

OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy

Cite as: J. Chem. Phys. **155**, 204103 (2021); <https://doi.org/10.1063/5.0061990>

Submitted: 01 July 2021 • Accepted: 26 October 2021 • Accepted Manuscript Online: 27 October 2021 • Published Online: 23 November 2021

 Anders S. Christensen,  Sai Krishna Sirumalla,  Zhuoran Qiao, et al.



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features](#)

The Journal of Chemical Physics **153**, 124111 (2020); <https://doi.org/10.1063/5.0021955>

[Perspective on integrating machine learning into computational chemistry and materials science](#)

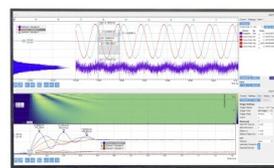
The Journal of Chemical Physics **154**, 230903 (2021); <https://doi.org/10.1063/5.0047760>

[SchNet - A deep learning architecture for molecules and materials](#)

The Journal of Chemical Physics **148**, 241722 (2018); <https://doi.org/10.1063/1.5019779>

Challenge us.

What are your needs for periodic signal detection?



Zurich
Instruments

OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy

Cite as: J. Chem. Phys. 155, 204103 (2021); doi: 10.1063/5.0061990

Submitted: 1 July 2021 • Accepted: 26 October 2021 •

Published Online: 23 November 2021



View Online



Export Citation



CrossMark

Anders S. Christensen,¹ Sai Krishna Sirumalla,¹ Zhuoran Qiao,² Michael B. O'Connor,¹
Daniel G. A. Smith,¹ Feizhi Ding,¹ Peter J. Bygrave,¹ Animashree Anandkumar,^{3,4}
Matthew Welborn,¹ Frederick R. Manby,¹ and Thomas F. Miller III^{1,2,a)}

AFFILIATIONS

¹Entos, Inc., Los Angeles, California 90027, USA

²Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

³Division of Engineering and Applied Sciences, California Institute of Technology, Pasadena, California 91125, USA

⁴NVIDIA, Santa Clara, California 95051, USA

^{a)} Author to whom correspondence should be addressed: tom@entos.ai

ABSTRACT

We present OrbNet Denali, a machine learning model for an electronic structure that is designed as a drop-in replacement for ground-state density functional theory (DFT) energy calculations. The model is a message-passing graph neural network that uses symmetry-adapted atomic orbital features from a low-cost quantum calculation to predict the energy of a molecule. OrbNet Denali is trained on a vast dataset of 2.3×10^6 DFT calculations on molecules and geometries. This dataset covers the most common elements in biochemistry and organic chemistry (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I) and charged molecules. OrbNet Denali is demonstrated on several well-established benchmark datasets, and we find that it provides accuracy that is on par with modern DFT methods while offering a speedup of up to three orders of magnitude. For the GMTKN55 benchmark set, OrbNet Denali achieves WTMAD-1 and WTMAD-2 scores of 7.19 and 9.84, on par with modern DFT functionals. For several GMTKN55 subsets, which contain chemical problems that are not present in the training set, OrbNet Denali produces a mean absolute error comparable to those of DFT methods. For the Hutchison conformer benchmark set, OrbNet Denali has a median correlation coefficient of $R^2 = 0.90$ compared to the reference DLPNO-CCSD(T) calculation and $R^2 = 0.97$ compared to the method used to generate the training data (ω B97X-D3/def2-TZVP), exceeding the performance of any other method with a similar cost. Similarly, the model reaches chemical accuracy for non-covalent interactions in the S66x10 dataset. For torsional profiles, OrbNet Denali reproduces the torsion profiles of ω B97X-D3/def2-TZVP with an average mean absolute error of 0.12 kcal/mol for the potential energy surfaces of the diverse fragments in the TorsionNet500 dataset.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0061990>

I. INTRODUCTION

Theoretical chemistry is based on the strategy of using approximate methods to make predictions for chemical problems. Perhaps most notable within the field are approximations to the time-independent Schrödinger equation, yielding the electronic energy that governs chemical potential energy surfaces.

While high-level wavefunction methods calculate energies to high accuracy, these methods are often too slow for use in many

areas of chemical research. A pragmatic middle-ground is density functional theory (DFT), which can yield results close to chemical accuracy, often on the scale of minutes to hours of computational time. In high-throughput applications, force-field (FF) and semi-empirical quantum mechanics (SEQM) methods are often used, providing results within fractions of a second of computational time, but at the cost of a loss in the reliability of the predictions.

Recent developments in machine learning models for chemistry have led to many new strategies for making energy predictions

of molecules.^{1–29} Relatively few of these, however, have sufficient accuracy and chemical coverage to provide a viable replacement for quantum chemistry methods, such as DFT, without needing to retrain the model for specialized purposes. The difficulty in constructing such machine learning models that can make reliable predictions across a large fraction of chemical space is twofold: (i) the model must capture the underlying physics and (ii) relevant and well-curated training data covering the relevant chemical problems must be available.

The current paper introduces OrbNet Denali to address these issues. OrbNet Denali is a scale-up of the previously introduced orbital-based neural network potential, OrbNet, both in terms of model size and training data.^{27,28} On the technical side, this scale-up includes a number of practical and architectural engineering improvements to the OrbNet code, which, in turn, enables training on an expanded dataset with much a wider coverage of chemical space. OrbNet is a message-passing graph neural network³⁰ that encodes a molecular system in graphs based on features from a low-cost quantum calculation. In this work, specifically, the low-cost quantum calculation is based on the GFN1-xTB method,³¹ and the features are generated from the Fock, density, orbital centroid distance, core Hamiltonian, and overlap matrices evaluated in the symmetry-adapted atomic orbital (SAAO) basis.²⁷ Graph-node features correspond to diagonal SAAO matrix elements, and edge features correspond to off-diagonal SAAO matrix elements.

OrbNet Denali is trained on a new dataset of 2.3×10^6 molecules generated from a variety of sources. The combined dataset covers a range of organic molecules and chemistries, with protonation and tautomeric states, non-covalent interactions, common salts, and counterions, spanning the most common elements in biochemistry and organic chemistry (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I). These combined developments enable reliable energy prediction with an accuracy comparable to DFT, i.e., the same method used to calculate the energies corresponding to the geometries in the training data. We demonstrate the performance and transferability of the trained model by benchmarking OrbNet Denali across the many diverse chemical problems in the well-established collection of benchmark sets, GMTKN55.³² Furthermore, we demonstrate the performance of OrbNet Denali on several other essential tasks, such as conformer scoring, non-covalent interactions, and torsional profiles.^{33–35}

II. METHODS

A. Training data

The main advance of this work is the expansion of the OrbNet Denali training set. The collection of training data is as described below.

1. ChEMBL molecule conformers

The ChEMBL27 database³⁶ was downloaded from the ChEMBL web service.^{37,38} All SMILES strings containing 50 or fewer atoms of the elements C, O, N, F, S, Cl, Br, I, P, Si, B, Na, K, Li, Ca, or Mg and no isotope specifications were kept. SMILES strings that did not resolve to a closed-shell Lewis structure were discarded. All SMILES strings corresponding to molecules in the

Hutchison conformer benchmark set³³ were removed from our training dataset. (In Sec. III B, these molecules will be used to validate the model.)

From this subset, a final surviving selection of 116 943 unique SMILES strings corresponding to neutral molecules was randomly chosen. Up to four conformers for each SMILES string were initially generated through the ENTOS BREEZE version 0.1.5 conformer generator and optimized at the GFN1-xTB level.³¹ For each of these four energy-minimized conformers, non-equilibrium geometries were generated using ENTOS BREEZE through either normal-mode sampling⁹ at 300 K or *ab initio* molecular dynamics (AIMD) sampling for 200 fs at 500 K; in both cases at the GFN1-xTB level of theory. These thermalization methods were selected randomly for each molecule with equal weight. This procedure resulted in a total of 1 771 191 equilibrium and non-equilibrium geometries.

2. Protonation states and tautomers

A subset of 26 186 SMILES strings were randomly chosen from the list of filtered ChEMBL SMILES strings described in the above text. For each of these, up to 128 unique protonation states were identified using DIMORPHITE-DL³⁹ version 1.2.4 and four of these protonation states were selected at random. The same conformer generation algorithm and non-equilibrium geometry sampling algorithms as described above were applied to the four protonation states, resulting in a total of 215 866 unique geometries. Additionally, the protonation-state sampling was applied to the SMILES strings of the molecules in the QM7b dataset, resulting in 9413 unique molecule graphs and 30 622 geometries.^{40,41}

3. Salt complexes and non-bonded interactions

From the list of filtered ChEMBL SMILES strings, a number of SMILES strings were selected and randomly paired with between one to three salt molecules from the list of common salts in the ChEMBL structure pipeline.⁴² This procedure resulted in a total of 21 735 salt complexes. For each of these complexes, four conformers were created through our conformer pipeline, and normal-mode sampling as described above was used to generate four non-equilibrium geometries for each conformer. This resulted in 271 084 unique geometries. Additionally, the structures in the JSCH-2005⁴³ and the sidechain–sidechain interaction (SSI) subset of the BioFragment Database⁴⁴ (BFDb) were added to the dataset.

4. Small molecules

To avoid biasing the datasets to represent only large drug-like molecules, a list of common chemical moieties and bonding patterns in organic molecules was created and combined to generate small molecules with up to 21 heavy atoms and relatively “exotic” compositions, resulting in around 15 000 SMILES strings. For each of these, SMILES strings were created by randomly substituting hydrogen atoms for halogens and carbon for silicon. This procedure resulted in a total of 40 565 SMILES strings, for which conformers were generated through our conformers pipeline, resulting in a total of 94 588 unique geometries.

B. QM calculations

All DFT single-point calculations for the OrbNet Denali training set were carried out in ENTOS QCORE version 0.8.17 at the

TABLE I. Comparison of models reported in this work to previously reported OrbNet models.

Model	No. trainable parameters	No. training data	Description of training data	Multitask?
Qiao <i>et al.</i> ²⁷	$\sim 1.8 \times 10^6$	$\sim 0.2 \times 10^6$	QM7b-T, QM9, GDB-13-t, DrugBank-T	No
Qiao <i>et al.</i> ²⁸	$\sim 1.8 \times 10^6$	$\sim 0.2 \times 10^6$	QM7b-T, QM9, GDB-13-t, DrugBank-T	Yes
“Denali (10%)” (this work)	$\sim 21 \times 10^6$	$\sim 0.2 \times 10^6$	10% of data described in Sec. II A	No
“Denali” (this work)	$\sim 21 \times 10^6$	$\sim 2.3 \times 10^6$	All data described in Sec. II A	No

ω B97X-D3/def2-TZVP level of theory using in-core density fitting with the neese = 4 DFT integration grid.^{45–49}

In order to gauge the performance of OrbNet Denali, a number of additional reference calculations on the various test sets presented in Sec. III were carried out using a number of low-cost models. B97-3c calculations were carried out in ENTOS_QCORE,^{47,50} while reference GFN*n*-xTB calculations were carried out in the xTB program.^{31,51,52} ANI-1ccx and ANI-2x calculations were carried out using the TORCHANI implementation,^{10,53,54} and finally, MMFF94 calculations were carried out using the Open Babel command-line tool.^{55,56}

C. Training details

PYTORCH⁵⁷ v1.7.1 and Deep Graph Library (DGL)⁵⁸ v0.6 were used to implement and train OrbNet Denali. PyTorch’s Distributed Data Parallel (DDP)⁵⁹ strategy was used to train the model on multiple graphics processing units (GPUs) using data parallelism. OrbNet Denali was trained on the OLCF Summit supercomputer using 96 NVIDIA V100-SXM2 (32G) GPUs with a batch size of 4/GPU for 300 epochs, totaling 6912 GPU-hours of training. The learning rate was linearly warmed-up for the first 100 epochs and cosine annealed⁶⁰ to zero for the remaining 200 epochs. During this process, the maximum learning rate was 3×10^{-4} and the model parameters were optimized with the Adam optimizer.⁶¹ The training dataset of size 1.8 TB was randomly split into four shards of equal sizes. Each Summit node, comprising six GPUs, was assigned to one of these four shards such that each shard was used on 1/4 of the nodes allocated for the training job. For example, training with 16 nodes results in each shard assigned to four nodes.

An additional model (“OrbNet Denali 10%”) was trained on a 10% subset of the training data, constructed by randomly sampling 10% of molecules or complexes in each of the various subsets of the full training dataset. All other training details were the same. Finally, we use a small set of randomly selected geometries from each subset described in Sec. II A to validate and track the model convergence during training. Table I provides a comparison of models presented in this work to previously reported OrbNet models.

The OrbNet Denali model and feature generation employed in this work are based on the model previously reported in Ref. 28. Compared to that work, four changes were made. First, the attention mechanism was replaced with the FAVOR + mechanism.⁶² This resulted in greatly decreased memory usage and negligible test accuracy degradation. Second, the number of message passing steps was increased from two to three. Third, batch normalization layers⁶³ were replaced with layer normalization layers⁶⁴ without affine transformations to facilitate multi-GPU training. Finally, the regression labels were modified to account for charged molecules.

The regression labels are

$$E^{\text{ML}} \approx E^{\text{DFT}} - E^{\text{GFN1}} - \Delta E_{\text{atoms}}^{\text{fit}}, \quad (1)$$

where E^{DFT} is the reference DFT (i.e., ω B97X-D3/def2-TZVP) energy and E^{GFN1} is the GFN1-xTB energy.⁶⁵ In this work, $\Delta E_{\text{atoms}}^{\text{fit}}$ is given by

$$\Delta E_{\text{atoms}}^{\text{fit}} = \sum_i \Delta E_{Z_i}^{\text{fit}} - \Delta E_{e^-q}^{\text{fit}}, \quad (2)$$

where i indices atoms within a molecule, Z_i is the atomic number of atom i , and q is the total charge of the molecule. $\Delta E_{Z_i}^{\text{fit}}$ and $\Delta E_{e^-q}^{\text{fit}}$ are parameters and are fit to $E^{\text{DFT}} - E^{\text{GFN1}}$ with ordinary least squares prior to training.

The loss function used here is similar to that of Ref. 66 and contains two terms. The first term minimizes the error on the absolute energy of the every sample in the training set, while the second term relates to the relative energies of different geometries of the same molecule,

$$\mathcal{L} = \mathcal{L}_2(\hat{E}(\eta, b_\eta), E(\eta, b_\eta)) + \beta \mathcal{L}_2(\hat{E}(\eta, b_\eta) - \hat{E}(\eta, \hat{b}_\eta), E(\eta, \hat{b}_\eta) - E(\eta, \hat{b}_\eta)). \quad (3)$$

Here, \mathcal{L}_2 denotes the L2-loss function $\mathcal{L}_2(\hat{y}, y) = \|\hat{y} - y\|_2^2$, while b_η is the geometry of a molecule sampled from the pool of geometries $\{b_\eta\}$ for each molecule, η , within each minibatch, and \hat{b}_η denotes a different randomly sampled geometry from the same molecule, η . We used $\beta = 9$ in this work.

III. RESULTS

A. The GMTKN55 collection

The General Main-group Thermochemistry, Kinetics, and Non-covalent Interactions 55 (GMTKN55) dataset³² is a collection of 55 datasets aimed at probing the accuracy of quantum mechanical (QM) methods across a variety of chemical problems, ranging from reaction energies and electronic properties to non-covalent interaction energies and conformational properties. The dataset consists of 55 individual subsets with a total of 1505 relative energies based on 2462 single-point calculations. The high-level reference energies for the molecules in GMTKN55 are best-estimates calculated using a range of extrapolative protocols based on CCSD and CCSD(T) calculations collected from several different sources and most recently updated in Ref. 32.

Commonly, the performance of QM methods on GMTKN55 is presented via aggregated scores based on weighting of the mean absolute deviation to a reference, the WTMAD-1 or WTMAD-2

scores, with the difference between the two being the relative weighting of the individual subsets.³²

We note that we are comparing QM models and ML models that do not always provide predictions for every subset of GMTKN55, for example, due to the lack of capabilities to handle certain elements and charge- or spin-states in a model. In order to enable a fair comparison between models with varying coverage, we present WTMAD scores that are calculated only over the GMTKN55 subsets that contain elements and charge- and spin-states that are supported for each model and set the weight to 0 for the MAD for unsupported subsets. For the GFN n -xTB methods, we also exclude those datasets that contain spin-states other than singlets. Although we do note that while it is possible to run xTB calculations with unpaired electrons, these will not yield, for example, singlet-triplet splittings.⁵² An overview of the supported GMTKN55 subsets of each ML and xTB model compared in this text is given in the [supplementary material](#), Table S3.

Calculating the WTMAD-1 and WTMAD-2 scores for OrbNet Denali as described above, we find the WTMAD-1 and WTMAD-2 scores to be 7.19 and 9.84 against the high-level reference energies, respectively. We note that a number of the subsets included in these weighted metrics contain examples of chemistry that are not present in the dataset used to train OrbNet Denali, for example, water clusters, C₆₀ molecules, and transition state geometries. If these “extrapolative” subsets (PA26, NBPRC, ALK8, C60ISO, WCPT18, PX13, INV24, BHROT27, BHPERI, BHDI10, WATER27, and IDISP) are removed from the pool, the WTMAD-1 and WTMAD-2 scores are slightly changed by -0.60 and $+0.19$ – 6.59 and 10.03 , respectively, which indicates that OrbNet Denali is able to yield reasonable predictions for many chemical problems that are necessarily not covered by the training set.

When the weighted scores are calculated against ω B97X-D3/def2-TZVP reference energies (the same method used to generate the OrbNet Denali training data), the WTMAD-1 and WTMAD-2 scores are 6.42 and 8.40, respectively. For the version of OrbNet Denali trained on 10% of the data, the WTMAD-1 and WTMAD-2 scores are 7.40 and 11.41, respectively, against the ω B97X-D3/def2-TZVP reference, demonstrating positive effects of increasing the dataset size.

We note that WTMAD-1 and WTMAD-2 between ω B97X-D3/def2-TZVP and the high-level reference energies are 3.37 and 5.87, respectively, which in some sense constitutes an upper bound for the accuracy vs high-level reference energies of an OrbNet model trained on ω B97X-D3/def2-TZVP data.

A graphical overview of the performance of the OrbNet Denali and the models discussed in the following compared to ω B97X-D3/def2-TZVP reference data is displayed in [Fig. 1](#) and in the [supplementary material](#), Fig. S1.

In comparison, the popular semi-low-cost DFT method B97-3c⁵⁰ has WTMAD-1 and WTMAD-2 values for GMTKN55 of 5.76 and 10.22, respectively, compared to the high-level references, very close to the OrbNet Denali scores. For this dataset, OrbNet Denali is roughly 100 times faster than B97-3c.

Another popular family of low-cost QM methods is the series of GFN n -xTB ($n \in \{0, 1, 2\}$) methods.^{31,51,52} For these methods, we find the WTMAD-1 values to be 24.8, 14.7, and 12.0 for GFN0-xTB,⁶⁷ GFN1-xTB,³¹ and GFN2-xTB,⁵² respectively, with

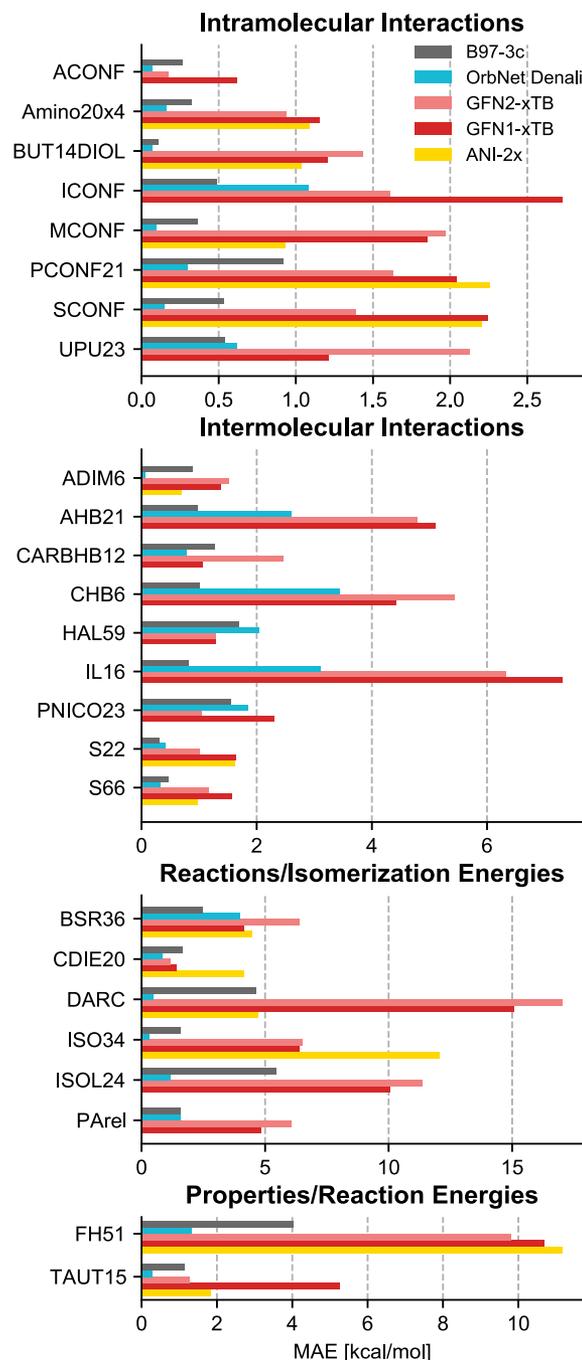


FIG. 1. The chart shows the MAE in kcal/mol for the subsets of the GMTKN55 that are covered by the OrbNet Denali training data relative to ω B97X-D3/def2-TZVP. For ANI-2x, values are left out for those subsets that contain elements or charge states that are not allowed by ANI-2x.

the same series of WTMAD-2 numbers being 38.0, 22.6, and 18.9. We point out that GFN1-xTB is the baseline method used to generate the input for OrbNet Denali and that OrbNet Denali yields a substantial improvement to this level of theory, resulting

in DFT-quality energy predictions across the covered parts of GMTKN55.

For the machine learning potentials ANI-1ccx and ANI-2x, we calculate the WTMAD- n scores over the subsets that only contain neutral singlet molecules with the elements that are covered by the individual methods (see [supplementary material](#), Table S3).^{10,53,68,69} For the ANI model that is parameterized against CCSD(T) reference data, namely, ANI-1ccx,⁵³ we find the WTMAD-1 and WTMAD-2 values to be 15.5 and 24.2, respectively, compared to CCSD(T) reference data. For the ANI-2x model,¹⁰ which, similarly to OrbNet Denali, is parameterized on DFT-level data, WTMAD-1 and WTMAD-2 are 14.2 and 23.9, respectively, with respect to CCSD(T) reference data. These WTMAD numbers are also found in [Fig. 2](#). A graphical overview of the performance of ANI-2x can be found in [Fig. 1](#) compared to ω B97X-D3/def2-TZVP reference data. We note that ANI-2x is parameterized from data calculated at the ω B97X/6-31G(d) level, and therefore, the comparison in [Fig. 1](#) does not reflect the true performance of ANI-2x compared to its underlying method since differences between the two DFT methods are expected.

In terms of coverage of common chemistry problems to which a general-purpose machine learning potential can be applied, OrbNet Denali has the broadest coverage of GMTKN55, to our knowledge, of any currently existing ML-based potential. OrbNet Denali covers 37 out of the 55 subsets; the OrbNet Denali training set does not cover the elements He, Be, and Al and a few heavy metals nor spin states other than singlets, for example, used to calculate ionization potentials and electron affinities. We note that even when extrapolating out of the training distribution to these other subsets, OrbNet Denali provides reasonable, but less accurate, results due to its baseline method, GFN1-xTB. The corresponding numbers for covered GMTKN55 subsets for ANI-1ccx and ANI-2x are 14 and 20, respectively. ANI-1ccx covers only neutral, singlet molecules with the elements H, C, N, and O, while ANI-2x extends this coverage to the elements F, Cl, and S.^{10,69} The family of GFN n -xTB methods has been

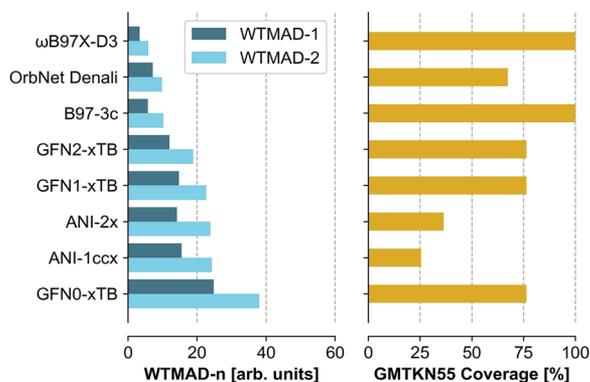


FIG. 2. Statistics over the accuracy and coverage of the GMTKN55 dataset³² is shown for a selection of methods, sorted by WTMAD-2 scores, relative to reference high-level³² estimates. The left panel shows the aggregated WTMAD-1 and WTMAD-2 metrics in arbitrary units, calculated on only those subsets that are covered by each method. The right panel shows the percentage of GMTKN55 subsets consisting only of molecules with elements, charged states, and spin states that are allowed within each model. The def2-TZVP basis set⁴⁶ was used for the ω B97X-D3 calculations.

parameterized on data containing elements up to Rn ($Z = 86$) but handles systems with odd numbers of electrons in a spin-averaged way and therefore covers 42 of the 55 GMTKN subsets. The basis for these numbers is also summarized in the [supplementary material](#), Table S3.

A graphical overview of WTMAD- n values and the coverage of GMTKN55 subsets for each method can be found in [Fig. 1](#).

A full overview of the performance of every method mentioned in this section compared to both CCSD(T) and ω B97X-D3/def2-TZVP reference values for all the 37 relevant GMTKN55 subsets is presented in the [supplementary material](#), Figs. S1 and S2.

B. Conformer scoring

Accurate determination of the ensemble of thermally accessible conformers is key to modeling molecules. In this section, we present results for a recent benchmark of conformer energetics.³³ This benchmark contains up to ten poses for each of ~ 700 drug-like molecules. Each molecule is comprised of C, H, N, O, S, Cl, F, P, Br, and I and contains between nine and fifty heavy atoms with a total charge between -1 and $+2$.

Accuracy in this benchmark for a given method is reported as the median R^2 value and is determined as follows. For every molecule, the correlation coefficient (R^2) is computed between the conformer energies of that molecule and the reference DLPNO-CCSD(T) energies. The median is then taken over the set of R^2 values corresponding to all molecules in the benchmark set.

[Figure 3](#) provides a direct comparison of OrbNet Denali to a representative sample of computational chemistry methods, including force fields, machine learning, semi-empirical, density functional theory, and wavefunction theory.^{27,33} For methods other than OrbNet Denali, a strong correlation is observed between accuracy

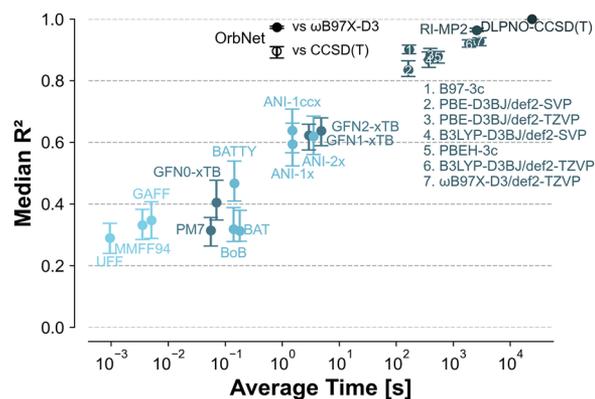


FIG. 3. A comparison between computational cost and the resulting accuracy for a number of methods for the Hutchison conformer benchmark set. The horizontal axis shows the average time of a single conformer energy computation, while the vertical axis denotes the median R^2 correlation coefficient for the molecules in the dataset. Error bars denote the 95% confidence interval for this number and are obtained through bootstrapping. The median correlation coefficient vs DLPNO-CCSD(T) is shown for all methods (filled circles). Additionally, the median correlation coefficient vs ω B97X-D3/def2-TZVP reference energies is shown for OrbNet Denali (open circle). This reference corresponds to the level of theory used to train the model. All data shown come from Ref. 33, with the exception of the two OrbNet Denali data points. For timings, all methods are run on a CPU, following the procedure described in Ref. 27.

and the logarithm of the average execution time of the method. OrbNet Denali breaks this trend, providing a median R^2 of 0.90 ± 0.02 vs the reference DLPNO-CCSD(T) at an average execution time of approximately one second per molecule. Here, the uncertainty refers to the 95% confidence interval and is obtained by bootstrapping the dataset. GFN1-xTB, the method used to generate input for OrbNet Denali, provides a median R^2 of 0.62 ± 0.04 with a similar execution time to OrbNet Denali. The median R^2 between OrbNet Denali and ω B97X-D3/def2-TZVP, the same method used to generate the training data, is 0.973 ± 0.004 , highlighting that OrbNet Denali is able to learn its underlying method to high accuracy. The OrbNet model train on 10% of the data reaches a lower median R^2 value of 0.94 ± 0.01 compared to ω B97X-D3/def2-TZVP, demonstrating the positive effects of increasing the amount of training data. Compared to the ω B97X-D3/def2-TZVP level of theory, which provides a similar accuracy vs DLPNO-CCSD(T) with a median R^2 of 0.92 ± 0.02 , OrbNet Denali results in a more than thousandfold speedup. This number also serves as an upper bound for the accuracy of a model trained on ω B97X-D3/def2-TZVP data and suggests that to increase the median R^2 for OrbNet Denali compared to DLPNO-CCSD(T), it is necessary to train on data that exceed the accuracy of DFT.

C. Non-covalent interactions (S66x10)

A standard benchmark for the accuracy of non-covalent interaction is the S66x10 benchmark set.^{34,70,71} This dataset consists of 66 different molecular dimers and their equilibrium geometries, along with nine additional displacements along the center-of-mass axis and corresponding CCSD(T)/CBS extrapolated binding energies.

For OrbNet Denali, we find the MAE and root mean square error (RMSE) to CCSD(T)/CBS to be 0.75 and 1.01 kcal/mol, respectively. We note that these numbers are very close the MAE and RMSE for the method used to generate the training data, ω B97X-D3/def2-TZVP, at 0.70 and 0.91 kcal/mol. Comparing OrbNet Denali to ω B97X-D3/def2-TZVP, we find smaller MAE and RMSE values at 0.46 and 0.65 kcal/mol, respectively. For the OrbNet model trained on only 10% of the data, these numbers increase to 0.67 and 0.85 kcal/mol compared to ω B97X-D3/def2-TZVP, respectively, suggesting that the increased training data size is, indeed, beneficial but that it is impossible for the model to substantially surpass the accuracy of the training data. The numbers referred to in this section are summarized in Table II.

D. Torsional profiles of drug-like molecules

Another common benchmark for potential energy surfaces is the accuracy with which torsional profiles can be reproduced. The recently published TorsionNet500 benchmark³⁵ compiles torsional profiles of 500 chemically diverse fragments containing the elements H, C, N, O, F, S, and Cl. For these torsion profiles, we have computed reference energies at the ω B97X-D3/def2-TZVP level, corresponding to the level of theory used to train OrbNet Denali. Following the work of Rai *et al.*,³⁵ we benchmark the performance of OrbNet Denali by comparing several different measures of accuracy discussed in the following text. An overview can be found in Table III. The first measure is the number of torsion profiles where the Pearson correlation coefficient (R) between the reference energies and the predicted energies is greater than 0.9. For OrbNet Denali, we

TABLE II. The MAE and RMSE binding energies for the S66x10 benchmark vs CCSD(T)/CBS reference binding energies.^{34,70} For rows marked with an asterisk (*), OrbNet Denali predictions are compared to binding energies calculated at the ω B97X-D3/def2-TZVP level. The latter reference corresponds to the same method used to generate the training data for OrbNet Denali.

Model	MAE (kcal/mol)	RMSE (kcal/mol)
OrbNet Denali	0.75	1.01
ω B97X-D3/def2-TZVP	0.70	0.91
B97-3c	0.49	0.64
GFN2-xTB	1.19	1.75
GFN1-xTB	1.35	1.87
GFN0-xTB	2.35	3.10
ANI-2x	1.44	1.96
ANI-1ccx	2.82	3.95
*OrbNet Denali	0.46	0.65
*OrbNet Denali (10%)	0.67	0.85

find this to be true for 99.4% of the profiles, while for OrbNet Denali (10%), the corresponding number is 98.8%, with average Pearson R values of 0.995 and 0.988, respectively. Second, the average MAE and RMSE for the torsion profiles are 0.12 and 0.18 kcal/mol for the full OrbNet Denali models and 0.23 and 0.34 kcal/mol for OrbNet Denali (10%). Finally, both OrbNet Denali models correctly predict the location of the global minimum to within 20° and its energy to within 1 kcal/mol for all 500 profiles. (These thresholds are taken from Ref. 35.) We note that these excellent results are achieved despite that the OrbNet Denali training set contains no torsion profiles.

For OrbNet Denali's baseline method, GFN1-xTB,³¹ the same figures are lower with 65.6% of the profiles having $R > 0.9$ and with an average R value of 0.832 and the average MAE and RMSE of 0.94 and 1.3 kcal/mol while capturing a good minimum for 89.4% of the predicted profiles.³¹ Figure 4 presents 25 torsion energy profiles of OrbNet Denali vs GFN1-xTB, stratified by the error of OrbNet Denali. OrbNet Denali reproduces every point along every torsion within a chemical accuracy of 1 kcal/mol in every case. Indeed, for all but the five worst cases, the OrbNet Denali torsion profile is quantitatively identical to that of the reference method. On the other hand, GFN1-xTB shows large errors for the several torsion profiles. In some cases, the shape of the GFN1-xTB profile is qualitatively incorrect. Overall, this suggests that the accuracy of OrbNet Denali is not solely due to the use of a good baseline method, but rather that OrbNet Denali is correctly able to capture the subtle differences in the torsion profiles.

We also compare torsion profiles calculated using another well-tested and reliable DFT method, B97-3c, to our reference profiles.⁵⁰ For B97-3c, the MAE and RMSE with respect to the ω B97X-D3 profile are 0.29 and 0.43 kcal/mol. While these numbers do not answer which is the more accurate DFT method, they do highlight the fact that OrbNet Denali is almost three times closer to the DFT reference than the variation between these two DFT methods. OrbNet Denali can therefore be considered on-par with DFT methods for this application.

Finally, we compare OrbNet Denali to the Merck Molecular Mechanics Force Field 94 (MMFF94) and the two ML-based

TABLE III. The performance of eight methods on the TorsionNet500 benchmark set. The reference energies are calculated at the ω B97X-D3/def2-TZVP level of theory, except for rows marked with an asterisk (*) where values are obtained from the work of Rai *et al.*,³⁵ benchmarked against a B3LYP/6-31G* reference. For a number of methods, the following statistics are shown: the percentage of the 500 torsion profiles for which the Pearson correlation coefficient (R) is greater than 0.9, the average Pearson R over the torsion profiles, the MAE and RMSE of the relative energies of the torsion profiles, and, finally, the percentage of torsion profiles where the global minimum of the profile is correct to within 20° and 1 kcal/mol.

Model	% profiles with Pearson $R > 0.9$ (%)	Average profile Pearson R	MAE (kcal/mol)	RMSE (kcal/mol)	% local minima with $\Delta\theta < 20^\circ$ and $\Delta E \leq 1$ kcal/mol (%)
OrbNet Denali	99.4	0.995	0.12	0.18	100.0
OrbNet Denali (10%)	98.8	0.988	0.23	0.34	100.0
B97-3c	97.4	0.985	0.29	0.43	100.0
GFN2-xTB	76.4	0.881	0.73	1.0	94.0
GFN1-xTB	65.6	0.832	0.94	1.3	89.4
GFN0-xTB	52.6	0.748	1.2	1.7	85.0
MMFF94	54.2	0.760	2.4	5.2	75.2
ANI-2x	73.2	0.899	1.3	1.9	91.8
* ANI-2x ³⁵	54	0.75	1.4	2.0	66
* TorsionNet ³⁵	79	0.91	0.7	1.3	83

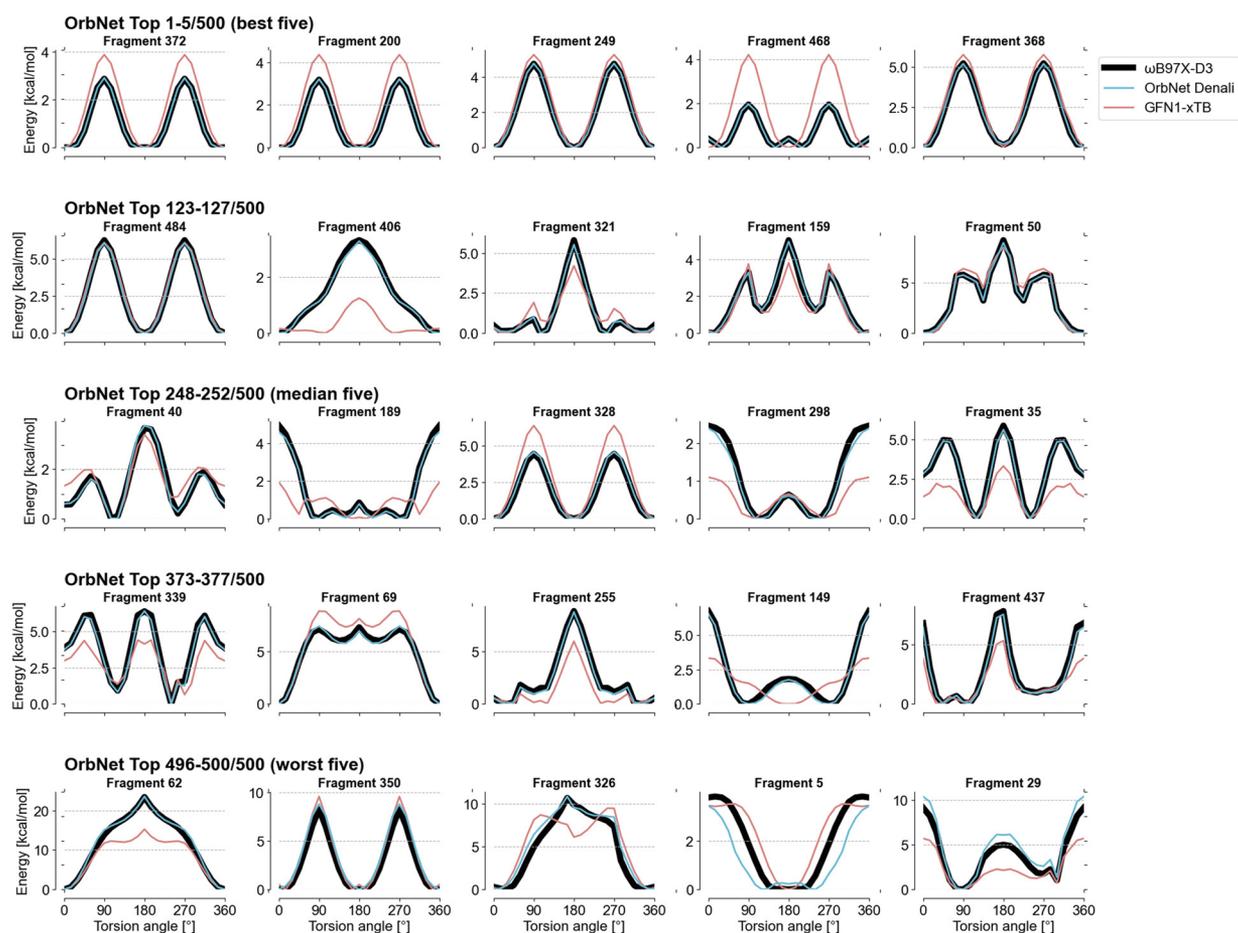


FIG. 4. The charts show torsional profiles of 25 drug-like molecules from the TorsionNet500 database,³⁵ stratified to represent quintiles of the OrbNet Denali error relative to the same torsional profiles calculated at the ω B97X-D3/def2-TZVP level of theory, shown as a reference (black). Additionally, the same torsion profiles are shown for OrbNet Denali's baseline method, GFN1-xTB (red).

methods, ANI-2x and TorsionNet.^{10,35,55} The MMFF94 force field is found to have the lowest accuracy of capturing the ω B97X-D3/def2-TZVP predicted minima, only finding the right minimum within the tolerance 75.2% of the time and with higher MAE and RMSE across the torsion profiles, at 1.4 kcal/mol and 5.2 kcal/mol, respectively. For ANI-2x, we find that a low-energy minimum is captured within the 20° tolerance with a 91.8% success rate, compared to the ω B97X-D3/def2-TZVP reference torsion profiles, which is better than MMFF94, GFN0-xTB, and GFN1-xTB. We note that ANI-2x is parameterized against ω B97X/6-31G(d) reference data, and therefore, it is possible that if that combination of DFT functional and basis set had been used to generate the reference curves, ANI-2x would have improved statistics in this test.

ANI-2x—while having better accuracy at finding the low-energy minima correctly—comes out with a larger MAE and RMSE than GFN0-xTB and GFN1-xTB due to the underestimation of the rotational barriers.

In addition to these numbers, we highlight the same benchmarks from Ref. 35, which compares ANI-2x and TorsionNet on the same structures, but against B3LYP/6-31G(d) energies. We preface this by noting, again, that ANI-2x is parameterized against ω B97X/6-31G(d) reference data, while TorsionNet is parameterized against B3LYP/6-31G(d) reference data, so it is possible that our reference data provide a more fair reference for ANI-2x.^{10,35} Against the B3LYP/6-31G(d) reference, TorsionNet is able to locate the low-energy minima with 83% success and ANI-2x with 66% success. The MAE and RMSE for TorsionNet against its torsion profile calculated at its own reference level of theory are 0.7 and 1.3 kcal/mol, respectively, while the MAE and RMSE for ANI-2x are 1.4 and 2.0 kcal/mol, respectively, which is within 0.1 kcal/mol from the same values vs the ω B97X-D3/def2-TZVP reference.

IV. CONCLUSIONS

We have presented a machine learning model, OrbNet Denali, that enables highly efficient energy prediction across broad swaths of chemical space relevant to biochemistry and organic chemistry applications.

The computational cost of running OrbNet Denali consists of the time spent to run the GFN1-xTB featurization plus additional overhead of running the inference. In our current central processing unit (CPU)-based implementation, the cost of inference is roughly comparable to that of running the GFN1-xTB featurization for molecules of the size presented in this paper. We note that large speedups on these steps are expected for a GPU-based implementation of OrbNet inference.

The accuracy of OrbNet Denali is comparable to that of modern DFT functionals, such as B97-3c, as demonstrated on several benchmark datasets. For the GMTKN55 benchmark set, 37 of the subsets can be evaluated with OrbNet Denali, only limited by the Periodic Table coverage and ability to handle non-singlet spin states. For these 37 subsets, the WTMAD-1 and WTMAD-2 scores were found to be 7.19 and 9.84, respectively, which is on par with many DFT functionals.³² For non-covalent interactions, OrbNet Denali reaches chemical accuracy for the dissociation curves of the S66x10 datasets. We find that OrbNet Denali provides conformer scoring on the Hutchison benchmark set with an accuracy of 0.90 ± 0.02 , which is better than some of the common DFT functional tested.

Finally, we report that OrbNet Denali reproduces the torsional profiles of 500 drug molecules in accordance with the reference model (ω B97X-D3/def2-TZVP) to an MAE of only 0.12 kcal/mol and, in all cases, was able to identify the correct global minimum within 20°.

OrbNet Denali learns the difference between the GFN1-xTB baseline energy and ω B97X-D3/def2-TZVP reference energies. Here, we note that, even in cases where the baseline method gives a qualitatively wrong description of the chemistry (such as incorrect torsional minima for some TorsionNet500 fragments and very large MAE in several GMTKN55 subsets), we observe that OrbNet Denali yields consistent energy predictions in good agreement with the reference DFT energies and therefore is not contingent on the accuracy of the baseline model.

Although our training data do not contain molecular systems similar to many of the systems in the GMTKN55 dataset, dissociation curves, or torsion profiles, OrbNet Denali, nevertheless, captures the correct physical behavior with very little error when compared to the reference model. This observation inspires confidence in future development of new OrbNet models. Natural developments will include training a larger model on more data, spanning even broader swaths of chemical space (and of the Periodic Table). Since the reference model could in many cases be reproduced with an error much smaller than the inherent error of the reference model itself, another potential path of development seems to be retraining the model on data obtained using higher level methods, such as coupled cluster theory, for example, through transfer learning.

We emphasize that even without these further developments, OrbNet Denali provides an exceptionally efficient drop-in replacement for DFT energy predictions for biological and organic chemistry.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for additional data on GMTKN55 benchmarks, including the MAE for every GMTKN55 subset for every method discussed in the main text, where applicable, and a table containing the GMTKN55 coverage for these methods.

ACKNOWLEDGMENTS

Z.Q. acknowledges graduate research funding from Caltech and partial support from the Amazon-Caltech AI4Science fellowship. T.F.M. and A.A. acknowledge partial support from the Caltech DeLogi fund, and A.A. acknowledges support from a Caltech Bren professorship. The authors acknowledge NVIDIA, including Abe Stern, Thorsten Kurth, Josh Romero, and Tom Gibbs, for helpful discussions regarding GPU implementations of graph neural networks. Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the DOE Office of Science, under Contract No. DE-AC02-05CH11231. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

AUTHOR DECLARATIONS

Conflict of Interest

Nine of the authors (A.S.C., S.K.S., M.B.O., D.G.A.S., F.D., P.J.B., M.W., F.R.M., and T.F.M.) are employees of Entos, Inc., or its affiliates.

Author Contributions

A.S.C. and S.K.S. contributed equally to this work.

DATA AVAILABILITY

The 2.3×10^6 geometries and energy labels in the Orb-Net Denali training set are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.14883867>.

REFERENCES

- S. Lorenz, A. Gross, and M. Scheffler, *Chem. Phys. Lett.* **395**, 210 (2004).
- J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- J. Behler, *J. Comput. Phys.* **145**, 170901 (2016).
- A. P. Bartók and G. Csányi, *Int. J. Quantum Chem.* **115**, 1051 (2015).
- T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, *npj Comput. Mater.* **3**, 37 (2017).
- Z. Li, J. R. Kermode, and A. De Vita, *Phys. Rev. Lett.* **114**, 096405 (2015).
- A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *J. Comput. Phys.* **285**, 316 (2015).
- A. Glielmo, P. Sollich, and A. De Vita, *Phys. Rev. B* **95**, 214302 (2017).
- J. S. Smith, O. Isayev, and A. E. Roitberg, *Sci. Data* **4**, 170193 (2017).
- C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev, and A. E. Roitberg, *J. Chem. Theory Comput.* **16**, 4192 (2020).
- S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, *Sci. Adv.* **3**, e1603015 (2017).
- S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, *Nat. Commun.* **9**, 3887 (2018).
- R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J. L. Klepeis, and D. E. Shaw, *J. Chem. Phys.* **147**, 161725 (2017).
- A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, *ACS Cent. Sci.* **5**, 57 (2019).
- A. Glielmo, C. Zeni, and A. De Vita, *Phys. Rev. B* **97**, 184307 (2018).
- K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, *J. Chem. Phys.* **148**, 241722 (2018).
- K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, *J. Chem. Theory Comput.* **15**, 448 (2019).
- B. Anderson, T.-S. Hy, and R. Kondor, [arXiv:1906.04015](https://arxiv.org/abs/1906.04015) (2019).
- J. Klicpera, J. Groß, and S. Günnemann, in International Conference on Learning Representations, 2019.
- O. T. Unke and M. Meuwly, *J. Chem. Theory Comput.* **15**, 3678 (2019).
- M. Welborn, L. Cheng, and T. F. Miller III, *J. Chem. Theory Comput.* **14**, 4772 (2018).
- L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, *J. Chem. Phys.* **150**, 131103 (2019).
- F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, *J. Chem. Phys.* **148**, 241717 (2018).
- A. S. Christensen and O. A. von Lilienfeld, *Chimia* **73**, 1028 (2019).
- A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, *J. Chem. Phys.* **152**, 044107 (2020).
- Y. Chen, L. Zhang, H. Wang, and W. E, [arXiv:2005.00169](https://arxiv.org/abs/2005.00169) (2020).
- Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller III, *J. Chem. Phys.* **153**, 124111 (2020).
- Z. Qiao, F. Ding, M. Welborn, P. J. Bygrave, D. G. A. Smith, A. Anandkumar, F. R. Manby, and T. F. Miller III, [arXiv:2011.02680](https://arxiv.org/abs/2011.02680) (2020).
- L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Phys. Rev. Lett.* **120**, 143001 (2018).
- J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in Proceedings of the 34th International Conference on Machine Learning, [arXiv:1704.01212](https://arxiv.org/abs/1704.01212) (2017).
- S. Grimme, C. Bannwarth, and P. Shushkov, *J. Chem. Theory Comput.* **13**, 1989 (2017).
- L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, *Phys. Chem. Chem. Phys.* **19**, 32184 (2017).
- D. Folmsbee and G. Hutchison, *Int. J. Quantum Chem.* **121**, e26381 (2020).
- J. Rezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.* **7**, 2427 (2011).
- B. Rai, V. Sresht, Q. Yang, R. J. Unwalla, M. Tu, A. M. Mathiowetz, and G. A. Bakken, [chemRxiv.13483185.v1](https://chemrxiv.org/abs/13483185.v1) (2020).
- CHEMBL database release 27, 2020.
- M. Davies, M. Nowotka, G. Papadatos, N. Dedman, A. Gaulton, F. Atkinson, L. Bellis, and J. P. Overington, *Nucleic Acids Res.* **43**, W612 (2015).
- D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, and A. R. Leach, *Nucleic Acids Res.* **47**, D930 (2018).
- P. J. Ropp, J. C. Kaminsky, S. Yablonski, and J. D. Durrant, *J. Cheminf.* **11**, 14 (2019).
- L. C. Blum and J.-L. Reymond, *J. Am. Chem. Soc.* **131**, 8732 (2009).
- G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *New J. Phys.* **15**, 095003 (2013).
- A. P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij, and A. R. Leach, *J. Cheminf.* **12**, 51 (2020).
- P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).
- L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, and C. D. Sherrill, *J. Chem. Phys.* **147**, 161727 (2017).
- Y.-S. Lin, G.-D. Li, S.-P. Mao, and J.-D. Chai, *J. Chem. Theory Comput.* **9**, 263 (2013).
- F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- F. Manby, T. Miller, P. Bygrave, F. Ding, T. Dresselhaus, F. Batista-Romero, A. Buccheri, C. Bungey, S. Lee, R. Meli, K. Miyamoto, C. Steinmann, T. Tsuchiya, M. Welborn, T. Wiles, and Z. Williams, [chemRxiv:7762646.v2](https://chemrxiv.org/abs/7762646.v2) (2019).
- R. Polly, H.-J. Werner, F. R. Manby, and P. J. Knowles, *Mol. Phys.* **102**, 2311 (2004).
- F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 73 (2011).
- J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, *J. Chem. Phys.* **148**, 064104 (2018).
- C. Bannwarth, S. Ehlert, and S. Grimme, *J. Chem. Theory Comput.* **15**, 1652 (2019).
- C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **11**, e01493 (2020).
- J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, *Nat. Commun.* **10**, 2903 (2019).
- X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith, and A. E. Roitberg, *J. Chem. Inf. Model.* **60**, 3408 (2020).
- T. A. Halgren, *J. Comput. Chem.* **17**, 490 (1996).
- N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *J. Cheminf.* **3**, 33 (2011).
- A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems*, edited by H. Wallach,

- H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), Vol. 32, pp. 8024–8035.
- ⁵⁸M. Wang, L. Yu, D. Zheng, Q. Gan, Y. Gai, Z. Ye, M. Li, J. Zhou, Q. Huang, C. Ma, Z. Huang, Q. Guo, H. Zhang, H. Lin, J. Zhao, J. Li, A. J. Smola, and Z. Zhang, [arXiv:1909.01315](https://arxiv.org/abs/1909.01315) (2019).
- ⁵⁹S. Li, Y. Zhao, R. Varma, O. Salpekar, P. Noordhuis, T. Li, A. Paszke, J. Smith, B. Vaughan, P. Damania *et al.*, [arXiv:2006.15704](https://arxiv.org/abs/2006.15704) (2020).
- ⁶⁰I. Loshchilov and F. Hutter, [arXiv:1608.03983](https://arxiv.org/abs/1608.03983) (2016).
- ⁶¹D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- ⁶²K. Choromanski, V. Likhoshesterov, D. Dohan, X. Song, A. Gane, T. Sarlós, P. Hawkins, J. Davis, A. Mohiuddin, L. Kaiser, D. Belanger, L. Colwell, and A. Weller, [arXiv:2009.14794](https://arxiv.org/abs/2009.14794) (2020).
- ⁶³S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) (2015).
- ⁶⁴J. L. Ba, J. R. Kiros, and G. E. Hinton, [arXiv:1607.06450](https://arxiv.org/abs/1607.06450) (2016).
- ⁶⁵R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- ⁶⁶Z. Qiao, A. S. Christensen, M. Welborn, F. R. Manby, A. Anandkumar, and T. F. Miller III, “UNiTE: Unitary n-body tensor equivariant network with applications to quantum chemistry,” [arXiv:2105.14655](https://arxiv.org/abs/2105.14655) [cs.LG] (2021).
- ⁶⁷P. Pracht, E. Caldeweyher, S. Ehlert, and S. Grimme, [ChemRxiv:8326202.v1](https://chemrxiv.org/abs/202108.03262) (2019).
- ⁶⁸J. S. Smith, O. Isayev, and A. E. Roitberg, *Chem. Sci.* **8**, 3192 (2017).
- ⁶⁹J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, and S. Tretiak, *Sci. Data* **7**, 134 (2020).
- ⁷⁰G. A. DiLabio, E. R. Johnson, and A. Otero-de-la-Roza, *Phys. Chem. Chem. Phys.* **15**, 12821 (2013).
- ⁷¹D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, *J. Phys. Chem. Lett.* **7**, 2197 (2016).