



Published in final edited form as:

*J Chem Theory Comput.* 2019 July 09; 15(7): 4113–4121. doi:10.1021/acs.jctc.9b00001.

## Predicting Molecular Energy using Force-Field Optimized Geometries and Atomic Vector Representations Learned from Improved Deep Tensor Neural Network

Jianing Lu<sup>1</sup>, Cheng Wang<sup>1</sup>, and Yingkai Zhang<sup>1,2,\*</sup>

<sup>1</sup>Department of Chemistry, New York University, New York, New York 10003

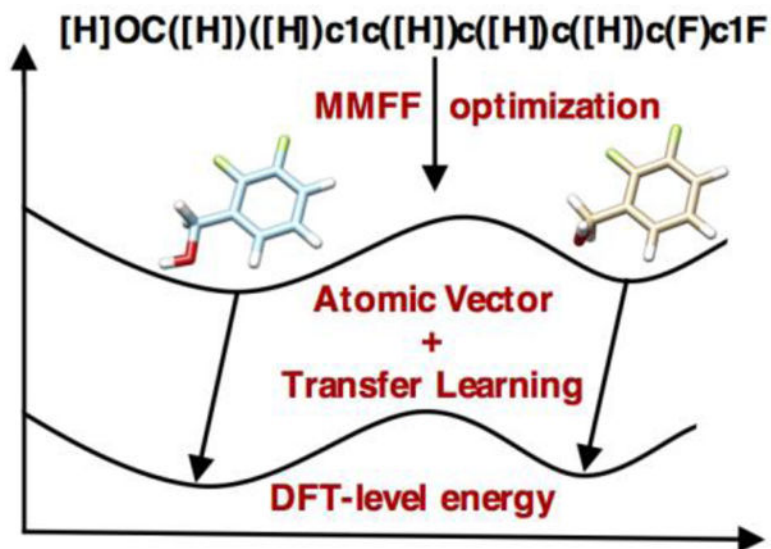
<sup>2</sup>NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

### Abstract

The use of neural networks to predict molecular properties calculated from high level quantum mechanical calculations has made significant advances in recent years, but most models need input geometries from DFT optimizations which limit their applicability in practice. In this work, we explored how machine learning can be used to predict molecular atomization energies and conformation stability using optimized geometries from Merck Molecular Force Field (MMFF). Based on the recently introduced deep tensor neural network (DTNN) approach, we first improved its training efficiency and performed an extensive search of its hyperparameters, and developed a DTNN\_7ib model which has a test accuracy of 0.34 kcal/mol mean absolute error (MAE) on QM9 dataset. Then using atomic vector representations in the DTNN\_7ib model, we employed transfer learning (TL) strategy to train readout layers on the QM9<sub>M</sub> dataset, in which QM properties are the same as in QM9 [calculated at the B3LYP/6–31G(2df,p) level] while molecular geometries are corresponding local minima optimized with MMFF94 force field. The developed TL\_QM9<sub>M</sub> model can achieve an MAE of 0.79 kcal/mol using MMFF optimized geometries. Furthermore, we demonstrated that the same transfer learning strategy with the same atomic vector representation can be used to develop a machine learning model that can achieve an MAE of 0.51 kcal/mol in molecular energy prediction using MMFF geometries for an eMol9\_C<sub>M</sub> conformation dataset, which consists of 9959 molecules and 88,234 conformations with energies calculated at the B3LYP/6–31G\* level. Our results indicate that DFT-level accuracy of molecular energy prediction can be achieved using force-field optimized geometries and atomic vector representations learned from deep tensor neural network, and integrated molecular modeling and machine learning would be a promising approach to develop more powerful computational tools for molecular conformation analysis.

### Graphical Abstract

\*To whom correspondence should be addressed. yingkai.zhang@nyu.edu.



## I. INTRODUCTION

Molecular conformation analysis is essential in elucidating molecular structure-property relationship and is often a prerequisite for structure based molecular design<sup>1-3</sup>. Reliable identification of low-energy conformers for simple small molecules can be achieved with high-level quantum mechanical calculations<sup>4</sup>, but would be computationally demanding for more complicated drug-like molecules<sup>3</sup>. Currently in structure based rational drug design, computational estimation of ligand conformation stability is mostly dependent on molecular mechanical force fields<sup>3, 5-7</sup>, which is computationally efficient but is limited by the accuracy of force fields.

In recent years, machine-learning based methods have made remarkable progresses in molecular energy predictions<sup>8-52</sup>. One significant advance is deep tensor neural network (DTNN)<sup>10</sup>, which employs atomic number and interatomic distance matrix as the input and utilizes a flexible graph neural network to predict molecular properties based on atomic vector representations. In DTNN, initial atomic vector  $c_i^{(0)}$  for an atom  $i$  is initialized based on its atomic number and then iteratively updated to final atomic vector  $c_i^{(T)}$  by using  $T$  interaction blocks (ib). Each ib is a tensor layer to model intermolecular interactions with other atoms in the molecule, and  $T = 2$  or  $3$  has been utilized in DTNN.  $c_i^{(T)}$  is then fed into two fully connected readout layers to predict an atomic energy contribution  $E_i$ , and the sum of all atom-wise energies in a molecule is the total molecular energy  $E$ . With this relatively simple and flexible neural network, DTNN achieved a test accuracy of 0.84 kcal/mol MAE for the QM9 dataset<sup>53-55</sup>. Subsequently, several deeper neural network models with more sophisticated network architectures have been developed and the state of the art performance in predicting molecular energies is progressing rapidly. For example, SchNet<sup>56-57</sup>, which utilizes continuous-filter convolutions with filter-generating subnetworks to model the interaction terms and more atom-wise layers in interaction block, has achieved a much

smaller MAE on QM9 dataset. However, such impressive test accuracy has so far only achieved with DFT optimized geometries, which limits the efficiency of these neural network based models in practical applications, such as for identifying low-energy conformers in high-throughput screening. Thus, we are motivated to investigate whether and how machine learning can be used to accurately predict DFT calculated molecular atomization energies and conformation stability using optimized geometries with Merck Molecular Force Field (MMFF)<sup>5</sup>. In 2017, Raghunathan et al<sup>16</sup>. introduced a machine learning method which linearly combines low-level quantum chemical properties and machine learning corrections to get highly accurate predictions. Different from their work, no baseline atomization energy of MMFF is needed in our current work, and transfer learning (TL) strategy<sup>58-59</sup> together with atom representation learned from deep-learning model are utilized to directly evaluate DFT-level atomization energies based on MMFF optimized geometries.

Our overall workflow is illustrated in Figure 1. First, we demonstrated importance of training efficiency improvement, hyperparameter search and data augmentation in the optimization of machine learning models. Based on the same DTNN architecture, we developed a DTNN\_7ib model which is a DTNN model with 7 interaction blocks and has a much-improved test accuracy of 0.34 kcal/mol MAE on the QM9 dataset. Then, we built two new data sets QM9<sub>M</sub> and eMol9\_C<sub>M</sub>, in which geometries are optimized with the MMFF94 force field and molecular energies are calculated with a DFT method at the corresponding DFT-minimized geometry in same local minima, and explored transfer learning strategy using atomic vector representations from the DTNN\_7ib model. The results from TL\_QM9<sub>M</sub> and TL\_eMol9\_C<sub>M</sub>, which are two models developed with QM9<sub>M</sub> and eMol9\_C<sub>M</sub> dataset respectively, indicate that our presented machine learning strategy can achieve DFT-level accuracy of molecular energy prediction using force-field optimized geometries. Finally, we generated another new data set Plati\_C<sub>M</sub> beyond nine heavy atoms to test the generality and applicability of the developed TL\_eMol9\_C<sub>M</sub> model in molecular conformation analysis. All of our presented machine learning models, corresponding source codes as well as data sets are freely available on the web at: <https://www.nyu.edu/projects/yzhang/IMA>.

## II. DATASET

A critical component in developing machine learning models is dataset. In this work, besides employing the QM9 dataset, we have generated three new data sets: QM9<sub>M</sub>, eMol9\_C<sub>M</sub>, and Plati\_C<sub>M</sub>, as listed in Table 1 and described in detail below.

QM9, which includes ~134k structures with 9 or less heavy atoms (C, O, N, F) from GDB-9 dataset, has become the most widely-used dataset for developing machine-learning models to predict quantum-chemical molecular energies. In the QM9 dataset, all molecular geometries are optimized at B3LYP/6-31G(2df,p) level and corresponding molecular properties are calculated at the same level<sup>53</sup>. We use this dataset for developing and benchmarking the DTNN\_7ib model from DTNN model.

QM9<sub>M</sub> dataset is built based on QM9 dataset, and the sole difference is in the molecular geometries. The molecular geometries in QM9<sub>M</sub> dataset were obtained from the optimization of the molecular geometries in QM9 dataset with MMFF94 force field using RDKit<sup>60</sup>. Molecular properties in QM9<sub>M</sub> dataset are kept the same as in QM9 dataset. The QM9<sub>M</sub> dataset is developed to explore whether DTNN\_7ib model can be directly applied for energy prediction with force-field optimized geometries.

eMol9\_C<sub>M</sub> dataset is a newly generated conformation dataset, and it is designed to develop models for energy prediction of low-energy conformers with force-field optimized geometries. This dataset is built based on eMolecules<sup>61</sup> dataset, which contains SMILES (Simplified Molecular-Input Line-Entry System) of purchasable compounds, and QM9 dataset. Firstly, the molecules are selected from the overlap set of the eMolecules and QM9. For each selected compound, we employed RDKit to generate up to 300 conformations from SMILES using ETKDG<sup>62</sup> method. Similar conformations have been removed after Butina<sup>63</sup> clustering with 0.2 Å RMSD cutoff, and redundant mirror-image conformations have been cleaned after the RMSD calculation between each pair of conformations by ArbAlign<sup>64</sup> with consideration of symmetry. Secondly, each conformation was minimized with MMFF94, and B3LYP/6-31G\* minimization<sup>65</sup> was conducted on MMFF optimized geometry to get corresponding DFT-level energy. Here, 6-31G\* basis set has been applied since it speeds up the computation in comparison with the 6-31G(2df, p) basis set which was used for developing QM9 dataset. Since DFT energies in the eMol9\_C<sub>M</sub> dataset are calculated by different basis set as in QM9 dataset, it would be a more stringent applicability test for transfer learning with atomic embedding obtained in the DTNN\_7ib model. The eMol9\_C<sub>M</sub> dataset includes 88,234 conformations from 9,959 molecules, and its distribution of RMSD between each pair of MMFF optimized geometry and DFT optimized geometry is shown in Figure S1.

Plati\_C<sub>M</sub> dataset is created to evaluate the extrapolation performance of TL\_eMol9\_C<sub>M</sub> model and it includes 4076 conformations for 74 molecules with 10 to 12 number of heavy atoms from Platinum dataset<sup>66</sup>, which is a dataset of protein-bound ligand conformations whose high-quality X-ray structures have been determined and are available from the PDB database<sup>67</sup>. Conformation generation procedure for Plati\_C<sub>M</sub> dataset is the same as the procedure used for eMol9\_C<sub>M</sub>.

### III. METHOD

#### A. Neural Network Model Architecture

As illustrated in Figure 2, DTNN\_7ib has the same network architecture as DTNN. The difference between two models is hyperparameter setting, which has been shown in Table 2. Inputs of DTNN\_7ib for a molecule with  $n$  atoms include an atomic number vector  $Z = [Z_1 \dots Z_n]$  and an interatomic distance matrix  $D$  computed from their 3D coordinates. Each interatomic distance is described by a vector from the following Gaussian function:

$$\hat{d}_{ij} = \left[ \exp\left(-\frac{(D_{ij} - k\Delta\mu)^2}{2\sigma^2}\right) \right]_{0 \leq k \leq \mu_{max}/\Delta\mu}$$

where  $\mu_{max}$ ,  $\Delta\mu$  and  $\sigma$  are the hyperparameters of a DTNN model. Here, we set  $\sigma$  has same value as  $\Delta\mu$ , which can also be different.

For initial atomic vector  $c_i^{(0)}$  of atom  $i$ , it is initialized randomly based on its atomic number, and optimized during training. At  $t(1 \leq t \leq T)$  interaction block, an updated atomic vector  $c_i^{(t)}$  is obtained by aggregating its atomic environment term  $v_{ij}^{(t)} = \sum_{j \neq i} v_{ij}^{(t)}$  with  $c_i^{(t-1)}$ :

$$c_i^{(t)} = c_i^{(t-1)} + \sum_{j \neq i} v_{ij}^{(t)}$$

The interaction vector  $v_{ij}^{(t)}$  between atom  $i$  and atom  $j$  is computed as flowing:

$$v_{ij}^{(t)} = f_{act} \left[ W_{fc}^{(t)} \left( \left( W_{cf}^{(t)} c_j^{(t-1)} + b_{cf}^{(t)} \right) \circ \left( W_{df} \hat{d}_{ij} + b_{df} \right) \right) \right]$$

where  $c_j^{(t-1)}$  is the current atom vector for neighboring atom  $j$  ( $j \neq i$ ),  $\hat{d}_{ij}$  is the expanded distance vector between atom  $i$  and atom  $j$ ,  $\circ$  is the element-wise multiplication of two matrices, and  $f_{act}$  is the activation function. After generating the final atomic vector  $c_i^{(T)}$ , 2 fully connected layers are added as readout layers to predict the atomic energy for atom  $i$ , and the sum of atomic energies of all atoms in the molecule is the molecular energy.

Using atomic vector representations in the DTNN\_7ib model, we employed transfer learning strategy and retrained readout layers for QM9<sub>M</sub> and eMol9\_C<sub>M</sub> datasets to develop two new machine learning models: TL\_QM9<sub>M</sub> and TL\_eMol9\_C<sub>M</sub>, respectively, which use MMFF optimized geometries as inputs. For both TL models, the number of readout layers has been increased to 4 and activation function in readout layers has been changed into shifted softplus (ssp).

## B. Neural Network Model Implementation, Training and Evaluation

All neural network models in this work were implemented with TensorFlow (version 1.4.1)<sup>68</sup>. To improve data-input efficiency, all input data were saved in the tfrecord format and data import pipelines were built using tf.data API. Meanwhile, cyclic cosine annealing training schedule<sup>69</sup> together with Adam optimizer was utilized to improve training efficiency, in which the learning rate at  $i$  iteration is as following:

$$\alpha(i) = \frac{\alpha_0}{2} \left( \cos \left( \frac{\pi \text{mod}(i-1, \lceil \frac{I}{M} \rceil)}{\lceil \frac{I}{M} \rceil} \right) + 1 \right)$$

where  $\alpha_0$  is the initial learning rate,  $I$  is the number of total iterations,  $M$  is the number of total cycles. Here,  $\alpha_0 = 0.001$  and  $M = 8$ .

To train the DTNN\_7ib model, QM9 dataset have been split randomly into training (99k), validation(1k) and test set (33,885), and 400k iterations with batch size equals to 100 have been employed. In order to improve atomic vector representations, we have augmented the training data set with 1.5 k distant diatomic systems, in which two atoms have distance between 20 Å and 30 Å and their molecular energy is assumed to be the sum of their atomic reference energies. Thus, the total training set includes 100.5k molecules. MAE of validation set has been checked every 1k iterations and the model with the lowest MAE has been picked as our final model. The average performance of five models trained upon data split by different random seed have also been computed.

The training/validation/test split for developing transfer learning models are also shown in Table 1. To avoid any training-testing overlap bias, the split for developing TL\_QM9M is the same as in optimizing DTNN\_7ib, and eMol9\_CM dataset has been divided by molecule types. Both models were trained using the same training schedule as DTNN\_7ib. To evaluate model performance, besides normal absolute error ( $\text{Error}_A$ ), which is the error between predicted energy and target energy for each conformation, we also computed relative error ( $\text{Error}_R$ ) to consider model's ability on conformation stability calculation. The relative error can be computes as

$$\text{Error}_R = \frac{\sum_i^n \text{error\_metric}(E_i^{(c)} - E_i^{\min})}{n}$$

where error\_metric can be MAE or RMSE,  $E_i^{(c)}$  is the energy of conformation  $c$  for molecule  $i$  and  $E_i^{\min}$  is the lowest energy of the molecule  $i$ , and  $n$  is the number of total molecules.

This relative error also enables us to compare different methods without the limitation of different reference energy. Success rate for finding the right lowest conformation for all molecules in the test set has also been calculated.

## IV. RESULT and DISCUSSION

### A. DTNN\_7ib

One significant advance in developing neural network models to predict DFT molecular energies is DTNN, which achieved a test accuracy of 0.84 kcal/mol MAE for the QM9 dataset. Subsequently, several more recent works, including SchNet, have achieved much-improved test accuracy with more sophisticated network architectures. Thus, one interesting question is whether DTNN itself can be significantly improved by hyperparameter searching, which is an important element of machine learning model development. However, one key challenge is its training efficiency since it took about a week to optimize the published DTNN model (162h for  $ib = 3$  on an NVIDIA Tesla K40 GPU), while hyperparameter searching needs to train many models. Here we accelerated training data input efficiency by using tfrecord and tf.data API, and our reimplemented DTNN code can achieve a speed-up of more than ten times in comparison with using the original code downloaded from <https://github.com/atomistic-machine-learning/dttnn>. Meanwhile, we employed a cyclic cosine annealing training schedule (cyclic training schedule) instead of



commonly-used learning rate exponential decay training schedule. As shown in Figure S2, the model training efficiency and performance can be significantly improved by using cyclic training schedule.

With more efficient model training, we first examined the atomic vector size (number of basis functions) on model performance. As shown in Figure S3, in comparison with an atomic vector size of 30 used in the DTNN model, a better performance can be achieved by increasing the size to 256, which has more flexibility to encode atomic environment. With the atomic vector size of 256, we then explored hyperparameter space of interaction block by grid search, as shown in Table 2 and Figure S4. Hyperparameters we searched include: shared/distinct weights for different interaction blocks (ibs),  $\mu_{max}$ ,  $\Delta\mu$ , number of interaction blocks and activation functions. As shown in Figure S4 (A), using distinct weights for different interaction blocks would reduce the validation error. With the increasing number of interaction blocks, different  $\mu_{max}$  and  $\Delta\mu$  have been tested. DTNN used 20 Å for  $\mu_{max}$ , which would limit model efficiency and scalability for large molecules. Smaller  $\mu_{max}$  would be preferred for the sake of potential computational efficiency but would be expected to sacrifice some accuracy. Thus, the result in Figure S4 (B) is quite surprising: a model with 7 for ib and 3 Å for  $\mu_{max}$  has the lowest validation error among all combinations tested.

Meanwhile, models trained with  $\Delta\mu = 0.1$  Å consistently perform better than corresponding ones with  $\Delta\mu$  as 0.2 Å (Figure S4(C)). Activation function has been changed into shifted softplus ( $\text{ssp}(x) = \ln(0.5e^x + 0.5)$ ) to alleviate the vanishing-gradient problem, and it performs better than using tanh for deeper models, as shown in Figure S4 (D). Although with different architectures, the optimized hyperparameter set for DTNN\_7ib becomes similar to the one in SchNet, including larger number of basis functions in atomic representations (64 in SchNet vs. 256 in DTNN\_7ib), deeper interaction blocks (6 in SchNet vs. 7 in DTNN\_7ib), distinct weights for different interaction blocks, and shifted softplus activation function.

Training data set has been augmented with 15k distant diatomic systems to further improve atomic vector representations. Comparing the atomization energy distribution for heavy atoms of two models trained with and without distant diatomic systems (Figure S5), we found that the energy ranges for N, O, F become much narrower after training data augmentation. Also, the peak position of C has been moved into lower energy region and can be easily distinguished from N and O. This indicates that model trained with distant diatomic systems can perform better in distinguishing different atom types.

By enhancing the training efficiency, searching hyperparameters and incorporating distant diatomic systems into our training set, as presented above, we have developed our improved model DTNN\_7ib. Mean absolute errors (MAE) of QM9 dataset for DTNN\_7ib trained with different training set sizes are summarized in Table 3 and compared with the performance of DTNN and SchNet. The learning curves have also been shown in Figure S6. As show in Table 3 and Figure S6(B), DTNN\_7ib model can achieve consistently better performance than DTNN and equally good performance as SchNet. DTNN\_7ib requires shorter training epochs to achieve 0.34 kcal/mol MAE (around 304 epochs) than original SchNet<sup>56</sup> (750 – 2400 epochs) and the training process can be finished in 13 hours on an NVIDIA Tesla P40

GPU. It should be noted that the training of SchNet has been sped up a lot (12h, NVIDIA Tesla P100 GPU) in SchNetPack<sup>57</sup> by using a decay factor in learning rate reducing process if the validation loss doesn't change in a given number of epochs. Thus, the training efficiency of our DTNN\_7ib can be further improved by applying a better training schedule in the future.

To examine whether final atomic vector representation  $c_i^{(T)}$  learned by DTNN\_7ib is chemically meaningful, we used pairwise cosine distance of atomic vector representations as the atomic similarity measure and divided atoms into different categories with hierarchical clustering for two molecules from QM9 dataset: 4-ethylphenol and 4-(hydroxymethyl)imidazole, which mimic amino acid side chains of tyrosine and histidine respectively. Meanwhile, general AMBER force field (gaff)<sup>71</sup> atom types have been assigned for each atom which accounts for its atomic number, hybridization and bonding environment. From Figure 3, we can see that using embedding from DTNN\_7ib, atoms in both molecules can be hierarchically clustered into different categories consistent with chemical intuition: heavy atoms and hydrogens immediately divided into two different clusters; different types of heavy atoms have been discriminated by DTNN\_7ib and the results are consistent with gaff atom type. It should be noted that this is not a trivial task: we have also trained a DTNN model by using original hyperparameters without distant diatomic molecules and a DTNN\_7ib model without distant diatomic molecules. The results in Figure S7 and Figure S8 indicate that clustering from the resulted embedding of these two models is much less chemically meaningful, which confuses among atoms with different atom types and same atom type in different environments. These results demonstrated that improved DTNN\_7ib model can yield more chemically meaningful atomic vector representation  $c_i^{(T)}$  which reflects both atom type and bonding environment.

## B. Transfer Learning Models: TL\_QM9<sub>M</sub> and TL\_eMol9\_C<sub>M</sub>

Using atomic vector representations in the DTNN\_7ib model, we have employed transfer learning strategy and retrained readout layers for a new machine learning model TL\_QM9<sub>M</sub> with the QM9<sub>M</sub> dataset, which has same molecular energies as in the QM9 dataset but with input geometries from the corresponding MMFF minima. The test results are shown in Table 3, in which we also included performance on different training set sizes. It is not surprising that DTNN\_7ib leads to a significantly large MAE of 4.55 kcal/mol on QM9<sub>M</sub>, which demonstrates the necessity to retrain a model using MMFF optimized geometries. However, for our transfer learning TL\_QM9<sub>M</sub>, its MAE on QM9<sub>M</sub> dataset is 0.79 kcal/mol (100k training references), which is better than 1 kcal/mol (1 kcal/mol is typically considered as chemical accuracy<sup>11</sup>) and even better than the original model DTNN (0.84 kcal/mol with DFT geometry as input). These results indicate that transfer learning strategy with atomic vector representations in the DTNN\_7ib model is a promising direction to develop machine learning models to accurately predict DFT calculated molecular atomization energies using optimized geometries with Merck Molecular Force Field (MMFF).

To identify low-energy conformers for small molecule efficiently is a prerequisite in structure based molecular design, and is typically carried out after the conformation



generation process. Accurately molecular energy prediction for diverse conformations of the same molecule is often more difficult since the difference between conformations could be very small. To examine whether machine learning can be used to build an accurate model for conformation analysis, we created the eMol9\_C<sub>M</sub> dataset and developed TL\_eMol9\_C<sub>M</sub> model using same transfer learning strategy. It should be noted that target molecular energy in eMol9\_C<sub>M</sub> has been calculated by a different basis set from QM9, which introduces additional difficulty.

Table 5 shows performances of our MMFF-based models on eMol9\_C<sub>M</sub> dataset. In terms of absolute error, TL\_eMol9\_C<sub>M</sub> can achieve 0.51 kcal/mol MAE, which is close to the performance of our DFT-based model DTNN\_7ib on QM9 dataset. On the other hand, MAE and RMSE of TL\_QM9<sub>M</sub> are much worse than TL\_eMol9\_C<sub>M</sub> model because of the different basis set in energy calculation, which indicates the necessity to retrain the model to adapt to the change of energy calculation method. Meanwhile, in terms of relative error, we can compare the performance of MMFF94, TL\_QM9<sub>M</sub> and TL\_eMol9\_C<sub>M</sub> together, and our results indicate that both TL\_QM9<sub>M</sub> and TL\_eMol9\_C<sub>M</sub> can achieve much better performance than MMFF method on MAE, RMSE and success rate, which indicates that it is promising to apply MMFF-based machine learning models to search low-energy conformers and improve conformation stability prediction.

Additionally, to consider the impact of difference between MMFF optimized geometries and DFT optimized geometries on prediction accuracy, the performances on test sets with different RMSD cutoffs have been calculated. As shown in Figure 4 and Figure S9, the performance of TL\_eMol9\_C<sub>M</sub> is quite robust given different RMSD cutoffs while MMFF94 method's performances deteriorate significantly when RMSD cutoff increases.

Besides good performance, to develop MMFF-based models based on transfer learning needs much less training time (TL\_QM9<sub>M</sub>: 5h, TL\_eMol9\_C<sub>M</sub>: 6h).

In order to further investigate the applicability and limitation of the TL\_eMol9\_C<sub>M</sub> model, we have built a Plati\_C<sub>M</sub> dataset, which includes molecules with more heavy atoms (10–12) than molecules in our training set (9). Our results show that the overall performance of TL\_eMol9\_C<sub>M</sub> become worse (relative MAE from 0.55 to 1.31 kcal/mol), but it still performs better than MMFF94 (Table 5). To analyze origin of significant performance decreasing of TL\_eMol9\_C<sub>M</sub>, we checked the average error of each molecule and grouped the results based on the number of heavy atoms. Error calculations with consideration of RMSD cutoff have also been conducted (Figure S11). Figure 5 shows the error distribution, and the peaks of two MAE distributions moves from small error region into large error region when the number of heavy atoms increases. It should be noted that there are two molecules with 10 heavy atoms having very large error (> 20 kcal/mol, Figure S10). After computing the cosine similarities between atomic vectors of these two molecules and atomic vectors from our training set, the nitrogen atoms in N(sp<sup>3</sup>)-N(sp<sup>3</sup>) group in these two molecules have been identified with lowest similarities. With the help of HasSubstructMatch command in RDKit, we checked the SMILES of whole QM9 dataset, and no similar N-N group has been found. Therefore, our results indicate that the main limitation of the TL\_eMol9\_C<sub>M</sub> model comes from its training data, which needs to be much larger and more

diverse for the development of more robust machine learning models in predicting molecular energies and conformation stabilities.

## V. CONCLUSION

In this work, we demonstrated importance of training efficiency improvement, hyperparameter search and data augmentation in the optimization of machine learning models, and improved DTNN performance from 0.84 kcal/mol to 0.34 kcal/mol on MAE with no change of model architecture. Our newly developed model DTNN\_7ib has deeper learning blocks and can generate chemical meaningfully atomic vectors which reflects both atom type and atomic environment. To address the application limitation caused by DFT optimized geometries, three datasets with both DFT calculated properties and MMFF optimized geometries have been created and two MMFF-based models have been developed by implementing transfer learning strategy on atomic vectors learned from DTNN\_7ib. TL\_QM9<sub>M</sub> can achieve better than chemical accuracy performance on QM9<sub>M</sub> dataset, and TL\_eMol9\_C<sub>M</sub> can achieve 0.51 kcal/mol MAE on eMol9\_C<sub>M</sub> dataset. In addition, our work indicates that although the overall presented strategy looks promising, one key challenge is the data set, which needs significantly larger and diverse training set for the development of more robust machine learning models for molecular conformation analysis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We would like to acknowledge the support by NIH (R35-GM127040) and computing resources provided by NYU-ITS.

## REFERENCES

1. Perola E; Charifson PS, Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem* 2004, 47, 2499–2510. [PubMed: 15115393]
2. Kuhn B; Guba W; Hert J; Banner D; Bissantz C; Ceccarelli S; Haap W; Korner M; Kuglstatte A; Lerner C; Mattei P; Neidhart W; Pinard E; Rudolph MG; Schulz-Gasch T; Wokering T; Stahl M, A Real-World Perspective on Molecular Design. *J. Med. Chem* 2016, 59, 4087–4102. [PubMed: 26878596]
3. Hawkins PCD, Conformation Generation: The State of the Art. *J. Chem. Inf. Model* 2017, 57, 1747–1756. [PubMed: 28682617]
4. Rossi M; Chutia S; Scheffler M; Blum V, Validation Challenge of Density-Functional Theory for Peptides-Example of Ac-Phe-Ala(5)-LysH. *J. Phys. Chem. A* 2014, 118, 7349–7359. [PubMed: 24405171]
5. Halgren TA, Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem* 1996, 17, 490–519.
6. Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; MacKerell AD, CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem* 2010, 31, 671–690. [PubMed: 19575467]

7. Rappe AK; Casewit CJ; Colwell KS; Goddard WA; Skiff WM, Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *J. Am. Chem. Soc* 1992, 114, 10024–10035.
8. Rupp M; Tkatchenko A; Muller KR; von Lilienfeld OA, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett* 2012, 108, 058301. [PubMed: 22400967]
9. Goh GB; Hodas NO; Vishnu A, Deep learning for computational chemistry. *J. Comput. Chem* 2017, 38, 1291–1307. [PubMed: 28272810]
10. Schutt KT; Arbabzadah F; Chmiela S; Muller KR; Tkatchenko A, Quantum-chemical insights from deep tensor neural networks. *Nat. Commun* 2017, 8, 13890. [PubMed: 28067221]
11. Faber FA; Hutchison L; Huang B; Gilmer J; Schoenholz SS; Dahl GE; Vinyals O; Kearnes S; Riley PF; von Lilienfeld OA, Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput* 2017, 13, 5255–5264. [PubMed: 28926232]
12. Montavon G; Rupp M; Gobre V; Vazquez-Mayagoitia A; Hansen K; Tkatchenko A; Muller KR; von Lilienfeld OA, Machine learning of molecular electronic properties in chemical compound space. *New J. Phys* 2013, 15, 095003.
13. Hansen K; Montavon G; Biegler F; Fazli S; Rupp M; Scheffler M; von Lilienfeld OA; Tkatchenko A; Muller KR, Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput* 2013, 9, 3404–3419. [PubMed: 26584096]
14. Schutt KT; Glawe H; Brockherde F; Sanna A; Muller KR; Gross EKV, How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B* 2014, 89, 205118.
15. Faber F; Lindmaa A; von Lilienfeld OA; Armiento R, Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem* 2015, 115, 1094–1101.
16. Ramakrishnan R; Dral PO; Rupp M; von Lilienfeld OA, Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach. *J. Chem. Theory Comput* 2015, 11, 2087–2096. [PubMed: 26574412]
17. Hansen K; Biegler F; Ramakrishnan R; Pronobis W; von Lilienfeld OA; Muller KR; Tkatchenko A, Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett* 2015, 6, 2326–2331. [PubMed: 26113956]
18. Faber FA; Lindmaa A; von Lilienfeld OA; Armiento R, Machine Learning Energies of 2 Million Elpasolite (ABC(2)D(6)) Crystals. *Phys. Rev. Lett* 2016, 117, 135502. [PubMed: 27715098]
19. Hirn M; Mallat S; Poilvert N, Wavelet Scattering Regression of Quantum Chemical Energies. *Multiscale Model. Simul* 2017, 15, 827–863.
20. Huo HY; Rupp M, Unified Representation of Molecules and Crystals for Machine Learning 2018, arXiv:1704.06439v3.
21. Ericksen M; Exarchakis G; Hirn M; Mallat S, Solid Harmonic Wavelet Scattering: Predicting Quantum Molecular Energy from Invariant Descriptors of 3D Electronic Densities. *Adv. Neural Inf. Process. Syst* 2017, 6543–6552.
22. Ryczko K; Mills K; Luchak I; Homenick C; Tamblyn I, Convolutional neural networks for atomistic systems. *Comput. Mater. Sci* 2018, 149, 134–142.
23. Mills K; Ryczko K; Luchak I; Domurad A; Beeler C; Tamblyn I, Extensive deep neural networks 2018, arXiv:1708.06686v2.
24. Behler J; Parrinello M, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett* 2007, 98, 146401. [PubMed: 17501293]
25. Behler J, Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys* 2011, 134, 074106. [PubMed: 21341827]
26. Bartok AP; Payne MC; Kondor R; Csanyi G, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett* 2010, 104, 136403. [PubMed: 20481899]
27. Bartok AP; Kondor R; Csanyi G, On representing chemical environments. *Phys. Rev. B* 2013, 87, 184115.
28. Shapeev AV, Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul* 2016, 14, 1153–1173.

29. Chmiela S; Tkatchenko A; Sauceda HE; Poltavsky I; Schutt KT; Muller KR, Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv* 2017, 3, No. e1603015. [PubMed: 28508076]
30. Brockherde F; Vogt L; Li L; Tuckerman ME; Burke K; Muller KR, Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun* 2017, 8, 872. [PubMed: 29021555]
31. Smith JS; Isayev O; Roitberg AE, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci* 2017, 8, 3192–3203. [PubMed: 28507695]
32. Podryabinkin EV; Shapeev AV, Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci* 2017, 140, 171–180.
33. Rowe P; Csanyi G; Alfe D; Michaelides A, Development of a machine learning potential for graphene. *Phys. Rev. B* 2018, 97, 054303.
34. Sinititskiy AV; Pande VS, Deep Neural Network Computes Electron Densities and Energies of a Large Set of Organic Molecules Faster than Density Functional Theory(DFT) 2018, arXiv: 1809.02723v1.
35. Zhang LF; Han JQ; Wang H; Car R; E, W., Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett* 2018, 120, 143001. [PubMed: 29694129]
36. Pronobis W; Schutt KT; Tkatchenko A; Muller KR, Capturing intensive and extensive DFT/TDDFT molecular properties with machine learning. *Eur. Phys. J. B* 2018, 91, 178.
37. Lubbers N; Smith JS; Barros K, Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys* 2018, 148, 241715. [PubMed: 29960311]
38. Pronobis W; Tkatchenko A; Muller KR, Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules. *J. Chem. Theory Comput* 2018, 14, 2991–3003. [PubMed: 29750522]
39. Jørgensen PB; Jacobsen KW; Schmidt MN, Neural Message Passing with Edge Updates for Predicting Properties of Molecules and Materials 2018, arXiv:1806.03146.
40. Wang RB, Significantly Improving the Prediction of Molecular Atomization Energies by an Ensemble of Machine Learning Algorithms and Rescanning Input Space: A Stacked Generalization Approach. *J. Phys. Chem. C* 2018, 122, 8868–8873.
41. Wu ZQ; Ramsundar B; Feinberg EN; Gomes J; Geniesse C; Pappu AS; Leswing K; Pande V, MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci* 2018, 9, 513–530. [PubMed: 29629118]
42. Butler KT; Davies DW; Cartwright H; Isayev O; Walsh A, Machine learning for molecular and materials science. *Nature* 2018, 559, 547–555. [PubMed: 30046072]
43. Zhang LF; Han JQ; Wang H; Saidi WA; Car R; E, W., End-to-end Symmetry Preserving Interatomic Potential Energy Model for Finite and Extended Systems 2018, arXiv:1805.09003.
44. Ferre G; Haut T; Barros K, Learning molecular energies using localized graph kernels. *J. Chem. Phys* 2017, 146, 114107. [PubMed: 28330348]
45. Han JQ; Zhang LF; Car R; E, W., Deep Potential: a general representation of a many-body potential energy surface 2017, arXiv:1707.01478.
46. Yao K; Herr JE; Toth DW; Mckintyre R; Parkhill J, The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci* 2018, 9, 2261–2269. [PubMed: 29719699]
47. Unke OT; Meuwly M, A reactive, scalable, and transferable model for molecular energies from a neural network approach based on local information. *J. Chem. Phys* 2018, 148, 241708. [PubMed: 29960298]
48. Gilmer J; Schoenholz SS; Riley PF; Vinyals O; Dahl GE, Neural Message Passing for Quantum Chemistry 2017, arXiv:1704.01212.
49. Smith JS; Nebgen B; Lubbers N; Isayev O; Roitberg AE, Less is more: Sampling chemical space with active learning. *J. Chem. Phys* 2018, 148, 241733. [PubMed: 29960353]
50. Tsubaki M; Mizoguchi T, Fast and Accurate Molecular Property Prediction: Learning Atomic Interactions and Potentials with Neural Networks. *J. Phys. Chem. Lett* 2018, 9, 5733–5741. [PubMed: 30081630]

51. Faber FA; Christensen AS; Huang B; von Lilienfeld OA, Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys* 2018, 148, 241717. [PubMed: 29960351]
52. Chmiela S; Sauceda HE; Muller KR; Tkatchenko A, Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun* 2018, 9, 3887. [PubMed: 30250077]
53. Ramakrishnan R; Dral PO; Rupp M; von Lilienfeld OA, Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* 2014, 1, 140022. [PubMed: 25977779]
54. Blum LC; Reymond JL, 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc* 2009, 131, 8732–8733. [PubMed: 19505099]
55. Reymond JL, The Chemical Space Project. *Acc. Chem. Res* 2015, 48, 722–730. [PubMed: 25687211]
56. Schutt KT; Sauceda HE; Kindermans PJ; Tkatchenko A; Muller KR, SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys* 2018, 148, 241722. [PubMed: 29960322]
57. Schutt KT; Kessel P; Gastegger M; Nicoli KA; Tkatchenko A; Muller KR, SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput* 2019, 15, 448–455. [PubMed: 30481453]
58. Pratt LY, Discriminability-based transfer between neural networks. *Adv. Neural Inf. Process. Syst* 1993, 204–211.
59. Pratt LY; Thrun S; Thomas GD, Special issue on inductive transfer. *Mach Learn* 1997, 28.
60. The RDKit: Open-Source Cheminformatics Software <http://www.rdkit.org/> (accessed Mar 2, 2018).
61. eMolecules. <https://www.emolecules.com/>.
62. Riniker S; Landrum GA, Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model* 2015, 55, 2562–2574. [PubMed: 26575315]
63. Butina D, Unsupervised data base clustering based on Daylight’s fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci* 1999, 39, 747–750.
64. Temelso B; Mabey JM; Kubota T; Appiah-Padi N; Shields GC, ArbAlign: A Tool for Optimal Alignment of Arbitrarily Ordered Isomers Using the Kuhn-Munkres Algorithm. *J. Chem. Inf. Model* 2017, 57, 1045–1054. [PubMed: 28398732]
65. Frisch MJT, G. W.; Schlegel HB; Scuseria GE; Robb MA; Cheeseman JR; Scalmani G; Barone V; Mennucci B; Petersson GA; Nakatsuji H; Caricato M; Li X; Hratchian HP; Izmaylov AF; Bloino J; Zheng G; Sonnenberg JL; Hada M; Ehara M; Toyota K; Fukuda R; Hasegawa J; Ishida M; Nakajima T; Honda Y; Kitao O; Nakai H; Vreven T; Montgomery JA Jr.; Peralta JE; Ogliaro F; Bearpark M; Heyd JJ; Brothers E; Kudin KN; Staroverov VN; Kobayashi R; Normand J; Raghavachari K; Rendell A; Burant JC; Iyengar SS; Tomasi J; Cossi M; Rega N; Millam JM; Klene M; Knox JE; Cross JB; Bakken V; Adamo C; Jaramillo J; Gomperts R; Stratmann RE; Yazyev O; Austin AJ; Cammi R; Pomelli C; Ochterski JW; Martin RL; Morokuma K; Zakrzewski VG; Voth GA; Salvador P; Dannenberg JJ; Dapprich S; Daniels AD; Farkas O; Foresman JB; Ortiz JV; Cioslowski J; Fox DJ Gaussian 09, Gaussian Inc.: Wallingford, CT, 2009.
66. Friedrich NO; Meyder A; Kops CD; Sommer K; Flachsenberg F; Rarey M; Kirchmair J, High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators. *J. Chem. Inf. Model* 2017, 57, 529–539. [PubMed: 28206754]
67. Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE, The Protein Data Bank. *Nucleic Acids Res* 2000, 28, 235–242. [PubMed: 10592235]
68. Abadi MA, A.; Barham P; Brevdo E; Chen Z; Citro C; Corrado GS; 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; Mane D; Monga R; Moore S; Murray D; Olah C; Schuster M; Shlens J; Steiner B; Sutskever IT, K.; Tucker P; Vanhoucke V; Vasudevan V; Viegas F; Vinyals O; Warden P; Wattenberg M; Wicke M; Yu Y; Zheng X TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 2015, available via the Internet at: <https://www.tensorflow.org/>.
69. Huang G; Li YX; Pleiss G; Liu Z; Hopcroft JE; Weinberger KQ, Snapshot ensembles: Train 1, get M for free. *ICLR* 2017.

70. Schutt KT; Kindermans PJ; Saucedo HE; Chmiela S; Tkatchenko A; Muller KR, SchNet: A continuous-filter convolutional neural network for modeling quantum interactions 2017, arXiv: 1706.08566.
71. Wang JM; Wolf RM; Caldwell JW; Kollman PA; Case DA, Development and testing of a general amber force field. *J. Comput. Chem* 2004, 25, 1157–1174. [PubMed: 15116359]

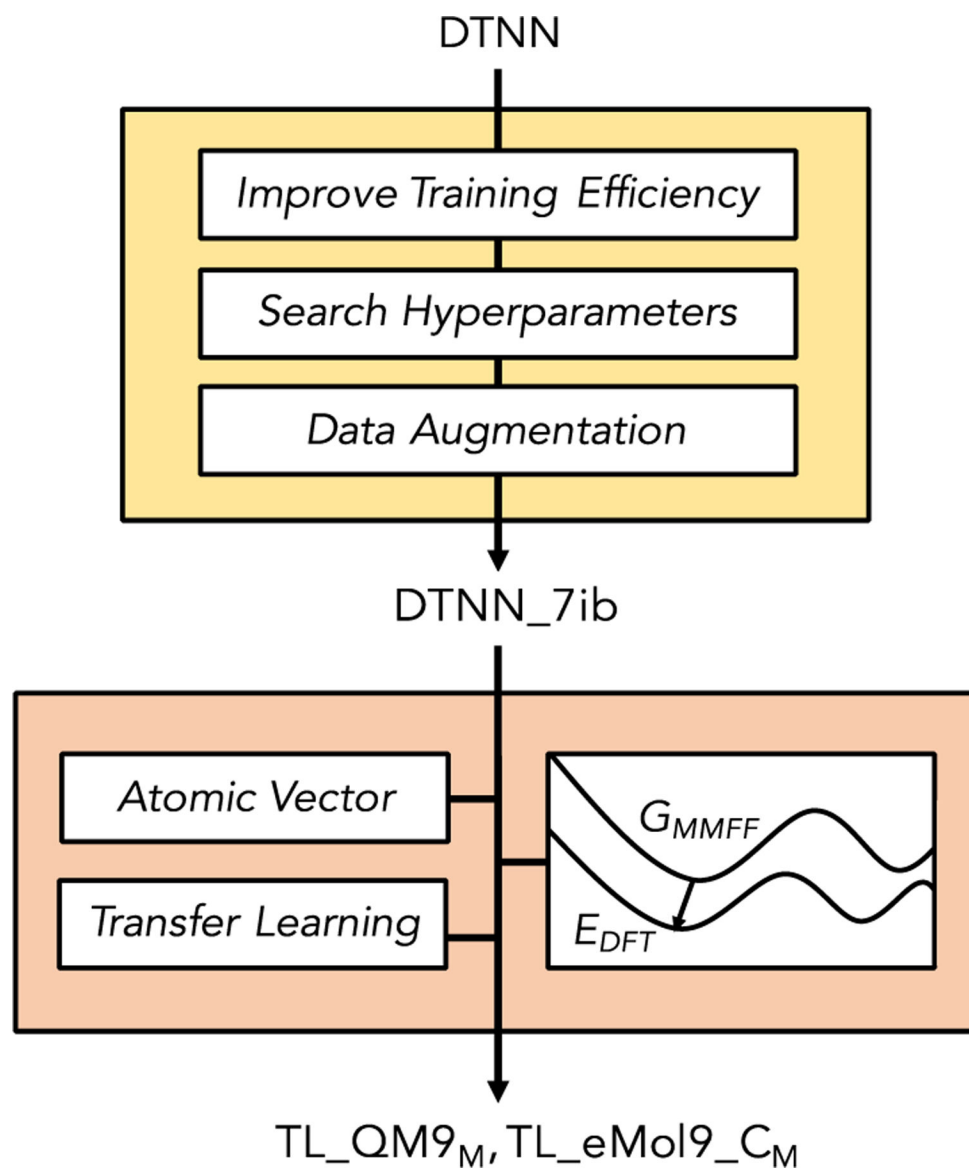
Author Manuscript

Author Manuscript

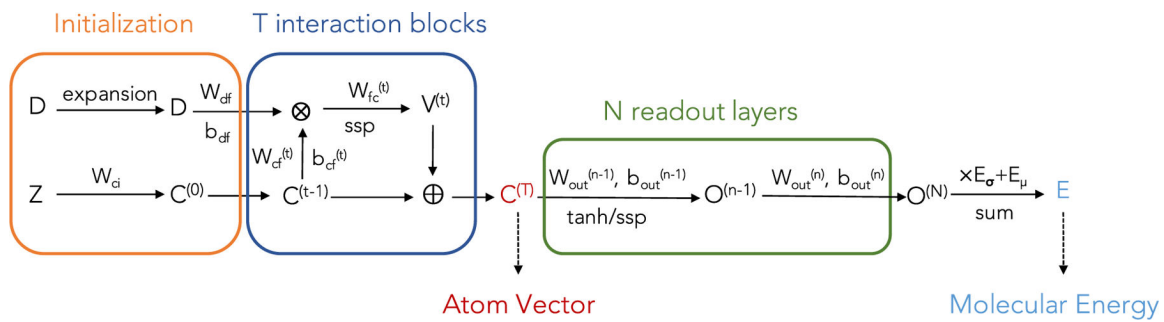
Author Manuscript

Author Manuscript

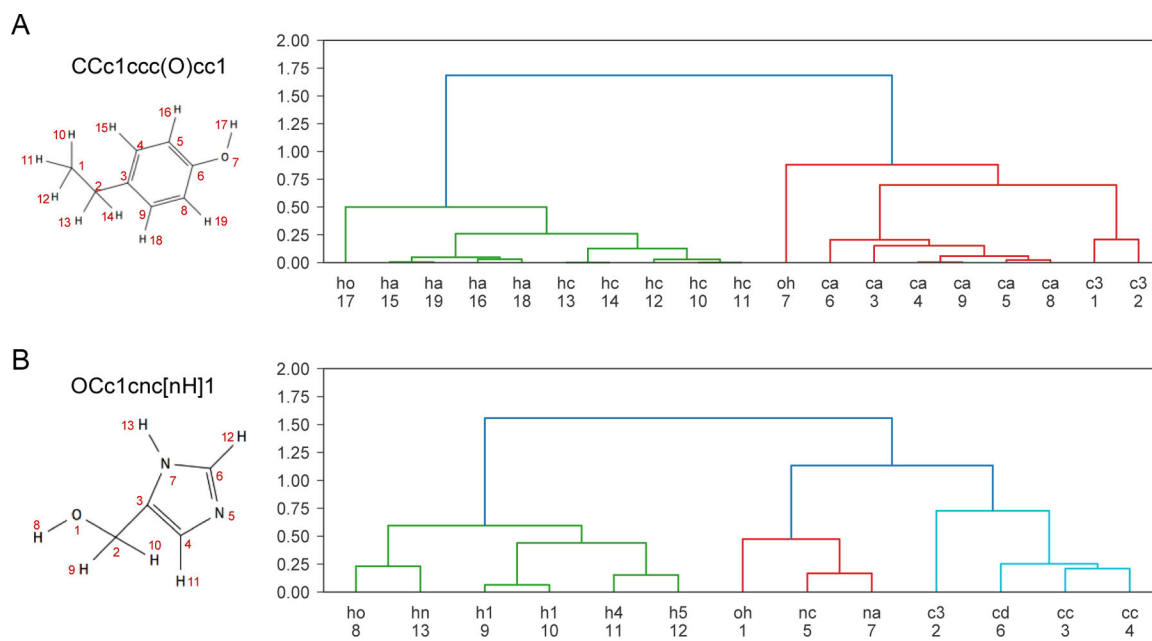




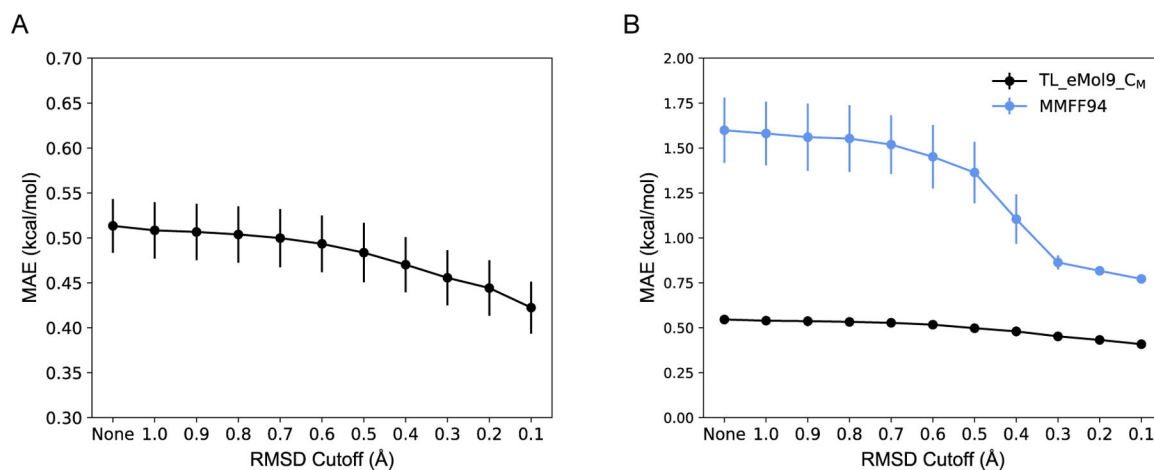
**Figure 1.** Overall workflow of exploring how machine learning can be used to predict molecular atomization energies and conformation stability using force-field optimized geometries.



**Figure 2.**  
Illustration of the model architecture.

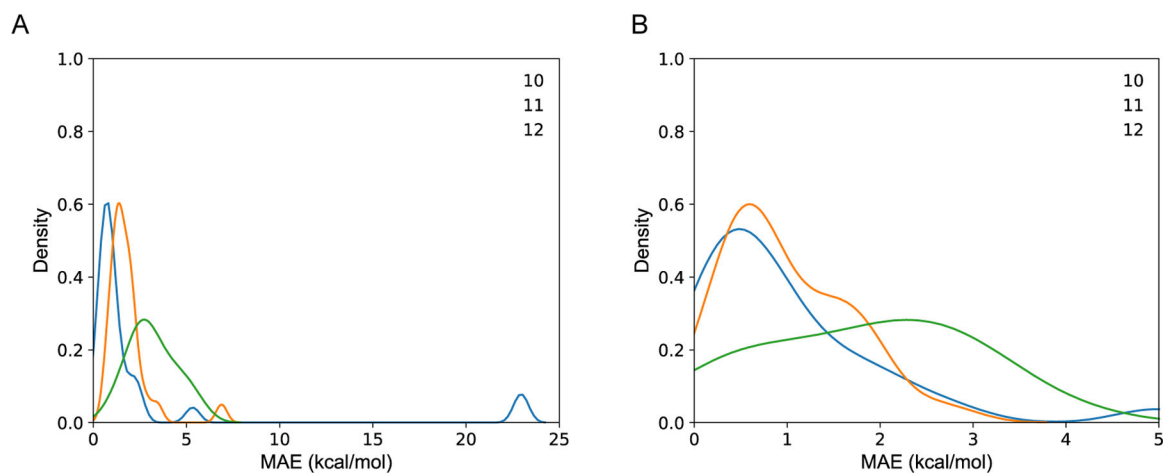
**Figure 3.**

Atomic vector interpretation. (A) 4-ethylphenol and (B) 4-(hydroxymethyl)imidazole. Left are molecular graphs with SMILES. Right are hierarchical clustering results. The first row of x-axis is gaff atom type and the second row of x-axis is atom indexes numbers which are same as in molecular graph.



**Figure 4.**

(A) Absolute MAE and (B) relative MAE on test sets with different RMSD cutoffs. “None” means the performance is for total test set, “value” means the performance is for subset, which only includes conformations with RMSD less than “value”. Average and standard deviation of performances from five models trained with random split data have been shown. Performances of TL\_eMol9\_CM, MMFF94 are colored as black and blue, respectively.



**Figure 5.** (A) Absolute MAE and (B) relative MAE distributions for molecules with different number of heavy atoms. Distribution for molecules with 10, 11, 12 heavy atoms is colored as blue, orange, and green, respectively.

**Table 1.**

Datasets<sup>a</sup> used for machine learning model development and evaluation.

Dataset	Geometry	Energy	#Molecule <sup>b</sup>	#Conformations <sup>b</sup>	#Heavy Atoms
QM9	B3LYP/6-31G(2df,p)	B3LYP/6-31G(2df,p)	99,000/1,000/33,885	99,000/1,000/33,885	[1, 9]
QM9 <sub>M</sub>	MMFF94	B3LYP/6-31G(2df,p)	99,000/1,000/33,885	99,000/1,000/33,885	[1, 9]
eMol9_C <sub>M</sub>	MMFF94	B3LYP/6-31G*	8111/500/1,348	~66,000/~6,000/~16,000 <sup>c</sup>	[1, 9]
Plati_C <sub>M</sub>	MMFF94	B3LYP/6-31G*	0/0/74	0/0/4,076	[10, 12]

<sup>a</sup>QM9 dataset is generated by Ramakrishnan et al<sup>53</sup>. Other three datasets are prepared by ourselves.

<sup>b</sup>Number of molecules and conformations in the training/validation/test sets are shown respectively.

<sup>c</sup>eMol9\_C<sub>M</sub> has been random split into train/validation/test sets using five different random seeds based on molecule types. Thus, the numbers of conformations for train/validation/test sets in different splits are different.



**Table 2.**

Hyperparameters searched for interaction blocks during the model optimization.

Hyperparameters	Test Values <sup>a</sup>
$\mu_{max}$	3, 4, 5, 6, 8, 10, <u>20</u>
$\Delta\mu$	<b>0.1</b> , <u>0.2</u>
number of interaction blocks	2, <u>3</u> , 4, 5, 6, 7, 8
share/distinct weights	<u>shared</u> , <b>distinct</b>
activation function	<u>tanh</u> , <b>shifted softplus</b>

<sup>a</sup>Hyperparameters used in DTNN\_7ib are in bold, and hyperparameters used in the original DTNN model are underlined.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

QM9 performance (MAE in kcal/mol) for various models with different training set sizes.

# Training Data	DTNN_7ib <sup>a</sup>	TL_QM9 <sub>M</sub> <sup>b</sup>	DTNN	SchNet
25k	0.69 ± 0.02	1.49 ± 0.03	1.04 ± 0.02	-
50k	0.46 ± .02	1.01 ± 0.03	0.94 ± 0.02	0.59
100k	0.34 ± 0.01	0.79 ± 0.03	0.84 ± 0.01	0.34 <sup>c</sup>

<sup>a</sup>MAE is the average of five models trained with data split by different random seed.

<sup>b</sup>TL\_QM9<sub>M</sub> performance is on QM9<sub>M</sub> dataset. The input coordinates are from MMFF optimized geometries.

<sup>c</sup>SchNet performance is the one trained with 100k molecules<sup>70</sup>. In recently published paper<sup>56–57</sup>, SchNet performance on QM9 with 110k training set is 0.31 kcal/mol and 0.26 kcal/mol, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4.**Performance on eMol9\_C<sub>M</sub> dataset<sup>a</sup>

		MMFF94	TL_QM9 <sub>M</sub>	TL_eMol9_C <sub>M</sub>
Error <sub>A</sub> <sup>b</sup>	MAE	None	18.30 ± 0.08	<b>0.51 ± 0.03</b>
	RMSE	None	18.66 ± 0.10	<b>1.27 ± 0.28</b>
Error <sub>R</sub> <sup>b</sup>	MAE	1.60 ± 0.18	1.04 ± 0.18	<b>0.55 ± 0.02</b>
	RMSE	2.70 ± 0.32	1.35 ± 0.18	<b>0.72 ± 0.04</b>
Success Rate		0.59 ± 0.01	0.62 ± 0.00	<b>0.72 ± 0.01</b>

<sup>a</sup>Best performances are in bold.<sup>b</sup>Unit is kcal/mol.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5.**Performance on Plati\_C<sub>M</sub> dataset<sup>a</sup>

		MMFF94 <sup>c</sup>	TL_eMol9_C <sub>M</sub>
Error <sub>A</sub> <sup>b</sup>	MAE	None	<b>2.73 ± 0.35</b>
	RMSE	None	<b>4.74 ± 0.39</b>
Error <sub>R</sub> <sup>b</sup>	MAE	2.65	<b>1.31 ± 0.09</b>
	RMSE	4.40	<b>1.59 ± 0.09</b>
Success Rate		0.49	<b>0.57 ± 0.05</b>

<sup>a</sup>Best performances are in bold.<sup>b</sup>Unit is kcal/mol.<sup>c</sup>There is no standard deviation in MMFF94 method since we used all conformations in Plati\_C<sub>M</sub> as test set, but for TL\_eMol9\_C<sub>M</sub> we showed performances on five models.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript