$

(a)



(b)

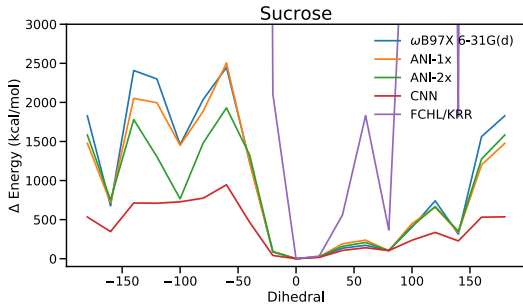


Figure 3: Dihedral energy predictions for (a) biphenyl and (b) sucrose in kcal/mol.

Dihedral scans demonstrate how small conformational changes in the molecule can affect the potential energy surface. The 2D torsion scans in Figures 4 and 5 compare ML performance to that of  $\omega$ B97X and FFs, MMFF94 and GAFF. ANI-1x, ANI-2x, and Colorful CNN retain the resolution of some of the higher energy  $\phi$  and  $\psi$  between  $-100^\circ$  to  $100^\circ$  while FCHL

predicts these to be lower energy conformations similar to both FF methods. In lower energy conformations both BAND and BOB/KRR methods over-estimate these energy differences.

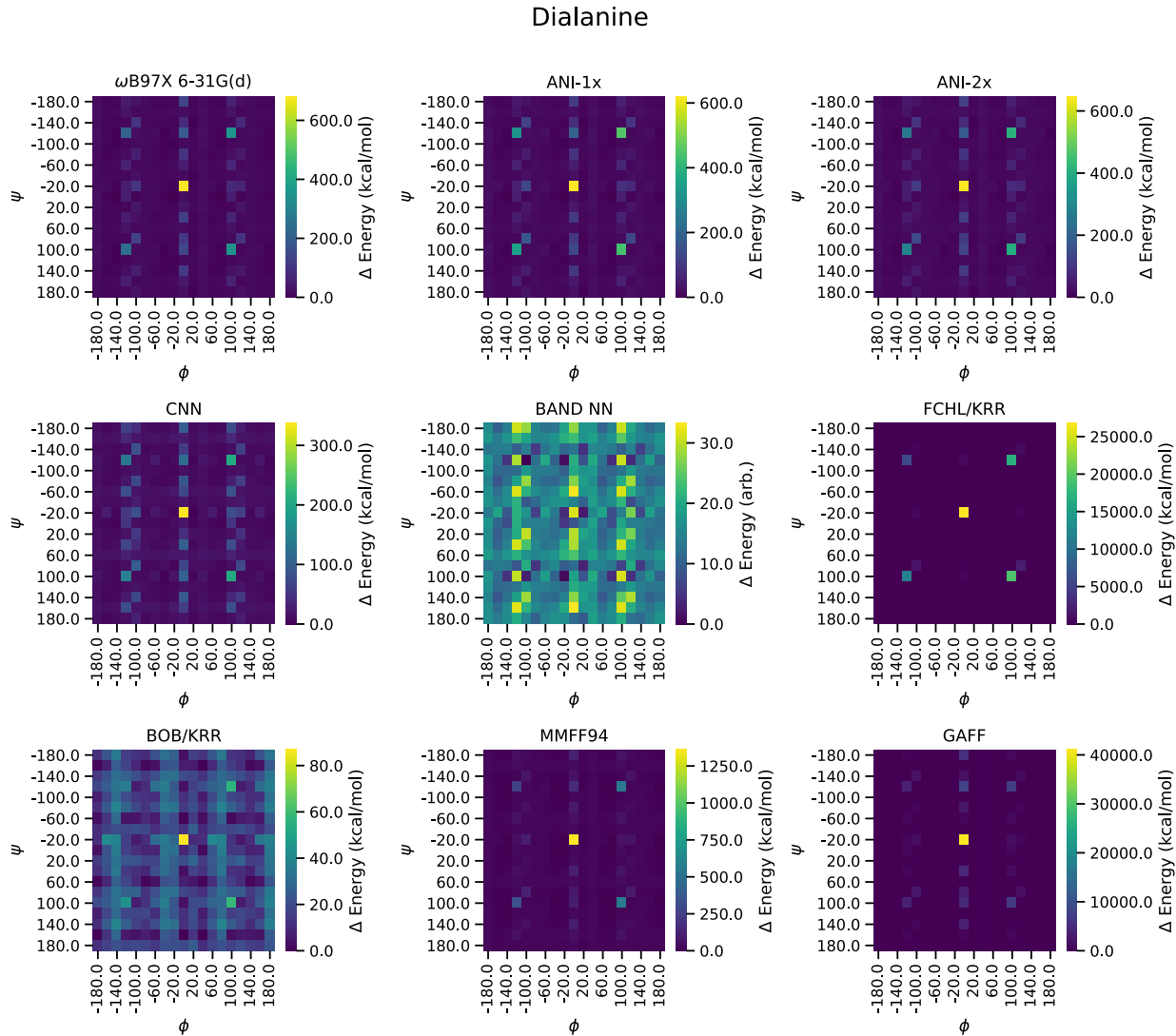


Figure 4: 2D torsion scans of dialanine in kcal/mol unless otherwise stated. Methods were tested at the geometries obtained with  $\omega$ B97X 6-31G(d) from the torsion scan. Note that color schemes differ, due to large differences in energy scales.

### Diglycine

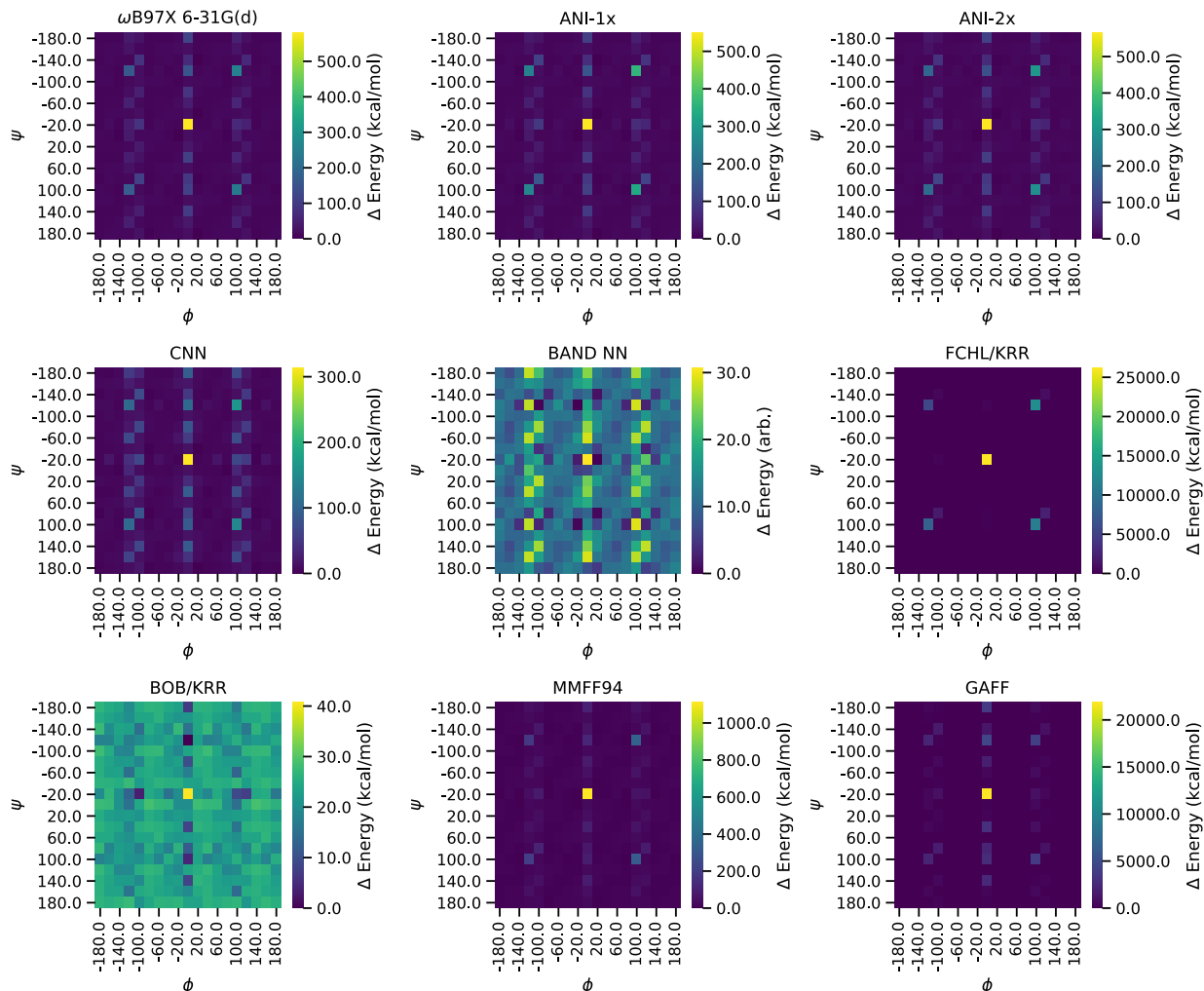


Figure 5: 2D torsion scans of diglycine in kcal/mol unless otherwise stated. Methods were tested at the geometries obtained with  $\omega$ B97X 6-31G(d) from the torsion scan. Note that color schemes differ, due to large differences in energy scales.

The additional torsion training in ANI-2x provided a beneficial reduction in the MAE for both dialanine and diglycine, seen in Table 3, by roughly 35% from ANI-1x. Additional torsion sampling for methods Colorful CNN and FCHL should also provide a decrease in MAE for predicting dihedral angle energies. This could improve accuracy for the Colorful CNN method that is already qualitatively adequate.

As an example, the ANI-2x training includes additional torsion sampling, and the method

**Table 3: Mean absolute error (MAE) in kcal/mol of 2D torsion scans for the top performing methods.**

Methods	Dialanine MAE $\Delta$ Energy (kcal/mol)	Diglycine MAE $\Delta$ Energy (kcal/mol)
ANI-2x	1.89	1.71
ANI-1x	3.01	2.52
Colorful CNN	7.10	6.07
FCHL/KRR	252.17	200.86

shows improved accuracy over ANI-1x. Providing additional torsion sampling training sets should improve ML method accuracy across multiple methods.

A prevailing pitfall of ML methods stems from the training set. At the end of the day, the machine learning method is only as good as the training set. As seen with  $H_2$ , models struggle with chemical motifs outside of the training set. Current ML training sets largely consist of a subset of the molecules generated in the GDB-17<sup>20</sup> set, typically containing at least H, C, O, and N. While these training sets are a noble starting point for covering small organic molecules, they lack a diversity of atom species needed for applications such as protein binding and DNA sequencing. Additional data sets such as PubChemQC<sup>46</sup> could help to further expand the snapshot of chemical space ML methods are trained on.

## Conclusions

Much work has focused on the use of machine learning methods as surrogates for computationally-intensive density functional and quantum chemical methods. Often such efforts train and test on single-point energies of optimized structures. An important step is to evaluate ML methods across potential energy curves and surfaces for tasks such as geometry optimization.

ML methods such as ANI-2x, Colorful CNN, and FCHL perform decently near the well of the potential energy curve while struggling to properly predict repulsive regions and particularly long-range attractive forces. While this poor performance outside the domain of the training set is expected, these methods show promise with further improvements through the addition of stretched bonds in training data helping to improve model performance in

this area. Increased torsion sampling for training ANI-2x improved the model’s performance over ANI-1x and should provide improvements for models like Colorful CNN and FCHL.

In general, there is still the issue of applying ML to the prediction of molecules too far outside the scope of the training set. The inclusion of additional elements and an increase in diversity of molecules in the training set from diverse data sets such as PubchemQC should alleviate some of these challenges.

## Acknowledgement

We acknowledge the National Science Foundation (CHE-1800435) for support and the University of Pittsburgh Center for Research Computing through the computational resources provided.

## Supporting Information Available

Figures of all bond stretch potential energy curves, dihedral potential energy scans for all molecules and methods considered. All raw data, Python notebooks, and the trained Colorful CNN model can be found at <https://github.com/hutchisonlab/ml-benchmark>.

## References

- (1) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters* **2007**, *98*.
- (2) 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. *Journal of Chemical Theory and Computation* **2017**, *13*, 5255–5264, PMID: 28926232.

- (3) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **2017**, *8*, 3192–3203.
- (4) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *The Journal of Chemical Physics* **2018**, *148*, 241733.
- (5) Devereux, C.; Smith, J. S.; Davis, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202, PMID: 32543858.
- (6) von Lilienfeld, O. A.; Burke, K. Retrospective on a decade of machine learning for chemical discovery. *Nature Communications* **2020**, *11*.
- (7) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2336–2347.
- (8) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics* **2020**, *153*, 124111.
- (9) Sinitskiy, A. V.; Pande, V. S. Deep Neural Network Computes Electron Densities and Energies of a Large Set of Organic Molecules Faster than Density Functional Theory (DFT). 2018.
- (10) Sinitskiy, A. V.; Pande, V. S. Physical machine learning outperforms ”human learning” in Quantum Chemistry. 2020.
- (11) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet

- A deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **2018**, *148*, 241722.
- (12) Wu, Z.; Ramsundar, B.; Feinberg, E.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513–530.
- (13) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- (14) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters* **2015**, *6*, 2326–2331.
- (15) Huang, B.; von Lilienfeld, O. A. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *The Journal of Chemical Physics* **2016**, *145*, 161102.
- (16) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *The Journal of Chemical Physics* **2018**, *148*, 241717.
- (17) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; von Lilienfeld, O. A. FCHL revisited: Faster and more accurate quantum machine learning. *The Journal of Chemical Physics* **2020**, *152*, 044107.
- (18) Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732.



- (19) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data* **2014**, *1*.
- (20) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875, PMID: 23088335.
- (21) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. **2019**,
- (22) Laghuvarapu, S.; Pathak, Y.; Priyakumar, U. D. BAND NN: A Deep Learning Framework for Energy Prediction and Geometry Optimization of Organic Small Molecules. *Journal of Computational Chemistry* **2020**, *41*, 790–799.
- (23) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific Data* **2017**, *4*, 170193.
- (24) Folmsbee, D.; Hutchison, G. Assessing conformer energies using electronic structure and machine learning methods. *International Journal of Quantum Chemistry* *n/a*, e26381.
- (25) DuBay, K. H.; Hall, M. L.; Hughes, T. F.; Wu, C.; Reichman, D. R.; Friesner, R. A. Accurate Force Field Development for Modeling Conjugated Polymers. *Journal of Chemical Theory and Computation* **2012**, *8*, 4556–4569.
- (26) Wildman, J.; Repiščák, P.; Paterson, M. J.; Galbraith, I. General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution. *Journal of Chemical Theory and Computation* **2016**, *12*, 3813–3824.
- (27) Kanai, I. Y.; Keith, J. A.; Hutchison, G. R. A sobering assessment of small-molecule force field methods for low energy conformer predictions. *International Journal of Quantum Chemistry* **2017**, *118*, e25512.

- (28) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics* **2008**, *128*, 084106.
- (29) Neese, F. The ORCA program system. *WIREs Computational Molecular Science* **2012**, *2*, 73–78.
- (30) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 490–519.
- (31) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry* **1996**, *17*, 520–552.
- (32) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 553–586.
- (33) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 587–615.
- (34) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *Journal of Computational Chemistry* **1996**, *17*, 616–641.
- (35) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- (36) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 33.
- (37) Christensen, A.; Faber, F.; Huang, B.; Bratholm, L.; Tkatchenko, A.; Müller, K.; von

- Lilienfeld, O. QML: A Python Toolkit for Quantum Machine Learning. 2017; <https://github.com/qmlcode/qml>.
- (38) Folmsbee, D.; Upadhyay, S.; Dumi, A.; Hiener, D.; Mulvey, D. chemreps/chemreps: Molecular Machine Learning Representations. 2019; <https://doi.org/10.5281/zenodo.3333856>.
- (39) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, *50*, 742–754, PMID: 20426451.
- (40) Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. *Journal of Cheminformatics* **2015**, *7*, 26.
- (41) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- (42) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein–ligand scoring with convolutional neural networks. *Journal of chemical information and modeling* **2017**, *57*, 942–957.
- (43) Jiménez, J.; Skalic, M.; Martinez-Rosell, G.; De Fabritiis, G. K deep: Protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of chemical information and modeling* **2018**, *58*, 287–296.
- (44) Sunseri, J.; Koes, D. R. libmolgrid: Graphics Processing Unit Accelerated Molecular Gridding for Deep Learning Applications. *Journal of Chemical Information and Modeling* **2020**, *60*, 1079–1084.
- (45) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.;

- Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data* **2020**, *7*, 1–10.
- (46) Nakata, M.; Shimazaki, T. PubChemQC Project: A Large-Scale First-Principles Electronic Structure Database for Data-Driven Chemistry. *Journal of Chemical Information and Modeling* **2017**, *57*, 1300–1308, PMID: 28481528.

# Graphical TOC Entry

