Elevating Density Functional Theory to Chemical Accuracy for Water Simulations through a Density-Corrected Many-Body Formalism

Saswata Dasgupta,[1,6] Eleftherios Lambros,[1,6] John P. Perdew[2,3] and Francesco Paesani[1,4,5]

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

[2]*Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, United States*

[3]*Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States*

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

[5]*San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, United States*

[6]*Contributed equally*

Kohn-Sham density functional theory (DFT) has been extensively used to model the properties of water. Albeit maintaining a good balance between accuracy and efficiency, no density functional has so far achieved the degree of accuracy necessary to correctly predict the properties of water across the entire phase diagram. The recent development of the strongly constrained and appropriately normed (SCAN) functional has renewed the interest in *ab initio* simulations of liquid water, yielding promising results that are, however, still unable to reproduce all the experimental data. Here, we present density-corrected SCAN (DC-SCAN)

calculations for water which, minimizing density-driven errors, elevate the accuracy of the SCAN functional to that of coupled cluster theory, the "gold standard" for chemical accuracy. Building upon the accuracy and efficiency of DC-SCAN within a many-body formalism, we introduce a data-driven many-body potential energy function, the MB-SCAN(DC) PEF, that is able to quantitatively reproduce coupled cluster reference values for interaction, binding, and individual many-body energies of water clusters. Importantly, the properties of liquid water calculated from molecular dynamics simulations carried out with the MB-SCAN(DC) PEF are found to be in excellent agreement with the experimental data, which thus demonstrates that MB-SCAN(DC) is effectively the first DFT-based model that correctly describes water from the gas to the condensed phase. Since the many-body formalism adopted by the present MB-SCAN(DC) PEF for water is general, we believe it can open the door to the routine development of data-driven many-body PEFs for predictive simulations of generic (small) molecules in the gas, liquid, and solid phases.

Its anomalous behavior[1] and importance to life[2] make water one of the most studied chemical compounds. Among its many unique properties is the high value of the heat capacity which allows water to resist sudden temperature changes, thus permitting living organisms to survive without experiencing significant temperature fluctuations.[3] In addition, the dynamic nature of the water hydrogen-bond network plays a central role in several fundamental processes, including transport and diffusion in bulk solutions and at interfaces, and hydration of hydrophobic and hydrophilic solutes.[4] For example, protein folding is thought to be driven by the hydrophobic effect.[5] Finally, countless chemical reactions involving charged species take place efficiently in liquid water due to

its high dielectric constant.[6–11] It thus not surprising that a myriad of simulation studies have been devoted to developing a fundamental understanding of both chemical and physical properties of water in different environments and under different thermodynamic conditions.[12–14]

Density functional theory (DFT)[15,16] is one the most important tools available to computational chemists and physicists for *ab initio* simulations of molecular systems in the condensed phase since it offers a good balance between accuracy and computational cost.[17,18] However, as discussed in the "Methods" section, the accuracy of a DFT calculation depends upon the accuracy of the underlying exchange-correlation (XC) functional, which allows for recasting the many-body electronic structure problem into a (self-consistent) single-particle problem formulated in terms of the Kohn-Sham equations.[19] For example, the simplest XC functional, which was built upon the local density approximation (LDA),[19–21] was shown to correctly predict the structure of metallic crystals under pressure,[22–24] but was unable to fulfill its promises for water simulations, overestimating the strength of the hydrogen bonds and, consequently, predicting a too packed and overstructured liquid phase.[25,26] These limitations hindered the ability of the LDA functional to describe the properties of water, even qualitatively.

Climbing the Jacob's ladder of DFT approximations,[27] the next generation of XC functionals, which were developed within the generalized gradient approximation (GGA),[28–30] dominated the scene of *ab initio* simulations of water for a long time, owing to their higher accuracy compared to LDA and affordable computational cost. Initial successes of the GGA functionals included relatively accurate binding energies for various water clusters and a reasonable description of the

structure of liquid water.[25,31–33] However, it became soon evident that serendipitous error cancellation was the primary reason behind the apparent accuracy of GGA simulations of liquid water, making the predictive power more accidental than consistent.[34–37] For example, it was found that GGA functionals generally underestimate the density of liquid water, while predicting denser ice phases.[38]

The third rung of the Jacob's ladder of DFT approximations includes meta-GGA functionals[39,40] that perform significantly better than both LDA and GGA functionals due to the inclusion of the kinetic energy density. Among them, the strongly constrained and appropriately normed (SCAN) functional has gained particular attention because it satisfies all 17 known exact constraints that can be satisfied by a meta-GGA functional.[41] Without being fitted to any bonded system, SCAN was shown to enable accurate predictions for various properties of molecules and solids.[42] In particular, for molecular dynamics (MD) simulations of liquid water, SCAN was found to outperform its predecessor GGA functionals.[43,44] Importantly, accounting for intermediate-range van der Waals interactions, the SCAN functional allows for a more accurate description of the energy differences among water clusters and ice phases,[42,44] while, when used in MD simulations, it predicts a density of liquid water which is appreciably closer to the experimental value compared to values obtained with GGA functionals.[43]

Despite its relatively higher accuracy, SCAN, as all GGA and meta-GGA functionals, is still prone to density-driven errors (defined in the "Methods" section), including self-interaction[45] and delocalization errors.[46–49] It was shown that self-interaction errors in the SCAN functional primar-

ily affect 2-body contributions (defined in the "Methods" section) to the interaction energies of water clusters.[50] On the other hand, inclusion of a fraction of Hartree-Fock exchange (also known as exact exchange) in a SCAN hybrid was found to partially reduce density-driven errors in calculations for various water systems.[51,52] However, it was found that increasing the fraction of Hartree-Fock exchange beyond 15% did not improve the accuracy of hybrid SCAN functionals, progressively shifting the structure of liquid water towards that of ice.[51] A systematic analysis of hybrid SCAN functionals with varying fractions of Hartree-Fock exchange demonstrated the inability of these functionals to accurately represent 2-body interactions between water molecules, with errors up to $\sim$5 kcal/mol for the water hexamer relative to reference values calculated using coupled cluster theory with single, double, and perturbative triple excitations, i.e., CCSD(T), in the complete basis set (CBS) limit,[51] the 'gold standard" method for molecular interactions.[53] In this context, a neural-network potential, NNP-SCAN0, was recently trained on a modified SCAN0 functional that incorporates 10% Hartree-Fock exchange.[52] (It is worth noting that, in its original formulation, the SCAN0 functional mixes 25% Hartree-Fock exchange with 75% SCAN exchange.[54]) Despite providing better agreement with experimental data than SCAN for several properties of liquid water measured at ambient conditions, this improved agreement was achieved by actually performing the NNP-SCAN0 simulations at 330 K.[52]

While all previous studies suggest that SCAN is overall one of the most accurate XC functionals, they also indicate that any further improvement of the accuracy of DFT models for water requires removing, at least partially, the associated density-driven errors. To this end, we introduce here a data-driven many-body potential energy function (PEF) for water, MB-SCAN(DC),

which is rigorously derived within a many-body formalism applied to density-corrected SCAN (DC-SCAN) data for individual many-body contributions to the interaction energies between water molecules. Density-corrected DFT (DC-DFT),[55–62] where the Hartree-Fock density is used instead of the Kohn-Sham density, is known to mitigate density-driven errors in GGA and meta-GGA functionals. Here we show that both binding and interaction energies calculated with the DC-SCAN functional for various water clusters are close to the CCSD(T)/CBS reference values, with DC-SCAN correctly reproducing each individual many-body contribution to the interaction energies. Importantly, we demonstrate that the MB-SCAN(DC) PEF preserves the accuracy of DC-SCAN and enables simulations of liquid water with significantly higher accuracy than all previous DFT-based models reported in the literature (including both *ab initio* and neural network models), predicting structural, thermodynamic, and dynamical properties in quantitative agreement with experiment.

Results

2-body interactions in water. Our analysis of the ability of the SCAN functional to represent the interactions between water molecules begins with the comparison in Fig. 1 between the total 2-body (2B) energies calculated for the low-energy isomers of the water hexamer using the (self-consistent) SCAN and SCAN0 functionals, and the corresponding (density-corrected) DC-SCAN and DC-SCAN0 functionals. Also shown for reference are the CCSD(T)/CBS values reported in ref. 63. It should be noted that the hexamer holds a special space along the path that connects individual water molecules in the gas phase to liquid water since it is the smallest water cluster for

which the low-energy isomers are characterized by three-dimensional arrangements that are reminiscent of the three-dimensional structure of the hydrogen-bond network found in the liquid phase. In addition, the large number of low-energy isomers makes the hexamer cluster the prototypical system to assess the ability of different water models to correctly reproduce many-body interactions in water.[14] Fig. 1 shows that the SCAN functional displays fairly large errors compared to the reference values, with a maximum unsigned error (MUE) of 4.59 kcal/mol. In contrast, DC-SCAN predicts 2-body energies that are in quantitative agreement with the CCSD(T)/CBS values, resulting in a MUE of only 0.08 kcal/mol. By effectively eliminating the errors in the representation of 2-body interactions, the application of the density correction thus addresses the main shortcoming of the SCAN functional applied to water.[50] Fig. 1 also shows that SCAN0, the hybrid variant of

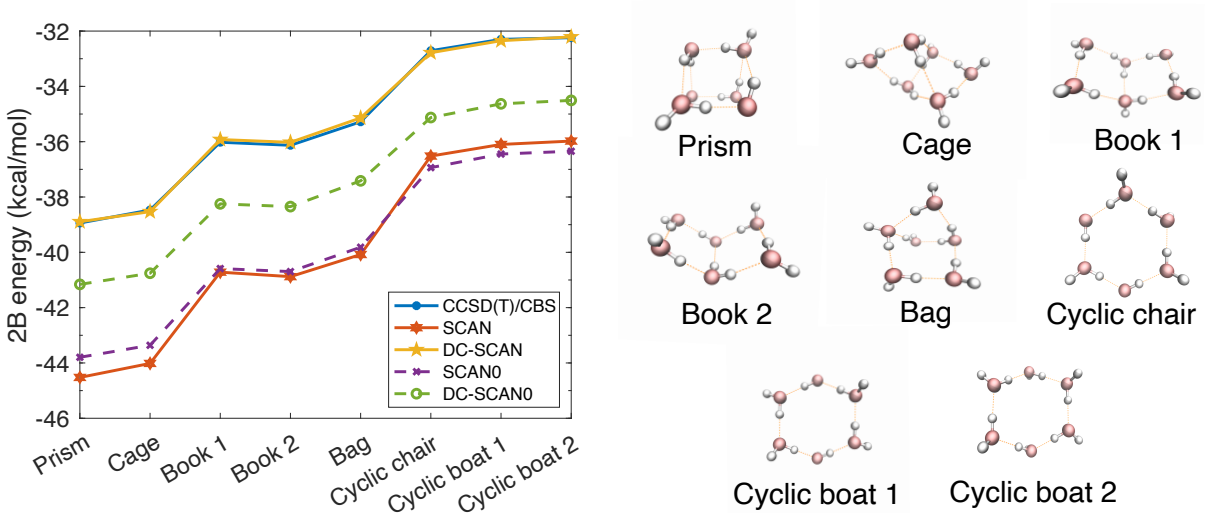

Figure 1: 2-body energies calculated for the first eight low-energy isomers of the water hexamer using SCAN, DC-SCAN, SCAN0 (with 25% exact exchange), and DC-SCAN0, along with the corresponding CCSD(T)/CBS reference values from ref. 63.

SCAN with a 25% fraction of Hartree-Fock exchange, only provides a minor improvement in the representation of the 2-body energies, resulting in a MUE of 4.48 kcal/mol. It should be noted that SCAN0 provides a slightly more accurate description of the three-dimensional isomers (i.e., prism and cage isomers) but a worse description of the planar isomers (i.e., cyclic isomers) compared to SCAN. Importantly, the density correction applied to SCAN0 does not result in a similarly dramatic improvement as found for SCAN, with DC-SCAN0 still displaying a relatively large MUE of 2.26 kcal/mol.

The analysis of the effects associated with various dispersion corrections, which is reported in the Supporting Information (Fig. S1), indicates that the addition of any form of dispersion energy worsens the accuracy of both SCAN and DC-SCAN. Specifically, all the dispersion-corrected SCAN functionals considered in our analysis, SCAN-D3(0), SCAN-D3(BJ), and SCAN-VV10, are found to overbind the hexamer isomers, which results in larger deviations from the CCSD(T)/CBS values compared to their dispersion-free counterparts. It is also worth noting that neither DC-SCAN+dispersion nor DC-SCAN0 perform as well as DC-SCAN. This suggests that the addition of the dispersion correction and/or a fraction of Hartree-Fock exchange actually worsens the functional-driven error of SCAN for water.

To further investigate the impact of the density correction on the energetics of various water systems, in Fig. 2 we analyze the interaction energies of dimers extracted from a classical MD simulation of liquid water carried out in the isobaric-isothermal (NPT) ensemble at ambient conditions using the MB-pol PEF.[64–66] (The interaction energy is the binding energy without its 1-body con-

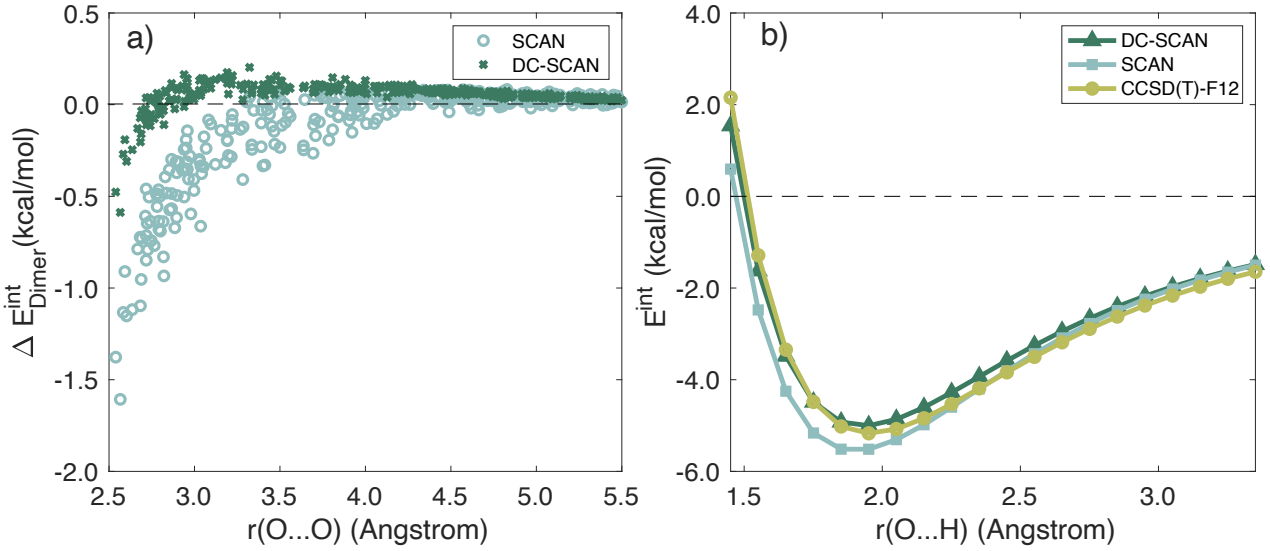

Figure 2: a) Errors in 2-body energies calculated with SCAN and DC-SCAN relative to CCSD(T)-F12b values for dimers with an oxygen-oxygen distance within 5.5 Å which were extracted from an NPT simulation of liquid water carried out with MB-pol[64–66] at ambient conditions. b) CCSD(T)-F12b, SCAN and DC-SCAN interaction energies calculated for an unrelaxed scan of the water dimer along the O$\cdots$H distance.

tribution.) For this analysis, we consider dimers with an oxygen-oxygen (O$\cdots$O) distance within 5.5 Å, which approximately corresponds to the radius of the first two solvation shells in liquid water.[66–68] It should be noted that, by definition, the interaction energy of a water dimer exactly corresponds to the associated 2-body energy. Fig. 2a shows the errors, ΔE, in 2-body energies calculated with SCAN and DC-SCAN relative to the corresponding reference values calculated at the CCSD(T)-F12b level of theory. As expected, DC-SCAN exhibits significantly smaller errors compared to SCAN for all dimers, independent of the O$\cdots$O distance. Specifically, the maximum

error associated with DC-SCAN is -0.47 kcal/mol, which must be compared with a maximum error of -1.38 kcal/mol calculated with SCAN. It is also important to analyze the errors as a function of the O$\cdots$O distance since they directly affect the ability of the SCAN and DC-SCAN functionals to correctly predict the cohesive energy, and thus the structure, of liquid water. Fig. 2a shows that the 2-body energies calculated with SCAN only start to approach the CCSD(T)-F12b values at $\sim$4.5 Å, with a MUE of 0.25 kcal/mol associated with dimers with an O$\cdots$O distance up to 4.5 Å, and a MUE of 0.16 kcal/mol for all dimers up to an O$\cdots$O distance of 5.5 Å. In contrast, the 2-body energies calculated with DC-SCAN converge to the CCSD(T)-F12b values at 3.5Å, with a MUE of 0.09 kcal/mol obtained for dimers with an O$\cdots$O distance up to 3.5 Å, which decreases to 0.07 kcal/mol when all dimers up to an O$\cdots$O distance of 5.5 Å are considered.

Fig. 2b shows a comparison between the interaction energies calculated at the CCSD(T)-F12b, SCAN, and DC-SCAN levels of theory for an unrelaxed scan of the water dimer along the O$\cdots$H distance, starting from the dimer optimized geometry. This comparison provides further evidence for DC-SCAN predicting 2-body energies in close agreement with the CCSD(T)-F12b values. In contrast, SCAN systematically overbinds the water dimer, which is particularly evident in the minimum-energy region (r(O$\cdots$H) $\sim$ 1.9 Å). It is worth noting that SCAN and DC-SCAN perform similarly for large values of r(O$\cdots$H), slightly deviating from the CCSD(T)-F12b values due to the lack of a proper description of long-range dispersion energy in both functionals.

Binding energies of water clusters. It is known that the binding energies of the low-energy isomers of the water hexamer lie within a few kcal/mol from each other[69] while the two most stable

isomers (D_{2d} and S_4) of the water octamer are degenerate.[70] Table 1 shows that the SCAN functional predicts significantly different binding energies relative to the CCSD(T)-F12b results of ref. 71 for both sets of clusters, with an overall MUE of 6.64 kcal/mol. Interestingly, the error per water molecule is higher for the 3-dimensional isomers (prism and cage isomers) than for the planar isomers of the water hexamer, and increases for the two isoenergetic isomers of the octamer.

Table 1 also includes the binding energies calculated in ref. 50 with SCAN corrected for the self-interaction energy using the Fermi–Lowdin orbital self-interaction correction (FLOSIC) scheme.[72] Relative to SCAN, FLOSIC-SCAN is able to reduce the errors in the binding energies

Table 1: Errors (in kcal/mol) in binding energies relative to the CCSD(T)-F12b values of ref. 71 calculated for representative isomers of the water hexamer and octamer using SCAN, FLOSIC-SCAN (from ref. 50) and DC-SCAN. The values in parentheses correspond to the errors per molecule. The last row reports the corresponding MUEs and MUEs per molecule.

	SCAN	FLOSIC-SCAN	DC-SCAN
Hexamer: Prism	6.09 (1.02)	-1.62 (-0.27)	-0.20 (-0.03)
Hexamer: Cage	6.12 (1.02)	-1.66 (-0.28)	-0.05 (-0.01)
Hexamer: Book 2	5.49 (0.92)	-1.56 (-0.26)	-0.20 (-0.03)
Hexamer: Cyclic boat 2	4.13 (0.69)	-2.43 (-0.41)	-0.65 (-0.08)
Octamer: D_{2d}	9.00 (1.13)	-1.31 (-0.16)	-0.42 (-0.05)
Octamer: S_4	9.00 (1.13)	-1.31 (-0.16)	-0.42 (-0.05)
MUE	6.64 (0.99)	1.65 (0.26)	0.32 (0.04)

of all water clusters analyzed in Table 1, resulting in a MUE of 1.65 kcal/mol. The error per water molecule remains nearly constant for the prism, cage, and book-2 isomers of the hexamer but increases for the cyclic boat-2 isomer. The FLOSIC-SCAN error per molecule is smaller for the two isoenergetic isomers of the water octamer. The comparisons reported in Table 1 show that DC-SCAN performs better than FLOSIC-SCAN, with an overall MUE of -0.32 kcal/mol relative to CCSD(T)-F12b. As found with FLOSIC-SCAN, also in the case of DC-SCAN the error per molecule remains constant for the prism, cage, and book-2 isomers of the water hexamer and increases for the cyclic boat-2 isomer. However, contrary to FLOSIC-SCAN, DC-SCAN predicts a larger error per molecule for the two isoenergetic octamer isomers than for the prism, cage, and book-2 isomers of the hexamer. The overall MUE per molecule of 0.04 kcal/mol indicates that the binding energies predicted by DC-SCAN are in excellent agreement with the CCSD(T)-F12b reference values for all clusters analyzed in Table 1.

Many-body interactions in water. Although the results presented in Fig. 1 and Table 1 demonstrate that, by correcting density-driven errors, DC-SCAN is able to accurately reproduce the interaction energies of small water clusters, the analyses of the previous sections do not provide any direct information about the ability of DC-SCAN to correctly describe many-body effects in water. The competition and interplay of many-body effects have been shown to play a critical role in determining structural, thermodynamic, and dynamical properties of aqueous systems, from small clusters to bulk solutions and interfaces.[69,73–76]

To investigate the impact of the density correction on individual n-body (nB) contributions

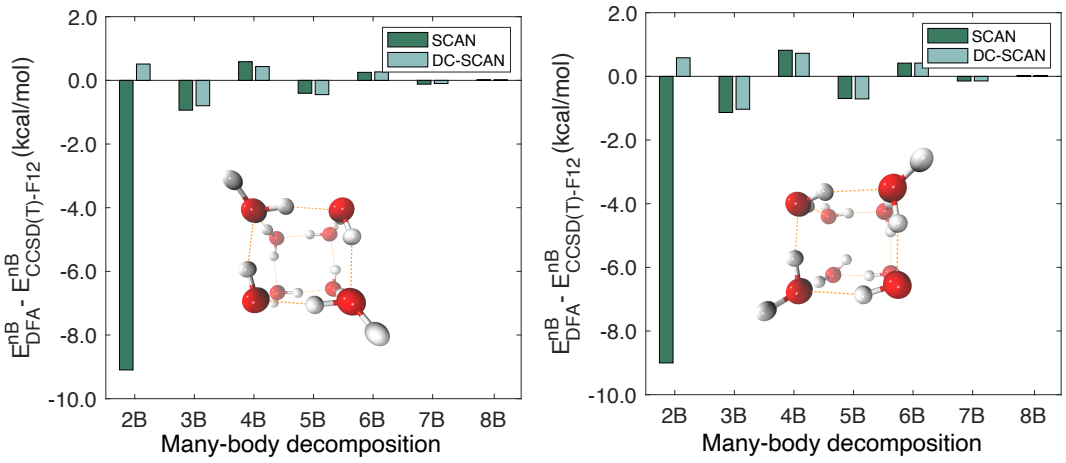

Figure 3: Errors relative to CCSD(T)-F12b reference values for each nB energy contribution to the interaction energies calculated for the two isoenergetic isomers (D_{2d} and S_4) of the water octamer using the SCAN and DC-SCAN functionals.

to the interactions between water molecules, many-body decomposition analyses were carried out for the two isoenergetic isomers of the octamer. Errors relative to the CCSD(T)-F12b reference values are shown in Fig. 3 for each n-body energy calculated with the SCAN and DC-SCAN functionals. This analysis provides further evidence for the density-driven errors in the SCAN functional primarily affecting 2-body energies, with SCAN displaying large negative deviations from the CCSD(T)-F12b values, which confirms the tendency of the SCAN functional to overbind water clusters.[50] After application of the density correction, the errors in the 2-body energies reduce to only $\sim$0.3 kcal/mol for calculations carried out with the DC-SCAN functional. Importantly, Fig. 3 shows that the impact of the density correction is minimal for all nB energies with $n > 2$.

After demonstrating that, by removing density-driven errors, the DC-SCAN functional ef-

fectively provides chemical accuracy for binding, interaction, and many-body energies of various water clusters, the 2B, 3B, and 4B energies, as well as the total interaction energies of the low-energy isomers of the water hexamer calculated using the SCAN and DC-SCAN functionals are compared in Fig. 4 with the analogous values calculated with the corresponding MB-SCAN and MB-SCAN(DC) potential energy functions (PEFs) described in the "Methods" section. Also shown for reference are the CCSD(T)/CBS values reported in ref. 63. As already discussed in the case of the octamer isomers, density-driven errors are most pronounced at the 2B level, with MUEs of 4.59 kcal/mol and 0.08 kcal/mol associated with SCAN and DC-SCAN, respectively. The MUEs reduce to 0.59 kcal/mol and 0.38 kcal/mol at the 3-body level. The comparisons shown in Fig. 4a and Fig. 4b demonstrate that both the MB-SCAN and MB-SCAN(DC) PEFs are able to quantitatively reproduce the 2-body and 3-body energies calculated *ab initio* with the corresponding SCAN and DC-SCAN functionals. Since, by construction, nB energies with $n > 3$ in the MB-SCAN and MB-SCAN(DC) PEFs are entirely represented by a classical polarization term, the errors associated with these energies are not strictly related to those calculated *ab initio* with the corresponding SCAN and DC-SCAN functionals. In this regard, Fig. 4c shows that the 4-body energies predicted by the MB-SCAN and MB-SCAN(DC) PEFs tend to underbind the hexamer isomers relative to CCSD(T)/CBS, whereas the 4-body energies calculated with the SCAN and DC-SCAN functionals tend to overbind the same clusters. However, it should be noted that in both cases the 4-body errors are small for all eight isomers, with SCAN and MB-SCAN providing MUEs of 0.17 kcal/mol and 0.35 kcal/mol, respectively. The corresponding MUEs for DC-SCAN and MB-SCAN(DC) are 0.16 kcal/mol and 0.21 kcal/mol, respectively.

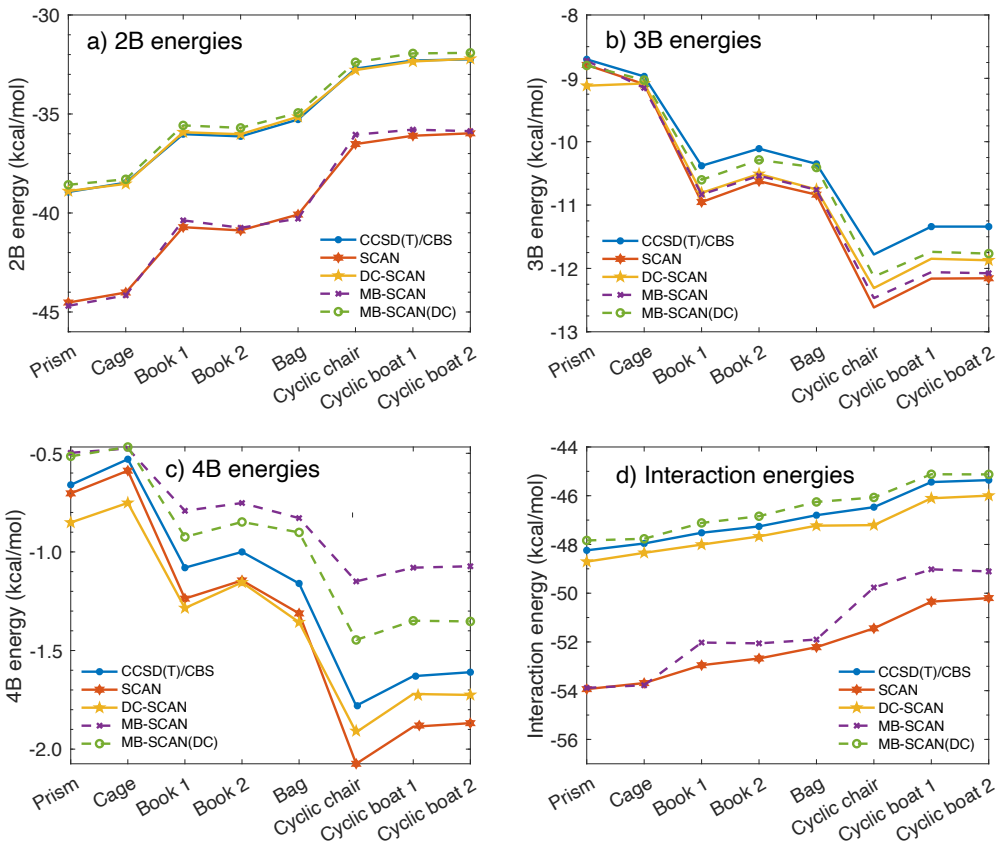

Figure 4: a) 2-body (2B) b) 3-body (3B) c) 4-body (4B) and d) total interaction energies of the first eight isomers of the water hexamer calculated using SCAN, DC-SCAN, MB-SCAN, MB-SCAN(DC), along with the CCSD(T)/CBS reference values of ref. 63.

The total interaction energies of the eight low-energy hexamer isomers calculated with the SCAN and DC-SCAN functionals, and the corresponding MB-SCAN and MB-SCAN(DC) PEFs are compared with the CCSD(T)/CBS reference values in Fig. 4d. Both DC-SCAN and MB-SCAN(DC) provide excellent agreement with the CCSD(T)/CBS reference values, displaying MUEs of 0.53 kcal/mol and 0.36 kcal/mol, respectively. In contrast, suffering from large density-driven errors at the 2-body level, SCAN and MB-SCAN systematically overbind all eight isomers.

Structural and dynamical properties of liquid water. The last question that remains to be addressed is whether the high accuracy displayed by the MB-SCAN(DC) PEF in reproducing the multidimensional energy landscape of water clusters is sufficient to correctly predict the properties of liquid water. To this end, classical MD simulations for a periodic box containing 256 molecules were carried out with the MB-SCAN(DC) PEF in the NPT ensemble at 1 atm and various temperatures between $T = 250$ K and $T = 350$ K. The lengths of the MD trajectories were 2.6 ns for $T < 298$ K and 2 ns for $T \geq 298$ K. Fig. 5 shows that the MB-SCAN(DC) PEF correctly reproduces the temperature-dependence of the density of liquid water at 1 atm, underestimating the experimental values by only $\sim$0.01 g/cm^3 at all temperatures. At 298 K, MB-SCAN(DC) predicts a density of 0.986 g/cm^3, which is in close agreement with the experimental value of 0.997 g/cm^3. The temperature of maximum density calculated by fitting a 5$^{\text{th}}$-order polynomial to the MB-SCAN(DC) results is 280 K, in nearly quantitative agreement with the experimental value of 277 K. The MB-SCAN(DC) results are compared in Fig. 5 with those reported in the literature from MD simulations with SCAN[43] (SCAN-AIMD) as well as with NNPs trained on SCAN[77] (SCAN-NNP) and SCAN0[52] (SCAN0-NNP) data. These comparisons demonstrate that the MB-SCAN(DC) PEF predicts a liquid density at 330 K which is in significantly closer agreement with experiment than the value calculated in ref. 43 from *ab initio* MD simulations with SCAN.

Particularly interesting is the comparison of the MB-SCAN(DC) PEF with the two NNPs models trained on SCAN[77] and SCAN0[52] data. Fig. 5 shows that, despite being trained on SCAN data, the SCAN-NNP model is unable to correctly reproduce the density value calculated from *ab initio* MD simulations with SCAN at 330 K. In addition, the SCAN-NNP model predicts a more

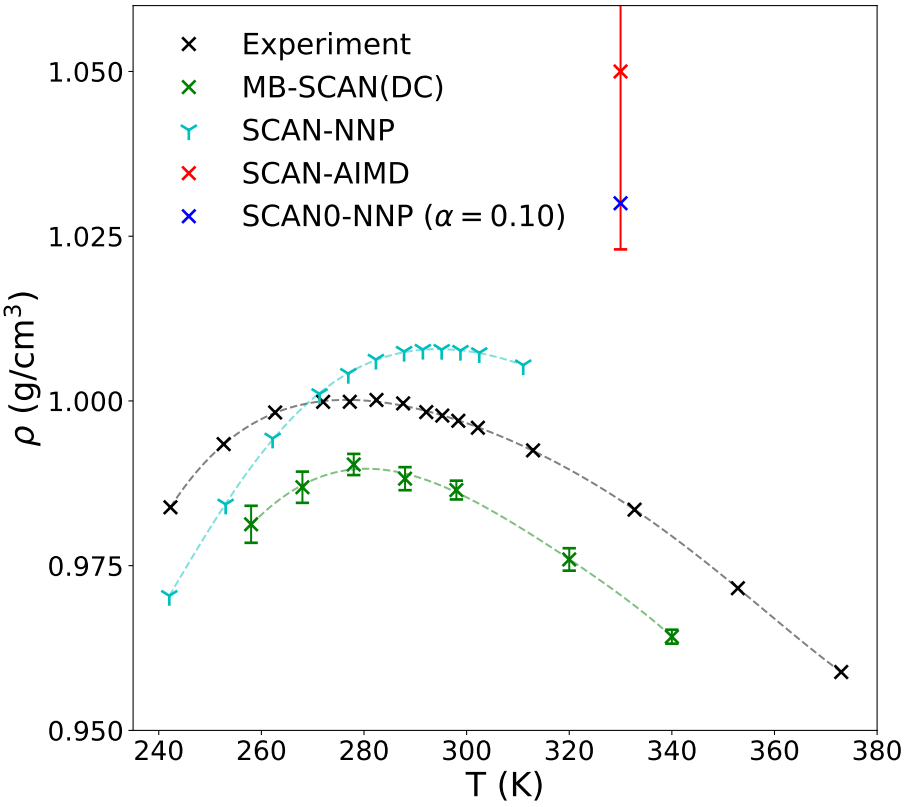

Figure 5: Temperature-dependence of the density of liquid water at 1 atm calculated from classical NPT simulations carried out with MB-SCAN(DC) along with the results from SCAN-AIMD,[43]), SCAN-NNP,[77]) and SCAN0-NNP (with 10% HF exchange)[52]) simulations. The experimental data are from the NIST Chemistry WebBook.[78]

pronounced temperature-dependence of the liquid density compared to experiment, overestimating both the value and the temperature of the density maximum.[77] A slightly more accurate prediction of the liquid density at 330 K is provided by the SCAN0-NNP model,[52] although no *ab initio* MD simulations with SCAN0 have been reported to compare with. Given the increased popularity of NNPs trained on DFT data, we believe that the differences between SCAN-AIMD and SCAN-

NNP results deserve further investigation to assess the ability of NNPs to faithfully represent the target DFT models. In this context, it should be noted that in a previous study[51] we found that MD simulations carried out with the MB-SCAN PEF predict a liquid density of 1.14 g/cm^3 at 298 K, which is significantly different from the value of 1.05 g/cm^3 obtained from *ab initio* simulations with SCAN.[43] This difference is not due to the different size of the water systems studied in the two sets of simulations (256 molecules for MB-SCAN[51] and 64 molecules for SCAN-AIMD[43]). An explanation for this difference, proposed in ref. 51, considers that any PEF rigorously derived from the many-body expansion of the energy (MBE) is strictly faithful to its parent quantum-mechanical method only when the latter does not display spurious delocalization of the electron density which affects the convergence of the MBE in an unphysical manner. We believe that the present analysis of the SCAN and DC-SCAN functionals, along with the corresponding MB-SCAN and MB-SCAN(DC) PEFs, provides support for the interpretation presented in ref. 51 that density-driven errors are responsible for the differences between MD simulations carried out with the SCAN functional and the MB-SCAN PEF.

Fig. 6 compares the oxygen-oxygen (g_{OO}) radial distribution function (RDF) calculated from MD simulations carried out with the MB-SCAN and MB-SCAN(DC) PEFs at 298 K and 1 atm with the corresponding experimental data.[67,68] The MB-SCAN(DC) PEF provides excellent agreement with the experimental RDF, slightly overestimating the height of the first peak while underestimating the height of the "valley" between the first two peaks. These small differences can be attributed to the neglect of nuclear quantum effects in classical MD simulations. Nuclear quantum effects were indeed shown to slightly lower the height of the first peak and raise the "valley"

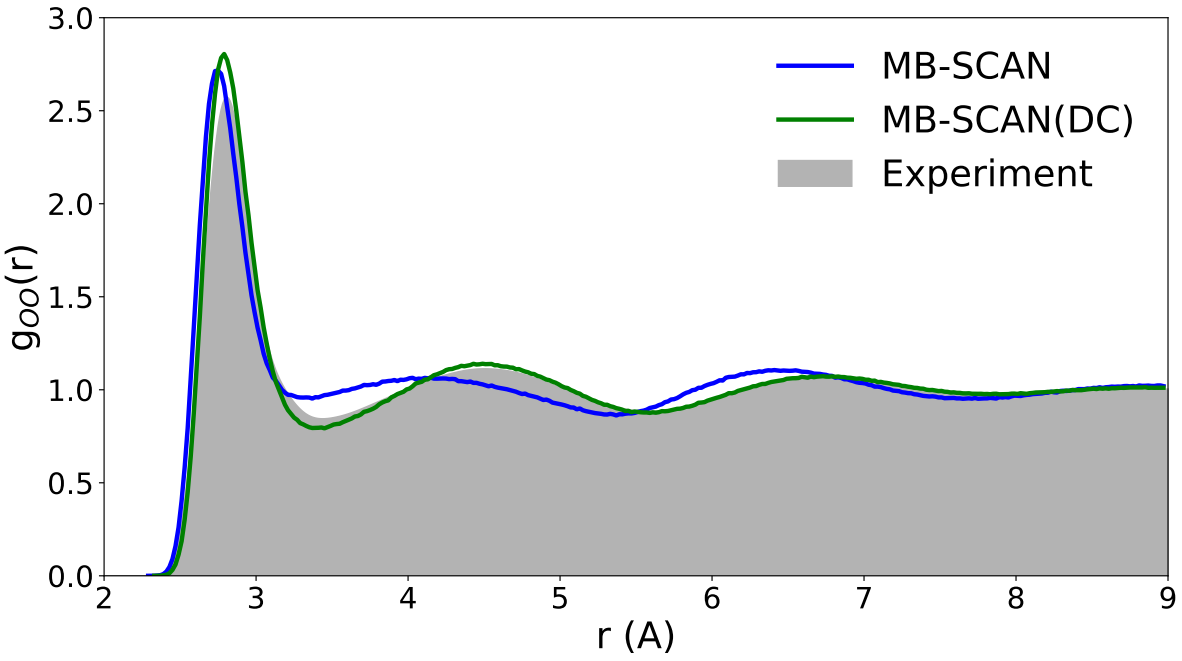

Figure 6: Oxygen-oxygen (g_{OO}) radial distribution function (RDF) calculated from NPT simulations carried out with the MB-SCAN(DC) PEF at 298K and 1 atm. The experimental RDFs at 295 K are taken from refs. 67 and 68.

between 3.2 Å and 4.0 Å of the g_{OO} calculated with the MB-pol PEF.[66] In contrast, as already discussed in ref. 51, the MB-SCAN PEF predicts a denser and more unstructured liquid. Based on the analyses discussed above, the differences between the MB-SCAN and MB-SCAN(DC) oxygen-oxygen RDFs can be unambiguously attributed to density-driven errors that affect SCAN many-body energies, particularly at the 2-body level, which are used to train the corresponding MB-SCAN PEF.

To provide further insights into the ability of the MB-SCAN(DC) PEF to describe the properties of liquid water, we also calculated the temperature dependence of the self-diffusion coeffi-

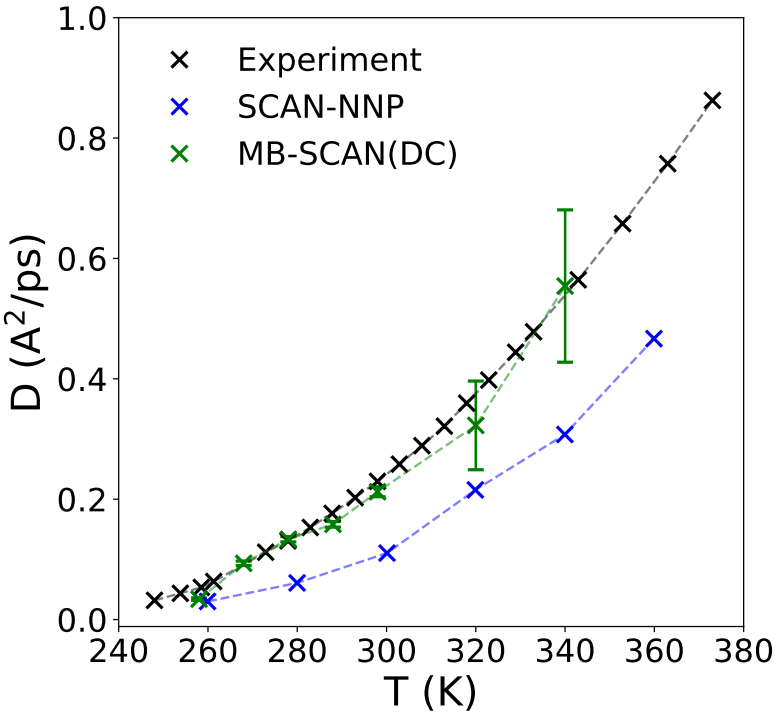

Figure 7: Temperature-dependence of the self-diffusion coefficient of liquid water calculated from NVE simulations carried out with the MB-SCAN(DC) PEF. The SCAN-NNP data are from ref. 79, while the experimental data are from refs. 80, 81 and 82.

cient, D, from a 500 ps-long MD simulation carried out in the microcanonical (NVE) ensemble for a periodic box containing 256 molecules using the equilibrium density determined from the corresponding NPT simulations. D was calculated from the velocity autocorrelation function of the center of mass of each water molecule according to

$$D = \frac{1}{3}\int_0^\infty \langle v_i(t) v_i(0) \rangle dt, \tag{1}$$

where v_i is the center of mass velocity of the ith water molecule. Fig. 7 shows that the MB-SCAN(DC) PEF is able to correctly predict the diffusion coefficient between 250 K and 350 K.

In particular, at 298 K, the diffusion coefficient predicted by the MB-SCAN(DC) PEF is 0.212 A^2/ps, which is in excellent agreement with the experimental value of 0.229 A^2/ps. This is in contrast to the value of 0.106 A^2/ps obtained in ref. 79 from MD simulations carried out with an adaptive neural network model trained on SCAN data (SCAN-NNP in Fig. 7). In contrast to the MB-SCAN(DC) PEF, the SCAN-NNP model severely underestimates the diffusion coefficient of liquid water over the entire temperature range, although the agreement with experiment apparently improves as the temperature decreases.

Discussion

An *ab initio* representation of water across all the different phases has been an elusive goal since the early days of computer simulations.[83–86]. Although models based on correlated wavefunction theories (WFT) can, in principle, provide such a long-sought after *ab initio* representation of water without resorting to *ad hoc* approximations or empirical parameterizations, the associated computational cost precludes the application of WFT models to systems containing more than a handful of water molecules. This effectively leaves DFT as the only viable approach to *ab initio* simulations of water.[13] However, it has been shown that existing XC functionals are not particularly accurate in their predictions of the properties of water,[14,37] suffering from both functional-driven and density-driven errors.

While the search for the "optimal" XC functional for water has been, in most cases, guided by evaluating the accuracy of a given functional based on its construction and/or its ability to

reproduce experimental data,[43,44,52,87–90] more complete assessments, which also include an understanding of density-driven errors, have shed light on both merits and shortcomings of existing XC functionals at the fundamental level.[46,47,50,51,59,62,91,92] In particular, the recent development of the nonempirical SCAN functional has sparked renewed interest in *ab initio* simulations of water since, by satisfying all 17 known constraints for meta-GGA functionals, it largely reduces functional-driven errors.[41,42] However, as all GGA and meta-GGA functionals, SCAN still suffers from density-driven errors that are intrinsic to computationally-efficient semilocal density functional approximations. Specifically, previous studies have shown that the main source of density-driven errors in SCAN simulations of water are the 2-body energies.[50,51] In this context, it has recently been shown that using the Hartree-Fock density in the density-corrected DFT formalism (DC-DFT) mitigates density-driven errors, particularly in the case of nonempirical functionals, because the standard fitting procedures for empirical functionals conflate density-driven errors with functional-driven errors.[62]

In this study, we have demonstrated that the density-corrected SCAN (DC-SCAN) functional effectively removes density-driven errors from the water 2-body energies, which brings both binding and interaction energies of different water clusters very close to reference values calculated at CCSD(T)/CBS level of theory. Although not as pronounced as for the 2-body energies, the density correction also reduces density-driven errors in all higher-body terms of the many-body expansion (MBE) of the energy calculated for water using the DC-SCAN functional, with each individual many-body term being in quantitative agreement with the corresponding CCSD(T)/CBS reference values. In this context, it should be noted that a previous study[50] found a significant but less

complete improvement for water clusters (Table 1) via a self-consistent FLOSIC self-interaction correction to SCAN. Ref. 50, however, did not find evidence for a major improvement from density correction, probably because the FLOSIC density is less localized than the Hartree-Fock and exact densities are. Although it should be kept in mind that the DC-SCAN functional, as does the parent SCAN functional, still suffers from functional-driven errors, which can be large for some chemical systems such as stretched H_2^+, the analyses presented here demonstrate that these functional-driven errors are negligible for water. In the future, it would be important to test the performance of DC-SCAN for more-general chemical applications. Importantly, our analyses suggest that, in principle, *ab initio* MD simulations with the DC-SCAN functional should be able to provide a consistently accurate description of the properties of water. However, the requirement of using the Hartree-Fock density in a non-self-consistent SCAN calculation at each MD step would make *ab initio* MD simulations with DC-SCAN not straightforward to implement and expensive to perform.

While *ab initio* MD simulations with DC-SCAN are currently not feasible, we have shown that the improved accuracy of the DC-SCAN functional can be exploited to develop a data-driven many-body potential energy function, the MB-SCAN(DC) PEF, which indeed provides a highly accurate representation of water, from small clusters in the gas phase to the liquid phase. MB-SCAN(DC) is rigorously derived from the DC-SCAN MBE and adopts a hybrid data-driven/physics-based scheme, where a data-driven model, which captures (short-range) quantum-mechanical interactions arising from the overlap of the electron densities of individual molecules at the 2-body and 3-body levels (e.g., Pauli repulsion, and charge transfer and penetration), is integrated with

a physics-based model of many-body interactions, which is represented by classical many-body electrostatics. Importantly, we have demonstrated that the MB-SCAN(DC) PEF achieves high accuracy by quantitatively reproducing each individual term of the DC-SCAN MBE for water, providing a correct representation of both short- and long-range many-body contributions. Since the DC-SCAN functional exhibits chemical accuracy for each individual term of the MBE for water and the MB-SCAN(DC) PEF quantitatively reproduces the DC-SCAN many-body energies, the MB-SCAN(DC) PEF effectively provides the first demonstration of a DFT-based model that correctly describes the properties of water, at the computational cost of advanced polarizable force fields.[14] Future applications of the MB-SCAN(DC) PEF will focus on modeling the phase diagram of water, which was shown to be only qualitatively reproduced by NNPs trained on SCAN data.[77,93] We expect MB-SCAN(DC) to be especially well suited to modeling the liquid/vapor equilibrium, where most hydrogen bonds are broken.

Finally, we want to emphasize that the many-body formalism adopted by the MB-SCAN(DC) PEF for water is general and has already been used in the development of data-driven many-body PEFs for various aqueous systems[94,95] and molecular fluids[96,97] which were trained on (expensive) CCSD(T) data. It thus follows that the significantly lower computational cost associated with DC-SCAN calculations can enable the routine development of MB-SCAN(DC) PEFs for generic (small) molecules which are trained on DC-SCAN data but effectively display CCSD(T) accuracy. In this context, it should be noted that the MB-Fit software infrastructure[98] for many-body PEFs combined with the MBX many-body energy/force calculator[99] interfaced with i-PI[100] and LAMMPS[101] already provides a robust platform for MD simulations of generic molecules in the

gas, liquid, and solid phases using MB-SCAN(DC) PEFs.

Methods

Density functional theory. In ground-state Kohn-Sham DFT,[19] the energy is self-consistently minimized as:

$$E = \min_{n} \left\{ F[n] + \int d^3r \, n(\mathbf{r}) v(\mathbf{r}) \right\} \tag{2}$$

where the minimizing $n(\mathbf{r})$ is the ground-state density, $v(\mathbf{r})$ is the external potential, and $F[n]$ includes the exact non-interacting kinetic and Hartree electrostatic energy terms plus an exchange-correlation (XC) energy. Since the exact XC functional is unknown, different DFT approximations have been developed to solve eq 2.

Density correction. The total-energy error ΔE associated with different DFT approximations can be written as the sum of the functional-driven error, ΔE_F, and the density-driven error, ΔE_D:[59]

$$\Delta E = \Delta E_{\mathrm{F}} + \Delta E_{\mathrm{D}} \tag{3}$$

The functional-driven error $\Delta E_{\mathrm{F}} = E_{\mathrm{XC}}^{\mathrm{approx}}[n_{\mathrm{exact}}] - E_{\mathrm{XC}}^{\mathrm{exact}}[n_{\mathrm{exact}}]$ arises from the difference between the approximate XC functional, $F[n]$, and the (unknown) exact functional, while the density-driven error arises from using an approximate density $n(\mathbf{r})$ to solve eq 2. In most systems, the functional-driven error is the main contribution to the total error.[56,59] By many measures, the best nonempirical functionals predict more accurate densities for neutral atoms than either the nonempirical functionals or even Hartree-Fock theory.[102] But they still make density-driven delocalization errors[103,104] that can dominate the total error under special conditions.[57,105]

Independent of the specific form and parametrization, standard approximate XC functionals still deviate from the piecewise-linear behavior of the exact functional for fractional charges,[103] causing excess charge delocalization and resulting in incorrect densities.[103,104] For certain systems, the density-driven error thus become the dominant contributor to the total error.[59,106] This error can be understood by considering that the classical electrostatic repulsion term represented by the integral in eq. 2 contains a self-interaction contribution due to each electron interacting with itself.[45,105] While this self-interaction contribution should, in theory, 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 over-delocalize in order to minimize the many-electron self-interactions,[48,91,107] leading to fractional charges that underestimate the energy predicted by the piecewise-linear behavior of the exact functional.[103,108]

Using a more accurate density can mitigate errors due to the over-delocalization of the electron density.[56,60,109] However, obtaining an accurate density from wavefunction theories, such as Møller-Plesset peturbation theory and coupled cluster theory, is computationally significantly more expensive than the corresponding DFT calculations. An approximate, yet efficient, approach to reducing density-driven errors in DFT calculations consists in using the Hartree-Fock density, $n^{\mathrm{HF}}(\mathbf{r})$ because, by construction, it does not suffer from either electron over-delocalization or self-interaction errors.[56,59–62] The resulting density-corrected DFT (DC-DFT) energy can then be written as:

$$E^{\mathrm{DC-DFT}} \approx E^{\mathrm{HF}} + \left(E_{\mathrm{XC}}^{\mathrm{approx}}\left[n^{\mathrm{HF}}\right] - E_{\mathrm{X}}^{\mathrm{HF}}\right) \tag{4}$$

Many-body expansion. Eq. 4 can be used to calculate individual n-body energies, $\epsilon^{n\mathrm{B}}$, from one-body (1B) to N-body (NB), which enter the many-body expansion (MBE) of the energy for a system containing N (atomic or molecular) monomers:[110]

$$E_N(1,\ldots,N) = \sum_{i=1}^{N} \epsilon^{1\mathrm{B}}(i) + \sum_{i<j}^{N} \epsilon^{2\mathrm{B}}(i,j) + \sum_{i<j<k}^{N} \epsilon^{3\mathrm{B}}(i,j,k) + \ldots + \epsilon^{\mathrm{NB}}(1,\ldots,N), \quad (5)$$

In the case of water, $\epsilon^{1\mathrm{B}}(i)$ in eq 5 corresponds to the distortion energy of the ith water molecule in the system from the equilibrium geometry of the corresponding free molecule, and all higher-order n-body energies, $\epsilon^{n\mathrm{B}}(2,\ldots,n)$, can be calculated recursively from the lower-order terms.[111] Building upon the demonstrated accuracy of the MB-pol PEF for water[63–66] and following the same theoretical/computational approach employed in development of DFT-based many-body PEFs,[51,75,92] we used eq. 5 to develop a data-driven many-body PEF, MB-SCAN(DC), that consistently reproduces each term of the MBE for water calculated using the DC-SCAN functional. Briefly, MB-SCAN(DC) includes explicit representations of 1B, 2B, and 3B energies, and describes all higher-order nB energy terms ($n > 3$) through classical many-body polarization. Specifically, $\epsilon^{1\mathrm{B}}$ in eq. 5 is represented by the Partridge-Schwenke PEF,[112] while $\epsilon^{2\mathrm{B}}$ and $\epsilon^{3\mathrm{B}}$ are represented by terms describing permanent elecrostatics, dispersion energy, and induction, which are combined with short-range permutationally invariant polynomials (PIPs)[113] fitted to reproduce 2B and 3B energies calculated with DC-SCAN for the same training sets of water dimers and trimers used in the development of MB-pol.[64,65] A detailed description of the theoretical and computational framework adopted in the development of data-driven many-body PEFs for water can be found in the original references.[51,64,65,75,92] It should be noted that, since our many-body PEFs directly target the underlying molecular interactions, differences in the representation of the 1-body (1B)

term of eq. 5 have been found to be negligible for modeling the properties of liquid water[92] and the air/water interface.[114]

All DFT calculations were performed with the aug-cc-pVQZ basis set[115,116] using Q-Chem[117] quantum chemistry packages. Since the SCAN functional is particularly sensitive to the real-space grid, all SCAN and DC-SCAN calculations are performed on the highly dense Euler-Maclaurin-Lebedev (99,590) grid[118,119] (58,410 points per atom). In this regard, the results of a sensitivity analysis reported in Table S1 suggest that the SG2 grid[120] ($\sim$8,000 points per atom) should also be sufficient to converge SCAN calculations. In case only smaller grids are available, we recommend to use r^{2}SCAN,[121] which often achieves an accuracy similar to SCAN. Single-point energy calculations using explicitly correlated coupled cluster, CCSD(T)-F12b, theory[122] were performed in the CBS limit by extrapolating [123,124] 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[125,126] using the ORCA quantum chemistry package.[127]

Data availability

Any data generated and analyzed for this study that are not included in this Article and its Supplementary Information are available from the authors upon request.

Code availability

The MB-SCAN and MB-SCAN(DC) PEFs are available in MBX,[99] and can be used in MD simulations with LAMMPS[101] and i-PI.[100] All computer codes used in the analysis presented in this study are available from the authors upon request.

Acknowledgements We thank Kieron Burke for helpful and stimulating discussions. This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, through grant no. DE-SC0019490 (F.P), and by the U.S. National Science Foundation through grant no. DMR-1939528, with a contribution from CTMC (J.P.P.). 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).

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Francesco Paesani (email: fpaesani@ucsd.edu).

1. Gallo, P. *et al.* Water: A tale of two liquids. *Chem. Rev.* **116**, 7463–7500 (2016).

2. Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108 (2008).

3. Franks, F. *Water: A Matrix of Life*, vol. 21 (Royal Society of Chemistry, 2000).

4. Eisenberg, D., Kauzmann, W. & Kauzmann, W. *The Structure and Properties of Water* (Oxford University Press, 2005).

5. Tanford, C. The hydrophobic effect and the organization of living matter. *Science* **200**, 1012–1018 (1978).

6. Jencks, W. P. General acid-base catalysis of complex reactions in water. *Chem. Rev.* **72**, 705–718 (1972).

7. Savage, P. E. Organic chemical reactions in supercritical water. *Chem. Rev.* **99** (1999).

8. Lindström, U. M. Stereoselective organic reactions in water. *Chem. Rev.* **102**, 2751–2772 (2002).

9. Akiya, N. & Savage, P. E. Roles of water for chemical reactions in high-temperature water. *Chem. Rev.* **102**, 2725–2750 (2002).

10. Li, C.-J. & Chen, L. Organic chemistry in water. *Chem. Soc. Rev.* **35**, 68–82 (2006).

11. Simon, M.-O. & Li, C.-J. Green chemistry oriented organic synthesis in water. *Chem. Soc. Rev.* **41**, 1415–1427 (2012).

12. Vega, C. & Abascal, J. L. Simulating water with rigid non-polarizable models: A general perspective. *Phys. Chem. Chem. Phys.* **13**, 19663–19688 (2011).

13. Hassanali, A. A., Cuny, J., Verdolino, V. & Parrinello, M. Aqueous solutions: State of the art in *ab initio* molecular dynamics. *Philos. Trans. R. Soc. A* **372**, 20120482 (2014).

14. Cisneros, G. A. *et al.* Modeling molecular interactions in water: From pairwise to many-body potential energy functions. *Chem. Rev.* **116**, 7501–7528 (2016).

15. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864 (1964).

16. Kohn, W. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **71**, 1253 (1999).

17. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471 (1985).

18. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87**, 897 (2015).

19. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).

20. Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566 (1980).

21. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **45**, 13244 (1992).

22. Glötzel, D. & McMahan, A. Relativistic effects, phonons, and the isostructural transition in cesium. *Phys. Rev. B* **20**, 3210 (1979).

23. Skriver, H. L. Crystal structure from one-electron theory. *Phys. Rev. B* **31**, 1909 (1985).

24. Moriarty, J. A. & McMahan, A. High-pressure structural phase transitions in Na, Mg, and Al. *Phys. Rev. Lett.* **48**, 809 (1982).

25. Laasonen, K., Csajka, F. & Parrinello, M. Water dimer properties in the gradient-corrected density functional theory. *Chem. Phys. Lett.* **194**, 172–174 (1992).

26. Laasonen, K., Parrinello, M., Car, R., Lee, C. & Vanderbilt, D. Structures of small water clusters using gradient-corrected density functional theory. *Chem. Phys. Lett.* **207**, 208–213 (1993).

27. Perdew, J. P. & Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conf. Proc.*, vol. 577, 1–20 (American Institute of Physics, 2001).

28. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098 (1988).

29. 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* **37**, 785 (1988).

30. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

31. Sim, F., St. Amant, A., Papai, I. & Salahub, D. R. Gaussian density functional calculations on hydrogen-bonded systems. *J. Am. Chem. Soc.* **114**, 4391–4400 (1992).

32. Tuckerman, M. E. Ab initio molecular dynamics: Basic concepts, current trends and novel applications. *J. Condens. Matter Phys.* **14**, R1297 (2002).

33. 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.* **127**, 184104 (2007).

34. Kuo, I.-F. W. *et al.* Liquid water from first principles: Investigation of different sampling approaches. *J. Phys. Chem. B* **108**, 12990–12998 (2004).

35. 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.* **120**, 300–311 (2004).

36. VandeVondele, J. *et al.* The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.* **122**, 014515 (2005).

37. Gillan, M. J., Alfe, D. & Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **144**, 130901 (2016).

38. 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.* **134**, 024516 (2011).

39. 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.* **82**, 2544 (1999).

40. Adamo, C., Ernzerhof, M. & Scuseria, G. E. The Meta-GGA functional: Thermochemistry with a kinetic energy density dependent exchange-correlation functional. *J. Chem. Phys.* **112**, 2643–2649 (2000).

41. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).

42. Sun, J. *et al.* Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nat. Chem.* **8**, 831 (2016).

43. Chen, M. *et al.* Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. USA* **114**, 10846–10851 (2017).

44. Zheng, L. *et al.* Structural, electronic, with dynamical properties of liquid water by ab initio molecular dynamics based on scan functional within the canonical ensemble. *J. Chem. Phys.* **148**, 164505 (2018).

45. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048 (1981).

46. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Development of exchange-correlation functionals with minimal many-electron self-interaction error. *J. Chem. Phys.* **126**, 191109 (2007).

47. 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.* **100**, 146401 (2008).

48. 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.* **129**, 204112 (2008).

49. 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.* **114**, 053001 (2015).

50. Sharkas, K. *et al.* Self-interaction error overbinds water clusters but cancels in structural energy differences. *Proc. Natl. Acad. Sci. USA* **117**, 11283–11288 (2020).

51. 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.* **17**, 3739–3749 (2021).

52. Zhang, C. *et al.* Modeling liquid water by climbing up Jacob's ladder in density functional theory facilitated by using deep neural network potentials. *arXiv:2104.14410.* .

53. Rezac, J. & Hobza, P. Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard," CCSD(T) at the complete basis set limit? *J. Chem. Theory Comput.* **9**, 2151–2155 (2013).

54. Hui, K. & Chai, J.-D. SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters. *J. Chem. Phys.* **144**, 044114 (2016).

55. Gordon, R. G. & Kim, Y. S. Theory for the forces between closed-shell atoms and molecules. *J. Chem. Phys.* **56**, 3122–3133 (1972).

56. Scuseria, G. E. Comparison of coupled-cluster results with a hybrid of Hartree–Fock and density functional theory. *J. Chem. Phys.* **97**, 7528–7530 (1992).

57. 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.* **100**, 6550–6561 (1994).

58. Janesko, B. G. & Scuseria, G. E. Hartree–Fock orbitals significantly improve the reaction barrier heights predicted by semilocal density functionals. *J. Chem. Phys.* **128**, 244112 (2008).

59. Kim, M.-C., Sim, E. & Burke, K. Understanding and reducing errors in density functional calculations. *Phys. Rev. Lett.* **111**, 073003 (2013).

60. Kim, M.-C., Sim, E. & Burke, K. Ions in solution: Density corrected density functional theory (DC-DFT). *J. Chem. Phys.* **140**, 18A528 (2014).

61. Vuckovic, S., Song, S., Kozlowski, J., Sim, E. & Burke, K. Density functional analysis: The theory of density-corrected DFT. *J. Chem. Theory Comput.* **15**, 6636–6646 (2019).

62. Song, S., Vuckovic, S., Sim, E. & Burke, K. Density sensitivity of empirical functionals. *J. Phys. Chem. Lett.* **12**, 800–807 (2021).

63. Reddy, S. K. *et al.* 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.* **145**, 194504 (2016).

64. 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.* **9**, 5395–5403 (2013).

65. 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.* **10**, 1599–1607 (2014).

66. 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.* **10**, 2906–2910 (2014).

67. Skinner, L. B. *et al.* Benchmark oxygen-oxygen pair-distribution function of ambient water from X-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **138**, 074506 (2013).

68. Skinner, L. B., Benmore, C., Neuefeind, J. C. & Parise, J. B. The structure of water around the compressibility minimum. *J. Chem. Phys.* **141**, 214507 (2014).

69. Brown, S. E. *et al.* Monitoring water clusters "melt" through vibrational spectroscopy. *J. Am. Chem. Soc.* **139**, 7082–7088 (2017).

70. Xantheas, S. S. & Aprà, E. The binding energies of the D_{2d} and S_4 water octamer isomers: High-level electronic structure and empirical potential results. *J. Chem. Phys.* **120**, 823–828 (2004).

71. 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.* **13**, 3136–3152 (2017).

72. Pederson, M. R., Ruzsinszky, A. & Perdew, J. P. Communication: Self-interaction correction with unitary invariance in density functional theory. *J. Chem. Phys.* **140**, 121103 (2014).

73. Elrod, M. J. & Saykally, R. J. Many-body effects in intermolecular forces. *Chem. Rev.* **94**, 1975–1997 (1994).

74. Pérez, C. *et al.* Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy. *Science* **336**, 897–901 (2012).

75. 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.* **10**, 8211–8218 (2019).

76. 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.* **10**, 406–412 (2019).

77. 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.* **17**, 3065–3077 (2021).

78. Lemmon, E. W., McLinden, M. O. & Friend, D. G. Thermophysical properties of fluid systems. In Linstrom, P. & Mallard, W. (eds.) *NIST Chemistry WebBook* (National Institute of Standards and Technology, Gaithersburg (MD)).

79. Yao, Y. & Kanai, Y. Temperature dependence of nuclear quantum effects on liquid water via artificial neural network model based on SCAN meta-GGA functional. *J. Chem. Phys.* **153**, 044114 (2020).

80. Holz, M., Heil, S. R. & Sacco, A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements. *Phys. Chem. Chem. Phys.* **2**, 4740–4742 (2000).

81. Easteal, A. J., Price, W. E. & Woolf, L. A. Diaphragm cell for high-temperature diffusion measurements. tracer diffusion coefficients for water to 363 K. *J. Chem. Soc., Faraday Trans. 1* **85**, 1091–1097 (1989).

82. Mills, R. Self-diffusion in normal and heavy water in the range 1-45^{o}. *J. Phys. Chem.* **77**, 685–688 (1973).

83. Matsuoka, O., Clementi, E. & Yoshimine, M. CI study of the water dimer potential surface. *J. Chem. Phys.* **64**, 1351–1361 (1976).

84. Lie, G. & Clementi, E. Molecular-dynamics simulation of liquid water with an ab initio flexible water–water interaction potential. *Phys. Rev. A* **33**, 2679 (1986).

85. Evans, M., Refson, K., Swamy, K., Lie, G. & Clementi, E. Molecular-dynamics simulation of liquid water with an ab initio flexible water–water interaction potential. II. The effect of internal vibrations on the time correlation functions. *Phys. Rev. A* **36**, 3935 (1987).

86. Niesar, U., Corongiu, G., Clementi, E., Kneller, G. & Bhattacharya, D. Molecular dynamics simulations of liquid water using the NCC ab initio potential. *J. Phys. Chem.* **94**, 7949–7956 (1990).

87. Morrone, J. A. & Car, R. Nuclear quantum effects in water. *Phys. Rev. Lett.* **101**, 017801 (2008).

88. Galib, M. *et al.* Mass density fluctuations in quantum and classical descriptions of liquid water. *J. Chem. Phys.* **146**, 244501 (2017).

89. 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.* **9**, 5009–5016 (2018).

90. Marsalek, O. & Markland, T. E. Quantum dynamics and spectroscopy of ab initio liquid water: The interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.* **8**, 1545–1551 (2017).

91. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for density functional theory. *Chem. Rev.* **112**, 289–320 (2012).

92. Lambros, E. *et al.* General many-body framework for data-driven potentials with arbitrary quantum mechanical accuracy: Water as a case study. *ChemRxiv, DOI: 10.26434/chemrxiv.14710815.v1.* .

93. Zhang, L., Wang, H., Car, R. & Weinan, E. Phase diagram of a deep potential water model. *Phys. Rev. Lett.* **126**, 236001 (2021).

94. 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.* **12**, 2698–2705 (2016).

95. 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.* **147**, 161715 (2017).

96. 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.* **16**, 2246–2257 (2020).

97. 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* **124**, 11207–11221 (2020).

98. GitHub. MB-Fit: Software infrastructure for data-driven many-body potential energy functions. https://github.com/paesanilab/MB-Fit. URL `https://github.com/paesanilab/{MB}-{F}it`.

99. MBX: A many-body energy and force calculator. `http://paesanigroup.ucsd.edu/software/mbx.html`.

100. i-PI: A universal force engine. `http://ipi-code.org`.

101. LAMMPS molecular dynamics simulator. `http://lammps.sandia.gov`.

102. 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* **355**, 49–52 (2017).

103. 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.* **49**, 1691 (1982).

104. Zhang, Y. & Yang, W. A challenge for density functionals: Self-interaction error increases for systems with a noninteger number of electrons. *J. Chem. Phys.* **109**, 2604–2608 (1998).

105. Engel, E. & Dreizler, R. M. Exchange-correlation energy functional. In *Density Functional Theory*, 109–217 (Springer, 2011).

106. 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.* **137**, 224113 (2012).

107. Ruzsinszky, A., Perdew, J. P. & Csonka, G. I. Binding energy curves from nonempirical density functionals. I. Covalent bonds in closed-shell and radical molecules. *J. Phys. Chem. A* **109**, 11006–11014 (2005).

108. Hait, D. & Head-Gordon, M. Delocalization errors in density functional theory are essentially quadratic in fractional occupation number. *J. Phys. Chem. Lett.* **9**, 6280–6288 (2018).

109. Lee, D., Furche, F. & Burke, K. Accuracy of electron affinities of atoms in approximate density functional theory. *J. Phys. Chem. Lett.* **1**, 2124–2129 (2010).

110. Hankins, D., Moskowitz, J. & Stillinger, F. Water molecule interactions. *J. Chem. Phys.* **53**, 4544–4554 (1970).

111. Góra, U., Podeszwa, R., Cencek, W. & Szalewicz, K. Interaction energies of large clusters from many-body expansion. *J. Chem. Phys.* **135**, 224102 (2011).

112. Partridge, H. & Schwenke, D. W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *J. Chem. Phys.* **106**, 4618–4639 (1997).

113. Braams, B. J. & Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **28**, 577–606 (2009).

114. Muniz, M. C. *et al.* Vapor–liquid equilibrium of water with the mb-pol many-body potential. *J. Chem. Phys.* **154**, 211103 (2021).

115. Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).

116. 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.* **96**, 6796–6806 (1992).

117. Shao, Y. *et al.* Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **113**, 184–215 (2015).

118. Murray, C. W., Handy, N. C. & Laming, G. J. Quadrature schemes for integrals of density functional theory. *Mol. Phys.* **78**, 997–1014 (1993).

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

120. Dasgupta, S. & Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **38**, 869–882 (2017).

121. Furness, J. W., Kaplan, A. D., Ning, J., Perdew, J. P. & Sun, J. Accurate and numerically efficient r^{2}SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **11**, 8208–8215 (2020).

122. Adler, T. B., Knizia, G. & Werner, H.-J. A simple and efficient CCSD(T)-F12 approximation. *J. Chem. Phys.* **127**, 221106 (2007).

123. 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.* **129**, 184116 (2008).

124. Helgaker, T., Klopper, W., Koch, H. & Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **106**, 9639–9646 (1997).

125. Yousaf, K. E. & Peterson, K. A. Optimized auxiliary basis sets for explicitly correlated methods. *J. Chem. Phys.* **129**, 184108 (2008).

126. Yousaf, K. E. & Peterson, K. A. Optimized complementary auxiliary basis sets for explicitly correlated methods: aug-cc-pvnz orbital basis sets. *Chem. Phys. Lett.* **476**, 303–307 (2009).

127. Neese, F. Software update: The ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.* **8**, e1327:1–6 (2017).