ROYAL SOCIETY
OF CHEMISTRY

PAPER

Check for updates

A neural network potential with rigorous treatment of long-range dispersion†

Nguyen Thien Phuc Tu, ID *a Nazanin Rezajooei, ID c Erin R. Johnson ID *b and Christopher N. Rowley ID *a

Neural Network Potentials (NNPs) have quickly emerged as powerful computational methods for modeling large chemical systems with the accuracy of quantum mechanical methods but at a much smaller computational cost. To make the training and evaluation of the underlying neural networks practical, these methods commonly cut off interatomic interactions at a modest range (*e.g.*, 5.2 Å), so longer-range interactions like London dispersion are neglected. This limits the accuracy of these models for intermolecular interactions. In this work, we develop a new NNP designed for modeling chemical systems where dispersion is an essential component. This new NNP is extended to treat dispersion interactions rigorously by calculating atomic dispersion coefficients through a second set of NNs, which is trained to reproduce the coefficients from the quantum-mechanically derived exchange-hole dipole moment (XDM) model. The NNP with this dispersion correction predicts intermolecular interactions in very good agreement with the QM data, with a mean absolute error (MAE) of 0.67 kcal mol^{-1} and a coefficient of determination (R^2) of 0.97. The dispersion components of these intermolecular interactions are predicted in excellent agreement with the QM data, with a mean absolute error (MAE) of 0.01 kcal mol^{-1} and an R^2 of 1.00. This combined dispersion-corrected NNP, called ANIPBE0-MLXDM, predicts intermolecular interaction energies for complexes from the DES370K test set with an MAE of 0.69 kcal mol^{-1} and an R^2 of 0.97 relative to high-level *ab initio* results (CCSD(T)), but with a computational cost that is billions of times smaller. The ANIPBE0-MLXDM method is effective for simulating large-scale dispersion-driven systems, such as molecular liquids and gas adsorption in porous materials, on a single computer workstation.

1 Introduction

Computer simulation provides an invaluable tool in understanding and designing new complex chemical systems. Such simulations require accurate models of the intramolecular and intermolecular forces that determine the stability of a chemical structure. Neural networks provide a revolutionary strategy to calculate the potential energy of chemical systems. Several successful neural network potentials (NNPs) have been reported.[1–8] These NNPs allow the potential energy of a chemical structure to be calculated with comparable accuracy to the quantum mechanical methods they were trained to reproduce.[9,10] Notably, the Accurate Neural network engIne (ANI)

aCarleton University, 1125 Colonel By Dr, Ottawa, ON, Canada. E-mail: christopherrowley@cunet.carleton.ca; Tel: +1 613-520-2600 ext. 1647

bDepartment of Chemistry, Dalhousie University, 6274 Coburg Rd, Halifax, B3H 4R2, Nova Scotia, Canada. E-mail: erin.johnson@dal.ca

cInterdisciplinary Program in Scientific Computing, Memorial University of Newfoundland, 230 Elizabeth Ave, St. John's, A1C 5S7, Newfoundland and Labrador, Canada

† Electronic supplementary information (ESI) available: The details of the computational methods and supplementary figures. See DOI: https://doi.org/10.1039/d2dd00150k

NNPs are efficient, general-purpose NNPs that provide molecular reaction and conformational energies with a similar level of accuracy as first-principles quantum mechanical models, but at a computational cost that can be billions of times smaller.[11–13] These NNPs can be used to model chemical systems of arbitrary composition without retraining, allowing them to be used as general-purpose tools.[14,15]

Although existing NNPs are very effective for describing intramolecular interactions,[15] they only include the effect of interatomic interactions within a short cutoff radius (*e.g.*, 5.2 Å in ANI-1× NNP) due to the construction of the atomic environment variables.[16–18] Long-range interactions outside of this range, namely electrostatic and dispersion, are neglected and act as a source of noise for short-range NNPs.[18–20]

Several NNPs have been developed that have descriptions of long range electrostatic interactions. Morawietz *et al.*[16,21] used artificial neural networks to predict the point charge of the system (*i.e.*, 3rd generation NNP). The scheme showed good results when training with large biological systems.[22] Yao *et al.* used NNs to assign partial atomic charges and used these charges to calculate a long-range electrostatic energy term using a damped coulombic potential.[23] Applying the charge equilibration scheme from Faraji *et al.*,[24] Ko *et al.*[25,26] introduced a 4th generation NNP that includes

long-range electrostatic interactions. Metcalf *et al.*[27] showed that a message passing neural network was effective for assigning partial atomic charges consistently even in large molecules. A recent review by Anstine and Isayev provides a more comprehensive summary of these methods.[18]

London dispersion is another significant long-range intermolecular interaction. These interactions are usually neglected by NNPs or have been included through a Grimme D3 correction.[3,6] Recently, NNs were used to calculate the dispersion coefficients in a condensed-phase simulation of C_{60}.[28] Piquemal and coworkers used NNs to add a dispersion correction to density-functional theory (DFT) calculations,[29,30] suggesting that NNs could be an effective means to account for dispersion interaction in NNPs more generally.

The dispersion interaction between a pair of atoms can be approximated as a multipole expansion of the oscillating electric moments between atomic pairs,[31] yielding

$$\nu_{\text{disp},ij}(R_{ij}) = -\frac{C_{6,ij}}{R_{ij}^{6}} - \frac{C_{8,ij}}{R_{ij}^{8}} - \frac{C_{10,ij}}{R_{ij}^{10}}\cdots, \tag{1}$$

where R_{ij} is the distance between the atomic pair i and j, and the series expansion is usually truncated after the C_8 or C_{10} term.

Calculation of the dispersion energy using this method requires the determination of the C_6, C_8, and C_{10} dispersion coefficients for each pair. The exchange-hole dipole moment (XDM) method[32,33] provides a physically-motivated method for calculating these dispersion coefficients. Paired with a damping function, this model has been successfully used to correct density functionals that neglect dispersion interactions.[34] In the XDM model, the coefficients are related to atomic polarizabilities (α) and expectation values of the squares of the atomic exchange-hole multipole moments, $\langle M_\ell^2 \rangle$. The C_6 dispersion coefficient for the pair of atoms i and j is

$$C_{6,ij} = \alpha_i \alpha_j \frac{\langle M_1^2 \rangle_i \langle M_1^2 \rangle_j}{\alpha_i \langle M_1^2 \rangle_j + \alpha_j \langle M_1^2 \rangle_i}, \tag{2}$$

where $\langle M_1^2 \rangle_i$ is the expectation value of the square of the exchange-hole dipole moment for atom i, and α_i is the atom-in-molecule polarizability. This is approximated by scaling the free-atom polarizability by a ratio of the atom-in-molecule and free-atom Hirshfeld volumes:

$$\alpha_i = \frac{V_i}{V_{i,\text{free}}} \alpha_{i,\text{free}}. \tag{3}$$

Similar relations provide the C_8 and C_{10} dispersion coefficients:

$$C_{8,ij} = \frac{3}{2} \alpha_i \alpha_j \frac{\langle M_1^2 \rangle_i \langle M_2^2 \rangle_j + \langle M_2^2 \rangle_i \langle M_1^2 \rangle_j}{\alpha_j \langle M_1^2 \rangle_i + \alpha_i \langle M_1^2 \rangle_j} \tag{4}$$

$$C_{10,ij} = \alpha_i \alpha_j \frac{2\langle M_1^2 \rangle_i \langle M_3^2 \rangle_j + 2\langle M_3^2 \rangle_i \langle M_1^2 \rangle_j + \frac{21}{5}\langle M_2^2 \rangle_i \langle M_2^2 \rangle_j}{\alpha_j \langle M_1^2 \rangle_i + \alpha_i \langle M_1^2 \rangle_j} \tag{5}$$

where $\langle M_2^2 \rangle_i$ and $\langle M_3^2 \rangle_i$ are expectation values involving the square of the exchange-hole quadrupole and octupole moments, respectively. The dispersion coefficients differ from free-atom values depending on the local chemical environment because of their dependency on the electron density and its derivatives *via* the exchange hole.[35,36]

Substituting eqn (2), (4) and (5) into eqn (1) and performing a sum over all atom pairs gives

$$\nu_{\text{disp},ij}(R_{ij}) = -\sum_i^N \sum_{j>i} \left(\frac{C_{6,ij}}{R_{ij}^{6} + R_{\text{vdW},ij}^{6}} + \frac{C_{8,ij}}{R_{ij}^{8} + R_{\text{vdW},ij}^{8}} \right. $$
$$\left. + \frac{C_{10,ij}}{R_{ij}^{10} + R_{\text{vdW},ij}^{10}} \right). \tag{6}$$

Here, the van der Waals (vdW) radii, $R_{\text{vdW},ij}$, are introduced to damp the dispersion energy at small interatomic distances. They depend linearly on the critical radii, $R_{\text{critical},ij}$, at which the various terms in the series expansion of the dispersion energy become equal:

$$R_{\text{vdW},ij} = a_1 R_{\text{critical},ij} + a_2. \tag{7}$$

The values of the two constants, a_1 and a_2, are determined empirically for each density functional (and basis set) and kept fixed thereafter. The critical radius is calculated from averaging ratios of the C_6, C_8, and C_{10} coefficients:

$$R_{\text{critical},ij} = \frac{1}{3}\left[\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{\frac{1}{2}} + \left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{\frac{1}{4}} + \left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{\frac{1}{2}} \right]. \tag{8}$$

XDM relies on the electron density distribution (and its derivatives) to determine the dispersion coefficients, so it can only be used after a DFT calculation is performed first. However, if the XDM terms are estimated using a machine-learning approach instead, the XDM dispersion correction (eqn (6)) can be calculated without the DFT calculation. This NN-based dispersion correction can be added to a NNP to account for the long-range dispersion interactions.

In this work, we develop a new NN to calculate the XDM dispersion coefficients that is based on those used in the ANI-type NNPs. These coefficients are then used to add a dispersion correction to a separate NNP that is trained to predict the non-dispersion components of the potential energy. The new NNP is designed for the calculation of intermolecular complexes of organic molecules involving H, C, N, and O atoms. The short-range component of the chemical interactions is calculated from a conventional NNP, while the dispersion terms are evaluated by a second NN. The resulting dispersion correction includes 6th, 8th, and 10th order contributions with physically correct asymptotic long-range behavior. With this approach, the atomic dispersion coefficients are propagated dynamically and adapt to changes in the system, allowing for simulations of arbitrary complex chemical systems using energy minimization, molecular dynamics, and Monte Carlo methods.

2 Theory and computational methods

2.1 Construction of the NNP

To develop a method where dispersion can be evaluated separately from other interactions, a new NNP was trained to reproduce energies and forces of density-functional theory calculations using the PBE0 exchange–correlation functional.[37] This functional completely neglects dispersion interactions, but provides a well-validated and physically-motivated description of other electronic interactions.[38,39] The new ANIPBE0 NNP was constructed in an analogous way to the ANI-1×/2× NNPs developed by Roitberg and coworkers,[11,12] which calculate the total energy of a chemical system as the sum of atomic energies,

$$\nu_{\text{NNP}} = \sum_{i=1}^{N_{\text{atoms}}} \nu_{\text{NNP},i}. \tag{9}$$

This type of NNP encodes the chemical environment of an atom in the form of an Atomic Environment Vector (AEV), which is constructed from a set of modified Behler–Parrinello symmetry functions.[4,40] These AEVs serve as the input features of element-specific NNs with 3 hidden layers that calculate the atomic energies. The query-by-committee method was used,[11,41] where an ensemble of 8 NNPs was trained and their average is used as the final energy for the ensemble.

2.2 Training set

A large training set of potential energies (PBE0 electronic energies) and forces (gradients with respect to changing nuclear positions) was generated using the Gaussian 16 program.[42] The NNP was trained with a loss function that was a combination of the root-mean-square error of the molecular energies and the norm of the gradients. The NNPs were implemented, trained, and executed through the TorchANI library.[43] The test set was composed of structures from the ANI-1× data set, augmented with randomly generated intermolecular complexes selected through an active learning process. 51 active learning cycles were performed, in which the standard deviation of the NNP ensemble was used to select intermolecular configurations where the NNP was under-trained, similar to the procedure developed by Smith *et al.* for unimolecular structures.[11] The final data set contained QM energies, forces, and dispersion coefficients for 2.1 million chemical structures. This data set can be downloaded from FigShare. The full details of the methods are described in ESI.†

2.3 Construction of the dispersion coefficient NNs

Separate NNs were trained to predict the atomic moments, $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and volumes, V, for the elements H, C, N, and O. The training data was calculated from the DFT (PBE0/aug-cc-pVTZ) density using the XDM model for the chemical structures in the NNP dataset. For consistency, the same input features and network architecture were used for the dispersion coefficient NNs as were used for the ANIPBE0 NNP, which allows the modified Behler–Parrinello symmetry function AEVs to be reused as the features to predict these terms.
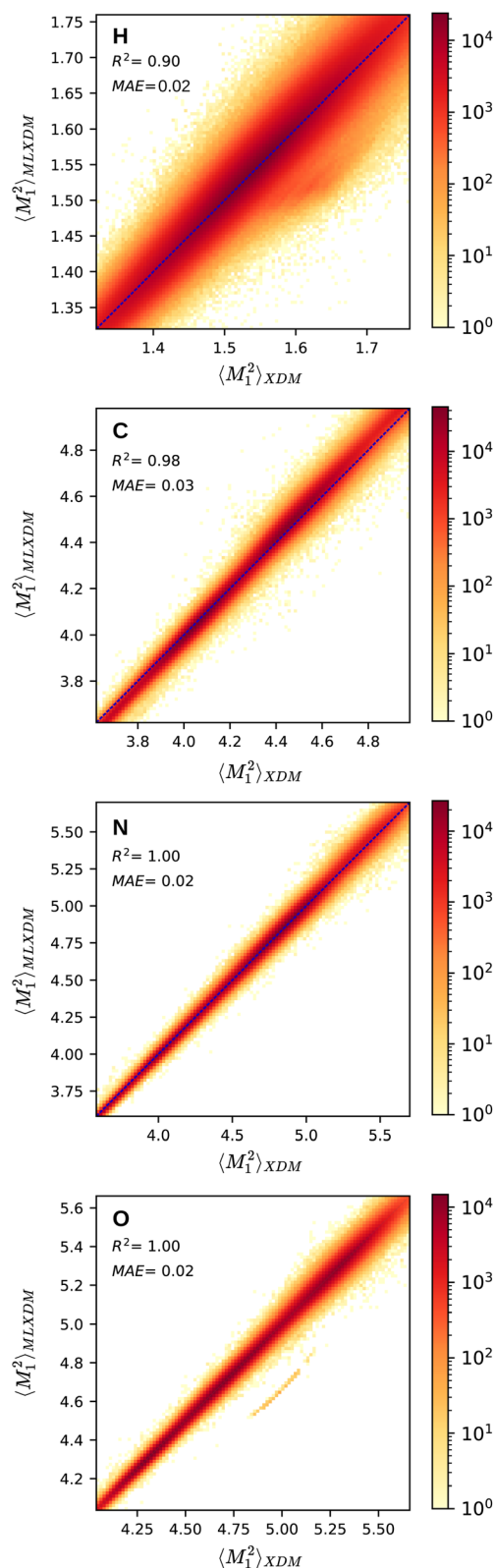

Fig. 1 Prediction of the XDM atomic $\langle M_1^2 \rangle$ moment integrals by MLXDM for the elements H, C, N, and O. The color shows the density of predictions with that correlation. Analogous plots for $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V are included in the ESI.†

The atomic exchange-hole moment integrals, $\langle M_\ell^2 \rangle$, have significantly different ranges depending on the element and order (ℓ). To maximize the performance of the NN, the moments and volumes were standardized such that the NN predicts the deviation of the quantity from its mean value, $\overline{\langle M_l^2 \rangle}$ or $\bar{V}$, scaled by the standard deviation, $\sigma_{\langle M_\ell \rangle^2}$ or σ_V (z-score standardization):

$$\langle M_l^2 \rangle' = \frac{\overline{\langle M_l^2 \rangle} - \langle M_l^2 \rangle}{\sigma_{\langle M_l^2 \rangle}}. \tag{10}$$

Fig. 1 shows the NN prediction of the standardized MLXDM atomic $\langle M_1^2 \rangle$ expectation values compared to the reference XDM results.‡ The coefficients of determination (R^2) and mean absolute errors (MAE) for C, N, O are all excellent.[44] MLXDM also predicts atomic $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V averages very effectively for all four elements. The correlation plots for these terms and the errors as a function of the value are included in the ESI.† MLXDM uses the same damping coefficients as PBE0/aug-cc-pVTZ: $a_1 = 0.4186$ and $a_2 = 2.6791$ Å.[45]

XDM can be used with other DFT functionals if different coefficients in the damping function are used. Likewise, MLXDM could be used with an NNP trained to reproduce a different DFT method provided appropriate damping coefficients were used. The XDM coefficients are only moderately dependent on the DFT functional used to calculate them, so the PBE0-trained NNs reported here could likely be used with NNPs trained using data for other DFT methods.

3 Results and discussion

3.1 Stability

A key design objective of MLXDM was for this method to be practical for use in molecular simulations of a broad range of chemical systems using standard molecular simulation methods (*e.g.*, geometry optimization, molecular dynamics, Monte Carlo…). An advantage of MLXDM's construction is that every parameter of the potential energy function (*i.e.*, ν_i, $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V) is calculated from differentiable neural networks, so analytical gradients can calculated using the auto-differentiation feature of the PyTorch library. Based on its construction and training, ANIPBE0-MLXDM can be applied to arbitrary chemical systems provided that they have a neutral charge and are composed of elements C, N, O, and H.

We tested the ANIPBE0-MLXDM NNP by performing 100 ps Langevin molecular dynamics (MD) simulations of 100 structures with more than 20 non-hydrogen atoms chosen at random from the ChEMBL 31 database. All these structures retained standard chemical connectivity and geometry. 99 of these molecules retained their original molecular structure, while CHEMBL139573 underwent a spontaneous proton transfer from a carboxylic acid to the imine moiety of a guanidine group. We also performed 100 ps microcanonical molecular dynamics
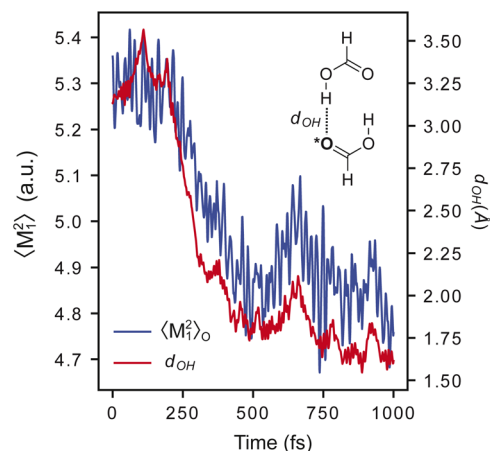
---

‡ The band of outlying underestimated predictions of $\langle M_1^2 \rangle$ values for oxygen corresponds to complexes containing carbon monoxide, which was not included in the training set. This is discussed in more detail in the ESI.†

Fig. 2 The H−O distance (blue) and $\langle M_1^2 \rangle_O$ (red) over the course of a 1 ps molecular dynamics simulation a pair of formic acid molecules. $\langle M_1^2 \rangle_O$ fluctuates significantly over the course of the trajectory and decreases systematically as the hydrogen bond forms.

(MD) simulations of these molecules and found that simulations conserve the total energy within the error of the numerical integration ($\mathcal{O}$ (10^{-7} eV)) shift in total energy per timestep, indicating that the calculated potential energy is consistent with its gradients.

A critical advantage of MLXDM over methods with fixed dispersion coefficients is that the coefficients propagate dynamically with the atomic positions, so they adapt to changes in the chemical environment over the course of a simulation. This is evident in Fig. 2, which shows the trajectory of a 1 ps MD simulation of a complex of two molecules of formic acid. In this trajectory, the oxygen–hydrogen distance (d_{OH}) decreases from ~3.5 Å in the initial configuration to ~1.6 Å when the two molecules form a hydrogen-bonded complex. In parallel to this, the fluctuation of the first moment of the exchange-dipole moment of the hydrogen-bond accepting oxygen, $\langle M_1^2 \rangle_O$, decreases from 5.3 a.u. to 4.7 a.u. due to the confinement of the electron density around the oxygen atom. This decrease in $\langle M_1^2 \rangle_O$ contributes to an attenuation in the dispersion component of the interaction energy of this complex.

3.2 Comparison to reference DFT data

We used the DES370K data set of interaction energies of intermolecular complexes[46] to test the new ANIPBE0-MLXDM NNP and its components. This test set includes 191 824 neutral intermolecular complexes composed of the elements H, C, N, and O. The performance of the uncorrected ANIPBE0 NNP was validated by comparing the ANIPBE0 NNP interaction energies to the PBE0 reference values. As shown in Fig. 3, it was found to provide generally good accuracy, with an R^2 of 0.97 and an MAE of 0.67 kcal mol^{-1}. Impressively, the dispersion components of the interaction energies calculated using MLXDM were found to be in almost perfect agreement with the DFT-derived XDM results, with an R^2 of 1.0 and an MAE of 0.01 kcal mol^{-1}. Finally, the total dispersion-corrected ANIPBE0-MLXDM energies are also in excellent agreement with the reference PBE0-XDM
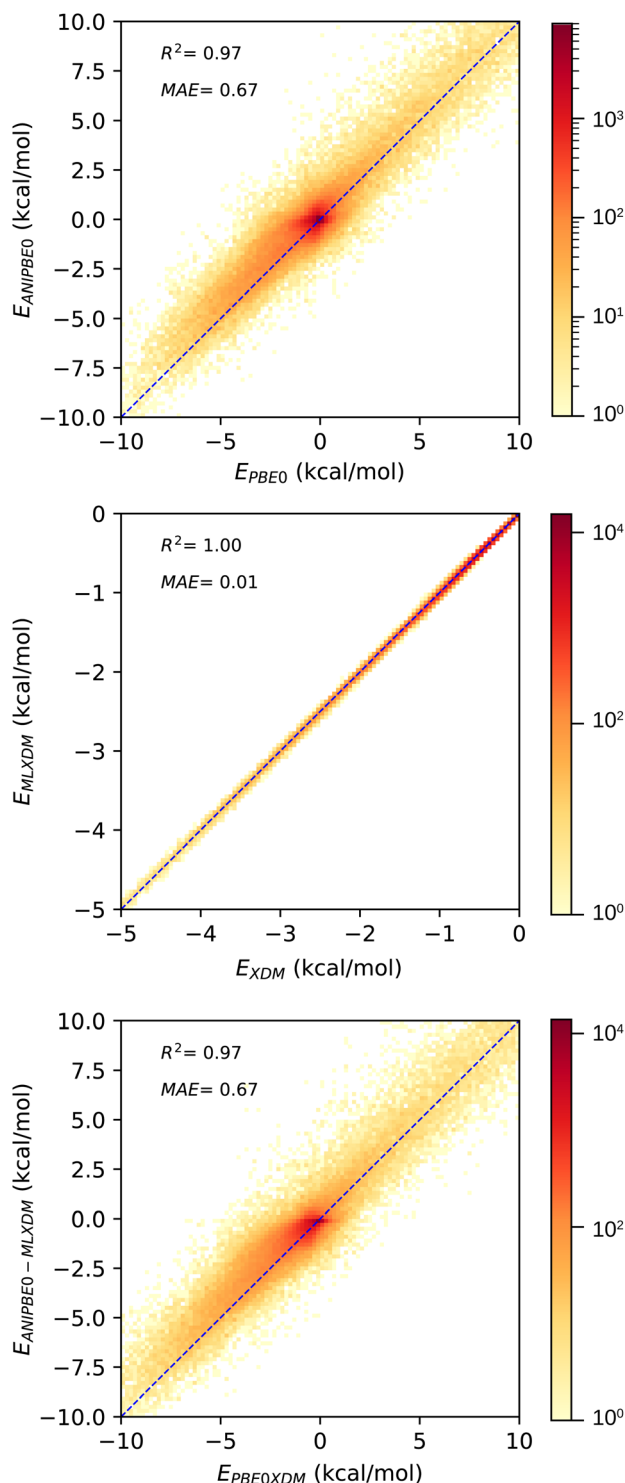
Fig. 3 Validation of the NNPs for the DES370K data set[46] compared to DFT reference data. Top: prediction of the PBE0 interaction energies using the ANIPBE0 NNP. Middle: prediction of the XDM dispersion component of the interaction energy by MLXDM. Bottom: prediction of the PBE0-XDM interaction energies by the combined ANIPBE0 NNP and MLXDM correction (ANIPBE0-MLXDM).

calculated with a higher degree of accuracy. Generally, these results show that molecular interaction energies from dispersion-corrected DFT calculations can be accurately approximated by ML approaches that estimate the DFT and dispersion energies separately.

3.3 Comparison to high-level correlated-wavefunction data

Although ANIPBE0-MLXDM is effective in predicting the inter-molecular interaction energies calculated by the PBE0-XDM dispersion-corrected DFT method it was trained to reproduce, the DFT data are also approximate. Thus, the true accuracy of the NNP cannot be assessed by this comparison alone. The DES370K test set provides interaction energies for the CCSD(T) correlated-wavefunction method, which intrinsically captures dispersion with a high level of accuracy. To judge how ANIPBE0-MLXDM compares with this more accurate but highly computationally
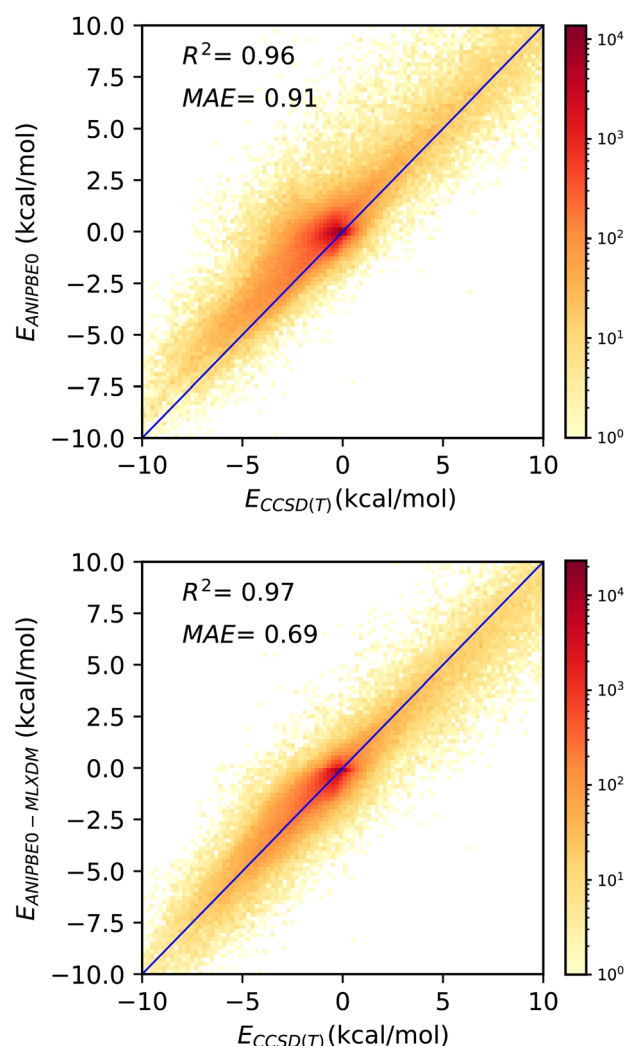

Fig. 4 Validation of the NNPs for the DES370K data set[46] compared to CCSD(T) reference data. Top: prediction of the CCSD(T) interaction energies by ANIPBE0; the interaction energies are systematically underestimated due to the neglect of London dispersion. Bottom: prediction of the CCSD(T) interaction energies by ANIPBE0-MLXDM; the correlation and MAE are improved by the MLXDM dispersion correction.

energies, with an R^2 of 0.97 and an MAE of 0.67 kcal mol^{-1}. This is consistent with the ANIPBE0 NNP being the largest source of error in this method, while the dispersion component can be

intensive method, we have compared the ANIPBE0 and ANIPBE0-MLXDM results to this CCSD(T) reference data, as shown in Fig. 4.

The MAE is significantly larger when the uncorrected ANIPBE0 NNP interaction energies are compared to the CCSD(T) interaction energies (Fig. 4, top), (MAE = 0.91 kcal mol^{-1}, $R^2 = 0.96$). The neglect of long-range dispersion is evident here, as the attractive components of the interaction energies are systematically too weak, with a mean signed error of 0.47 kcal mol^{-1}. Adding the MLXDM dispersion correction to the ANIPBE0 NNP (Fig. 4, bottom) improves the performance considerably (MAE = 0.69 kcal mol^{-1}, $R^2 = 0.97$) and the systematic underestimation is largely eliminated (mean signed error = 0.14 kcal mol^{-1}). The ANIPBE0-MLXDM NNP has sub-kcal mol^{-1} accuracy in predicting CCSD(T) interaction energies at a dramatically lower computational cost. While there are other DFT-based formulations of the dispersion energy,[47–55] the success of MLXDM in bringing an NNP into alignment CCSD(T) indicates that this atomic-pairwise description of the dispersion energy, with C_6, C_6, and C_{10} terms, is effective.

3.4 Use cases

The rigorous inclusion of long-range dispersion within a NNP will allow for routine simulation of chemical systems where dispersion is an important feature. To illustrate the widespread utility of ANIPBE0-MLXDM, we have applied it to simulations of molecular liquids and gas adsorption in porous materials, as highlighted in Fig. 5.

3.4.1 Covalent Organic Frameworks. Covalent Organic Frameworks (COFs) are porous materials formed from repeating organic units. Small molecules can be adsorbed into these pores, making COFs attractive materials for the sequestration and storage of gases. In particular, COF-320 is noted for its high adsorption of methane gas.[56] We performed a Grand Canonical Monte Carlo (GCMC) simulation of the adsorption of methane into this COF using the ANIPBE0 and ANIPBE0-MLXDM models (Fig. 5A). Dispersion is an important component of guest–host and guest–guest interactions. At saturating pressures (80 bar), the inclusion of the MLXDM dispersion correction increases the average number of adsorbed methane molecules from 3.5 to 16.6, indicating that dispersion plays a key role in the adsorption capacity of this material.

3.4.2 Molecular liquids. NNPs are uniquely powerful for simulating the liquid state of complex materials because they allow for the efficient calculation of interatomic forces, which is necessary for molecular dynamics simulations. Five 100 ps molecular dynamics simulation of liquid toluene were performed using ANIPBE0-MLXDM. The dynamic structure factor allows the liquid structure predicted by ANIPBE0-MLXDM to be validated, and it is found to be in very good agreement with the results of neutron scattering experiments,[57] as shown in Fig. 5B.

Decomposition of the dispersion energy into C_6, C_8, and C_{10} components (see Fig. 5B) shows that, although the contribution to the total dispersion energy from the C_6 term accounts for 61% for the intermolecular dispersion interactions in liquid
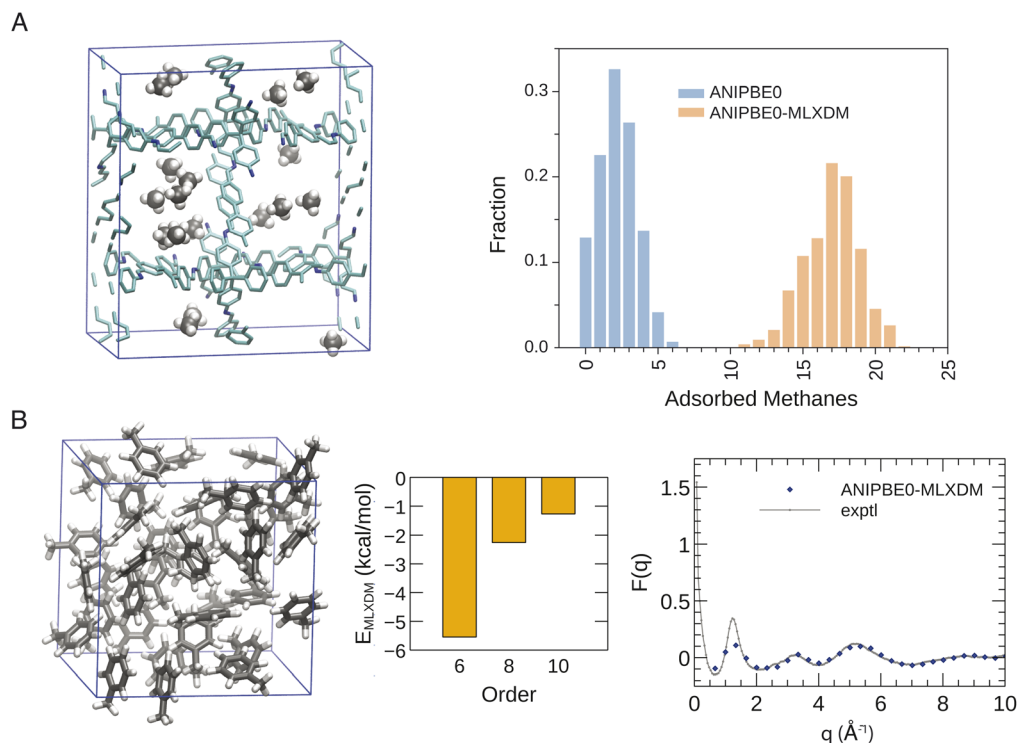
Fig. 5 (A) Snapshot from a GCMC simulation of methane adsorption into COF-320 at 80 bar (left) and a histogram comparing the distribution of adsorbed methane molecules with and without the inclusion of MLXDM dispersion (right). (B) Snapshot from a molecular dynamics simulation of liquid toluene (left), decomposition of the intermolecular dispersion interaction energy per molecule into 6th, 8th, and 10th order components (center), and comparison the calculated and experimental neutron scattering curves (right).

toluene, the C_8 and C_{10} dispersion terms are also significant, accounting for 25% and 14%, respectively. Most molecular-mechanical and some DFT-based dispersion corrections only include the leading-order C_6 term. These models will either underestimate the strength of dispersion, or will have to use anomalously large C_6 dispersion coefficients to capture the full dispersion energy,[58] resulting in a model for dispersion that is skewed to be erroneously weaker at short range but stronger at long range.

3.5 Comparison to XDM using fixed coefficients

The moments and atomic volumes calculated using MLXDM are calculated relative to their mean values over the training set. This suggests a simpler method that uses the average values of $\langle M_\ell^2 \rangle_i$ and V_i for each element to calculate the dispersion coefficients for the interaction between a pair of atoms without using the NNs, akin to dispersion corrections that use constant coefficients. We implemented this method in TorchANI as XDM-CC (exchange-hole dipole moment – constant coefficients). The XDM-CC method performs very well overall, with only incrementally higher errors in predicting the XDM dispersion interaction energy for the DES370K set than MLXDM ($R^2 = 0.99$ MAE = 0.03 kcal mol^{-1}, correlation plots are included in the ESI†).

At first glance, it may be surprising that XDM-CC yields very similar dispersion interaction energies to MLXDM because the atomic moments used to calculate the XDM dispersion interaction difference 6 to 18% from their elemental average value, on average (Table 6, ESI†), so the coefficients generated by XDM and MLXDM are significantly different. However, because the dispersion interaction energy between a pair of molecules is the sum of the dispersion interaction between many pairs of atoms, there is a tendency for the errors in the XDM-CC interaction energies to cancel (at least for light elements, Table 7, ESI†). As a result, the relative deviation between the MLXDM and XDM-CC interaction energies tend to be smaller than the deviations of the atomic coefficients; the final dispersion interaction energy is only 3.2% different from the MLXDM energy. The D3-BJ dispersion correction developed by Grimme and coworkers[47] gives comparable accuracy to XDM-CC (see ESI†), although it does not include a C_{10} term and does not have the same capacity as XDM to adjust the dispersion coefficients to reflect the chemical environment. The performance of D3-BJ is described in more depth in the ESI.†

Although XDM-CC gives similar dispersion interaction energies to MLXDM on average, there are instances where XDM-CC deviates significantly from XDM/MLXDM. Notably, the fixed coefficients used by XDM-CC gives dispersion interaction energies that are up to 10% different from the MLXDM coefficients in hydrogen-bonded complexes. This is evident in Fig. 2, where the $\langle M_1^2 \rangle$ for the hydrogen-bond donating oxygen atom of the formic-acid dimer fluctuates over a range significantly different from its average value for the whole data set. A similar effect is evident along the potential energy surface of the hydrogen-bonded complex of N-methylacetamide and piperidine (Fig. 6). Here, the XDM dispersion component of the
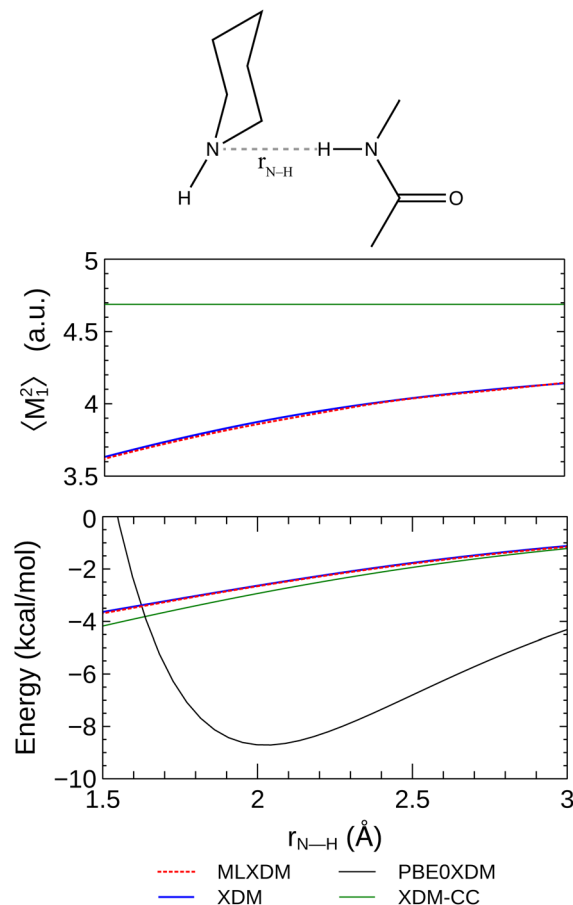

Fig. 6 Bottom: potential energy surface for the hydrogen-bonding interaction between N-methylacetamide and piperidine as a function of the distance between the amide hydrogen and the piperidine nitrogen. The XDM-CC model, which approximates all atoms of an element to have same volumes and exchange-dipole moments, overestimates the XDM dispersion energy. Top: the $\langle M_1^2 \rangle_N$ of the piperidine nitrogen along the potential energy surface predicted using MLXDM (red) vs. XDM (blue). XDM-CC uses the average $\langle M_1^2 \rangle_N$ of 4.7 a.u. (green).

interaction energy is overestimated by 10–15% when constant coefficients are used. Conversely, when the MLXDM NN method is used, the dispersion energy is predicted in nearly perfect agreement with the XDM values. This effect is largely due to changes in the $\langle M_1^2 \rangle$ moment of the piperidine nitrogen, which systematically decreases as the amide hydrogen approaches it.

We anticipate that the advantage of MLXDM adjusting the dispersion interactions of atoms based on their chemical environment will be more significant when the method is extended to elements that can adopt multiple oxidation states (e.g., phosphorous, sulfur, and transition metals).[59–61] The partial success of XDM-CC also suggests the computational cost of MLXDM could be mitigated by approximating some of the less significant XDM terms as constants rather than predicting them using NNs.

3.6 Performance

The inclusion of long-range dispersion interactions using MLXDM results in a significantly higher computational cost

than uncorrected ANIPBE0 calculations; a benchmark MD simulation of 10 pentane molecules required 4.3 times more computing time using ANIPBE0-MLXDM in comparison to ANIPBE0. The computational cost of MLXDM grows quadratically with the number of atoms because of the pair-wise sum, while computational cost of the original ANI model grows linearly.

Nevertheless, ANIPBE0-MLXDM still has a much lower computational cost than the DFT calculations it was trained to reproduce. In a benchmark calculation on a cluster of seven pentane molecules, the ANIPBE0-MLXDM NNP is roughly 4.5 million times faster than the equivalent calculation using the highly efficient electronic structure code TURBOMOLE.[62] This performance difference is expected to increase at least quadratically as the size of the system increases. NNPs are also amenable to parallel and GPU computing, so simulations with thousands or millions of atoms are possible.[63]

4 Conclusion

The ANIPBE0-MLXDM neural network potential was developed to explicitly include long-range London dispersion interactions through a pairwise correction. It includes 6th, 8th, and 10th order terms in the asymptotic expansion of the dispersion energy, unlike many popular models for dispersion that only include the leading, 6th order term. The C_6, C_8, and C_{10} coefficients for each interatomic dispersion interaction are calculated from atomic volumes and exchange-hole multipole moment integrals. These terms are calculated from NNs trained to reproduce their quantum-mechanical values derived from the XDM model. As with XDM, the MLXDM dispersion coefficients are intrinsically dependent on the chemical environment, in contrast to other models that use fixed coefficients. MLXDM predicts the dispersion component of intermolecular interaction energies in excellent agreement with the QM model it is trained to reproduce ($R^2 = 1.00$, MAE = 0.01 kcal mol^{-1}). It can be applied to arbitrary neutral systems composed of the elements H, C, N and O without further parameterization. By including the MLXDM correction for long-range dispersion interactions, and training a separate NNP (ANIPBE0) to account for the short-range interactions described by the PBE0 density functional, the combined ANIPBE0-MLXDM method is capable of predicting intermolecular interaction energies with a mean absolute error of 0.69 kcal mol^{-1} relative to high-level correlated-wavefunction methods.

To assess the importance of environment dependence, we compared MLXDM to XDM-CC. This is alternative method where constant values are used for $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V, calculated by averaging over these quantities for atoms of each element in the training set. Since there is comparatively little change in the dispersion coefficients with environment for H, C, N and O, the XDM-CC model performed remarkably well in comparison to MLXDM and it may be useful for adding a simple dispersion correction to DFT and NNP calculations without evaluating any additional NNs. However, we found the dispersion component of the interaction energies in hydrogen-bonded complexes were overestimated by XDM-CC.

The ANIPBE0-MLXDM method can be used to simulate large chemical systems with the same accuracy as dispersion-corrected DFT, but is far less computationally expensive than the PBE0-XDM method it approximates. Even for systems with hundreds of atoms, it was feasible to sample millions of configurations on a single GPU-equipped workstation. MLXDM offers a rigorous and computationally efficient approach to describing long-range London dispersion interactions within simulations using neural network potentials. We anticipate that this approach can be widely used in large-scale NNP simulations.

5 Data availability

This method has been implemented into the TorchANI code.[43] The source code for this modified version is deposited on GitHub (**https://github.com/RowleyGroup/MLXDM**). The data set used to train and validate this method is available for download from FigShare **https://figshare.com/articles/dataset/ANIPBE0-MLXDM_dataset/19790524**. The ANIPBE0 and MLXDM NNs can be trained and evaluated using this modified TorchANI code and this data set. Instructions for installation and execution are included in the GitHub repository. The DES370K data set used to test this method is available to download from a separate FigShare repository **https://figshare.com/articles/dataset/MLXDM_test_set_data/22223221**. The ANIPBE0 and MLXDM NNs can be trained and evaluated using this modified TorchANI code and this data set. Instructions for installation and execution are included in the GitHub repository.

Author contributions

Tu implemented MLXDM into TorchANI and developed the active learning protocol. Rezajooei generated part of the training data and trained the first versions of ANIPBE0-MLXDM. Rowley devised the concept, generated part of the training data, provided supervision, contributed to the writing of the manuscript, and trained the NNs. Johnson helped devise the concept, developed the theory, and contributed to the writing of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 V. L. Deringer and G. Csányi, *Phys. Rev. B*, 2017, **95**, 094203.

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

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

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

5 R. Zubatyuk, J. S. Smith, J. Leszczynski and O. Isayev, *Sci. Adv.*, 2019, **5**, eaav6490.

6 Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney and C. D. Sherrill, *J. Chem. Phys.*, 2020, **153**, 044112.

7 J. R. Cendagorta, H. Shen, Z. Bačić and M. E. Tuckerman, *Adv. Theory Simul.*, 2021, **4**, 2000258.

8 L. Zhang, H. Wang, M. C. Muniz, A. Z. Panagiotopoulos, R. Car and W. E, *J. Chem. Phys.*, 2022, **156**, 124107.

9 E. Kocer, T. W. Ko and J. Behler, *Annu. Rev. Phys. Chem.*, 2022, **73**, 163–186.

10 H. Gokcan and O. Isayev, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2022, **12**, e1564.

11 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**, 241733.

12 C. Devereux, J. S. Smith, K. K. Davis, K. Barros, R. Zubatyuk, O. Isayev and A. E. Roitberg, *J. Chem. Theory Comput.*, 2020, **16**, 4192–4202.

13 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Nat. Commun.*, 2019, **10**, 2903.

14 S.-L. J. Lahey and C. N. Rowley, *Chem. Sci.*, 2020, **11**, 2362–2368.

15 S.-L. J. Lahey, T. N. Thien Phuc and C. N. Rowley, *J. Chem. Inf. Model.*, 2020, **60**, 6258–6268.

16 N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 1–4.

17 D. P. Metcalf, A. Koutsoukas, S. A. Spronk, B. L. Claus, D. A. Loughney, S. R. Johnson, D. L. Cheney and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 074103.

18 D. M. Anstine and O. Isayev, *J. Phys. Chem. A*, 2023, **127**(11), 2417–2431.

19 A. Grisafi and M. Ceriotti, *J. Chem. Phys.*, 2019, **151**, 204105.

20 S. Yue, M. C. Muniz, M. F. Calegari Andrade, L. Zhang, R. Car and A. Z. Panagiotopoulos, *J. Chem. Phys.*, 2021, **154**, 034111.

21 T. Morawietz, V. Sharma and J. Behler, *J. Chem. Phys.*, 2012, **136**, 064103.

22 K. Kato, T. Masuda, C. Watanabe, N. Miyagawa, H. Mizouchi, S. Nagase, K. Kamisaka, K. Oshima, S. Ono, H. Ueda, A. Tokuhisa, R. Kanada, M. Ohta, M. Ikeguchi, Y. Okuno, K. Fukuzawa and T. Honma, *J. Chem. Inf. Model.*, 2020, **60**, 3361–3368.

23 K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre and J. Parkhill, *Chem. Sci.*, 2018, **9**, 2261–2269.

24 S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker and M. Amsler, *Phys. Rev. B*, 2017, **95**, 104105.

25 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Nat. Commun.*, 2021, **12**, 398.

26 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Acc. Chem. Res.*, 2021, **54**, 808–817.

27 D. P. Metcalf, A. Jiang, S. A. Spronk, D. L. Cheney and C. D. Sherrill, *J. Chem. Inf. Model.*, 2021, **61**, 115–122.

28 H. Muhli, X. Chen, A. P. Bartók, P. Hernández-León, G. Csányi, T. Ala-Nissila and M. A. Caro, *Phys. Rev. B*, 2021, **104**, 054106.

29 P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardère and J.-P. Piquemal, *J. Phys. Chem. Lett.*, 2022, **13**, 4381–4388.

30 P. P. Poier, L. Lagardère and J.-P. Piquemal, *Generalized Many-Body Dispersion Correction through Random-phase Approximation for Chemically Accurate Density Functional Theory*, 2022, https://arxiv.org/abs/2210.09784.

31 A. Stone and A. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, UK, 2013.

32 A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 154108.

33 E. R. Johnson, *Non-Covalent Interactions in Quantum Chemistry and Physics*, Elsevier, Amsterdam, NL, 2017, pp. 169–194.

34 A. J. A. Price, K. R. Bryenton and E. R. Johnson, *J. Chem. Phys.*, 2021, **154**, 230902.

35 M. Mohebifar, E. R. Johnson and C. N. Rowley, *J. Chem. Theory Comput.*, 2017, **13**, 6146–6157.

36 E. R. Johnson, *J. Chem. Phys.*, 2011, **135**, 234109.

37 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

38 A. L. Hickey and C. N. Rowley, *J. Phys. Chem. A*, 2014, **118**, 3678–3687.

39 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science*, 2017, **355**, 49–52.

40 J. Behler, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 12828–12840.

41 H. Seung, M. Opper and H. Sompolinsky, *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, New York, NY, USA, 1992, pp. 287–294.

42 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma,

O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.

43 X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, *J. Chem. Inf. Model.*, 2020, **60**, 3408–3415.

44 The prediction for the coefficients of H atoms is slightly poorer ($R^2 = 0.96$ MAE $= 0.14$ a.u.), although the absolute performance is actually similar for all elements because these metrics are for the prediction of the standardized coefficients and magnitude and range of coefficients for H are smaller than the other elements.

45 A. Otero-de-la Roza and E. R. Johnson, *J. Chem. Phys.*, 2013, **138**, 204109.

46 A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K.-H. Law, B. A. Gregersen, J.-L. Li, K. Palmo, K. Siva, M. Bergdorf, J. L. Klepeis and D. E. Shaw, *Sci. Data*, 2021, **8**, 55.

47 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

48 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

49 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.

50 S. N. Steinmann and C. Corminboeuf, *J. Chem. Theory Comput.*, 2011, **7**, 3567–3577.

51 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.

52 A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.

53 A. Ambrosetti, A. M. Reilly, R. A. DiStasio Jr and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 18A508.

54 T. Gould, S. Lebègue, J. G. Ángyán and T. Bučko, *J. Chem. Theory Comput.*, 2016, **12**, 5920–5930.

55 M. Kim, W. J. Kim, T. Gould, E. K. Lee, S. Lebègue and H. Kim, *J. Am. Chem. Soc.*, 2020, **142**, 2346–2354.

56 Y.-B. Zhang, J. Su, H. Furukawa, Y. Yun, F. Gándara, A. Duong, X. Zou and O. M. Yaghi, *J. Am. Chem. Soc.*, 2013, **135**, 16336–16339.

57 T. F. Headen, C. A. Howard, N. T. Skipper, M. A. Wilkinson, D. T. Bowron and A. K. Soper, *J. Am. Chem. Soc.*, 2010, **132**, 5735–5742.

58 A. Otero-de-la Roza, L. M. LeBlanc and E. R. Johnson, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8266–8276.

59 E. R. Johnson, *J. Chem. Phys.*, 2011, **135**, 234109.

60 M. S. Christian, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2016, **12**, 3305.

61 A. Otero-de-la-Roza, L. M. LeBlanc and E. R. Johnson, *J. Phys. Chem. Lett.*, 2020, **11**, 2298–2302.

62 *TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007*, available from **http://www.turbomole.com**.

63 S. Desai, S. T. Reeve and J. F. Belak, *Comput. Phys. Commun.*, 2022, **270**, 108156.