BAND NN: A Deep Learning Framework For Energy Prediction and Geometry Optimization of Organic Small Molecules

Siddhartha Laghuvarapu[†], Yashaswi Pathak[†], U Deva Priyakumar[*]

November 30, 2019

Abstract

Recent advances in artificial intelligence along with development of large datasets of energies calculated using quantum mechanical (QM)/density functional theory (DFT) methods have enabled prediction of accurate molecular energies at reasonably low computational cost. However, machine learning models that have been reported so far requires the atomic positions obtained from geometry optimizations using high level QM/DFT methods as input in order to predict the energies, and do not allow for geometry optimization. In this paper, a transferable and molecule-size independent machine learning model (BAND NN) based on a chemically intuitive representation inspired by molecular mechanics force fields is presented. The model predicts the atomization energies of equilibrium and non-equilibrium structures as sum of energy contributions from bonds (B), angles (A), nonbonds (N) and dihedrals (D) at remarkable accuracy. The robustness of the proposed model is further validated by calculations that span over the conformational, configurational and reaction space. The transferability of this model on systems larger than the ones in the dataset is demonstrated by performing calculations on select large molecules. Importantly, employing the BAND NN model, it is possible to perform geometry optimizations starting from non-equilibrium structures along with predicting their energies.

Address: Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500 032, India

[†]Authors contributed equally

[*]Corresponding Author Email: deva@iiit.ac.in; Phone: +91 40 6653 1161

1 INTRODUCTION

Accurate estimation of molecular energies is important for reliable modeling of various chemical and biological phenomena in general. Quantum mechanical (QM) and density functional theory (DFT) methods are the methods of choices for the calculation of accurate molecular energies and physicochemical properties. However, application of these methods to molecular systems is computationally expensive and is impractical for large systems. For modeling such systems, one resorts to the use of molecular mechanics (MM) force fields methods which are computationally tractable.[1–3] Force fields provide the potential energy of a molecule as a function of nuclear positions and has empirical parameters that are derived based on their ability to reproduce certain experimental and QM data via a detailed optimization procedure.[4–6] Though the force field methods in general are widely used to model biological macromolecules to study their dynamics, structural and thermodynamic properties, they are considered less accurate compared to *ab initio* or DFT methods.

In an attempt to develop new methods for predicting energies that are of DFT quality but are comparable to MM in terms of the computational cost, energy predictions have become an important application of supervised machine learning algorithms.[7–10] These algorithms have been shown to efficiently recognize patterns on training data which can be applied on unseen data. Traditionally, various regression techniques using kernel based methods[11] were used that convert 3-dimensional coordinates of a molecule into fixed length feature coordinates.[12–14] Recently, deep learning has become the sought after method for various supervised learning tasks due to their superior performance in several fields, primarily computer cision and natural language processing.[15–17] Various computational chemistry tasks[18] including quantum mechanical property prediction,[19–22] protein structure prediction,[23–25] protein-protein interactions[26], material property prediction,[7,27,28] retrosynthesis[29] and drug discovery[30–33] have been the targets of the machine learning methods and more recently deep learning applications.[8]

In order to provide a molecule as an input to a supervised learning algorithm, accurate description of a molecule as a vector is required.[22,34] In other words, it is helpful to have a vector representation that captures as much chemical information as possible. The descrip-

tor should precisely capture the atomic environment of each atom and should be sensitive to small changes in relative atomic positions. As hypothesized by,[35] molecular descriptors should follow these properties - rotational and translational invariance, invariance with respect to the permutation of atoms, provide a unique description of the atomic positions. Molecular descriptors in general suffer from inconsistency in terms of the size of molecules since most supervised learning algorithms require a fixed length representation of the input. Various approaches were proposed to tackle this problem. These approaches[12–14] extend the descriptor of every molecule in the set to the largest length descriptor by appending zeros at the end. These methods are not readily applicable to molecules larger than the ones trained with. Recent approaches have expressed total energy in terms of contributions from individual atoms[20,21,36] or has total energy broken down into contribution from individual bonds[37] where the individual feature vectors have fixed sizes.

The recent ML based methods generate DFT-level accurate potential energy surfaces, but their feature vectors are derived by transforming the nuclear coordinates of the constituent atoms, rather than explicit chemically intuitive terms. Smith et al.[36] used modification of symmetry functions originally developed by Behler and Parinello[38] to represent the local environment of each atom that are further used as inputs for the neural networks. Bartók et al. used smooth overlap of atomic positions (SOAP) to generate feature vectors.[39] Schütt et al. in their works[21,20] used nuclear charges (Z) and a matrix of inter-atomic distances as input to their model to find the energy of the molecule.

Although methods have been proposed that explicitly build feature vectors based on the bond topology of a molecule,[13,14] to the best of our knowledge they have not been demonstrated to generate potential energy surfaces or work on molecules larger than the ones present in the data set. In this paper, we propose a novel molecular descriptor inspired by classical force fields terms[1] - bonds (B), angles (A), non-bonded (N) interactions and dihedrals (D), which is named as BAND in this manuscript. A molecule is broken down into these terms and energy contribution from each of these terms is measured through several feed-forward neural networks. The sum of energies from each of the terms gives the total energy of the molecule. Through a series of studies that span over the conformational and configurational space, we show that our model can predict energies and potential energy

surfaces accurate to DFT-level. The applicability can be extended to molecules larger than the ones trained in the data set. We also demonstrate the ability of our model to perform geometry optimization of molecules to minimum energy when provided with an approximate structure over a defined bond topology. This is possible due to the nature of our molecular descriptor which is built taking into consideration the explicit bond topology of the molecule.

2 THEORY

Deep learning[40] has been shown to learn complex nonlinear functions through artificial neural networks. BAND NN proposed here uses feed-forward fully connected deep neural networks. These consist of multiple layers of nodes - an input layer, one or more hidden layers and an output layer. Each node is activated through weighted inputs from the previous layer and a non-linear activation function. The 'weights' are the optimizable parameters which can be trained through back-propagation of derivatives of an objective function with respect to each of them. The objective or cost function is a measure of deviation of the predicted output from the ground truth. As mentioned earlier, neural networks (NN) need a fixed length input feature vector. This creates a fundamental problem of obtaining accurate feature vectors starting from typical molecular representations such as internal and Cartesian coordinates whose dimensions change with respect to the number of atoms. Such a fixed length representation can further be used to train the NNs to predict molecular properties. The following subsections describe the feature vector/molecular representation, their relationship with classical force fields and the ML model used here.

2.1 BAND Molecular descriptor

A molecular descriptor that captures the essence of typical MM force field equations is used here. Each molecule is broken down into bonded pairs (atoms that are adjacent) and non-bonded pairs (atom pairs that are not adjacent). From this, lists of angles identified as two consecutive bonds forming an angle and lists of dihedrals identified as three consecutive bonds forming a dihedral angle were created. Each atom is represented by an eight dimensional feature vector: first four dimensions representing the atom name (the dataset used

here involves only four atoms C, N, O and H) and the second four dimensions representing the atom type in terms of how many of the C, N, O and H atoms are connected to it (Figure 1) essentially capturing the atom type as referred to in force fields.[1–3] Each bond is represented by a 17-dimensional vector which is the concatenation of the vectors representing the two atoms (eight-dimensions each) that form the bond followed the bond length. For the angle, it is the combination of the three atomic representations (twenty four) followed by the bond angle and two bond lengths making it a 27-dimensional vector. Similarly for the dihedral angle, it is a 38-dimensional vector made by four atomic representations followed by the dihedral angle, two angles and three bond lengths as given in Figure 1. The nonbond pair representation is similar to bonds where the bond length is replaced by the internuclear distance.

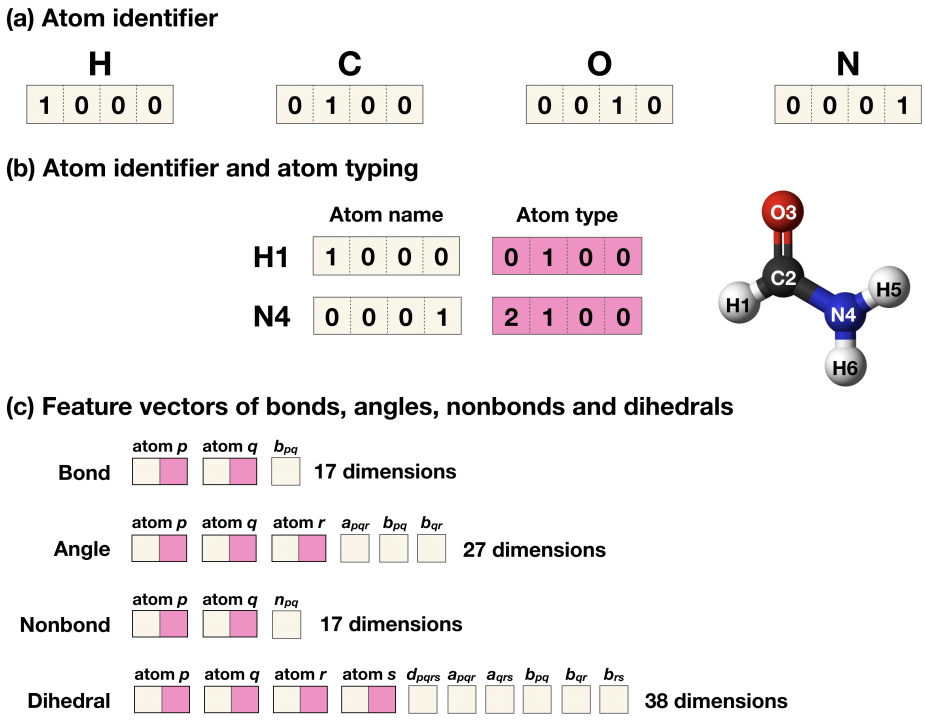

Figure 1: (a) Four dimensional feature vector for the atom name. (b) Eight dimensional feature vector for atom name and type. The atomic representation of two select atoms in formaldehyde is shown. (c) Schematic representation of the feature vectors of bonds, angles, nonbonds and dihedrals.

2.2 Resemblance to classical force field equations

A typical force field[1] equation is represented as the sum of energy contributions from the bonded (E_{bonded}) and non-bonded terms ($E_{nonbonded}$)). The (E_{bonded}) term usually involves energy as a function of bond lengths, bond angles and dihedrals angles in addition to other terms like Urey-Bradley and improproper dihedral terms depending on the force field, and the ($E_{nonbonded}$) term is typically a combination of an electrostatic and Lennard-Jones terms.

$$E_{total} = E_{bonded} + E_{nonbonded} \tag{1}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals} \tag{2}$$

The molecular representation proposed here is inspired by the force field equations where the total energy is expressed as sum of individual contributions from the bonded (bonds, angles and dihedrals) and non-bonded terms. In the force fields, the individual terms of the equation are expressed as a function of the nuclear coordinates in terms of bond lengths, angles, internuclear distances, etc. along with their characteristic constants. For eg, E_{bonds} is given as

$$E_{bonds} = \sum_{bonds} k_b(b - b_0)^2 \tag{3}$$

Here the constant k_b is the force constant that is characteristic of bond formed by the two participating atom types, b is the bond length and b_0 is the equilibrium bond length. The atom type typically captures the nature of the atom which comprise the atomic number and its connectivity. The molecular representation used here captures this by the eight dimensional vector for each atom. One modification is the implicit consideration of coupling between stretching and bending, and stretching, bending and rotation about single bonds akin to the class II force fields[41] (see Figure 1).

2.3 The Model

In this model, the atomization energy (difference between the molecular energy and that of the constituent atoms calculated at the DFT level) is expressed as the sum of the contributions from bonds, angles (coupled with bonds), dihedrals (coupled with bonds and angles), and non-bonds. More specifically, the contribution from each of these is estimated using a

feed-forward fully connected neural network. Four different models were trained for measuring contributions from bonds, angles, dihedrals and non-bonded terms.(See Figure 2). Each of these bonds, angles, dihedrals and non-bonded terms share the same weights and different types of these are differentiated by their feature vectors. This allows the model to be scalable with the number of atoms (or other bonded/non-bonded terms) as the final energy is only expressed as sum of the individual contribution from each term as given below.

$$E = \sum_{bonds} E_B + \sum_{angles} E_A + \sum_{nonbonds} E_N + \sum_{dihedrals} E_D \tag{4}$$

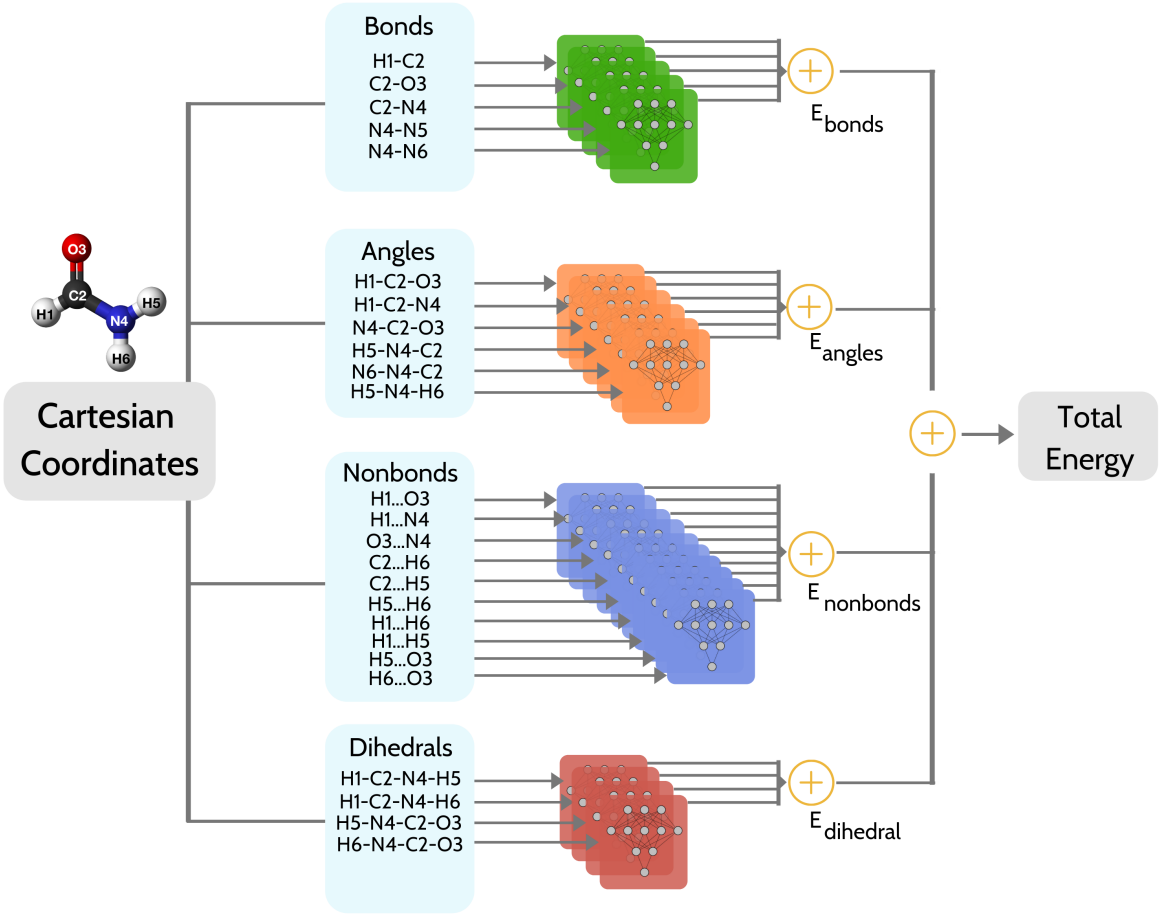

Figure 2: Schematic representation of the neural network architecture used for BAND NN. As an example, the list of bonds, angles, nonbonds and dihedrals for formaldehyde along with the number of neural networks used to predict the energies of each of these are shown.

3 METHODOLOGY

3.1 Data Selection

A subset of ANI-1 dataset[42] which is a large data set of non-equilibrium DFT total energy calculations for organic molecule with about 22 million molecular structures for 57,462 minimum energy structures was used for developing the ML model. These molecules were picked from the GDB-11 dataset[43,44] that has up to 8 heavy atoms containing only H, C, N and O. In addition to the equilibrium geometries obtained by performing geometry optimizations on ∼57k molecules at the ωB97X/6-31G(d), Smith et al. have used normal mode sampling to generate hundreds of non-equilibrium structures for each of the equilibrium structures resulting in ∼22 million data points. Single point energies of these configurations were calculated using the same method.[45] Although most methods use the QM-9 data set,[46] the conformation space is limited to equilibrium structures only and hence does not allow for calculating energies of non-equilibrium structures and hence geometry optimizations. All the equilibrium configurations along with each of their non-equilibrium structures whose relative energies with respect to the corresponding minimum energy structure are less than 30 kcal/mol were used for this study. The rationale are that, (a) most of the structure generation software (such as Gaussview[47]) are able to give initial geometries that are not too far away from the minimum, and (b) most of the drug design/biomolecular simulations do not aim to model bond breaking/forming. Hence optimization of structures generated using standard visualization software programs and for the purposes of such molecular modeling exercises, the chosen subset of the dataset is deemed adequate.

3.2 Data Pre-processing

Initial task is to make a list of all bonds, angles, nonbonds and dihedral angles for each of the configurations in the dataset for representing along the feature vectors proposed here. For a given molecule, the equilibrium structure was chosen to derive the molecular representation of its own and all its non-equilibrium structures. The list of bonds were generated using RDKit[48] based on the atomic coordinates of equilibrium structure which are extended to

corresponding non-equilibrium configurations. Once the list of bonds were derived, the lists of angles were generated by taking all possible 1,3 neighbors that are connected to 2, and similarly all 1,4 neighbors where 2 and 3 are connected were taken as dihedrals. For the non-bonded lists, all pairs except 1,2 whose distances are less than 6 Å in the equilibrium structure were considered.

3.3 Training

Keras deep learning framework[49] with TensorFlow[50] backend was used for all training and validation purposes. Fully connected networks were used for bonds, angles, non-bonds and dihedrals. Each network has an input layer, three hidden layers for each type and an output layer that measures the energy contribution from that term. Table 1 gives the dimensions of bond, angle, non-bond and dihedral networks used for BAND NN model. The output layer is a one dimensional vector that predicts the energy contribution from that particular network. The total energy contribution is the sum of energy predictions from all the networks. A train-test-validation randomly split in the ratio of 80-10-10 was used in this work. This resulted in ∼6.1 million data points in the training set and ∼760,000 data points each in the test and validation sets. Adam optimizer was used for updating weights with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ as suggested by Kingma and Ba.[51] Learning rate was set at 0.01 initially which was then gradually decreased to $10^{-}5$ by a factor of 10. All the intermediate layers were activated using the ReLU activation function.[52] The objective minimization function is the mean squared error between the predicted and actual atomization energies. The training data was iterated for 20 epochs until no notable increase in validation accuracy was observed.

3.4 Geometry optimization

As mentioned above, the ANI-1 dataset includes non-equilibrium structures that span over conformational and configurational space. This enables accurate prediction of energies at regions not limited to only minima on the potential energy surface but also higher energy structures. Geometry optimization involves finding the least energy structure of a molecule (minimum on its potential energy surface) given an approximate structure over a defined

Table 1: Dimensions of the input and hidden layers of the network architecture of BAND NN. Output dimension for each of the network is one.

Type of Network	**Input dimensions**	**Hidden Layer Dimensions**
Bonds	17	128-256-128
Angles	27	128-350-128
Non-bonds	17	128-256-128
Dihedrals	38	128-512-128

bond topology. In this study, the suitability of the proposed BAND NN model to be used for geometry optimization is demonstrated. The optimization technique used is the Nelder-Mead's method,[53] which is a popularly used direct search method for nonlinear optimization. The method is initialized by construction of a simplex by randomly sampling points on the target surface. The method propagates through generation of a sequence of simplices by repeatedly replacing the worst point on the simplex with better ones. The algorithm terminates either when the working simplex is sufficiently small or when the differences in function values on the vertices of the simplex is less than a threshold. The implementation of Nelder-Mead's optimizer method in the Scikit-learn library with the default parameters[54] was used for the results reported in the paper. Algorithm 1 (see below) describes the procedure followed for optimization of a molecule starting from its Cartesian coordinates and a defined bond topology.

The scripts along with example files for generation of molecular feature, training the model and prediction of energies are provided here: https://github.com/devalab/BAND-NN

Algorithm 1: Procedure for Geometry Optimization

Input: atomic coordinates, bond connectivity list

Initialise x to a z-matrix computed from atomic coordinates

Initialise $T \leftarrow 3$. This is a hyperparameter

Initialise $history \leftarrow [(\infty, x), (\infty, x), ...T \text{ times}], terminate \leftarrow False$

f is the function that takes z-matrix as the input and returns energy computed from BAND NN

while $terminate = False$ **do**

  $energy, x = History[0]$

  Set x to a different representation of z-matrix randomly

  Minimize $f(x)$ using Nelder-Mead's optimization procedure. This step returns $energy', x'$ at minima of f

  Append $(energy', x')$ to $history$ and sort $history$

  Set $worst\_performer$ to the last element of $history$

  if $worst\_performer = energy', x'$ **then**

    Set $terminate = True$

  else

    Delete last element from history

  end

end

4 RESULTS AND DISCUSSION

In this section, the accuracy of the model to predict atomization energies of molecules in the dataset and slightly larger molecules are presented. Following this, the ability of the BAND NN model to effectively learn the configurational and conformational space is demonstrated by predicting relative energies of isomers $C_{11}H_{22}$ and by performing potential energy scans on large drug molecules. This is followed by discussions on the predictive ability of the model for reaction energies of common organic reactions. Finally the importance of including the three-body and four-body terms for accurate predictions and the capability of BAND NN model for utilization in geometry optimizations are presented.

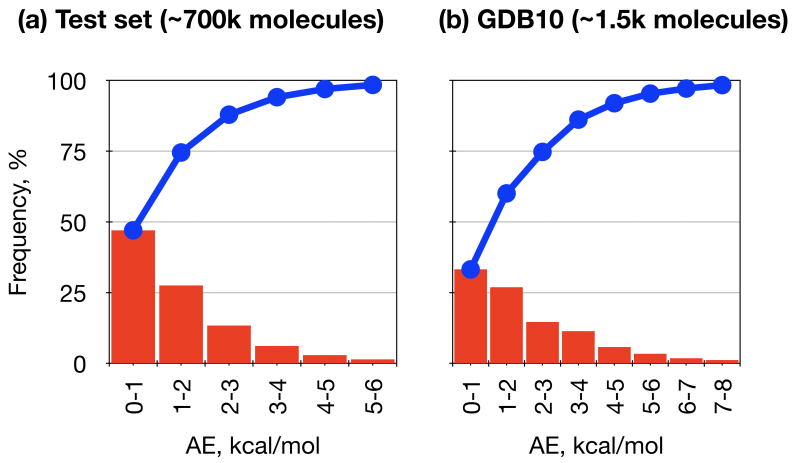

Figure 3: The histograms and the cumulative distributions of the absolute errors (in kcal/mol) calculated on (a) test set and (b) GDB-10 test set.

4.1 Accuracy of the BAND NN model

As mentioned in the section above, all the conformers that were under 30 kcal/mol in the ANI-1 data set from the corresponding minimum energy structure were chosen for this study. This data set had about 7.6 million conformers, and a 80-10-10 split for training, testing and validation was done on the dataset. A mean absolute error of 1.45 kcal/mol on the test set was obtained, which is expected to be significantly better than the small molecule force fields in general. The distribution of the absolute errors calculated for the test set comprising about 700,000 structures is given in Figure 3a. Predicted atomization energies of about 75% of structures in the test dataset are within 2 kcal/mol. To test the transferability of the BAND NN model to molecules with number of atoms more than that present in the training dataset, energies of molecules and their high energy structures with 10 heavy atoms were calculated (calculations on much larger systems are discussed later). Smith et al. performed normal mode sampling on 134 randomly chosen molecules with 10 heavy atoms from GDB-11 dataset.[43,44] From these, we picked all structures whose relative energies are under 30 kcal/mol with respect to their corresponding minimum. This resulted in 1500 structures and the mean absolute error of the atomization energies predicted using BAND NN for this set was found to be 2.1 kcal/mol, which demonstrates the transferability of the model to molecules larger than the ones trained with. The distribution of the absolute errors for this

set of structures are given in Figure 3b.

BAND NN is based on a feature vector inspired from classical force field terms, direct comparisons is more appropriate with models based on comparable feature vectors. The Bag of Bonds approach reports a mean absolute error of 1.5 kcal/mol on 7000 molecules from GDB-7 and 2.0 kcal/mol on 30% of QM9.[13] Bonds-in-Molecules neural network reports mean absolute error of 0.94 kcal/mol on the QM9 dataset.[37] Other recent approaches such as SchNet report a mean absolute error of 0.31 on QM9.[20] Recently, PhysNet model was proposed by Unke and Meuwly which reports a mean absolute error of 0.19 kcal/mol.[55] It is to be noted that all of these methods have only been validated on datasets containing equilibrium structures. The ANAKIN-ME approach reports a root mean squared error of 1.3 kcal/mol when trained on the entire ANI-1 dataset.[42] On the molecules from GDB-10 benchmark dataset prepared by Smith et. al, ANAKIN-ME reports mean absolute error of 0.83 kcal/mol for molecules with relative energies under 30 kcal/mol from their respective ground state conformer.[36]

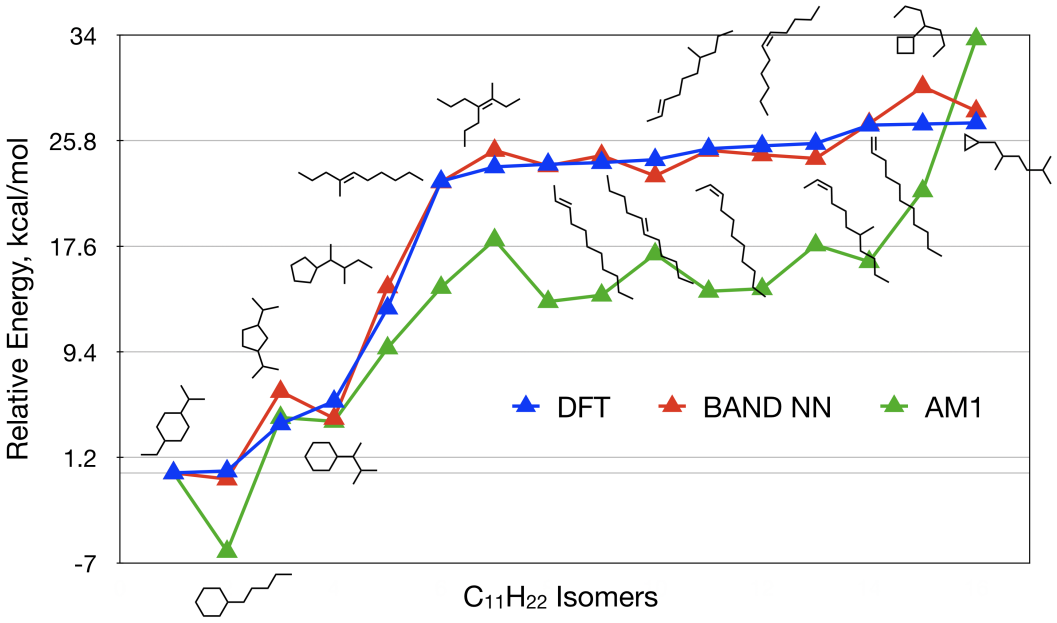

Figure 4: The relative energies (in kcal/mol) of select isomers of C_11H_22 relative to the least energy isomer calculated using the ωB97X/6-31G(d) level of theory, AM1 semiempirical method and using BAND NN.

4.2 Structural and Geometric Isomers

The accuracy of the proposed model in satisfactorily predicting the relative energies of structural and geometric isomers is examined here. Several isomers of $C_{11}H_{22}$ spanning diverse structural and geometric space, namely, linear chains, *cis-trans* isomers, varying ring sizes (three to six), etc. were chosen. The energies of the optimized geometries of these isomers were calculated using the ωB97X/6-31G(d) level of theory using the Gaussian 09 program[56]. Despite the diverse set of molecules considered for this evaluation, quantitative agreement between the DFT and BAND NN methods is observed (Figure 4). It is also found that the neural network model significantly outperforms the semiempirical quantum mechanical AM1 method.[57] This further indicates that machine learning based methods developed with molecular size invariant featurizations are capable of accurate modeling of molecular systems at the fraction of the computational expense that DFT or *ab initio* calculations would require.

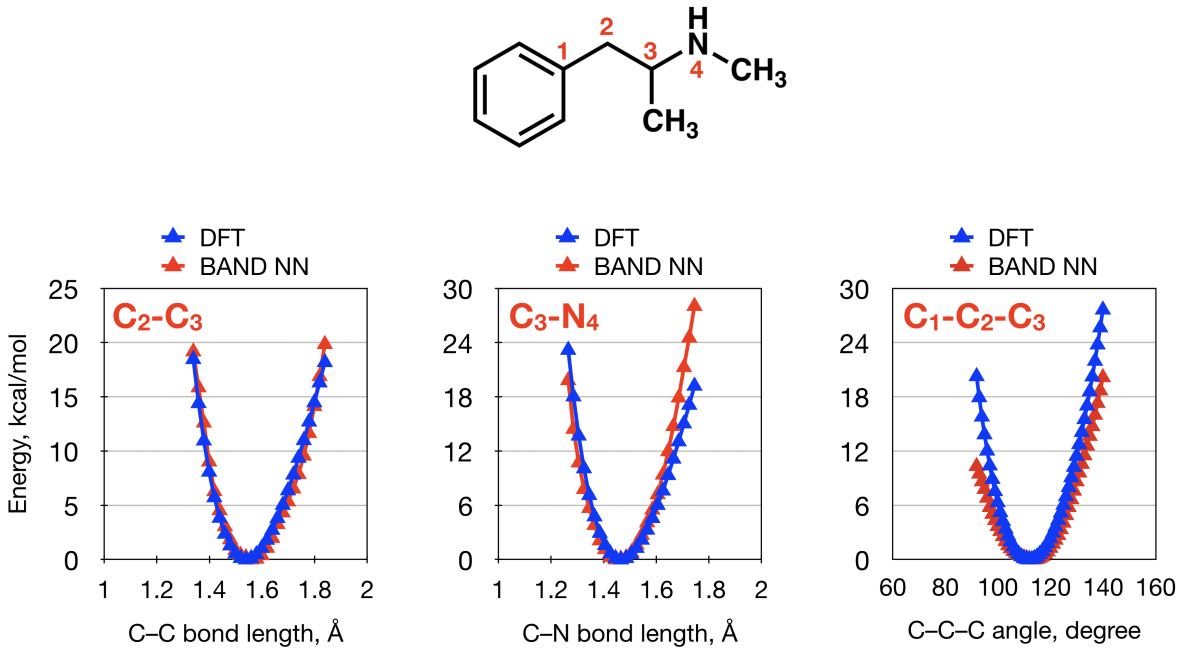

Figure 5: Potential energy surface (in kcal/mol) corresponding to C-C and C-N bond stretching and C-C-C angle bending of methamphetamine calculated using the ωB97X/6-31G(d) level of theory and BAND NN. The structure of the molecule along with the labels of atoms that were used for calculating the potential energies are given above the plots.

4.3 Potential energy surfaces

From the above discussions, it is apparent that the BAND NN model is capable of prediction atomization energies of small organic molecules very well. However, it is also important that models such as the one proposed in this paper are able to represent the potential energy surface of molecular systems and not just the energies for select points on the potential energy surface. Such a proper behavior of the model is necessary for it to be useful for performing energy minimizations, conformational analysis and force calculations in molecular dynamics simulations. Potential energy scans with respect to bonds and angles were performed on molecules that are significantly larger than those in the training set. Figure 5 gives the potential energy surfaces corresponding to C-C and C-N bond lengths calculated using the ωB97X/6-31G(d) level and BAND NN. For both the bonds, the positions of the minima are predicted accurately and the curves maintain a smooth curvature. Similarly, the potential energy scan for a C-C-C angle indicates very good agreement between the DFT results and the BAND NN data. To further show the chemical accuracy of the model, we performed conformational analysis for the central C-C bond of decane molecule and found very good agreement. The positions of the minima and maxima are predicted reasonably well along with the energies of different conformers and transition state with a mean absolute error of only 0.6 kcal/mol (Figure 6).

4.4 Reaction energies

In this section, the ability of the BAND NN model to predict reaction energies of simple organic reactions is examined. Some of the most simple and common reactions in organic chemistry (conformational differences stabilized by intramolecular hydrogen bonds, hydrogenation, Diels-Alder reaction, aldol condensation, esterification and electrocyclic ring closing reaction) were chosen for this analysis. The reaction energies calculated for these using the ωB97X/6-31G(d) level, AM1 method and BAND NN model along with the schematic diagrams of the reactions are given in Figures 7. All the reaction energies obtained using the BAND NN model are comparable to the DFT results. Among the six reactions, largest difference between the DFT and the BAND NN model was observed for the hydrogenation

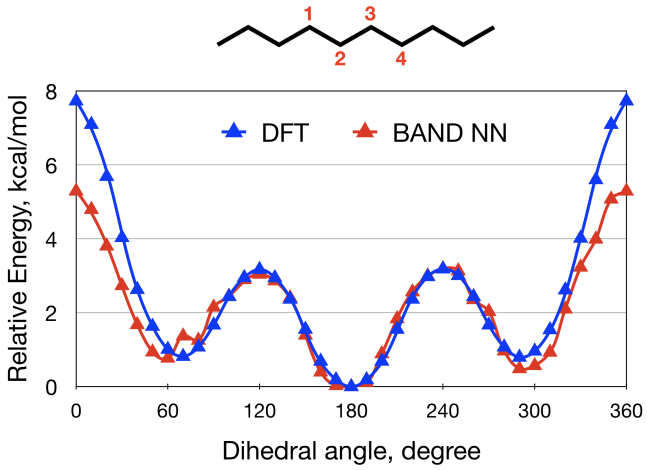

Figure 6: Potential energy surface (in kcal/mol) corresponding to the rotation about the central C-C single bond of *n*-decane calculated at the ωB97X/6-31G(d) level of theory and using BAND NN.

reaction. Notably, no data pertaining to the H_2 system was present in the training dataset. Similar to the prediction of relative energies of $C_{11}H_{22}$, the reaction energies computed using the BAND NN model outperfoms the AM1 level of theory.

4.5 Importance of 3,4 body terms

Most of the machine learning models for QM/DFT energy predictions have been done by including only two-body terms.[13,37] In this study, the energy is given as the sum of the energy contributions from all the bonds, angles, dihedral angles and nonbonded pairs. Two other models, one excluding the dihedrals (referred to as BAN NN model) and another excluding the angles and dihedrals (referred to as BN NN model) were trained using the same procedure as the BAND NN model to investigate the importance of including the 3- and 4-body terms. The distributions of the absolute error obtained from these models are given in Figure 8. The atomization energies are predicted within 2 kcal/mol for only about 50% and 60% of the molecules in the dataset in the BN NN and BAN NN models respectively. The mean absolute errors are 2.7 and 2.4 kcal/mol (1.45 kcal/mol for the BAND NN model). The performances of these models are inferior compared to the BAND NN model. Previous studies that utilized 'bag of bonds' feature involved the prediction of energies of molecules that are in their

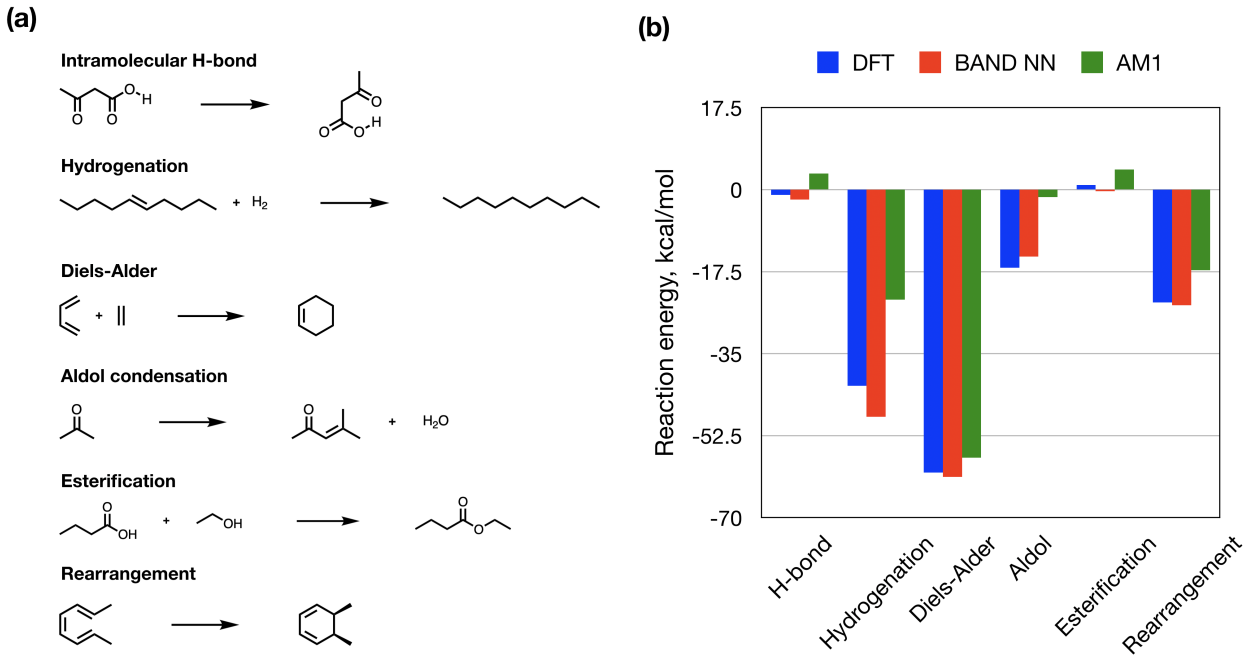

Figure 7: (a) Select organic reactions chosen for the calculation of reaction energies. (b) Reaction energies (kcal/mol) calculated using the ωB97X/6-31G(d) and AM1 levels of theory, and those predicted using BAND NN.

minimum energy states.[13] In other words, all the angles and dihedrals in these molecule are in their equilibrium values and hence the variances of the angles and dihedrals in the dataset are not large. In this study, we consider high energy configurations for each of the minimum energy structures for which the angles and dihedral angles are away from the minimum on the potential energy surface and hence sample a larger configurational/conformational space. This requires that the energy of the molecules is expressed as a function of angles and dihedral angles as well. Hence, the BAND molecular representation proposed in this manuscript is well suited for handling non-equilibrium structures compared to those that include only 2-body terms.

4.6 Geometry Optimization

Though there has been quite a few ML models to predict atomization energies of small organic molecules have been published in the last two years, there are few shortfalls. Some of these models cannot be applied to molecules larger than the ones in the training set, most of

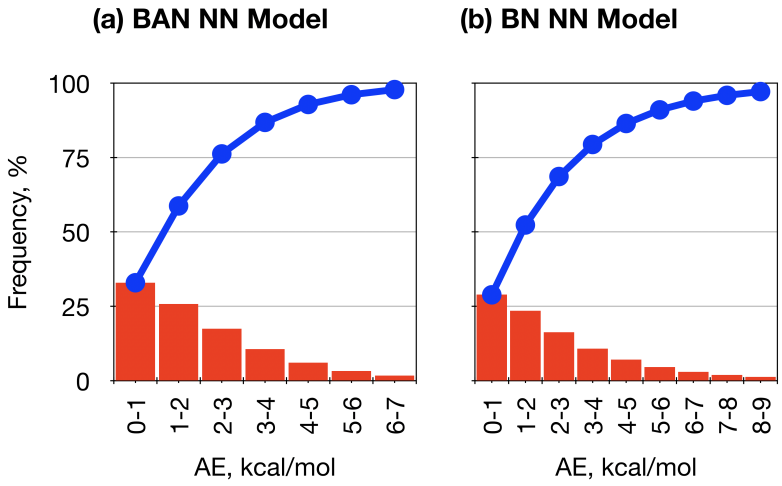

Figure 8: The histograms and the cumulative distributions of the absolute errors (in kcal/mol) calculated using the (a) BAN NN and (b) BN NN models.

Table 2: Input structure: Difference (kcal/mol) between the single point energies on the initial structure and the DFT optimized structure obtained at the ωB97X/6-31G(d) level. BAND optimized: Difference (kcal/mol) between the single point energies on the BAND NN optimized structure and the DFT optimized structure obtained at the ωB97X/6-31G(d) level. The structures of the molecules are given in Figure10

Molecule Name	**Input Structure**	**BAND Optimized**
1	5.5	1.7
2	10.7	3.1
3	4.9	1.5
4	9.1	3.4
5	17.5	7.6

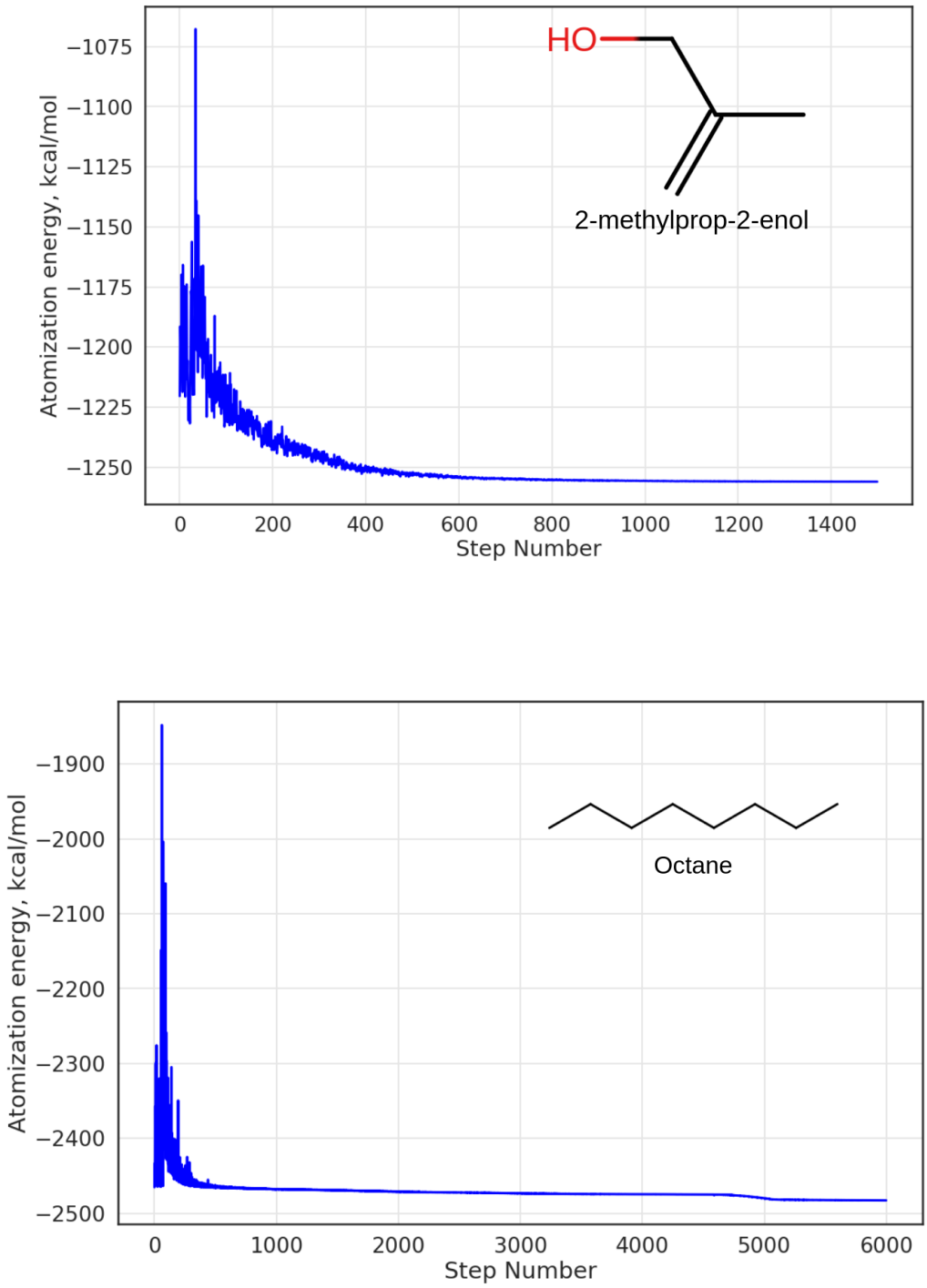

Figure 9: BAND NN atomization energies (kcal/mol) of 2-methylprop-2-enol and octane with respect to the optimization step number.

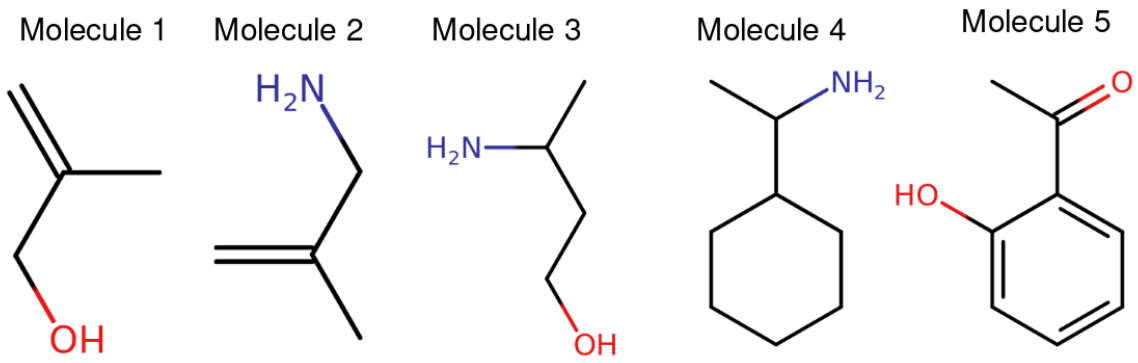

Figure 10: Molecules that were optimized starting from initial geometries generated using the GaussView program. Energies are presented in Table2

them cannot be applied to structures that are not in their minima on the potential energy surface and they have not been used for geometry optimizations. The condition that the geometry optimized using the DFT level has to be provided for the ML model to predict the energy is not desirable, since the geometry optimization involves calculation of the DFT energy. The next useful step in applying machine learning for molecular systems is to be able to develop models that allow for geometry optimization such that one could start from a structure away from the minimum and use the model along with an optimization method to reach the minimum. BAND NN model has been trained on high energy structures with explicit topology of the molecule as defined by the featurization used here. Nelder-Mead's optimization method has been used for updating the geometric parameters starting from a non-equilibrium structure. Starting from a reasonable guess structure of ocatane and 2-methylprop-2-enol, geometry optimization was performed. Figure 9 gives the energy of these molecules with respect to the optimization step number. The energies of the two molecules gradually decrease with respect to the optimization step and reaches convergence. For another test, few structures were generated using the GaussView program[47] (as an acceptable way of generating initial geometries in electronic structure theory calculations), and optimizations were performed using the Nelder-Mead's optimization employing the BAND NN. The single point energies of the initial and optimized geometries obtained using the ωB97X/6-31G(d) level are given in Table 2. In all the cases, the optimizer converged the molecules to structures whose energies are significantly lower than those of the initial struc-

ture. Though the results are not perfect for all the systems, it is clear that it is possible to use an appropriate molecular representation that will allow for geometry optimizations and that optimal structures can be obtained from this method. Implementation of gradient based methods may further improve the efficiency of the geometry optimization process.

CONCLUSIONS

A chemically intuitive molecular descriptor inspired from classical force field equation has been developed for prediction of atomization energy of small organic molecules. BAND NN model was trained on a subset of ANI-1 data set by choosing molecules that were at most 30 kcal/mol higher than the corresponding minimum. It was shown to accurately predict atomization energies with a mean absolute error of 1.45 kcal/mol on the test set. It accurately predicted the atomization energies of molecules randomly sampled from GDB-10, which are larger than the molecules in the data set. The model was demonstrated to be sensitive to structural and geometric isomers, generate accurate potential energy surfaces and predict reaction energies to DFT level accuracy on larger molecules. These experiments demonstrate that the model is transferable to larger molecules. In recent years, several methods have been proposed to predict atomization energy for ground state molecules, but for a model to be practically useful it should also be able to predict potential energy surfaces accurately. BAND NN model proposed in this work not only predicts the atomization energy for equilibrium and off-equilibrium structures but also can be used to perform geometry optimization. Further work in this area to develop robust transferable models using deep learning methods aimed at predicting accurate potential energy surfaces of molecular systems is expected to be more fruitful for state of the art problems in computational chemistry.

ACKNOWLEDGMENTS

We thank the DST-SERB (grant no. EMR/2016/007697) for the financial support.

References

1. MacKerell Jr, A. D., *J. Comput. Chem.*, 2004, **25**, 1584–1604.

2. Hollingsworth, S. A. and Dror, R. O., *Neuron*, 2018, **99**, 1129–1143.

3. Vanommeslaeghe, K.; Guvench, O. and MacKerell, A. D. J., *Curr. Pharm. Des.*, 2014, **20**, 3281–3292.

4. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M. and MacKerell Jr, A. D., *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

5. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. and Case, D. A., *J. Comput. Chem.*, 2004, **25**, 1157–1174.

6. Lemkul, J. A.; Huang, J.; Roux, B. and MacKerell, A. D., *Chem. Rev.*, 2016, **116**, 4983–5013.

7. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O. and Walsh, A., *Nature*, 2018, **559**, 547.

8. Goh, G. B.; Hodas, N. O. and Vishnu, A., *J. Comput. Chem.*, 2017, **38**, 1291–1307.

9. Ramakrishnan, R. and von Lilienfeld, O. A., *Rev. Comp. Chem.*, 2017, **30**, 225–256.

10. Mater, A. C. and Coote, M. L., *J. Chem. Inf. Model.*, 2019, **59**, 2545–2559.

11. Schölkopf, B.; Smola, A. J.; Bach, F. and others, , *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, 2002.

12. Rupp, M.; Tkatchenko, A.; Müller, K.-R. and Von Lilienfeld, O. A., *Phys. Rev. Lett.*, 2012, **108**, 058301.

13. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Muller, K.-R. and Tkatchenko, A., *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.

14. Huang, B. and Von Lilienfeld, O. A., *J. Chem. Phys.*, 2016, **145**, 161102.

15. Krizhevsky, A.; Sutskever, I. and Hinton, G. E. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

16. Bahdanau, D.; Cho, K. and Bengio, Y., *arXiv preprint arXiv:1409.0473*, 2014.

17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. and Bengio, Y. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

18. Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K. and Pande, V., *Chem. Sci.*, 2018, **9**, 513–530.

19. Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O. and Dahl, G. E. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272, 2017.

20. Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A. and Müller, K.-R., *J. Chem. Phys.*, 2018, **148**, 241722.

21. Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R. and Tkatchenko, A., *Nat. Commun.*, 2017, **8**, 13890.

22. Swann, E.; Sun, B.; Cleland, D. and Barnard, A., *Mol. Sim.*, 2018, **44**, 905–920.

23. Lyons, J.; Dehzangi, A.; Heffernan, R.; Sharma, A.; Paliwal, K.; Sattar, A.; Zhou, Y. and Yang, Y., *J. Comput. Chem.*, 2014, **35**, 2040–2046.

24. Jiang, Q.; Jin, X.; Lee, S.-J. and Yao, S., *J. Mol. Graph. Model.*, 2017, **76**, 379–402.

25. Hanson, J.; Paliwal, K.; Litfin, T.; Yang, Y. and Zhou, Y., *Bioinformatics*, 2018, **10**.

26. Romero-Molina, S.; Ruiz-Blanco, Y. B.; Harms, M.; Münch, J. and Sanchez-Garcia, E., *J. Comput. Chem.*, 2019, **40**, 1233–1242.

27. Raccuglia, P.; Elbert, K. C.; Adler, P. D.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J. and Norquist, A. J., *Nature*, 2016, **533**, 73.

28. Ward, L.; Agrawal, A.; Choudhary, A. and Wolverton, C., *npj Comput. Mater.*, 2016, **2**, 16028.

29. Segler, M. H.; Preuss, M. and Waller, M. P., *Nature*, 2018, **555**, 604.

30. Dral, P. O., *J. Comput. Chem.*, 2019.

31. Guimaraes, G. L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P. L. C. and Aspuru-Guzik, A., *arXiv preprint arXiv:1705.10843*, 2017.

32. Segler, M. H.; Kogej, T.; Tyrchan, C. and Waller, M. P., *ACS Cent. Sci.*, 2017, **4**, 120–131.

33. Y. Pathak, S. Laghuvarapu, S. M. and Priyakumar, U. D., *ChemRxiv*, 2019.

34. Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F. and von Lilienfeld, O. A., *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.

35. Behler, J., *Int. J. Quantum Chem.*, 2015, **115**, 1032–1050.

36. Smith, J. S.; Isayev, O. and Roitberg, A. E., *Chem. Sci.*, 2017, **8**, 3192–3203.

37. Yao, K.; Herr, J. E.; Brown, S. N. and Parkhill, J., *J. Phys. Chem. Lett.*, 2017, **8**, 2689–2694.

38. Behler, J. and Parrinello, M., *Phys. Rev. Lett.*, 2007, **98**, 146401.

39. Bartók, A. P.; Kondor, R. and Csányi, G., *Phys. Rev. B*, 2013, **87**, 184115.

40. LeCun, Y.; Bengio, Y. and Hinton, G., *Nature*, 2015, **521**(7553), 436.

41. Halgren, T. A., *J. Comput. Chem.*, 1996, **17**, 490–519.

42. Smith, J. S.; Isayev, O. and Roitberg, A. E., *Sci. Data*, 2017, **4**, 170193.

43. Fink, T.; Bruggesser, H. and Reymond, J.-L., *Angew. Chem. Int. Ed.*, 2005, **44**, 1504–1508.

44. Fink, T. and Reymond, J.-L., *J. Chem. Inf. Model.*, 2007, **47**, 342–353.

45. Chai, J.-D. and Head-Gordon, M., *J. Chem. Phys.*, 2008, **128**, 084106.

46. Ramakrishnan, R.; Dral, P. O.; Rupp, M. and Von Lilienfeld, O. A., *Sci. Data*, 2014, **1**, 140022.

47. Dennington, R.; Keith, T.; Millam, J. and others, , Gaussview, version 5, 2009.

48. Landrum, G., Rdkit: Open-source cheminformatics.

49. Chollet, F. and others, , Keras, 2015.

50. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y. and Zheng, X., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.

51. Kingma, D. P. and Ba, J., *arXiv preprint arXiv:1412.6980*, 2014.

52. Nair, V. and Hinton, G. E. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

53. Nelder, J. A. and Mead, R., *Comput. J.*, 1965, **7**, 308–313.

54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M. and Duchesnay, E., *J. Mach. Learn. Res*, 2011, **12**, 2825–2830.

55. Unke, O. T. and Meuwly, M., *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693.

56. 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.; Nakatsuji, H.; Caricato,

M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J. and Fox, D. J., Gaussian09 Revision E.01, 2009.

57. Dewar, M. J.; Zoebisch, E. G.; Healy, E. F. and Stewart, J. J., *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.