Assessing the accuracy of the SCAN functional for water through a many-body analysis of the adiabatic connection formula

Eleftherios Lambros,[*,†] Jie Hu,[†] and Francesco Paesani[*,†,‡,¶]

†*Department of Chemistry and Biochemistry, University of California San Diego,*
La Jolla, California 92093, United States
‡*Materials Science and Engineering, University of California San Diego,*
La Jolla, California 92093, United States
¶*San Diego Supercomputer Center, University of California San Diego,*
La Jolla, California 92093, United States

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

Abstract

We present a systematic analysis of the accuracy of a series of SCANα functionals for water, with varying fractions (α) of exact exchange, which are constructed through the adiabatic connection formula. Our results indicate that that all SCANα functionals exhibit substantial errors in the representation of the water 2-body energies. Importantly, the inclusion of exact exchange is found to have opposite effects on the ability of the SCANα functionals to describe the interaction energies of water clusters with 2-dimensional and 3-dimensional hydrogen-bonding arrangements. These errors are found to directly affect the ability of the SCANα functionals to describe the structure of liquid water at ambient conditions, which is investigated using explicit many-body models (MB-SCANα) derived from the corresponding SCANα data. In particular, it is found that all MB-SCANα models predict a more compact

first hydration shell, which results in a denser liquid with a more ice-like structure. These apparent opposite trends can be explained by the inability of all SCANα functionals to provide a balanced description of the water 2B and 3B energies at the fundamental level. The analyses presented in this study provide new insights that can guide future developments of improved exchange-correlation functionals for water.

1   Introduction

Density functional theory (DFT) approaches have been used to simulate molecular systems ranging from small molecules to liquids and solids when more expensive, post-Hartree-Fock methods become prohibitively inefficient.[1–4] Originally built upon the Hohenberg and Kohn theorems,[5] DFT was later formulated in its modern version by Kohn and Sham who introduced the exchange-correlation energy functional in the expression of the total energy functional, thus reducing the original many-body problem of interacting electrons to solving for the ground-state energy and density of a fictitious system of non-interacting electrons.[6] Although in principle exact, the solution of the Kohn-Sham equations becomes approximate in practice because the exact exchange-correlation energy functional is unknown.[7]

The earliest approximation to the exact density functional, known as the local density approximation (LDA), was inspired by work by Slater who, a decade before, had proposed an approximation to the exact exchange term in Hartree-Fock theory based on a functional of the electron density.[8] Formally, the LDA is derived from the corresponding terms of the homogeneous electron gas, and assumes that the electron density is purely "local". This assumption of locality implies that the LDA approximation holds only for systems with very slowly varying densities as, at every point, the density is assumed to be uniform.[6] Besides its tendency to overestimate atomization energies,[9] the LDA also tends to overestimate the strength of hydrogen bonds in aqueous systems, predicting shorter oxygen-oxygen distances in small water clusters.[10,11] First order corrections to the LDA, initially proposed in 1969 by Herman, Van Dyke, and Ortenburger, were based on an analytic approach that eventually became the modern-day generalized gradient approximation

(GGA).[12]

Modern GGA functionals, such as the PBE functional,[13] perform relatively better than the LDA functional for aqueous systems, but still result in broad disagreement with various experimental data, including the over-structuring of liquid water at ambient conditions.[14,15] These GGA functionals were further extended to incorporate second order corrections to the LDA using the Laplacian of the electron density, leading to meta-GGA functionals.[16,17] By the late 90s, several meta-GGA functionals had been developed,[16,18–20] which eventually evolved into more recent implementations, such as B97M-V[21] and SCAN[22] functionals. Both SCAN, which satisfies all 17 known constraints on the exact meta-GGA functional, and B97M-V are able to partially overcome the over-structuring of liquid water and provide better agreement to experiment for various properties,[23–26] albeit often the comparisons with the experimental data are made with results obtained from simulations carried out at higher temperature.[23,25]

In an attempt to improve the accuracy of DFT models, hybrid exchange-correlation functionals were developed starting from the the adiabatic connection formula,,[27]

$$E_{xc}[\rho] = \int_0^1 E_{xc,\lambda} d\lambda \tag{1}$$

In practical applications, the hybrid exchange-correlation functionals approximate the integral in eq 1 as a fractional sum of the non-interacting limit, which is described by the exact exchange, and the fully interacting limit, which is described by the exchange-correlation energy calculate with the a GGA or meta-GGA density functional (DFA).[27] The original expression for $E_{xc}[\rho]$ of hybrid functionals was later refined into a 3-parameter fractional sum, with separate coefficients for the exact exchange, and the exchange and correlation terms represented by GGA or meta-GGA functionals.[28] Following the rationale introduced by Perdew and Ernzerhof in 1996,[29] PBE0 and later on, SCAN0 were proposed as hybrid versions of the corresponding base functionals (i.e., PBE and SCAN, respectively) without using any fitted parameters and adopting 25% exact exchange.[30,31] Bundled with SCAN0, the double-hybrid SCAN0-2 functional was also shown to

provide an accurate description of various test systems.[31]

Recently, the SCAN functional has been gaining popularity as a functional suitable for studies of liquid water. Yao and Kanai studied the temperature dependence of nuclear quantum effects in liquid water.[26] Sharkas *et. al.* used small water clusters to show that the self-interaction error in the SCAN functional primarily affects the two-body term of the many-body expansion (MBE) of the energy.[32] Other studies used SCAN to analyze the equilibrium of high and low density liquid water at negative pressure.[33,34] More recently, Xu *et. al.* studied the isotope effects in liquid H_2O and D_2O using a deep neural network potential (NNP) trained on the SCAN functional.[35] A similar NNP was used to investigate the properties of supercooled water,[36] as well as the equilibrium between liquid water and ice Ih and Ic,[37] and the ice Ih/XI transition.[38]

Building upon these previous studies, particularly the many-body analysis reported in ref 32, we provide here an assessment of the accuracy of the SCAN functional in predicting the properties of water. Our analysis uses the adiabatic connection formula to systematically investigate the effects of the exact exchange on the individual terms of the MBE as well as on the structure of liquid water at ambient conditions and the description of hydrogen bonding in the water dimer. We demonstrate that by modulating the fraction of exact exchange, small, and even negligible errors in the representation of the low-order terms of the MBE can lead to substantial changes in the structure of liquid water. Interestingly, our analysis indicates that the inclusion of exact exchange has opposite effects on the description of different hydrogen-bonding motifs, improving the accuracy of SCAN for 3-dimensional structures, while deteriorating its performance for 2-dimensional arrangements of water molecules.

2   Theory and Computational Details

All electronic structure calculations were performed with the Q-Chem software package[39] using the SCAN functional[22] with the aug-cc-pVQZ basis set.[40,41] The amount of exact exchange was specified using the "general" exchange functional section in Q-Chem. All molecular dynamics

(MD) simulations were carried out in the isobaric-isothermal (NPT) ensemble for 1 ns with a 0.2 fs timestep using in-house software based on the DL_POLY2 simulation package,[42] which was modified to include the MB-DFT models.[43] The temperature was controlled via Nosé–Hoover chains (NHC) of four thermostats coupled to each degree of freedom.[44] The NPT ensemble was generated according to the algorithm described in ref. 45. A radial atom– atom cutoff distance of 9.0 Å was applied to the nonbonded interactions and the Ewald sum was used to treat the long-range electrostatic interactions.[46] The reference structures for the water hexamers, with their associated many-body interaction energies calculated at the CCSD(T)/CCSD(T)-F12 level of theory within the SAMBA scheme[47] are taken from Ref. 48.

2.1 Hybrid Density Functionals

Hybrid density functionals are derived from the adiabatic connection formula:[49–53]

$$E_{xc}[\rho] = \int_0^1 E_{xc,\lambda} d\lambda \tag{2}$$

In eq 2, the electron-electron coupling goes through a continuum of partially interacting states, where the integration variable, λ, varies from 0 to 1, changing the system from a non-interacting Kohn-Sham reference system to a fully interacting system, with each intermediate state having the same ground-state electron density. The integrand in eq 2 is formally expressed as the exchange-correlation potential energy at a given λ,

$$E_{xc,\lambda} = \langle \Psi_\lambda | V_{ee} | \Psi_\lambda \rangle - \frac{e^2}{2} \int \int \frac{\rho(r)\rho(r')}{|r - r'|} dr \, dr' \tag{3}$$

At the non-interacting limit ($\lambda = 0$), the integrand becomes the exact exchange energy of the Kohn-Sham orbitals, while at the fully interacting limit the integrand is calculated using a density functional approximation (DFA). As discussed in the introduction, the integral in 2 is approximated by a linear interpolation of the $\lambda = 0$ and $\lambda = 1$ states, and becomes a fractional sum of the two.[27]

The most general formulation of this sum is given by

$$E_{xc} = \alpha E_x^{exact} + (1 - \alpha)E_x^{DFA} + E_c^{DFA} \tag{4}$$

where E_{xc}^{exact} is the exact exchange, and E_x^{DFA} and E_c^{DFA} are the DFA exchange and correlation terms. Eq. 4 can be rewritten as[29]

$$E_{xc} = \frac{1}{n}E_x^{exact} + \left(1 - \frac{1}{n}\right)E_x^{DFA} + E_c^{DFA} \tag{5}$$

for an integer n corresponding to the lowest order of Moller-Plesset perturbation theory needed to reproduce the atomization energies. Given the overall good performance of 4th order Moller-Plesset perturbation theory (MP4), PBE0 and SCAN0 use $n = 4$ (corresponding to α=0.25).[30,31] In this work, we refer to the hybrid SCAN functionals with varying fractions of exact exchange, α, as SCANα.

2.2 Explicit Many-Body Models: MB-pol and MB-DFT

MB-pol[54–56] and MB-DFT[43] are explicit many-body models derived from the many-body expansion (MBE) of the energy, which represents the energy of a system of N "bodies" (i.e., distinct atoms or molecules) as the sum of all the individual n-body contributions, with $n \leq N$.[57] The MBE is formally expressed as

$$E_N(r_1,..,r_N) = \sum_{i=1}^{N} \varepsilon_{1B}(r_i) + \sum_{i<j}^{N} \varepsilon_{2B}(r_i,r_j) + \sum_{i<j<k}^{N} \varepsilon_{3B}(r_i,r_j,r_k) + ... + \varepsilon_{NB}(r_1,..,r_N) \tag{6}$$

where ε_{1B} represents the distortion energy of an individual monomer from its equilibrium geometry, and each ε_{nB} with $n > 1$ represents the n-body energy which is defined recursively as

$$\varepsilon_{nB} = \varepsilon_n(1,...,n) - \sum_{i=1}^{N} \varepsilon_{1B}(r_i) - \sum_{i<j}^{N} \varepsilon_{2B}(r_i,r_j) - \sum_{i<j<k<...<n-1}^{N} \varepsilon_{(n\text{-}1)B}(r_i,r_j,..r_{n-1}). \tag{7}$$

The MBE converges quickly for atoms and molecules with localized electron densities and sufficiently large band gaps, such as water[47] and other molecular fluids.[58,59]

The MB-pol and MB-DFT models include explicit representations for the first three terms of the MBE, implicitly accounting for all higher order, $n > 3$, body terms via classical polarization,

$$E_N(r_1,..,r_N) = \sum_{i=1}^{N} \varepsilon_{IB}(r_i) + \sum_{i>j}^{N} \varepsilon_{2B}(r_i,r_j) + \sum_{i>j>k}^{N} \varepsilon_{3B}(r_i,r_j,r_k) + E_{pol} \tag{8}$$

The 1B term in both MB-pol and MB-DFT models is described by the Partridge-Schwenke potential energy surface calculated at the configuration interaction level of theory and subsequently refined to quantitatively reproduce the experimental rovibrational transitions of a water molecule.[60] The 2-body term consists of three contributions:

$$\varepsilon_{2B} = E_{2B}^{sr} + E_{elec} + E_{disp} \tag{9}$$

where E_{2B}^{sr} is a 4th-degree permutationally invariant polynomial (PIP)[61] representing the short-range interactions within a water dimer. E_{elec} is described by a Coulomb potential between geometry dependent point charges that reproduce the *ab initio* dipole moment surface of a water molecule.[60] Finally, the 2-body dispersion interaction is expressed as

$$E_{disp} = -\sum_{i,j} f(\delta_{ij}) \frac{C_{6,ij}}{R_{ij}^6} \tag{10}$$

where i and j are indices for two atoms on two separate water molecules, $f(\delta_{ij})$ is the Tang-Toennies damping function with a fitted parameter δ_{ij},[62] and $C_{6,ij}$ are the dispersion coefficients calculated from the asymptotic reference energies of the water dimer as originally introduced in the CC-pol model.[63] The explicit 3-body term,

$$\varepsilon_{3B} = E_{3B}^{sr} \tag{11}$$

7

describes short-range 3-body interactions and is represented by a 4th-degree PIP.[61] In MB-pol, the 2B and 3B PIPs are fitted on top of an implicit many-body polarizable potential (E_{pol} in Eq. 8) in order to reproduce 2B and 3B energies calculated at the CCSD(T) level of theory in the complete basis set (CBS) limit.[54,55] Similarly, the explicit ε_{2B} and ε_{3B} terms of the MB-DFT models are fitted to 2B and 3B energies calculated with a given DFA, using the same procedure described in ref 43.

In this study, we introduce a series of MB-DFT models for water, denoted as MB-SCANα, fitted to 2B and 3B energies calculated with the SCANα functionals, including different fractions (α) of exact exchange. The accuracy of each MB-SCANα model is assessed through a systematic analysis of the MBE, the structure of liquid water at ambient conditions, and the potential energy profile along the hydrogen bonding coordinate in the water dimer.

3   Results and Discussion

3.1   Many-Body Analysis of the Water Hexamer

As a first demonstration of the role played by exact exchange in determining the strength of the interactions between water molecules, figure 2 plots the errors relative to the CCSD(T)/CBS reference values which are associated with the n-body (with $n = 2 - 6$) energies calculated using the SCANα functionals for the first eight low-energy isomers of the water hexamer displayed in figure 1. Figure 3 shows the corresponding total interaction energies (panel a) and the relative interaction energies (panel b) as a function of α. The relative interaction energies are calculated by minimizing the total signed error with respect to the corresponding CCSD(T)/CBS reference values. By removing energy contributions due to monomer distortions, the analysis of the interaction energies allows for the direct comparison of the molecular interactions specific to each fraction of exact exchange. As discussed in section 2, the interaction energies were calculated at the SCANα/aug-cc-pVQZ level of theory with varying fractions of exact exchange corresponding to α=0.00 (pure SCAN functional), 0.05, 0.10, 0.15, 0.20, and 0.25 (hybrid SCAN0 functional).
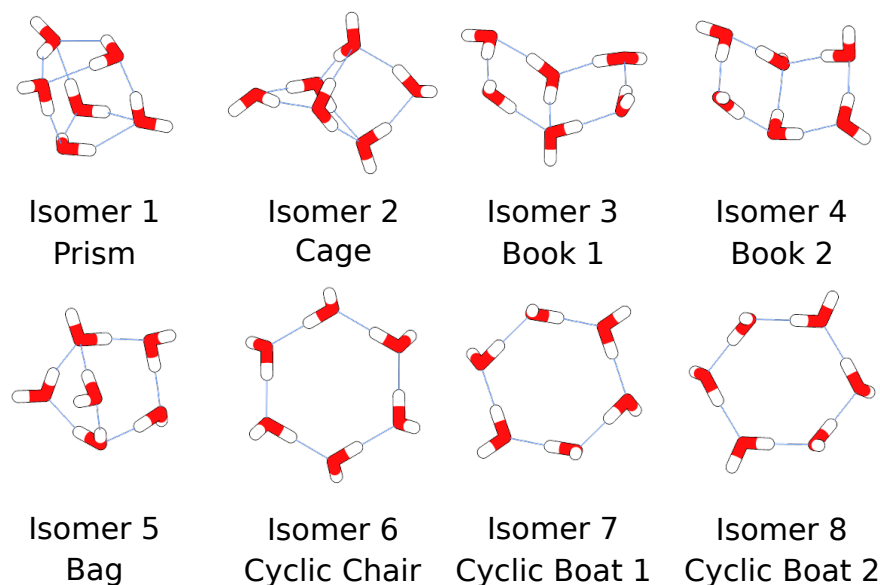
Figure 1: Structures of the first eight low-energy isomers of the water hexamer, $(H_2O)_6$.

In the analysis of the MBE for the prism isomer shown figure 2a, the 2B term calculated with α=0.00 exhibits an error of -5.74 kcal/mol. The error progressively decreases with increasing α, reaching a value of -4.96 kcal/mol at α=0.25. The 2B term for the cage isomer displays errors ranging from -5.69 kcal/mol for SCANα with α=0.00 to -4.98 kcal/mol for SCANα with α=0.25. Both book isomers display similar 2B errors of -4.83 and -4.90 kcal/mol for α=0.00, and reduced errors of -4.65 and -4.66 kcal/mol when α=0.25. Interestingly, the lowest errors (-4.63 and -4.65 kcal/mol for book-1 and book-2, respectively) are obtained when α=0.20. The bag isomer displays a 2B error of -4.95 kcal/mol when α=0.00, with an error of -4.63 kcal/mol when α=0.25. The cyclic isomers display opposite behavior as a function of α. Specifically, the 2B error in the cyclic chair is the smallest (-3.95 kcal/mol) at α=0.00 and the largest (-4.31 kcal/mol) at α=0.25. Similarly, the two cyclic boat isomers exhibit errors of -3.92 and -3.87 kcal/mol at α=0.00, and -4.22 and -4.19 kcal/mol at α=0.25. This general trend indicates that the accuracy of the 2B term in the SCANα functionals increases with α in isomers that display 3-dimensional hydrogen-bonding arrangements (i.e., the prism and cage isomers). The opposite trend is observed for ring-like structures (i.e., the cyclic isomers) for which the addition of any fraction of exact exchange deteriorates the accuracy of the 2B term relative to the pure SCAN functional.
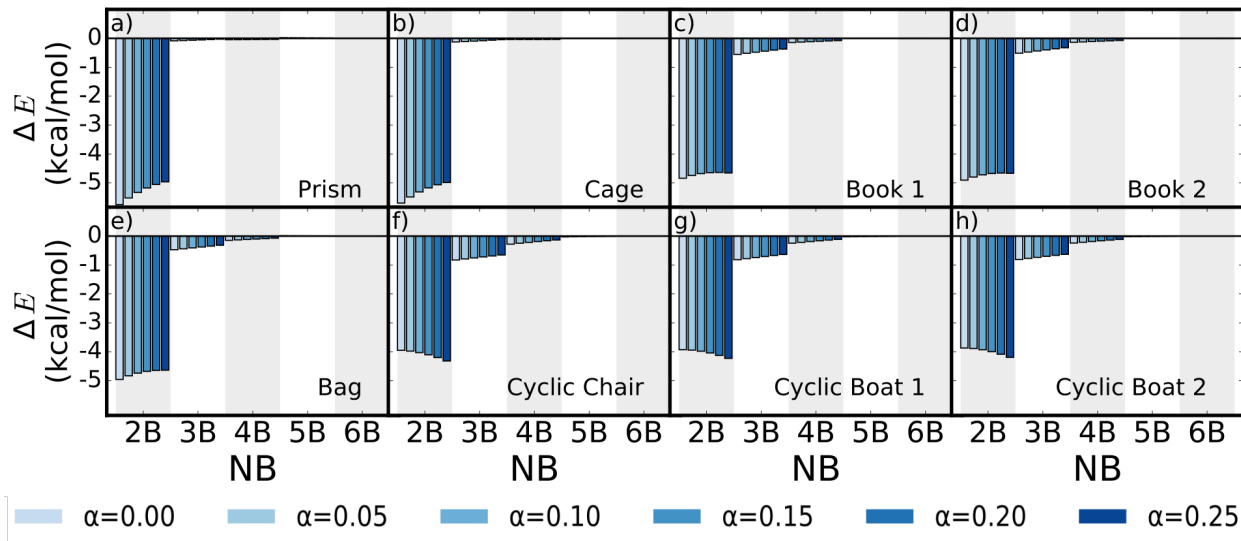
Figure 2: Errors in the individual terms of the MBE for the first eight low-energy isomers of the water hexamer calculated with the SCANα functionals relative to the CCSD(T)/CBS reference values of ref 48. The alternating grey and white shaded sectors separate each individual nB term of the MBE, with the corresponding errors associated with each SCANα functional shown as bars with varying shades of blue.

All SCANα functionals provide significantly smaller errors for all higher-body terms, which systematically decrease in magnitude as α increases. In particular, both the prism and cage isomers, which are characterized by fully 3-dimensional hydrogen-bonded structures, display maximum errors of -0.08 and -0.12 kcal/mol for the 3B term, and a similarly small maximum error of -0.04 kcal/mol for the 4B term at α=0.00. Relatively larger 3B and 4B errors (on the order of -0.5 kcal/mol and 0.15 kcal/mol, respectively) are found for the book-1, book-2, and bag isomers described by the pure SCAN functional, with errors decreasing as α increases. All cyclic isomers exhibit 3B and 4B errors of approximately -0.8 kcal/mol and -0.25 kcal/mol when $\alpha = 0.00$, with errors decreasing as the fraction of exact exchange increases from α=0.00 to α=0.25. For all isomers, independently of α, the 5B and 6B errors are always small, ranging from 10^{-4} to 10^{-2} kcal/mol.

This analysis highlights the limitations of the pure SCAN functional (α=0.00) in providing a balanced representation of 2B and higher-body energies, with the former being largely overestimated. Importantly, this "asymmetry" is not corrected by including any fraction of exact exchange
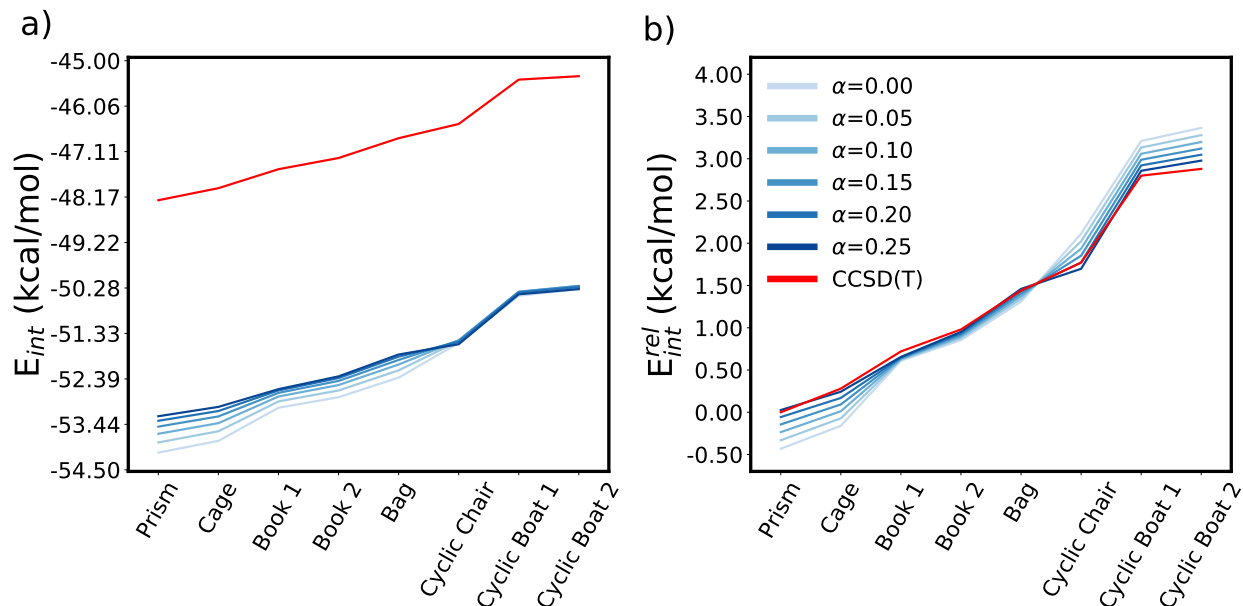
Figure 3: Total (panel a) and relative (panel b) interaction energies of the first eight low-energy isomers of the water hexamer calculated with the SCANα functionals. The CCSD(T)/CBS reference values are taken from ref 48. The relative interaction energies for each SCANα functional are aligned to the CCSD(T)/CBS reference values by minimizing the total signed error.

(up to $\alpha = 0.25$). In particular, while it is found to improve the description of 2B energies in the hexamer isomers with 3-dimensional hydrogen-bonding arrangements, the inclusion of any fraction of exact exchange deteriorates (approximately by the same amount) the description of 2B interactions in cyclic structures.

The errors in the individual terms of the MBE directly translate into the ability of the SCANα in predicting the relative stability of the hexamer isomers. With respect to the total interaction energies, figure 3a shows the total interaction energy for all isomers as a function of α. Starting from the prism isomer, the 3-dimensional isomers display substantially distinct interaction energies which become less negative with increasing α. Moving from the cage to the book and bag isomers, the interaction energies become progressively more clustered, and closer to the values obtained with $\alpha=0.25$. Once the 2-dimensional isomers (the cyclic chair and boat isomers) are reached, all SCANα functionals effectively predict the same interaction energy obtained with $\alpha=0.25$.

As shown in figure 3b, the SCANα functional with $\alpha=0.25$ predict relative interaction energies that closely follow the CCSD(T)/CBS reference values, while all other values of α display progres-

11

sively divergent curves as α decreases. Importantly, the SCANα functionals with $alpha < 0.25$ are found to progressively overbind the low-energy isomers and underbind the high-energy isomers. This lead the pure SCAN functional ($\alpha = 0.00$) to predict an energy difference of $\sim$4 kcal/mol between the prism and cyclic-boat-2 isomers, which is $\sim$1 kcal/mol larger than the corresponding CCSD(T)/CBS value.

The similarity of the total interaction energies between all values of α found in figure 3a for the three cyclic isomers may be explained by the opposing trend displayed by the corresponding 2B and higher-body terms. For these isomers, figure 2 shows that the lower error in the 2B term found for small values of α is countered by higher errors in the 3B and 4B terms, and vice versa as α increases. Finally, it should be noted that the 2B terms for the isomers with 3-dimensional structures show the largest variance in the associated errors, which correspondingly elicits the relatively wider distribution of interaction energies seen in figure 3b. These results suggest that the fraction of exact exchange included in the SCANα functionals acts primarily on the 2B term of the MBE, leaving all other (higher-order) terms less dependent on α.

3.2   Liquid Water

While the previous section examines the effects of modulating the fraction of exact exchange within the SCAN functional on the energetics of small water clusters, this section seeks to connect the differences amongst the SCANα functionals to the ability of these functionals to describe the structure of liquid water at ambient conditions. For this purpose, MD simulations were carried out at 298.15 K and 1 atm in the isothermal-isobaric (NPT) ensemble with the MB-SCANα models fitted to the corresponding SCANα 2B and 3B energies. Based on the analysis in the previous section, as the errors in the SCANα 4B and higher terms are effectively equivalent to those associated with MB-pol,[48] the MB-SCANα models provide an accurate representation of the SCANα functionals, despite only the corresponding 2B and 3B energies are explicitly used in their parameterization (see Figure S1 in the Supporting Information). Following ref 43, three distinct MB-SCANα models are constructed for each SCANα functional by progressively replacing the

12

MB-pol 2B and 3B terms with the SCANα counterparts. The MB-SCANα models are thus labeled as (2B+3B)-MB-SCANα, (2B)-MB-SCANα, and (3B)-MB-SCANα to indicate that both the 2B and 3B, only the 2B, and only the 3B terms are represented by fits to the corresponding SCANα 2B and 3B energies, with all other terms being represented as in MB-pol. As demonstrated in ref 43, the MB-DFT models are able to reproduce, at a fraction of the associated computational cost, the results obtained from the corresponding *ab initio* MD simulations.

Figure 4 displays the oxygen-oxygen (O-O) radial distribution functions (RDFs) calculated with the (2B+3B)-MB-SCANα (panel a), (2B)-MB-SCANα (panel b), and (3B)-MB-SCANα (panel c) models. All (2B+3B)-MB-SCANα models (panel a) predict O-O RDFs that are systematically shifted (by ~0.3Å) to the left of the corresponding experimental curve,[64,65] indicating overly attractive interactions between the water molecules. This is a direct consequence of all SCANα functionals overestimating, especially at the 2B level, the strength of the low-order many-body interactions between water molecules as shown in figures 2 and 3. Besides this common feature, significant variation exists in the ability of the different (2B+3B)-MB-SCANα models to reproduce the experimental O-O RDF. In particular, the (2B+3B)-MB-SCANα model with α=0.00 predicts a significantly less structured second solvation shell. As α increases, the (2B+3B)-MB-SCANα models predict an increasingly more structured liquid. In this context, a value of α=0.15 appears to be the "optimal" compromise, leading to an O-O RDF that is in closer agreement with
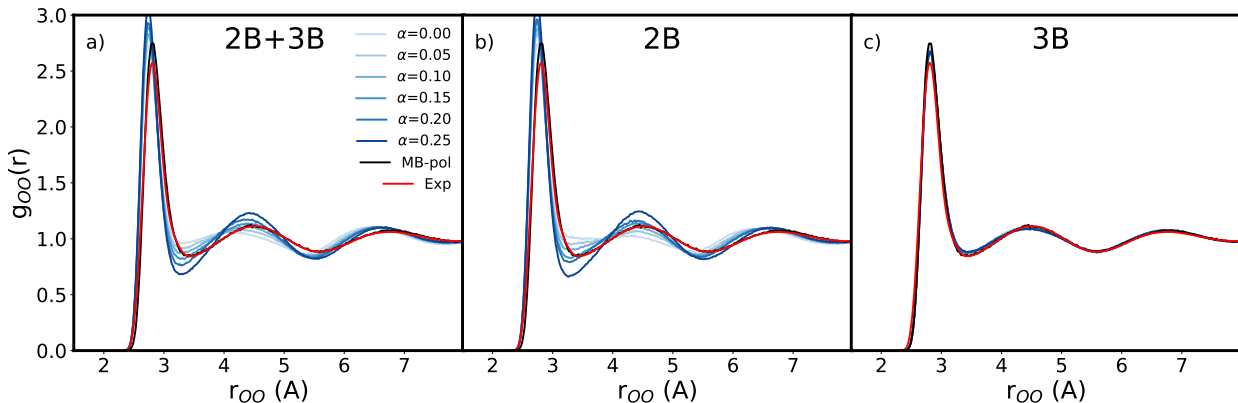

Figure 4: O-O RDFs calculated from NPT MD simulations carried out at 298 K and 1 atm using the (2B+3B)-MB-SCANα (panel a), (2B)-MB-SCANα (panel b), and (3B)-MB-SCANα (panel c) models. The experimental RDFs at 295 K are taken from refs 64 and 65.

the experimental curve, although the peaks of the solvation shells are consistently higher and the "valleys" between consecutive shells are consistently deeper than in the experimental RDF.

The general trend observed in the O-O RDFs calculated with the (2B+3B)-MB-SCANα models is also reproduced by the O-O RDFs calculated with the (2B)-MB-SCANα models (figure 4b), where only the 2B term of MB-pol is replaced with the 2B term fitted to the corresponding SCANα 2B energies. On the other hand, all (3B)-MB-SCANα models, which replace only the 3B term of MB-pol with the corresponding SCANα 3B, closely reproduce the experimental O-O RDF (figure 4c), effectively achieving the same accuracy of the MB-pol model. The close correspondence between the O-O RDFs calculated with the (2B+3B)-MB-SCANα and (2B)-MB-SCANα models combined with the agreement between the (3B)-MB-SCANα and experimental RDFs provide clear evidence that the inability of the pure SCAN functional ($\alpha = 0.00$) to reproduce the shape of the experimental RDF is primarily due to intrinsic deficiencies of the SCAN functional in correctly predicting the 2B energies as shown in Figure 2. Similar conclusions can be drawn from the analyses of the O-H and H-H RDFs reported in the Supporting Information.

The differences in the O-O RDFs calculated with the three sets of MB-SCANα models directly translate in the differences seen in the corresponding O-O cumulative distribution functions (CDFs) shown in Figure 5 and densities listed in Table 1. In particular, both the (2B+3B)-MB-SCANα and
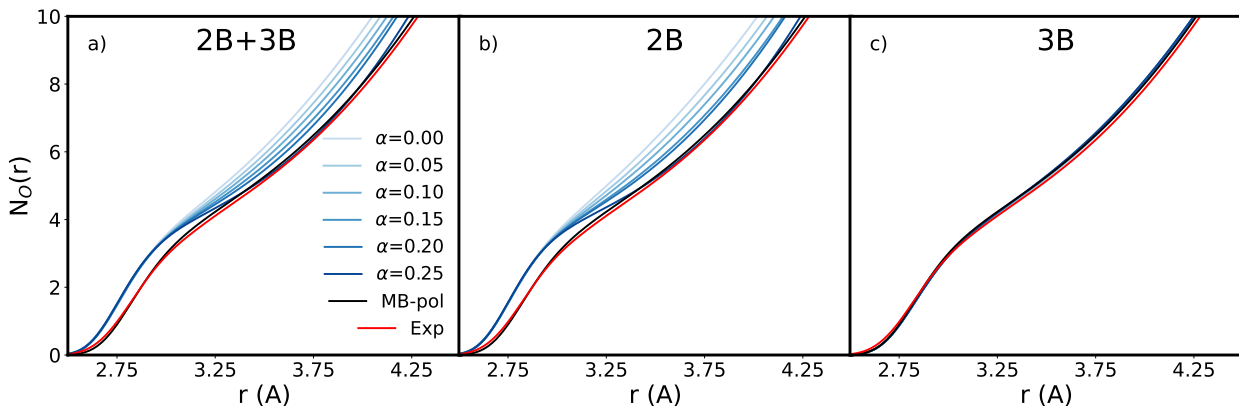

Figure 5: O-O CDFs calculated from NPT MD simulations carried out at 298 K and 1 atm using the (2B+3B)-MB-SCANα (panel a), (2B)-MB-SCANα (panel b), and (3B)-MB-SCANα (panel c) models. The experimental RDFs at 295 K are taken from refs 64 and 65, and 66 at 295K.

Table 1: Density (in g/cm^3) of liquid water calculated from NPT MD simulations carried out at 298 K and 1 atm using the three sets of MB-SCANα models. As a reference, the experimental density is 0.997 g/cm^3, while the MB-pol density is 1.007 g/cm^3.[48]

α	(2B+3B)-MB-SCANα	(2B)-MB-SCANα	(3B)-MB-SCANα
0.00	1.14	1.16	1.00
0.05	1.12	1.14	1.00
0.10	1.10	1.12	1.00
0.15	1.08	1.09	1.01
0.20	1.07	1.08	1.01
0.25	1.04	1.04	1.01

(2B)-MB-SCANα models predict CDFs that deviate significantly from the experimental and MB-pol curves, systematically overestimating the coordination number between 2.5 Åand 4.5 Å. On the contrary, all (3B)-MB-SCANα models are effectively indistinguishable from each other and in quantitative agreement with the experimental and MB-pol curves. As a consequence, both the (2B+3B)-MB-SCANα and (2B)-MB-SCANα models predict significantly higher densities, which systematically decrease as the fraction of exact exchange increases. On the other hand, all (3B)-MB-SCANα models predict densities that are in quantitative agreement with the experimental value of 0.997 g/cm^3, independently of α.

It should be noted that, while the density calculated with the (2B+3B)-MB-SCANα model with $\alpha = 0.00$ is in line with the value obtained in ref 23 using *ab initio* MD simulations carried out with the pure SCAN functional at 330 K, it differs from the value of 1.00 g/cm^3 that has been reported from MD simulations carried out with a DNN model trained on SCAN data.[37] In this context, it has recently been shown that DNN models based on short-range representations display some limitations in describing the underlying molecular interactions.[67] While the neglect of long-range interactions in the DNN model of ref 37 may be a possible reason for the differences with the (2B+3B)-MB-SCANα model with $\alpha = 0.00$, further investigations are needed to systematically determine both differences and similarities between DNN and MB-DFT models.

While the RDFs report on the distributions of inter-atomic distances between pairs of atoms, direct insights into the 3-dimensional arrangement of the water molecules in the liquiid can be

gained from the analysis of the tetrahedral order parameter, q_{tet}, defined as[68]

$$q_{tet} = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(cos(\psi_{jk}) + \frac{1}{3} \right)^2 \quad (12)$$

Here, ψ_{jk} is the angle between the O atom of the central water molecule and the O atoms of two neighboring water molecules. A q_{tet} value of 1 indicates a perfectly tetrahedral arrangement, while smaller values correspond to more disordered arrangements.

The analysis of q_{tet} calculated with the (2B+3B)-MB-SCANα models and shown in figure 6a demonstrates that these models predict a progressively more tetrahedral structure as α increases, as indicated by leftward shift and heightening of the main peak at $q_{tet} \sim 0.75$. For small values of α, the observed larger values of q_{tet} relative to MB-pol can be explained by the larger coordination numbers predicted by the MB-SCANα models (figure 5a) and the higher densities (up to 1.14 g/cm^3 for $\alpha = 0.00$) reported in Table 1. Since, by definition, q_{tet} only consider the four nearest water molecules, without accounting for any additional molecule in the solvation shell, the q_{tet} distribution for the MB-SCANα model with $\alpha = 0.00$, which corresponds to the pure SCAN functional, indicates a more compact, but geometrically ice-like structure of liquid water at ambient temperature. The results shown in figure 5a also indicate that, as α increases, the (2B+3B)-MB-SCANα models approach a less compact structure, which is accompanied by a decrease in density
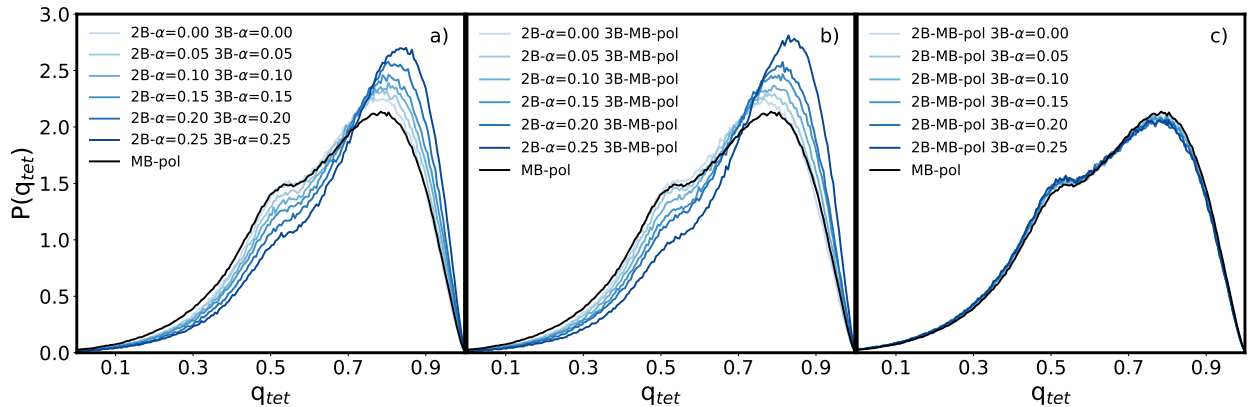

Figure 6: Distributions of the tetrahedral order parameter, q_{tet}, calculated from NPT MD simulations carried out at 298 K and 1 atm using the (2B+3B)-MB-SCANα (panel a), (2B)-MB-SCANα (panel b), and (3B)-MB-SCANα (panel c) models.

16

and, at the same time, a more pronounced ice-like signature. These results are in line with MD simulations carried out with a DNN model trained on SCAN data which predicts the melting point of SCAN to be $\sim$310 K.[37]

While an analogous trend is predicted by the (2B)-MB-SCANα models (figure 6b), the (3B)-MB-SCANα models (figure 6c) closely reproduce the MB-pol distribution. The close correspondence between the q_{tet} distributions for the (2B+3B)-MB-SCANα and (2B)-MB-SCANα models, along with the agreement between the (3B)-MB-SCANα and the MB-pol data further suggests that the compact, ice-like structure predicted by the (2B+3B)-MB-SCANα models with $\alpha = 0.00$ is rooted in the inability of the SCAN functional to correctly reproduce the 2B energies of the MBE for water (Figure 2).

3.3   Hydrogen Bonding in the Water Dimer

In an attempt to rationalize the differences seen in the O-O RDF, density, and tetrahedral order parameter calculated with the three sets of MB-SCANα models, this section examines the ability of the SCANα functionals to describe hydrogen bonding in the water dimer. To this purpose, Figure 7 shows the deviations, $\Delta\tilde{\nu}$, associated with the harmonic frequencies of the water monomer and dimer calculated with the SCANα functionals relative to the reference CCSD(T):MP2 harmonic frequencies reported in refs 69 and 70. This analysis indicates that the deviations associated with the OH harmonic frequencies (marked in blue) of both H_2O and $(H_2O)_2$ systematically move from negative to positive values as α increases In particular, for the water monomer, the pure SCAN functional (α=0.00) predicts a redshift of $\sim$30 cm^{-1}, while a blueshift of $\sim$75 cm^{-1} is associated with SCAN0 (α=0.25). While the same shifts are effectively found for harmonic vibrations involving the free OH bonds in the water dimer, the harmonic frequency of the hydrogen bonded OH (labeled as OH* in the figure) is appreciably underestimated by the pure SCAN functional, resulting in $\Delta\tilde{\nu} = 75$ cm^{-1}. This redshift, indicative of a relatively stronger hydrogen bond compared to the reference CCSD(T):MP2 value, directly correlates with the higher liquid density obtained from NPT simulations carried out with the (2B+3B)-MB-SCANα model with $\alpha = 0.00$. The
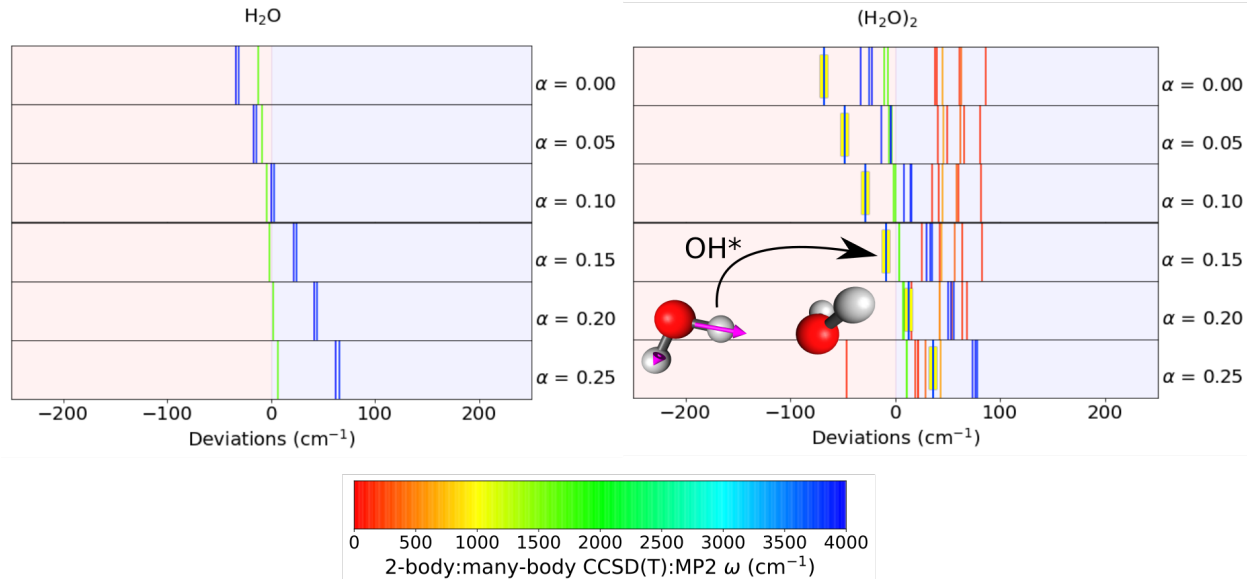
Figure 7: Deviations from reference 2-body:many-body harmonic frequencies of $(H_2O)_n$ clusters, with $n = 1 - 2$,[69,70] calculated using α=0.00-0.25. The light red and blue backgrounds correspond to red and blue shifts relative to the reference values. Each harmonic mode, indicated by a stick line, is colored according to the associated frequency according to the colorbar displayed at the bottom of the figure. For each value of α the harmonic frequency for the hydrogen bonded OH is highlighted with a yellow box, as indicated by the inset graphic.

progressive blueshift observed in the harmonic frequency of the hydrogen-bonded OH parallels the decrease in the liquid density predicted by the corresponding (2B+3B)-MB-SCANα models. Importantly, Figure 7 shows that the closest agreement between the CCSD(T):MP2 and SCANα harmonic frequency of the hydrogen-bonded OH is obtained with $\alpha = 0.15$, which is the same fraction of exact exchange that provides the closet agreement with the experimental RDFs in the analysis of the (2B+3B)-MB-SCANα models (Figure 4).

4   Conclusion

In this study, we presented a systematic analysis of the role played by the exact exchange in describing the water interactions by investigating the accuracy of a series of SCANα functionals, with varying fractions (α) of exact exchange, which are constructed through the adiabatic connection formula. Our analysis demonstrates that all SCANα functionals are unable to correctly

represent the 2B energies between water molecules, with errors relative to the CCSD(T)/CBS reference values that are on the order of 4-5 kcal/mol. Importantly, we show that the inclusion of exact exchange has opposite effects on the ability of the SCANα functionals to describe the interaction energies of water clusters with 2-dimensional and 3-dimensional hydrogen-bonding arrangements. Specifically, while the errors relative to the CCSD(T)/CBS reference values decrease going from $\alpha = 0.00$ (pure SCAN functional) to $\alpha = 0.25$ (hybrid SCAN0 functional) for the prism and cage isomers of the water hexamer, the opposite trend is found for the planar isomers (i.e., cyclic chair and cyclic boat isomers). These deficiencies are found to directly related to the ability of the SCANα functionals to describe the structure of liquid water at ambient conditions, which is investigated through MD simulations carried out with the corresponding MB-SCANα models. In particular, we have found that the MB-SCANα model with $\alpha = 0.00$, which corresponds to the pure SCAN functional, predicts a significantly denser liquid, with a more compact first hydration shell. On the other hand, the MB-SCANα model with $\alpha = 0.25$, which corresponds to the hybrid SCAN0 functional, predicts a more ice-like structure. These differences can be explained by the different accuracy with which the SCANα functionals describe hydrogen bonding and the individual many-body contributions to the interaction energies at the fundamental level as determined from the analysis of the harmoniic frequencies of the water dimer and the energetics of the hexamer isomers. The analysis of the dimer and hexamer clusters and the structure of liquid water at ambient conditions indicates that the SCANα functional with α=0.15 provides the closer agreement with the CCSD(T)/CBS reference energies and experimental O-O RDFs. However, it should be noted that "optimal" MB-SCANα model with $\alpha = 0.25$ still provides a more attractive O-O RDF, which results in an appreciably higher density, although the overall structure of liquid water is predicted to be more ice-like. These apparent opposite trends can be explained by the inability of all SCANα functionals to provide a balanced description of both 2B and 3B energies. While these findings suggest that some caution should be exercised when the SCAN functional are used in computer simulations of liquid water, ice, and other aqueous solutions,[23,26,35,37,38,71] they also provide fundamental insights in both merits and shortcomings of the SCAN functional which can

guide future theoretical developments.[32,72]

Acknowledgement

Supporting Information Available

Analyses of the O-O, O-H, and H-H radial distribution functions, as well as a comparison between MB-SCAN and an AIMD SCAN simulation from Ref. 26 in the NVT ensemble.

References

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

(2) Tse, J. S. Ab Initio Molecular Dynamics With Density Functional Theory. *Annu. Rev. Phys. Chem.* **2002**, *53*, 249–290.

(3) Carloni, P.; Rothlisberger, U.; Parrinello, M. The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems. *Acc. Chem. Res.* **2002**, *35*, 455–464.

(4) Hassanali, A. A.; Cuny, J.; Verdolino, V.; Parrinello, M. Aqueous Solutions: State of the Art in Ab Initio Molecular Dynamics. *Philos. Trans. Royal Soc. A* **2014**, *372*, 20120482.

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

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

(7) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press, 1994.

(8) Slater, J. C. A Simplification of the Hartree-Fock Method. *Phys. Rev.* **1951**, *81*, 385.

(9) Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140*, 18A301.

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

(11) 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.

(12) Herman, F.; Van Dyke, J. P.; Ortenburger, I. B. Improved Statistical Exchange Approximation for Inhomogeneous Many-Electron Systems. *Phys. Rev. Lett.* **1969**, *22*, 807.

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

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

(15) Forster-Tonigold, K.; Groß, A. Dispersion Corrected RPBE Studies of Liquid Water. *J. Chem. Phys.* **2014**, *141*, 064501.

(16) Perdew, J. P. Accurate Density Functional for the Energy: Real-Space Cutoff of the Gradient Expansion for the Exchange Hole. *Phys. Rev. Lett.* **1985**, *55*, 1665.

(17) Ghosh, S. K.; Parr, R. G. Phase-Space Approach to the Exchange-Energy Functional of Density-Functional Theory. *Phys. Rev. A* **1986**, *34*, 785.

(18) Van Voorhis, T.; Scuseria, G. E. A Novel Form for the Exchange-Correlation Energy Functional. *J. Chem. Phys.* **1998**, *109*, 400–410.

(19) Proynov, E.; Ruiz, E.; Vela, A.; Salahub, D. Determining and Extending the Domain of Exchange and Correlation Functionals. *Int. J. Quantum Chem.* **1995**, *56*, 61–78.

(20) Becke, A. D. A New Inhomogeneity Parameter in Density-Functional Theory. *J. Chem. Phys.* **1998**, *109*, 2092–2098.

(21) Mardirossian, N.; Head-Gordon, M. Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97m-V. *J. Chem. Phys.* **2015**, *142*, 074111.

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

(23) 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., et al. Ab Initio Theory and Modeling of Water. *Proc. Natl. Acad. Sci.* **2017**, *114*, 10846–10851.

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

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

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

(27) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(28) Becke, A. Density-Functional Thermochemistry. III. The Role of Exact Exchange Year. *J. Chem. Phys 98*, 5648.

(29) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for Mixing Exact Exchange With Density Functional Approximations. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

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

(31) Hui, K.; Chai, J.-D. SCAN-based Hybrid and Double-Hybrid Density Functionals From Models Without Fitted Parameters. *J. Chem. Phys.* **2016**, *144*, 044114.

(32) 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.* **2020**, *117*, 11283–11288.

(33) Zhao, G.; Shi, S.; Xie, H.; Xu, Q.; Ding, M.; Zhao, X.; Yan, J.; Wang, D. Equation of State of Water Based on the SCAN Meta-Gga Density Functional. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4626–4631.

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

(35) Xu, J.; Zhang, C.; Zhang, L.; Chen, M.; Santra, B.; Wu, X. Isotope Effects on Molecular Structures and Electronic Properties of Liquid Water via Deep Potential Molecular Dynamics Based on SCAN Functional. *Phys. Rev. B* **2020**, *102*, 214113.

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

(37) Piaggi, P. M.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. Phase Equilibrium of Water With Hexagonal and Cubic Ice Using the SCAN Functional. *arXiv:2101.04806* **2021**,

(38) Piaggi, P. M.; Car, R. Enhancing the Formation of Ionic Defects to Study the Ice Ih/Xi Transition With Molecular Dynamics Simulations. *arXiv:2101.09308* **2021**,

(39) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kús, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock III, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio Jr., R. A.; Dop, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Pieniazek, P. A.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sergueev, N.; Sharada, S. M.; Sharmaa, S.;

Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Vanovschi, V.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhou, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard III, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer III, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xua, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184–215.

(40) 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.

(41) 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.

(42) Smith, W.; Forester, T. DL_POLY_2. 0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graph.* **1996**, *14*, 136–141.

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

(44) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(45) Martyna, G. J.; Hughes, A.; Tuckerman, M. E. Molecular Dynamics Algorithms for Path Integrals at Constant Pressure. *J. Chem. Phys.* **1999**, *110*, 3275–3290.

(46) Allen, M.; Tildesley, D. *Computer Simulations of Liquids*; Oxford University Press, 1987.

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

(48) 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.

(49) Harris, J.; Jones, R. The Surface Energy of a Bounded Electron Gas. *J. Phys. F Met. Phys.* **1974**, *4*, 1170.

(50) Langreth, D. C.; Perdew, J. P. The Exchange-Correlation Energy of a Metallic Surface. *Solid State Commun.* **1975**, *17*, 1425–1429.

(51) Gunnarsson, O.; Lundqvist, B. I. Exchange and Correlation in Atoms Molecules and Solids by the Spin-Density-Functional Formalism. *Phys. Rev. B* **1976**, *13*, 4274.

(52) Langreth, D. C.; Perdew, J. P. Exchange-Correlation Energy of a Metallic Surface: Wave-Vector Analysis. *Phys. Rev. B* **1977**, *15*, 2884.

(53) Harris, J. Adiabatic-Connection Approach to Kohn-Sham Theory. *Phys. Rev. A* **1984**, *29*, 1648.

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

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

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

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

(58) Riera, M.; Yeh, E. P.; Paesani, F. Data-Driven Many-Body Models for Molecular Fluids: CO_2/H_2O Mixtures as a Case Study. *J. Chem. Theory Comput.* **2020**, *16*, 2246–2257.

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

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

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

(62) Tang, K.; Toennies, J. P. An Improved Simple Model for the Van Der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

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

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

(65) Skinner, L. B.; Benmore, C.; Neuefeind, J. C.; Parise, J. B. The Structure of Water Around the Compressibility Minimum. *J. Chem. Phys.* **2014**, *141*, 214507.

(66) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535.

(67) Yue, S.; Muniz, M. C.; Calegari Andrade, M. F.; Zhang, L.; Car, R.; Panagiotopoulos, A. Z. When Do Short-Range Atomistic Machine-Learning Models Fall Short? *J. Chem. Phys.* **2021**, *154*, 034111.

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

(69) Howard, J. C.; Gray, J. L.; Hardwick, A. J.; Nguyen, L. T.; Tschumper, G. S. Getting Down to the Fundamentals of Hydrogen Bonding: Anharmonic Vibrational Frequencies of $(HF)_2$ and $(H_2O)_2$ From Ab Initio Electronic Structure Computations. *J. Chem. Theory Comput.* **2014**, *10*, 5426–5435.

(70) Howard, J. C.; Tschumper, G. S. Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters:$(H_2O)_n$. *J. Chem. Theory Comput.* **2015**, *11*, 2126–2136.

(71) Xu, J.; Sun, Z.; Zhang, C.; DelloStritto, M.; Lu, D.; Klein, M. L.; Wu, X. Importance of Nuclear Quantum Effects on the Hydration of Chloride Ion. *Phys. Rev. Mater.* **2021**, *5*, L012801.

(72) Wagle, K.; Santra, B.; Bhattarai, P.; Shahi, C.; Pederson, M. R.; Jackson, K. A.; Perdew, J. P. Self-Interaction Correction in Water-Ion Clusters. *arXiv:2012.13469* **2020**,

(73) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(74) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

TOC Figure

$$E_{xc}[\rho] = \int_0^1 E_{xc,\lambda} d\lambda \approx \alpha E_x^{exact} + (1-\alpha) E_x^{DFA} + E_c^{DFA}$$