# General many-body framework for data-driven potentials with arbitrary quantum mechanical accuracy: Water as a case study

Eleftherios Lambros,*,†,§ Saswata Dasgupta,*,†,§ Etienne Palos,† Steven Swee,† Jie Hu,† and Francesco Paesani*,†,‡,¶

†*Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, United States*

‡*Materials Science and Engineering, University of California San Diego, La Jolla, California 92093, United States*

¶*San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, United States*

§*Contributed equally to this work*

E-mail: elambros@ucsd.edu; s1dasgupta@ucsd.edu; fpaesani@ucsd.edu

**Abstract**

We present a general framework for the development of data-driven many-body (MB) potential energy functions (MB-QM PEFs) that represent the interactions between small molecules at an arbitrary quantum-mechanical (QM) level of theory. As a demonstration, a family of MB-QM PEFs for water are rigorously derived from density functionals belonging to different rungs across Jacob's ladder of approximations within density functional theory (MB-DFT) as well as from Møller-Plesset perturbation theory (MB-MP2). Through a systematic analysis of individual many-body contributions to the interaction energies of water clusters, we

demonstrate that all MB-QM PEFs preserve the same accuracy as the corresponding *ab initio* calculations, with the exception of those derived from density functionals within the generalized gradient approximation (GGA). The differences between the DFT and MB-DFT results are traced back to density-driven errors that prevent GGA functionals from accurately representing the underlying molecular interactions for different cluster sizes and hydrogen-bonding arrangements. We show that this shortcoming may be overcome, within the many-body formalism, by using density-corrected functionals that provide a more consistent representation of each individual many-body contribution. This is demonstrated through the development of a MB-DFT PEF derived from density-corrected PBE-D3 data, which more accurately reproduce the corresponding *ab initio* results.

# 1 Introduction

Molecular mechanics (MM) models, with tunable parameters, are the workhorse of computer simulations. While the earliest MM models were parameterized using simple point-charge and pairwise potentials to reproduce experimental observables, recent MM models are constructed with first-principles approaches using parameters derived from high level *ab initio* reference data. Polarizable force fields (FFs)[1–3] and machine learning (ML) models[4–14] have now taken the center stage as the models of choice for molecular dynamics (MD) and Monte Carlo (MC) simulations.

Among ML models, neural networks (NNs) have become increasingly popular in computational molecular sciences. Atomistic NNs describe the target multidimensional potential energy surface (PES) using a set of descriptors that represent the immediate local environment around each atom.[4,5,8,9] NN models rely on using regression algorithms to train flexible potential energy functions (PEFs) on large sets of reference data that are calculated at the highest level of theory compatible with the system's size and complexity.[4,5,8,9,15–17] While the restrictions on the functional forms used to represent the target PES are somewhat loose, ML models must satisfy rotational, translational, and permutational invariance, and must be able to uniquely describe a molecular configuration.[4,5,9,18–24] Some ML models are trained on gas-phase reference data and,

while useful for studying individual molecules and small clusters, are not adequate to accurately describe systems in the condensed phase where long-range many-body effects may play important roles.[25] On the other hand, since gas-phase reference data are generated for small systems, they are not only relatively cheap and fast to calculate, but can also be computed using higher levels of theory compared to models trained on large condensed-phase systems.[12] For small molecules the reference data are usually calculated using coupled cluster (CC) theory, including single, double, and perturbative triple excitations, i.e., CCSD(T), often in the complete basis set (CBS) limit, currently the "gold standard" for molecular interactions. Towards this end, some ML approaches, which include Δ-machine learning procedures applied to permutationally invariant polynomials (PIPs),[24] allow for a ML model to be trained to an effective higher level of theory by training a correction to a core potential generated at a lower level of theory, using a sparser number of higher-level data points.[26,27] ML approaches based on NNs and PIPs have also been developed to model chemical reactions in the gas-phase and at solid surfaces.[19–22,28–32] In order to implicitly include many-body effects, some ML models are trained on entire sets of condensed-phase configurations, which are expensive to generate and must consequently use a lower level of theory, commonly density functional theory (DFT), for the reference data.[33–35]

An alternative way to represent condensed-phase systems using ML approaches is to adopt a hybrid data-driven/physics-based scheme where a data-driven model, which captures (short-range) quantum-mechanical interactions, is integrated with a physics-based model of many-body interactions, which are represented by classical expressions.[36–38] Examples of hybrid data-driven/physics-based models are the MB-pol and related MB-nrg PEFs that are able to accurately predict the properties of water[39–42] and various aqueous systems,[43–50] as well as molecular fluids,[51,52] from the gas to condensed phases. Both MB-pol and MB-nrg PEFs are rigorously derived from the many-body expansion (MBE) of the energy and use PIPs[24] to capture short-range quantum-mechanical effects arising from the overlap of the electron densities of individual monomers (e.g., Pauli repulsion, and charge transfer and penetration), which are missing in conventional force fields. The MB-DFT PEFs generalize the MBE formalism adopted to develop the MB-pol PEF of water by replacing the

2-body (2B) and 3-body (3B) terms, which were originally calculated at the CCSD(T)/CBS level of theory, with corresponding terms calculated using an arbitrary density functional.[53] It has been shown that the MB-DFT PEFs closely reproduce the structural properties of liquid water calculated from fully *ab initio* MD simulations carried out with the same density functional.[53]

The MB-DFT family of hybrid data-driven/physics-based PEFs is particularly appealing for quantum mechanics/molecular mechanics (QM/MM) simulations of systems that are either too computationally expensive to be treated at a fully quantum-mechanical (QM) level or cannot be described using molecular mechanics (MM) models since they involve rearrangements of chemical bonds. Likewise, while ML approaches can, by construction, model bond rearrangements, their behavior is highly dependent on the composition the datasets and level of theory used in the training process. In principle, these problems do not affect QM/MM simulations where the total system is divided into a smaller QM subsystem, which includes all molecular species that are involved in the chemical transformation, and a MM region, which describes environmental effects.[54,55] However, conventional QM/MM approaches suffer from energy discontinuities at the boundary between the QM and the MM region which appear due to the different accuracy of the QM and MM models in representing the underlying molecular interactions. Another shortcoming of conventional QM/MM calculations arises when they are applied to studying chemical reactions in solutions where, due to diffusion, molecules initially assigned to the MM region may enter the QM region and, vice versa, molecules initially assigned to the QM region may enter the MM region during the MD simulation. To overcome this problem, several adaptive QM/MM schemes have been proposed where the QM region is dynamically repartitioned at every MD step in order to prevent diffusive breakup. This is generally accomplished by introducing transition layers that smooth over the QM/MM boundary discontinuity.[56,57,57,58,58–76] This suggests that "elevating" the accuracy of the MM model to the same level as the QM model would effectively remove these discontinuities and significantly improve the realism and predictive power of QM/MM calculations.[77]

In this context, pure ML models of the MM region do not lend themselves well to QM/MM simulations since point charges must, at least, be assigned to the MM atoms to capture Coulom-

bic interactions between the MM region and the QM electronic density through a minimal electrostatic embedding scheme.[78] In contrast, MB-DFT PEFs[53] and other polarizable models like AMOEBA[79,80] can directly couple to the QM region using a polarizable embedding scheme which allows the environment to dynamically respond to the QM density via the polarization of inducible multipole moments on the MM atoms.[77,81–89] In principle, when used in QM/MM calculations, the MB-DFT PEFs can not only elevate the accuracy of the MM region to be consistent with that of the QM region but, by construction, their 2B and 3B PIPs help recover quantum-mechanical effects in the QM/MM interaction which are missing in QM/MM calculations with conventional (either polarizable or nonpolarizable) force fields, thus allowing the entire QM/MM system to be consistently treated at the same level of theory.[77] However, in the case of water, the original MB-DFT PEFs[44] were built upon the same Thole-type scheme adopted by MB-pol to represent permanent electrostatics and polarization derived from high-level QM calculations,[39–41] which implies that they are not able to strictly reproduce many-body energies calculated *ab initio* using the corresponding DFT models. This is a problem in QM/MM calculations in water where the MM region is described by a MB-DFT PEF and the QM region by the corresponding DFT model because the associated QM/MM electrostatic interactions may be substantially different from those predicted by the original MB-pol electrostatic model underlying the MB-DFT PEFs, which results in an unbalanced representation of short-range many-body effects between the QM and MM regions.[77] This problem is particularly evident in configurations with strong polarization. For example, we observed that when the hydrogen-bond donor in the water dimer is assigned to the QM region described by a DFT model and the hydrogen-bond acceptor is described by the corresponding MB-DFT PEF, the underlying QM/MM electrostatic interactions are very similar to the pure MM electrostatic interactions provided by the MB-DFT PEF. This allows the 2B and 3B PIPs of the MB-DFT PEFs to correctly recover the reference QM energy of the water dimer. In contrast, when the QM/MM partition is flipped such that the hydrogen-bond donor is assigned to the MM region represented by the MB-DFT PEF, the QM/MM electrostatic interactions become appreciably different from the pure MM electrostatic interactions provided by the MB-DFT PEF, which prevents

the PIPs of the MB-DFT PEFs from correctly recovering the *ab initio* DFT energy.[77] Furthermore, the higher-body $n$B terms ($n \geq 4$) of the MB-DFT PEFs are fully described by the polarization term of MB-pol, which implies that they are identical in all MB-DFT PEFs, regardless to which density functional a particular MB-DFT PEF is derived from.[53] As a consequence, if a given density functional displays sufficiently different polarization effects from those represented by MB-pol, the original MB-DFT PEFs[53] are not able to correctly reproduce higher-body $n$B interactions ($n \geq 4$) provided by that density functional. Finally, the 1B term of the original MB-DFT PEFs for water, which describes the intramolecular distortions in an isolated water molecule, is represented by the same Partridge-Schwenke PEF[90] used in MB-pol,[39–41] which was derived from high-level QM calculations and further refined to reproduce the rovibrational transitions of an isolated water molecule in the gas phase. This implies that the intramolecular distortion energies in the original MB-DFT PEFs do not properly reproduce the corresponding DFT 1B energies.[53]

In this study, we introduce an efficient theoretical/computational framework for the development of MB-QM PEFs that consistently reproduce each individual many-body energy contribution calculated *ab initio* using the corresponding QM model. While MB-QM PEFs can be developed for generic (small) molecular fluids at an arbitrary QM level of theory, the present study focuses on MB-QM PEFs for water derived from *ab initio* data calculated using various density functionals belonging to different rungs across the Jacob's ladder of DFT approximations as well as from second-order Møller-Plesset (MP2) perturbation theory. While these MB-QM PEFs are intended for use in future QM/MM simulations, this first study focuses on the theoretical details and assessment of the accuracy of the MB-QM PEFs through a systematic analysis of the energetics of small water clusters. By analyzing each many-body contribution to the interaction energies, we find that density functionals derived within the generalized gradient approximation (GGA) and hybrid density functionals suffer from relatively large density-driven errors. This leads to the over-delocalization of the electron density which, in turn, results in (unphysically) larger higher-body energies whose magnitude cannot correctly be captured by the purely classical, many-body polarization term adopted by the MB-DFT PEFs. We show that, within the theoretical frame-

work based on the many-body formalism that is introduced here, it is possible to develop density-corrected MB-DFT PEFs that remove, at least partially, these density-driven errors and provide a more accurate description of individual many-body contributions to the interactions between water molecules.

# 2 Theory

## 2.1 MB-QM

The MBE expresses the total energy, $E_{tot}$, of an $N$-body system as the sum of individual $n$-body energy contributions, $\varepsilon_{nB}$, where $n \leq N$,[91]

$$E_{tot}(r_1,..,r_N) = \sum_{i=1}^{N} \varepsilon_{1B}(r_i) + \sum_{i<j}^{N} \varepsilon_{2B}(r_i,r_j) + \sum_{i<j<k}^{N} \varepsilon_{3B}(r_i,r_j,r_k) + ... + \varepsilon_{NB}(r_1,..,r_N) \qquad (1)$$

Here, $\varepsilon_{1B}$ represents the energy of an isolated monomer, and the $n$-body energies are defined recursively as

$$\varepsilon_{nB} = \varepsilon_n(1,...,n) - \sum_{i=1}^{N} \varepsilon_{1B}(r_i) - \sum_{i<j}^{N} \varepsilon_{2B}(r_i,r_j) - \sum_{i<j<k<...<n-1}^{N} \varepsilon_{(n\text{-}1)B}(r_i,r_j,..r_{n-1}). \qquad (2)$$

It has been shown that the MBE converges quickly for systems with localized electron densities and large band gaps. For example, the sum of 2B and 3B energies contributes to $\sim$96-99% of the total interaction energy in water.[91–98] Exploiting the fast convergence of the MBE, several many-body PEFs for water[39–41,99,100] and other molecular fluids[51,52,101,102] have been derived from high-level QM data. In this study, we generalize the many-body formalism originally adopted by the MB-pol PEF[39–41] to the development of many-body PEFs for generic (small) molecules at an arbitrary QM level of theory. As examples, we introduce several MB-DFT PEFs for water derived from various density functionals belonging to different rungs across the Jacob's ladder of DFT approximations (MB-DFT) as well as from MP2 perturbation theory (MB-MP2).

Briefly, the MB-QM PEFs approximate eq 1 to the sum of explicit 1B, 2B, and 3B terms along with a polarization term implicitly representing classical many-body interactions:

$$E_{tot}(r_1,..,r_N) = \sum_{i=1}^{N} \varepsilon_{1B}(r_i) + \sum_{i>j}^{N} \varepsilon_{2B}(r_i,r_j) + \sum_{i>j>k}^{N} \varepsilon_{3B}(r_i,r_j,r_k) + E_{pol} \tag{3}$$

Each term of eq 3 is fitted to reproduce the corresponding *ab initio* data calculated at the same QM level of theory. Specifically, the 1B term is represented by a PIP[24] that accurately describes intramolecular distortions. The 2B term incorporates three different energy contributions:

$$\varepsilon_{2B} = E_{2B}^{sr} + E_{elec} + E_{disp} \tag{4}$$

where $E_{2B}^{sr}$ describes short-range 2B interactions and is represented by a $4^{th}$-degree PIP.[39,53] $E_{elec}$ represents permanent electrostatics between point charges that reproduce the *ab initio* dipole moment of an isolated molecule. Following ref 77, in the MB-QM PEFs for water introduced here, the point charges are kept fixed to the values that reproduce the dipole moment of a water molecule in its equilibrium geometry. However, the present MB-QM scheme is general and allows for using geometry-dependent charges if deemed necessary for achieving higher accuracy. The last term in eq 4, $E_{disp}$, describes the 2B dispersion energy that is expressed as

$$E_{disp} = -\sum_{i,j} f(\delta_{ij}) \frac{C_{6,ij}}{R_{ij}^6} \tag{5}$$

where $f(\delta_{ij})$ is the Tang-Toennies damping function,[103] $i$ and $j$ are atom indices on two separate water monomers, and $C_{6,ij}$ are the dispersion coefficients. The 3B term of the MB-QM PEFs describes short-range 3-body interactions and, as in the case of $\varepsilon_{2B}$, is represented by a $4^{th}$-degree PIP,

$$\varepsilon_{3B} = E_{3B}^{sr}. \tag{6}$$

## 2.2 Density-corrected DFT

The Kohn-Sham theory allows for minimizing the exchange-correlation (XC) functional given by[104]

$$F[n] = T_S[n] + E_H[n] + E_{XC}[n]. \tag{7}$$

where $n(r)$ is the ground-state density of the system, $T_S$ is the non-interacting kinetic energy, and $E_H$ and $E_{XC}$ define Hartree and exchange-correlation energies, respectively. The exact $F[n]$ gives rise to derivative discontinuities for total energies at integer values of the number of electrons ($N$),[105] a feature that is missing in GGA and hybrid GGA density functionals which, by construction, provide smooth functions of $N$. As a consequence, the derivatives of these density functionals are inaccurate,[105] which results in an incorrect Kohn-Sham potential and, in turn, poor orbital energies. As the potential generated by these density functionals show a constant shift with respect to the exact potential, these errors do not affect the density. Unsurprisingly, for this reason, most density functionals engender accurate electron densities, which suggests that DFT errors are primarily functional-driven rather than density-driven. In this regard, it should, however, be noted that density functionals constructed in unconstrained forms may give rise to unphysical electron densities despite providing a good representation of the energies.[106]

Probably the most significant contributor to density-driven errors is the self-interaction error (SIE). In wavefunction theory, Coulomb interactions are incorporated as pairwise two-electron potentials, where an electron does not interact with itself. However, in Kohn-Sham theory, since the energy is a functional of the one-electron density, it is impossible to remove the interaction of the electrons with themselves as each electron interacts with the entire density. As a consequence, most density functionals are unable to satisfy the following conditions for the one-electron system,[107]

$$T_s = \int \mathrm{d}^3 r \frac{|\nabla n|^2}{8n}, \quad E_X = -E_H, \quad E_C = 0. \tag{8}$$

where $E_X$ and $E_C$ are the exchange and correlation energies. The inability to satisfy eq 8 gives rise to the SIE. Although for many-electron systems the self-interaction energy can be formally

defined as the sum of the energies of each orbital interacting with itself,[108] correcting for the SIE has been found not to be straightforward.[109] For hybrid density functionals, mixing a fraction of Hartree-Fock exchange with the XC functional leads to incomplete cancellation of the SIE, thus giving rise to smaller density-driven errors compared to pure functionals.[110–112]

Besides the SIE, another contributor to density-driven errors is the delocalization error, which is a many-electron effect resulting in unphysical delocalization of the electrons and low energies for the delocalized electrons.[109,113–115]

One possible workaround to minimize DFT density-driven errors is the use of an accurate electron density. However, this effectively makes the minimization of density-driven errors impractical because finding a highly accurate electron density is more expensive than the DFT calculation itself. In this context, with the absence of derivative discontinuities in the energy provided by the GGA functionals, the use of the Hartree-Fock density, $n^{\mathrm{HF}}$, in the orbital-dependent functionals is free from density-driven errors[107] and makes the overall computational cost of the density-corrected energy, $E^{\mathrm{DC-DFT}}$, similar to that of the original DFT energy,

$$E^{\mathrm{DC-DFT}} \approx E^{\mathrm{HF}} + \left(\tilde{E}_{\mathrm{XC}}\left[n^{\mathrm{HF}}\right] - E_{\mathrm{X}}^{\mathrm{HF}}\right) \tag{9}$$

Although approximate, using the Hartree-Fock density, $n^{\mathrm{HF}}$, to calculate $E^{\mathrm{DC-DFT}}$ in a non-self-consistent way has been shown to minimize density-driven errors in DFT calculations of various properties (e.g., electron affinities,[116] noncovalent interactions,[117–120] spin gaps for coordination compounds[120]) with minimal additional computational cost. It should be noted that other approaches have also been proposed in the literature to correct the SIE in DFT calculations.[121,122]

# 3 Computational details

All MB-DFT PEFs were fitted to *ab initio* data calculated with GGA density functionals (BLYP[123,124] and PBE[125] belonging to rung 2 functionals), meta-GGA density functionals (B97M-rV[126] belonging to rung 3 functionals), hybrid density functionals (B3LYP,[124,127] PBE0,[128] and M06-2X,[129]

belonging to rung 4 functionals) and range-separated density functionals ($\omega$B97M-V[130] belonging to rung 4 functionals). The empirical D3 correction was added to all density functionals that do not account for dispersion energy.[131] Similarly, the MB-MP2 PEF was derived from the corresponding data calculated at the MP2 level of theory. Both DFT and MP2 calculations were carried out with the aug-cc-pVQZ basis set[132,133] using Q-Chem[134] and ORCA[135] quantum chemistry packages. The Euler-Maclaurin-Lebedev (99,590) grid[136,137] was used in all DFT calculations to minimize possible grid errors since it has been shown that the accuracy of more recent density functionals are particularly sensitive to the choice of the integration grid.[138]

The 1B, 2B, and 3B training sets used in the development of each MB-DFT and MB-MP2 PEF comprise 5000 monomers, 42508 dimers, and 12347 trimers, respectively. The 1B training set was generated using in-house software following the same procedure described in previous studies,[51,52] while we used the same 2B and 3B training sets used in the development of the MB-pol PEF.[39,40]

The dipole polarizabilities for the free O and H atoms were computed at the corresponding DFT and MP2 levels of theory. The effective polarizabilities used in the MB-DFT and MB-MP2 PEFs were calculated as:

$$\alpha_{eff} = \alpha_{free} \left( \frac{V_{eff}}{V_{free}} \right)^{4/3} \tag{10}$$

where $V_{eff}$ and $V_{free}$ are the effective and free volumes of the O and H atoms in $H_2O$ calculated via the XDM model[139–141] as implemented in Q-Chem.[142] The atomic charges were calculated using the CM5 method[143] and then distributed on the actual sites representing the MB-QM $H_2O$ molecule (i.e., the fictitious site along the bisector of the HOH angle and the two H atoms)[39] according to the procedure described in ref 144.

All the density-corrected DFT calculations were performed in two steps, where a fully self-consistent Hartree-Fock calculation was performed prior to a single non-self-consistent DFT calculation where the orbitals of the HF calculation were used to generate the DFT density.

In the analysis of the energetics of the water clusters presented in section 4, the binding energies are defined as

$$E_{bind} = E^{\text{cluster}} - nE_{opt}^{\text{H}_2\text{O}} \tag{11}$$

where $E^{\text{cluster}}$ is the total energy of $n$-molecule cluster and $E_{opt}^{H_2O}$ is the energy of an isolated water molecule in the optimized geometry. The corresponding interaction energies are defined as

$$E_{int} = E^{\text{cluster}} - \sum_i E_i^{H_2O} \tag{12}$$

where $E^{\text{cluster}}$ is the same total energy as in eq 11 and $E_i^{H_2O}$ is the energy of the $i$-th water molecule in the same geometry as in the cluster. All MP2 optimized geometries used in the analyses of the energetics of the water clusters are taken from ref 42.

# 4 Results and discussions

## 4.1 Many-body analysis of the MB-QM PEFs

In this section we analyze the ability of the MB-DFT and MB-MP2 PEFs to reproduce results obtained from *ab initio* calculations carried out at the corresponding QM level of theory. Correlation plots between the 2B and 3B energies calculated with the different QM methods and the corresponding values obtained with the MB-QM PEFs are shown in Figures 1 and 2, respectively. Analogous correlation plots for the 1B energies are included in the Supporting Information.

Independently of the QM level of theory, the MB-QM PEFs are able to quantitatively reproduce the 1B, 2B, and 3B reference energies, with RMSDs of $\sim$0.08, $\sim$0.12, and $\sim$0.03 kcal/mol respectively. In the case of the 2B and 3B energies, the high correlation between the QM data and the corresponding MB-QM values indicates that the 4th-degree PIPs used to supplement the representation of short-range 2B and 3B interactions in the MB-QM PEFs are sufficiently flexible to "capture" quantum-mechanical contributions (e.g., Pauli repulsion, charge transfer and penetration) which, arising from the overlap of the monomers' electron densities, cannot be quantitatively represented by classical expressions commonly used in conventional force fields.

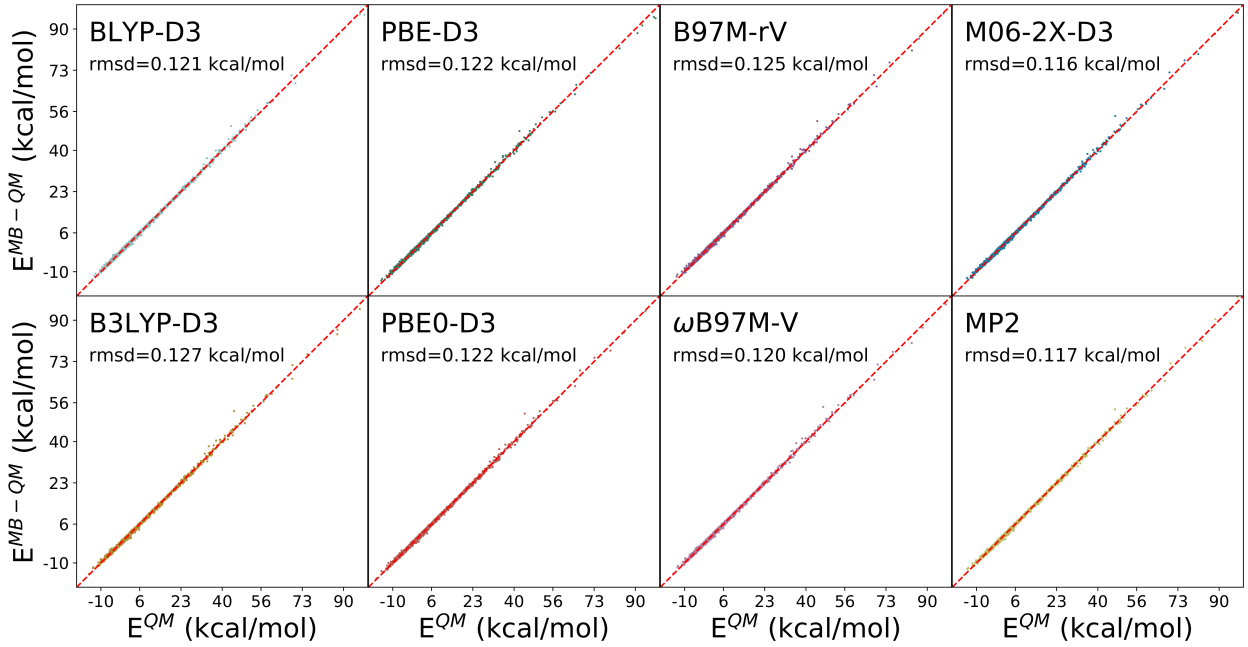


Figure 1: Correlation plots between the 2B QM reference energies and the values obtained with the corresponding MB-QM PEFs calculated for the 42508 configurations of the 2B training set.

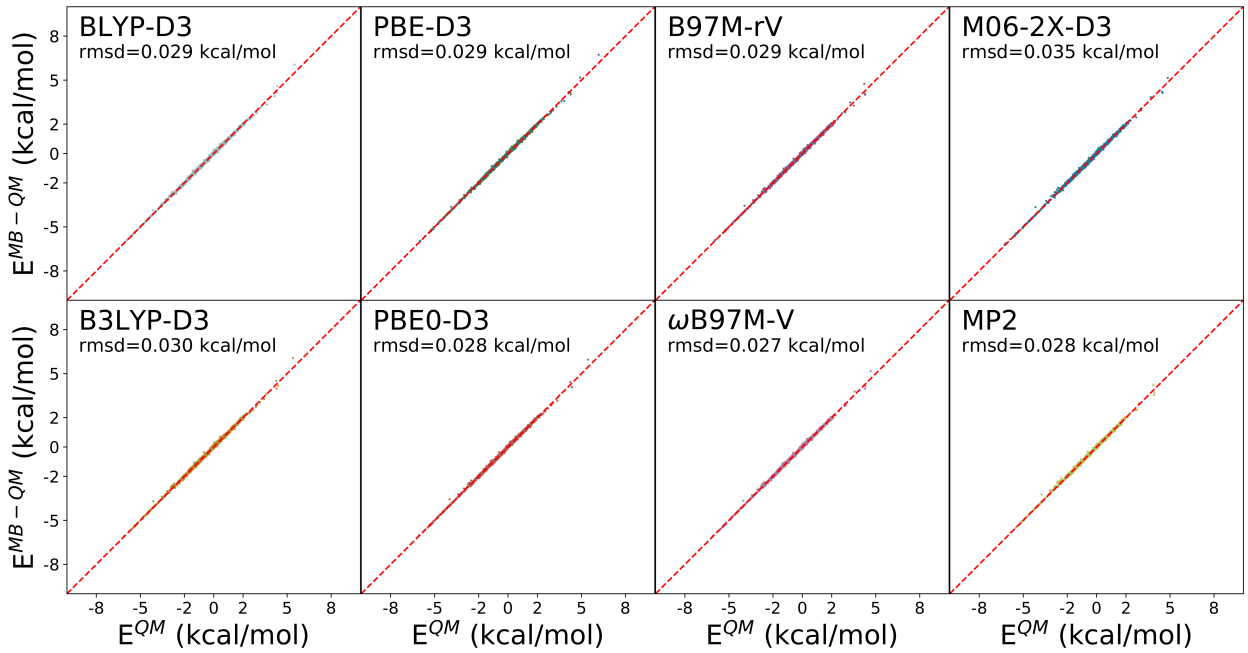


Figure 2: Correlation plots between the 3B QM reference energies and the values obtained with the corresponding MB-QM PEFs calculated for the 12347 configurations of the 3B training set.

As discussed in previous studies,[1,42] the analysis of individual many-body contributions to the interaction energies of small water clusters allows for an overall assessment of the accuracy of a given water model. In this context, it should be emphasized that, since the MB-QM PEFs are trained on data that contain information only up to 3B term of the MBE, they do not have any prior knowledge of systems containing more than three water molecules. Figure 3 displays the MB-QM errors relative to the corresponding QM values for each individual many-body contribution (i.e., from 2B to 6B energies) to the interaction energy of the prism isomer that corresponds to the minimum-energy isomer of the water hexamer.[145–148] While the corresponding analyses for the first eight low-energy hexamer isomers are reported in the Supporting Information, Table 1 lists the MB-DFT and MB-MP2 errors for each individual *n*B energy averaged over all eight isomers. As discussed in section 3, since the interaction energies, by definition, remove all contributions due to monomer distortions, the present many-body analysis allows for a direct and quantitative comparison of the MB-DFT and MB-MP2 PEFs in their ability to reproduce the corresponding QM *n*B contributions to the interaction energy.

Figure 3 shows that the absolute magnitude of the 2B and 3B errors associated with the MB-QM PEFs range between 0.1 and 0.3 kcal/mol, with the MB-DFT PEFs fitted to M06-2X-D3 and

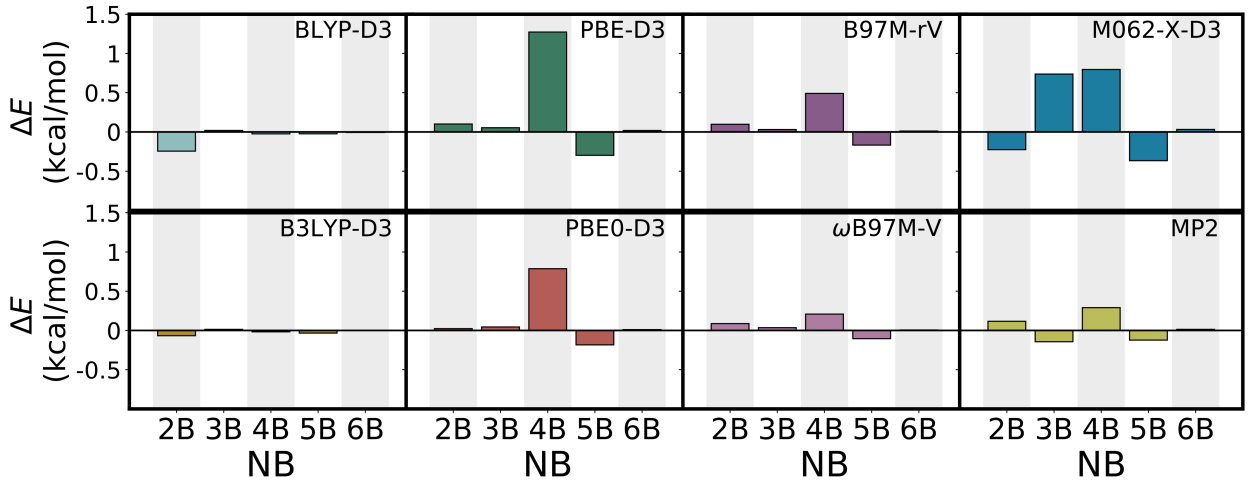


Figure 3: Errors in individual many-body energies, *n*B, associated with the MB-QM PEFs relative to the corresponding QM reference values for the prism isomer of the water hexamer. The MP2 optimized geometry of the prism isomer is from ref 42.

BLYP-D3 data displaying the largest 2B errors. The overall low errors for the 2B and 3B energies are consistent with the low RMSDs reported in Figures 1 and 2. In the case of M06-2X-D3, relatively larger errors are not only associated with the 3B term but are also found at the 4B and 5B levels, as discussed in more detail in ref 1. The large average 3B error (0.7 kcal/mol) associated with M06-2X-D3 can be explained by considering that the water hexamer contains 20 distinct trimers and M06-2X-D3 displays the largest 3B RMSD of 0.035 kcal/mol (Figure 2). It should be noted that, although similar trends are also found for the other isomers as shown in Figures S3 - S10 of the Supporting Information, the more planar hexamer isomers (i.e., the book, boat, and chair isomers) display significantly larger 2B and 3B errors compared to the three-dimensional isomers (i.e., the prism and cage isomers). This difference is primarily due to error accumulation in the highly symmetric planar isomers, where repeated dimer and trimer subunits contribute many-body energies with same signed errors, resulting in the total error to add up.

As shown in Table 1, the average 4B error is disproportionately large relative to the magnitude of the total error for all MB-QM PEFs, except those derived from B3LYP-D3 and MP2 data. The largest 4B error (1.07 kcal/mol) is associated with the MB-DFT PEF fitted to PBE-D3 data. Importantly, despite 4B effects generally contributing less than 5% to the total interaction energy in water clusters,[1,93,94,149–151] the error in the 4B term accounts for most of the total error associated with the MB-DFT PEFs. By construction (see section 2), 4B energies in

Table 1: Average errors associated with the QM methods considered in this study relative to the reference QM values calculated for the first low-energy isomers of the water hexamers. The MP2 optimized geometries of the water hexamer isomers are from ref 42.

| | $\Delta E_{avg}^{NB}$ (kcal/mol) | | | | | |
|---|---|---|---|---|---|---|
| Method | 2B | 3B | 4B | 5B | 6B | Total |
| BLYP-D3 | 0.19 | 0.08 | 0.45 | 0.06 | 0.00 | 0.79 |
| B3LYP-D3 | 0.36 | 0.08 | 0.24 | 0.03 | 0.01 | 0.72 |
| PBE-D3 | 0.46 | 0.07 | 1.07 | -0.06 | 0.02 | 1.57 |
| PBE0-D3 | 0.38 | 0.07 | 0.72 | -0.03 | 0.01 | 1.16 |
| M06-2X-D3 | -0.10 | 0.70 | 0.67 | -0.08 | 0.01 | 1.20 |
| B97M-rV | 0.27 | 0.07 | 0.56 | -0.03 | 0.00 | 0.89 |
| $\omega$B97M-V | 0.31 | 0.07 | 0.42 | -0.00 | 0.00 | 0.81 |
| MP2 | 0.35 | -0.06 | 0.33 | 0.01 | -0.01 | 0.62 |

the MB-QM PEFs are represented completely by a classical many-body polarization term that is expressed according to a modified Thole-type model.[152] Table I shows that MB-QM PEFs fitted to $\omega$B97M-V and MP2 data are associated with relatively small 4B errors, similar to those found for MB-pol relative to CCSD(T)/CBS,[42] while large 4B errors are associated with MB-DFT PEFs fitted to PBE-D3, PBE0-D3, and M06-2X-D3 data. Given the trend observed in the 4B errors associated with MB-DFT PEFs fitted to density functionals belonging to different rungs across the Jacob's ladder of DFT approximations, we posit that the magnitude of the 4B errors stems from density-driven errors (i.e., self-interaction and delocalization errors) which plague, to various extent, all density functionals. As discussed in section 4, the electron self-repulsion is explicitly non-local and should, in principle, be removed via the exchange-correlation energy. In practice, most exchange-correlation functionals contain substantial local components and are thus unable to correctly compensate for the interactions of the electrons with themselves. This results in (unphysical) over-delocalization of the electron density in order to minimize the electron self-repulsion. In turn, this over-delocalization of the electron density causes the higher-order terms of the MBE to be (artificially) more quantum-mechanical in nature and, consequently, not amenable to the purely classical representation based on many-body polarization which is adopted by the MB-QM PEFs.

## 4.2 Density-driven errors in the DFT description of molecular interactions in water

In this section we analyze density-driven errors in DFT models of water with a particular focus on the dependence of these erorrs on both the rung of the density functional and the cluster size.

### Interaction energies of water clusters

Our analysis indicates that common GGA functionals tend to overbind the water clusters, likely due to the presence of relatively large density-driven errors (Figure S11 and S12). To test this hypothesis, we use HF orbitals in non-self-consistent DFT calculations of the interaction and many-body energies of water clusters, which was shown to reduce the impact of density-driven errors.[153]

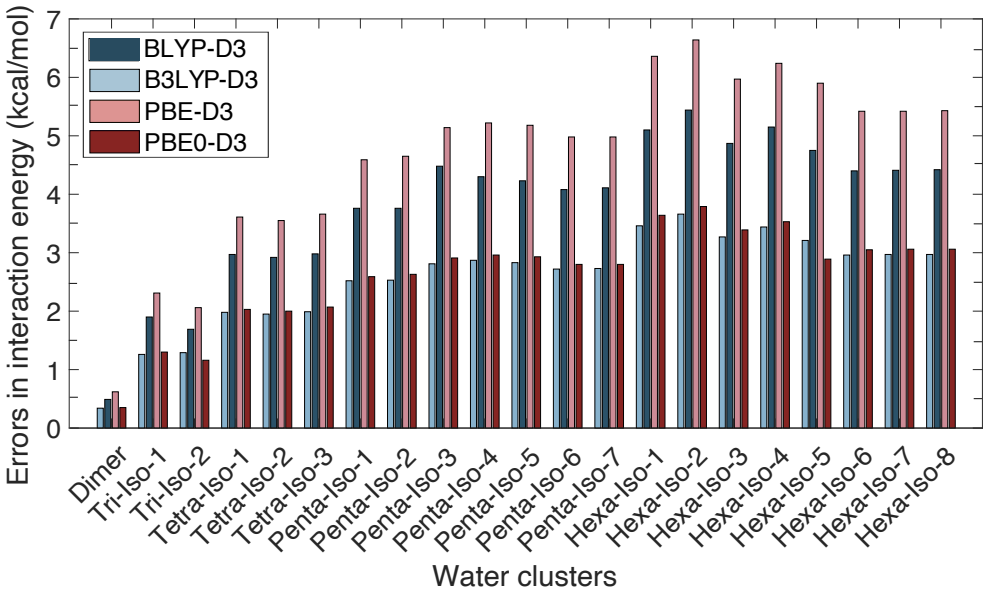


Figure 4: Density-driven errors in the interaction energies of $(H_2O)_n$ clusters, with $n = 2-6$, associated with BLYP-D3, B3LYP-D3, PBE-D3, PBE0-D3 relative to the corresponding density-corrected functionals. The errors are defined as $\Delta E_{int}^{DC} = E_{int}^{DC} - E_{int}^{SC}$, where $E_{int}^{DC}$ and $E_{int}^{SC}$ are the interaction energies calculated with the density-corrected and self-consistent functionals, respectively.

Figure 4 shows the errors in the interaction energies ($\Delta E_{int}^{DC}$) calculated for various water clusters, from the dimer to the hexamer, using both the self-consistent (SC) and density-corrected (DC) versions of the GGA BLYP-D3 and PBE-D3 functionals, and corresponding hybrid GGA B3LYP-D3 and PBE0-D3 functionals. For all clusters, PBE-D3 is found to be the most susceptible to the density-driven errors. Independently of the density functional, the density-driven error defined as $\Delta E_{int}^{DC} = E_{int}^{DC} - E_{int}^{SC}$ increases from the dimer to the hexamer. In particular, $\Delta E_{int}^{DC}$ per molecule lies within 0.73 and 0.82 kcal/mol for BLYP-D3 and decreases to 0.39-0.44 kcal/mol for B3LYP-D3,[124,127] which contains 20% Hartree-Fock exchange (Figure S13). A similar trend is found for the PBE-D3 and PBE0-D3 functionals, with the former displaying $\Delta E_{int}^{DC}$ per molecule between 0.93 and 1.06 kcal/mol, and the latter between 0.44 and 0.56 kcal/mol. It should be noted that PBE0-D3 contains 25% Hartree-Fock exchange.[128]

To assess the overall accuracy of the density-corrected functionals, Table 2 lists the errors associated with each density-corrected functional relative to the CCSD(T)/CBS reference values[42] for the interaction energies of the first eight low-energy isomers of the water hexamer. For this

Table 2: Errors (in kcal/mol) in interaction energies for the first eight low-energy isomers of the water hexamer associated with the density-corrected functionals PBE-D3(DC), PBE0-D3(DC), BLYP-D3(DC), B3LYP-D3(DC), M06-2X-D3(DC) and $\omega$B97M-V(DC) relative to the CCSD(T)/CBS reference values. The errors are defined as $\Delta E_{int}^{CCSD(T)} = E_{int}^{DC} - E_{int}^{CCSD(T)}$, where the reference $E_{int}^{CCSD(T)}$ values are taken from ref 42.

| Isomer | PBE-D3 (DC) | PBE0-D3 (DC) | BLYP-D3 (DC) | B3LYP-D3 (DC) | M06-2X-D3 (DC) | $\omega$B97M-V (DC) |
|---|---|---|---|---|---|---|
| Prism | 3.76 | 1.05 | 5.44 | 2.82 | -1.16 | 0.50 |
| Cage | 3.32 | 0.67 | 5.26 | 2.54 | -0.66 | 0.73 |
| Book1 | 2.55 | -0.01 | 4.92 | 2.06 | -0.21 | 0.65 |
| Book2 | 2.51 | -0.03 | 4.91 | 2.06 | -0.12 | 0.68 |
| Bag | 2.48 | -0.05 | 4.91 | 2.05 | -0.20 | 0.58 |
| Cyclic chair | 1.97 | -0.56 | 4.65 | 1.62 | -0.40 | 0.42 |
| Cyclic boat 1 | 1.92 | -0.54 | 4.57 | 1.61 | -0.32 | 0.45 |
| Cyclic boat 2 | 1.98 | -0.50 | 4.61 | 1.63 | -0.33 | 0.43 |
| MUE | 2.56 | 0.43 | 4.91 | 2.05 | 0.42 | 0.56 |

analysis, the errors are defined as $\Delta E_{int}^{CCSD(T)} = E_{int}^{DC} - E_{int}^{CCSD(T)}$, where $E_{int}^{DC}$ and $E_{int}^{CCSD(T)}$ are the interaction energies calculated with the density-corrected functionals and at CCSD(T)/CBS level of theory, respectively. A large mean unsigned error (MUEs) of 4.91 kcal/mol is found for BLYP-D3(DC), whereas PBE-D3(DC) and B3LYP-D3(DC) provide comparable MUEs of 2.56 kcal/mol and 2.05 kcal/mol, respectively. Despite being a hybrid functional as B3LYP-D3(DC), PBE0-D3(DC) provides a much smaller MUE of 0.43 kcal/mol, which is comparable to that found for M06-2X-D3(DC) (MUE = 0.42 kcal/mol) and $\omega$B97M-V(DC) (MUE = 0.56 kcal/mol).

It should be noted that the hexamer isomers can be broadly classified in two distinct groups, one containing three-dimensional structures, where each water molecule is involved in three hydrogen bonds (i.e., the prism and cage isomers), and one containing two-dimensional structures, where each water molecule is involved in two hydrogen bonds (i.e., the cyclic isomers), with the book-type isomers being in between and sharing features that are common to both groups. In this context, it is worth mentioning that, with the exception of BLYP-D3(DC) and $\omega$B97M-V(DC), all density-corrected functionals analyzed in this study exhibit significantly lower accuracy for the three-dimensional isomers.

### Many body decomposition analysis

It has been shown that 2B and 3B effects contribute to 96-99% of the total interaction energy in water,[1,93,154] which implies that the ability of a given water model to correctly reproduce 2B and 3B energies mainly determines the accuracy with which the model is able to describe the total interaction energy and relative stability of different water clusters.

Figure 5 shows the density-driven errors, $\Delta E_{nB}^{DC} = E_{nB}^{DC} - E_{nB}^{SC}$ $(n = 2-6)$, calculated for each individual $n$B contribution to the interaction energies of the first eight low-energy isomers of the water hexamer. Here, $E_{int}^{DC}$ and $E_{int}^{SC}$ are the interaction energies calculated with the density-corrected and self-consistent functionals, respectively. Due to relatively large density-driven errors, the MUE for $\Delta E_{2B}^{DC}$ is 6.22 kcal/mol for PBE-D3 and 4.69 kcal/mol for BLYP-D3. Adding a fraction of Hartree-Fock exchange systematically reduces these errors, with the corresponding MUEs being 3.45 kcal/mol and 3.25 kcal/mol for PBE0-D3 and B3LYP-D3, respectively. More recent functionals show significantly smaller errors, resulting in MUEs of 1.08 kcal/mol and 0.93 kcal/mol for M06-2X-D3 and $\omega$B97M-V, respectively, which confirms that density-driven errors are less severe

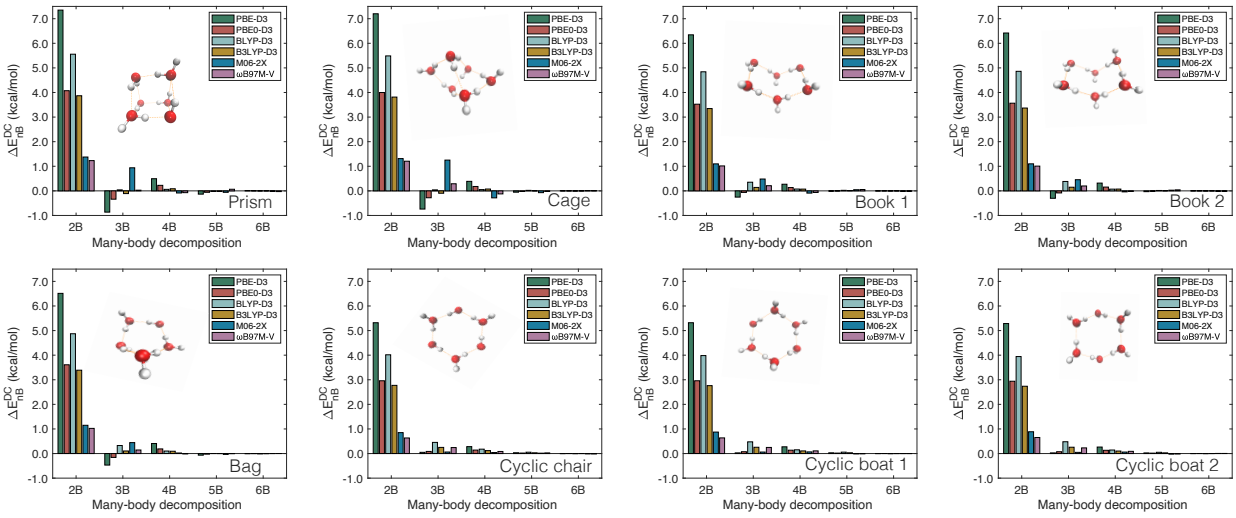


Figure 5: Errors in $n$B energies $(n = 2-6)$ associated with PBE-D3, PBE0-D3, BLYP-D3, B3LYP-D3, M06-2X-D3 and $\omega$B97M-V relative to the corresponding density-corrected values calculated for the first eight low-energy isomers of the water hexamer. Errors are defined as $\Delta E_{nB}^{DC} = E_{nB}^{DC} - E_{nB}^{SC}$, where $E_{int}^{DC}$ and $E_{int}^{SC}$ are the interaction energies calculated with the density-corrected and self-consistent functionals, respectively.

for hybrid, meta-GGA and range-separated functionals.

The analysis of $\Delta E_{2B}^{DC}$ shown in Figure 5 also indicates that, at the 2B level, density-driven errors tend to overbind the water clusters, as the 2B errors are positive for all density functionals examined in this study. This is in agreement with similar results recently obtained with the SCAN functional.[155] The trend in 3B density-driven errors is not as consistent as for the 2B errors. The largest MUE (0.47 kcal/mol) is found for M06-2X-D3, with the MUE for all other density functionals ranging from 0.14 kcal/mol to 0.34 kcal/mol. Interestingly, with the exception of the cyclic isomers, PBE-D3 and PBE0-D3 display negative $\Delta E_{3B}^{DC}$ for all hexamer isomers, which implies that 3B density-driven errors tend to underbind these clusters. Independently of the cluster structure, all other density functionals display positive 3B density-driven errors, except B3LYP-D3 that provide negative $\Delta E_{3B}^{DC}$ for the prism and cage isomers.

The comparisons with the reference CCSD(T)/CBS $n$-body energies shown in Figure 6 indicate that the PBE0-D3(DC) and M06-2X-D3(DC) functionals provide the smallest MUE at both 2B (0.27 kcal/mol and 0.65 kcal/mol, respectively) and 3B (0.27 kcal/mol and 0.53 kcal/mol, re-

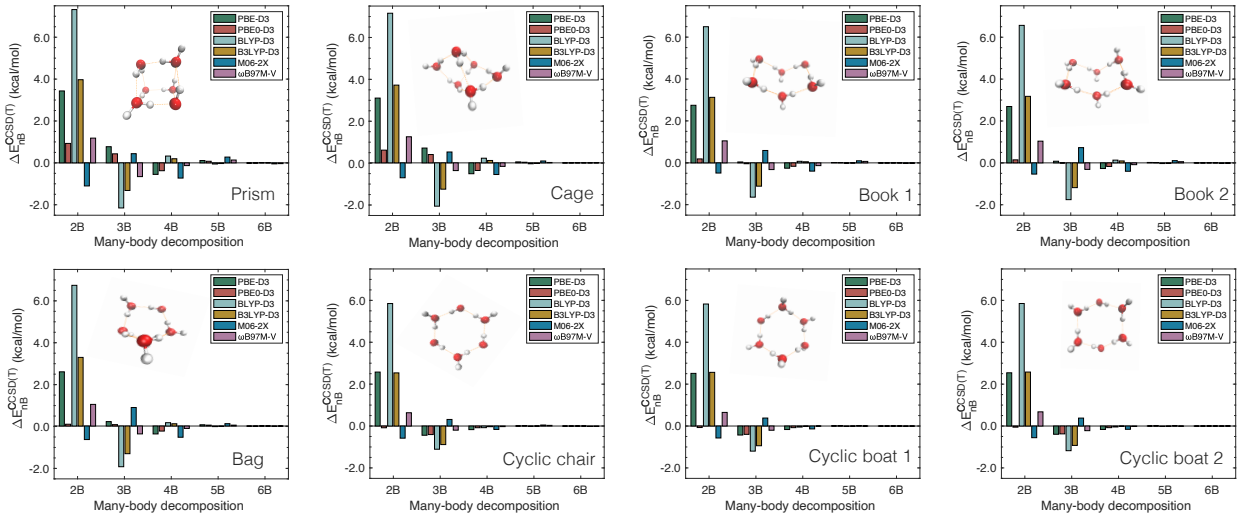


Figure 6: Errors in $n$B energies ($n = 2-6$) associated with the density corrected PBE-D3(DC), PBE0-D3(DC), BLYP-D3(DC), B3LYP-D3(DC), M06-2X-D3(DC) and $\omega$B97M-V(DC) relative to the corresponding CCSD(T)/CBS reference values[42] calculated for the first eight low-energy isomers of the water hexamer. Errors are defined as $\Delta E_{nB}^{CCSD(T)} = E_{nB}^{DC} - E_{nB}^{CCSD(T)}$, where $E_{nB}^{DC}$ and $E_{nB}^{CCSD(T)}$ are the $n$B energies calculated with the density-corrected functionals and at the CCSD(T)/CBS level of theory,[42] respectively.

spectively) levels. Relatively large 2B errors are still found for the GGA functionals (although significantly smaller than those associated with the corresponding self-consistent density functionals), with MUEs of 2.77 kcal/mol and 6.47 kcal/mol for the PBE-D3(DC) and BLYP-D3(DC) functionals, respectively. The 2B MUE for the $\omega$B97M-V(DC) functional is 0.94 kcal/mol for the 2B. Interestingly, independently of the isomer structure, all density-corrected functionals are associated with 2B and 3B errors of opposite signs relative to the CCSD(T)/CBS reference values. This results in fortuitous error compensation between 2B and 3B energies which, in turn, leads to apparently better agreement with the CCSD(T)/CBS total interactions energies for all hexamer isomers. Considering that the density-corrected functionals provide an approximate, yet reliable, representation of the interactions which is, at least, partially "free" of density-driven errors, the remaining deviations from the CCSD(T)/CBS values are likely due to inaccuracies in the density functionals, along with the localization errors associated with using Hartree-Fock densities in DFT calculations.[156]

**Binding energies of the water hexamer isomers**

Since the binding energies of all hexamer isomers lie within $\sim$1 kcal/mol, following ref 155, we focus our attention to the prism, cage, book-2, and cyclic chair-2 isomers. Based on the CCSD(T)/CBS reference values, the binding energies for these four hexamer isomers follow this order: prism $<$ cage $<$ book-2 $<$ cyclic chair-2, with the prism isomer being the most strongly bound isomer. The energy difference between the prism and cyclic chair-2 isomers is 0.88 kcal/mol at the CCSD(T)/CBS level. As shown in Figure 7a, the GGA functionals largely overbind all hexamer isomers, with MUE from the reference CCSD(T)/CBS data being 5.19 kcal/mol and 2.15 kcal/mol for PBE-D3 and BLYP-D3, respectively. The hybrid GGA functionals provides smaller MUE of 3.24 kcal/mol and 1.69 kcal/mol for PBE0-D3 and B3LYP-D3, respectively. Importantly, all these functionals provide the incorrect energy ordering of the isomers, predicting the cage isomer to be the most stable isomer. On the other hand, the more recent functionals (i.e., M06-2X and $\omega$B97M-V) predict the correct energy ordering, with the $\omega$B97M-V binding energies being

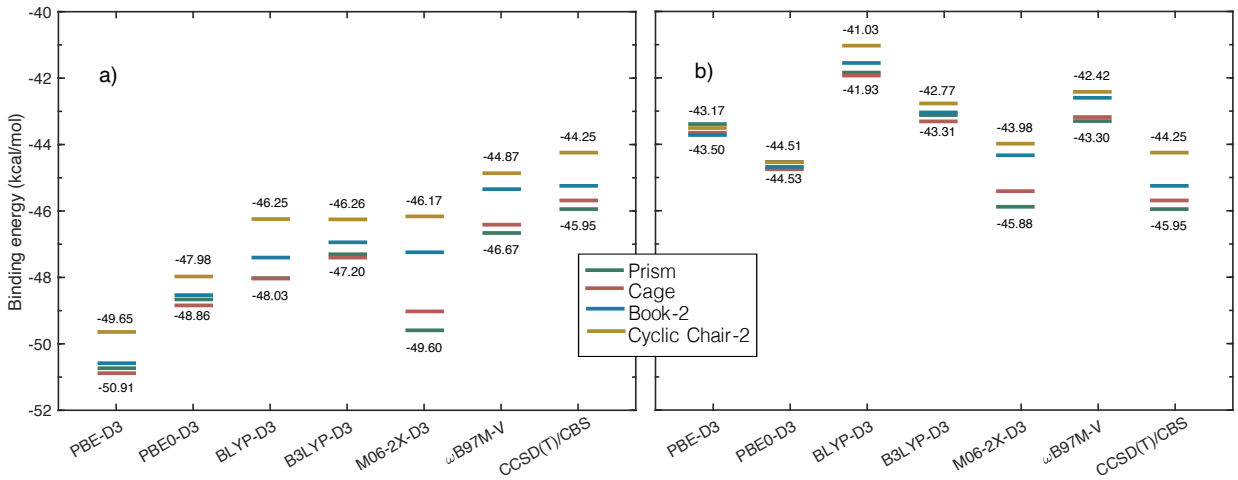


Figure 7: Binding energies of the prism, cage, book-2 and cyclic chair-2 isomers of the water hexamer calculated with PBE-D3, PBE0-D3, BLYP-D3, B3LYP-D3, M06-2X-D3 and $\omega$B97M-V, along with the corresponding CCSD(T)/CBS reference values. Panels a) and b) report binding energies calculated with the self-consistent and density-corrected functionals, respectively.

very close to the reference CCSD(T)/CBS values, which results in a MUE of 0.58 kcal/mol. The relatively large MUE of 2.71 kcal/mol associated with the M06-2X-D3 functional is mainly due to this functional largely overbinding the prism and cage isomers.

Figure 7b shows that the overbinding tendency associated with all self-consistent density functionals is removed upon applying the density correction. Overall, the density-corrected functionals provide binding energies that are in significantly closer agreement with the reference CCSD(T)/CBS values, although they tend to slightly underbind all hexamer isomers, which may be related to localization errors associated with using Hartree-Fock densities.[156] Specifically, the density correction decreases the MUE associated with the PBE-D3 and PBE0-D3 functionals to 1.72 kcal/mol and 0.81 kcal/mol, respectively. Relatively large MUE are still obtained with the density-corrected BLYP-D3 (3.70 kcal/mol) and B3LYP-D3 (2.23 kcal/mol) functionals. As for the corresponding self-consistent density functionals, none of the density-corrected PBE-D3, PBE0-D3, BLYP-D3, and B3LYP-D3 functionals is able to reproduce the correct energy ordering of the hexamer isomers, with all four density functionals predicting the cage isomer to be the lowest-energy isomer.

Since hybrid, meta-GGA and range-separated functionals are less prone to density-driven errors, the differences in binding energies calculated with the self-consistent and density-corrected

versions of M06-2X-D3 and $\omega$B97M-V are significantly smaller than those obtained with the GGA and hybrid GGA functionals. Specifically, the MUE associated with the density corrected M06-2X-D3 and $\omega$B97M-V functionals are 0.39 kcal/mol and 2.41 kcal/mol, respectively. Interestingly, this analysis indicates that applying the density correction somewhat deteriorates the ability of $\omega$B97M-V to reproduce the CCSD(T)/CBS binding energies of the water hexamer isomers.

## Density-corrected MB-DFT PEFs

To further investigate the effects of density-driven errors on the ability of density functionals to reproduce many-body interactions in water, we developed a density-corrected MB-DFT PEF, dubbed MB-PBE(DC), which was trained on 1B, 2B, and 3B energies calculated with the corresponding density-corrected PBE-D3(DC) functional. Due to the lack of a suitable machinery for calculating atomic charges, polarizabilities and $C_6$ coefficients using density-corrected functionals, the corresponding PBE0-D3 values were used for these quantities, which seems to be a good compromise as the presence of a fraction of Hartree-Fock exchange partially reduces density-driven errors.

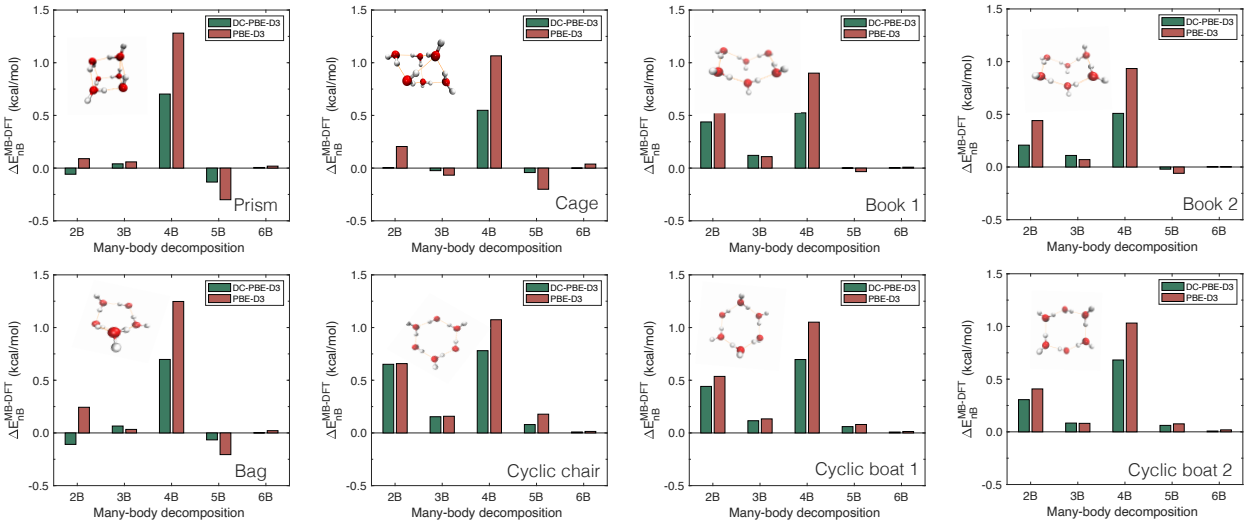


Figure 8: Errors in $n$B energies ($n = 2-6$) associated with MB-PBE and MB-PBE(DC) relative to the corresponding PBE-D3 and PBE-D3(DC) values calculated for the first eight low-energy isomers of the water hexamer. Errors are defined as $\Delta E_{nB}^{MB-DFT} = E_{nB}^{MB-DFT} - E_{nB}^{DFT}$, where $E_{nB}^{MB-DFT}$ and $E_{nB}^{DFT}$ are the $n$B energies calculated with the MB-DFT PEFs and corresponding (self-consistent and density-corrected) functionals, respectively.

Figure 8 shows the errors associated with the MB-PBE and MB-PBE(DC) PEFs for each individual many-body contribution to the interaction energies of the first eight low-energy hexamer isomers relative to the corresponding values calculated with the self-consistent PBE-D3 and density-corrected PBE-D3(DC) functionals, respectively. As the density-driven errors are minimized in the MB-PBE(DC) PEF, the corresponding 4B energies, which, as discussed in section 2, are entirely represented by classical polarization in the MB-DFT PEFs, are significantly closer to the reference CCSD(T)/CBS values. Specifically, the 4B MUE associated with the MB-PBE(DC) PEF is 0.64 kcal/mol which must be compared to a value of 1.07 kcal/mol reported in Table 1 for the analogous MB-PBE PEF trained on PBE-D3 data. Importantly, Figure 8 shows that not only the 4B error but also the errors for all other *n*B terms are significantly smaller when the density correction is taken into account as demonstrated by the higher accuracy provided by the MB-PBE(DC) PEF compared to the analogous MB-PBE PEF.

Figure 9 shows the interaction energies calculated with PBE-D3, PBE-D3(DC) and their cor-

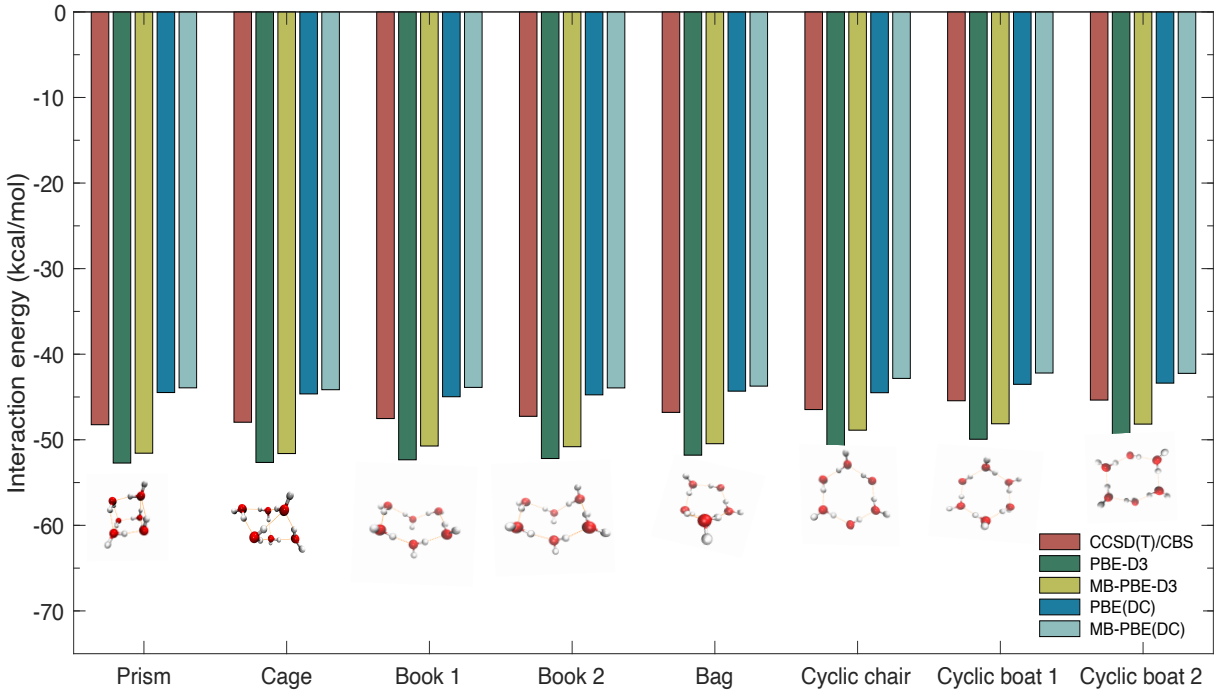


Figure 9: Interaction energies of the first eight low-energy isomers of the water hexamer calculated with the PBE-D3 and PBE-D3(DC) functionals, and the corresponding MB-PBE and MB-PBE(DC) PEFs along with the reference CCSD(T)/CBS values.

responding MB-PBE and MB-PBE(DC) PEFs, respectively, along with the CCSD(T)/CBS reference values. As density-driven errors tend to overbind the isomers of the water hexamer, PBE-D3 predicts lower interaction energies than CCSD(T)/CBS. On the other hand, PBE-D3(DC) systematically underbinds all isomers. The density correction significantly improves the description of the interaction energies of the MB-PBE(DC) PEF, by effectively reducing the errors associated with all $n$B contributions to the interaction energy (Figure 8). The MB-PBE PEF displays a MUE of -1.51 kcal/mol relative to PBE-D3 whereas the density-corrected MB-PBE(DC) PEF displays a MUE of -0.96 kcal/mol relative to PBE-D3(DC). A direct comparison of the errors are shown in Figure S14.

It should be noted that, while PBE-D3(DC) displays a significantly lower MUE (2.56 kcal/mol) relative to CCSD(T)/CBS than PBE-D3 (4.67 kcal/mol), the MB-PBE(DC) PEF is associated with a slightly higher MUE (3.52 kcal/mol) relative to CCSD(T)/CBS than the MB-PBE PEF (3.16 kcal/mol). This is due to the signs of $\Delta E^{CCSD(T)}$ (Figure 6) and $\Delta E^{MB-DFT}$ (Figure 8), which make these errors add up for the MB-PBE(DC) PEF while partially cancelling out for the MB-PBE PEF.

# 5 Conclusions

In this study, we have presented a general theoretical framework for the development of data-driven many-body PEFs (MB-QM) in which the individual terms of the many-body expansion of the energy are rigorously derived from electronic structure data calculated at an arbitrary QM level of theory. As a demonstration, we have introduced a family of MB-QM PEFs for water which are rigorously derived from density functionals belonging to different rungs across Jacob's ladder of DFT approximations (MB-DFT) as well as from Møller-Plesset perturbation theory (MB-MP2).

All MB-QM PEFs, except those derived from the GGA functionals, are shown to retain the same accuracy of the corresponding QM methods for the energetics of small water clusters and associated many-body energy contributions. Due to the presence of relatively large density-driven

errors, *ab initio* DFT calculations carried out with the GGA functionals analyzed in this study (PBE-D3 and BLYP-D3) yield appreciable smaller 4B energies compared to analogous calculations carried out with the corresponding MB-DFT PEFs. Since in the MB-DFT PEFs all higher-than-3B contributions are described by classical many-body polarization, the difference between the MB-DFT and DFT results can be traced back to density-driven errors that prevent density functionals, in particular those derived within the GGA, from accurately representing individual many-body contributions. In this regard, the analysis of the energetics of the isomers of the water hexamer shows that density-driven errors in the DFT representations of the interactions in water primarily affect 2B and 3B energies.

To further investigate the effects of density-driven errors in the DFT descriptions of the interactions in water, for each density functional considered in this study, we have analyzed a corresponding density-corrected functional obtained by using Hartree-Fock densities in non-self-consistent DFT calculations. Except for $\omega$B97M-V, all density-corrected functionals provide better agreement with the reference CCSD(T)/CBS data. Motivated by the improved performance of the density-corrected functionals, we have developed a density-corrected MB-DFT PEF, MB-PBE(DC) from density-corrected PBE-D3 data, and shown that it more closely reproduces the corresponding *ab initio* many-body energies than the analogous MB-DFT derived from self-consistent PBE-D3 data. The different performance of self-consistent and density-corrected functionals in reproducing the energetics of water clusters indicates that, while not essential for hybrid and range-separated functionals, accounting for density-driven errors is a necessary requirement in the development of MB-DFT PEFs derived from GGA functionals. These findings suggest that density-driven errors affect the convergence of the many-body expansion in water and indicate that the accuracy of a given density functional may vary significantly depending on both the size and hydrogen-bonding arrangements of the system under examination.

## Acknowledgements

This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, through grant no. DE-SC0019490. This research used 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, the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1548562, and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC) Supercomputer Center (SDSC).

## References

(1) 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*, 7501–7528.

(2) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48*, 371–394.

(3) Inakollu, V. S.; Geerke, D. P.; Rowley, C. N.; Yu, H. Polarisable Force Fields: What Do They Add in Biomolecular Simulations? *Curr. Opin. Struct. Biol.* **2020**, *61*, 182–190.

(4) Behler, J. Neural Network Potential-Energy Surfaces in Chemistry: A Tool for Large-Scale Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

(5) Behler, J. Representing Potential Energy Surfaces by High-Dimensional Neural Network Potentials. *J. Phys. Condens. Matter* **2014**, *26*, 183001.

(6) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-

Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.

(7) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840.

(8) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, e1603015.

(9) 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*, 3192–3203.

(10) 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*, 1–8.

(11) Haghighatlari, M.; Hachmann, J. Advances of Machine Learning in Molecular Modeling and Simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51–57.

(12) Manzhos, S.; Carrington Jr, T. Neural Network Potential Energy Surfaces for Small Molecules and Reactions. *Chem. Rev.* **2020**, https://doi.org/10.1021/acs.chemrev.0c00665.

(13) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390.

(14) Gkeka, P.; Stoltz, G.; Barati Farimani, A.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H.; Christine, P.; Pietrucci, F.; Silveira, A.; Tkatchenko, A.; Trstanova, Z.; Wiewiora, R.; Lelièvre, T. Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757–4775.

(15) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet

– A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(16) Unke, O. T.; Meuwly, M. A Reactive, Scalable, and Transferable Model for Molecular Energies from a Neural Network Approach Based on Local Information. *J. Chem. Phys.* **2018**, *148*, 241708.

(17) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Phys.* **2019**, *15*, 3678–3693.

(18) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 1–8.

(19) Jiang, B.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. *J. Chem. Phys.* **2013**, *139*, 054112.

(20) Li, J.; Jiang, B.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. II. Four-Atom Systems. *J. Chem. Phys.* **2013**, *139*, 204103.

(21) Jiang, B.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. III. Molecule-Surface Interactions. *J. Chem. Phys.* **2014**, *141*, 034109.

(22) Xie, C.; Zhu, X.; Yarkony, D. R.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. IV. Coupled Diabatic Potential Energy Matrices. *J. Chem. Phys.* **2018**, *149*, 144107.

(23) Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting Potential Energy Surfaces with Fundamental Invariant Neural Network. *J. Chem. Phys.* **2016**, *145*, 071101.

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

(25) Rosenberger, D.; Smith, J. S.; Garcia, A. E. Modeling of Peptides with Classical and Novel Machine Learning Force Fields: A Comparison. *J. Phys. Chem. B* **2021**,

(26) Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. Breaking the Coupled Cluster Barrier for Machine-Learned Potentials of Large Molecules: The Case of 15-Atom Acetylacetone. *J. Phys. Chem. Lett.* **2021**, *12*, 4902–4909.

(27) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ-Machine Learning for Potential Energy Surfaces: A PIP Approach to Bring a DFT-Based PES to CCSD(T) Level of Theory. *J. Chem. Phys.* **2021**, *154*, 051102.

(28) Li, J.; Varga, Z.; Truhlar, D. G.; Guo, H. Many-Body Permutationally Invariant Polynomial Neural Network Potential Energy Surface for $N_4$. *J. Chem. Theory Comput.* **2020**, *16*, 4822–4832.

(29) Guan, Y.; Xie, C.; Guo, H.; Yarkony, D. R. Neural Network Based Quasi-diabatic Representation for S0 and S1 States of Formaldehyde. *J. Phys. Chem. A* **2020**, *124*, 10132–10142.

(30) Jiang, B.; Li, J.; Guo, H. High-Fidelity Potential Energy Surfaces for Gas-Phase and Gas–Surface Scattering Processes from Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 5120–5131.

(31) del Cueto, M.; Zhou, X.; Zhou, L.; Zhang, Y.; Jiang, B.; Guo, H. New Perspectives on $CO_2$–Pt (111) Interaction with a High-Dimensional Neural Network Potential Energy Surface. *J. Phys. Chem. C* **2020**, *124*, 5174–5181.

(32) Wang, Y.; Guan, Y.; Guo, H.; Yarkony, D. R. Enabling Complete Multichannel Nonadiabatic Dynamics: A Global Representation of the Two-Channel Coupled, 1, 2 $^1$A and $1^3$A States of $NH_3$ Using Neural Networks. *J. Chem. Phys.* **2021**, *154*, 094121.

(33) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How Van der Waals Interactions Determine the Unique Properties of Water. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8368–8373.

(34) Yao, Y.; Kanai, Y. Temperature Dependence of Nuclear Quantum Effects on Liquid Water Via Artificial Neural Network Model Based on SCAN Meta-GGA Functional. *J. Chem. Phys.* **2020**, *153*, 044114.

(35) Gartner, T. E.; Zhang, L.; Piaggi, P. M.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Signatures of a Liquid–Liquid Transition in an Ab Initio Deep Neural Network Model for Water. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 26040–26046.

(36) 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*, 241725.

(37) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261–2269.

(38) Shi, W.; Jia, T.; Li, A. Quasi-Classical Trajectory Analysis with Isometric Feature Mapping and Locally Linear Embedding: Deep Insights Into the Multichannel Reaction on an $NH_3^+(^4A)$ Potential Energy Surface. *Phys. Chem. Chem. Phys.* **2020**, *22*, 17460–17471.

(39) 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*, 5395–5403.

(40) 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*, 1599–1607.

(41) 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*, 2906–2910.

(42) 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*, 194504.

(43) 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*, 2698–2705.

(44) 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*, 161715.

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

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

(47) 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. Lett.* **2019**, *123*, 2843–2852.

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

(49) Bajaj, P.; Wang, X.-G.; CarringtonJr, T.; Paesani, F. Vibrational Spectra of Halide–Water Dimers: Insights on Ion Hydration from Full-Dimensional Quantum Calculations on Many-Body Potential Energy Surfaces. *J. Chem. Phys.* **2017**, *148*, 102321.

(50) 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*, 406–412.

(51) 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**, *16*, 2246–2257.

(52) Riera, M.; Hirales, A.; Ghosh, R.; Paesani, F. Data-Driven Many-Body Models with Chemical Accuracy for $CH_4$/$H_2O$ Mixtures. *J. Phys. Chem. B* **2020**, *124*, 11207–11221.

(53) 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*, 8211–8218.

(54) Warshel, A.; Karplus, M. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.

(55) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(56) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM Simulation Method Applied To the Solution of $Li^+$ in Liquid Ammonia. *Chem. Phys.* **1996**, *211*, 313–323.

(57) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: An Extension of the ONIOM Method for Molecular Simulation in Condensed Phase. *Chem. Phys. Lett.* **2002**, *355*, 257–262.

(58) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.

(59) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.;

Csányi, G. QM/MM Simulation of Liquid Water With an Adaptive Quantum Region. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646–656.

(60) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive Partitioning in Combined Quantum Mechanical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale Simulations. *J. Phys. Chem. B.* **2007**, *111*, 2231–2241.

(61) Park, K.; Götz, A. W.; Walker, R. C.; Paesani, F. Application of Adaptive QM/MM Methods to Molecular Dynamics Simulations of Aqueous Systems. *J. Chem. Theory Comput.* **2012**, *8*, 2868–2877.

(62) Pezeshki, S.; Lin, H. Adaptive-Partitioning Redistributed Charge and Dipole Schemes for QM/MM Dynamics Simulations: On-the-Fly Relocation of Boundaries That Pass Through Covalent Bonds. *J. Chem. Theory Comput.* **2011**, *7*, 3625–3634.

(63) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. The Number-Adaptive Multiscale QM/MM Molecular Dynamics Simulation: Application To Liquid Water. *Chem. Phys. Lett.* **2012**, *524*, 56–61.

(64) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. Multiscale Modeling of Chemistry in Water: Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 5567–5577.

(65) Waller, M. P.; Kumbhar, S.; Yang, J. A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014**, *15*, 3218–3225.

(66) Watanabe, H. C.; Kubař, T.; Elstner, M. Size-Consistent Multipartitioning QM/MM: A Stable and Efficient Adaptive QM/MM Method. *J. Chem. Theory Comput.* **2014**, *10*, 4242–4252.

(67) Boereboom, J. M.; Potestio, R.; Donadio, D.; Bulo, R. E. Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures Using Many-Body Potentials. *J. Chem. Theory Comput.* **2016**, *12*, 3441–3448.

(68) Zheng, M.; Waller, M. P. Adaptive Quantum Mechanics/Molecular Mechanics Methods. *WIREs Comput. Mol. Sci.* **2016**, *6*, 369–385.

(69) Duster, A. W.; Wang, C.-H.; Garza, C. M.; Miller, D. E.; Lin, H. Adaptive Quantum/Molecular Mechanics: What Have We Learned, Where Are We, and Where Do We Go From Here? *WIREs Comput. Mol. Sci.* **2017**, *7*, e1310.

(70) Dohm, S.; Spohr, E.; Korth, M. Developing Adaptive QM/MM Computer Simulations for Electrochemistry. *J. Comput. Chem.* **2017**, *38*, 51–58.

(71) Field, M. J. An Algorithm for Adaptive QC/MM Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 2342–2351.

(72) Boereboom, J. M.; Fleurat-Lessard, P.; Bulo, R. E. Explicit Solvation Matters: Performance of QM/MM Solvation Models in Nucleophilic Addition. *J. Chem. Theory Comput.* **2018**, *14*, 1841–1852.

(73) Duster, A. W.; Wang, C.-H.; Lin, H. Adaptive QM/MM for Molecular Dynamics Simulations: 5. On the Energy-Conserved Permuted Adaptive-Partitioning Schemes. *Molecules* **2018**, *23*, 2170.

(74) Watanabe, H. C.; Cui, Q. Quantitative Analysis of QM/MM Boundary Artifacts and Correction in Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 3917–3928.

(75) Zheng, M.; Waller, M. P. Yoink: An Interaction-Based Partitioning API. *J. Comput. Chem.* **2018**, *39*, 799–806.

(76) Duster, A. W.; Garza, C. M.; Aydintug, B. O.; Negussie, M. B.; Lin, H. Adaptive Partitioning QM/MM for Molecular Dynamics Simulations: 6. Proton Transport Through a Biological Channel. *J. Chem. Theory Comput.* **2019**, *15*, 892–905.

(77) Lambros, E.; Lipparini, F.; Cisneros, G. A.; Paesani, F. A Many-Body, Fully Polarizable Approach to QM/MM Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7462–7472.

(78) Shen, L.; Wu, J.; Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 4934–4946.

(79) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(80) 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*, 9423–9437.

(81) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(82) Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.

(83) Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. TINKTEP: A Fully Self-Consistent, Mutually Polarizable QM/MM Approach Based on the AMOEBA Force Field. *J. Chem. Phys.* **2016**, *145*, 124106.

(84) Dziedzic, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. Mutually Polarizable QM/MM Model With in Situ Optimized Localized Basis Functions. *J. Chem. Phys.* **2019**, *150*, 074103.

(85) Loco, D.; Polack, E.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

(86) Olsen, J. M. H.; Kongsted, J. In *Advances in Quantum Chemistry*; Sabin, J. R., Brändas, E., Eds.; Elsevier, 2011; pp 107–143.

(87) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(88) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM Molecular Dynamics With AMOEBA Polarizable Embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.

(89) Loco, D.; Jurinovich, S.; Cupellini, L.; Menger, M. F.; Mennucci, B. The Modeling of the Absorption Lineshape for Embedded Molecules Through a Polarizable QM/MM Approach. *Photochem. Photobiol. Sci.* **2018**, *17*, 552–560.

(90) 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*, 4618–4639.

(91) Hankins, D.; Moskowitz, J.; Stillinger, F. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.

(92) Del Bene, J.; Pople, J. Intermolecular Energies of Small Water Polymers. *Chem. Phys. Lett.* **1969**, *4*, 426–428.

(93) Clementi, E.; Kołos, W.; Lie, G.; Ranghino, G. Nonadditivity of Interaction in Water Trimers. *Int. J. Quantum Chem.* **1980**, *17*, 377–398.

(94) Kim, K.; Dupuis, M.; Lie, G.; Clementi, E. Revisiting Small Clusters of Water Molecules. *Chem. Phys. Lett.* **1986**, *131*, 451–456.

(95) Xantheas, S. S.; Dunning Jr, T. H. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, $n = 1 - 6$. I. Optimal Structures and Vibrational Spectra. *J. Chem. Phys.* **1993**, *99*, 8774–8792.

(96) Xantheas, S. S.; Dunning Jr, T. H. The Structure of the Water Trimer from Ab Initio Calculations. *J. Chem. Phys.* **1993**, *98*, 8037–8040.

(97) Ojamäe, L.; Hermansson, K. Ab Initio Study of Cooperativity in Water Chains: Binding Energies and Anharmonic Frequencies. *J. Phys. Chem.* **1994**, *98*, 4271–4282.

(98) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction Energies of Large Clusters from Many-Body Expansion. *J. Chem. Phys.* **2011**, *135*, 224102.

(99) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Van der Avoird, A. Predictions of the Properties of Water From First Principles. *Science* **2007**, *315*, 1249–1252.

(100) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Flexible, Ab Initio Potential, and Dipole Moment Surfaces for Water. I. Tests and Applications for Clusters up to the 22-mer. *J. Chem. Phys.* **2011**, *134*, 094509.

(101) Conte, R.; Qu, C.; Bowman, J. M. Permutationally Invariant Fitting of Many-Body, Non-Covalent Interactions with Application to Three-Body Methane–Water–Water. *J. Chem. Theory Comput.* **2015**, *11*, 1631–1638.

(102) Wang, Q.; Bowman, J. M. Two-Component, Ab Initio Potential Energy Surface for $CO_2$—$H_2O$, Extension to the Hydrate Clathrate, $CO_2@(H_2O)_{20}$, and VSCF/VCI Vibrational Analyses of Both. *J. Chem. Phys.* **2017**, *147*, 161714.

(103) Tang, K.; 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*, 3726–3741.

(104) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.

(105) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz Jr, J. L. Density-functional theory for fractional particle number: derivative discontinuities of the energy. *Phys. Rev. Lett.* **1982**, *49*, 1691.

(106) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density Functional Theory is Straying from the Path Toward the Exact Functional. *Science* **2017**, *355*, 49–52.

(107) Wasserman, A.; Nafziger, J.; Jiang, K.; Kim, M.-C.; Sim, E.; Burke, K. The Importance of Being Inconsistent. *Annu. Rev. Phys. Chem.* **2017**, *68*, 555–581.

(108) Engel, E.; Dreizler, R. M. *Density Functional Theory*; Springer, 2011; pp 109–217.

(109) Zhang, Y.; Yang, W. A Challenge for Density Functionals: Self-Interaction Error Increases for Systems with a Noninteger Number of Electrons. *J. Chem. Phys.* **1998**, *109*, 2604–2608.

(110) Liu, F.; Proynov, E.; Yu, J.-G.; Furlani, T. R.; Kong, J. Comparison of the Performance of Exact-Exchange-Based Density Functional Methods. *J. Chem. Phys.* **2012**, *137*, 114104.

(111) Tsuneda, T.; Hirao, K. Self-Interaction Corrections in Density Functional Theory. *J. Chem. Phys.* **2014**, *140*, 18A513.

(112) Bao, J. L.; Gagliardi, L.; Truhlar, D. G. Self-Interaction Error in Density Functional Theory: An Appraisal. *J. Phys. Chem. Lett.* **2018**, *9*, 2353–2358.

(113) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Delocalization Errors in Density Functionals and Implications for Main-Group Thermochemistry. *J. Chem. Phys.* **2008**, *129*, 204112.

(114) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.

(115) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. Binding Energy Curves from Nonempirical Density Functionals. I. Covalent Bonds in Closed-Shell and Radical Molecules. *J. Phys. Chem. A* **2005**, *109*, 11006–11014.

(116) Kim, M.-C.; Sim, E.; Burke, K. Ions in Solution: Density Corrected Density Functional Theory (DC-DFT). *J. Chem. Phys.* **2014**, *140*, 18A528.

(117) Kim, M.-C.; Park, H.; Son, S.; Sim, E.; Burke, K. Improved DFT Potential Energy Surfaces via Improved Densities. *J. Phys. Chem. Lett.* **2015**, *6*, 3802–3807.

(118) Kim, M.-C.; Sim, E.; Burke, K. Communication: Avoiding Unbound Anions in Density FFunctional Calculations. *J. Chem. Phys.* **2011**, *134*, 171103.

(119) Kim, Y.; Song, S.; Sim, E.; Burke, K. Halogen and Chalcogen Binding Dominated by Density-Driven Errors. *J. Phys. Chem. Lett.* **2018**, *10*, 295–301.

(120) Song, S.; Kim, M.-C.; Sim, E.; Benali, A.; Heinonen, O.; Burke, K. Benchmarks and Reliable DFT Results for Spin Gaps of Small Ligand Fe(II) Complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2304–2311.

(121) Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Phys. Rev. B* **1981**, *23*, 5048.

(122) Withanage, K. P.; Akter, S.; Shahi, C.; Joshi, R. P.; Diaz, C.; Yamamoto, Y.; Zope, R.; Baruah, T.; Perdew, J. P.; Peralta, J. E.; Jackson, K. A. Self-Interaction-free Electric Dipole Polarizabilities for Atoms and their Ions Using the Fermi-Löwdin Self-Interaction Correction. *Phys. Rev. A* **2019**, *100*, 012505.

(123) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(124) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(125) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximations Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868, Erratum: *ibid*. **78**, 1396 (1997).

(126) Mardirossian, N.; Ruiz Pestana, L.; Womack, J. C.; Skylaris, C.-K.; Head-Gordon, T.; Head-Gordon, M. Use of the rVV10 Nonlocal Correlation Functional in the B97M-V Density

Functional: Defining B97M-rV and Related Functionals. *J. Phys. Chem. Lett.* **2017**, *8*, 35–40.

(127) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(128) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(129) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two new Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(130) Mardirossian, N.; Head-Gordon, M. $\omega$B97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.* **2016**, *144*, 214110:1–23.

(131) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parameterization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104:1–19.

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

(133) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(134) 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.; Kús, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock

III, 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 Jr., 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.; 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.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Pieniazek, P. A.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sergueev, N.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Vanovschi, V.; 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 III, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer III, 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*, 184–215.

(135) Neese, F. Software Update: The ORCA Program System, version 4.0. *WIREs Comput. Mol. Sci.* **2017**, *8*, e1327:1–6.

(136) Murray, C. W.; Handy, N. C.; Laming, G. J. Quadrature Schemes for Integrals of Density

Functional Theory. *Mol. Phys.* **1993**, *78*, 997–1014.

(137) Lebedev, V. I. Quadratures on a Sphere. *USSR Comput. Math. & Math. Phys.* **1976**, *16*, 10–24.

(138) Dasgupta, S.; Herbert, J. M. Standard Grids for High-Precision Integration of Modern Density Functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38*, 869–882.

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

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

(141) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(142) Kong, J.; Gan, Z.; Proynov, E.; Freindorf, M.; Furlani, T. R. Efficient Computation of the Dispersion Interaction with Density-Functional Theory. *Phys. Rev. A* **2009**, *79*, 042510.

(143) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *jctc* **2012**, *8*, 527–541.

(144) Fanourgakis, G. S.; Xantheas, S. S. The Flexible, Polarizable, Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.

(145) Pérez, C.; Muckle, M. T.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy. *Science* **2012**, *336*, 897–901.

(146) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The Water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116–11119.

(147) Babin, V.; Paesani, F. The Curious Case of the Water Hexamer: Cage vs. Prism. *Chem. Phys. Lett.* **2013**, *580*, 1–8.

(148) Brown, S. E.; Götz, A. W.; Cheng, X.; Steele, R. P.; Mandelshtam, V. A.; Paesani, F. Monitoring Water Clusters "Melt" Through Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 7082–7088.

(149) Pedulla, J. M.; Vila, F.; Jordan, K. Binding Energy of the Ring Form of $(H_2O)_6$: Comparison of the Predictions of Conventional and Localized-Orbital MP2 Calculations. *J. Chem. Phys.* **1996**, *105*, 11091–11099.

(150) Cui, J.; Liu, H.; Jordan, K. D. Theoretical Characterization of the $(H_2O)_{21}$ Cluster: Application of an n-body Decomposition Procedure. *J. Phys. Chem. B* **2006**, *110*, 18872–18878.

(151) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction Energies of Large Clusters from Many-Body Expansion. *jcp* **2011**, *135*, 224102.

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

(153) Vuckovic, S.; Song, S.; Kozlowski, J.; Sim, E.; Burke, K. Density Functional Analysis: The Theory of Density-Corrected DFT. *J. Chem. Theory Comput.* **2019**, *15*, 6636–6646.

(154) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, n= 1–6. II. Analysis of Many-Body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

(155) Sharkas, K.; Wagle, K.; Santra, B.; Akter, S.; Zope, R. R.; Baruah, T.; Jackson, K. A.; Perdew, J. P.; Peralta, J. E. Self-Interaction Error Overbinds Water Clusters but Cancels in Structural Energy Differences. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11283–11288.

(156) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights Into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.

TOC graphic