Assessing the Interplay Between Functional-Driven and Density-Driven Errors in DFT Models of Water

Etienne Palos,[*,†] Eleftherios Lambros,[*,†] Steven Swee,[†,§] Jie Hu,[†,§] Saswata Dasgupta,[†] 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: epalos@ucsd.edu; elambros@ucsd.edu; fpaesani@ucsd.edu

Abstract

We investigate the interplay between functional-driven and density-driven errors in different density functional theory (DFT) approximations, and the implications of these errors for simulations of water with DFT-based data-driven many-body potentials. Specifically, we quantify density-driven errors in two widely used dispersion-corrected functionals derived within the generalized gradient approximation (GGA), namely BLYP-D3 and revPBE-D3, and two modern meta-GGA functionals, namely SCAN and B97M-rV. The effects of functional-driven and density-driven errors on the interaction energies are assessed for the water clusters of the BEGDB dataset. Further insight into the nature of functional-driven errors is gained from applying the absolutely localized molecular orbital energy decomposition analysis (ALMO-EDA) to the interaction energies, which demonstrates that functional-driven errors are strongly correlated with the nature of the interactions. We discuss cases where density-corrected DFT (DC-DFT) models display higher accuracy than the original DFT models, and cases where reducing the density-driven errors leads to larger deviations from the reference energies due to the presence of large functional-driven errors. Finally, molecular dynamics simulations are performed with data-driven many-body potentials derived from DFT and DC-DFT data to determine the effect that minimizing density-driven errors has on the description of liquid water. Besides rationalizing the performance of widely used DFT models of water, we believe that our findings unveil fundamental relations between the shortcomings of some common DFT approximations and the requirements for accurate descriptions of molecular interactions, which will aid the development of a consistent, DFT-based framework for data-driven simulations of condensed-phase systems.

1 INTRODUCTION

Accomplishing a unified, molecular-level understanding of the properties of water, from small gas-phase clusters to the thermodynamic limit, has been a paramount objective in the physical sciences since the first computer simulations.[1,2] Despite its simple molecular structure, the physics of water is complex, as it displays several anomalous properties in the condensed phase.[3] Liquid water is a topologically disordered system, where the perfectly tetrahedral hydrogen-bond network found in ice[4] is lost due to thermal fluctuations causing the hydrogen-bond network to rearrange within the picosecond to nanosecond timescales.[5,6] As a result, the free-energy landscape associated with the hydrogen-bond network in liquid water is determined by the subtle competition between energetic and entropic contributions, which is modulated by short-range low-order many-body interactions[7] that are difficult to represent using purely classical models.[8,9]

In principle, within density functional theory (DFT), the multidimensional Born-Oppenheimer potential energy surface of water can be exactly described by solving the Kohn-Sham equations. However, the exact density functional is unknown, which has motivated and continues to motivate the quest for a density functional approximation (DFA) capable of providing a quantitative description of both the molecular and thermodynamic properties of water.[10–33] It was found early on that the simplest DFA, the local spin density approximation (LSDA),[34–36] systematically overestimates the strength of hydrogen bonds and, consequently, is unable to provide even a qualitative description of the structure of liquid water.[37] Density functionals based on the general gradient approximation (GGA), such as the Perdew-Burke-Erzerhoff (PBE) functional,[38] yield more accurate energetics but do not eradicate the over-structuring of liquid water.[37,39] The systematic overbinding predicted by the LSDA and GGA functionals originates from their inability to exactly cancel self-Coulomb and self-exchange-correlation effects, an error known as the self-interaction error (SIE).[40]

Exchange-correlation functionals that include a dependence on both the gradient of the electron density and the kinetic energy density, known as meta-GGA functionals,[41] can often provide improved description of water clusters and have, recently, become accessible for *ab initio* molecu-

lar dynamics (AIMD) simulations, with and without including nuclear quantum effects (NQE).[32] To this end, meta-GGAs such as the range-separated B97M-rV[31,32,42] and the strongly constrained and appropriately normed (SCAN)[33,43] functionals have received particular attention in recent years.[44,45] For instance, it has been shown that classical and path-integral AIMD simulations of water with the B97M-rV functional display accuracy comparable to the more expensive dispersion-corrected hybrid revPBE0-D3 functional.[32] Several studies have been reported focusing on assessing the accuracy of SCAN and SCAN-based data-driven models for various properties of water, including interaction, binding, and many-body energies of water clusters, as well as the structure, density, and diffusion of liquid water.[46–51]

In recent years, there has also been significant progress in understanding how different DFAs deviate from the exact density functional. Rigorously, one can quantify how a DFA deviates from the exact density functional by decomposing the energy difference into functional-driven and density-driven errors.[52] This is the essence of density-corrected DFT (DC-DFT), a framework that has emerged as a generalization of HF-DFT,[53,54] for improving the accuracy of DFAs through mitigation of density-driven errors in a cost-effective manner.[55–57] It has recently been demonstrated that DC-DFT may be particularly beneficial for the treatment of diversely-bonded systems.[58,59] The most common flavor of DC-DFT is HF-DFT in which the Kohn-Sham equations are solved using the system's density computed within the Hartree-Fock approximation. Since the HF orbitals are SIE-free and HF scales as $O(N^4)$, HF-DFT is closer in computational cost to DFT ($O(N^3)$) compared to Møller-Plesset perturbation theory[60,61] and coupled cluster (CC) theory.[62]

Building upon recent studies that focus on improving DFT-based representations of water, such as the development of the generalized MB-QM[63] framework and the many-body analysis of DC-SCAN reported in ref,[50] we investigate here the effect of functional-driven and density-driven errors on describing the properties of water from the gas to the condensed phase within different DFAs. Specifically, we asses the density-driven errors of various semi-local functionals, and apply the MB-DFT formalism[7,49,50,63] to density-corrected functionals in order to derive a set of data-driven many-body potential energy functions (PEFs) for water simulations. We demonstrate that in

cases where density-driven errors dominate, the derived MB-DFT(DC) PEFs elevate the accuracy for water simulations for GGA and meta-GGA functionals. While this study is focused on water, we believe that our findings provide fundamental insights into the shortcomings of some popular DFAs which have direct implications for the accuracy of DFT-based data-driven models that are nowadays routinely used in computer simulations of aqueous systems.

2 THEORY

2.1 Many-Body Potential Energy Functions

Consider an N-body system composed of N monomers (atoms or molecules). The potential energy of this system, E_N, is rigorously defined by the many-body expansion (MBE) which expresses E_N as the sum of n-body energy contributions,[64]

$$E_N(\mathbf{r}_1,..,\mathbf{r}_N) = \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) + \sum_{i=1}^{N}\sum_{j>i}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) + \sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{k>j>i}^{N} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + ... + \varepsilon_{\mathrm{NB}}(\mathbf{r}_1,..,\mathbf{r}_\mathrm{N}) \tag{1}$$

Here, $\varepsilon_{1\mathrm{B}}$ represents the distortion energy of an isolated monomer, such that $\varepsilon_{1\mathrm{B}}(\mathbf{r}_i) = E(\mathbf{r}_i) - E(\mathbf{r}_{\mathrm{eq},i})$ where $E(\mathbf{r}_{\mathrm{eq},i})$ is the energy of the ith monomer in its equilibrium geometry ($\mathbf{r}_{\mathrm{eq},i}$). The n-body energies, $\varepsilon_{\mathrm{nB}}$, are defined recursively for $1 \leq n \leq N$ by the expression

$$\begin{aligned}\varepsilon_{\mathrm{nB}} = E_\mathrm{n}(1,...,n) - \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) - \sum_{i=1}^{N}\sum_{i<j}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) - \sum_{i=1}^{N}\sum_{i<j}^{N}\sum_{i<j<k} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) - \cdots \\ \cdots - \sum_{i<j<k<...}^{N} \varepsilon_{(\mathrm{n}-1)\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k,...)\end{aligned} \tag{2}$$

It has been shown that the MBE converges quickly for nonmetallic systems, e.g., molecular systems with large band gaps and/or localized electron densities. Water, our system of interest, has a band gap of ~ 9 eV, and the sum of the two-body (ε_{2B}) and three-body (ε_{3B}) energy contributions correspond to $\sim 90-95\%$ of the total interaction energy.[64–71] Given its relatively large band gap and the fast convergence of the MBE, several many-body potentials derived from high-level

quantum-mechanical data have been reported in the literature for water,[72–78] and more recently for various aqueous systems.[79–84]

In order to asses how the mitigation of density-driven errors in different DFAs translates into the ability of a given density functional to represent the properties of water from gas-phase clusters to the liquid phase, we develop a family of density-corrected many-body PEFs called MB-DFT(DC) PEFs. Briefly, a MB-DFT(DC) PEF approximates eq 1 to the sum of explicit 1B, 2B, and 3B terms, along with V_{pol}, an implicit many-body term based on classical polarization representing all n-body interactions for $n > 3$:

$$E_N^{\mathrm{MB-DFT(DC)}}(\mathbf{r}_1,..,\mathbf{r}_N) = \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) + \sum_{i>j}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) + \sum_{i>j>k}^{N} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + V_{\mathrm{pol}} \tag{3}$$

Within the MB-DFT(DC) formalism, each term of eq 3 is fitted to reproduce the value calculated with the corresponding DC-DFT model. Further details regarding the calculations of each ε_{nB} term of the MB-DFT(DC) PEFs can be found in the Supporting Information as well as in our previous work.[63]

2.2 Density-Corrected DFT

The form of the exact functional $E[\rho(\mathbf{r})]$ that determines the ground-state electronic energy is unknown. For this reason, all practical applications of DFT employ approximations of the functional,

$$\tilde{E}[\rho(\mathbf{r})] = \tilde{F}[\rho(\mathbf{r})] + \int d\mathbf{r} V_{\mathrm{ext}}(\mathbf{r})\rho(\mathbf{r}) \tag{4}$$

The approximate one-electron density functional $\tilde{F}[\rho]$ is independent of the second term of eq 4 where $V_{\mathrm{ext}}(\mathbf{r})$ is the external potential acting on the N-body system, and is defined as

$$\tilde{F}[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + \tilde{V}_{\mathrm{XC}}[\rho(\mathbf{r})] \tag{5}$$

Here, $T_s[\rho(\mathbf{r})]$ is the Kohn-Sham (noninteracting) kinetic energy, and $J[\rho(\mathbf{r})]$ is the Coulomb electron-electron interaction. The last term of eq 5, $\tilde{V}_{\mathrm{XC}}[\rho(\mathbf{r})]$, is the exchange-correlation potential containing the non-classical electron-electron interaction terms. Note that approximating $F[\rho]$ implies that all physical quantities determined by $\rho(\mathbf{r})$ will also be approximate.

Let $\tilde{\rho}$ and $\tilde{E}$ be the approximate ground-state electron density and energy, respectively. The deviation of a DFA from the exact treatment can be quantified through the classification of functional-driven and density-driven errors, with the overall error on the energy given by

$$\Delta E = \tilde{E}[\tilde{\rho}] - E[\rho] = \Delta E_{\mathrm{F}} + \Delta E_{\mathrm{D}} \tag{6}$$

Here, ΔE_{F} defines the functional-driven error, i.e., the error made by $\tilde{F}$ on $\rho(\mathbf{r})$:

$$\Delta E_{\mathrm{F}} = \tilde{F}[\rho] - F[\rho] \tag{7}$$

The error due to $\tilde{\rho}(\mathbf{r})$ defines the density-driven error ΔE_D:

$$\Delta E_{\mathrm{D}} = \tilde{E}[\tilde{\rho}] - \tilde{E}[\rho] \tag{8}$$

In principle, provided with the exact density, DC-DFT can enable a systematic analysis of functional-driven and density-driven errors in any DFA. Since the exact density is unknown, the most common flavor of DC-DFT is HF-DFT, where ρ^{HF}, the density of the system computed at the Hartree-Fock level of theory, is used to solve the Kohn-Sham equations in a non-self-consistent fashion. Herein, DC-DFT is employed in the form of HF-DFT, and we preserve the DC-DFT notation as the results discussed in Section 3 do not strictly depend on ρ^{HF}. The density-corrected energy E^{DC} is then approximated non-self-consistently as:

$$E^{\mathrm{DC}} \simeq E^{\mathrm{HF}} + \left(\tilde{E}_{\mathrm{XC}}[\rho^{\mathrm{HF}}] - E_X^{\mathrm{HF}} \right) \tag{9}$$

Self-interaction induces the delocalization error, which combined constitute the density-driven errors. By construction, ρ^{HF} is SIE-free, which implies that E^{DC} in eq 9 can be assumed to be, to a good approximation, free from density-driven errors.[85] However, it should be noted that the lack of correlation in HF introduces artificial overlocalization to the electron density.

2.3 Second-Generation ALMO-EDA

For a given density functional $\tilde{F}[\rho(\mathbf{r})]$, the second generation absolutely localized molecular orbital energy decomposition analysis (ALMO-EDA) method (often referred to as EDA2) decomposes the computed intermolecular interaction energy, E_{int}, into a sum of physical contributions.[86,87] In brief, the strength of the second-generation ALMO-EDA method lies in the orthogonal decomposition of the density matrix associated with the frozen wavefunction of the system into a sum over a set of individual fragments with modified density matrices. The interaction energy can then be separated into independent contributions as

$$E_{\mathrm{int}} = E_{\mathrm{pol}} + E_{\mathrm{frz}} + E_{\mathrm{CT}} \tag{10}$$

where E_{pol}, E_{frz}, and E_{CT} are the polarization, unrelaxed frozen, and charge-transfer energies of the overall system, respectively.

In eq 10, E_{pol} describes the induced electrostatic interactions resulting from the relaxation of $\tilde{\rho}_k$ such that each monomer in k polarizes every other monomer within its polarization subspace.[86,87] E_{frz} can be further separated into dispersion-free frozen (E_{DF}) and dispersion (E_{disp}) contributions:

$$E_{\mathrm{frz}} = \underbrace{E_{\mathrm{elec}} + E_{\mathrm{Pauli}}}_{E_{\mathrm{DF}}} + E_{\mathrm{disp}} \tag{11}$$

Here, E_{elec} is the Coulomb interaction between distorted fragment densities $\tilde{\rho}_k$, E_{Pauli} is the Pauli repulsion, and E_{disp} is the 2B dispersion energy, evaluated as the difference between the total exchange-correlation energy and the dispersion-free portion which is calculated using an auxil-

iary dispersion-free functional. In this work, E_{disp} is isolated from the other two energy terms that constitute E_{DF}. The charge transfer energy (E_{CT}), which describes the donor-acceptor orbital interactions between fragments, is calculated by subtracting E_{frz} and E_{pol} from the interaction energy. For further details regarding the ALMO-EDA method, we refer the reader to the original references.[86,87]

2.4 COMPUTATIONAL DETAILS

2.4.1 Electronic Structure Calculations

All electronic structure calculations were performed with a development version of the Q-Chem 5 software[88] using the aug-cc-pVQZ basis set[89,90] The DC-DFT calculations were carried out in two steps: an initial self-consistent Hartree-Fock calculation was performed to generate the HF orbitals that were then used in the subsequent non-self-consistent DFT calculation to obtain the corresponding DC-DFT energy. The exchange-correlation functionals considered in this study are: the Becke-exchange Lee-Yang-Parr-correlation (BLYP)[91,92] and revised Perdew-Burke-Ernzerhof (revPBE)[38,93] GGA functionals, with empirical corrections to the 2B dispersion energy calculated using the Grimme D3 parameters[94] as well as the optimized-power D3 parameters,[95] and the SCAN[43] and the B97M-rV[42] meta-GGA functionals.

The $(H_2O)_n = 2-10$ structures used in the analyses presented in Sections 3.1 and 3.4 were obtained from the BEGDB dataset,[96] with the structures optimized at the RI-MP2/aug-cc-pVDZ level of theory.[97] The $(H_2O)_6$ geometries used in the analysis of Section 3.4 were obtained from ref 98, and were originally optimized at the MP2 level of theory.

The interaction energies discussed in Sections 3.1 and 3.4 are defined as the difference between the total energy of the $(H_2O)_n$ cluster ($E_{n-\text{mer}}$) and the sum of the energies of the individual n water molecules in the same distorted geometries as in the cluster (E^{H_2O}),

$$E_{int} = E_{n\text{-mer}} - \sum_{i=1}^{n} E_i^{H_2O} \tag{12}$$

2.4.2 MB-DFT and MB-DFT(DC) Potential Energy Functions

Four MB-DFT and their corresponding MB-DFT(DC) PEFs were developed for the BLYP-D3, revPBE-D3, B97M-rV and SCAN functionals, and the density-corrected functionals DC-BLYP-D3, DC-revPBE-D3, DC-B97M-rV, and DC-SCAN, respectively. The 1B, 2B, and 3B training sets used in the development of both the MB-DFT and MB-DFT(DC) PEFs comprise 5000 monomers, 42508 dimers, and 12347 trimers, respectively. The 1B training set was generated using the MB-Fit software,[99] following the procedure described in our previous studies.[63,83,84] The 2B and 3B terms of eq 3 were developed using the same training sets used in the development of the MB-pol PEF.[76,77] Further details related to the development of the MB-DFT and MB-DFT(DC) PEFs can be found in the Supporting Information.

2.4.3 Many-Body Molecular Dynamics Simulations

All molecular dynamics (MD) simulations were performed in periodic boundary conditions for $N = 256$ water molecules in the canonical (NVT: constant number of molecules, volume, and temperature) and isobaric-isothermal (NPT: constant number of molecules, pressure, and temperature) ensembles using the MBX/i-PI interface.[100] NVT simulations were carried out at a temperature of 298 K, and the system's density was fixed to the experimental value of 0.997 g·cm^{-3} using a cubic box of length $\ell = 19.7295$ Å. The NPT simulations were carried out at a pressure 1 atm and a temperature of 298 K. In both NVT and NPT of simulations, the temperature was controlled by a Langevin thermostat with a relaxation time of 25 fs, while the pressure in the NPT simulations was controlled by the Bussi-Zykova-Timan-Parrinello barostat[101] with a relaxation time of 0.25 ps. The equations of motion were integrated with a time step of 0.5 fs. After equilibration, all MD simulations were carried out for 100 ps.

Besides calculating the radial distribution functions (RDFs), the structure of liquid water predicted by the MD simulations carried out with the different MB-DFT and MB-DFT(DC) PEFs is

characterized by analyzing the tetrahedral order parameter q_{tet} defined by:[102]

$$q_{\text{tet}} = 1 - \frac{3}{8} \cdot \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\psi_{jk}) + \frac{1}{3} \right) \tag{13}$$

Here, ψ_{jk} is the angle between the oxygen of the central water molecule and the oxygen atoms of the two neighboring water molecules. When $q_{\text{tet}} = 1$, the water molecules are in a perfect tetrahedral arrangement, and $q_{\text{tet}} = 0$ represents the ideal gas limit.

3 Results and Discussion

3.1 Interaction Energies for the $(H_2O)_{n=2-10}$ Clusters

The interplay between functional-driven and density-driven errors effectively determines the accuracy of a density functional. Correcting the density-driven errors can then reveal, to a good approximation, how the functional-driven errors affect the ability of a given density functional to correctly describe a molecular system. We begin to develop our understanding of these errors for water simulations by calculating the interaction energies E_{int} of the 38 water clusters, $(H_2O)_n$ where $2 \leq n \leq 10$, contained in the BEGDB dataset.[96,97] The interaction energy E_{int} is calculated using selected density functionals that are routinely used in water simulations, namely, BLYP-D3(0), BLYP-D3(op), revPBE-D3(0), revPBE-D3(op), B97M-rV and SCAN, as well as their density-corrected counterparts, DC-BLYP-D3(0), DC-BLYP-D3(op), DC-revPBE-D3(0), DC-revPBE-D3(op), DC-B97M-rV and DC-SCAN. The errors in E_{int} associated with the DFT and DC-DFT calculations are defined with respect to the corresponding CCSD(T) reference values of the BEGDB dataset. As defined in Section 2, in the case of dispersion-corrected GGA functionals, we evaluate the effect of the dispersion correction on E_{int}, comparing the energies obtained with the original Grimme D3 parameters, DFT-D3(0),[94] with the more recently developed optimized-power D3 parameterization, DFT-D3(op).[95]

The interaction energies calculated with the DFT and DC-DFT functionals considered in this

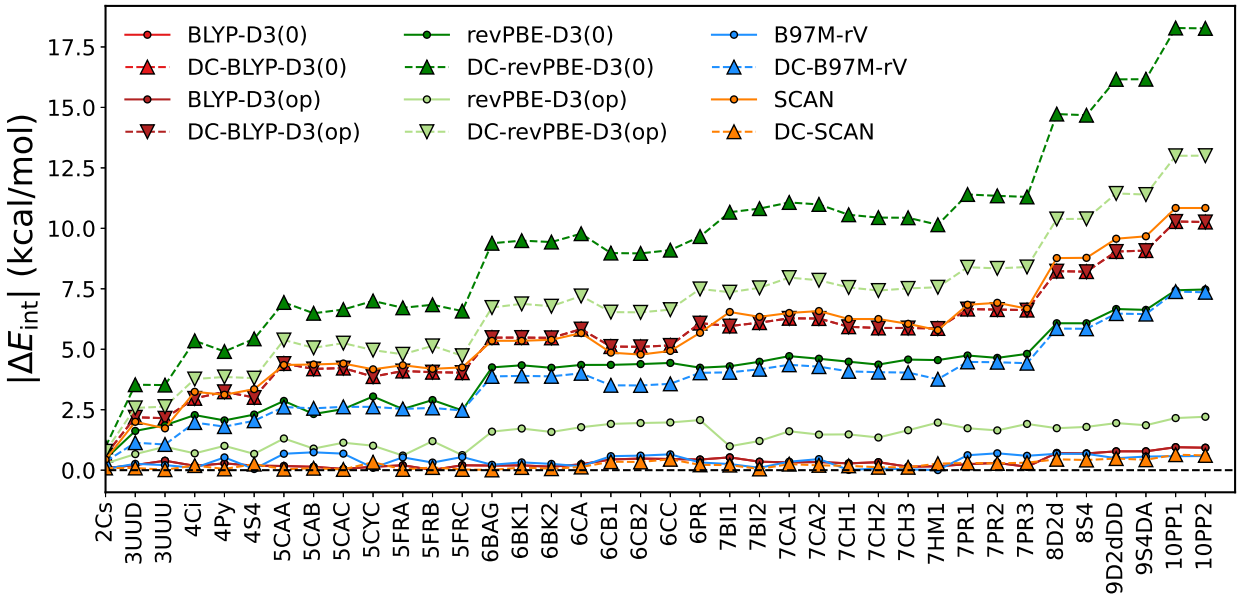

Figure 1: Absolute error of the interaction energy, $|\Delta E_{\text{int}}| = |E_{\text{int}}^{\text{model}} - E_{\text{int}}^{\text{ref}}|$ for the 38 $(H_2O)_{n=2-10}$ structures within the BEGDB data set.

study are reported in Figure 1 for all 38 clusters in the BEGDB dataset. All DFT and DC-DFT models predict E_{int} within ~ 0.5 kcal/mol of the CCSD(T)/CBS value for the water dimer. However, larger deviations become apparent starting from the trimer. Figure 1 shows that BLYP-D3(0) provides the closest agreement with the CCSD(T) results for all the 38 clusters, followed by B97M-rV, consistently yielding interaction energies that are $\sim$0.25 kcal/mol higher than the reference values. In the case of revPBE-D3(0), E_{int} is consistently overestimated by $\sim$3 kcal/mol, while, on the other end, SCAN systematically underestimates E_{int}. In addition, the D3(op) dispersion correction quantitatively improves the accuracy of revPBE, compared to the standard "zero-damping" revPBE-D3(0). The difference between revPBE-D3(op) and revPBE-D3(0) has far-reaching consequences, including significant effects on individual n-body energies which are discussed in the following sections. While D3(op) also quantitatively improves the performance of BLYP, the effect of the D3(op) damping is minimal. The implications of these results are discussed in detail in Section 3.3.

The density-correction has the same general effect on all density functionals examined in this study, shifting the interaction energies such that $E_{\text{int}}[\rho^{\text{HF}}] > E_{\text{int}}[\tilde{\rho}]$ is always true, as the density-

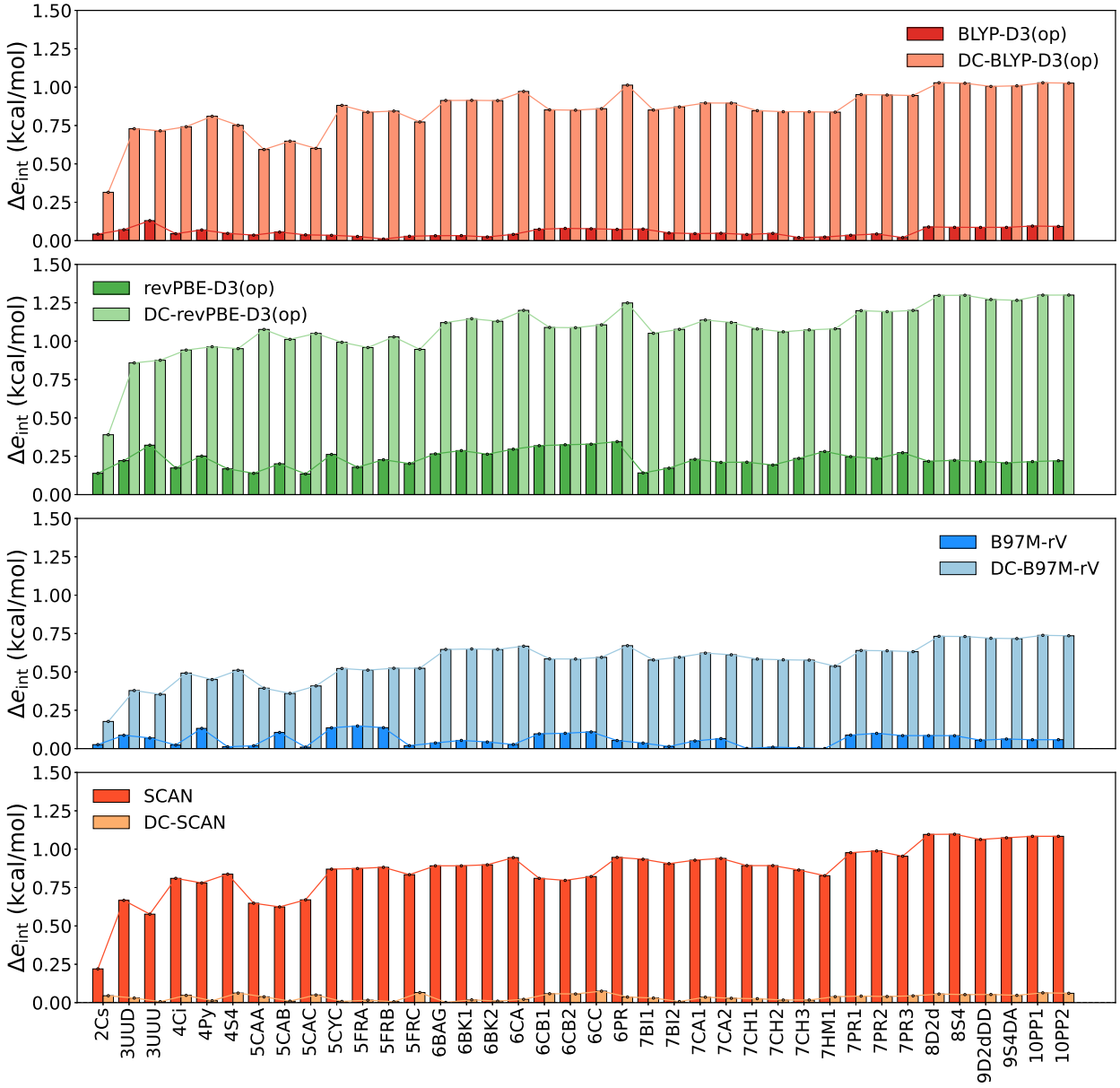

Figure 2: Absolute error per monomer for the total interaction energy $|\Delta e_{\text{int}}|$ calculated relative to CCSD(T)/CBS reference values using all density functionals considered in this study. $|\Delta e_{int}|$ is shown for all the 38 structures, $(H_2O)_{n=2-10}$, of the BEGDB dataset.

driven errors artificially overstabilize the water clusters by means of SIE and delocalization errors, albeit it does not necessarily imply a systematic trend of error increase/decrease. Minimizing the density-driven errors drives the energetics of BLYP-D3, revPBE-D3, and B97M-rV further away from the CCSD(T)/CBS reference. Interestingly, the effect on SCAN is a remarkable increase in accuracy, such that DC-SCAN shows excellent agreement with CCSD(T)/CBS as already noted in

ref 50.

To gain further insight into the results of this analysis, in Figure 2 we report the absolute error per monomer for self-consistent and density-corrected GGA-D3(op), B97M-rV, and SCAN functionals. Here, the absolute error per monomer Δe_{int} for a given $H_2O)_n$ cluster is given by

$$\Delta e_{\text{int}} = \frac{\left| E_{\text{int}}^{\text{model}} - E_{\text{int}}^{\text{ref}} \right|}{n} \tag{14}$$

BLYP-D3(op) consistently displays small Δe_{int} across the BEGDB dataset. Minimizing the density-driven error increases the magnitude of Δe_{int} for BLYP-D3, revPBE-D3, and B97M-rV. In this regard, the Hartree-Fock orbitals seem to have the same effect on these three functionals, resulting in the following trend: $E_{\text{int}}^{\text{DC-revPBE-D3}} > E_{\text{int}}^{\text{DC-BLYP-D3}} > E_{\text{int}}^{\text{DC-B97M-rV}}$. Within both DFT and DC-DFT, revPBE-D3 (with either the D3(0) or the D3(op) correction) displays the largest Δe_{int}, among these three functionals. Remarkably, DC-SCAN represents a $\sim$ 20-fold increase in accuracy over SCAN, consistently predicting interaction energies at "gold standard" CCSD(T)/CBS accuracy.[50]

The statistics of the analysis shown in Figure 2 are summarized in Table 1, where the mean absolute errors (MAE) per monomer, are reported in kcal/mol for $(H_2O)_n$ for $n = 2-10$. The MAE per monomer $\langle \Delta e_{\text{int}} \rangle$ is given by

$$\langle \Delta e_{\text{int}} \rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta e_{\text{int},i}}{n} \tag{15}$$

where n is the number of water molecules (i.e., monomers) in a given $(H_2O)_n$ cluster, and N is the number of isomers of a given $(H_2O)_n$ cluster in the BEGDB dataset. Consistently, SCAN displays the largest MAE within DFT, but the lowest MAE within DC-DFT. As an example, we consider the water hexamer, where $\langle \Delta e_{\text{int}} \rangle$ is 0.87 kcal/mol for SCAN and 0.03 kcal/mol for DC-SCAN. Interestingly, the opposite trend is observed for the revPBE results, for which $\langle \Delta e_{\text{int}} \rangle$ is 0.30 kcal/mol for revPBE-D3(op) and 1.14 kcal/mol for DC-revPBE-D3(op). Although not explicitly summarized in Table 1, a remark is to be made about revPBE-D3(0). All the $\langle \Delta e_{\text{int}} \rangle$ values obtained with revPBE-D3(0) are by far the largest among the functionals examined in this study, with

Table 1: Mean absolute errors (MAE) per monomer in the interaction energies, $\langle \Delta e_{\text{int}} \rangle$, for (H_2O)n clusters with $n = 2-10$ calculated relative to the CCSD(T)/CBS reference values using all the DFT and DC-DFT models examined in this study. $\langle \Delta e_{\text{int}} \rangle$ is reported in kcal/mol.

Cluster	Method	BLYP-D3(op)	revPBE-D3(op)	B97M-rV	SCAN
$(H_2O)_2$	DFT	0.04	0.13	0.02	0.22
	DC-DFT	0.31	0.39	0.17	0.05
$(H_2O)_3$	DFT	0.10	0.27	0.08	0.62
	DC-DFT	0.35	0.86	0.36	0.02
$(H_2O)_4$	DFT	0.05	0.19	0.06	0.81
	DC-DFT	0.76	0.95	0.48	0.04
$(H_2O)_5$	DFT	0.03	0.19	0.08	0.77
	DC-DFT	0.74	1.00	0.46	0.02
$(H_2O)_6$	DFT	0.05	0.30	0.06	0.87
	DC-DFT	0.91	1.14	0.63	0.03
$(H_2O)_7$	DFT	0.04	0.22	0.04	0.91
	DC-DFT	0.88	1.12	0.60	0.03
$(H_2O)_8$	DFT	0.09	0.22	0.08	1.10
	DC-DFT	1.03	1.29	0.73	0.05
$(H_2O)_9$	DFT	0.09	0.21	0.06	1.06
	DC-DFT	1.01	1.27	0.72	0.05
$(H_2O)_{10}$	DFT	0.09	0.22	0.05	1.08
	DC-DFT	1.03	1.30	0.74	0.06

$\langle \Delta e_{\text{int}} \rangle$ of ~ 2 kcal/mol within DFT and ~ 1.5 kcal/mol within DC-DFT. The notable difference between revPBE-D3(0) and revPBE-D3(op) is due to the contribution of the dispersion correction to the total energy, indicating that the parameters in the empirical D3(0) correction for the revPBE-D3(0) functional are inappropriate for water, a point that will be further illustrated in Section 3.3. While minimizing the density-driven errors in revPBE-D3(0) improves the interaction energies, DC-revPBE-D3(0) still displays a larger MAE (by ~ 0.5 kcal/mol) than DC-revPBE-D3(op).

In the case of BLYP, a minimal effect is seen when comparing the D3(0) and D3(op) corrections, as their Δe_{int} values are effectively identical (0.05 kcal/mol within DFT and 0.85 kcal/mol within DC-DFT), with the difference between the MAE of BLYP-D3(0) and BLYP-D3(op) being less than 0.01 kcal/mol in both DFT and DC-DFT. Similar to revPBE, the energetics of BLYP worsens upon applying the density correction, indicating that the errors in BLYP calculations for water are primarily functional-driven.

Despite both being meta-GGA functionals, SCAN and B97M-rV show different behavior. First, B97M-rV outperforms SCAN in self-consistent calculations of E_{int}, evidenced by an $\langle \Delta e_{\text{int}} \rangle$ of ~ 0.06 kcal/mol for the $(H_2O)_6$ cluster, which is ~ 14 times lower than that of SCAN. On average, B97M-rV displays a MAE that is more than 10 times smaller than that of SCAN for clusters with $n = 2-10$ water molecules. These results are in agreement with previous assessments where B97M-rV was found to be, overall, the most accurate meta-GGA functional.[103] In particular, B97M-rV was shown to be very accurate for non-covalent interactions.[42] Figure 2 shows that both B97M-rV and BLYP-D3(op) display similar behavior, in that they predict E_{int} within fair agreement with the CCSD(T)/CBS reference values. Hence, upon using a more accurate density, the worsening of the interaction energy predictions indicates error compensation between E_{D} and E_{F} for both BLYP-D3(op) and B97M-rV. Similar trends have been observed for rev-PBE-D3(op) as well, where the density correction deteriorates the energetics with respect to the reference values, suggesting a similar error cancellation.

Unlike the other semi-local functionals, SCAN behaves differently as the SCAN interaction energies are significantly worse than BLYP-D3(op), revPBE-D3(op) and B97M-rV, displaying $\langle \Delta e_{\text{int}} \rangle$ within 0.62-1.08 kcal/mol for (H_2O)n clusters with $n = 3-10$. Use of the Hartree-Fock orbitals leads to remarkable agreement with the reference values, as $\langle \Delta e_{\text{int}} \rangle$ lies within 0.02 to 0.06 kcal/mol for all (H_2O)n clusters with $n = 2-10$. This implies that the errors is SCAN calculations for water are primarily dominated by density-driven errors (E_D in eq 6).

The magnitude of density-driven errors can be approximated by taking the difference between the DFT and DC-DFT energies. From Figure 2, it is apparent that E_{D} is smallest for B97M-rV among the investigated functionals. In this spirit, it is also evident that SCAN is the only functional where it is clear that the magnitude of the density-driven errors is significantly larger than that of the functional-driven errors. At this point, it is relevant to ask what makes SCAN different from B97M-rV, a contemporary meta-GGA functional, in terms of the magnitude of functional- and density-driven errors for water. The primary difference between these two meta-GGA functionals is that SCAN is a non-empirical density functional,[43] whereas B97M-rV is a combinatorially op-

timized semi-empirical functional.[42] The SCAN functional was constructed to satisfy all 17 exact constraints known for semi-local meta-GGA functionals. Thus, if a more accurate density is provided, SCAN performs remarkably well due to its small functional-driven errors, as it satisfies all the necessary constraints and was not *a priori* optimized for any particular class of system. On the other hand, B97M-rV was developed with 12 fitted parameters: 4 exchange parameters, 4 same-spin correlation parameters, and 4 opposite-spin correlation parameters. These parameters were optimized using an extensive range of datasets. However, it is yet to be seen if B97M-rV satisfies all the required constraints known for meta-GGA functionals. The fact that $\Delta E_{\mathrm{F}}^{\mathrm{SCAN}} < \Delta E_{\mathrm{F}}^{\mathrm{B97M-rV}}$ indicates that, while B97M-rV is significantly more accurate than SCAN in predicting the interaction energies of water clusters (or similar systems), the physics behind SCAN is potentially more robust, resulting in the possibility of systematically improving its accuracy for a wider class of diversely bonded systems.

Finally, our analysis reveals a non-linear dependence of both functional-driven and density-driven errors on system's size (see Figures S2 and S3 in the Supporting Information). In the case of self-consistent BLYP-D3, revPBE-D3 and B97M-rV, $\langle \Delta e_{\mathrm{int}} \rangle$ does not increase with n, while it does for self-consistent SCAN. Using ρ^{HF} to calculate the interaction energies allows for effectively approximating E_{F} because the remaining density-driven error is in principle $E[\tilde{\rho}^{\mathrm{HF}}] - E[\rho^{\mathrm{exact}}] \ll \Delta E_{\mathrm{D}}$, where ΔE_{D} is defined in eq 8. In this regard, our analysis indicates that ΔE_{F} depends on the system's size for DFAs where $\Delta E_{\mathrm{F}} > E_{\mathrm{D}}$. The SCAN functional applied to water, however, is a special case since $\Delta E_{\mathrm{D}} > \Delta E_{\mathrm{F}}$ such that the density-driven errors drive the size-dependence of the error on the interaction energy. As a consequence, $\langle \Delta e_{\mathrm{int}}^{\mathrm{DC-SCAN}} \rangle$ is effectively independent on the system's size. The size-dependence of functional-driven and density-driven errors in semi-local functionals shines light into the origin of the error cancellation that allows certain functionals to be successful in representing some properties of liquid water while, at the same time, failing to predict the properties of water clusters.[9,104–108]

3.2 Density Analysis

The magnitude of density-driven errors is reflected in ΔE_{D}, which defines how well a given $\tilde{F}[\rho]$ approximates the electron density of the system in question. It is well known that semi-local functionals do not manage to cancel out the entirety of self-exchange and self-correlation, often leading to large SIE. Hybrid functionals reduce the SIE by describing $\tilde{V}_{\mathrm{XC}}[\rho]$ as a fractional sum of exact exchange and the exchange-correlation energy calculated by a GGA or meta-GGA functional. Because the SIE constitutes a significant fraction of ΔE_{D}, it is reasonable to assume that the density-driven error of a hybrid functional[109] is smaller than that of its semi-local parent functional.

In this section, we aim to understand density-driven errors directly from the electronic densities. To this end, we approximate the density-error as the difference between the ground-state electronic density predicted by a given functional with respect to the Hartree-Fock density, i.e., $\Delta\rho(\mathbf{r}) \approx \tilde{\rho}(\mathbf{r}) - \rho^{\mathrm{HF}}(\mathbf{r})$. Figure 3a-b show 3-dimensional representations of $\Delta\rho(r)$ for the water

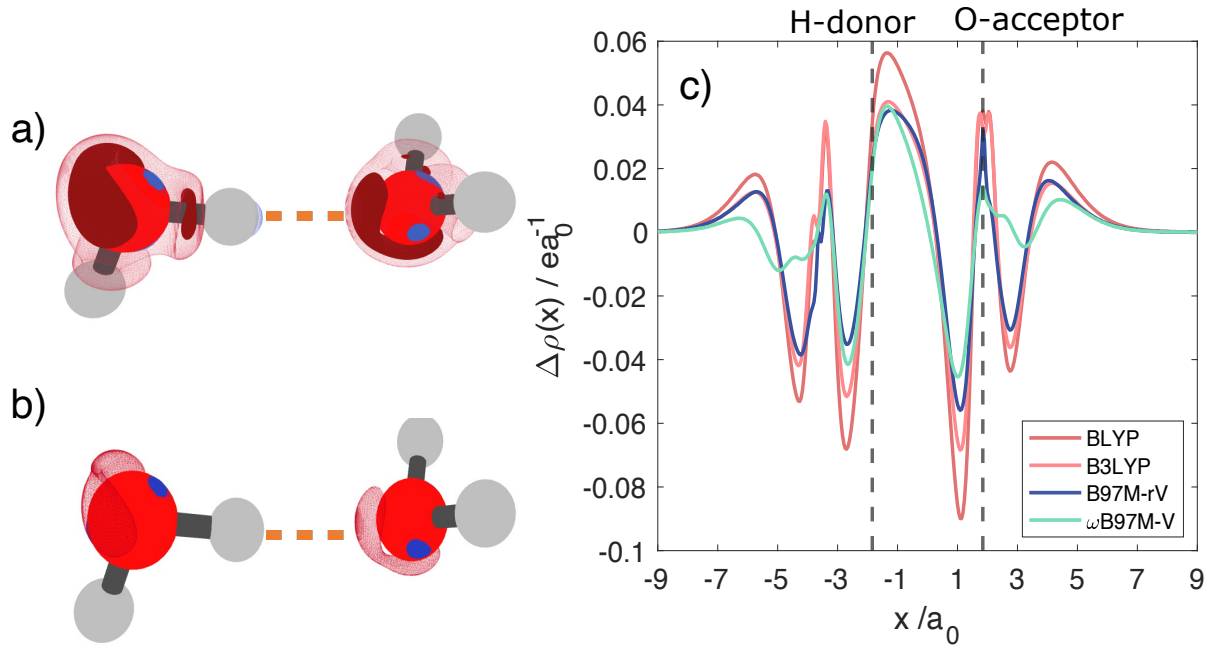

Figure 3: Illustration of the density-difference in the donor-acceptor pair of the water dimer for a) BLYP relative to B3LYP, and b) B97M-rV relative to ωB97M-V. The density error, defined as $\Delta\rho(x) \approx \tilde{\rho}(x) - \rho^{\mathrm{HF}}$ is shown in c) for the four functionals in units of e/a_0.

dimer calculated with BLYP (transparent iso-surface) and B3LYP (solid iso-surface) (panel a), and B97M-rV (transparent iso-surface) and ωB97M-V (solid iso-surface). The iso-surface $|\Delta\rho(r)|$ = 0.008 e/a_0^3 is shown in both cases, where the red shade corresponds to negative values, i.e., it indicates the withdrawal of electron density, and blue shades corresponds to positive values, i.e., it indicates the deposition of electron density.

The BLYP iso-surface has a surface area S of 37.35 a_0^2 which gets contracted to 17.47 a_0^2 for the B3LYP functional. Compared to B3LYP, BLYP provides a higher electron density on the donor hydrogen of the water molecule which acts as the hydrogen-bond donor. In constrast, the electron density decreases near the oxygen atoms, which becomes more evident for BLYP. This leads to comparatively stronger hydrogen bonds for BLYP, which leads to over-binding and, consequently, artificially more dense liquid phase.[110] Addition of a fraction of exact exchange substantially reduces the density-driven errors , which improves the description of hydrogen-bonding interactions in water.[63] Climbing up to the third rung of Jacob's ladder of DFAs partially reduces over-binding, as S is equal to 12.55 a_0^2 for the iso-surface calculated with B97M-rV, and is further reduced to $\sim$ 7.54 a_0^2 for the range-separated and hybrid ωB97M-V functional. The smaller density-driven error associated with B97M-rV is in line with the results of Figure 2, which show that B97M-rV displays the lowest density-driven errors among all the functionals examined in this study.

Figure 3c shows $\Delta\rho$ projected along the hydrogen-bond axis ($\hat{x}$) between the H atom of the donor molecule and the O atom of the acceptor molecules. Similar to Figures 3a-b, the H atom of the donor molecule shows an inflated electron density, which is drastically overestimated by BLYP. The other three density functionals show a somewhat similar $\Delta\rho$ profiles near the H atom of the donor molecule. For all density functionals, $\Delta\rho$ shows a dip near the O atom of the acceptor molecule ($x \sim$3 a_0). This dip represents partial removal of electron density and follows the trend $\Delta\rho(x)_{\mathrm{BLYP}} < \Delta\rho(x)_{\mathrm{B3LYP}} < \Delta\rho(x)_{\mathrm{B97M-rV}} < \Delta\rho(x)_{\omega\mathrm{B97M-V}}$.

Overall, the analyses of Figure 3 demonstrate that the self-consistent GGA density provides a poor description of the electron density along the hydrogen bond of the water dimer, which results in an incorrect description of hydrogen-bonding interactions in water.

3.3 ALMO-EDA calculations for $(H_2O)_{n=2-10}$

The accuracy of BLYP-D3, revPBE-D3, B97M-rV, and SCAN is further assessed by means of second-generation ALMO-EDA calculations that provide fundamental insights into the ability of these functionals to accurately represent E_{int} as a sum of individual energy contributions as defined in Section 2.3. To this end, we performed ALMO-EDA calculations for $(H_2O)_{n=2-10}$ clusters to determine possible relationships between the construction of a density functional and its ability to model a given water system as a function of its size. All the energy components were benchmarked against the ωB97M-V/auc-cc-pVQZ level of theory, as ωB97M-V has not only been successful to reproduce the liquid water's structure[7] but also provides many-body energies with minimal loss of accuracy compared to the CCSD(T)/CBS.[7,63]

In Figure 4, we show the signed error per monomer, $\Delta e = (E^{\text{model}} - E^{\text{ref}})/n$, of each individual energy contribution to E_{int} for all 38 structures in the BEGDB dataset. In general, all four density functionals examined in this study display similar behavior in describing E_{pol} when compared to the reference values. For small clusters the four functionals predict very similar values for E_{pol}, but tend to deviate from ωB97M-rV at different rates as the cluster size increases. Nonetheless it should be noted that Δe_{pol} is always small compared to the errors associated with the other components of the interaction energy.

The distinct nature of the four functionals is highlighted in their ability to describe the dispersion-free frozen energy term, $E_{\text{DF}} = E_{\text{elec}} + E_{\text{Pauli}}$. Figure 4b shows that none of the functionals examined here are able to completely reproduce the ωB97M-V values for E_{DF}, with B97M-rV providing the closest agreement for all the 38 structures. On the other hand, SCAN systematically underestimates E_{DF} by ~ 0.4 kcal/mol per monomer. In the case of the GGA functionals (BLYP-D3 and revPBE-D3), the ALMO-EDA calculations indicate that the observed over-binding of the water clusters (Figure 1) arises primarily from the systematic over-estimation of E_{DF}. Furthermore, Δe_{DF} appears to depend on the cluster's size, with all functionals providing similar values for the 2Cs structure, but then predicting larger deviations as the cluster size increases. For example, Δe_{DF} has a value of 0.22 kcal/mol (BLYP-D3) and 0.18 kcal/mol (revPBE-D3) for the water dimer, and

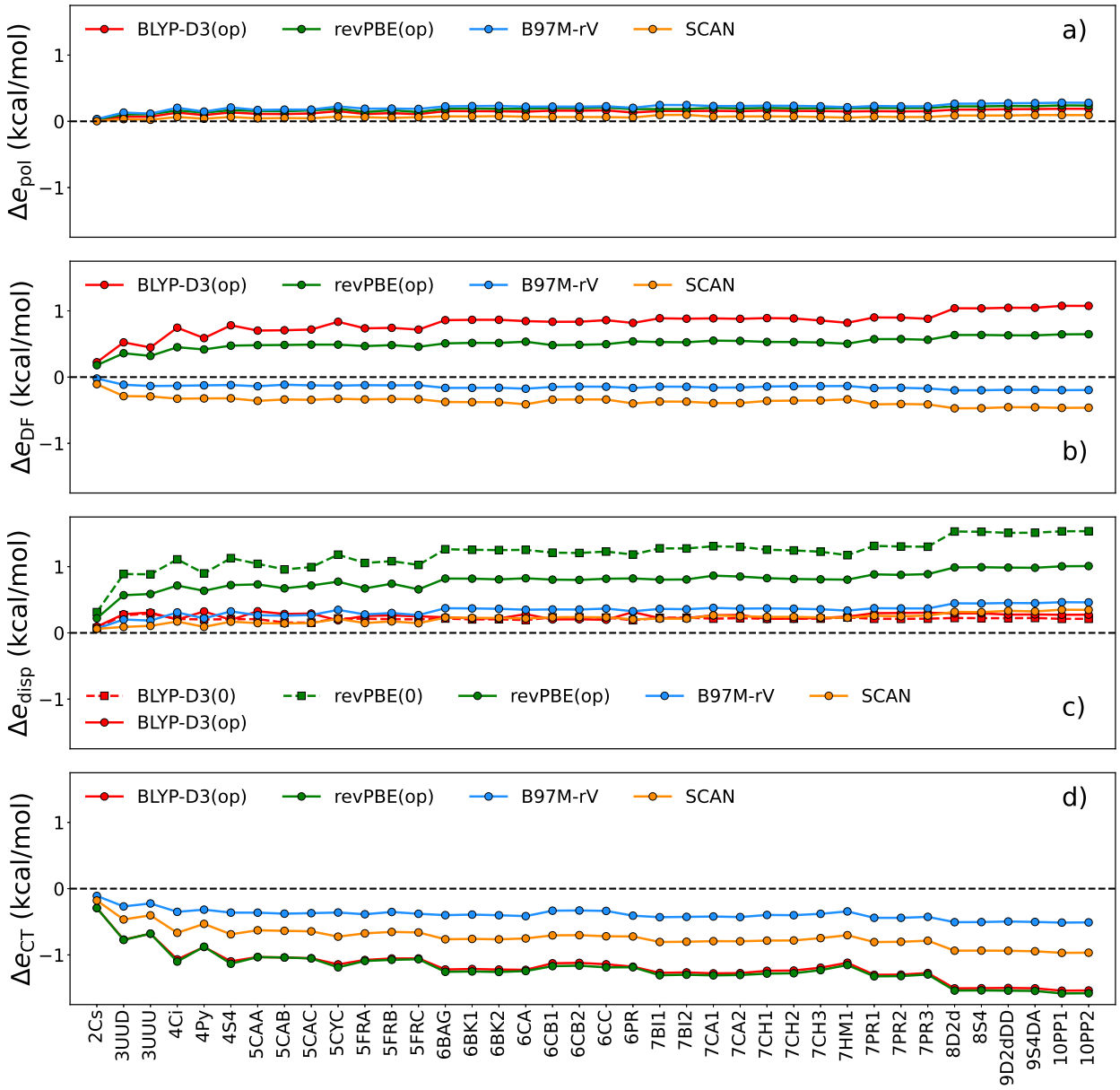

Figure 4: Signed error per monomer associated with each individual contribution to E_{int} as defined by the ALMO-EDA method. a) Δe_{pol}, b) Δe_{DF}, c) Δe_{disp}, and d) Δe_{CT}. All values are reported in kcal/mol for the 38 $(H_2O)_{n=2-10}$ clusters of the BEGDB data set. See main text for details.

reaches 1.07 (BLYP-D3) and 0.71 kcal/mol (revPBE-D3) at the water decamer. In contrast, independently of the system's size, B97M-rV and SCAN accurately predict E_{DF}, as $\langle \Delta e_{\text{DF}}^{\text{B97M-rV}} \rangle$ and $\langle \Delta e_{\text{DF}}^{\text{SCAN}} \rangle$ are 0.15 and 0.36 kcal/mol, respectively.

In the case of revPBE, the ALMO-EDA calculations provide further insights into the importance of an accurate selection of the dispersion correction. From Figure 4, it is clear that the

dispersion correction applied to revPBE-D3 results in large deviations from the reference data, with revPBE-D3(op) performing relatively better than revPBE-D3(0), as already noted in Section 3.1. The ALMO-EDA results demonstrate that the difference in E_{int} between revPBE-D3(op) and revPBE-D3(0) originates from the choice of dispersion parameters adopted by the two density functionals. For example, for the 6BAG structure, the lowest-energy hexamer in the BEGDB dataset, ΔE_{disp} is equal to 7.56 kcal/mol (1.26 kcal/mol per monomer) with revPBE-D3(0), and is reduced to 4.86 kcal/mol (0.81 kcal/mol per monomer) with revPBE-D3(op). The Δe_{disp} of both PBE-D3(op) and PBE-D3(0) are shown for the BEGDB $(H_2O)_6$ structures in the Supplementary Information. It should be noted that the D3(0) correction is the most common form of dispersion correction used in AIMD simulations of water with the revPBE functionals.[104–108] The present analysis indicates that it would be beneficial to move beyond the D3(0) correction in AIMD simulations of water and suggests that some caution should be exercised when selecting an appropriate dispersion correction for the development of data-driven and machine-learned models of water based on DFT simulations.

All density functionals considered in this study underestimate E_{CT} relative to the ωB97M-V reference value, along the following trend: B97M-rV $>$ SCAN $>$ revPBE-D3 $\simeq$ BLYP-D3. In describing the charge-transfer contribution, BLYP-D3 and revPBE-D3 become indistinguishable.

The ALMO-EDA analyses presented here also shine light into the surprising accuracy obtained using the BLYP-D3 functional for the interaction energies of the BEGDB dataset (Figure 1 and 2). In particular, Figure 4 suggests that there is a pronounced error cancellation between E_{DF} and E_{CT} calculated with BLYP-D3. Partial error cancellation also occurs in the case of revPBE-D3, while both ΔE_{DF} and ΔE_{CT} calculated with SCAN have the same sign, significantly contributing to the over-binding observed for the water clusters.[49]

Finally, in order to gain insight into the reliability of DFT predictions for water, we further analyze the two energy contributions the electrostatic energy obtained using orthogonal frozen decomposition (E_{elec}) and the charge-transfer energy (E_{CT}) that display the largest errors. In Figure 5a, we show the mean absolute errors associated with BLYP-D3(op), revPBE-D3(op), B97M-rV, and

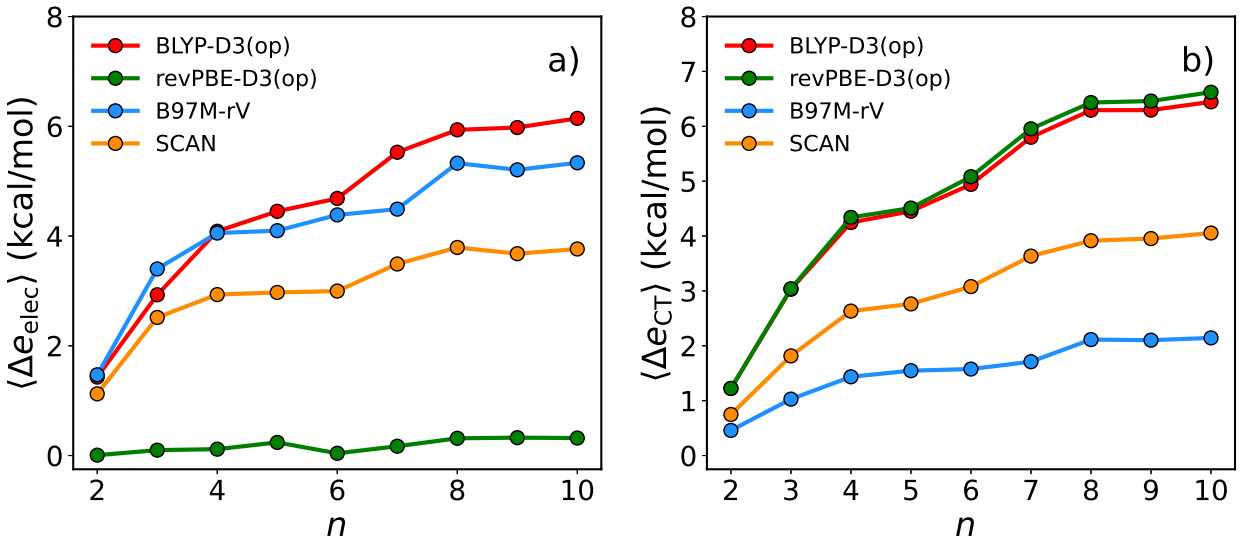

Figure 5: Mean absolute errors per monomer of independent energy contributions as a function of water molecules. In (a) $\rangle\Delta e_{\text{elec}}$ is shown for permanent electrostatics, and (b) $\rangle\Delta e_{\text{CT}}$ is shown charge-transfer energies, for (H_2O)n, with $n = 2-10$, relative to the ωB97M-V reference energies.

SCAN calculations of E_{elec} relative to the ωB97M-V reference values. revPBE-D3(op) consistently provides the closest agreement with ωB97M-V, displaying a $\langle e_{\text{elec}}\rangle$ of $\sim$0.16 kcal/mol that is effectively independent of the cluster's size. This indicates that, while revPBE suffers from large functional-driven and density-driven errors, it predicts electrostatic interactions of water within quantitative accuracy. Since hydrogen-bond forming and breaking in water are largely driven by electrostatic fluctuations in the hydrogen-bond network, these results may explain why revPBE provides a reasonably accurate description of the structure of liquid water at ambient conditions.[106] Furthermore, this analysis provides further support for the observation made in ref 58 that the DC-PBEα-D4 functional (where α is the percentage of HF exchange) outperforms DC-BαLYP-D4, DC-TPSSn, and DC-SCANα in three benchmark datasets where electrostatic interactions are dominant. In the case of $(H_2O)_{10}$, taking the 10PP1 structure as an example, E_{elec} predicted by revPBE-D3(op) lies within $\sim$ 0.32 kcal/mol of the reference value, while the $\langle\Delta e_{\text{elec}}\rangle$ associated with BLYP-D3(op) is larger than 6 kcal/mol. B97M-rV and SCAN predict E_{elec} with an absolute error of 5.33 kcal/mol and 3.76 kcal/mol, respectively. While Figure 5 shows the absolute mean errors associated with E_{elec}, it should be noted that the GGA and meta-GGA functionals

examined in this study differ qualitatively, with BLYP-D3(op) and revPBE-D3(op) systematically underestimating E_{elec}, and B97M-rV and SCAN systematically overestimating E_{elec}.

From Figure 5b, it is clear that the errors in the charge-transfer energy vary with the size of the water cluster for all four functionals. As the density-driven error determines the accuracy of charge-transfer energies, the two GGA functionals (BLYP-D3(op) and revPBE-D3(op)) predict E_{CT} with similar accuracy, while SCAN displays a larger error than B97M-rV as well as a more pronounced increase in error as a function of the cluster size. As a result, while the B97M-rV and SCAN display similar $\langle \Delta e_{\text{CT}} \rangle$ values of 0.46 kcal/mol and 0.75 kcal/mol, respectively, for the water dimer, the difference between the two functionals quadruples for the water decamer.

3.4 Analysis of Many-Body Energies for the Water Hexamer

3.4.1 Interaction Energies within MB-DFT(DC)

Since within DC-DFT the Kohn-Sham equations are not solved self-consistently and require $\rho^{\text{HF}}(\mathbf{r})$, the application of DC-DFT in AIMD simulations is not straightforward. However, with recent advances in the development of data-driven many-body PEFs,[49,63,80] DC-DFT-based molecular dynamics simulations are now possible, as demonstrated in ref. 50. To further rationalize the interplay between functional-driven and density-driven errors in modeling water from the gas to the liquid phase, in this section, we investigate the contributions of n-body energies to the interaction energies calculated using the DFT and DC-DFT functionals as well as the corresponding MB-DFT[7,63] and MB-DFT(DC)[50,63] PEFs. The fully-consistent MB-DFT and MB-DFT(DC) models were developed as described in Section 2.1 and 2.4.2, with dipole polarizabilities computed *ab initio*[111,112] from their corresponding DFT models, and the atomic charges[113] computed from their corresponding hybrid functionals (e.g., the atomic charges of MB-B97M-rV were obtained from ωB97M-V calculations), similar to what has been done in ref 63. The correlation plots to validate the accuracy of the 2B and 3B energies of the MB-DFT and MB-DFT(DC) models are included in the Supporting information (Figure S5). The MB-DFT(DC) PEFs reproduce the DC-DFT reference data for 1B, 2B, and 3B energies, with root mean square deviations (RMSDs) of ~ 0.08,

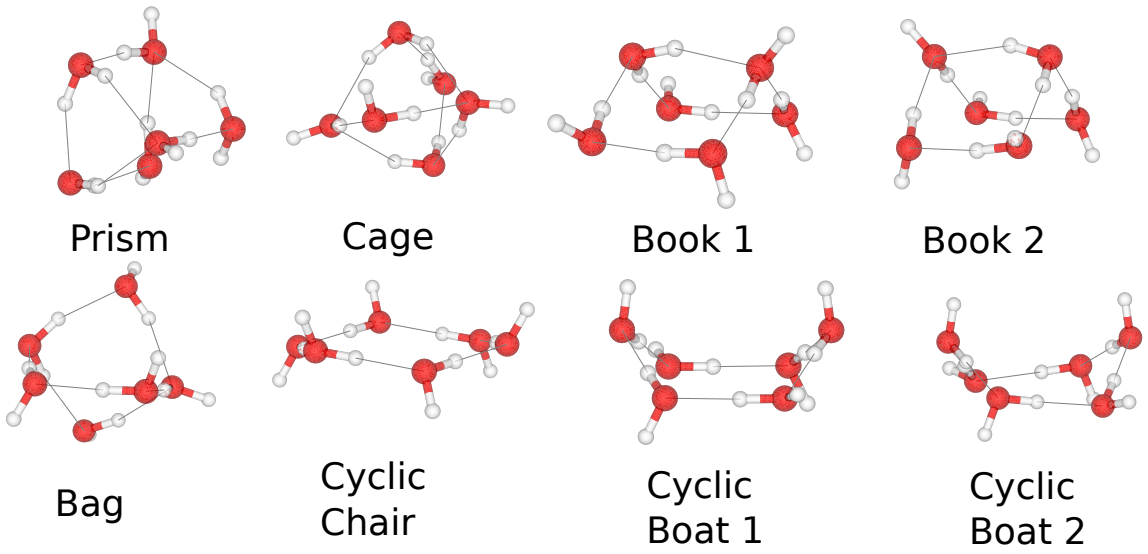

Figure 6: Illustrations of the first eight low-energy isomers of the water hexamer, $(H_2O)_6$, as considered in this study. Structures obtained from ref .[114]

~ 0.13, and ~ 0.03 kcal/mol, respectively. The reference geometries and n-body energies of the isomers of the water hexamer discussed in this section were taken from ref 114.

We begin by assessing the interaction energies of the first eight low-energy isomers of the $(H_2O)_6$ cluster shown in Figure 6. In Figure 7, the interaction energies are shown for both the DFT (panel a) and DC-DFT (panel b) models relative to the CCSD(T)/CBS reference energy of the prism isomer, i.e.,

$$E^{\text{rel}}_{\text{int,isomer}} = E^{\text{model}}_{\text{int,isomer}} - E^{\text{CCSD(T)}}_{\text{int,prism}}. \tag{16}$$

Consistent with our previous analyses, SCAN systematically underestimates the relative interaction energy.[50] For the prism isomer, SCAN displays an error of -5.84 kcal/mol, while BLYP-D3(op) provides the highest accuracy, with an error of -0.20 kcal/mol. The B97M-rV and revPBE-D3(op) functionals display errors of -0.94 kcal/mol and 1.43 kcal/mol, respectively. As shown in Section 3.1, the accuracy of BLYP improves going from the D3(0) to the D3(op) dispersion correction, with the error for the prism isomer decreasing by ~ 0.7 kcal/mol The same change in the dispersion correction results in a significantly larger improvement (~ 2.2 kcal/mol) in the case of the revPBE functional.

When the HF density is used to calculate the interaction energies as shown in Figure 7b, DC-SCAN displays a mean error of -0.3 kcal/mol, as already reported in ref 50. The performance of the other three functionals worsens upon applying the density correction, although the increase in error is higher for DC-BLYP-D3 and DC-revPBE-D3 than DC-B97M-rV, since B97M-rV displays smaller density-driven errors (Sections 3.1 and 3.2).

Figure 7 also shows that the MB-DFT (panel c) and MB-DFT(DC) (panel d) PEFs closely reproduce the interaction energies of their corresponding DFT and DC-DFT functionals. It is, however, noteworthy to mention that for cyclic geometries, the MB-BLYP-D3(0,op), MB-revPBE-D3(0,op), and MB-B97M-rV PEFs display slightly larger errors than their DFT counterparts. Overall, the $\langle \Delta E_{\text{int}} \rangle$ values for MB-BLYP-D3(op), MB-revPBE-D3(op), and MB-B97M-rV are $\sim$1.25, $\sim$3.18, and $\sim$ 0.96 kcal/mol, respectively. Interestingly, MB-SCAN has a mean error of -4.10 kcal/mol, which is $\sim$1 kcal/mol smaller than the mean error associated with SCAN. In Figure 7d, the high accuracy of MB-SCAN(DC) ($\langle \Delta E_{\text{int}}^{\text{MB}-\text{SCAN(DC)}} \rangle \sim 0.52$ kcal/mol) is evident compared to the other MB-DFT(DC) PEFs.

3.4.2 Many-body Decomposition Analysis

Further insight into the nature of molecular interactions in water can be gained through the decomposition of the interaction energies into individual many-body contributions. In Figure 8 the errors ($\Delta\varepsilon_{\text{nB}}$) associated to each term of the many-body expansion (eq 1) of the interaction energy of the prism isomer is plotted for both the DFT and DC-DFT functionals, as well as for their corresponding MB-DFT and MB-DFT(DC) PEFs.

It is clear that the selection of the D3 parameters has important consequences in the case of revPBE-D3, as the error in the 2B energy is $\sim$ 2 kcal/mol smaller for revPBE-D3(op) compared to revPBE-D3(0). Consistent with the analyses discussed in Section 3.1, SCAN displays the largest error at the 2B level, underestimating ε_{2B} by $\sim$ 5.8 kcal/mol. Interestingly, BLYP-D3(0) and BLYP-D3(op) are the only density functionals with 3B errors larger than 2 kcal/mol, which also suggests pronounced error cancellation in the corresponding interaction energy since the 2B and

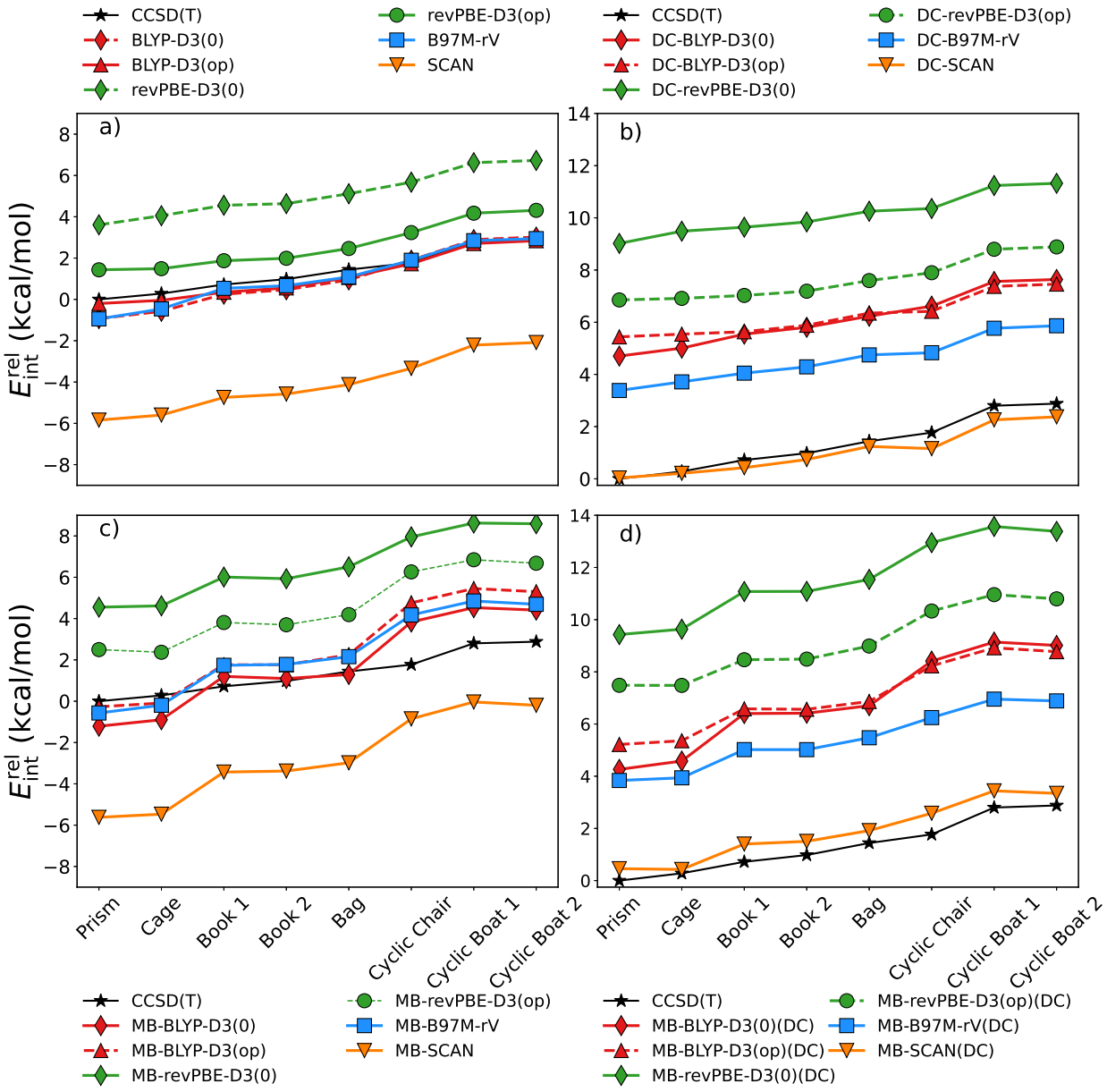

Figure 7: Interaction energies (in kcal/mol) for the first eight low-energy isomers of the $(H_2O)_6$ cluster calculated with the DFT (panel a) and DC-DFT (panel b) models, and corresponding MB-DFT (panel c) and DC-MB-DFT (panel d) PEFs. All interaction energies are reported relative to the CCSD(T)/CBS value of the interaction energy of the prism isomer.

3B energies have opposite signs.

Figure 8b shows the errors associated with the DC-DFT functionals, where only the D3(op) dispersion correction is considered given its improved performance. The effect of using the HF density on $\Delta\varepsilon_{2B}$ is consistent with the analyses reported in the previous sections, indicating that

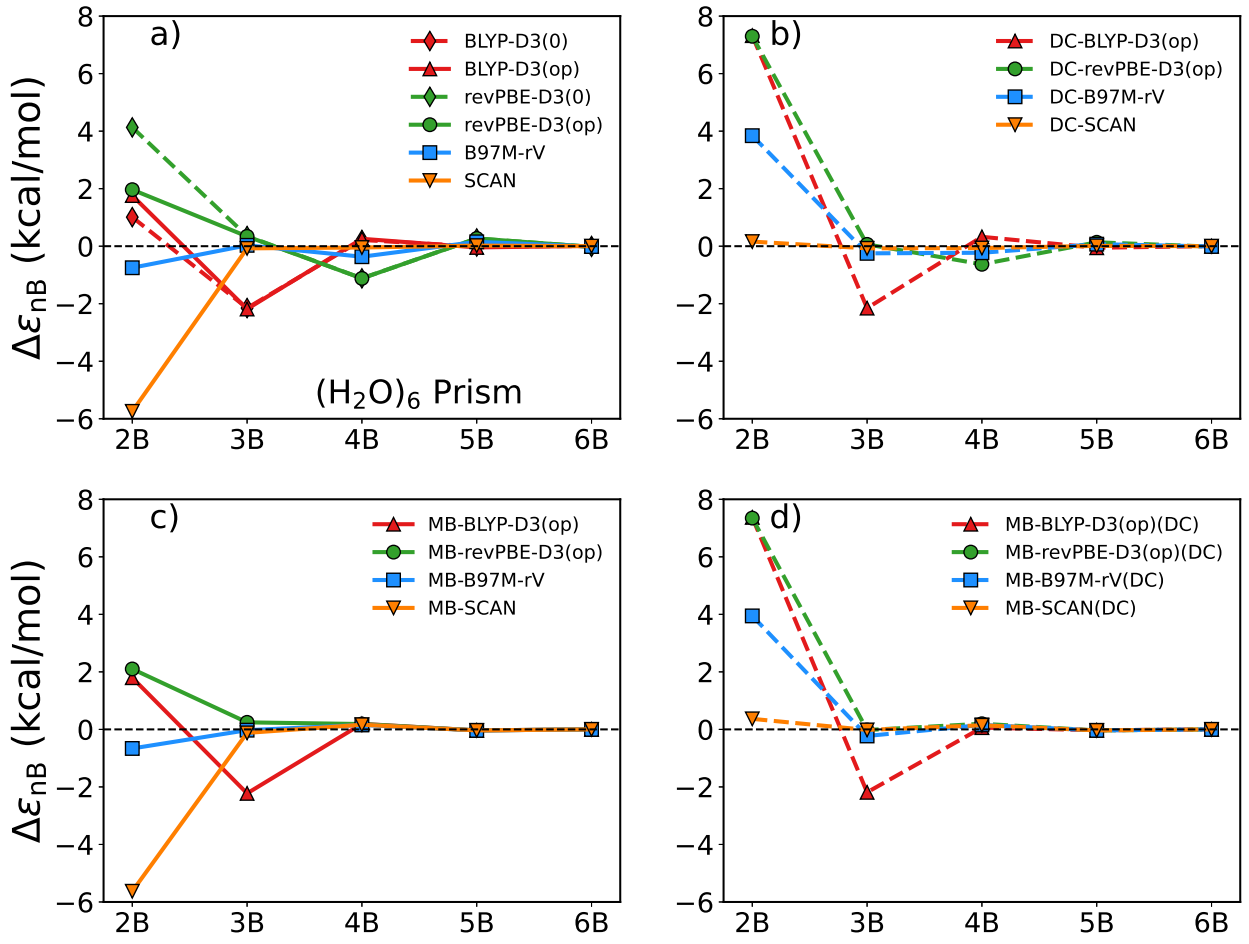

Figure 8: Errors (in kcal/mol) associated with individual many-body contributions to the interaction energy of the prism isomer of the water hexamer relative to the corresponding CCSD(T)/CBS reference values. a) DFT functionals, b) DC-DFT functionals, c) MB-DFT PEFs, and d) MB-DFT(DC) PEFs. See main text for details.

the error is reduced for SCAN but increases for all other functionals. However, the error in the 3B energy is virtually unaffected, except for the BLYP-D3(op) functional and its derivatives, with $\Delta\varepsilon_{2B} \sim 2.2$ kcal/mol for BLYP-D3(op), DC-BLYP-D3(op), MB-BLYP-D3(op) and MB-BLYP-D3(op)(DC). Since ε_{3B} constitutes $\sim$15% of E_{int} of the interaction energy in water, an accurate representation of the 3B energies is critical to a realistic description of the properties of liquid water within the MB-DFT formalism. It is, therefore, expected that MB-BLYP-D3(op) will be, among the MB-DFT PEFs examined here, the least accurate PEF in predicting the properties of liquid water. Overall, the accuracy of the DFT and DC-DFT functionals is preserved by their corresponding MB-DFT and MB-DFT(DC) PEFs.

Furthermore, the mean absolute errors of ε_{nB} with $n = 2,3,4$ were calculated for all DFT and DC-DFT functionals, and MB-DFT and MB-DFT(DC) PEFs investigated in this study. The B97M-rV functional predicts 2B energies within chemical accuracy, with $\langle \Delta\varepsilon_{2B} \rangle$ of $\sim$ 0.35 kcal/mol. As expected, the BLYP-D3(0) and BLYP-D3(op) functionals predict similar values of ε_{2B}, while the difference between revPBE-D3(0) and revPBE-D3(op) is not subtle. Since the 2B energy represents $\sim$80% of the total interaction energy in water, it is expected that the large difference ($\sim$ 3 kcal/mol) between revPBE-D3(0) and revPBE-D3(op) will lead to qualitatively different predictions for the structure of liquid water from MD simulations with the corresponding MB-revPBE-D3(0) and MB-revPBE-D3(op) models and, likely, from analogous AIMD simulations.

In addition, we highlight that the DFT functionals and MB-DFT(DC) PEFs based on BLYP display the largest $\langle \Delta\varepsilon_{3B} \rangle$, while all other functionals and PEFs predict ε_{3B} within chemical accuracy (1 kcal/mol). The 3B energy predicted by the BLYP functional, however, benefits from the density correction which is illustrated by the error reduction displayed by both DC-BLYP-D3(op) and MB-BLYP-D3(op)(DC). The increase in 2B error, but significant improvement in the representation of ε_{3B} when going from BLYP-D3(op) to DC-BLYP-D3(op) indicates that mitigating the density-driven errors in BLYP will improve its ability to describe the properties of liquid water, but only at the 3B level, since the deficiencies of BLYP for water are primarily due to intrinsic functional-driven errors.

Finally, all functionals examined here display minimal improvement in their ability to represent 4B energies upon applying the density correction except the revPBE-D3(0) and revPBE-D3(op) functionals, for which $\langle \Delta\varepsilon_{4B} \rangle$ reduces from $\sim$0.75 kcal/mol to $\sim$ 0.30 kcal/mol. A complete analysis of $\langle \Delta\varepsilon_{nB} \rangle$ with $n = 2,3,4$ is reported in the Supporting Information.

3.5 Liquid water

Functional-driven and density-driven errors in DFT are expected to affect simulations of liquid water carried out ab initio or with DFT-based potentials. In this section, we investigate the structural properties of liquid water predicted by MD simulations carried out with both the MB-DFT

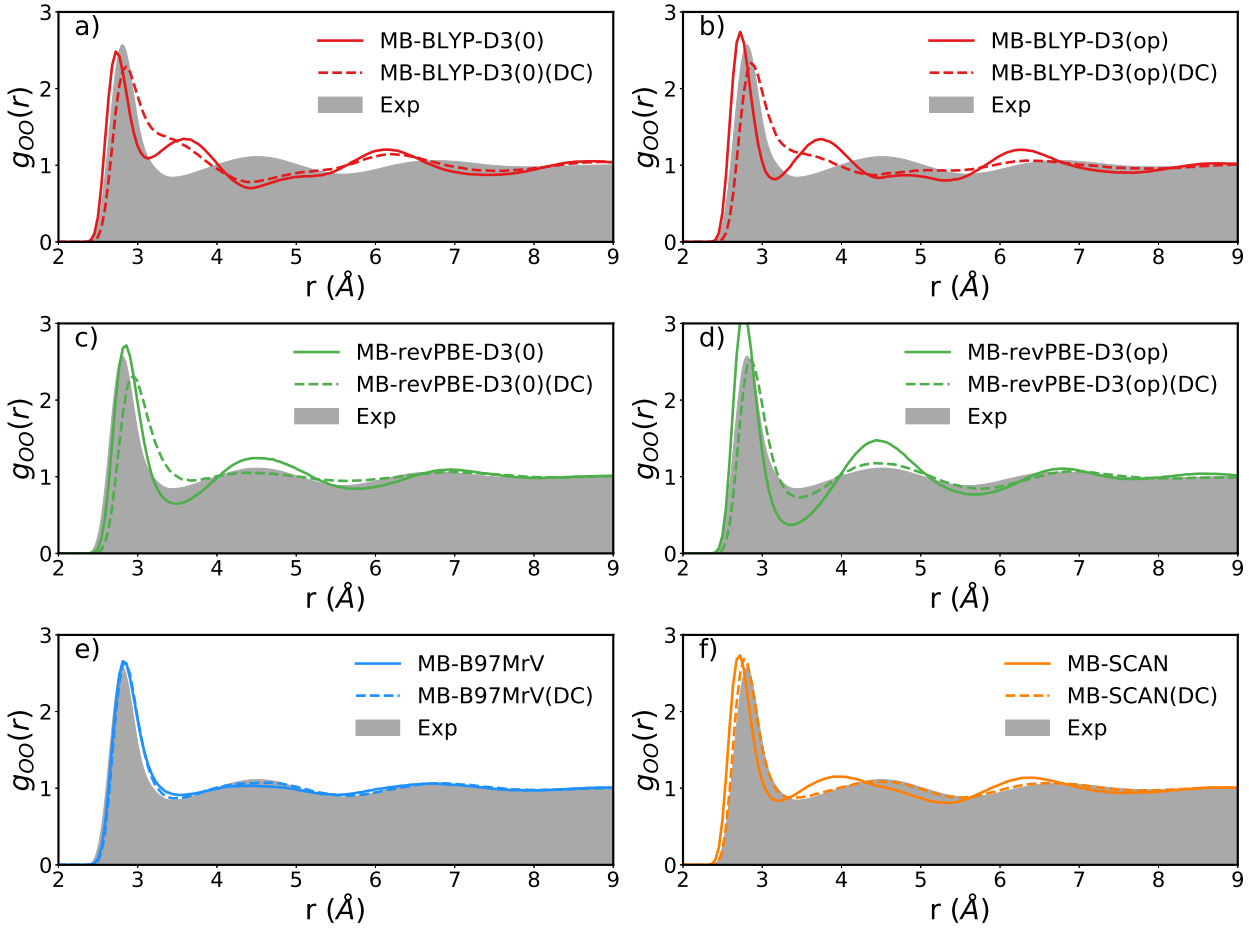

Figure 9: Oxygen-oxygen radial distribution function (g_{OO}) calculated from MD simulations carried out in the NPT ensemble at 298 K and 1 atm with: a) MB-BLYP-D3(0)(DC), b) MB-BLYP-D3(op)(DC), c) MB-revPBE-D3(0)(DC), d) MB-revPBE-D3(op)(DC), e) MB-B97M-rV(DC), and f) MB-SCAN(DC). The experimental g_{OO}) at 295 K is from ref 115.

and MB-DFT(DC) PEFs. The oxygen-oxygen RDFs, $g_{\mathrm{OO}}(r)$, calculated in the NPT ensemble at $T = 298$ K and 1 atm are shown in Figure 9, while the tetrahedral order parameter, q_{tet}, is shown in Figure 10.

Overall, at ambient pressure and temperature, the density correction "disorders" the structure of liquid water, and in the cases of MB-BLYP-D3(0)(DC), MB-BLYP-D3(op)(DC), and MB-revPBE-D3(0)(DC), breaks the solvation structure up to the first solvation shell, as suggested by the flattening of g_{OO} after the first peak, and a more disordered signature in the order parameter distribution where the peak at $\sim q_{\mathrm{tet}} = 0.5$ gains intensity. In this context it should be noted that both MB-BLYP-D3 PEFs at 298 K predict a qualitatively different g_{OO} from that obtained from AIMD

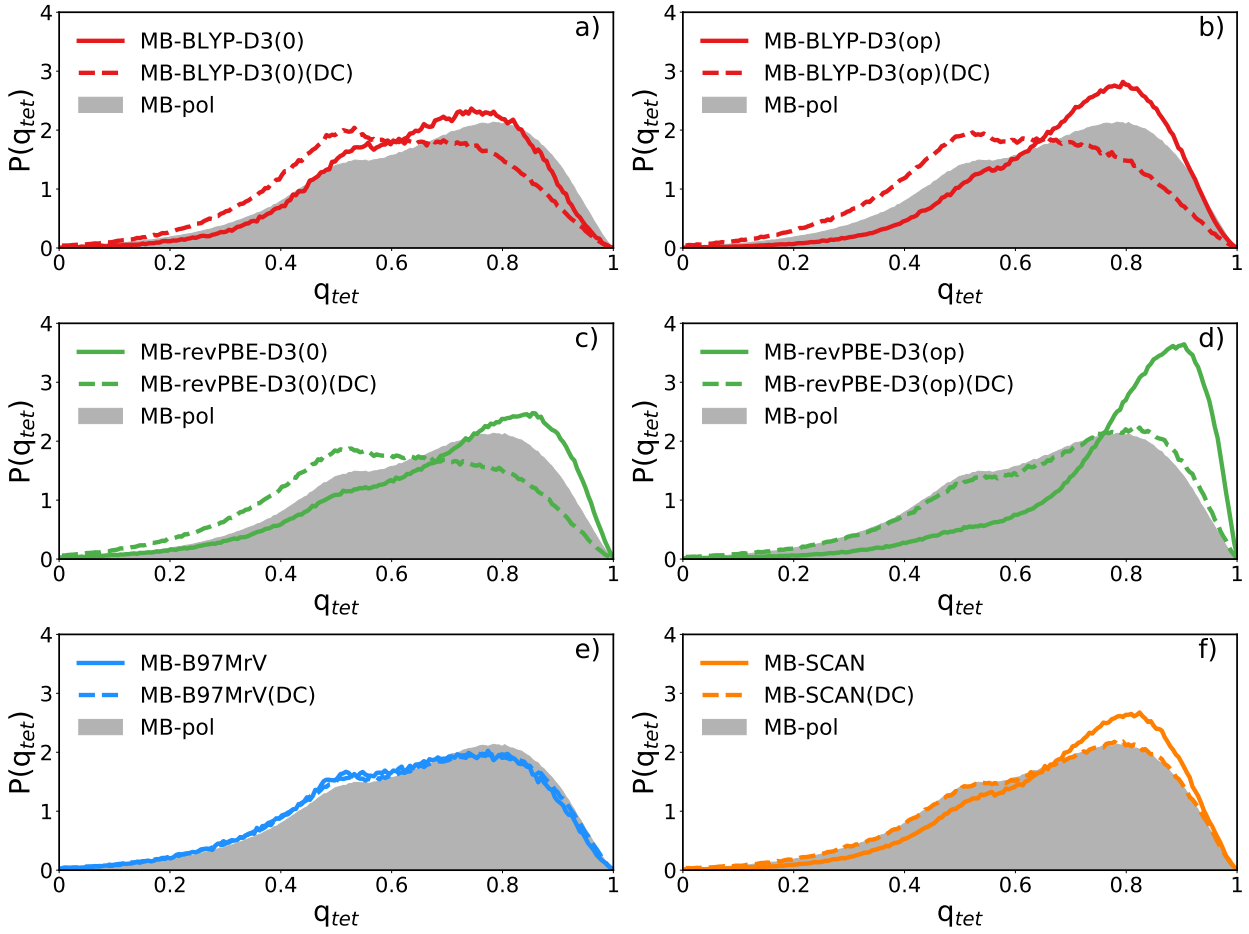

Figure 10: Distributions of the tetrahedral order parameter, q_{tet}, calculated from MD simulations in the NPT ensemble at 298 K and 1 atm with: a) MB-BLYP-D3(0) and MB-BLYP-D3(0)(DC), b) MB-BLYP-D3(op) and MB-BLYP-D3(op)(DC), c) MB-revPBE-D3(0) and MB-revPBE-D3(0)(DC), d) MB-revPBE-D3(op) and MB-revPBE-D3(op)(DC), e) MB-B97M-rV and MB-B97M-rV(DC), and f) MB-SCAN and MB-SCAN(DC).

simulations carried out with BLYP-D3 at 300 K,[116] and at 330 K in the NPT and NVT ensembles.[105,117] Based on the analyses presented in Sections 3.1-3.4, this discrepancy can be traced back to size dependent density-driven errors, whereby the magnitude of individual many-body energies for liquid water described by BLYP-D3 is fundamentally different from the magnitude of the same contributions calculated for gas-phase water clusters to which the corresponding MB-BLYP-D3 PEFs are trained on.

The MB-DFT PEFs based on meta-GGA functionals respond qualitatively differently to the density correction, with MB-B97MrV(DC) and particularly MB-SCAN(DC) accurately reproduc-

ing the experimental oxygen-oxygen RDF. Interestingly, the MB-DFT PEFs based on the GGA functionals show a particular sensitivity to the choice of the dispersion correction, with both MB-BLYP-D3(op) and MB-revPBE-D3(op) predicting a more structured liquid than their MB-BLYP-D3(0) and MB-revPBE-D3(0) counterparts. Evidenced by substantially higher peaks at $\sim q_{tet} = 0.9$, these results suggest that the D3(OP) dispersion correction favors a substantially more tetrahedral structure of liquid water than the D3(0) dispersion correction.

In the context of the density correction, while both MB-BLYP-D3(0)(DC) and MB-BLYP-D3(op)(DC) predict relatively similar g_{OO}, which are appreciably different from the experimental curves, MB-revPBE-D3(0)(DC) predicts a relatively flatter g_{OO} while MB-revPBE-D3(op)(DC) predicts a g_{OO} in good agreement with experiment. This is consistent with the analyses presented in Sections 3.1-3.4, where the D3(0) and D3(op) parameters for revPBE were found to provide significantly different results, and suggests that the choice of parameters for the D3 dispersion correction has substantial impact on the simulations of liquid water.

It should be noted that the MB-revPBE-D3(0) PEF accurately reproduces g_{OO} reported from AIMD simulations (e.g., see ref 106), which is in close agreement with experiment. However, the analyses presented above show that revPBE-D3(0) consistently displays the largest errors in the interaction and many-body energies of water clusters. In particular, the D3(0) correction applied to the revPBE functional leads to incorrect energetics as it does not appropriately correct for the 2B dispersion energy (Section 3.3), which consequently results in MB-revPBE-D3(0)(DC) predicting an under-structured g_{OO}. When the 2B dispersion is suitably corrected, g_{OO} becomes more structured as shown by the MB-revPBE-D3(op) PEF in Figure 9d.

As expected from the analyses discussed in Sections 3.1-3.4, the MB-B97M-rV PEF provides an accurate description of the structure of liquid water at ambient conditions, which is in agreement with the results obtained from AIMD simulations with the B97M-rV functional. Since density-driven errors in B97M-rV are small, the MB-B97M-rV(DC) PEF only provides minimal improvement. In contrast, as discussed in detail in ref 50, the application of the density-correction has a significant impact on the individual many-body energies calculated with SCAN, which results

in the MB-SCAN(DC) PEF clearly outperforming the MB-SCAN counterpart and providing an accurate description of various properties of liquid water.[50]

4 Conclusion

In this study, we presented a systematic investigation of the interplay between functional-driven and density-driven errors in four semi-local density functional approximations applied to the representation of molecular interactions in water, from the dimer to the liquid phase. Interaction and *n*-body energies predicted by the BLYP-D3, revPBE-D3, B97M-rV and SCAN density functional were analyzed within self-consistent DFT as well as density-corrected DFT (DC-DFT) in order to gain insight into the effect of density-driven errors, and identify reliable density functionals for water simulations. In addition, ALMO-EDA calculations were used to shine light into the ability of the different density functionals to reproduce individual physical contributions to the interaction energies.

We found that the magnitude of functional-driven and density-driven errors may sensitively depend on the size of the water cluster in question. In cases where density-driven errors dominate, correcting for density-driven errors improves the overall performance of a given density functional. In contrast, in cases where functional-driven errors dominate, the application of the density correction causes the total error to increase as a function of cluster size.

Among all functionals examined in this study, SCAN, which satisfies all 17 constraints known for meta-GGA functionals, represents a special case since it is strongly affected by density-driven errors. Application of the density-correction results in DC-SCAN being able to consistently predict the interaction energies of all 38 clusters of the BEGDB dataset within chemical accuracy. Notably, the MB-SCAN(DC) PEF derived from DC-SCAN many-body data preserves the same accuracy as DC-SCAN and enables MD simulations of liquid water which are found to be in agreement with experiment. B97M-rV, another modern meta-GGA density functional, displays small functional-driven and density-driven errors, and provides an accurate description of water, from the gas phase

to the liquid phase. To this end, B97M-rV does not significantly benefit from density corrections, and therefore can be generally recommended for AIMD simulations of liquid water.

On the other hand, our analyses indicate that large functional-driven errors in GGA functionals such as BLYP and revPBE systematically hinder them from providing accurate descriptions of both interaction and n-body energies for water clusters. In addition, the choice of the D3 dispersion parameters was found to play a significant role in the performance of the revPBE functional. Our analyses thus indicate that some caution should be exercised when deriving molecular insights from AIMD simulations of water carried our with GGA functionals since the interplay between functional-driven and density-driven errors may lead to erroneous interpretations of the underlying physical mechanisms.

Finally, we demonstrated that our theoretical/computational framework based on the many-body expansion of the interaction energies originally introduced with MB-pol[76–78] provides an efficient platform for the development of data-driven PEFs derived from density-corrected DFT data which, by disentangling functional-driven from density-driven errors, can enable accurate DFT-based simulations of generic molecular systems from the gas to the condensed phase.

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). E.P. acknowledges support from the Alfred P. Sloan Foundation Graduate Fellowship Program.

Data Availability

The data related to the analyses presented in this work are freely available at the following repository: `https://github.com/paesanilab/Data_Repository/tree/main/MB-DFT(DC)`

References

(1) Barker, J. A.; Watts, R. Structure of Water; A Monte Carlo Calculation. *Chem. Phys. Lett.* **1969**, *3*, 144–145.

(2) Rahman, A.; Stillinger, F. H. Molecular Dynamics Study of Liquid Water. *J. Chem. Phys.* **1971**, *55*, 3336–3359.

(3) Gallo, P.; Amann-Winkel, K.; Angell, C. A.; Anisimov, M. A.; Caupin, F.; Chakravarty, C.; Lascaris, E.; Loerting, T.; Panagiotopoulos, A. Z.; Russo, J.; Sellberg, J. A.; Stanley, H. E.; Tanaka, H.; Vega, C.; Xu, L.; Pettersson, L. G. M. Water: A Tale of Two Liquids. *Chem. Rev.* **2016**, *116*, 7463–7500.

(4) Petrenko, V. F.; Whitworth, R. W. *Physics of Ice*; OUP Oxford, 1999.

(5) Luzar, A.; Chandler, D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.* **1996**, *76*, 928.

(6) Fecko, C.; Eaves, J.; Loparo, J.; Tokmakoff, A.; Geissler, P. Ultrafast Hydrogen-Bond Dynamics in the Infrared Spectroscopy of Water. *Science* **2003**, *301*, 1698–1702.

(7) 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.

(8) Lambros, E.; Paesani, F. How Good Are Polarizable and Flexible Models for Water: Insights From a Many-Body Perspective. *J. Chem. Phys.* **2020**, *153*, 060901.

(9) 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.

(10) Prendergast, D.; Galli, G. X-Ray Absorption Spectra of Water From First Principles Calculations. *Phys. Rev. Lett.* **2006**, *96*, 215502.

(11) Morrone, J. A.; Car, R. Nuclear Quantum Effects in Water. *Phys. Rev. Lett.* **2008**, *101*, 017801.

(12) Silvestrelli, P. L.; Bernasconi, M.; Parrinello, M. Ab Initio Infrared Spectrum of Liquid Water. *Chem. Phys. Lett.* **1997**, *277*, 478–482.

(13) Silvestrelli, P. L.; Parrinello, M. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys. Rev. Lett.* **1999**, *82*, 3308–3311.

(14) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. "Ab Initio" Liquid Water. *J. Chem. Phys.* **1993**, *99*, 9080–9089.

(15) Trout, B. L.; Parrinello, M. the Dissociation Mechanism of H_2O in Water Studied by First-Principles Molecular Dynamics. *Chem. Phys. Lett.* **1998**, *288*, 343–347.

(16) Trout, B. L.; Parrinello, M. Analysis of the Dissociation of H_2O in Water Using First-Principles Molecular Dynamics. *J. Phys. Chem. B* **1999**, *103*, 7340–7345.

(17) Kühne, T. D.; Krack, M.; Parrinello, M. Static and Dynamical Properties of Liquid Water From First Principles by a Novel Car-Parrinello-Like Approach. *J. Chem. Theory Comput.* **2009**, *5*, 235–241.

(18) Kuo, I.-F. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. Liquid Water From First Principles: Investigation of Different Sampling Approaches. *J. Phys. Chem. B* **2004**, *108*, 12990–12998.

(19) Boero, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C.; Parrinello, M. Hydrogen Bonding and Dipole Moment of Water at Supercritical Conditions: A First-Principles Molecular Dynamics Study. *Phys. Rev. Lett.* **2000**, *85*, 3245–3248.

(20) Boero, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C.; Parrinello, M. Water at Supercritical Conditions: A First Principles Study. *J. Chem. Phys.* **2001**, *115*, 2219–2227.

(21) Boero, M.; Parrinello, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C. First-Principles Molecular-Dynamics Simulations of a Hydrated Electron in Normal and Supercritical Water. *Phys. Rev. Lett.* **2003**, *90*, 226403.

(22) Hura, G.; Russo, D.; Glaeser, R. M.; Head-Gordon, T.; Krack, M.; Parrinello, M. Water Structure as a Function of Temperature X-Ray Scattering Experiments and Ab Initio Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1981–1991.

(23) Marx, D.; Tuckerman, M. E.; Parrinello, M. Solvated Excess Protons in Water: Quantum Effects on the Hydration Structure. *J. Condens. Matter Phys.* **2000**, *12*, A153–A159.

(24) Sprik, M.; Hutter, J.; Parrinello, M. Ab Initio Molecular Dynamics Simulation of Liquid Water: Comparison of Three Gradient-Corrected Density Functionals. *J. Chem. Phys.* **1996**, *105*, 1142–1152.

(25) Fois, E.; Sprik, M.; Parrinello, M. Properties of Supercritical Water: An Ab Initio Simulation. *Chem. Phys. Lett.* **1994**, *223*, 411–415.

(26) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. the Influence of Temperature and Density Functional Models in Ab Initio Molecular Dynamics Simulation of Liquid Water. *J. Chem. Phys.* **2005**, *122*, 014515.

(27) Soper, A. K. The Quest for the Structure of Water and Aqueous Solutions. *J. Condens. Matter Phys.* **1997**, *9*, 2717–2730.

(28) Izvekov, S.; Voth, G. A. Car–Parrinello Molecular Dynamics Simulation of Liquid Water: New Results. *J. Chem. Phys.* **2002**, *116*, 10372–10376.

(29) Ortega, J.; Lewis, J. P.; Sankey, O. F. First Principles Simulations of Fluid Water: The Radial Distribution Functions. *J. Chem. Phys.* **1997**, *106*, 3696–3702.

(30) Lin, I.-C.; Seitsonen, A. P.; Tavernelli, I.; Rothlisberger, U. Structure and Dynamics of Liquid Water From Ab Initio Molecular Dynamics—Comparison of BLYP PBE and revPBE Density Functionals With and Without Van Der Waals Corrections. *J. Chem. Theory Comput.* **2012**, *8*, 3902–3910.

(31) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab Initio Molecular Dynamics Simulations of Liquid Water Using High Quality Meta-GGA Functionals. *Chem. Sci.* **2017**, *8*, 3554–3565.

(32) Ruiz Pestana, L.; Marsalek, O.; Markland, T. E.; Head-Gordon, T. the Quest for Accurate Liquid Water Properties From First Principles. *J. Phys. Chem. Lett.* **2018**, *9*, 5009–5016, PMID: 30118601.

(33) Chen, M.; Ko, H.-Y.; Remsing, R. C.; Calegari Andrade, M. F.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P.; Wu, X. Ab Initio Theory and Modeling of Water. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 10846–10851.

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

(35) Ceperley, D. M.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45*, 566–569.

(36) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.

(37) Laasonen, K.; Csajka, F.; Parrinello, M. Water Dimer Properties in the Gradient-Corrected Density Functional Theory. *Chem. Phys. Lett.* **1992**, *194*, 172–174.

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

(39) Laasonen, K.; Parrinello, M.; Car, R.; Lee, C.; Vanderbilt, D. Structures of Small Water Clusters Using Gradient-Corrected Density Functional Theory. *Chem. Phys. Lett.* **1993**, *207*, 208–213.

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

(41) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(42) 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.

(43) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(44) LaCount, M. D.; Gygi, F. Ensemble First-Principles Molecular Dynamics Simulations of Water Using the SCAN Meta-GGA Density Functional. *J. Chem. Phys.* **2019**, *151*, 164101.

(45) Li, M.; Chen, L.; Gui, L.; Cao, S.; Liu, D.; Zhao, G.; Ding, M.; Yan, J.; Wang, D. Born–Oppenheimer Molecular Dynamics Simulations on Structures of High-Density and Low-Density Water: A Comparison of the SCAN Meta-GGA and PBE GGA Functionals. *Phys. Chem. Chem. Phys.* **2021**, *23*, 2298–2304.

(46) 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. U.S.A.* **2020**, *117*, 11283–11288.

(47) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. Modeling Liquid Water by Climbing Up Jacob's Ladder in Density Functional Theory Facilitated by Using Deep Neural Network Potentials. *J. Phys. Chem. B* **2021**, *125*, 11444–11456.

(48) Zhang, L.; Wang, H.; Car, R.; E, W. Phase Diagram of a Deep Potential Water Model. *Phys. Rev. Lett.* **2021**, *126*, 236001.

(49) Lambros, E.; Hu, J.; Paesani, F. Assessing the Accuracy of the SCAN Functional for Water Through a Many-Body Analysis of the Adiabatic Connection Formula. *J. Chem. Theory Comput.* **2021**, *17*, 3739–3749.

(50) Dasgupta, S.; Lambros, E.; Perdew, J.; Paesani, F. Elevating Density Functional Theory to Chemical Accuracy for Water Simulations Through a Density-Corrected Many-Body Formalism. *Nat. Commun.* **2021**, *12*, 1–12.

(51) Piaggi, P. M.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. Phase Equilibrium of Water With Hexagonal and Cubic Ice Using the SCAN Functional. *J. Chem. Theory Comput.* **2021**, *17*, 3065–3077.

(52) Kim, M.-C.; Sim, E.; Burke, K. Understanding and Reducing Errors in Density Functional Calculations. *Phys. Rev. Lett.* **2013**, *111*, 073003.

(53) Scuseria, G. E. Comparison of Coupled-cluster Results With a Hybrid of Hartree–Fock and Density Functional Theory. *J. Chem. Phys.* **1992**, *97*, 7528–7530.

(54) Oliphant, N.; Bartlett, R. J. a Systematic Comparison of Molecular Properties Obtained

Using Hartree–Fock a Hybrid Hartree–Fock Density-functional-theory and Coupled-cluster Methods. *J. Chem. Phys.* **1994**, *100*, 6550–6561.

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

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

(57) 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.

(58) Santra, G.; Martin, J. M. What Types of Chemical Problems Benefit From Density-Corrected DFT? A Probe Using an Extensive and Chemically Diverse Test Suite. *J. Chem. Theory Comput.* **2021**, *17*, 1368–1379.

(59) Song, S.; Vuckovic, S.; Sim, E.; Burke, K. Density Sensitivity of Empirical Functionals. *J. Phys. Chem. Lett.* **2021**, *12*, 800–807, PMID: 33411542.

(60) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(61) Binkley, J. S.; Pople, J. A. Möller–Plesset Theory for Atomic Ground State Energies. *Int. J. Quantum Chem.* **1975**, *9*, 229–236.

(62) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(63) Lambros, E.; Dasgupta, S.; Palos, E.; Swee, S.; Hu, J.; Paesani, F. General Many-Body Framework for Data-Driven Potentials With Arbitrary Quantum Mechanical Accuracy: Water as a Case Study. *J. Chem. Theory Comput.* **2021**, *17*, 5635–5650.

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

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

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

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

(68) Xantheas, S. S.; Dunning Jr, T. H. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$ N. *J. Chem. Phys.* **1993**, *99*, 8774–8792.

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

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

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

(72) 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.

(73) 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.

(74) Wang, Y.; Bowman, J. M. Ab Initio Potential and Dipole Moment Surfaces for Water. II. Local-Monomer Calculations of the Infrared Spectra of Water Clusters. *J. Chem. Phys.* **2011**, *134*, 154510.

(75) Babin, V.; Medders, G. R.; Paesani, F. Toward a Universal Water Model: First Principles Simulations From the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.* **2012**, *3*, 3765–3769.

(76) 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.

(77) 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.

(78) 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.

(79) 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.

(80) 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.

(81) 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.

(82) Caruso, A.; Paesani, F. Data-Driven Many-Body Models Enable a Quantitative Description of Chloride Hydration From Clusters to Bulk. *J. Chem. Phys.* **2021**, *155*, 064502.

(83) 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.

(84) 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.

(85) 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.

(86) Horn, P. R.; Mao, Y.; Head-Gordon, M. Defining the Contributions of Permanent Electrostatics Pauli Repulsion and Dispersion in Density Functional Theory Calculations of Intermolecular Interaction Energies. *J. Chem. Phys.* **2016**, *144*, 114107.

(87) Horn, P. R.; Mao, Y.; Head-Gordon, M. Probing Non-Covalent Interactions With a Second Generation Energy Decomposition Analysis Using Absolutely Localized Molecular Orbitals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067–23079.

(88) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; Gan, Z.; Hait, D.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Kussmann, J.; Lange, A. W.; Lao, K. U.; Levine, D. S.; Liu, J.; McKenzie, S. C.; Morrison, A. F.; Nanda, K. D.; Plasser, F.; Rehn, D. R.; Vidal, M. L.; You, Z.-Q.; Zhu, Y.; Alam, B.; Albrecht, B. J.; Aldossary, A.; Alguire, E.; Andersen, J. H.; Athavale, V.; Barton, D.; Begam, K.; Behn, A.; Bellonzi, N.; Bernard, Y. A.; Berquist, E. J.; Burton, H. G. A.; Carreras, A.; Carter-Fenk, K.; Chakraborty, R.; Chien, A. D.; Closser, K. D.; Cofer-Shabica, V.; Dasgupta, S.; de Wergifosse, M.; Deng, J.; Diedenhofen, M.; Do, H.; Ehlert, S.; Fang, P.-T.; Fatehi, S.; Feng, Q.; Friedhoff, T.; Gayvert, J.; Ge, Q.; Gidofalvi, G.; Goldey, M.; Gomes, J.; González-Espinoza, C. E.; Gulania, S.; Gunina, A. O.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A.; Herbst, M. F.; Hernández Vera, M.; Hodecker, M.; Holden, Z. C.; Houck, S.; Huang, X.; Hui, K.; Huynh, B. C.; Ivanov, M.; Jász, Á.; Ji, H.; Jiang, H.; Kaduk, B.; Kähler, S.; Khistyaev, K.; Kim, J.; Kis, G.; Klunzinger, P.; Koczor-Benda, Z.; Koh, J. H.; Kosenkov, D.; Koulias, L.; Kowalczyk, T.; Krauter, C. M.; Kue, K.; Kunitsa, A.; Kus, T.; Ladjánszki, I.; Landau, A.; Lawler, K. V.; Lefrancois, D.; Lehtola, S.; Li, R. R.; Li, Y.-P.; Liang, J.;

Liebenthal, M.; Lin, H.-H.; Lin, Y.-S.; Liu, F.; Liu, K.-Y.; Loipersberger, M.; Luenser, A.; Manjanath, A.; Manohar, P.; Mansoor, E.; Manzer, S. F.; Mao, S.-P.; Marenich, A. V.; Markovich, T.; Mason, S.; Maurer, S. A.; McLaughlin, P. F.; Menger, M. F. S. J.; Mewes, J.-M.; Mewes, S. A.; Morgante, P.; Mullinax, J. W.; Oosterbaan, K. J.; Paran, G.; Paul, A. C.; Paul, S. K.; Pavošević, F.; Pei, Z.; Prager, S.; Proynov, E. I.; Rák, Á.; Ramos-Cordoba, E.; Rana, B.; Rask, A. E.; Rettig, A.; Richard, R. M.; Rob, F.; Rossomme, E.; Scheele, T.; Scheurer, M.; Schneider, M.; Sergueev, N.; Sharada, S. M.; Skomorowski, W.; Small, D. W.; Stein, C. J.; Su, Y.-C.; Sundstrom, E. J.; Tao, Z.; Thirman, J.; Tornai, G. J.; Tsuchimochi, T.; Tubman, N. M.; Veccham, S. P.; Vydrov, O.; Wenzel, J.; Witte, J.; Yamada, A.; Yao, K.; Yeganeh, S.; Yost, S. R.; Zech, A.; Zhang, I. Y.; Zhang, X.; Zhang, Y.; Zuev, D.; Aspuru-Guzik, A.; Bell, A. T.; Besley, N. A.; Bravaya, K. B.; Brooks, B. R.; Casanova, D.; Chai, J.-D.; Coriani, S.; Cramer, C. J.; Cserey, G.; DePrince, A. E.; DiStasio, R. A.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Goddard, W. A.; Hammes-Schiffer, S.; Head-Gordon, T.; Hehre, W. J.; Hsu, C.-P.; Jagau, T.-C.; Jung, Y.; Klamt, A.; Kong, J.; Lambrecht, D. S.; Liang, W.; Mayhall, N. J.; McCurdy, C. W.; Neaton, J. B.; Ochsenfeld, C.; Parkhill, J. A.; Peverati, R.; Rassolov, V. A.; Shao, Y.; Slipchenko, L. V.; Stauch, T.; Steele, R. P.; Subotnik, J. E.; Thom, A. J. W.; Tkatchenko, A.; Truhlar, D. G.; Van Voorhis, T.; Wesolowski, T. A.; Whaley, K. B.; Woodcock, H. L.; Zimmerman, P. M.; Faraji, S.; Gill, P. M. W.; Head-Gordon, M.; Herbert, J. M.; Krylov, A. I. Software for the Frontiers of Quantum Chemistry: An Overview of Developments in the Q-Chem 5 Package. *J. Chem. Phys.* **2021**, *155*, 084801.

(89) 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.

(90) 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.

(91) Becke, A. D. Density-Functional Exchange-Energy Approximation With Correct Asymp-

totic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(92) 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.

(93) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **1998**, *80*, 890–890.

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

(95) Witte, J.; Mardirossian, N.; Neaton, J. B.; Head-Gordon, M. Assessing DFT-D3 Damping Functions Across Widely Used Density Functionals: Can We Do Better? *J. Chem. Theory Comput.* **2017**, *13*, 2043–2052.

(96) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černỳ, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J., et al. Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems (www.begdb.com): A Users Manual and Examples. *Collect. Czechoslov. Chem. Commun.* **2008**, *73*, 1261–1270.

(97) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters With Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

(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) Bull-Vulpe, E.; Riera, M.; Goetz, A.; Paesani, F. MB-Fit: Software Infrastructure for Data-Driven Many-Body Potential Energy Functions. *J. Chem. Phys.* **2021**, *155*, 124801.

(100) MBX: A Many-Body Energy and Force Calculator. http://paesanigroup.ucsd.edu/software/mbx.html.

(101) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-Isobaric Molecular Dynamics Using Stochastic Velocity Rescaling. *J. Chem. Phys.* **2009**, *130*, 074101.

(102) Errington, J. R.; Debenedetti, P. G. Relationship Between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, *409*, 318–321.

(103) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115*, 2315–2372.

(104) Lin, I.-C.; Seitsonen, A. P.; Tavernelli, I.; Rothlisberger, U. Structure and Dynamics of Liquid Water From Ab Initio Molecular Dynamics—Comparison of BLYP PBE and revPBE Density Functionals With and Without Van Der Waals Corrections. *J. Chem. Theory Comput.* **2012**, *8*, 3902–3910.

(105) Bankura, A.; Karmakar, A.; Carnevale, V.; Chandra, A.; Klein, M. L. Structure Dynamics and Spectral Diffusion of Water From First-Principles Molecular Dynamics. *J. Phys. Chem. C* **2014**, *118*, 29401–29411.

(106) Galib, M.; Duignan, T. T.; Misteli, Y.; Baer, M. D.; Schenter, G. K.; Hutter, J.; Mundy, C. J. Mass Density Fluctuations in Quantum and Classical Descriptions of Liquid Water. *J. Chem. Phys.* **2017**, *146*, 244501.

(107) Ohto, T.; Dodia, M.; Imoto, S.; Nagata, Y. Structure and Dynamics of Water at the Water–Air Interface Using First-Principles Molecular Dynamics Simulations Within Generalized Gradient Approximation. *J. Chem. Theory Comput.* **2019**, *15*, 595–602.

(108) Dodia, M.; Ohto, T.; Imoto, S.; Nagata, Y. Structure and Dynamics of Water at the Wa-

ter–Air Interface Using First-Principles Molecular Dynamics Simulations. II. NonLocal vs Empirical Van Der Waals Corrections. *J. Chem. Theory Comput.* **2019**, *15*, 3836–3843.

(109) Becke, A. D. Density-functional Thermochemistry. I. The Effect of the Exchange-only Gradient Correction. *J. Chem. Phys.* **1992**, *96*, 2155–2160.

(110) Wang, J.; Román-Pérez, G.; Soler, J. M.; Artacho, E.; Fernández-Serra, M.-V. Density Structure and Dynamics of Water: The Effect of Van Der Waals Interactions. *J. Chem. Phys.* **2011**, *134*, 024516.

(111) Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction: High-Order Dispersion Coefficients. *J. Chem. Phys.* **2006**, *124*, 014104.

(112) Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction: High-Order Dispersion Coefficients. *J. Chem. Phys.* **2006**, *124*, 014104.

(113) 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. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(114) 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.

(115) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark Oxygen-Oxygen Pair-Distribution Function of Ambient Water From X-Ray Diffraction Measurements With a Wide Q-Range. *J. Chem. Phys.* **2013**, *138*, 074506.

(116) Ma, Z.; Zhang, Y.; Tuckerman, M. E. Ab Initio Molecular Dynamics Study of Water at

Constant Pressure Using Converged Basis Sets and Empirical Dispersion Corrections. *J. Chem. Phys.* **2012**, *137*, 044506.

(117) Schmidt, J.; VandeVondele, J.; Kuo, I.-F. W.; Sebastiani, D.; Siepmann, J. I.; Hutter, J.; Mundy, C. J. Isobaric-Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions. *J. Phys. Chem. B* **2009**, *113*, 11959–11964.

TOC graphic