Highly accurate many-body potentials for simulations of N_2O_5 in water: benchmarks, development, and validation

Vinícius Wilian D. Cruzeiro [1,2,], Eleftherios Lambros [2], Marc Riera [2], Ronak Roy [1], Francesco Paesani [1,2,3], Andreas W. Götz [1,*]*

[1] San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, United States

[2] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, United States

[3] Materials Science and Engineering, University of California San Diego, La Jolla, CA 92093, United States

ABSTRACT

Dinitrogen pentoxide (N_2O_5) is an important intermediate in the atmospheric chemistry of nitrogen oxides. Although there has been much research, the processes that govern the physical interactions between N_2O_5 and water are still not fully understood at a molecular level. Gaining quantitative insight from computer simulations requires going beyond the accuracy of classical force fields, while accessing length scales and time scales that are out of reach for high-level quantum chemical approaches. To this end we present the development of MB-nrg many-body potential energy functions for simulations of N_2O_5 in water. This MB-nrg model is based on electronic structure calculations at the coupled cluster level of theory and is compatible with the successful MB-pol model for water. It provides a physically correct description of long-range many-body interactions in combination with an explicit representation of up to three-body short-range interactions in terms of multidimensional permutationally invariant polynomials. In order to further investigate the importance of the underlying interactions in the model, a TTM-nrg model was also devised. TTM-nrg is a more simplistic representation that contains only two-body short-range interactions represented through Born-Mayer functions. In this work an active learning approach was employed to efficiently build representative training sets of monomer, dimer and trimer structures, and benchmarks are presented to determine the accuracy of our new models in comparison to a range of density functional theory methods. By assessing binding curves, distortion energies of N_2O_5, and interaction energies in clusters of N_2O_5 and water, we evaluate the importance of two-body and three-body short-range potentials. The results demonstrate that our MB-nrg model has high accuracy with respect to the coupled cluster reference, outperforms current density functional theory models, and thus enables highly accurate simulations of N_2O_5 in aqueous environments.

INTRODUCTION

Dinitrogen pentoxide (N_2O_5) is an important molecule in the atmospheric chemistry of nitrogen oxides (NO_x). Because N_2O_5 serves as a temporary reservoir for NO_x, its fate has important consequences for the abundance of atmospheric ozone, hydroxyl radicals, methane, and other species, with major implications for air quality and climate.[1–4] Up to 50% of N_2O_5 is thought to be removed from the atmosphere through reactive uptake by aqueous aerosol.[1] Because of its importance, N_2O_5 has been extensively studied over the past few decades, both theoretically and experimentally.[5–21] Notwithstanding all these efforts, the mechanism of reactive uptake of N_2O_5 to aqueous aerosol is still not fully understood at a molecular level. The initial step of this process is the physical accommodation of N_2O_5 in aqueous aerosol.[6, 12, 14, 22] Important steps towards the goal of understanding this mechanism include predictive computer simulations of N_2O_5 with small water clusters, at extended air/water interfaces, and in bulk water.

Classical force fields, which are widely used in computer simulations because of their computational efficiency, have been employed in molecular dynamics (MD) simulations to study the hydration properties of N_2O_5.[5, 23] However, due to their inherent approximations and limited accuracy, classical force fields might not be relied upon for quantitative predictions. In *ab initio* MD (AIMD), the configurations of a system are explored at a given thermodynamic condition with energies and forces computed on-the-fly from first-principles electronic structure calculations [24, 25]. AIMD simulations, accessing limited time scales, have been employed to explore the physical accommodation of N_2O_5 to bulk water.[5] While AIMD can be directly used without the need for any prior parametrizations, condensed phase simulations are typically restricted to density functional theory (DFT), which has limited accuracy and displays difficulties in correctly describing the structure of liquid water.[26] In addition, the high computational cost associated with

AIMD precludes sampling over long time scales, especially for systems that contain a large number of atoms.

With the recent advent of new machine learning techniques, electronic structure calculations can now be reproduced quite accurately at relatively low computational cost.[27–31] A popular class of current deep learning approaches for quantum chemistry makes use of symmetry functions[32] for the description of the chemical environment around an atom and transferable neural networks in order to describe the molecular interactions.[33–35] Using appropriate training data, such methods have been shown to accurately reproduce coupled cluster potential energy surfaces of small molecules in the gas-phase.[36] However, because these models rely on descriptors of the local atomic environment, they do not explicitly incorporate long-range interactions, which are essential to accurately describe extended systems or condensed phase systems.[37, 38]

This drawback can be addressed with an approach that is based on the many-body expansion (MBE)[39, 40] of the total energy as has been demonstrated with the successful MB-pol model for water.[41–43] In brief, MB-pol consists of a classical polarizable model with physically motivated representation of long-range interactions in combination with explicit data-driven corrections to the low-order terms of the MBE. MB-pol uses highly accurate coupled cluster electronic structure calculations as reference data for short-range interactions. Specifically, short-range two-body and three-body potentials are represented in terms of permutationally invariant polynomials (PIPs),[44–47] although it was shown that similar accuracy is obtained using other functions such as Gaussian approximation potentials or neural networks.[48] Importantly, since the classical induction is fully many-body in MB-pol, the MBE is not truncated, which differs from typical fragmentation approaches.[49–54] MB-pol thus accurately represents each term of the many-body expansion at all

orders, both at short- and at long-range. As a result, MB-pol is capable of correctly describing the properties of water across different phases.[37, 42, 43, 55, 56]

The MB-pol model has been extended to describe generic mixtures of molecules. These potential energy functions, termed MB-nrg (for "many-body energy"), employ the same underlying functional forms for the intermolecular many-body interactions as MB-pol. The first MB-nrg models have been devised to describe the interactions of water with monoatomic ions[57–59] and, more recently, to describe CO_2/water and CH_4/water mixtures.[60, 61] Similarly to MB-pol, the MB-nrg potential energy functions have been successfully used in studies across phases from clusters[62–67] to bulk mixtures.[60, 66] It is possible to replace the PIPs with simple atom-pairwise Born-Mayer functions[68, 69] to describe the two-body (2B) short-range repulsion between pairs of molecules in an approximate fashion, in which case the model is called TTM-nrg,[70, 71] because it is inspired from the Thole-Type Model (TTM).[72–76] TTM-nrg potentials have been developed for the same systems for which MB-nrg models are available.[60, 70, 71]

In the current work, we present and validate the development of MB-nrg and TTM-nrg models for accurate simulations of N_2O_5 and water. We have employed an active learning procedure that makes use of neural networks during the construction of our training sets in order to efficiently sample the multi-dimensional configurational space of N_2O_5 as well as all relevant two- and three-body configurations of dimers (N_2O_5 + H_2O) and trimers (N_2O_5 + 2 H_2O). In line with MB-pol, we have used coupled cluster theory with single and double excitations and perturbative treatment of triple excitations, CCSD(T), extrapolated to the complete basis set (CBS) limit as our reference level. We present a detailed analysis of the fit quality of the new MB-nrg potential energy functions and benchmark calculations that demonstrate superior accuracy as compared to a range of density functional theory (DFT) methods. Specifically, we assess distortion energies of N_2O_5, binding

curves in the (N_2O_5 + H_2O) dimer and (N_2O_5 + 2 H_2O) trimer, and interaction energies in (N_2O_5 + n H_2O) clusters (n < 5). We also investigate the relative importance of individual terms in the many-body expansion for interaction energies and cluster geometries.

THEORY AND METHODOLOGIES

The MB-nrg and TTM-nrg models

According to the many-body expansion (MBE) [39, 40], the total energy of a system with *N* bodies can be described in terms of contributions of one-body, two-body, three-body, etc., up to the *N*-body contribution:

$$E = \sum_{i}^{N} V^{1\mathrm{B}}(i) + \sum_{i<j}^{N} V^{2\mathrm{B}}(i,j) + \sum_{i<j<k}^{N} V^{3\mathrm{B}}(i,j,k) + \cdots + V^{N\mathrm{B}}(1,\ldots,N) \quad (1)$$

Here, $V^{1B}(i)$ corresponds to the one-body (1B) distortion energy of monomer i from its equilibrium geometry, and the *n*-body ($V^{n\mathrm{B}}$) energies are defined recursively starting from the 1B energies. By construction, this expansion is expected to converge rather quickly. The use of the MBE for larger systems, truncated at some order in which the errors are sufficiently low, becomes significantly more efficient than performing a single calculation considering the whole system.[39, 40] Instead of computing the individual terms of the MBE with electronic structure methods on-the-fly, further speedup can be obtained by using analytical functions that closely represent these terms. These underlying ideas are exploited in MB-nrg and MB-pol potential energy functions.

Both MB-pol and MB-nrg rely on an underlying polarizable classical potential that contains permanent electrostatics, dispersion, and polarization interactions. Many-body induction interactions are naturally taken into account by the classical potential without truncation at all

orders of the MBE due to the nature of the polarization effects. In order to capture short-range effects adequately, the lower order many-body interactions up to three-body (3B) are corrected using permutationally invariant polynomials (PIPs).[44–46] The many-body interactions are described in MB-nrg and MB-pol as follows:[41–43]

$$V^{1\mathrm{B}} = V_{\mathrm{SR}}^{1\mathrm{B}} \quad (2)$$

$$V^{2\mathrm{B}} = s_{2\mathrm{B}} V_{\mathrm{SR}}^{2\mathrm{B}} + V_{\mathrm{elec}}^{2\mathrm{B}} + V_{\mathrm{ind}}^{2\mathrm{B}} + V_{\mathrm{disp}}^{2\mathrm{B}} \quad (3)$$

$$V^{3\mathrm{B}} = s_{3\mathrm{B}} V_{\mathrm{SR}}^{3\mathrm{B}} + V_{\mathrm{ind}}^{3\mathrm{B}} \quad (4)$$

$$V^{>3\mathrm{B}} = V_{\mathrm{ind}}^{>3\mathrm{B}} \quad (5)$$

The terms labeled with short-range (SR) are described by the PIPs, except for $V_{\mathrm{SR}}^{1\mathrm{B}}$ in MB-pol, which is described by the accurate water monomer model devised by Partridge and Schwenke.[77] $V_{\mathrm{elec}}^{2\mathrm{B}}$ represents 2B permanent point-charge electrostatics and $V_{\mathrm{ind}}^{n\mathrm{B}}$ represents *n*-body polarization contributions that are described by an extended Thole-type model as originally introduced by the TTM4-F water model.[72] $V_{\mathrm{disp}}^{2\mathrm{B}}$ is the 2B dispersion energy that is represented by a sum of pairwise additive contributions with interatomic dispersion coefficients determined from electronic structure calculations, dampened at short-range with Tang-Toennies functions.[78] As can be seen, the PIPs in the 2B and 3B potentials are designed to correct for the difference between the classical potential and the reference electronic structure calculations.

TTM-nrg is a simpler model in which the short-range corrections go up to two-body ($V_{\mathrm{SR}}^{3\mathrm{B}} = 0$) and the PIPs in the short-range 2B term $V_{\mathrm{SR}}^{2\mathrm{B}}$ are replaced with a simple atom-pairwise Born-Mayer repulsion potential[68] between monomers M_1 and M_2 without a switching function so that $V^{2\mathrm{B}}$ becomes a Coulomb-Buckingham potential[69, 79] plus induction effects:

$$V_{\mathrm{SR}}^{2\mathrm{B}}(\text{TTM-nrg}) = \sum_{i\in \mathrm{M}_1, j\in \mathrm{M}_2} A_{ij} e^{-B_{ij} r_{ij}} \qquad (6)$$

TTM-nrg can be regarded as a simple polarizable potential which is expected to be less accurate than MB-nrg.

The switching functions $s_{2\mathrm{B}}$ and $s_{3\mathrm{B}}$ in Equations (2) and (3) are employed in order to guarantee a smooth transition when switching off the short-range terms at large distances. The cutoff distances are chosen such that the 2B and 3B interactions are accurately described by the underlying classical potential, hence saving computational time by not evaluating the short-range expressions for dimers and trimers at larger distances. In this work we are using the following switching functions:

$$s_{2\mathrm{B}}(r_{12}) = \begin{cases} 1, & r_{12} \le r_{\mathrm{in}} \\ 0, & r_{12} \ge r_{\mathrm{out}} \\ \dfrac{1 + cos\left[\left(\dfrac{r_{12} - r_{\mathrm{in}}}{r_{out} - r_{\mathrm{in}}}\right)\pi\right]}{2}, & \text{elsewhere} \end{cases} \qquad (7)$$

$$s_{3\mathrm{B}}(r_{12}, r_{13}, r_{23}) = f\left(s_{2\mathrm{B}}(r_{12}) + s_{2\mathrm{B}}(r_{13}) + s_{2\mathrm{B}}(r_{23})\right) \qquad (8)$$

$$f(x) = \begin{cases} 0, & x \le 1 \\ 1, & x \ge 2 \\ 1 - cos^2\left[(x-1)\dfrac{\pi}{2}\right], & \text{elsewhere} \end{cases} \qquad (9)$$

The 2B switching function $s_{2\mathrm{B}}$ is 1 when the dimer distance is below the inner cutoff value r_{in}, and then it smoothly transitions to 0 at the outer cutoff value r_{out}. The 3B switching function $s_{3\mathrm{B}}$ is 1 when the distance of at least two dimers inside the trimer is below the inner cutoff value r_{in}. Thus, $s_{3\mathrm{B}}$ is 1 for any interacting trimer, including the most extreme case of linear trimers, and then it smoothly transitions to 0 when at least two dimer distances exceed the outer cutoff distance r_{out}, that is, when one monomer is sufficiently separated from the two other monomers. It is important to emphasize that we can use different values of r_{in} and r_{out} for $s_{2\mathrm{B}}$ and $s_{3\mathrm{B}}$, and that the first derivatives of $s_{2\mathrm{B}}$ and $s_{3\mathrm{B}}$ are continuous. The expression used here for $s_{2\mathrm{B}}$ is identical to the one

used in MB-pol, while we employ a different expression for s_{3B} in this work. The value of s_{3B} in MB-pol can go up to 3.0, [48] which is not a problem if the switching function is included during the determination of the 3B PIPs. On the other hand, with the s_{3B} expression that we are proposing here, the values of r_{in} and r_{out} can be modified without refitting the 3B PIPs. The proposed expression for s_{3B} could even be employed in case electronic structure calculations would be directly employed in place of evaluating the fitted potential.

Building training sets with neural networks and active learning

A crucial point when developing many-body potentials for new systems is the generation of high-quality training sets that adequately sample the relevant high-dimensional intra- and inter-molecular configurational space. In this work, we have employed the active learning procedure via query-by-committee through cross-validation proposed in Ref. [80] in the process of selecting structures for the one-body (N_2O_5), two-body (N_2O_5 + H_2O), and three-body (N_2O_5 + 2 H_2O) training sets. In this procedure, ANI neural networks (NNs)[33] are constructed aiming at building a criterium for the selection of new structures for the training sets. Given that we only need to describe monomers, dimers, or trimers at short range, the lack of long-range effects inherent to NN potentials can be overcome by increasing the NN cutoff distance in such a way that all structures are fully considered.

In a nutshell, in this active learning procedure the initial training data set is split into N_g random groups of approximately equal size. Then, a total of N_g different NNs are constructed where the training data of each NN is composed of the total data set minus one of the random groups. Therefore, each NN does not contain "1/ N_g"-th of the data set.

Once this ensemble of NNs is constructed, it can be employed to determine whether new structures should be accepted or rejected for inclusion into the existing training data set. For each new structure *i* we use the following inclusion criterium:[80]

$$\rho_i = \frac{\sigma_i}{\sqrt{N_{\mathrm{at}}}} = \frac{1}{\sqrt{N_{\mathrm{at}}}} \sqrt{\frac{1}{N_g} \sum_{n=1}^{N_g} \left(E_{i,n} - \langle E_n \rangle\right)^2} \qquad (10)$$

where N_{at} is the number of atoms, σ_i is the standard deviation of $E_{i,n}$, the energy predicted for the new structure *i* by each neural network *n*, and $\langle E_n \rangle = \frac{1}{N_g} \sum_{n=1}^{N_g} E_{i,n}$ is the average energy predicted by the ensemble. The normalization coefficient $\sqrt{N_{\mathrm{at}}}$ has been introduced due to system size effects on the energy fluctuations, and allows a given value for the criterium ρ_i to be used for systems of different sizes. If the energy predictions of all NNs agree with each other, then ρ_i is small, and there is no need to add the new structure *i* into the training set. However, if ρ_i is larger than a given threshold, then the new structure *i* is added to the training set. Once a given number of new structures is fed to the ensemble of NNs, an active learning iteration step is completed. Then, the whole process can be repeated, starting from building a new ensemble of NNs with the new data set that contains the new structures that have been accepted. Active learning iteration steps are performed until the number of selected structures becomes sufficiently small. As shown in Ref. [80], this active learning procedure can be employed effectively to avoid redundancies in the training sets, which consequently leads to smaller training sets, thus significantly saving computational time during the calculation of reference energies with electronic structure methods. More details about this active learning approach can be found in Ref. [80].

COMPUTATIONAL DETAILS

Description of the MB-nrg model

We employ the MB-pol model to treat all water molecule properties (monomer distortion energies, atomic charges and polarizabilities) and all interactions that involve only water molecules. Molecular properties of N_2O_5 that are required for the classical potential were computed with the range-separated hybrid meta-GGA density functional ωB97M-V[81] and the aug-cc-pVTZ[82, 83] basis set at the optimized C_2 geometry of an isolated N_2O_5 monomer. The ChElPG model[84] as implemented in Q-Chem 5.0[85] was used to determine atomic charges. Atomic dipole polarizabilities (α_i^{free}) were taken from the literature[86] and scaled by the ratio of the effective atomic volumes (V_i^{eff}) of the atoms in N_2O_5 and the free volumes of the isolated N and O atoms (V_i^{free}):[87, 88]

$$\alpha_i^{\text{eff}} = \alpha_i^{\text{free}} \frac{V_i^{\text{eff}}}{V_i^{\text{free}}} \qquad (11)$$

Both V_i^{free} and V_i^{eff} were obtained using the exchange-hole dipole moment (XDM) model[89, 90] as implemented in Q-Chem 5.0.[85] The XDM model was also used to compute the C_6 dispersion coefficients that describe both intramolecular dispersion in N_2O_5 and intermolecular dispersion between N_2O_5 and water. The values of the N and O charges and polarizabilities are reported in Table S1 in the Supporting Information, and the C_6 dispersion coefficients in Table S2.

The PIPs that are used to represent $V_{\text{SR}}^{1\text{B}}$, $V_{\text{SR}}^{2\text{B}}$ and $V_{\text{SR}}^{3\text{B}}$ in Eqs. 2 to 4 depend on the distances d_{ij} between atom pairs *ij* involving the atoms of N_2O_5 and water, which can be intramolecular or intermolecular distances. A set of exponential functions of the type $\xi_{ij} = e^{-k_{ij}d_{ij}}$ is constructed for each atom pair and the PIP is written as a sum of products of these functions: $V_{\text{PIP}} = \sum_l c_l \eta_l$.

The terms η_l are symmetrized monomials, which are products of the functions ξ_{ij} up to a given degree. Symmetrization in this context means ensuring that V_{PIP} is invariant with respect to permutations of equivalent atoms. In N_2O_5 the four peripherical oxygen atoms (Op in Figure 3) were considered equivalent, as well as the two nitrogen atoms. In the water molecule, the two hydrogen atoms were considered equivalent. The c_l and k_{ij} terms are, respectively, the linear and non-linear parameters to be fitted based on the reference electronic structure calculations performed on the configurations in the training sets while satisfying equations 2, 3, and 4. See Refs. [41, 42, 48] for more details.

The number of terms and the degree of the PIPs devised to describe the 1B distortion energy of N_2O_5 and the 2B and 3B interaction energies between N_2O_5 and water were carefully tuned to find a balance between accuracy and efficiency. For the 1B potential, monomials up to 5$^{\mathrm{th}}$ degree were considered, giving a total of 2624 monomial terms and 5 non-linear parameters. For the 2B potential, monomials up to 3$^{\mathrm{rd}}$ degree were considered and all monomials that contain only intra-molecular distances were removed, which results in a total of 983 monomial terms with 13 non-linear parameters. For the 3B potential, monomials up to 2$^{\mathrm{nd}}$ degree were considered if distances involving any oxygen atom from N_2O_5 and one oxygen atom of water are present, and up to 3$^{\mathrm{rd}}$ degree otherwise. As for the 2B potential, monomials that contain only intra-molecular distances have been excluded. The resulting 3B PIP contains 2038 monomial terms with 16 non-linear parameters. The final 1B, 2B and 3B potentials have been included in the MBX software for MB-nrg simulations, which is publicly available on GitHub.[91]

Having a balance between accuracy and efficiency in mind, the following inner and outer cutoffs, respectively, (see equations 7 to 9) were used in combination with these potentials: 8.0 and 9.0 Å

for the 2B potential, and 4.0 and 5.0 Å for the 3B potential. In our implementation, the cutoff distance is computed with respect to the distances between Oc of N_2O_5 and Ow of water (see Figure 3).

Description of the TTM-nrg model

As mentioned above, the TTM-nrg model developed here uses the same 1B PIPs as the MB-nrg model. For the description of 2B interactions between N_2O_5 and one water molecule, the TTM-nrg potential contains 6 pairs of parameters A_{ij} and B_{ij} (see equation 6), one for each pair of inter-molecular atom types; thus, 6 linear and 6 non-linear parameters. As can be observed, the number of linear parameters is significantly smaller than in the 2B PIPs in MB-nrg.

Training and test sets for the MB-nrg model

In order to build our training sets, we performed active learning with an ensemble of 8 neural networks and the parameter ρ_i (see equation 10) set to 0.06 kcal/mol for the 1B training set and 0.12 kcal/mol for the 2B and 3B training sets. The neural networks were trained with data from electronic structure calculations performed at the ωB97X/6-31G* level of theory computed with Gaussian 09,[92] which is the same level of theory that was used for the development of the ANI-1 neural network.[33]

The structures for the initial 1B, 2B and 3B training sets were generated using normal mode sampling (NMS) [33, 93] at 600K for the optimized N_2O_5 monomer at its two minima, with C_2 and C_s symmetry, and at 300K for optimized (N_2O_5 + *n* H_2O) dimers and trimers (*n* = 2, 3), respectively. The geometry optimizations and frequency calculations for NMS were performed at the ωB97XD/aug-cc-pVDZ level using Gaussian 09.[92] For the geometry optimizations, a total of 400

dimers and 400 trimers were extracted from *ab initio* MD simulations of N_2O_5 in bulk water and N_2O_5 at a water liquid/vapor interface performed in Ref. [5]. Only the unique minima, as determined by a root-mean-square distance with a tolerance of 0.1 Å, were considered and submitted to frequency calculations. Not only structures from NMS were fed to the active learning selection, but also additional structures of monomers, dimers, and trimers from the *ab initio* MD trajectories with center-of-mass (COM) distances up to the cutoff distances of 9.0 and 7.0 Å, respectively for dimers and trimers.

The training sets as described above were used to generate an initial MB-nrg model. Simulations of N_2O_5 in bulk and at a water liquid/vapor interface were performed with this model, and monomers, dimers, and trimers were extracted from these simulations, using active learning selection, to refine the training sets.

Part of the final configurations selected were removed from the training sets and considered as test sets. The final training and test sets for N_2O_5 and water contain, respectively: 7118 and 358 monomers, 6723 and 642 dimers, and 15415 and 908 trimers. It is interesting to compare the number of structures in these training sets to the MB-pol training sets that contain 42508 water dimers and 12347 water trimers.[48] Given that N_2O_5 has more degrees of freedom than a water molecule, one would expect a larger number of structures in the N_2O_5 training sets, assuming that the same procedure would be used as employed during the construction of the MB-pol training sets. Therefore, our reduced number of dimers and a similar number of trimers in comparison to the MB-pol training sets is a good indication of the effectiveness of active learning, since the original MB-pol training sets have been devised without an active learning selection.

Training set for the TTM-nrg model

Given the simplicity and the small number of free parameters in the repulsive 2B Born-Mayer potential (see equation 6), a simpler training set was constructed for TTM-nrg. This training set is composed of random rotations of the optimized structures of the N_2O_5 monomer at C_2 symmetry and of a water molecule at varying center-of-mass distances between the molecules up to 4.5 Å using logarithmic progression. Shorter distances were favored because only repulsive interactions are described by the Born-Mayer potential used in TTM-nrg (see equation 6). A total of 200 structures, that adequately cover the relevant repulsive regions, were considered in the training set.

Benchmarks: choosing highly accurate reference levels of theory

Our goal is to train the MB-nrg model against highly accurate reference data with accuracy comparable to CCSD(T) calculations extrapolated to the complete basis set (CBS) limit, similarly to what has been done in MB-pol. To this end we have investigated a large set of DFT methods in addition to CCSD(T) with different correlation-consistent Dunning basis sets[82] in order to gain insight into the errors afforded by each method and to make an informed choice about the computationally cheapest approach that meets our target accuracy.

All single point energy calculations for these benchmarks and the final training sets were performed with Psi4 version 1.4a3.[94, 95] Interaction energies were corrected for the basis set superposition error (BSSE) by using the counterpoise approach.[96] Density fitting (DF) as implemented in Psi4 was employed for the DFT calculations. The frozen natural orbitals (FNO) method[97] in combination with density fitting[98] was employed for the CCSD(T) calculations. FNO-DF-CCSD(T) has been shown to significantly speed up the calculations with negligible errors.[97, 98] In our own benchmarks, we found that monomer deformation energies and 2B and 3B interaction

energies are reproduced with errors below 0.01 kcal/mol with respect to regular CCSD(T) calculations. All valence electrons have been correlated in the CCSD(T) calculations (frozen core approximation). CBS extrapolation of the FNO-DF-CCSD(T) energies was performed separately for the Hartree-Fock (HF) energy using an exponential extrapolation and for the correlation energy using a power form:[99, 100]

$$E_{\mathrm{HF}}^{(X)} = E_{\mathrm{HF}}^{(\infty)} + E_A e^{-\alpha X} \quad (12)$$

$$E_{\mathrm{corr}}^{(X)} = E_{\mathrm{corr}}^{(\infty)} + E_B X^{-\beta X} \quad (13)$$

Here, X is the cardinal number of the basis set. The extrapolated energies $E_{HF}^{(\infty)}$ and $E_{corr}^{(\infty)}$ can be computed from the energies at two consecutive basis sets given values for α and β. We have used a global value of $\alpha = 1.63$[99] for all cardinal numbers, the optimized values $\beta = 2.51$ for 2/3 extrapolation (double- and triple-zeta basis sets) and $\beta = 3.05$ for 3/4 extrapolation (triple- and quadruple-zeta basis sets), and the "ideal" value $\beta = 3.00$ for 4/5 extrapolation (quadruple- and quintuple-zeta basis sets).[100]

For the benchmarks illustrated in Figure 1, 10 monomers, 10 dimers, and 10 trimers have been selected from the *ab initio* MD simulations performed in Ref. [5]. A limited number of structures was considered due to the substantial computational cost associated with the FNO-DF-CCSD(T) calculations using the larger basis sets. In Figure 1, the mean absolute error (MAE) and its standard deviation is shown for 1B deformation energies and the 2B and 3B interaction energies with respect to the best FNO-DF-CCSD(T) extrapolation that has been performed. Table S3 in the Supporting Information contains additional data from DFT calculations with different basis sets, mean unsigned errors (MUE), and the average relative errors (ratio between the error and the reference energy).

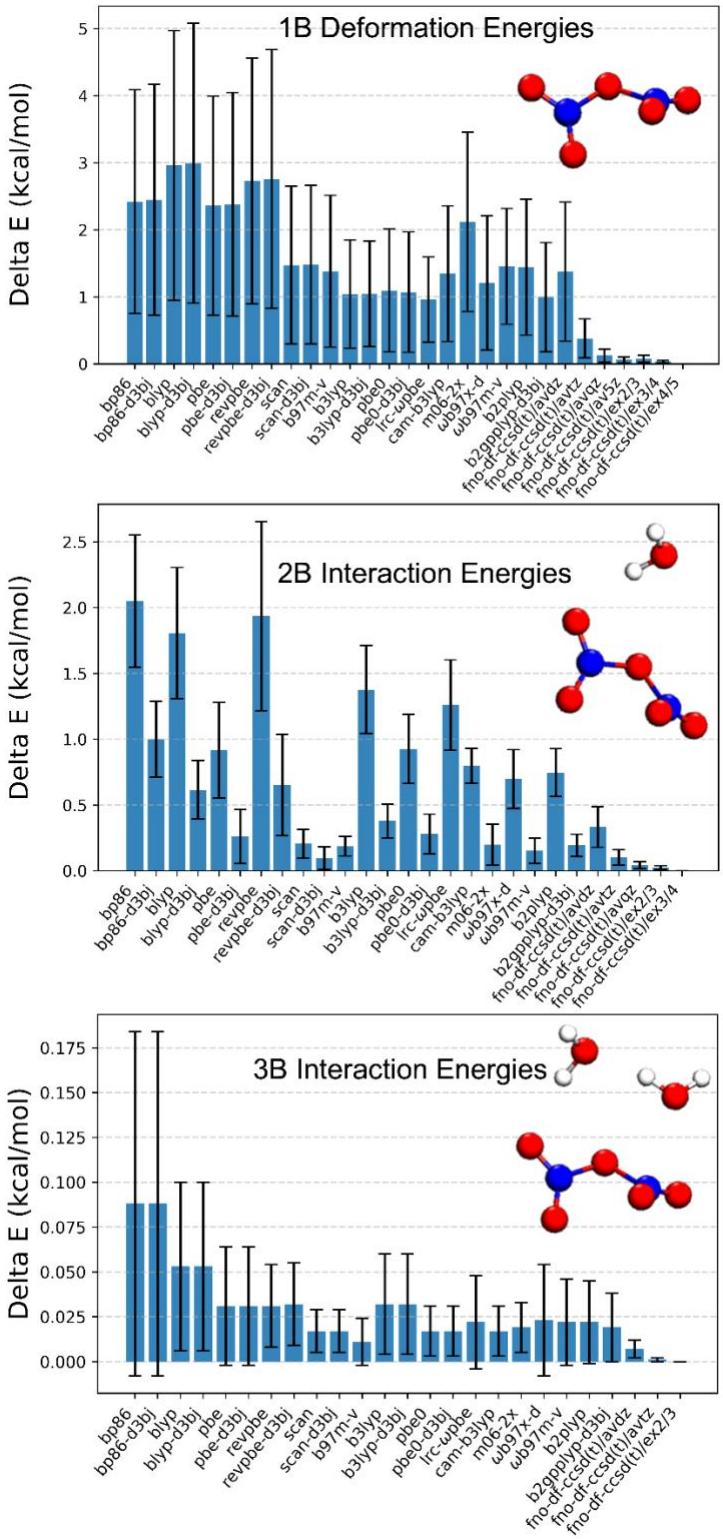

Figure 1. Mean absolute error (MAE) and its standard deviation for one-body deformation energies (top panel), and two-body (middle panel) and three-body (bottom panel) interaction energies with respect to the best CBS extrapolated FNO-DF-CCSD(T) reference data for 10

different structures of N_2O_5 monomers, N_2O_5 + H_2O dimers, and N_2O_5 + 2 H_2O trimers. All DFT calculations were performed with the aug-cc-pV5Z basis sets. aVXZ = aug-cc-pVXZ. exA/B = A/B CBS extrapolation.

As can be seen in Figure 1, the errors with respect to the reference FNO-DF-CCSD(T)/CBS level are overall higher for 1B deformation energies than for 2B and 3B interaction energies, and higher for 2B than for 3B interaction energies, correlating with the magnitude of the many-body energies (see Table S3). For 1B deformation energies, all explored DFT methods exhibit MAEs that exceed 0.95 kcal/mol, similarly to FNO-DF-CCSD(T) with the aug-cc-pVDZ basis set. For 2B interaction energies, a few DFT functionals like SCAN-D3BJ and ωB97M-V have a significantly smaller error than others; however, as shown in Table S3, this error is still relatively high in comparison with the magnitude of the reference interaction energies. The 3B interaction energies are smaller in magnitude than the 2B interactions energies, and most DFT methods exhibit MAEs smaller than 0.1 kcal/mol with B97M-V showing the best performance. Overall, SCAN, SCAN-D3BJ, and B97M-V are the DFT functionals with the best performance in predicting *n*B energies.

Having a balance between accuracy and efficiency in mind, Figure 1 and Table S3 allows us to choose FNO-DF-CCSD(T)/CBS with 2/3 extrapolation for both 1B distortion energies and 2B interaction energies, and FNO-DF-CCSD(T)/aug-cc-pVDZ for 3B interaction energies. As Table S3 shows, the errors corresponding to these choices are minimal in comparison to the magnitude of the reference values and outperform the DFT results.

RESULTS AND DISCUSSION

Accuracy of MB-nrg and TTM-nrg on training and test sets

Figure 2 shows the correlation plots between the FNO-DF-CCSD(T) reference energies and the MB-nrg 1B, 2B, and 3B energies, including MAE with standard deviation and maximum errors for both training sets and test sets (left column). The figure also shows the errors of the MB-nrg model with respect to the FNO-DF-CCSD(T) reference energies (central column), and, for comparison, correlation plots that show how the TTM-nrg model performs for the MB-nrg training and test sets data. The colors in the figure encode the density of data points ranging from red (low density) to yellow (high density).

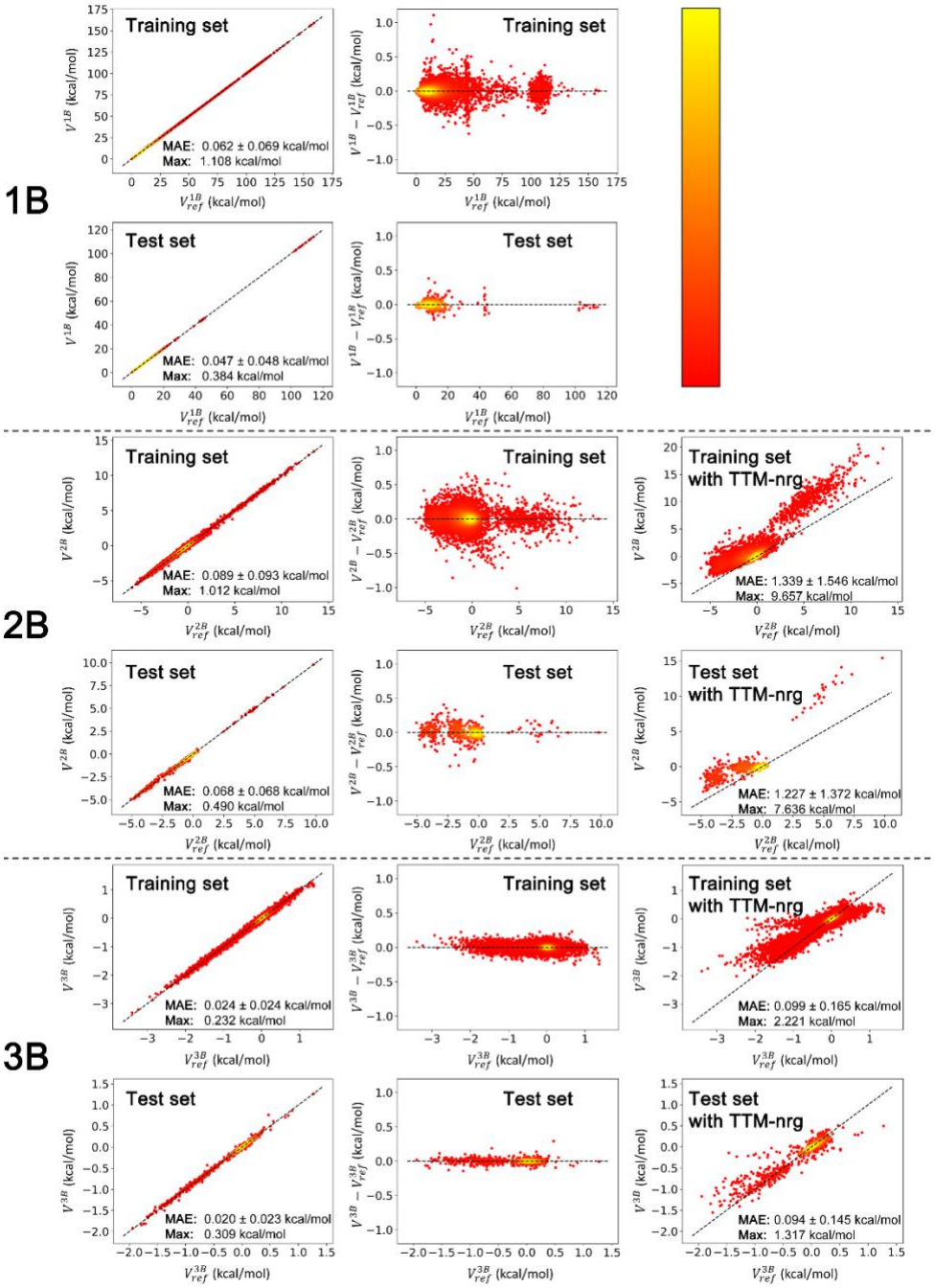

Figure 2. Correlation plots (left column) and errors (central column) between the FNO-DF-CCSD(T) reference data and MB-nrg for the MB-nrg 1B, 2B, and 3B training and test sets. For comparison, correlation plots are presented for TTM-nrg for the same training and test sets (right column). The colors represent the density of data points ranging from low (red) to high (yellow). Also shown are the mean absolute errors (MAE) with standard deviation, and the maximum errors (Max).

By evaluating Figure 2, it is noticeable that the MB-nrg errors are similar for the training and the test sets in all cases. It can also be observed that the PIPs in MB-nrg are very accurate: the MAE with standard deviation and maximum errors for the test sets are, respectively, 0.047±0.048 kcal/mol and 0.384 kcal/mol for 1B energies, 0.068±0.068 kcal/mol and 0.490 kcal/mol for 2B energies, and 0.020±0.023 kcal/mol and 0.309 kcal/mol for 3B energies. On the other hand, the TTM-nrg model exhibits errors that are significantly higher, especially for 2B energies, which demonstrates the limitations of a purely classical description of many-body effects at short range and indicates that a higher number of parameters is needed to accurately describe the 2B and 3B interactions between N_2O_5 and water.

Another interesting point to be observed in Figure 2 is the range of sampled energies in the training and test sets and the location of the majority of the data (see colors). For 3B interactions, it is noticeable that most structures have their interaction energy close to zero, which is an indication that 3B interactions might not be very relevant to describe the behavior of N_2O_5 in water.

Evaluation of multiple rigid scans

In order to further assess the accuracy of the MB-nrg and TTM-nrg models, several rigid scans for monomers, dimers, and trimers have been performed starting from different structures. The 1B, 2B, and 3B energies predicted by our models are compared with the reference level (FNO-DF-CCSD(T)/CBS(extrap.2/3) for 1B and 2B energies, and FNO-DF-CCSD(T)/aug-cc-pVDZ for 3B energies), and with MP2/aug-cc-pVTZ, ωB97XD/aug-cc-pVTZ, and SCAN/aug-cc-pVTZ. These DFT methods have been chosen for this analysis because of their overall performance in the prediction of *n*B energies: as Figure 1 shows, SCAN and ωB97XD have, respectively, good and intermediate performance. A comparison with these DFT functionals is expected to indicate at

least qualitatively how our MB-nrg model performs with respect to state-of-the-art DFT calculations.

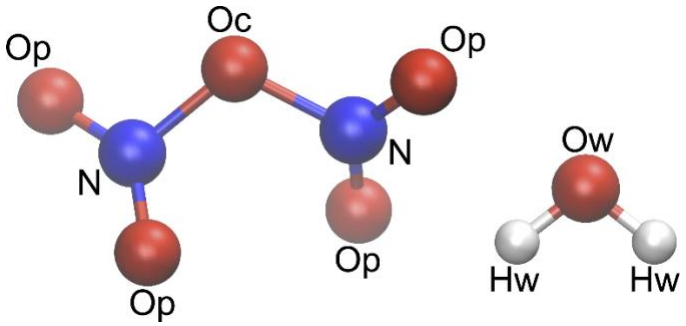

Figure 3. Atom type labels used in this work for the dinitrogen pentoxide (N_2O_5) and water molecules.

Figure 4 shows rigid scans performed for the N_2O_5 monomer, starting from structures with both C_s and C_2 symmetry. The figure presents changes caused by modifying one dihedral angle, one bond angle, and one bond distance.

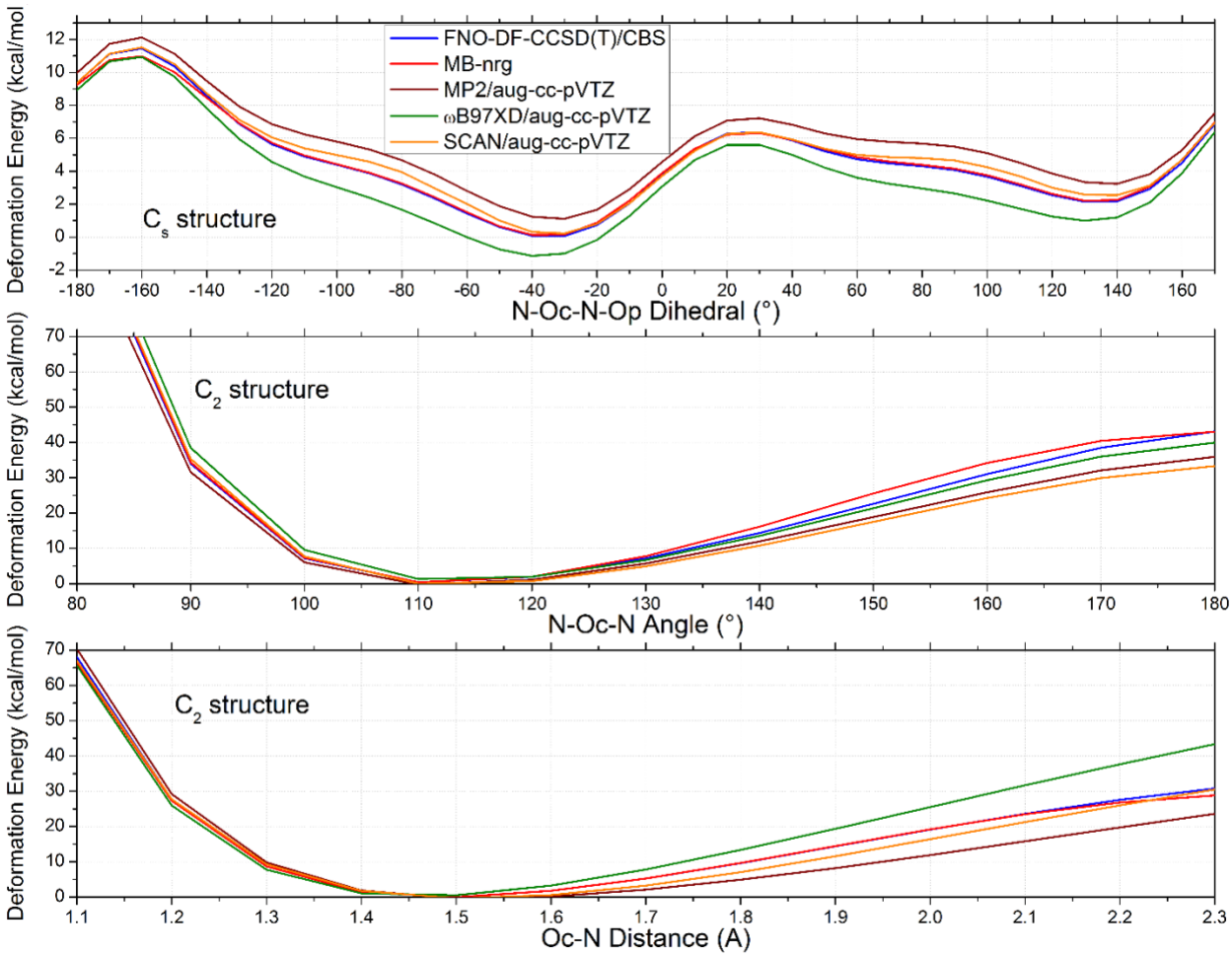

Figure 4. Rigid scans along the Oc-N bond (lower panel), around the N-Oc-N angle (central panel) and around the N-Oc-N-Op dihedral angle (top panel) for the N_2O_5 monomer, starting from structures with C_s and C_2 symmetry. Atomic labels can be found in Figure 3.

All deformation energies shown in Figure 4 were computed with respect to the same reference structure. The figure shows that MB-nrg closely reproduces the FNO-DF-CCSD(T)/CBS reference energies. Notably, MP2, ωB97XD, and SCAN do not reproduce the FNO-DF-CCSD(T)/CBS results as closely as MB-nrg.

Figures 5 and 6 present, respectively, rigid scans performed for dimers and trimers. For comparison, the interaction energies in the scans have also been computed with the polarizable AMOEBA force field.[101] Details about the AMOEBA parametrization are provided in the

Supporting Information. Figures 5 and 6 show the water molecule(s) that move during the scan (in green) and the direction of motion (dotted lines). The positions of the inner and outer cutoffs applied in the switching functions (see Equations 3 and 4) are presented as vertical dashed lines. The cutoffs are applied based on the distance between Oc of N_2O_5 and Ow of water (see Figure 3 and Equations 7 to 9).

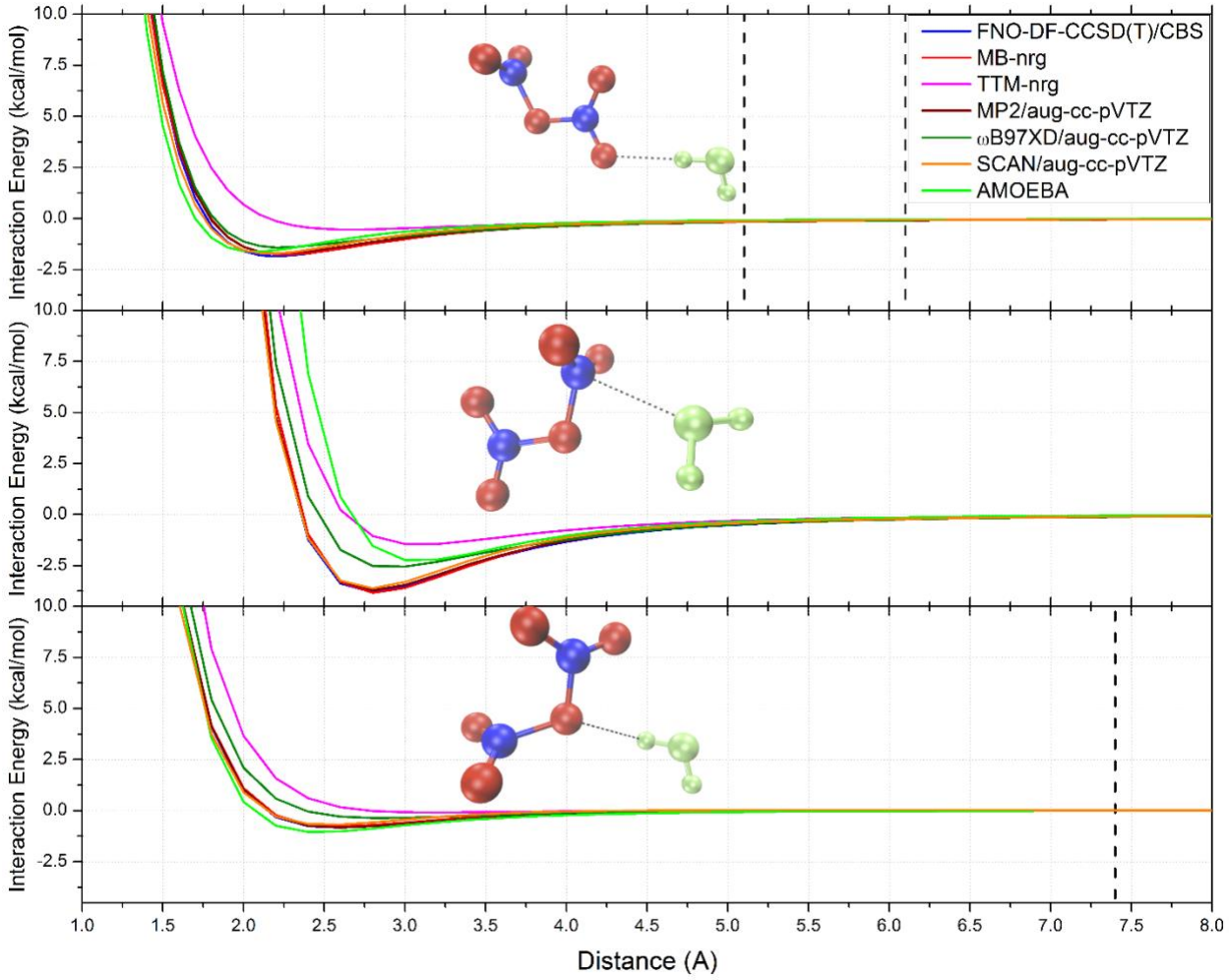

Figure 5. Rigid scans for dimers (N_2O_5 + H_2O). The water in green is the one that moves during the scan, and the dotted lines indicate the sampled distance and the direction of the motion. The vertical dashed lines represent the inner and outer 2B cutoffs.

According to Figure 5, MB-nrg, MP2 and SCAN closely reproduce the 2B interaction energy curves computed with FNO-DF-CCSD(T)/CBS, both at long- and short-range. ωB97XD performs

rather well for the stronger hydrogen bonding interaction (top panel) but exhibits significant deviations in the attractive and repulsive regions of the other two binding curves. AMOEBA is also considerably close to the reference for both hydrogen bonding interactions (top and lower panels) but significantly underestimates bonding between N and Ow atoms (central panel) with an onset of repulsion that is much too early. The TTM-nrg model does not perform well overall, which again indicates that the use of a Born-Mayer repulsion potential (see equation 6) instead of PIPs for 2B interactions is not sufficient for an accurate description in the relevant binding region.

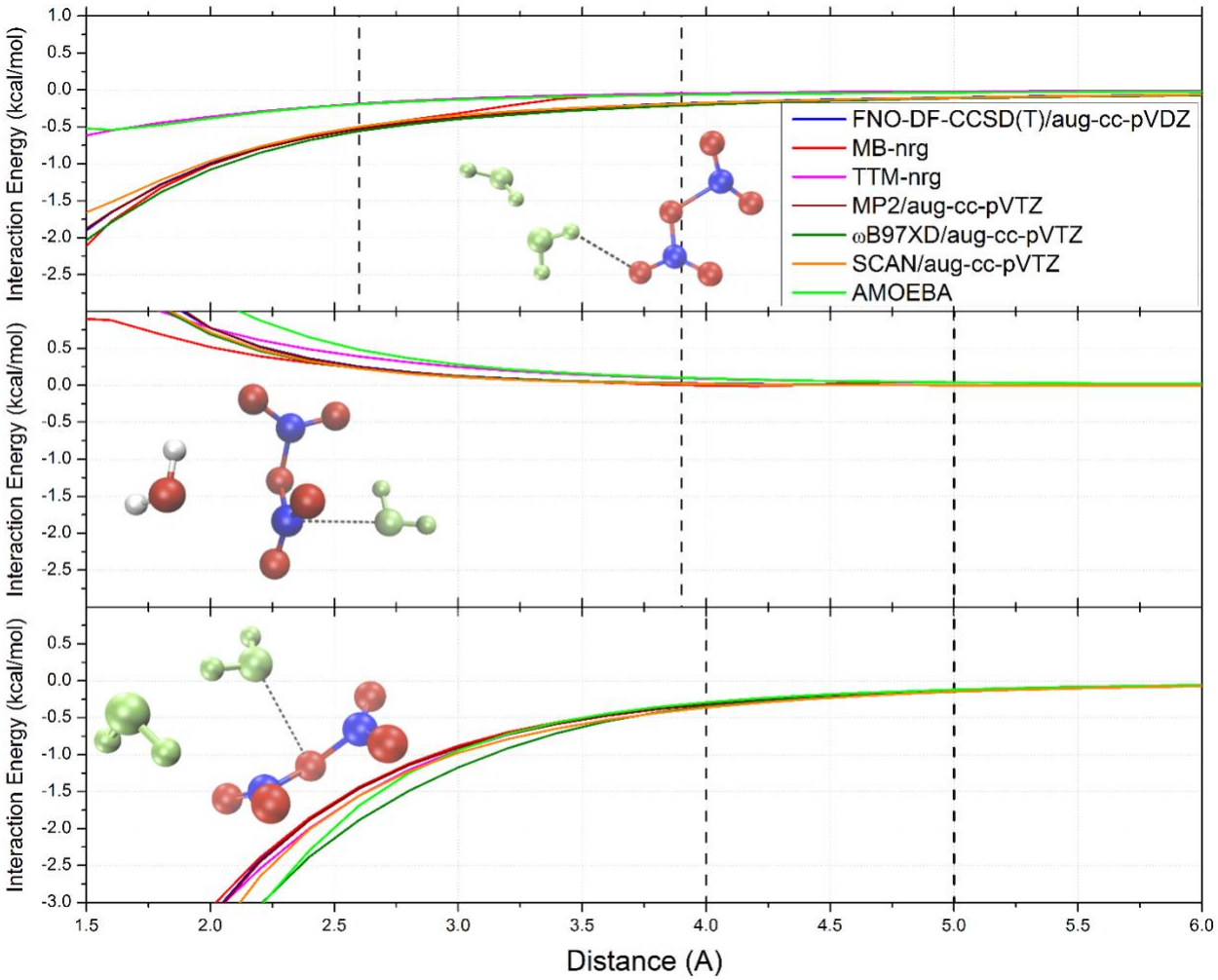

Figure 6. Rigid scans for trimers (N_2O_5 + 2 H_2O). The water molecules depicted in green are the ones that move during the scan, and the dotted lines indicate the sampled distance and the direction of the motion. The vertical dashed lines represent the inner and outer 3B cutoffs.

For the trimer scans, Figure 6 shows that MB-nrg, MP2 and SCAN closely reproduce FNO-DF-CCSD(T)/aug-cc-pVDZ. A disagreement between MB-nrg and the reference is observed for distances below 2.3 Å in the scan in the central panel. Figure S1 shows the 2B interaction energies between the N_2O_5 and the water molecule in green in the scan at the central panel of Figure 6. The figure shows that the 2B potential is very repulsive for the distances below 2.3 Å, hence such structures will never (or very rarely) be visited during MD at ambient condition, which means that our training sets will not contain many samples in this region. It also means that the observed discrepancy is not of relevance for our targeted applications.

Another aspect that can be seen from Figure 6 is in the top panel, which indicates that higher 3B cutoffs could be employed to improve agreement where the underlying classical potential does not yet fully reproduce the reference data, even though the current 3B cutoff values lead to perfect agreement in the other scans. As previously mentioned, the cutoff values have been chosen by having a balance between accuracy and efficiency in mind. Given that the number of trimers (N_2O_5 + 2 H_2O) scales with the square of the number of waters, whereas the number of dimers (N_2O_5 + H_2O) scales linearly with the number of waters, increasing the 2B cutoffs to larger values, as we did, has a smaller impact in the computational efficiency in comparison to increasing the 3B cutoffs.

It is noticeable in Figure 6, that TTM-nrg and AMOEBA but also ωB97XD, albeit to a lesser degree, show significant deviations from the reference 3B energies in at least one of the three scans that are presented. 3B interactions in TTM-nrg are completely represented through the classical polarization model. TTM-nrg performs significantly better for 3B interactions than for 2B interactions, as can be observed by comparing Figures 5 and 6. This observation agrees with the

expectation that the classical potential should be sufficient for the description of induction effects and long-range interactions.

Truncation effects in the explicit short-range corrections

We now evaluate if the inclusion of up to 2B or 3B explicit short-range corrections is sufficient for an accurate description of N_2O_5 in water. For this assessment, we have performed a many-body decomposition of the total energy of clusters of N_2O_5 with up to 4 water molecules and investigated the corresponding *n*B (n = 1 to 5) energies. The interaction energies predicted by the MB-nrg and TTM-nrg models were compared with results obtained with AMOEBA, MP2/aug-cc-pVTZ, ωB97XD/aug-cc-pVTZ, SCAN/aug-cc-pVTZ, and FNO-DF-CCSD(T). The clusters used in this analysis have been randomly selected from an MD trajectory of N_2O_5 in bulk water at ambient conditions performed with MB-nrg. 100 monomers, 100 dimers, 100 trimers, 10 tetramers, and 10 pentamers whose molecules are in closer proximity, with center-of-mass distances smaller than 4 Å, have been selected. FNO-DF-CCSD(T)/CBS(extrap.2/3) has been employed as the reference level for 1B and 2B energies, and FNO-DF-CCSD(T)/aug-cc-pVDZ for 3B, 4B, and 5B energies. The number of tetramers and pentamers had to be limited due to the large computational cost associated with the FNO-DF-CCSD(T) calculations for these systems. Table 1 shows the errors of the investigated methods in reproducing the *n*B energies of the clusters consisting of *n* molecules with respect to the reference data. The table also presents the range of *n*B energy values at the reference level.

Table 1. Mean absolute errors (MAEs) with standard deviation in the *n*B energies with respect to the FNO-DF-CCSD(T) reference energies (REs). Minimum, maximum, and average with standard deviation of the reference energies (REs) are also shown. Values are in kcal/mol. aVTZ refers to the aug-cc-pVTZ basis set. Averages are over 100 structures for monomers, dimers and trimers, and over 10 structures for tetramers and pentamers.

	MAEs in with respect to the reference							REs, FNO-DF-CCSD(T)		
	TTM-nrg	AMOEBA	MB-nrg (1B+2B)	MB-nrg (1B+2B+3B)	MP2/ aVTZ	ωB97XD/ aVTZ	SCAN/ aVTZ	Min.	Max.	Average
N_2O_5	0.031±0.027	N/A	0.031±0.027	0.031±0.027	1.843±1.056	1.196±0.946	1.483±0.676	2.257	40.345	9.193±5.109
$N_2O_5 + H_2O$	0.830±0.648	0.588±0.669	0.147±0.143	0.147±0.143	0.156±0.115	0.707±0.298	0.163±0.078	-3.828	1.528	-1.155±1.098
$N_2O_5 + 2\ H_2O$	0.047±0.042	0.039±0.041	0.047±0.042	0.037±0.035	0.011±0.013	0.036±0.057	0.022±0.028	-0.419	0.27	0.009±0.120
$N_2O_5 + 3\ H_2O$	0.002±0.002	0.002±0.001	0.002±0.002	0.002±0.002	0.001±0.001	0.001±0.001	0.002±0.001	-0.015	0.013	0.002±0.008
$N_2O_5 + 4\ H_2O$	0.001±0.001	0.001±0.001	0.001±0.001	0.001±0.001	0.001±0.001	0.001±0.001	0.001±0.001	-0.005	0.001	-0.001±0.002

By evaluating the right part of Table 1, it can be observed that, as expected, the *n*B energies decrease systematically from 1B to 5B. Noticeably, 4B and 5B energies are essentially zero, which indicates that the inclusion of these interactions may not be important for the description of N_2O_5 in bulk water at ambient conditions.

The left part of Table 1 indicates that the addition of 2B short-range corrections in MB-nrg is important as the results improve significantly in comparison to TTM-nrg; however, the addition of 3B short-range corrections is not as important since the improvement in accuracy is rather small compared to MB-nrg without 3B short-range corrections and TTM-nrg, which describes the 3B energies entirely through induction. Similarly to what has been observed in the previous subsection for the scans, MP2, ωB97XD and SCAN have higher errors than MB-nrg for 1B energies. Also, MP2, SCAN and MB-nrg have similar errors for 2B interaction energies whereas the ωB97XD errors are higher. Additionally, ωB97XD and MB-nrg have similar errors for 3B interaction energies while MP2 and SCAN perform slightly better for the selected clusters. For 4B and 5B interaction energies, TTM-nrg and MB-nrg correctly reproduce the reference energies.

Optimized clusters of N_2O_5 in water

We now evaluate if the conclusions drawn from the analyses of the scans and structures taken from MD trajectories still hold true in optimized clusters. Optimized clusters have been taken from Ref. [5], containing the N_2O_5 molecule in the C_2 symmetry and, respectively, one, two, three, and four water molecules. These structures have been submitted to geometry optimization at the MP2/aug-cc-pVDZ level in Gaussian 09.[92] Afterwards, these structures have been resubmitted to geometry optimization using the truncated Newton linear conjugate gradient method[102, 103] with the MB-nrg and TTM-nrg models. Our goal here is to understand how well the MB-nrg and TTM-nrg optimized geometries correlate with the minima predicted by MP2. Table 2 shows the RMSD values between the structure of each cluster as optimized with MP2/aug-cc-pVDZ versus the optimization with the MB-nrg and TTM-nrg models.

Table 2. RMSD (in Å) of geometries optimized with the MB-nrg and TTM-nrg with respect to geometries optimized with MP2/aug-cc-pVDZ.

	TTM-nrg	MB-nrg (1B+2B)	MB-nrg (1B+2B+3B)
N_2O_5	0.04	0.04	0.04
$N_2O_5 + H_2O$	0.22	0.17	0.17
$N_2O_5 + 2\ H_2O$	0.36	0.12	0.09
$N_2O_5 + 3\ H_2O$	0.22	0.14	0.13
$N_2O_5 + 4\ H_2O$	0.18	0.09	0.09

According to Table 2, the optimized structures predicted by MB-nrg are close to the optimized geometries obtained with MP2/aug-cc-pVDZ, which is an indication of the accuracy of the MB-nrg model if it is assumed that the MP2 optimized structures are close to CCSD(T) optimized geometries. As one would expect, the RMSD values are higher for TTM-nrg than for MB-nrg. Noticeably, the addition of explicit short-range 3B effects does not have a significant effect for the tetramer and the pentamer, but a closer agreement with MP2 is observed for the trimer.

The many-body energies of the clusters optimized at the MP2/aug-cc-pVDZ level have also been evaluated. Table 3 shows the interaction energies predicted for these clusters with MP2/aug-cc-pVTZ, ωB97XD/aug-cc-pVTZ, SCAN/aug-cc-pVTZ, AMOEBA, TTM-nrg and MB-nrg in comparison to FNO-DF-CCSD(T)/CBS(extrap.2/3) for 1B and 2B energies and FNO-DF-CCSD(T)/aug-cc-pVDZ for 3B to 5B energies.

Table 3. *n*B energies (in kcal/mol) with different levels of theory for optimized clusters of N_2O_5 in water obtained at the MP2/aug-cc-pVDZ level. aVTZ refers to the aug-cc-pVTZ basis set.

	TTM-nrg	AMOEBA	MB-nrg (1B+2B)	MB-nrg (1B+2B+3B)	MP2/ aVTZ	ωB97XD/ aVTZ	SCAN/ aVTZ	FNO-DF-CCSD(T)
N_2O_5	1.52	N/A	1.52	1.52	0.02	3.40	0.95	1.49
$N_2O_5 + H_2O$	-1.27	-2.88	-5.49	-5.49	-5.76	-4.36	-5.58	-5.62
$N_2O_5 + 2\ H_2O$	-1.06	-1.49	-1.06	-1.95	-2.03	-2.25	-2.14	-1.96
$N_2O_5 + 3\ H_2O$	-0.20	-0.41	-0.20	-0.20	-0.27	-0.31	-0.33	-0.28
$N_2O_5 + 4\ H_2O$	0.03	0.07	0.03	0.03	0.05	0.07	0.05	0.05

Table 3 indicates that MB-nrg with 1B, 2B, and 3B explicit short-range interactions closely reproduces the reference FNO-DF-CCSD(T) energies for all terms of the MBE for all optimized clusters. For the trimer, the result without explicit 3B short-range corrections does not accurately capture the reference energy. TTM-nrg does not reproduce the 2B interaction energy accurately for the dimer. The trends observed for MP2, ωB97XD, and SCAN are similar to what has been observed for structures from MD trajectories, shown in the previous subsection.

CONCLUSIONS

In this work, we have presented the development and validation of MB-nrg and TTM-nrg many-body potential energy functions for simulations of N_2O_5 in water with high accuracy. These potential energy functions employ MB-pol for water and use the same functional form for the classical part of the potential (permanent electrostatics, polarization, dispersion). While TTM-nrg relies on simple Born-Mayer functions to represent two-body short-range repulsion, MB-nrg instead features an explicit representation of up to three-body short-range interactions in terms of multidimensional permutationally invariant polynomials. Importantly, both MB-nrg and TTM-nrg correctly capture long-range interactions, which is important for extended systems.

We have employed an active learning approach that makes use of neural networks to guarantee an efficient selection of representative structures of monomers (N_2O_5), dimers (N_2O_5 + H_2O), and trimers (N_2O_5 + 2 H_2O) for the training sets used to describe short-range interactions. In order to choose an appropriate electronic structure method for the reference many-body energies of the training sets, we have performed benchmark calculations with different basis sets for a range of DFT methods and coupled cluster theory. In general, DFT with different exchange-correlation functionals exhibits non-uniform accuracy for 1B, 2B, and 3B energies with mean absolute errors exceeding 1 kcal/mol for 1B deformation energies, while only few functionals reach an accuracy of ~ 0.1-0.2 kcal/mol for 2B energies. Our benchmarks indicate that SCAN, SCAN-D3BJ, and B97M-V are the DFT functionals with the best performance in describing *n*B energies between N_2O_5 and water. For our training sets, we have therefore chosen to use FNO-DF-CCSD(T)/CBS with aug-cc-pVDZ and aug-cc-pVTZ basis sets for the extrapolation for 1B and 2B energies and FNO-DF-CCSD(T)/aug-cc-pVDZ for 3B energies. The corresponding errors in the many-body

energies associated with these choices are minimal in comparison to higher order FNO-DF-CCSD(T) calculations, and significantly outperform the DFT calculations.

We have assessed the accuracy of the new TTM-nrg and MB-nrg models by analyzing binding curves as well as optimized geometries and many-body energies of clusters (N_2O_5 + *n* H_2O) with $n \leq 4$. These analyses demonstrate that MB-nrg very closely reproduces the reference data at the coupled cluster level of theory for all individual many-body contributions. This is in contrast to MP2, ωB97XD, and SCAN, which exhibit significant deviations from the reference data for 1B energies and, in the case of ωB97XD, also for 2B energies. As expected, TTM-nrg is less accurate, somewhat comparable to the polarizable AMOEBA force field.

We also observe that the classical potential describes the 3B energies of trimers extracted from molecular dynamics simulations of N_2O_5 in bulk water quite well, the majority of which is close to zero. We conclude that the addition of an explicit representation for 3B short-range interactions is unlikely to play an important role in molecular dynamics simulations of N_2O_5 in bulk water at ambient conditions. However, this is not the case for optimized clusters, where an appropriate representation of explicit short-range 3B interactions has a significant effect on geometries and 3B energies. The 4B and 5B energies of tetramers and pentamers are significantly smaller and well described by MB-nrg and TTM-nrg.

In follow-up publications, we will present results obtained with the MB-nrg model for simulations of N_2O_5 in water clusters, bulk water, and at the water liquid/vapor interface, which are important steps to further the understanding of the reactive uptake of N_2O_5 by aerosol particles.

AUTHOR INFORMATION

Corresponding Authors

* E-mails: vcruzeiro@ucsd.edu, agoetz@sdsc.edu

Notes

The authors declare no competing financial interest.

All data related to this publication can be accessed from the NSF-CAICE Data Repository[104] at https://doi.org/10.6075/J0KK99B6

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/

PDF file containing an analysis of relative energies of C_2 and Cs optimized structures of N_2O_5 by different methods, details about the AMOEBA parametrization, the point charges, polarizabilities, and C_6 coefficients used for N_2O_5, complementary results to Figure 1, and one figure containing 2B interaction energies for structures in the central panel of Figure 6.

ZIP file containing the final 1B, 2B, and 3B training and test sets, structures and energies from the benchmarks, and the geometries of the optimized clusters.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the National Science Foundation (NSF) through the Center for Aerosol Impacts on Chemistry of the Environment (CAICE) under grant number CHE-1801971. This research was partially supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science through Grant No. DE-SC0019490 (F.P). This

work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the NSF under grant number ACI-1053575 (resources at the San Diego Supercomputer Center and the Texas Advanced Computing Center through award TG-CHE130010), as well as resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract DE- AC02-05CH11231, and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC). The authors thank Adrian Roitberg and Justin Smith for providing access to their active learning software and neural network infrastructure, Xiang Gao for insightful discussions, Benny Gerber and Barak Hirshberg for providing their ab initio MD trajectories, and Tim Bertram and Gil Nathanson for stimulating discussions about the role of N_2O_5 in the atmosphere which have motivated this work.

REFERENCES

[1] Dentener, F. J.; Crutzen, P. J. Reaction of N_2O_5 on Tropospheric Aerosols: Impact on the Global Distributions of NO_x, O_3, and OH. *J. Geophys. Res. Atmos.*, **1993**, *98* (D4), 7149–7163.

[2] Evans, M. J.; Jacob, D. J. Impact of New Laboratory Studies of N_2O_5 Hydrolysis on Global Model Budgets of Tropospheric Nitrogen Oxides, Ozone, and OH. *Geophys. Res. Lett.*, **2005**, *32* (9), L09813.

[3] Macintyre, H. L.; Evans, M. J. Sensitivity of a Global Model to the Uptake of N_2O_5 by Tropospheric Aerosol. *Atmos. Chem. Phys.*, **2010**, *10* (15), 7409–7414.

[4] Seinfeld, J. H.; Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 3rd ed.; Wiley, 2016.

[5] Hirshberg, B.; Rossich Molina, E.; Götz, A. W.; Hammerich, A. D.; Nathanson, G. M.;

Bertram, T. H.; Johnson, M. A.; Gerber, R. B. N_2O_5 at Water Surfaces: Binding Forces, Charge Separation, Energy Accommodation and Atmospheric Implications. *Phys. Chem. Chem. Phys.*, **2018**, *20* (26), 17961–17976.

[6] Chang, W. L.; Bhave, P. V.; Brown, S. S.; Riemer, N.; Stutz, J.; Dabdub, D. Heterogeneous Atmospheric Chemistry, Ambient Measurements, and Model Calculations of N_2O_5: A Review. *Aerosol Sci. Technol.*, **2011**, *45* (6), 665–695.

[7] Hanway, D.; Tao, F.-M. A Density Functional Theory and Ab Initio Study of the Hydrolysis of Dinitrogen Pentoxide. *Chem. Phys. Lett.*, **1998**, *285* (5–6), 459–466.

[8] Raff, J. D.; Njegic, B.; Chang, W. L.; Gordon, M. S.; Dabdub, D.; Gerber, R. B.; Finlayson-Pitts, B. J. Chlorine Activation Indoors and Outdoors via Surface-Mediated Reactions of Nitrogen Oxides with Hydrogen Chloride. *Proc. Natl. Acad. Sci.*, **2009**, *106* (33), 13647–13654.

[9] McNamara, J. P.; Hillier, I. H. Exploration of the Atmospheric Reactivity of N_2O_5 and HCl in Small Water Clusters Using Electronic Structure Methods. *Phys. Chem. Chem. Phys.*, **2000**, *2* (11), 2503–2509.

[10] Alecu, I. M.; Marshall, P. Computational Study of the Thermochemistry of N_2O_5 and the Kinetics of the Reaction N2O5 + H2O → 2 HNO3. *J. Phys. Chem. A*, **2014**, *118* (48), 11405–11416.

[11] McNamara, J. P.; Hillier, I. H. Structure and Reactivity of Dinitrogen Pentoxide in Small Water Clusters Studied by Electronic Structure Calculations. *J. Phys. Chem. A*, **2000**, *104* (22), 5307–5319.

[12] Bertram, T. H.; Thornton, J. A. Toward a General Parameterization of N_2O_5 Reactivity on Aqueous Particles: The Competing Effects of Particle Liquid Water, Nitrate and Chloride.

Atmos. Chem. Phys., **2009**, *9* (21), 8351–8363.

[13] McDuffie, E. E.; Fibiger, D. L.; Dubé, W. P.; Lopez-Hilfiker, F.; Lee, B. H.; Thornton, J. A.; Shah, V.; Jaeglé, L.; Guo, H.; Weber, R. J.; Michael Reeves, J.; Weinheimer, A. J.; Schroder, J. C.; Campuzano-Jost, P.; Jimenez, J. L.; Dibb, J. E.; Veres, P.; Ebben, C.; Sparks, T. L.; Wooldridge, P. J.; Cohen, R. C.; Hornbrook, R. S.; Apel, E. C.; Campos, T.; Hall, S. R.; Ullmann, K.; Brown, S. S. Heterogeneous N_2O_5 Uptake During Winter: Aircraft Measurements During the 2015 WINTER Campaign and Critical Evaluation of Current Parameterizations. *J. Geophys. Res. Atmos.*, **2018**, *123* (8), 4345–4372.

[14] Brown, S. S.; Stutz, J. Nighttime Radical Observations and Chemistry. *Chem. Soc. Rev.*, **2012**, *41* (19), 6405.

[15] Abbatt, J. P. D.; Lee, A. K. Y.; Thornton, J. A. Quantifying Trace Gas Uptake to Tropospheric Aerosol: Recent Advances and Remaining Challenges. *Chem. Soc. Rev.*, **2012**, *41* (19), 6555.

[16] Stirling, A.; Pápai, I.; Mink, J.; Salahub, D. R. Density Functional Study of Nitrogen Oxides. *J. Chem. Phys.*, **1994**, *100* (4), 2910–2923.

[17] Jitariu, L. C.; Hirst, D. M. Theoretical Investigation of the $N_2O_5 \rightleftharpoons NO_2$+$NO_3$ Equilibrium by Density Functional Theory and *Ab Initio* Calculations. *Phys. Chem. Chem. Phys.*, **2000**, *2* (4), 847–852.

[18] Grabow, J. -U.; Andrews, A. M.; Fraser, G. T.; Irikura, K. K.; Suenram, R. D.; Lovas, F. J.; Lafferty, W. J.; Domenech, J. L. Microwave Spectrum, Large-amplitude Motions, and *Ab Initio* Calculations for N_2O_5. *J. Chem. Phys.*, **1996**, *105* (17), 7249–7262.

[19] Snyder, J. A.; Hanway, D.; Mendez, J.; Jamka, A. J.; Tao, F.-M. A Density Functional Theory Study of the Gas-Phase Hydrolysis of Dinitrogen Pentoxide. *J. Phys. Chem. A*, **1999**,

103 (46), 9355–9358.

[20] Bianco, R.; Hynes, J. T. Theoretical Studies of Heterogeneous Reaction Mechanisms Relevant for Stratospheric Ozone Depletion. *Int. J. Quantum Chem.*, **1999**, *75* (4–5), 683–692.

[21] Voegele, A. F.; Tautermann, C. S.; Loerting, T.; Liedl, K. R. Toward Elimination of Discrepancies between Theory and Experiment: The Gas-Phase Reaction of N_2O_5 with H_2O. *Phys. Chem. Chem. Phys.*, **2003**, *5* (3), 487–495.

[22] Gržinić, G.; Bartels-Rausch, T.; Türler, A.; Ammann, M. Efficient Bulk Mass Accommodation and Dissociation of N_2O_5 in Neutral Aqueous Aerosol. *Atmos. Chem. Phys.*, **2017**, *17* (10), 6493–6502.

[23] Li, W.; Pak, C. Y.; Tse, Y.-L. S. Free Energy Study of H_2O, N_2O_5, SO_2, and O_3 Gas Sorption by Water Droplets/Slabs. *J. Chem. Phys.*, **2018**, *148* (16), 164706.

[24] Iftimie, R.; Minary, P.; Tuckerman, M. E. *Ab Initio* Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc. Natl. Acad. Sci.*, **2005**, *102* (19), 6654–6659.

[25] Tuckerman, M. E. *Ab Initio* Molecular Dynamics: Basic Concepts, Current Trends and Novel Applications. *J. Phys. Condens. Matter*, **2002**, *14* (50), R1297–R1355.

[26] Riera, M.; Lambros, E.; Nguyen, T. T.; Götz, A. W.; Paesani, F. Low-Order Many-Body Interactions Determine the Local Structure of Liquid Water. *Chem. Sci.*, **2019**, *10* (35), 8211–8218.

[27] Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature*, **2018**, *559* (7715), 547–555.

[28] Das Sarma, S.; Deng, D.-L.; Duan, L.-M. Machine Learning Meets Quantum Physics. *Phys. Today*, **2019**, *72* (3), 48–54.

[29] Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying Machine Learning and Quantum Chemistry with a Deep Neural Network for Molecular Wavefunctions. *Nat. Commun.*, **2019**, *10* (1), 5024.

[30] Cova, T. F. G. G.; Pais, A. A. C. C. Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. *Front. Chem.*, **2019**, *7*.

[31] Smith, J. S.; Roitberg, A. E.; Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med. Chem. Lett.*, **2018**, *9* (11), 1065–1069.

[32] Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.*, **2007**, *98* (14), 146401.

[33] Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.*, **2017**, *8* (4), 3192–3203.

[34] Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.*, **2019**, *5* (8), eaav6490.

[35] Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.*, **2018**, *120* (14), 143001.

[36] Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.*, **2019**, *10* (1), 2903.

[37] Paesani, F. Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions. *Acc. Chem. Res.*, **2016**,

49 (9), 1844–1851.

[38] Cirera, J.; Sung, J. C.; Howland, P. B.; Paesani, F. The Effects of Electronic Polarization on Water Adsorption in Metal-Organic Frameworks: H_2O in MIL-53(Cr). *J. Chem. Phys.*, **2012**, *137* (5), 054704.

[39] Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters. *J. Chem. Theory Comput.*, **2007**, *3*, 46.

[40] Richard, R. M.; Herbert, J. M. A Generalized Many-Body Expansion and a Unified View of Fragment-Based Methods in Electronic Structure Theory. *J. Chem. Phys.*, **2012**, *137* (6), 064113.

[41] Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.*, **2013**, *9* (12), 5395–5403.

[42] Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory Comput.*, **2014**, *10* (4), 1599–1607.

[43] Medders, G. R.; Babin, V.; Paesani, F. Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.*, **2014**, *10* (8), 2906–2910.

[44] Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.*, **2009**, *28* (4), 577–606.

[45] Qu, C.; Yu, Q.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces. *Annu. Rev. Phys. Chem.*, **2018**, *69* (1), 151–175.

[46] Brown, S. E. From *Ab Initio* Data to High-Dimensional Potential Energy Surfaces: A

Critical Overview and Assessment of the Development of Permutationally Invariant Polynomial Potential Energy Surfaces for Single Molecules. *J. Chem. Phys.*, **2019**, *151* (19), 194111.

[47] Marquardt, R.; Quack, M. Global Analytical Potential Hypersurfaces for Large Amplitude Nuclear Motion and Reactions in Methane. I. Formulation of the Potentials and Adjustment of Parameters to *Ab Initio* Data and Experimental Constraints. *J. Chem. Phys.*, **1998**, *109* (24), 10628–10643.

[48] Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions through Many-Body Expansions. *J. Chem. Phys.*, **2018**, *148* (24), 241725.

[49] Herbert, J. M. Fantasy versus Reality in Fragment-Based Quantum Chemistry. *J. Chem. Phys.*, **2019**, *151* (17), 170901.

[50] Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.*, **2015**, *115* (12), 5607–5642.

[51] Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. The Combined Fragmentation and Systematic Molecular Fragmentation Methods. *Acc. Chem. Res.*, **2014**, *47* (9), 2776–2785.

[52] Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the Many-Body Expansion for Large Systems. I. Precision Considerations. *J. Chem. Phys.*, **2014**, *141* (1), 014108.

[53] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.*, **2012**, *112* (1), 632–672.

[54] Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From

Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.*, **2016**, *116* (13), 7501–7528.

[55] Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the Accuracy of the MB-Pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid Water and Ice. *J. Chem. Phys.*, **2016**, *145* (19), 194504.

[56] Moberg, D. R.; Becker, D.; Dierking, C. W.; Zurheide, F.; Bandow, B.; Buck, U.; Hudait, A.; Molinero, V.; Paesani, F.; Zeuch, T. The End of Ice I. *Proc. Natl. Acad. Sci.*, **2019**, *116* (49), 24413–24419.

[57] Paesani, F.; Bajaj, P.; Riera, M. Chemical Accuracy in Modeling Halide Ion Hydration from Many-Body Representations. *Adv. Phys. X*, **2019**, *4* (1), 1631212.

[58] Riera, M.; Mardirossian, N.; Bajaj, P.; Götz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion–Water Interactions through Many-Body Representations. Alkali-Water Dimer Potential Energy Surfaces. *J. Chem. Phys.*, **2017**, *147* (16), 161715.

[59] Bajaj, P.; Götz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion–Water Interactions through Many-Body Representations. I. Halide–Water Dimer Potential Energy Surfaces. *J. Chem. Theory Comput.*, **2016**, *12* (6), 2698–2705.

[60] Riera, M.; Yeh, E. P.; Paesani, F. Data-Driven Many-Body Models for Molecular Fluids: CO_2/H_2O Mixtures as a Case Study. *J. Chem. Theory Comput.*, **2020**, acs.jctc.9b01175.

[61] Riera, M.; Hirales, A.; Ghosh, R.; Paesani, F. Data-Driven Many-Body Models with Chemical Accuracy for CH 4 /H 2 O Mixtures. *J. Phys. Chem. B*, **2020**, *124* (49), 11207–11221.

[62] Egan, C. K.; Paesani, F. Assessing Many-Body Effects of Water Self-Ions. II: $H_3O^+(H_2O)_n$ Clusters. *J. Chem. Theory Comput.*, **2019**, *15* (9), 4816–4833.

[63] Bajaj, P.; Zhuang, D.; Paesani, F. Specific Ion Effects on Hydrogen-Bond Rearrangements in the Halide–Dihydrate Complexes. *J. Phys. Chem. Lett.*, **2019**, *10* (11), 2823–2828.

[64] Bajaj, P.; Riera, M.; Lin, J. K.; Mendoza Montijo, Y. E.; Gazca, J.; Paesani, F. Halide Ion Microhydration: Structure, Energetics, and Spectroscopy of Small Halide–Water Clusters. *J. Phys. Chem. A*, **2019**, *123* (13), 2843–2852.

[65] Bajaj, P.; Richardson, J. O.; Paesani, F. Ion-Mediated Hydrogen-Bond Rearrangement through Tunnelling in the Iodide–Dihydrate Complex. *Nat. Chem.*, **2019**, *11* (4), 367–374.

[66] Zhuang, D.; Riera, M.; Schenter, G. K.; Fulton, J. L.; Paesani, F. Many-Body Effects Determine the Local Hydration Structure of Cs^+ in Solution. *J. Phys. Chem. Lett.*, **2019**, *10* (3), 406–412.

[67] Riera, M.; Brown, S. E.; Paesani, F. Isomeric Equilibria, Nuclear Quantum Effects, and Vibrational Spectra of $M^+(H_2O)n$ =1–3 Clusters, with M = Li, Na, K, Rb, and Cs, through Many-Body Representations. *J. Phys. Chem. A*, **2018**, *122* (27), 5811–5821.

[68] Born, M.; Mayer, J. E. Zur Gittertheorie Der Ionenkristalle. *Zeitschrift für Phys.*, **1932**, *75* (1–2), 1–18.

[69] Buckingham, R. A. The Classical Equation of State of Gaseous Helium, Neon and Argon. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, **1938**, *168* (933), 264–283.

[70] Arismendi-Arrieta, D. J.; Riera, M.; Bajaj, P.; Prosmiti, R.; Paesani, F. I-TTM Model for Ab Initio-Based Ion–Water Interaction Potentials. 1. Halide–Water Potential Energy Functions. *J. Phys. Chem. B*, **2016**, *120* (8), 1822–1832.

[71] Riera, M.; Götz, A. W.; Paesani, F. The I-TTM Model for Ab Initio-Based Ion–Water

Interaction Potentials. II. Alkali Metal Ion–Water Potential Energy Functions. *Phys. Chem. Chem. Phys.*, **2016**, *18* (44), 30334–30343.

[72] Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The Vibrational Proton Potential in Bulk Liquid Water and Ice. *J. Chem. Phys.*, **2008**, *128* (15), 154519.

[73] Burnham, C. J.; Xantheas, S. S. Development of Transferable Interaction Models for Water. IV. A Flexible, All-Atom Polarizable Potential (TTM2-F) Based on Geometry Dependent Charges Derived from an Ab Initio Monomer Dipole Moment Surface. *J. Chem. Phys.*, **2002**, *116* (12), 5115.

[74] Fanourgakis, G. S.; Xantheas, S. S. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-Type Model Potential (TTM3-F, v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.*, **2008**, *128* (7), 074506.

[75] Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. The Parametrization of a Thole-Type All-Atom Polarizable Water Model from First Principles and Its Application to the Study of Water Clusters (n =2–21) and the Phonon Spectrum of Ice Ih. *J. Chem. Phys.*, **1999**, *110* (9), 4566–4581.

[76] Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.*, **1981**, *59* (3), 341–350.

[77] Partridge, H.; Schwenke, D. W. The Determination of an Accurate Isotope Dependent Potential Energy Surface for Water from Extensive *Ab Initio* Calculations and Experimental Data. *J. Chem. Phys.*, **1997**, *106* (11), 4618–4639.

[78] Tang, K. T.; Toennies, J. P. An Improved Simple Model for the van Der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.*,

1984, *80* (8), 3726–3741.

[79] Migliorati, V.; Serva, A.; Terenzio, F. M.; D'Angelo, P. Development of Lennard-Jones and Buckingham Potentials for Lanthanoid Ions in Water. *Inorg. Chem.*, **2017**, *56* (11), 6214–6224.

[80] Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.*, **2018**, *148* (24), 241733.

[81] Mardirossian, N.; Head-Gordon, M. ωB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.*, **2016**, *144* (21), 214110.

[82] Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.*, **1989**, *90* (2), 1007–1023.

[83] Kendall, R. A.; Jr., Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.*, **1992**, *96* (9), 6796.

[84] Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.*, **1990**, *11* (3), 361–373.

[85] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang,

C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.*, **2015**, *113* (2), 184–215.

[86] Schwerdtfeger, P.; Nagle, J. K. 2018 Table of Static Dipole Polarizabilities of the Neutral Elements in the Periodic Table. *Mol. Phys.*, **2019**, *117* (9–12), 1200–1225.

[87] Brinck, T.; Murray, J. S.; Politzer, P. Polarizability and Volume. *J. Chem. Phys.*, **1993**, *98* (5), 4305–4306.

[88] Becke, A. D.; Johnson, E. R. A Unified Density-Functional Treatment of Dynamical, Nondynamical, and Dispersion Correlations. *J. Chem. Phys.*, **2007**, *127* (12), 124108.

[89] Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction. *J. Chem. Phys.*, **2005**, *122* (15), 154104.

[90] Johnson, E. R.; Becke, A. D. A Post-Hartree–Fock Model of Intermolecular Interactions. *J. Chem. Phys.*, **2005**, *123* (2), 024101.

[91] MBX GitHub Page. https://github.com/paesanilab/MBX.

[92] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V; Cioslowski, J.; and D. J. Fox. Gaussian 09 Revision A.02. Wallingford 2009.

[93] Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data*, **2017**, *4* (1), 170193.

[94] Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.*, **2017**, *13* (7), 3185–3197.

[95] Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: An Open-Source Ab Initio Electronic Structure Program. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2012**, *2* (4), 556–565.

[96] Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.*, **1970**, *19* (4), 553–566.

[97] DePrince, A. E.; Sherrill, C. D. Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals. *J. Chem. Theory Comput.*, **2013**, *9* (1), 293–299.

[98] DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t_1-Transformed Hamiltonian. *J. Chem. Theory Comput.*, **2013**, *9* (6), 2687–2696.

[99] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree–Fock Calculations. *Chem. Phys. Lett.*, **1999**, *302* (5–6),

437–446.

[100] Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated Ab Initio Methods? *J. Chem. Theory Comput.*, **2011**, *7* (1), 33–43.

[101] Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B*, **2015**, *119* (29), 9423–9437.

[102] Nash, S. G. Preconditioning of Truncated-Newton Methods. *SIAM J. Sci. Stat. Comput.*, **1985**, *6* (3), 599–616.

[103] Nash, S. G.; Nocedal, J. A Numerical Study of the Limited Memory BFGS Method and the Truncated-Newton Method for Large Scale Optimization. *SIAM J. Optim.*, **1991**, *1* (3), 358–372.

[104] Cruzeiro, V. W. D.; Lambros, E.; Riera, M.; Roy, R.; Paesani, F.; Götz, A. W. Data from: Highly Accurate Many-Body Potentials for Simulations of N_2O_5 in Water: Benchmarks, Development and Validation, Center for Aerosol Impacts on Chemistry of the Environment (CAICE). UC San Diego Library Digital Collections, 2021. https://doi.org/10.6075/J0KK99B6.