How Good is the Density-Corrected SCAN Functional for Neutral and Ionic Aqueous Systems, and What is so Right about the Hartree-Fock Density?

Saswata Dasgupta,[*,†] Chandra Shahi,[‡] Pradeep Bhetwal,[‡] John P. Perdew,[*,‡,¶] and Francesco Paesani[*,†,§,‖]

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

‡*Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, United States*

¶*Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, 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*

E-mail: s1dasgupta@ucsd.edu; perdew@temple.edu; fpaesani@ucsd.edu

Abstract

Density functional theory (DFT) is the most widely used electronic structure method, due to its simplicity and cost effectiveness. The accuracy of a DFT calculation depends not only on the choice of the density functional approximation (DFA) adopted but also on the electron density produced by the DFA. SCAN is a modern functional that satisfies all known constraints for meta-GGA functionals. The density-driven errors, defined as energy errors arising from errors of the self-consistent DFA electron density, can hinder SCAN from achieving chemical accuracy in some systems, including water. Density-corrected DFT (DC-DFT) can alleviate this shortcoming by adopting a more accurate electron density which, in most applications, is the electron density obtained at the Hartree-Fock level of theory due to its relatively low computational cost. In this work, we present extensive calculations aimed at determining the accuracy of the DC-SCAN functional for various aqueous systems. DC-SCAN (SCAN@HF) shows remarkable consistency in reproducing reference data obtained at the coupled cluster level of theory, with minimal loss of accuracy. Density-driven errors in the description of ionic aqueous clusters are thoroughly investigated. By comparison with the orbital-optimized CCSD density in the water dimer, we find that the self-consistent SCAN density transfers a spurious fraction of an electron across the hydrogen bond to the hydrogen atom ($H^\star$, covalently bound to the donor oxygen atom) from the acceptor (O_A) and donor (O_D) oxygen atoms, while HF makes a much smaller spurious transfer in the opposite direction, consistent with DC-SCAN (SCAN@HF) reduction of SCAN over-binding due to delocalization error. While LDA seems to be the conventional extreme of density delocalization error, and HF the conventional extreme of (usually much smaller) density localization error, these two densities do not quite yield the conventional range of density-driven error in energy differences. Finally, comparisons of the DC-SCAN results with those obtained with the Fermi-Löwdin orbital self-interaction correction (FLOSIC) method show that DC-SCAN represents a more accurate approach to reducing density-driven errors in SCAN calculations of ionic aqueous clusters.

1 INTRODUCTION

Water and solvated ions are not only widely studied in computational chemistry, but also in environmental sciences, electrochemistry, and biochemistry due to their importance in various chemical and biomolecular reactions.[1] The presence of ions affects the conformations and functions of proteins and nucleic acids.[2–4] The activity of enzymes and drugs often depends upon the concentration of ions,[5,6] while the presence of ions determines the electrostatic potentials, conductances, and permeabilities of cell membranes.[7,8] Ions affect both reaction rates and mechanisms, and are widely used in separation processes.[9–11] Autoionization of water leads to the generation of hydronium H_3O^+ and hydroxide OH^- ions which are the main pillars of acid-base chemistry in solution.[12]

Understanding the structural and dynamical properties of ion–water interactions requires careful theoretical and experimental investigation. Among all the available methods, Kohn-Sham density functional theory (DFT)[13,14] is widely used due to its trade-off between accuracy and computational cost. DFT plays an important role in *ab initio* studies of ion hydration, water autoionization, proton-transfer reactions, and various physical and chemical properties of hydrated ions.[15–17] In this context, the study of ion–water clusters provides fundamental insights into the multidimensional potential energy surface associated with ions in aqueous solutions.[18–22]

It is well known that the accuracy of any DFT calculation depends upon the choice of the "exchange-correlation" (XC) functional which transforms the many-body electronic structure problem into a single-particle problem.[23] Historically, DFT applications to aqueous systems have faced several roadblocks. The simplest XC functional, the local density approximation (LDA), was found to overestimate the strength of hydrogen bonds which resulted in predictions of an overstructured liquid phase.[24,25] Although having some initial success,[24,26–28] the next-generation functionals developed within the generalized gradient approximation (GGA)[29–31] failed to make a long-lasting impression, as serendipitous error cancellation often provided "the right answers for the wrong reasons".[15,32–35] Meta-GGA

functionals,[36,37] which belong to the third rung of Jacob's ladder of density functional approximations (DFAs), provide improved results compared to their GGA predecessors due to inclusion of a kinetic energy density term in the XC functional. Among the meta-GGA functionals, the strongly constrained and appropriately normed (SCAN) functional has gained particular attention since it satisfies all 17 exact constraints known for meta-GGA functionals.[38] SCAN has been found to outperform most of the other functionals when it is applied to aqueous systems.[39–41] However, like most other functionals, SCAN has been found to be sensitive to density-driven errors,[42–46] which artificially overstabilize hydrogen-bonded systems.[47,48] The inclusion of a fraction of the Hartree-Fock (HF) exchange energy (making a hybrid functional) partially removes density-driven errors in SCAN and often provides more accurate results, although the improvement is not always consistent[48] since the fraction of exact exchange is determined by fitting to bonded systems and not by the exact constraints that a meta-GGA can satisfy.

Density-driven errors are particularly prominent in anion–water interactions as LDA, GGA, and meta-GGA functionals are unable to correctly describe the excess electrons due to the excessive delocalization of the electron density.[49] The Perdew–Zunger self-interaction correction (PZ-SIC) addresses the self-interaction error (SIE) through orbital-by-orbital removal of the SIE for semi-local XC functionals.[42] However, PZ-SIC applied to the delocalized Kohn-Sham orbitals of a molecule is not size-extensive, and consequently requires a transformation to localized orbitals.[42] The Fermi-Löwdin orbitals as implemented in the FLOSIC method[50] are localized and unitarily equivalent to the Kohn-Sham orbitals, but can be constructed without the full panoply of a general unitary transformation. Wagle et al. recently reported promising results for ionic aqueous clusters which were obtained by applying the FLOSIC method in calculations with the PBE and SCAN functionals.[49] Although the FLOSIC method is exact for a one-electron system, its accuracy decreases for many-electron systems, which thus require proper scaling.

Instead of improving the FLOSIC method, in this study we follow an alternative route by

using density-corrected DFT (DC-DFT) to tackle density-driven errors in aqueous systems. DC-DFT[51–60] replaces the Kohn-Sham density with a more accurate density in a non-self-consistent fashion. In practice, the most common flavor of DC-DFT is HF-DFT where a DFA is evaluated on the HF electron density. Since, by construction, the HF density is free from self-interaction errors, its application in DC-DFT calculations mitigates density-driven errors in GGA and meta-GGA functionals. Herein, DC-DFT is employed in the form of HF-DFT, and we preserve the DC-DFT notation, since the HF density is a good proxy for the exact density in a water cluster. DC-SCAN has recently been found to improve the accuracy of DFT calculations for water,[61] effectively elevating the accuracy of the SCAN functional to that of coupled cluster theory, the "gold standard" for chemical accuracy.[62] DC-SCAN does not achieve this level of accuracy for all systems, but it is overall somewhat more accurate than self-consistent SCAN for a large and diverse data set of main-group molecular properties.[63]

2 THEORY AND COMPUTATIONAL DETAILS

In ground-state Kohn-Sham DFT,[23] the energy is self-consistently minimized as:

$$E = \min_{n} \left\{ F[n] + \int d^3r \, n(\mathbf{r}) v(\mathbf{r}) \right\} \tag{1}$$

where $n(\mathbf{r})$ is ground-state density, $v(\mathbf{r})$ is the external potential, and $F[n]$ is the internal part of energy functional. In the Kohn-Sham approach, orbitals are introduced to treat most of $F[n]$ exactly, leaving only the XC energy to be approximated. Since the exact XC functional is unknown, different DFAs have been developed to solve eq 1. The error associated with any DFA is given by the sum of the functional-driven error, ΔE_{F}, and the density-driven error, ΔE_{D}:[55]

$$\Delta E = \Delta E_{\mathrm{F}} + \Delta E_{\mathrm{D}} \tag{2}$$

The functional-driven error arises from the difference between the approximate XC functional, $F[n]$, and the (unknown) exact functional, with both evaluated on the exact density, while the density-driven error arises from using an approximate density $n(\mathbf{r})$ that solves eq 1 for the approximated $F[n]$. Since most functionals produce reasonably accurate electron densities, the functional error is, in many cases, the primary contributor to the total error.[52,53,55,64] In this context, it should be noted that non-empirical, constrained XC functionals tend to produce more accurate electron densities for neutral and cationic atoms than empirically parameterized XC functionals.[65] Independently of the specific parameterization, all (approximate) XC functionals, by construction, still deviate from the piecewise-linear behavior[66,67] of the exact functional for fractional charges, causing excess charge delocalization and resulting in incorrect densities. For certain systems, the density-driven error thus becomes the dominant contributor to the total error.[55,68]

The density-driven error can be understood by considering that the fully-nonlocal classical electrostatic repulsion term represented by the integral in eq 1 contains a self-interaction contribution due to each electron interacting with itself.[69] While this self-interaction contribution should, in principle, be compensated by the XC energy, approximate XC functionals contain substantial local components that prevent them from quantitatively removing electron self-interactions. As a result, the electron density thus tends to overdelocalize in order to minimize the approximated functional. This prevents approximate XC functionals from perfectly cancelling all one- and especially many-electron self-interactions,[45,70,71] leading to spurious fractional charges on separated atoms that underestimate the energy predicted by the piecewise-linear total-energy variation of the exact functional.[66,72] Semilocal approximations to the XC energy only need to be accurate for physical densities, but their XC potentials are functional derivatives that need to be accurate for "arbitrary" (and sometimes unphysical) infinitesimal density variations. It follows that a fully nonlocal dependence on electron density is more needed in the XC potential (which determines the electron density) than in the XC energy.

Since density-driven errors are expected from all approximate XC functionals and can vary in magnitude depending on the type of system, using a more accurate density can mitigate errors due to the overdelocalization of the self-consistent electron density.[52,56,73] However, obtaining an accurate density from wavefunction theories, such as Møller-Plesset perturbation theory and coupled cluster theory, is computationally more expensive than the corresponding DFT calculations. An approximate, yet efficient, approach to reducing density-driven errors in DFT calculations consists in using the spin-restricted Hartree-Fock density, $n^{\mathrm{HF}}(\mathbf{r})$, because, by construction, it does not suffer from either electron overdelocalization or self-interaction errors.[52,55–57,59,60] The resulting DC-DFT energy can then be written as:

$$E^{\mathrm{DC-DFT}} \approx E^{\mathrm{HF}} + \left(E_{\mathrm{XC}}^{\mathrm{approx}}\left[n^{\mathrm{HF}}\right] - E_{\mathrm{X}}^{\mathrm{HF}}\right) \tag{3}$$

An alternative approach developed for dealing with density-driven errors is provided by the Perdew-Zunger self-interaction correction (PZ-SIC) which is based on the orbital-by-orbital removal of the SIE according to:[42]

$$E^{\mathrm{PZ-SIC}} = E_{\mathrm{xc}}^{\mathrm{approx}}\left[n_{\uparrow}, n_{\downarrow}\right] - \sum_{i\sigma}\left(E_{\mathrm{XC}}^{\mathrm{approx}}\left[n_{i\sigma}, 0\right] + U\left[n_{i\sigma}\right]\right) \tag{4}$$

Here, $n_{i\sigma}$ is a single orbital density and $U\left[n_{i\sigma}\right]$ is the Hartree energy of that density. However, the PZ-SIC method is not size-extensive unless localized orbitals are used. The FLOSIC method starts from (and symmetrically orthogonalizes) Fermi orbitals,[49,50]

$$F_{i\sigma}(\mathbf{r}) = \frac{\sum_{j}^{N_\sigma} \psi_{j\sigma}^{\star}\left(\mathbf{a}_{i\sigma}\right)\psi_{j\sigma}(\mathbf{r})}{\sqrt{\sum_{j}^{N_\sigma}\left|\psi_{j\sigma}\left(\mathbf{a}_{i\sigma}\right)\right|^2}} \tag{5}$$

Here N_σ is the number of electrons with spin σ. The Fermi-orbitals, $F_{i\sigma}(\mathbf{r})$, are labelled by the position vectors $\mathbf{a}_{i\sigma}$ of the Fermi orbital descriptors, and are constructed from the non-interacting one-particle density matrix of orthonormal orbitals $\psi_{i\sigma}(\mathbf{r})$ spanning the occupied space.

Unless otherwise stated, all DFT calculations were performed with the aug-cc-pVQZ basis set[74,75] using Q-Chem 5.[76] The highly dense Euler-Maclaurin-Lebedev (99,590) grid[77,78] was used in all DFT calculations to minimize possible grid errors, since it has been shown that the accuracy of some recent meta-GGA and hybrid functionals is particularly sensitive to the choice of the integration grid.[79] The grid sensitivity of SCAN is strongly reduced without loss of overall accuracy in r^2SCAN.[80] Single-point energy calculations using explicitly correlated coupled cluster theory, with single, double, and perturbative triple excitations, i.e., CCSD(T)-F12b,[81] were performed in the complete basis set (CBS) limit by extrapolating[82,83] the energy values obtained with the cc-pVTZ-F12 and cc-pVQZ-F12 basis sets along with associated auxiliary and complementary auxiliary (CABS) basis sets[84,85] using the ORCA quantum chemistry package.[86]

In the analysis of the energetics of the aqueous clusters presented in Section 3, the binding energies are defined as

$$E_{\text{binding}} = E^{\text{cluster}} - nE^{\text{monomer}}_{\text{opt}} \tag{6}$$

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

$$E_{\text{interaction}} = E^{\text{cluster}} - \sum_i E^{\text{monomer}}_i \tag{7}$$

where E^{cluster} is the same total energy as in eq 6 and E^{monomer}_i is the energy of the ith monomer in the same distorted geometry as in the cluster. All MP2 optimized geometries used in the analyses of the energetics of the water clusters are taken from ref 87.

3 RESULTS AND DISCUSSION

3.1 Dependence of the binding energies on basis set and integration grid

Figure 1 shows the basis-set errors of the binding energies for various neutral, protonated, and deprotonated water clusters of the WATER27 dataset[88] calculated as a function of the basis set size. The neutral water clusters consist of water dimer to hexamer structures, whereas the protonated and deprotonated water clusters span from $H_3O^+(H_2O)$ to $H_3O^+(H_2O)_6$, and from $OH^-(H_2O)$ to $OH^-(H_2O)_6$, respectively. Figure 1 shows the errors in binding energies per water molecule ($\Delta\tilde{E}_{\text{bind}}$), where the tilde denotes "per monomer", calculated with both the SCAN and DC-SCAN functionals using the aug-cc-pVDZ (aDZ), aug-cc-pVTZ (aTZ), aug-cc-pVQZ (aQZ) and aug-cc-pV5Z (a5Z) basis sets relative to reference values obtained at the CCSD(T)/CBS level of theory.[89] To understand the convergence of the binding energies with respect to the basis set, ($\Delta\tilde{E}^{\text{basis}}_{\text{conv}}$) has been computed for SCAN and DC-SCAN, which defines the error in binding energies for a given basis set relative to the aug-cc-pV5Z basis:

$$\Delta\tilde{E}^{\text{basis}}_{\text{conv}} = \tilde{E}^{\text{basis}}_{\text{bind}} - \tilde{E}^{\text{aug-cc-pV5Z}}_{\text{bind}} \tag{8}$$

Here, $\tilde{E}^{\text{basis}}_{\text{bind}}$ is the binding energy per monomer of the cluster calculated using a given basis set (aDZ, aTZ or aQZ), and $\tilde{E}^{\text{aug-cc-pV5Z}}_{\text{bind}}$ is the binding energy per monomer of the cluster calculated using the a5Z basis set.

Unsurprisingly, the aDZ basis set is the worst performer in both SCAN and DC-SCAN calculations carried out for the neutral water clusters. In this case, the binding energies converge quickly going from the aDZ to the aTZ and aQZ basis sets. The average $|\Delta\tilde{E}_{\text{bind}}|$ associated with aDZ calculations for neutral water clusters is 0.97 kcal/mol for SCAN and 0.13 kcal/mol for DC-SCAN, whereas the average $|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ associated with aDZ calculations for neutral water clusters is 0.21 kcal/mol for SCAN and 0.15 kcal/mol for DC-SCAN.

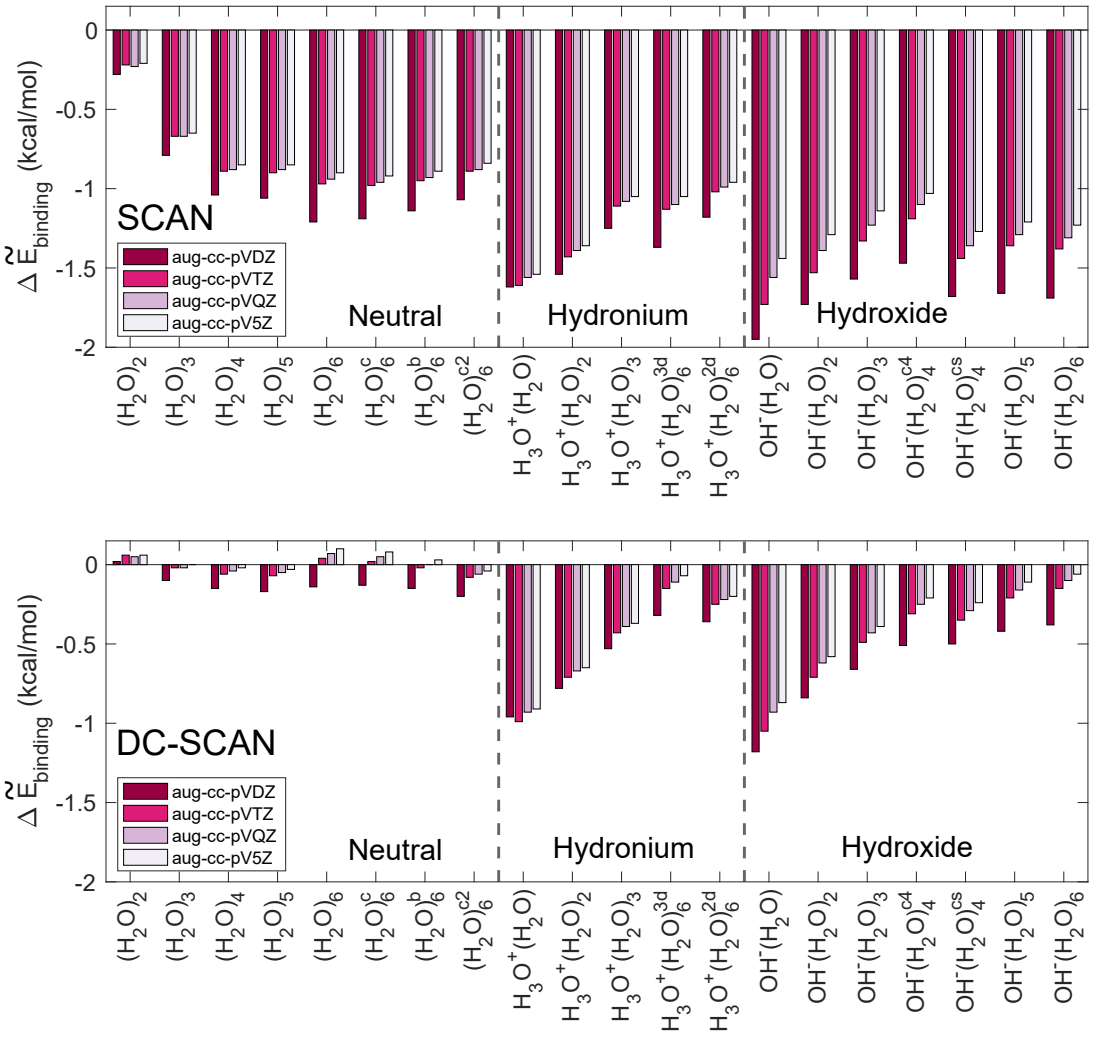

Figure 1: Convergence of the binding energies per monomer for the neutral $(H_2O)_{2-6}$, protonated $H_3O^+(H_2O)_{1-6}$, and deprotonated $OH^-(H_2O)_{1-6}$ water clusters of the WATER27 dataset[88] calculated with SCAN and DC-SCAN using the aDZ, aTZ, and aQZ basis sets relative to the CCSD(T)/CBS reference values.[89]

$|\Delta\tilde{E}_{\text{bind}}|$ converges to 0.81 kcal/mol and 0.04 kcal/mol for both the aTZ and aQZ calculations using SCAN and DC-SCAN, respectively. Conversely $|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ decreases to 0.04 kcal/mol and 0.03 kcal/mol for the aTZ and aQZ calculations, respectively, carried out with both SCAN and DC-SCAN. The SCAN results with the a5Z basis set display a mean unsigned error (MUE) of 0.76 kcal/mol for SCAN, which decreases to 0.05 kcal/mol for the DC-SCAN results relative to the CCSD(T)/CBS reference values. For the neutral water clusters

$|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ converges quickly going from aDZ to aTZ to aQZ for both SCAN and DC-SCAN.

The protonated water clusters show a similar trend, with $|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ being 0.27 kcal/mol for SCAN and 0.21 kcal/mol for DC-SCAN with the aDZ basis set, and decreasing to 0.04 kcal/mol for SCAN and 0.03 kcal/mol for DC-SCAN with the aQZ basis set. On the other hand, $|\Delta\tilde{E}_{\text{bind}}|$ displays a MUE of 1.39 and 0.59 kcal/mol for the aDZ basis set, which reduces monotonically to 1.19 and 0.44 kcal/mol when the larger aQZ basis set is used in the SCAN and DC-SCAN calculations, respectively.

It is worth mentioning that both SCAN and DC-SCAN binding energies calculated for neutral and protonated water clusters using the aDZ basis set display relatively larger errors for three-dimensional structures than planar structures.

In contrast, a different trend is found for the deprotonated water clusters, with the deviations from the reference binding energies calculated with the a5Z basis set being appreciably larger for both SCAN and DC-SCAN. This is particularly evident in the aDZ calculations for which the mean $|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ is 0.63 kcal/mol for SCAN and 0.40 kcal/mol for DC-SCAN. The addition of extra diffuse functions to the basis set results in similar trends as those observed for the neutral and protonated water clusters, with the mean $|\Delta\tilde{E}^{\text{basis}}_{\text{conv}}|$ reducing to 0.13 kcal/mol and 0.06 kcal/mol in SCAN and DC-SCAN calculations, respectively, carried out with the aQZ basis set. Figure 1 shows that the binding energies of the deprotonated water clusters calculated with DC-SCAN converge relatively more quickly with the size of the basis set than the corresponding SCAN values, as the MUE for $|\Delta\tilde{E}_{\text{bind}}|$ decreases from 1.68 kcal/mol to 1.23 kcal/mol going from the aDZ to the a5Z basis sets in the SCAN calculations, and from 0.64 kcal/mol to 0.35 kcal/mol for the same DC-SCAN calculations. The faster convergence observed for the DC-SCAN binding energies is related to the presence of density-driven errors that prevent SCAN from fully binding the excess electron in OH^- and $OH^-(H_2O)$ even when using large basis sets.[49] This problem is, at least partially, removed by using the more localized and SIE-free Hartree-Fock density, which thus explains the faster convergence of the DC-SCAN binding energies with the basis set size.

Table 1: Mean unsigned errors of SCAN binding energies for different integration grids with respect to the (250,974) quadrature for neutral water, hydronium and hydroxide water clusters. The cluster geometries have been extracted from the WATER27 dataset.[88] The neutral water clusters consist of water dimer to hexamer geometries whereas the hydronium and hydroxide clusters span from $H_3O^+\cdots H_2O$ to $H_3O^+\cdots(H_2O)_6$ and $OH^-\cdots H_2O$ to $OH^-\cdots(H_2O)_6$ respectively. All the individual binding energies using different integration grids have been presented in Figure S2

Quadrature	SG-1	SG-2	SG-3	(75,302)	(99,590)
MUE (kcal/mol)	0.11	0.04	0.02	0.06	0.04

It has been established that the choice of the integration grid is as important as the choice of the basis set for calculations carried out with some modern XC functionals. In particular, some recent meta-GGAs, such as SCAN, and some hybrid functionals, such as M11, require a relatively large number of grid points in order to properly integrate the kinetic energy density.[79] To determine the sensitivity of SCAN calculations on the grid size, the binding energies of the neutral, protonated, and deprotonated water clusters of the WATER27 dataset were calculated using the standard pruned grids of Q-Chem,[76] i.e., SG1,[90] SG2,[79] SG-3[79] as well as the unpruned (75,302), (99,590), and (250,974) grids, where the first number in the parenthesis defines the number of radial points using the Euler-Maclaurin quadrature[77] and the second number defines the number of angular points using the Lebedev quadrature.[78] It was shown that smaller grids can be used for r^2SCAN without loss of overall accuracy.[80]

Table 1 reports the MUE associated with different grids relative to the highly dense (250,974) grid which provides 243,500 grid points per atom. The SG-1 grid, which corresponds to the pruned version of the (50,194) grid and provides ∼5000 grid points per atom, is associated with a MUE of 0.11 kcal/mol. The MUE systematically decreases for larger pruned integration grids, with the SG-2 (∼7500 points per atom) and SG-3 (∼ 17000 points per atom) grids displaying MUEs of 0.04 kcal/mol and 0.02 kcal/mol, respectively. The MUEs of the pruned integration grids are comparable with the MUEs associated with the corresponding unpruned (75,302) and (99,590) grids, which provide 22,650 and 58,410 points

points per atom, respectively. This analysis suggests that the standard SG-2 grid provides an optimal compromise between accuracy and computational cost for SCAN calculations of aqueous clusters although for more grid-sensitive systems it might be necessary to employ the denser SG-3 grid. In this regard, it should be noted that both the SG-2 and SG-3 grids have been extensively tested on different datasets and non-covalent interactions.[79]

3.2 Interaction and binding energies for neutral water clusters

To study the effect of the density correction on neutral water clusters, we calculated the interaction and binding energies for the subset of neutral water clusters of the BEGDB dataset[91] which contains 38 low-energy isomers of the $(H_2O)_{n=2-10}$ clusters, with all structures optimized at the RI-MP2/aug-cc-pVDZ level of theory.[92] The different isomers correspond to global and local minima of the corresponding $(H_2O)_{n=2-10}$ clusters obtained by combining molecular dynamics sampling with high-level quantum-chemical calculations.[91] In the following analyses, the reference interaction and binding energies are taken from ref 89, while the corresponding values for SCAN and DC-SCAN were calculated with and without the empirical D3 dispersion correction, resulting in the corresponding SCAN-D3 and DC-SCAN-D3 functionals. The optimized D3 coefficients for the SCAN functional were adapted from ref 93.

Figure 2 shows the absolute errors in interaction (panel a) and binding (panel b) energies associated with the four SCAN-based functionals relative to the CCSD(T)-F12b reference values.[89] Note that these errors are for entire clusters, and not per molecule as in the previous section. The interaction energies calculated with SCAN increasingly deviate from the reference values as a function of the cluster size, ranging from 0.44 kcal/mol for the dimer ($2C_s$) to 10.84 kcal/mol for the decamer (10PP1 and 10PP2). By improving the quality of the electron density, these errors are nearly completely removed in the corresponding calculations carried out with DC-SCAN. Specifically, the DC-SCAN errors for the entire dataset comprising all 38 water clusters lie well within chemical accuracy (i.e., 1 kcal/mol), ranging

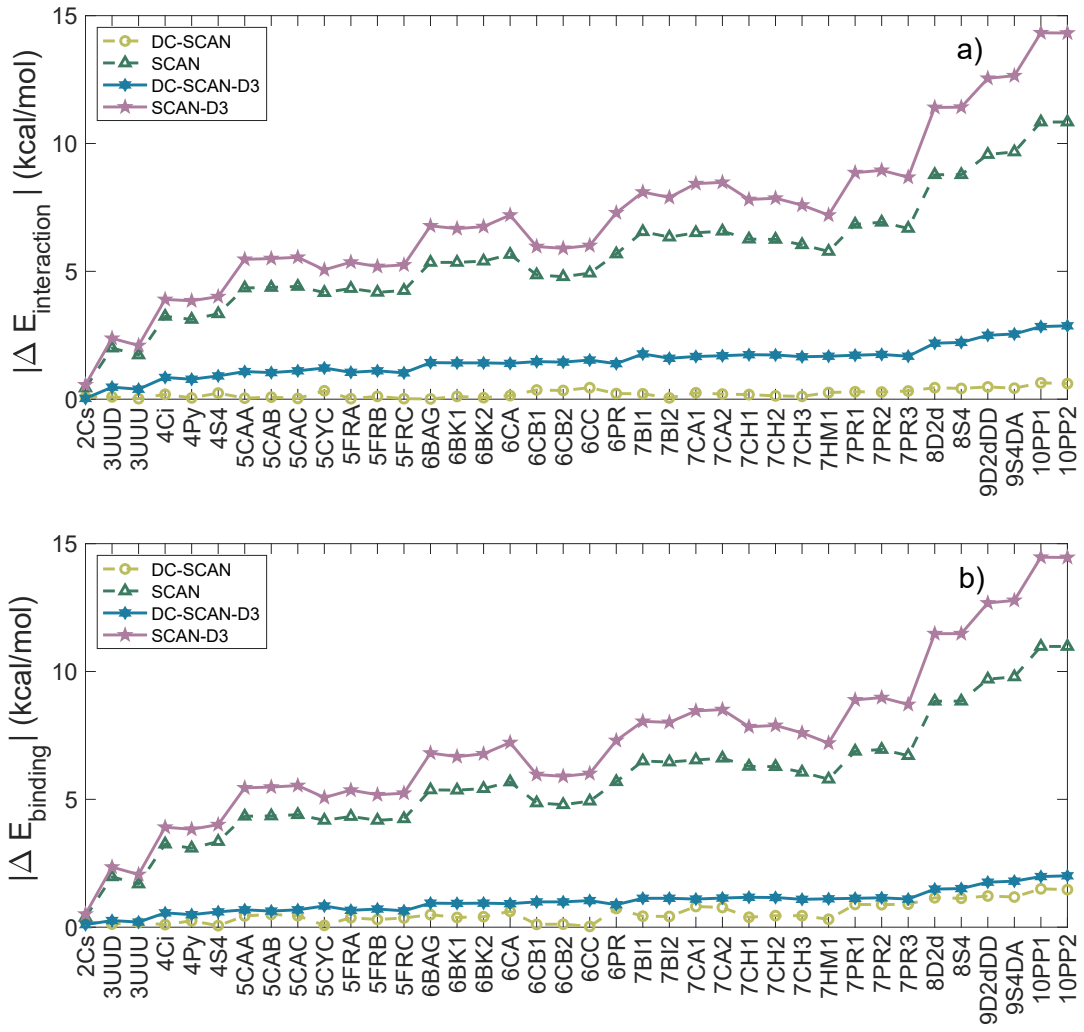

Figure 2: Absolute errors in a) interaction and b) binding energies calculated for the neutral water cluster subset of the BEGDB dataset using SCAN, DC-SCAN, SCAN-D3, and DC-SCAN-D3 with respect to the CCSD(T)/CBS reference values.[89]

from 0.02 kcal/mol for the pentamer (5FRC) to 0.64 kcal/mol for the decamer (10PP1).

The addition of the D3 dispersion correction deteriorates the energetics of both SCAN and DC-SCAN, with the MUE increasing from 5.66 kcal/mol for SCAN to 7.19 kcal/mol for SCAN-D3, and from 0.22 kcal/mol for DC-SCAN to 1.49 kcal/mol for DC-SCAN-D3.

Similar trends are also observed for the binding energies, with the MUE for the SCAN functional being 5.69 kcal/mol compared to 0.54 kcal/mol for DC-SCAN. As for the interaction energies, the addition of the D3 correction worsens the overall agreement with the

CCSD(T)/CBS reference values, with SCAN-D3 and DC-SCAN-D3 displaying MUEs of 7.21 kcal/mol and 0.99 kcal/mol, respectively. Importantly, the MUE per monomer for both interaction and binding energies remains almost constant in the DC-SCAN and DC-SCAN-D3 calculations for all 38 clusters (Figure S3 and S4), while it increases with the cluster size in the SCAN and SCAN-D3 calculations. This trend indicates that density-driven errors grow faster than the number of hydrogen bonds which can be qualitatively explained by considering that, as the cluster size increases, the electron density can delocalize over a larger space.

3.3 Binding energies for the WATER27 dataset

In order to investigate whether the high accuracy of DC-SCAN is only specific to water or it also transfers to other aqueous systems, we analyzed the binding energies of the WATER27 dataset.[88] The compiled database consists of 27 clusters optimized at the B3LYP/6-311++G(2d,2p) level of theory,[88] which includes a set of 14 neutral water clusters [$(H_2O)_n$, with $n = 2-6, 8, 20$], 5 protonated water clusters [$(H_3O^+(H_2O)_n$, with $n = 1-3, 6$], 7 deprotonated water clusters [$OH^-(H_2O)_n$, with $n = 1-6$], and 1 autoionized water cluster [$H_3O^+(H_2O)_4OH^-$]. The absolute errors of the binding energies were calculated with respect to the CCSD(T)-F12 reference values taken from ref 89. Figure 3 shows that, as for the BEGDB dataset, DC-SCAN performs remarkably well for the neutral water clusters of the WATER27 dataset, displaying a MUE of 1.43 kcal/mol which should be compared to a MUE of 9.89 kcal/mol for SCAN. The MUEs calculated for the protonated water clusters are 1.55 kcal/mol and 5.24 kcal/mol for DC-SCAN and SCAN, respectively. For the deprotonated water clusters, DC-SCAN displays a MUE of 1.39 kcal/mol that must be compared with a MUE of 5.91 kcal/mol obtained with SCAN. By construction, all DFAs generally predict large errors for the $H_3O^+(H_2O)_4OH^-$ cluster since this cluster exhibits a relatively strong multiconfiguration character which is thus difficult for DC-SCAN to describe. Without symmetry-breaking of the Hartree-Fock density, DC-SCAN makes a large functional-driven

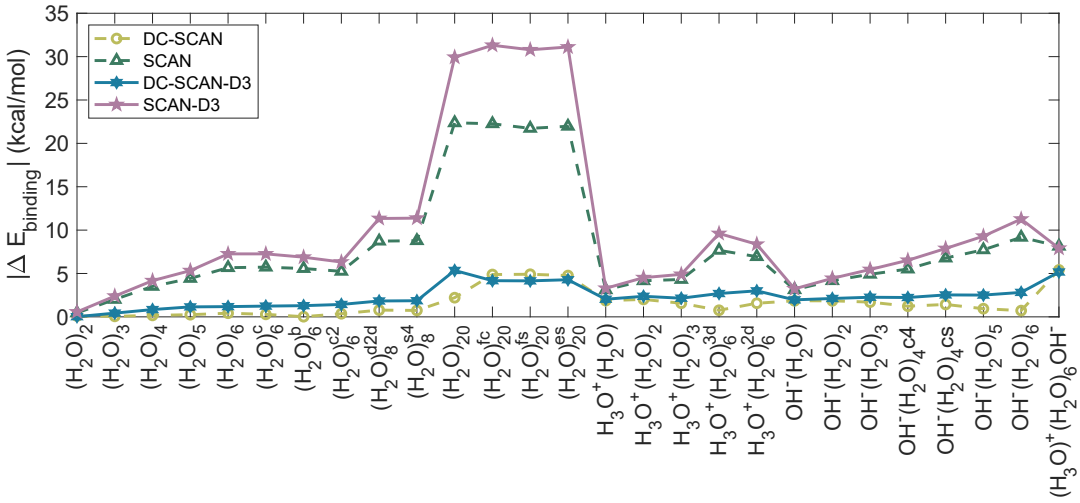

Figure 3: Absolute errors in binding energies calculated for the WATER27 dataset mers using SCAN, DC-SCAN, SCAN-D3, and DC-SCAN-D3, with respect to the CCSD(T)/CBS benchmark.[89]

error. In this case, the error associated with DC-SCAN is 5.41 kcal/mol, whereas SCAN displays an error of 8.14 kcal/mol. The MUE for the whole WATER27 dataset is 1.59 kcal/mol for DC-SCAN, which is significantly smaller than the MUE of 7.93 kcal/mol associated with SCAN. As expected, the addition of the D3 dispersion correction results in significant overbinding for all clusters, with SCAN-D3 and DC-SCAN-D3 being associated with MUEs of 10.10 kcal/mol and 2.34 kcal/mol, respectively. Overall, the analysis of the binding energies of the WATER27 dataset indicates that, by largely reducing density-driven errors, DC-SCAN is a highly accurate functional not only for neutral water, as demonstrated in ref 62, but also for protonated and deprotonated water.

To summarize the performance of the SCAN, SCAN-D3, DC-SCAN, DC-SCAN-D3 functionals, the corresponding error distributions associated with the binding energies calculated relative to the reference CCSD(T) values for the combined BEGDB and WATER27 datasets are shown as violin plots in Figure 4. A violin plot provides detailed information about a given distribution and is based on a Gaussian kernel density estimation using Scott's rule[94] as implemented in Matlab. Specifically, the black box inside the violin holds 50% of the

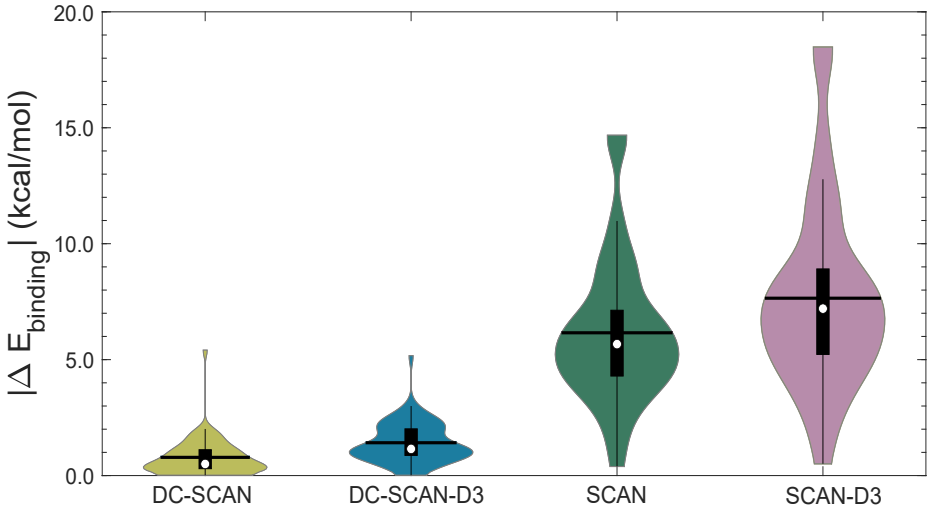

Figure 4: The violin-plot error distribution for binding energies of BEGDB+WATER27 dataset using SCAN, SCAN-D3, DC-SCAN, DC-SCAN-D3 with respect to reference CCSD(T) values. The violin plots represent the data distribution and are based on a Gaussian kernel density estimation using Scott's rule[94] as implemented in Matlab. In the box plot, the boxes hold 50% of the data, with equal number of data points above and below the median, marked by the diamond. The whiskers denote the range of data falling within 1.5×box-length beyond the upper and lower limits of the box. The horizontal line marks the mean.

data, with an equal number of data points above and below the median that is marked by the white circle. It follows that the bottom and top parts of the black box correspond to the first and third quartiles, respectively. The horizontal line in the violin marks the mean of the overall data, while the bottom and top points of the whisker define the first quartile minus 1.5× IQR and third quartile plus 1.5×IQR, respectively, where IQR is the interquartile range. The shape of the violin is proportional to the probability distribution of the data, i.e., a wider shape indicates higher frequency of the corresponding value in the distribution. From the statistics presented in Figure 4 it is evident that DC-SCAN outperforms SCAN on the combined BEGDB and WATER27 datasets, consistently providing lower values for the mean, median, and first and third quartiles of the error distribution. It is also worth noting that DC-SCAN yields a significantly narrower error distribution than SCAN, which can be traced back to the reduction of density-driven errors in DC-SCAN. Finally, the deleterious effect of adding the D3 dispersion correction discussed above becomes quite apparent

from the analysis of the corresponding violin plots, with SCAN-D3 providing the worst performance in terms of mean, median, and maximum error. Figures 2-4 demonstrate that using the localized and self-interaction-free Hartree-Fock density significantly improves the description of neutral, protonated, deprotonated, and autoionized water clusters, bringing the binding energies calculated with DC-SCAN closer to the CCSD(T) reference values.

In Sections 3.4 and 3.5 we will take a closer look at the various electron densities for the deprotonated, neutral, and protonated water dimers.

3.4 Density error and density sensitivity

For small atoms[65] and typical molecules[39] near equilibrium, including the water molecule, the meta-GGA SCAN functional[38] is more accurate for the electron density than the local spin density approximation[23] or the Perdew-Burke-Ernzerhof (PBE) GGA functional,[31] and (except in low-density tails of little importance to the energy) more accurate for the density than the Hartree-Fock approximation. In particular, SCAN is superior for dipole moments[95] and polarizabilities.[96] The energies of atoms and small molecules as well as the atomization energies of small molecules are rather insensitive to the density, and can be slightly worsened by using the Hartree-Fock density in place of the SCAN self-consistent density. The Hartree-Fock density is mainly useful for density correction in weakly-bonded clusters or in complexes with stretched bonds, including the transition states of chemical reactions, where the delocalization errors of semi-local approximations become energetically important.

Quintessential delocalization errors of semilocal approximations occur for two chemically-different open-shell atoms at infinite separation, which in exact theories are electrically neutral. It has been known[66,71,97] for more than 40 years that semi-local total energies of isolated open systems vary with non-integer electron number like upward-curving parabolas. For separated Na$\cdots$Cl, LDA minimizes the total energy by a spurious transfer of 0.4 electrons from Na to Cl, and makes the total energy lower than that of the isolated neutral atoms by about 5 kcal/mol (overbinding).[98] Unusually large overbinding by an accurate semilocal functional

like SCAN suggests density-driven error.

The exact density functional avoids spurious charge transfers by making the total energy of an isolated open system vary linearly[66,71] with electron number between adjacent integers, with derivative discontinuities at the integers. The Hartree-Fock approximation also produces neutral atoms at infinite separation, but it does so by being too localizing, with its total energy for an isolated open system varying between adjacent integer electron numbers like a downward-curving parabola.

Before making a density correction (using for example the Hartree-Fock density as a proxy for the exact one) to a particular energy difference evaluated with a particular DFA, it is recommended to compute the density sensitivity[57,59] ($|E_{DFA@HF} - E_{DFA@LDA}|$) or change of the energy difference from DFA@LDA (using the LDA density) to DFA@HF (using the HF density). Only if this difference is significantly large it is recommended to make a density correction.

To this end, Figure 5 compares the SCAN errors relative to the CCSD(T) reference values for the BEGDB clusters ($E_{SCAN} - E_{CCSD(T)}$) with the errors associated with SCAN calculations carried out non-self-consistently using the LDA density ($E_{SCAN@LDA} - E_{CCSD(T)}$). Also

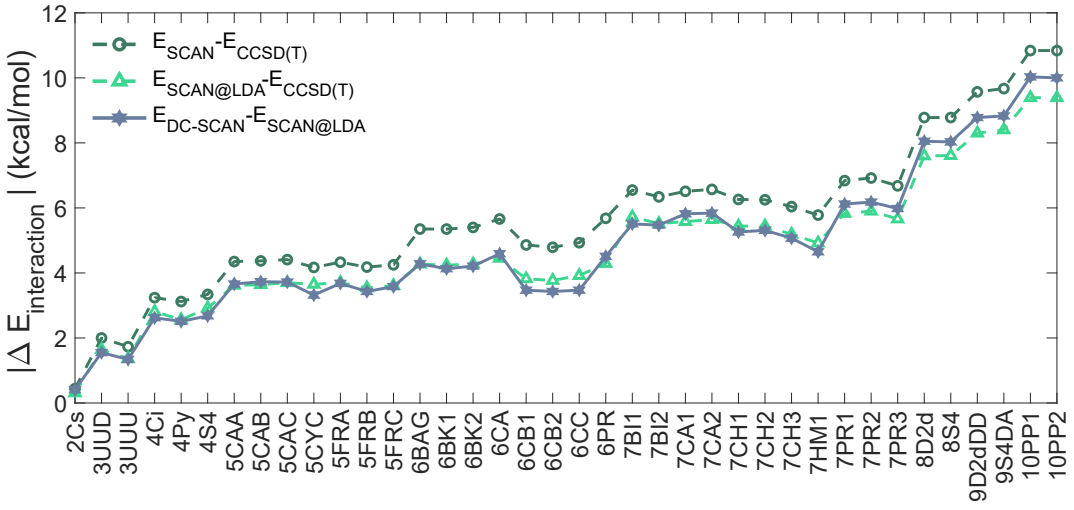

Figure 5: Absolute errors ($E_{SCAN} - E_{CCSD(T)}$, $E_{SCAN@LDA} - E_{CCSD(T)}$, and $E_{DC-SCAN} - E_{SCAN@LDA}$) in interaction energies calculated for the water clusters of the BEGDB dataset.

shown is the density sensitivity, i.e., the difference between the DC-SCAN and SCAN@LDA results ($E_{DC-SCAN}$ - $E_{SCAN@LDA}$). As expected from the discussion of density-driven errors in Sections 3.2, Figure 5 shows that the BEGDB dataset exhibits large density sensitivity, with a MUE of 4.82 kcal/mol.

Figure 5 shows that, relative to CCSD(T), the MUE of SCAN@LDA (4.78 kcal/mol) is less than the MUE of self-consistent SCAN (5.66 kcal/mol). This might suggest that the LDA density is somehow more accurate or less-delocalized than the SCAN density, which would be unexpected (and untrue, as shown in Section 3.5). The smaller MUE of SCAN@LDA can be explained by considering that the SCAN energy of a cluster is minimized by the SCAN density, which has a small degree of delocalization error that leads to overbinding. Suppose that delocalization error is only important in the cluster, and not in its constituent molecules. Then any variation from the SCAN degree of delocalization, whether it is toward greater localization (as in the Hartree-Fock density) or toward greater delocalization (as in LDA) can move the energy of the cluster above its parabolic minimum and so reduce the overbinding error of self-consistent SCAN. In other words, a change of binding energy in the right direction can occur for either a right reason or a wrong reason.

The better performance provided by SCAN@LDA compared to SCAN suggests that some caution should be exercised when using the original definition of density sensitivity[99] to measure deviations in DFT energies calculated with the Hartree-Fock and LDA electron densities as representative densities corresponding to the two extremes of the delocalization error.

3.5 Visualization and quantification of density error

It can be argued, with caveats like the one discussed at the end of Section 3.4, that the best measure of electron density error is the error it drives in an energy difference evaluated for a given density functional. After all, the energy difference is a single number, and often it is the needed number, while the density is defined by many numbers. Nevertheless, it

can also be of interest to investigate the density errors directly, as is done for the neutral, protonated, and deprotonated water dimers in this section. First the errors of the SCAN density relative to the HF density will be plotted in three dimensions and also projected onto the hydrogen-bond axis. Then the errors of the HF and SCAN densities relative to an accurate CCSD density will be reduced to single numbers. A single-number error measure can lose a lot of relevant information and yet still aid human understanding. Understanding density error is, however, still far from explaining the size of the density-driven error of an energy or energy difference, for which computer calculations of the energy seem unavoidable.

A more reliable density correction to the energy would employ the accurate CCSD density instead of the HF density. However, this is a difficult task. Finding the CCSD density itself is the first and lesser difficulty. The bigger problem is that the CCSD kinetic energy and kinetic energy density include kinetic energy of correlation, which would be double-counted since it is already present in the exact and approximate XC energies of Kohn-Sham and generalized Kohn-Sham theories. One might need to invert the CCSD density[100] to find its Kohn-Sham potential, which is not only demanding but also, in many approaches, insufficiently accurate.

Further insights into density errors can be gained from the analysis of $\Delta\rho(\mathbf{r})$ defined as

$$\Delta\rho(\mathbf{r}) \equiv \rho_{\text{approx}}(\mathbf{r}) - \rho_{\text{exact}}(\mathbf{r}) \implies \Delta\rho(\mathbf{r}) \simeq \rho_{\text{DFA}}(\mathbf{r}) - \rho_{\text{HF}}(\mathbf{r}) \tag{9}$$

where $\rho_{\text{approx}}(\mathbf{r})$ and $\rho_{\text{exact}}(\mathbf{r})$ are the approximate and (unknown) exact electron densities, respectively. As discussed in Section 2, in this study, $\rho_{exact}(\mathbf{r})$ is approximated by $\rho_{\text{HF}}(\mathbf{r})$ (unless otherwise stated), while $\rho_{\text{approx}}(\mathbf{r}) = \rho_{\text{DFA}}(\mathbf{r})$. Figure 6 shows 3-dimensional representations of $\Delta\rho(\mathbf{r})$ for the $OH^-(H_2O)$, $(H_2O)_2$, $H_3O^+(H_2O)$ dimers calculated using the SCAN and SCAN0 functionals. In both cases, the isosurfaces corresponding to $|\Delta\rho| = 0.006$ e/Bohr^3 are shown. In the case of $OH^-(H_2O)$, the isosurface of $\Delta\rho(\mathbf{r})$ calculated with SCAN has an area of 31.80 Bohr^2 which shrinks to 16.28 Bohr^2 when the calculations are carried out with SCAN0. Similar trends are observed for $(H_2O)_2$ and $H_3O^+(H_2O)$, with SCAN

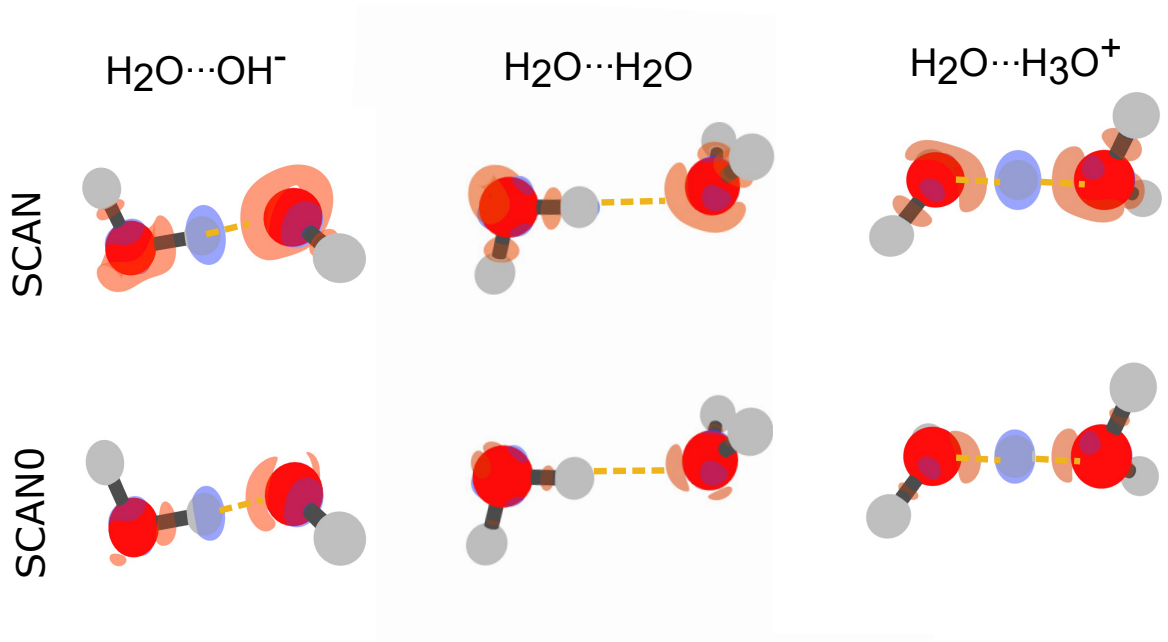

Figure 6: Isosurfaces for the density error corresponding to $|\Delta\rho| = 0.006$ e/Bohr^3 calculated using SCAN/aug-cc-pVQZ and SCAN0/aug-cc-pVQZ for $OH^- \cdots H_2O$, $H_2O \cdots H_2O$ and $H_3O^+ \cdots H_2O$. The orange and blue lobes represent the negative and positive isosurfaces of $\Delta\rho(\mathbf{r})$, respectively. The reference density is the HF density.

and SCAN0 providing isosurface areas of 22.89 Bohr^2 and 11.67 Bohr^2, respectively, for the neutral dimer, and 26.88 Bohr^2 and 15.72 Bohr^2, respectively, for the protonated dimer. The larger isosurface area obtained with SCAN for $OH^-(H_2O)$ provides further evidence for the SCAN calculations of negatively charged clusters suffering from relatively larger density-driven errors due to the SCAN density being overly delocalized and, therefore, unable to correctly bind the excess electron as discussed in Section 3.1. In SCAN, the extra electron is bound by the restriction to localized basis sets.

From the shape of the isosurfaces, it is evident that, compared to SCAN0, SCAN predicts relatively higher electron density on the hydrogen atom of the water molecule that acts as hydrogen-bond donor, which is accompanied by the reduction of the electron density near the oxygen atoms of both hydrogen-bond donor (O_D) and acceptor (O_A) species. This leads to artificially stronger intermolecular interactions and results in significant over-binding as already shown in Figures 2 and 3. By reducing the density-driven errors, the addition of a

fraction (25%) of HF exchange in SCAN0 results in a smaller delocalization of the electron density and, consequently, improved description of all three hydrogen-bonded dimers.

More quantitative insights into the magnitude of the density errors can be obtained by the analysis of $\Delta\rho(\mathbf{r})$ projected onto the hydrogen bond, which corresponds to the z-axis in Figure 7,

$$\Delta\tilde{\rho}(z) = \int dx \int dy\, \Delta\rho(x, y, z) \tag{10}$$

eq 10 is just eq 4 of ref 42 in the perfect resolution limit $a^{-1} \to 0$, since, when $a \to \infty$, $(\frac{a}{\sqrt{\pi}})exp[-a^2(z' - z)^2] \to \delta(z' - z)$. Figure 7 clearly shows the depletion of electron density along the hydrogen bond which is accompanied by the deposition of additional electron density on the hydrogen atom of the donor species.

Furthermore, Figure 7 demonstrates that density-driven errors are larger in SCAN than SCAN0 (as would be expected since SCAN0 has 25% of HF exchange).

It is also worth noting that the fluctuations of $\Delta\tilde{\rho}(z)$ are larger for the ionic systems (i.e., the protonated and deprotonated water dimers), which is consistent with the relatively larger density-driven errors associated with these systems.

Section 3.4 discussed a quintessential density-driven error in which atoms at infinite

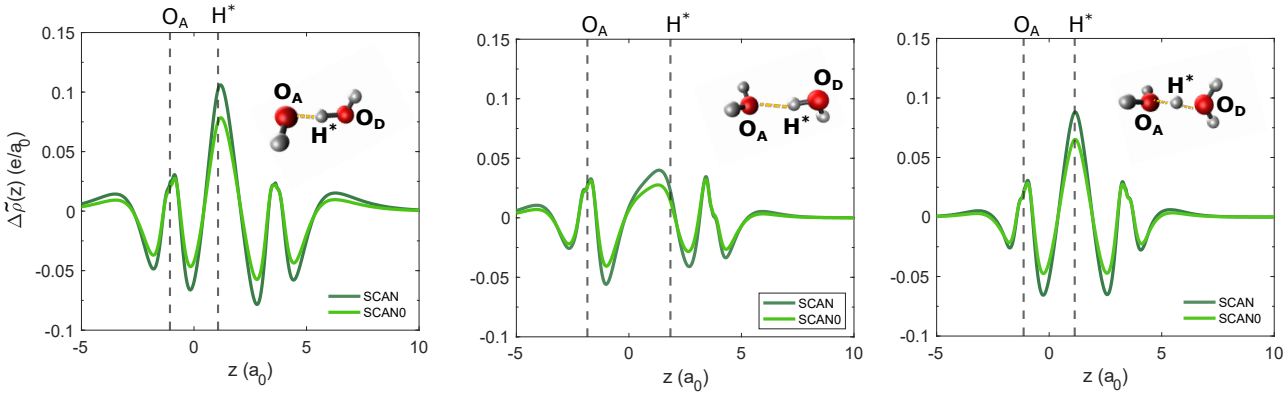

Figure 7: Projection of the density error ($\Delta\rho(\mathbf{r})$) onto the H$\cdots$O bond-axis for the $OH^-\cdots H_2O$, $H_2O\cdots H_2O$ and $H_3O^+\cdots H_2O$ dimers. In all three panels, the origin of the z-axis is placed at the midpoint between the H^* and O_A atoms. The reference density is the HF density.

Table 2: Benchmark CCSD(T)/CBS binding energies (BE) of the deprotonated, neutral, and protonated water dimers, and the errors of various approximations, in kcal/mol.

Cluster	BE	error in BE			
	CCSD(T)/CBS	urCCSD	DC-SCAN	SCAN	PBE
$OH^-\cdots H_2O$	-26.69	0.44	-1.87	-3.12	-2.72
$H_2O\cdots H_2O$	-4.97	0.21	0.13	-0.44	-0.19
$H_3O^+\cdots H_2O$	-33.74	0.40	-1.85	-3.11	-3.32

separation can be overbound through the spurious transfer of a fraction of an electron. It is thus interesting to ask if any residue of this effect can be found in the three 20-electron water dimers analyzed in Figures 6 and 7 whose binding energies are reported in Table 2. To try to answer this question, we calculated the CCSD density for each dimer and divided the space around each dimer by a plane perpendicular to the hydrogen-bond z-axis and positioned at $z = z_0$ so that the CCSD density has exactly 10 electrons on each side. It is then possible to quantify the density-driven errors associated with a given density relative to the CCSD density in terms of the error in the number of electrons on the H^* side of the plane, which is equal but of opposite sign to that on the O_A side, as

$$\int_{z_0}^{\infty} dz \Delta\tilde{\rho}(z) \tag{11}$$

The electron number errors associated with the HF, SCAN, and PBE densities are listed in Table 3. It should be noted that the standard or unrelaxed CCSD density is not variational and, therefore, does not satisfy the Hellmann-Feynman theorem. To address this potential problem, the reference density is taken from orbital-optimized[101] or partly relaxed CCSD. As shown in Table 3, the orbital optimization is not very important for the water dimers analyzed here.

The first thing to notice in Table 2 is that the deprotonated and protonated water dimers are much more strongly bound than the neutral water dimer. There is a standard hydrogen bond only in the neutral dimer.

Table 3: Deviations in various approximations (including unrelaxed or urCCSD) from 10 electrons on the H^* side of a plane perpendicular to the hydrogen-bond axis and positioned so that, in benchmark orbital-optimized or partly-relaxed CCSD, there are exactly 10 electrons on each side. $^{\Delta z}/_{d_{\mathrm{OH}}}$, where Δz is the distance along the hydrogen-bond from the H^* nucleus to that plane, and d_{OH} is the length of the hydrogen- or longest bond. Note that this plane is always between the O_A and $H^\star$ nuclei. The calculations for Table 3 were done with the ORCA code and the aug-cc-pVQZ basis sets. The dimensions (in bohrs) of the cube file were 25×25×25 for the deprotonated dimer, 25×25×28 for the neutral dimer, and 20×20×20 for the protonated dimer. 1210 grid points were placed along each axis.

Cluster	$^{\Delta z}/_{d_{\mathrm{OH}}}$	error in electron number			
		urCCSD	HF	SCAN	PBE
$OH^-\cdots H_2O$	0.0818	0.0006	-0.0034	0.0057	0.0102
$H_2O\cdots H_2O$	0.2932	0.0001	-0.0008	0.0033	0.0044
$H_3O^+\cdots H_2O$	0.0003	0.0000	-0.0001	0.0001	0.0001

This is confirmed by the intermolecular and intramolecular OH bond lengths ($O_A\cdots H^*$, $O_D{-}H^*$) which are (1.34 Å, 1.13 Å), (1.95 Å, 0.97 Å), and (1.20 Å, 1.20 Å) for the deprotonated, neutral, and protonated water dimers, respectively. Thus an analysis based on quintessential delocalization error is most likely to be relevant to the neutral dimer. The second thing to notice is that, while DC-SCAN is always an improvement over self-consistent SCAN, the correction is more effective in the neutral dimer than in the other two dimers, suggesting that the HF density might be best for the neutral dimer. While PBE overbinds the water dimer less than SCAN does, only SCAN correctly predicts[47] the small energy differences between the water hexamers that are more relevant to the bulk phases. Overbinding of the dimer by PBE or SCAN is not large, but it is amplified in larger clusters.

Table 3 shows that in all three dimers, the SCAN error of electron number on the $H^\star$ (donor) side is larger in magnitude than, and of opposite sign to, the corresponding HF density. This is consistent with the claim that SCAN is too delocalizing, and that HF is slightly too localizing. In all cases, PBE is more delocalizing than SCAN. For the neutral water dimer, LDA is still more delocalizing than PBE, as anticipated in Section 3.4, with an error of electron number equal to 0.0058.

In the neutral water dimer, the SCAN error in electron number is more than three times

larger in magnitude than the HF error, suggesting that the HF density should be a good basis for density correction. In the deprotonated water dimer, all the errors of electron number are larger in magnitude, and the HF error is closer in magnitude to the SCAN error, suggesting that density correction based on HF should be less successful. In the protonated water dimer, there is an approximate symmetry around the central H atom that makes all errors in electron number fall close to zero, but there is still substantial density error for SCAN, as shown in the more detailed Figures 6 and 7.

It is asserted here, not that these small errors of electron number by themselves explain the density-driven errors of the SCAN water-dimer binding energy, but only that they indicate a residue of the much larger SCAN delocalization error and also of the HF localization error found in many pairs of infinitely-separated and chemically-different atoms. The results in Table 3 are however consistent with the viewpoint of ref 102, which found that a GGA description of liquid water became much more accurate when an additional subsystem approximation was made, and attributed that to the fact that each subsystem was a water molecule constrained to have exactly 10 electrons.

3.6 Comparison between DC-DFT and FLOSIC-SCAN

As mentioned in the Introduction, an alternative approach to removing the SIE in DFT calculations is provided by FLOSIC. Several variations of FLOSIC are being explored, but we will only discuss the PZ-SIC[42] version of eq 4). Using this version, ref 47 found that self-interaction error in SCAN overbinds water clusters but nearly cancels out of structural energy differences, while ref 49 found large FLOSIC-SCAN errors for some water-ion clusters.

Although FLOSIC-SCAN is just as free of one-electron self-interaction error as Hartree-Fock theory, the FLOSIC-SCAN density sometimes displays more pronounced delocalization error. For the neutral water dimer, FLOSIC-SCAN makes an error of electron number on the H^* side of about 0.0024, which is similar to the value of 0.0033 calculated with SCAN (Table 2). This is surprising, since PZ-SIC typically reduces the spurious fractional charge

on dissociated atoms to zero.[97] The change in binding energy is negligible from SCAN to SCAN@FLOSIC, but FLOSIC-SCAN yields an accurate binding energy of 5.1 kcal/mol. For a different set of systems, small organic molecules adsorbed on a Cu^+ ion, PBE@FLOSIC and PBE@HF correct the PBE binding emergy to nearly the same extent.[103] Comparisons[104] of SCAN@HF to SCAN@FLOSIC are underway for the transition states of chemical reactions.

To compare the performance of the FLOSIC and DC methods, in Figure 8 we consider four different hydrogen-bonded systems, including protonated and deprotonated clusters, as well as ionic aqueous clusters containing a single alkali-metal or halide ion.[49] For each system, the errors relative to the CCSD(T) reference values were calculated using SCAN and DC-SCAN, while the FLOSIC-SCAN and SCAN@FLOSIC results were taken from ref 49. Following the same definitions introduced in ref 49, FLOSIC-SCAN corresponds to fully self-consistent FLOSIC calculations with the SCAN functional, while SCAN@FLOSIC corresponds to non-self-consistent calculations carried out using the self-interaction-free FLOSIC density and Fermi orbitals. The difference between the SCAN and SCAN@FLOSIC energies provides an alternative estimate of density-driven errors which is analogous to the difference between the SCAN and DC-SCAN energies.

Figure 8 shows that SCAN predicts relatively large errors for the protonated water clusters, with a MUE of 5.24 kcal/mol. The SCAN@FLOSIC results display similar errors, resulting in a MUE of 5.69 kcal/mol. As expected from the analyses presented in Sections 3.2 and 3.3, the use of the localized HF density in the DC-SCAN calculations reduces the error significantly, leading to a MUE of 1.56 kcal/mol. The fully self-consistent orbital-by-orbital self-interaction corrected FLOSIC-SCAN displays a MUE of 2.88 kcal/mol. It is worth mentioning that the accuracy of the self-interaction corrected FLOSIC-SCAN calculations for the protonated water clusters does not follow a monotonic trend as a function of the cluster size, while the error per monomer systematically decreases in the corresponding DC-SCAN calculations.

For the deprotonated water clusters, SCAN displays a MUE of 5.91 kcal/mol, which must

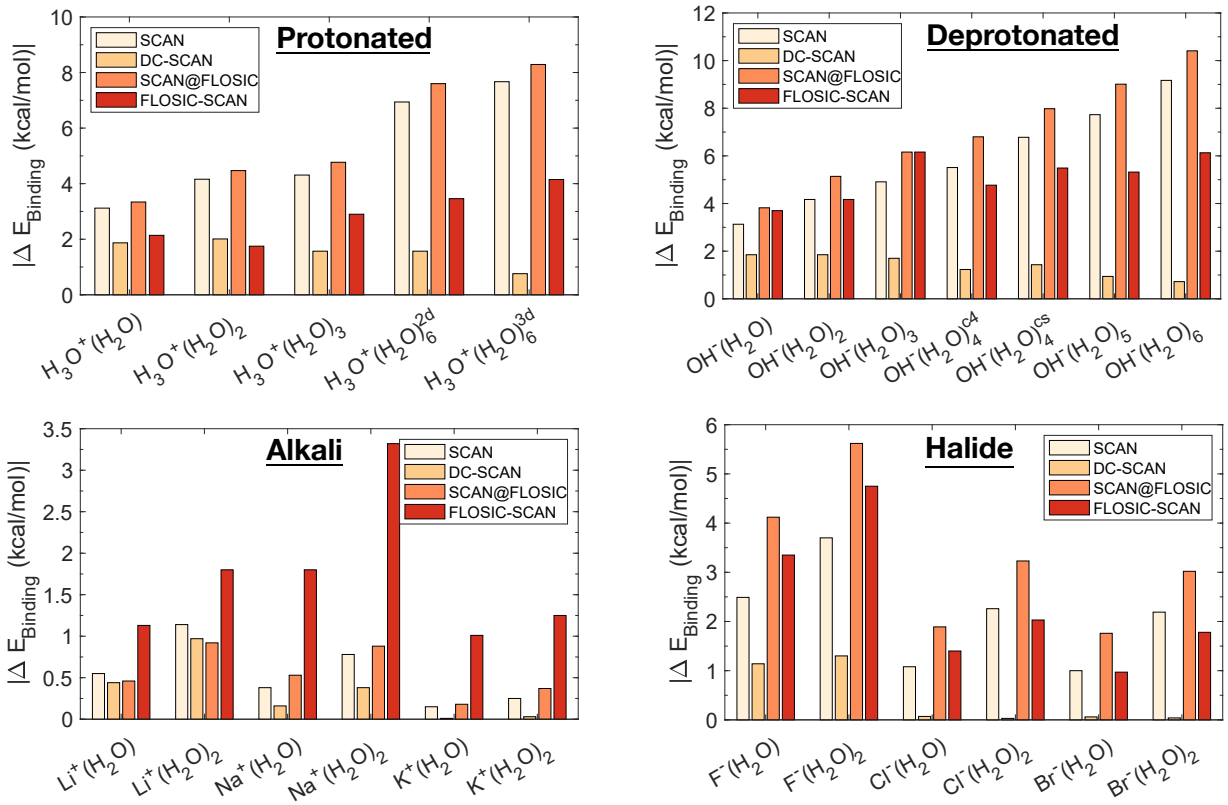

Figure 8: Error in the binding energies of the a) Protonated water clusters b) Deprotonated water clusters c) Alkali metal clusters d) Halide-water clusters using SCAN, DC-SCAN, SCAN@FLOSIC and FLOSIC-SCAN with respect to the benchmark CCSD(T) values. The SCAN@FLOSIC and FLOSIC-SCAN values are obtained from ref 49

be compared with the corresponding values of 5.10 kcal/mol and 1.39 kcal/mol obtained with FLOSIC-SCAN and DC-SCAN, respectively. It should be noted that, in the case of the deprotonated water clusters, the density-driven errors associated with the monomers partially compensate the density-driven errors associated with the entire cluster, which is the reason why the difference between the SCAN and FLOSIC-SCAN energies is not significant.

Since the alkali metal–water clusters are primarily stabilized by ion–dipole interactions, their binding energy decreases with increasing atomic number because the increasing radius of the alkali-metal ion reduces the ionic potential. Furthermore, increasing the cluster size reduces the ion-dipole interaction which, consequently, results in a decrease of the error per monomer. As already noted in ref 49, FLOSIC-SCAN displays a MUE of 1.72 kcal/mol

alkali metal–water clusters, which is larger than that associated with SCAN (0.54 kcal/mol). On the other hand, DC-SCAN improves significantly on SCAN, displaying a MUE of 0.33 kcal/mol.

Finally, similar to the other-aqueous ionic clusters, DC-SCAN performs well for the halide-water clusters, displaying MUE of 0.44 kcal/mol. Relatively large errors are instead associated with SCAN and FLOSIC-SCAN, which lead to MUEs of 2.12 kcal/mol and 2.38 kcal/mol, respectively.

Overall, the comparisons of Figure 8 demonstrate that DC-SCAN consistently improves on SCAN for all the four classes of hydrogen-bonded systems examined in this study, while the performance of FLOSIC-SCAN is somewhat erratic.

3.7 Many-body decomposition analysis

The many-body expansion (MBE) expresses the total energy, $E_{\rm tot}$, of an N-body system as the sum of individual n-body energy contributions, ϵ_{nB}, where $n \leq N$,[105]

$$E_{\rm tot}(r_1,..,r_N) = \sum_{i=1}^{N} \epsilon_{\rm 1B}(r_i) + \sum_{i<j}^{N} \epsilon_{\rm 2B}(r_i,r_j) + \sum_{i<j<k}^{N} \epsilon_{\rm 3B}(r_i,r_j,r_k) + ... + \epsilon_{\rm nB}(r_1,..,r_N) \quad (12)$$

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

$$\epsilon_{\rm nB} = \epsilon_{\rm n}(1,...,n) - \sum_{i=1}^{N} \epsilon_{\rm 1B}(r_i) - \sum_{i<j}^{N} \epsilon_{\rm 2B}(r_i,r_j) - \sum_{i<j<k<...<n-1}^{N} \epsilon_{\rm (n-1)B}(r_i,r_j,..r_{n-1}). \quad (13)$$

To investigate the impact of the density correction on each individual n-body contribution to the interaction energies of different ion–water systems, many-body decomposition analyses were carried out for two protonated water clusters, $H_3O^+(H_2O)_6^{(3d)}$ and $H_3O^+(H_2O)_6^{(2d)}$, and two deprotonated water clusters, $OH^-(H_2O)_5$ and $OH^-(H_2O)_6$. The errors associated with the SCAN and DC-SCAN nB energies were calculated relative to the CCSD(T)-F12b

reference values and are shown in Figure 9.

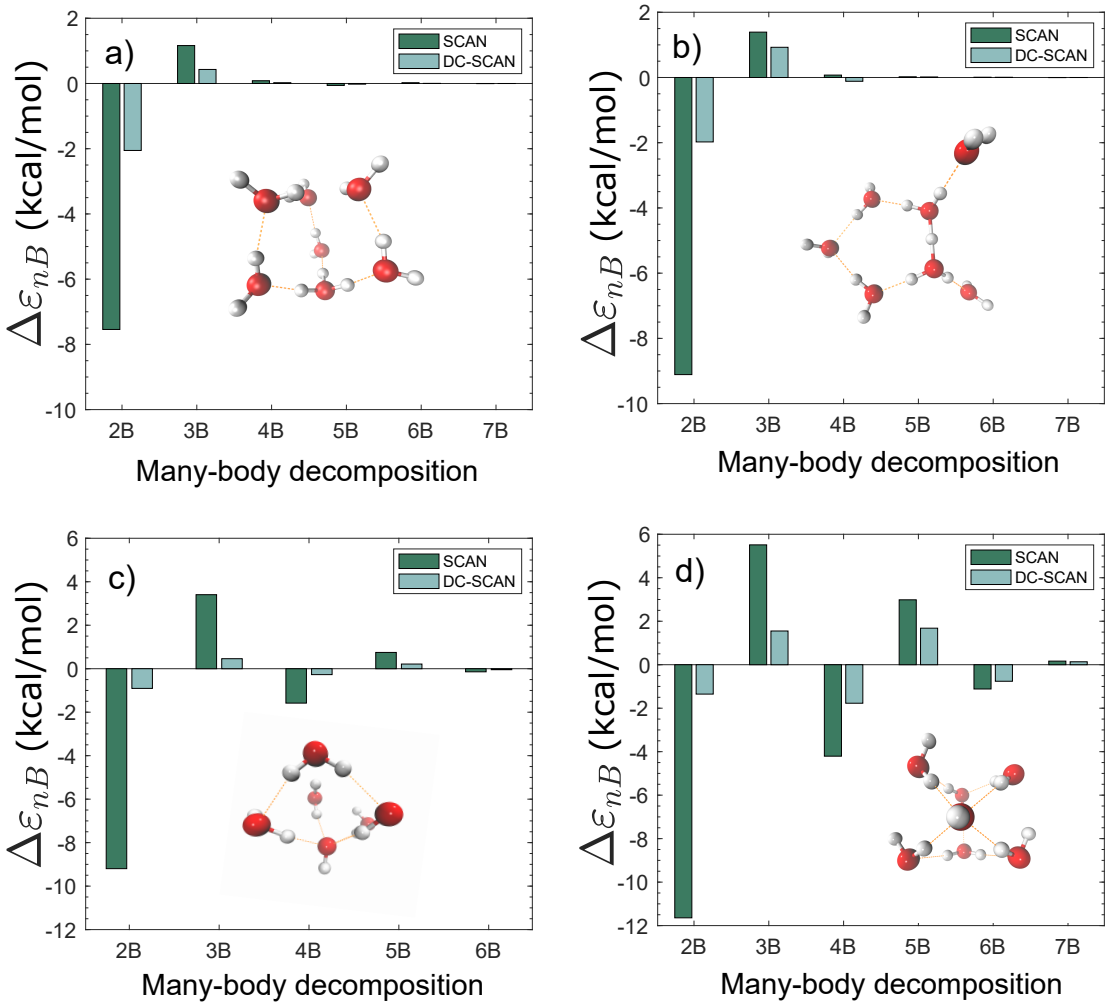

Figure 9: Errors relative to CCSD(T)-F12b reference values for each nB energy contribution to the interaction energies calculated with SCAN and DC-SCAN for the a) $H_3O^+(H_2O)_6^{(3d)}$ b) $H_3O^+(H_2O)_6^{(2d)}$ c) $OH^-(H_2O)_5$ d) $OH^-(H_2O)_6$

For the protonated water clusters, SCAN displays large 2B errors of -7.54 kcal/mol and -9.11 kcal/mol for $H_3O^+(H_2O)_6^{(2d)}$ and $H_3O^+(H_2O)_6^{(3d)}$, respectively, which confirms the tendency of SCAN to overbind aqueous clusters.[47,62] The error of the 2B energy reduces to only -2.05 and -1.97 kcal/mol in the corresponding DC-SCAN calculations. Figures 9a,b demonstrate that the impact of the density correction is effectively negligible for nB energies with $n > 2$ for the protonated water clusters. A similar trend for the 2B energies is also observed

for $OH^-(H_2O)_5$ and $OH^-(H_2O)_6$. Unlike the neutral[62] and protonated water clusters, SCAN is associated with large 3B and 4B errors for the deprotonated water clusters.

4 CONCLUSIONS

It is known that density-driven errors in DFT calculations are prominent for GGA and meta-GGA functionals, and the SCAN functional also makes such errors, but its density delocalization errors are less than those of PBE or LDA. Use of the Hartree-Fock density in the density-corrected DFT formalism (DC-DFT) has been shown to mitigate density-driven errors as the density-correction elevates the accuracy of the SCAN functional towards the CCSD(T) limit, not only for neutral water but also (to a lesser extent) for all the hydrated-ion systems.

The performance of density-corrected SCAN (DC-SCAN) is independent of the system size as it reproduces the CCSD(T) interaction and binding energies for water dimer to decamer with minimal loss of accuracy. Similarly, DC-SCAN has been succesful to reproduce the CCSD(T) binding energies for various protonated and deprotonated water clusters.

The inability of semi-local density functional approximations to bind a full extra electron in the smallest anionic water clusters is well known, as the excess electron leaks towards the continuum due to the large density-driven errors, give rise to an unstable SCF solution. The localized Hartree-Fock density minimizes this problem by binding the electron fully as the deprotonated-water clusters move towards convergence with the addition of diffuse basis function. The density-driven errors of water dimers using the SCAN functional can be visualised in terms of the density-deviation of the self-consistent SCAN density with respect to the density generated by the HF orbitals. These density-deviations can quantify the extent of electron delocalization as well as demonstrate the quality of the density generated by different density functionals.

For accuracy in water, DC-SCAN outperforms FLOSIC, which is an orbital-by-orbital

SIE removal method. DC-SCAN consistently improves the accuracy of the SCAN functional for all the protonated, deprotonated, alkali-metal and halide water clusters, whereas the performance of FLOSIC-SCAN have been found to be somewhat erratic, as it improves the accuracy of SCAN for all but the hydrated alkali-metal clusters.

In this study, we have demonstrated that density-corrected SCAN (DC-SCAN) effectively removes density-driven errors from the water 2B energies, which brings both binding and interaction energies of different water clusters very close to reference values calculated at the CCSD(T) level of theory. The density-driven errors in neutral and protonated water clusters have been found mostly in the 2B energies, which can be successfully mitigated by the density-correction. On the other hand the notorious deprotonated-water clusters give rise to large 3B errors along with the 2B due to their large density-driven errors, which can also be successfully minimized by DC-SCAN. It is worth mentioning that DC-SCAN is not a panacea. It can only alleviate energy errors driven by density errors, and is unable to elevate the accuracy when functional-driven error is dominant, *e.g.*, the binding energy of $H_3O^+(H_2O)_6(OH)^-$.

What's so right about the Hartree-Fock electron density? In neutral and positively charged atoms or small molecules with integer electron numbers, the Hartree-Fock density of an atom or small molecule is easily outperformed by non-empirical or few-parameter empirical functionals, whose accuracy improves from LDA to PBE to SCAN to the PBE0 hybrid.[65,95,96] The Hartree-Fock density of an atom is a best-performer only in the low-density tail, where the FLOSIC density is still comparably accurate. But more relevant to chemistry is how the density is shared among the atoms or molecules of the system. Here, for weakly-bound fragments, the Hartree-Fock density shines.

We believe that this has not been shown before, although recent work[106] on molecules has identified localizing steps[66,98] in the exact (but not in semi-local approximate) Kohn-Sham potentials, and some approximations[107] in condensed matter physics (motivated in part by refs 42 and 66) have been aimed at improving the partitioning.

While LDA seems to be the conventional extreme of density delocalization error, and HF the conventional extreme of (usually much smaller) density localization error, these two densities do not quite yield the conventional range of density-driven error in energy differences.

In view of the fact that PZ-SIC typically[97] eliminates fractional charges on separated atoms, it is surprising that it does not better partition the density in water clusters and some other systems. This might be related to the fact that the original PZ-SIC is more accurate for atoms than for molecules. Perhaps future refinements of PZ-SIC will be more successful in this regard.

Correctible density-driven overbinding by SCAN is not limited to water, and in particular it can be expected in other hydrogen-bonded systems. In this context, the hydrogen-bonded ionic dimers of the HB15 and the small- and medium-sized hydrogen-bonded systems of the HB49 datasets are overbound[108] in SCAN on average by 0.6 and 0.5 kcal/mol, respectively.

Acknowledgements

This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, through grants no. DE-SC0019490 (F.P.) and no. DE-SC0018331 (J.P.P. and C.S.). The work of J.P. and P.B. was also supported in part by the U.S. National Science Foundation under grant no. DMR-1939528. This research used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1548562, and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC) Supercomputer Center (SDSC). This research also includes calculations carried out on HPC resources supported in part by the National Science Foundation through major research instrumentation

grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. J.P.P. thanks Aaron Kaplan for finding ref 108.

References

(1) Marcus, Y. *Ion Solvation*; Wiley, 1985.

(2) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133–7155.

(3) Chalikian, T. V.; Völker, J.; Plum, G. E.; Breslauer, K. J. A More Unified Picture for the Thermodynamics of Nucleic Acid Duplex Melting: A Characterization by Calorimetric and Volumetric Techniques. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7853–7858.

(4) Rupley, J. A.; Careri, G. Protein Hydration and Function. *Adv. Protein Chem.* **1991**, *41*, 37–172.

(5) Sussman, F.; Weinstein, H. On the Ion Selectivity in Ca-Binding Proteins: The Cyclo (-L-Pro-Gly-)3 Peptide as a Model. *Proc. Natl. Acad. Sci.* **1989**, *86*, 7880–7884.

(6) Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833–835.

(7) Jordan, P. C. Ion-Water and Ion-Polypeptide Correlations in a Gramicidin-Like Channel. A Molecular Dynamics Study. *Biophys. J.* **1990**, *58*, 1133–1156.

(8) Katz, B. *Nerve, Muscle, and Synapse*; McGraw-Hill New York, 1966.

(9) Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M. Metal Ions in Biological Catalysis: From Enzyme Databases to General Principles. *J. Biol. Inorg. Chem.* **2008**, *13*, 1205–1218.

(10) Berthod, A.; Ruiz-Angel, M.; Carda-Broch, S. Ionic Liquids in Separation Techniques. *J. Chromatogr. A* **2008**, *1184*, 6–18.

(11) Son, D. H.; Hughes, S. M.; Yin, Y.; Alivisatos, A. P. Cation Exchange Reactions in Ionic Nanocrystals. *Science* **2004**, *306*, 1009–1012.

(12) Jencks, W. P. General Acid-Base Catalysis of Complex Reactions in Water. *Chem. Rev.* **1972**, *72*, 705–718.

(13) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.

(14) Kohn, W. Nobel Lecture: Electronic Structure of Matter—Wave Functions and Density Functionals. *Rev. Mod. Phys.* **1999**, *71*, 1253.

(15) 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.

(16) Agmon, N.; Bakker, H. J.; Campen, R. K.; Henchman, R. H.; Pohl, P.; Roke, S.; Thämer, M.; Hassanali, A. Protons and Hydroxide Ions in Aqueous Systems. *Chem. Rev.* **2016**, *116*, 7642–7672.

(17) Van Der Vegt, N. F.; Haldrup, K.; Roke, S.; Zheng, J.; Lund, M.; Bakker, H. J. Water-Mediated Ion Pairing: Occurrence and Relevance. *Chem. Rev.* **2016**, *116*, 7626–7641.

(18) Castleman Jr, A. W.; Keesee, R. G. Ionic Clusters. *Chem. Rev.* **1986**, *86*, 589–618.

(19) Ohtaki, H.; Radnai, T. Structure and Dynamics of Hydrated Ions. *Chem. Rev.* **1993**, *93*, 1157–1204.

(20) Bieske, E.; Maier, J. Spectroscopic Studies of Ionic Complexes and Clusters. *Chem. Rev.* **1993**, *93*, 2603–2621.

(21) Bieske, E. J.; Dopfer, O. High-Resolution Spectroscopy of Cluster Ions. *Chem. Rev.* **2000**, *100*, 3963–3998.

(22) Robertson, W. H.; Johnson, M. A. Molecular Aspects of Halide Ion Hydration: The Cluster Approach. *Annu. Rev. Phys. Chem.* **2003**, *54*, 173–213.

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

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

(25) 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.

(26) Sim, F.; St. Amant, A.; Papai, I.; Salahub, D. R. Gaussian Density Functional Calculations on Hydrogen-Bonded Systems. *J. Am. Chem. Soc.* **1992**, *114*, 4391–4400.

(27) Tuckerman, M. E. Ab Initio Molecular Dynamics: Basic Concepts, Current Trends and Novel Applications. *J. Condens. Matter Phys.* **2002**, *14*, R1297.

(28) Santra, B.; Michaelides, A.; Scheffler, M. On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters: Benchmarks Approaching the Complete Basis Set Limit. *J. Chem. Phys.* **2007**, *127*, 184104.

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

(30) 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.

(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(32) 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.

(33) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. Towards an Assessment of the Accuracy of Density Functional Theory for First Principles Simulations of Water. *J. Chem. Phys.* **2004**, *120*, 300–311.

(34) 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.

(35) Gillan, M. J.; Alfe, D.; Michaelides, A. Perspective: How Good is DFT for Water? *J. Chem. Phys.* **2016**, *144*, 130901.

(36) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation. *Phys. Rev. Lett.* **1999**, *82*, 2544.

(37) Adamo, C.; Ernzerhof, M.; Scuseria, G. E. The Meta-GGA Functional: Thermochemistry with a Kinetic Energy Density Dependent Exchange-Correlation Functional. *J. Chem. Phys.* **2000**, *112*, 2643–2649.

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

(39) Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P. Accurate First-Principles Structures and Energies of Diversely Bonded Systems from an Efficient Density Functional. *Nat. Chem.* **2016**, *8*, 831.

(40) Chen, M.; Ko, H.-Y.; Remsing, R. C.; Andrade, M. F. C.; 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. USA* **2017**, *114*, 10846–10851.

(41) Zheng, L.; Chen, M.; Sun, Z.; Ko, H.-Y.; Santra, B.; Dhuvad, P.; Wu, X. Structural, Electronic, with Dynamical Properties of Liquid Water by Ab Initio Molecular Dynamics Based on SCAN Functional Within the Canonical Ensemble. *J. Chem. Phys.* **2018**, *148*, 164505.

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

(43) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Development of Exchange-Correlation Functionals with Minimal Many-Electron Self-Interaction Error. *J. Chem. Phys.* **2007**, *126*, 191109.

(44) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.

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

(46) Li, C.; Zheng, X.; Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Local Scaling Correction for Reducing Delocalization Error in Density Functional Approximations. *Phys. Rev. Lett.* **2015**, *114*, 053001.

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

(48) 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.

(49) Wagle, K.; Santra, B.; Bhattarai, P.; Shahi, C.; Pederson, M. R.; Jackson, K. A.; Perdew, J. P. Self-Interaction Correction in Water–Ion Clusters. *J. Chem. Phys.* **2021**, *154*, 094302.

(50) Pederson, M. R.; Ruzsinszky, A.; Perdew, J. P. Communication: Self-Interaction Correction with Unitary Invariance in Density Functional Theory. *J. Chem. Phys.* **2014**, *140*, 121103.

(51) Gordon, R. G.; Kim, Y. S. Theory for the Forces Between Closed-Shell Atoms and Molecules. *J. Chem. Phys.* **1972**, *56*, 3122–3133.

(52) 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.

(53) 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.

(54) Janesko, B. G.; Scuseria, G. E. Hartree–Fock Orbitals Significantly Improve the Reaction Barrier Heights Predicted by Semilocal Density Functionals. *J. Chem. Phys.* **2008**, *128*, 244112.

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

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

(58) Jana, S.; Patra, A.; Śmiga, S.; Constantin, L. A.; Samal, P. Insights from the Density Functional Performance of Water and Water–Solid Interactions: SCAN in Relation to Other Meta-GGAs. *J. Chem. Phys.* **2020**, *153*, 214116.

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

(60) Rana, B.; Coons, M.; Herbert, J. Detection and correction of delocalization errors for electron- and hole-polarons using density-corrected DFT. *ChemRxiv* **2022**,

(61) 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.

(62) Dasgupta, S.; Lambros, E.; Perdew, J. P.; 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.

(63) 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.

(64) Gill, P. M.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. An Investigation of the Performance of a Hybrid of Hartree-Fock and Density Functional Theory. *Int. J. Quantum Chem.* **1992**, *44*, 319–331.

(65) Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density

Functional Theory Is Straying from the Path Toward the Exact Functional. *Science* **2017**, *355*, 49–52.

(66) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz Jr, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691.

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

(68) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. Density Functional Theory Embedding for Correlated Wavefunctions: Improved Methods for Open-Shell Systems and Transition Metal Complexes. *J. Chem. Phys.* **2012**, *137*, 224113.

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

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

(71) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Density Functionals That Are One- and Two- Are Not Always Many-Electron Self-Interaction-Free, as Shown for H_2^+, He_2+, LiH^+, and Ne_2^+. *J. Chem. Phys.* **2007**, *126*, 104102.

(72) Hait, D.; Head-Gordon, M. Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number. *J. Phys. Chem. Lett.* **2018**, *9*, 6280–6288.

(73) Lee, D.; Furche, F.; Burke, K. Accuracy of Electron Affinities of Atoms in Approximate Density Functional Theory. *J. Phys. Chem. Lett.* **2010**, *1*, 2124–2129.

(74) 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.

(75) 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.

(76) 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.

(77) Murray, C. W.; Handy, N. C.; Laming, G. J. Quadrature Schemes for Integrals of Density Functional Theory. *Mol. Phys.* **1993**, *78*, 997–1014.

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

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

(80) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J.phys.Chem Lett.* **2020**, *11*, 8208–8215.

(81) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(82) Zhong, S.; Barnes, E. C.; Petersson, G. A. Uniformly Convergent n-tuple-ζ Augmented Polarized (nZaP) Basis Sets for Complete Basis Set Extrapolations. I. Self-Consistent Field Energies. *J. Chem. Phys.* **2008**, *129*, 184116.

(83) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(84) Yousaf, K. E.; Peterson, K. A. Optimized Auxiliary Basis Sets for Explicitly Correlated Methods. *J. Chem. Phys.* **2008**, *129*, 184108.

(85) Yousaf, K. E.; Peterson, K. A. Optimized Complementary Auxiliary Basis Sets for Explicitly Correlated Methods: aug-cc-pVnZ Orbital Basis Sets. *Chem. Phys. Lett.* **2009**, *476*, 303–307.

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

(87) 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.

(88) Bryantsev, V. S.; Diallo, M. S.; Van Duin, A. C.; Goddard III, W. A. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016–1026.

(89) Manna, D.; Kesharwani, M. K.; Sylvetsky, N.; Martin, J. M. Conventional and Explicitly Correlated Ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets. *J. Chem. Theory Comput.* **2017**, *13*, 3136–3152.

(90) Gill, P. M.; Johnson, B. G.; Pople, J. A. A Standard Grid for Density Functional Calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.

(91) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černỳ, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. 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.

(92) 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.

(93) Brandenburg, J.; Bates, J.; Sun, J.; Perdew, J. Benchmark Tests of a Strongly Constrained Semilocal Functional With a Long-Range Dispersion Correction. *Phys. Rev. B* **2016**, *94*, 115144.

(94) Scott, D. W. Scott's rule. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 497–502.

(95) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.

(96) Hait, D.; Head-Gordon, M. How Accurate Are Static Polarizability Predictions From Density Functional Theory? An Assessment Over 132 Species at Equilibrium Geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800–19810.

(97) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. Spurious Fractional Charge on Dissociated Atoms: Pervasive and Resilient Self-Interaction Error of Common Density Functionals. *J. Chem. Phys.* **2006**, *125*, 194112.

(98) Perdew, J. P. *Density Functional Methods in Physics*; Springer, 1985; pp 265–308.

(99) Sim, E.; Song, S.; Burke, K. Quantifying Density Errors in DFT. *J. Phys. Chem. Lett.* **2018**, *9*, 6385–6392.

(100) Nam, S.; Song, S.; Sim, E.; Burke, K. Measuring Density-Driven Errors Using Kohn–Sham Inversion. *J. Chem. Theory Comput.* **2020**, *16*, 5014–5023.

(101) Sherrill, C. D.; Krylov, A. I.; Byrd, E. F.; Head-Gordon, M. Energies and Analytic Gradients for a Coupled-Cluster Doubles Model Using Variational Brueckner Orbitals: Application to Symmetry Breaking in O_4^+. *J. Chem. Phys.* **1998**, *109*, 4171–4181.

(102) Genova, A.; Ceresoli, D.; Pavanello, M. Avoiding Fractional Electrons in Subsystem DFT Based Ab-Initio Molecular Dynamics Yields Accurate Models for Liquid Water and Solvated OH Radical. *J. Chem. Phys.* **2016**, *144*, 234105.

(103) Withanage, K.; Sharkas, K.; Johnson, J.; Perdew, J. P.; Peralta, J.; Jackson, K. Fermi-Löwdin Orbital Self-Interaction Correction of Adsorption Energies of Transition Metal Ions. *J. Chem. Phys.* **2022**, under review.

(104) Shahi, C.; Bhetwal, P.; Perdew, J. P.; Song, S.; Sim, E.; Burke, K. **2022**, to be submitted.

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

(106) Kraisler, E.; Hodgson, M. J. P.; Gross, E. K. U. From Kohn–Sham to Many-Electron Energies via Step Structures in the Exchange-Correlation Potential. *J.Chem. Theory and Comput.* **2021**, *17*, 1390–1407.

(107) Ferretti, A.; Dabo, I.; Cococcioni, M.; Marzari, N. Bridging Density-Functional and Many-Body Perturbation Theory: Orbital-Density Dependence in Electronic-Structure Functionals. *Phys. Rev. B* **2014**, *89*, 195134.

(108) 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.

TOC graphic