Data-Driven Many-Body Potentials from Density Functional Theory for Aqueous Phase Chemistry

Etienne Palos,[1, a)] Saswata Dasgupta,[1] Eleftherios Lambros,[1] and Francesco Paesani[1, 2, 3, b)]

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

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

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

Density functional theory (DFT) has been applied to modeling molecular interactions in water for over three decades. The ubiquity of water in chemical and biological processes demands a unified understanding of its physics, from the single-molecule to the thermodynamic limit and everything in between. Recent advances in the development of data-driven and machine-learning potentials have accelerated simulation of water and aqueous systems with DFT accuracy. However, the anomalous properties of water in the condensed phase, where a rigorous treatment of both local and non-local many-body interactions is in order, is often unsatisfactory or partially missing in DFT models of water. In this review, we discuss the modeling of water and aqueous systems based on DFT, and provide a comprehensive description of a general theoretical/computational framework for the development of data-driven many-body potentials from DFT reference data. This framework, coined MB-DFT, readily enables efficient many-body molecular dynamics (MB-MD) simulations of small molecules, in both gas and condensed phases, while preserving the accuracy of the underlying DFT model. Theoretical considerations are emphasized, including the role that the delocalization error plays in MB-DFT potentials of water, and the possibility to elevate DFT and MB-DFT to near-chemical-accuracy through a density-corrected formalism. The development of the MB-DFT framework is described in detail, along with its application in MB-MD simulations and recent extension to the modeling of reactive processes in solution within a quantum mechanics/many-body molecular mechanics (QM/MB-MM) scheme, using water as a prototypical solvent. Finally, we identify open challenges and discuss future directions for MB-DFT and QM/MB-MM simulations in condensed phases.

[a]Electronic mail: epalos@ucsd.edu
[b]Electronic mail: fpaesani@ucsd.edu

CONTENTS

I. INTRODUCTION

Water, often referred to as "the matrix of life",[1] is a substance with a simple molecular structure that has been present on Earth for approximately four billion years.[2] However, it is only in this past half-century that progress has been made toward a fundamental understanding of the physical properties of water, with microscopic detail.[3–13] Theorists worldwide have tackled "obtaining" water vapor, liquid water, and ice from computer models founded on classical mechanics,[14,15] quantum mechanics,[16–18], or hybrid approaches that combine the former to some extent.[19] The first numerical simulations of water employed Monte Carlo (MC)[20] and molecular dynamics (MD)[14] techniques, and represented water as a system of point charges interacting through pairwise additive potentials.[21] While these simulations represent a cornerstone in the history of computer modeling of water, it became apparent that pairwise additive potentials alone would not be able to fully capture the complex physics of water in the condensed phase.[4]

In ice, hydrogen bonds arrange into low-entropy configurations that give rise to the formation of a highly symmetrical tetrahedral network.[22] However, in the liquid state, thermal fluctuations that occur within the picosecond to nanosecond timescale[23,24] are the driving force for the recurring formation and dissolution of hydrogen bonds that makes liquid water topologically disordered.[25] From a molecular standpoint, this implies that predicting the properties of water through computer simulations requires the ability to rationalize the interplay between the energetic and entropic contributions that mold the free-energy landscape.[3] This may be achieved provided there is a rigorous representation of many-body interactions that shape the free-energy landscape and determine the local structure of liquid water, along with all thermodynamic and dynamical properties.[4]

Within density functional theory (DFT),[26] the quantum-mechanical electronic ground state of a water molecule is stored in a three-dimensional particle density, as opposed to a thirty-dimensional function required by wavefunction theories (WFT).[27] A $\sim \mathscr{O}(N^3)$ scaling in combination with reasonable accuracy has consolidated DFT as the most widely used methodology in the computational physical sciences for the modeling of complex systems.[28,29] Nevertheless the universal density functional is unknown, which has motivated a three-decade long navigation in the vast sea of density functional approximations (DFAs), searching for a DFT representation of water that can accurately predict the properties of water from the single molecule to the thermodynamic limit. Since the 1980s, many have contributed to our understanding of water within DFT and DFT-based *ab initio* molecular dynamics (AIMD) simulations.[16–18,30–52] Beginning on the first rung of John Perdew's "Jacob's ladder",[53] the local density approximation (LDA),[26,54,55] systematically predicts too strong and, consequently, too short hydrogen bonds that preclude a realistic description of the structure of liquid water.[16] Subsequently, it was found that DFAs defined within the general gradient approximation (GGA), such as the Perdew-Burke-Erzerhoff (PBE) functional,[56] yield more accurate energetics for water clusters but ultimately still predict an over-structuring of liquid water.[16–18] The systematic over-binding predicted by the LDA and GGA functionals originates from their inability to exactly cancel self-Coulomb and self-exchange-correlation effects, an error known as the self-interaction error (SIE).[57] However, GGA exchange-correlation functionals continue to have strong presence in water simulations, with two popular DFAs being the empirical Becke-Lee-Yang-Parr (BLYP) functional and the revised Perdew-Burke-Ernzerhof (revPBE) functional, in combination with corrections for van der Waals forces.[58–68] While dispersion corrections can improve the accuracy of these functionals for water, it can often also amplify error cancellation[69–71] and lead to poor descriptions of water across different phases.

Exchange-correlation functionals that include a dependence on both the gradient of the electron density and the kinetic energy density, dubbed meta-GGA functionals,[72] can provide an improved description of water clusters in terms of computed energies and geometries, leading to qualitatively reasonable descriptions of the structure of liquid water at ambient conditions, without the higher computational cost of orbital dependent DFAs.[48] However, not all is resolved by meta-GGA DFAs, as classical and path-integral AIMD simulations of water performed using the B97M-rV functional with non-local correlation[73] systematically leads to an overly-repulsive local structure of liquid water, tethered by weaker hydrogen bonds as evidenced by a less-structured oxygen-oxygen radial distribution function (RDF).[50,52] Nonetheless, a recent assessment of over 200 functionals suggests that B97M-rV is currently the most accurate meta-GGA functional on average for diversely bonded systems,[74] exhibiting an overall accuracy for water that is comparable to that of the more expensive revPBE0-D3, a dispersion-corrected hybrid functional containing 25% Hartree-Fock exchange.[52]

Hybrid DFAs that contain a fraction of Hartree-Fock exchange have become quite popular in *ab initio* calculations due to their improved accuracy relative to GGA functionals.[75–78] Several popular hybrid functionals have been used in recent years to study molecular interactions in water,[48,52,79–86] focusing on the transferability of accuracy from static to dynamic properties,[80] as well the modulation of Hartree-Fock exchange and van der Waals corrections to probe their effects on the predicted hydrogen-bonding network in liquid water.[48] Hybrid functionals belong to the 4$^{\text{th}}$ rung of Jacob's ladder, while double-hybrid functionals define the fifth rung, where a fraction of second-order perturbative correlation is included in addition to the fraction of Hartree-Fock exchange.[87] While a few electronic properties of water have been reported using double-hybrids,[88–90] their $\sim \mathcal{O}(N^5)$-scaling deems them currently intractable for AIMD simulations of condensed-phase systems.

The quest for a physically robust and efficient description of water from DFT simulations has resulted in the continuing rise in popularity of the meta-GGA Strongly Constrained and Appropriately Normed (SCAN) functional[51,91] along with its hybrid,[92] regularized,[93] regularized-and-restored relatives,[94] and associated variants.[95,96] SCAN is a non-empirical density functional that was derived to satisfy the 17 exact constraints known for a meta-GGA functional.[91,97] Although it is semi-local in nature, SCAN is capable of capturing short- to medium-range dispersion interactions.[51,70,71,91] For this reason, SCAN has been found to provide a reasonable description of both gas-phase water clusters and liquid water, albeit still quantitatively overestimating hydrogen-bond strengths even when approximating the inclusion of nuclear quantum effects.[51] In an effort to complete the *density functional theory of water*, several studies have been reported focusing on assessing the accuracy of SCAN, as well as SCAN-based data-driven potentials (DDPs) and machine-learned potentials (MLPs) for various properties of water,[70,98–102] and ionic solutions.[103,104]

In this review, we focus on discussing data-driven many-body (MB) potentials for applications in aqueous phase chemistry. While we note that several classes of DFT-based DDPs have been proposed in recent years for water and various aqueous systems, we restrict the main discussion to a class of DFT-based DDPs that are rigorously derived from the many-body expansion of the energy (MBE), a theoretical/computational framework coined MB-DFT.

II. DATA-DRIVEN MODELING OF WATER AND AQUEOUS SYSTEMS

Data-driven modeling plays an increasingly important role in theoretical and computational chemistry for its promise of accelerating the prediction of physical properties with a quantum-mechanical accuracy,[105–117] with one of the main applications being the development of interatomic and intermolecular potentials.[112,118–144] In the context of potentials for condensed-phase simulations, DDPs and MLPs with complex analytical forms are flexible enough to effectively capture the intricate interactions between molecules at the accuracy of the underlying *ab initio* reference data.[145] Regarding water and aqueous systems, such models are broadly divided into two distinct categories: A) models whose entire representation is machine learned,[99,100,102,141,143,146–161] and B) data-driven physically-motivated that integrate a machine-learned representation with an underlying physical model.[144,162–172]

A. Machine-learned potentials

Over the past fifteen years, high-dimensional machine-learned models have been developed representing molecular interactions in water through neural network potentials (NNPs),[99,118–121,141,143,145,155,173–176] gaussian approximation potentials (GAPs),[105] permutationally invariant polynomials (PIPs).[147–150,163–166,177–179] In general, MLPs trained using various regression algorithms rely on using a set of descriptors which account for the immediate environment around a molecule in order to describe the potential energy surface.[119,121,125,126,180–182] Because these models are strongly dependent on the dataset to which they are trained on, they are typically limited in their transferability across different phases.[141,151,152,154,159,183–187] On the one hand, MLPs trained purely on gas-phase data have difficulties in properly describing condensed-phase systems where long-range and many-body effects play significant roles.[141,188] Recent studies have proposed avenues to overcome the challenge of long-range physics in NNPs.[151,152,154,183–185,187,189–192] For condensed-phase systems, however, NNPs are usually trained on configurations of liquid water extracted from AIMD simulations and, consequently, are able to provide an effective description of long-range interactions within the bounds of the simulation box. In addition, such condensed phase MLPs currently account for many-body effects only in an implicit fashion and, therefore, are not guaranteed to correctly reproduce each individual n-body contribution to the energy of a system containing N water molecules.[161] As the field of MLPs, and particularly NNPs continues to mature, a hierarchy has been defined to classify the NNPs into "generations" based on the physical content of the models. In this regard, the development of high-dimensional NNPs that take into account the intricate interplay between short-range and long-range many-body interactions is still in its early stages.[138,145]

PIPs provide a different type of ML representation that satisfies permutational, rotational, and translational invariance,[177] and has become popular in the development of MLPs for various molecular systems.[131] For example, PIP-based MLPs for water include the early HBB0,[147] HBB1,[148] HBB2,[148], and WHBB[149,150] models, which fit 2-body and 3-body energies to high-level *ab initio* data, where the the 2-body PIP is "range-separated" as it smoothly transitions into a long-range potential described by classical electrostatics. Going beyond the 3-body term, the q-AQUA model was recently introduced, which includes PIP representations for 2-body, 3-body, and 4-body energies, but neglects all n-body contributions with $n > 4$.[160]

Besides the applications mentioned above, PIPs have also been used in a variety of general approaches such as PIP-NN models, which use permutationally invariant monomials to guarantee proper symmetry relations in the NN representation of the target potential energy surface,[193–196] and Δ-ML approaches, which train an ML model on top of a lower-level (e.g., DFT) core potential to elevate the overall accuracy of the model.[197–199] PIPs and PIP-based ML approaches have also been used to develop high-dimensional potential energy surfaces of polyatomic molecules (with up to 15 atoms) in the gas phase, a step toward the "first principles" modeling of the physical properties of complex organic and biological molecules.[198,200–202]

B. Physics-based data-driven potentials

Physics-based DDPs are analytical expressions representing the multidimensional potential energy surface of an N-body system, where only a few contributions to the interaction energy are predicted from ML. In essence, physics-based DDPs capture quantum-mechanical short-range interactions using an ML framework which is integrated with a physics-based representation of classical many-body interactions.[166,203] In this context, CC-pol is a relatively simpler water model developed from high-level *ab initio* data integrated with a classical representation of electrostatic interactions.[162,204–211] CC-pol provided some of the first accurate predictions of water properties in both gas and condensed phases.

In recent years, a class of explicit many-body DDPs, namely the MB-pol,[163–165] MB-nrg,[167,168,171,172] and generalized MB-QM[140,212] potential energy functions (PEFs), have been shown to quantitatively reproduce the properties of water,[213] including gas-phase clusters,[214–216] liquid water,[217,218] the vapor/liquid interface,[219–221] and ice,[222–225] as well as various aqueous molecular[226,227] and ionic systems.[228–234] These many-body PEFs use PIPs[177] integrated with a classical many-body polarizable model[235] to represent short-range quantum-mechanical effects that arise from density overlap between molecules (i.e., exchange-repulsion, charge transfer, and charge penetration), which cannot be represented by classical expressions adopted by conventional force fields. Recently, this class of data-driven many-body potentials has been extended to covalently bonded molecules.[236] This class of DDPs, particularly the MB-DFT potentials, will be described in detail in the remainder of this work.

III. MANY-BODY INTERACTIONS IN MOLECULAR SYSTEMS

A. Theory

Consider a system of N interacting atoms or molecules. At rest, the energy of this N-body system is simply the potential energy, $E_N(\mathbf{r}_1 \ldots \mathbf{r}_N)$. By definition, $E_N(\mathbf{r}_1 \ldots \mathbf{r}_N)$ is a many-body function that contains all information regarding n-body energies, with $1 \leq n \leq N$. Therefore, the energy of the system is rigorously defined by the many-body expansion (MBE) given by:

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

where $\{\mathbf{r}_i\}$ collectively defines the coordinates of the ith monomer and $\varepsilon_{1\mathrm{B}}(\mathbf{r}_i)$ is the 1-body energy of the ith monomer,

$$\varepsilon_{1\mathrm{B}}(\mathbf{r}_i) = \begin{cases} 0 & \text{atomic monomer} \\ E(\mathbf{r}_i) - E(\mathbf{r}_{\mathrm{eq},i}) & \text{molecular monomer} \end{cases} \tag{2}$$

$\varepsilon_{1\mathrm{B}}(\mathbf{r}_i)$ represents the distortion energy of the ith isolated monomer, with $E(\mathbf{r}_{\mathrm{eq},i})$ being the energy of the ith monomer in its equilibrium configuration ($\mathbf{r}_{\mathrm{eq},i}$). In an N-molecule system, $\varepsilon_{1\mathrm{B}}(\mathbf{r}_i)$ is related to the geometrical frustration[237] of the ith monomer due to competing many-body effects that determine the minima in the underlying multidimensional energy landscape.

From Eq. 1, the 2-body and 3-body energies are obtained recursively as follows:

$$\begin{aligned} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) &= E_2(\mathbf{r}_i,\mathbf{r}_j) - \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) - \varepsilon_{1\mathrm{B}}(\mathbf{r}_j), \\ \varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) &= E_3(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) - \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) - \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_k) - \varepsilon_{2\mathrm{B}}(\mathbf{r}_j,\mathbf{r}_k) \\ &\quad - \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) - \varepsilon_{1\mathrm{B}}(\mathbf{r}_j) - \varepsilon_{1\mathrm{B}}(\mathbf{r}_k) \end{aligned} \tag{3}$$

where $E_\mathrm{n}(\mathbf{r}_1,\ldots,\mathbf{r}_n)$ is the energy of a subsystem containing n molecules. It follows that any n-body contribution $\varepsilon_{\mathrm{nB}}$ can be written as:

$$\begin{aligned} \varepsilon_{\mathrm{nB}} = E_\mathrm{n}(1,...,n) &- \sum_{i=1}^{N}\varepsilon_{1\mathrm{B}}(\mathbf{r}_i) - \sum_{i=1}^{N}\sum_{i<j}^{N}\varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) - \sum_{i=1}^{N}\sum_{i<j}^{N}\sum_{i<j<k}^{N}\varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) - \\ \ldots &- \sum_{i<j<k<\ldots}^{N}\varepsilon_{(\mathrm{n}-1)\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k,\ldots,\mathbf{r}_{n-1}). \end{aligned} \tag{4}$$

The convergence of the MBE is dependent on the intrinsic electronic structure of the system in question. Herein, we focus on aqueous systems that have localized electron densities, which implies the presence of large electronic band gaps, a feature common to all insulating systems.[238–240] For these systems, the MBE converges rapidly. Water, for example, has a band gap of ~ 9 eV, and the sum of 2-body and 3-body energies correspond to $\sim 90-95\%$ of the interaction energy.[241–250]

From a modeling standpoint, the fast convergence of the MBE makes developing data-driven PEFs based on quantum mechanics for molecular systems a feasible task, enabling the rigorous characterization of molecular systems while achieving numerical efficiency. In recent years, the synergy between high-performance computing, machine learning, and advances in theoretical chemistry has led to the development of several data-driven many-body PEFs for water[149,150,162–165,251] and other molecular systems.[167,168,171,172,232,234] We will limit this technical review to the discussion of the theoretical foundations, developments, and applications of data-driven many-body PEFs derived from arbitrary density functional approximations within DFT: MB-DFT PEFs.

B. Components and architecture of MB-DFT potential energy functions

The MB-DFT PEFs arise form the generalization of the MB-pol formalism for water[163–165] to density functionals developed across Jacob's ladder of approximations, from local functionals to semi-local, hybrid, and range-separated functionals.[140,169]

Within the MB-DFT formalism, the functional form of a data-driven many-body PEF is given by

$$E_N^{\mathrm{MB}}(\mathbf{r}_1,..,\mathbf{r}_N) = \sum_{i=1}^{N} V^{1\mathrm{B}}(\mathbf{r}_i) + \sum_{i>j}^{N} V^{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) + \sum_{i>j>k}^{N} V^{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + V_{\mathrm{pol}}(\mathbf{r}_1,..,\mathbf{r}_N) \quad (5)$$

where the $V^{1\mathrm{B}}, V^{2\mathrm{B}}, V^{3\mathrm{B}}$ are analytical representations of 1-body, 2-body, and 3-body energies fitted to the corresponding reference quantum-mechanical data. In the case of the MB-DFT PEFs, $V^{1\mathrm{B}}$ is represented by

$$V^{1\mathrm{B}}(\mathbf{r}_i) = \begin{cases} V_{\mathrm{PS}}(r_1,r_2,\theta) & \text{water} \\ V_{\mathrm{PIP}}(\{\xi_{1\mathrm{B}}\}) & \text{generic system} \end{cases} \quad (6)$$

where $V^{\mathrm{PS}}(r_1,r_2,\theta)$ is the Partridge-Schwenke PEF for the water monomer,[252] while $V_{\mathrm{PIP}}(\{\xi_{1\mathrm{B}}\})$ is a PIP representing the 1-body energy of a generic molecule, with $\{\xi_{1\mathrm{B}}\}$ collectively defining exponential functions of the interatomic distances. The second term on the left hand side of Eq. 5 is

$$V^{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) = V^{2\mathrm{B}}_{\mathrm{sr}}(\mathbf{r}_i,\mathbf{r}_j) + V^{2\mathrm{B}}_{\mathrm{elec}}(\mathbf{r}_i,\mathbf{r}_j) + V^{2\mathrm{B}}_{\mathrm{pol}}(\mathbf{r}_i,\mathbf{r}_j) + V^{2\mathrm{B}}_{\mathrm{disp}}(\mathbf{r}_i,\mathbf{r}_j) \quad (7)$$

where $V^{2\mathrm{B}}_{\mathrm{sr}}(\mathbf{r}_i,\mathbf{r}_j)$ describes short-range 2-body interactions and is represented by the product of a 2-body PIP with a switching function that smoothly tends to zero as the distance between a pair of monomers reaches a predefined cutoff limit. $V^{2\mathrm{B}}_{\mathrm{elec}}(\mathbf{r}_i,\mathbf{r}_j)$ and $V^{2\mathrm{B}}_{\mathrm{pol}}(\mathbf{r}_i,\mathbf{r}_j)$

represent permanent electrostatics between point charges that reproduce the *ab initio* dipole moment of an isolated monomer, and 2-body polarization, respectively. In the actual implementation of the MB-DFT PEFs, $V^{2B}_{\mathrm{pol}(\mathbf{r}_i,\mathbf{r}_j)}$ is implicitly included in the N-body polarization term, $V_{\mathrm{pol}}(\mathbf{r}_1,..,\mathbf{r}_N)$, of Eq. 5. Following Refs. 71,140, the point charges are kept fixed to the values that reproduce the dipole moment of a molecule in its equilibrium geometry ($\mathbf{r}_i = \mathbf{r}_{\mathrm{eq},i}$). The last term in Eq 7, $V^{2B}_{\mathrm{disp}}(\mathbf{r}_i,\mathbf{r}_j)$, describes the 2-body dispersion energy which is expressed as

$$V^{2B}_{\mathrm{disp}}(\mathbf{r}_i,\mathbf{r}_j) = -\sum_{i,j} f_k(r_{ij}) \frac{C_{6,ij}}{r^6_{ij}} \tag{8}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $f_k(r_{ij})$ is the Tang-Toennies damping function[253],

$$f_k(r_{ij}) = 1 - \left(\sum_{i=0}^{k} \frac{(\delta \mathbf{r}_{ij})^k}{k!} \right) \exp\left(-\delta \mathbf{r}_{ij}\right), \tag{9}$$

Here, k is the order of the function, r_{ij} is the separation between two atoms on two distinct monomers, δ is the coefficient that determines the effective length of damping obtained during the fitting procedure,[171,172] and $C_{6,ij}$ is the corresponding dispersion coefficient derived from calculations carried out within the exchange-hole dipole moment (XDM) model.[60,254,255]

The third term on the left hand side of Eq. 5, V^{3B}, is defined as

$$V^{3B}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) = V^{3B}_{\mathrm{SR}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + V^{3B}_{\mathrm{pol}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) \tag{10}$$

Analogous to $V^{2B}_{\mathrm{SR}}(\mathbf{r}_i,\mathbf{r}_j)$, $V^{3B}_{\mathrm{SR}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k)$ is represented by the product of a 3-body PIP with a switching function that smoothly tends to zero as the distance between any pair of monomers in a trimer reaches a predefined cutoff limit. Similarly, $V^{3B}_{\mathrm{pol}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k)$ represents 3-body polarization that is implicitly included in the N-body polarization term, $V_{\mathrm{pol}}(\mathbf{r}_1,..,\mathbf{r}_N)$, of Eq. 5. The 2-body and 3-body PIPs effectively recover quantum-mechanical interactions arising from the overlap of monomer's electron densities (i.e., exchange-repulsion, charge transfer, and charge penetration) which cannot be represented by classical expressions.[256] For further details regarding the explicit form of $V_{2B,\mathrm{SR}}(\mathbf{r}_i,\mathbf{r}_j)$, $V_{3B,\mathrm{SR}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k)$, including the definition of the switching functions, the reader is referred to Ref. 144.

The final term of Eq. 5, $V_{\mathrm{pol}}(\mathbf{r}_1,..,\mathbf{r}_N)$, implicitly represents N-body classical polarization. It is important to emphasize that the inclusion of $V_{\mathrm{pol}}(\mathbf{r}_1,..,\mathbf{r}_N)$ in addition to the explicit 1-body, 2-body and 3-body terms guarantees that the MB-DFT PEFs are not truncated, enabling them to rigorously account for both short-range and long-range many-body interactions.

C. Implementation

Data-driven many-body PEFs can be readily developed for generic molecules in an automated fashion using the MB-Fit software infrastructure introduced in Ref. 144. Briefly, MB-Fit is an open-source package that is designed to streamline the development of data-driven PEFs, such as the MB-nrg family of PEFs for aqueous ionic systems and molecular fluids derived from CCSD(T)) data, and the focus of this technical review, MB-DFT PEFs.

MB-Fit readily enables the development of training and test sets, the generation of the PIPs, the calculations of the reference *ab initio* n-body energies (currently supporting Q-Chem[257] and PSI4[258]), the evaluation of the fits, as well as the generation of the codes necessary to perform molecular dynamics simulations using the developed many-body PEF in LAMMPS[259] or i-PI[260] through the MBX[261] interface. LAMMPS, i-PI, and MBX are freely available to the public. A schematic overview of the MB-DFT framework is shown in Figure 1. Figure 1 shows the two main pillars of MB-DFT: theory and implementation (a-b) and applications (b-c). Panel (a) shows the Jacob's ladder of DFAs, suggesting the robustness of the MBE with respect to the different levels of DFAs (b). Panel (b) specifically summarizes the MB-DFT development workflow according to the general MB-nrg framework introduced in Ref. 144. Briefly, the MB-DFT V^{nB} terms, with n = 1-3, contain explicit PIP terms given by

$$P(\xi_1, \xi_2, \ldots, \xi_N) = \sum_{l=0}^{L} A_l S[\xi_1^{a_{l1}}, \xi^{a_{l2}}, \ldots, \xi^{a_{lN}}]. \tag{11}$$

Here, L is the total number of monomials in the PIP, A_l is a linear fitting coefficient for monomial l, and $S[\{\xi\}]$ is an operator that symmetrizes each monomial l to guarantee invariance to permutations. In addition, ξ_i is a variable defined as a function of distance between sites, which include both physical atoms and fictitious sites of the monomers contributing to the 1-body, 2-body, or 3-body PIPs. Four different functional forms are available for ξ_i in MB-Fit, including Coulomb and Morse variables.[144]

The fitting procedure, using Tikhonov regularization (also known as ridge regression),[262] is discussed in detail in Ref. 144, and further details regarding the PIPs can be found in the literature.[177]

As shown in Figure 1, MB-DFT enables (b) non-reactive molecular dynamics simulations as well as (d) fully polarizable quantum mechanics/many-body molecular mechanics (QM/MB-MM) simulations with arbitrary levels of theory, enabling a seamless multiscale

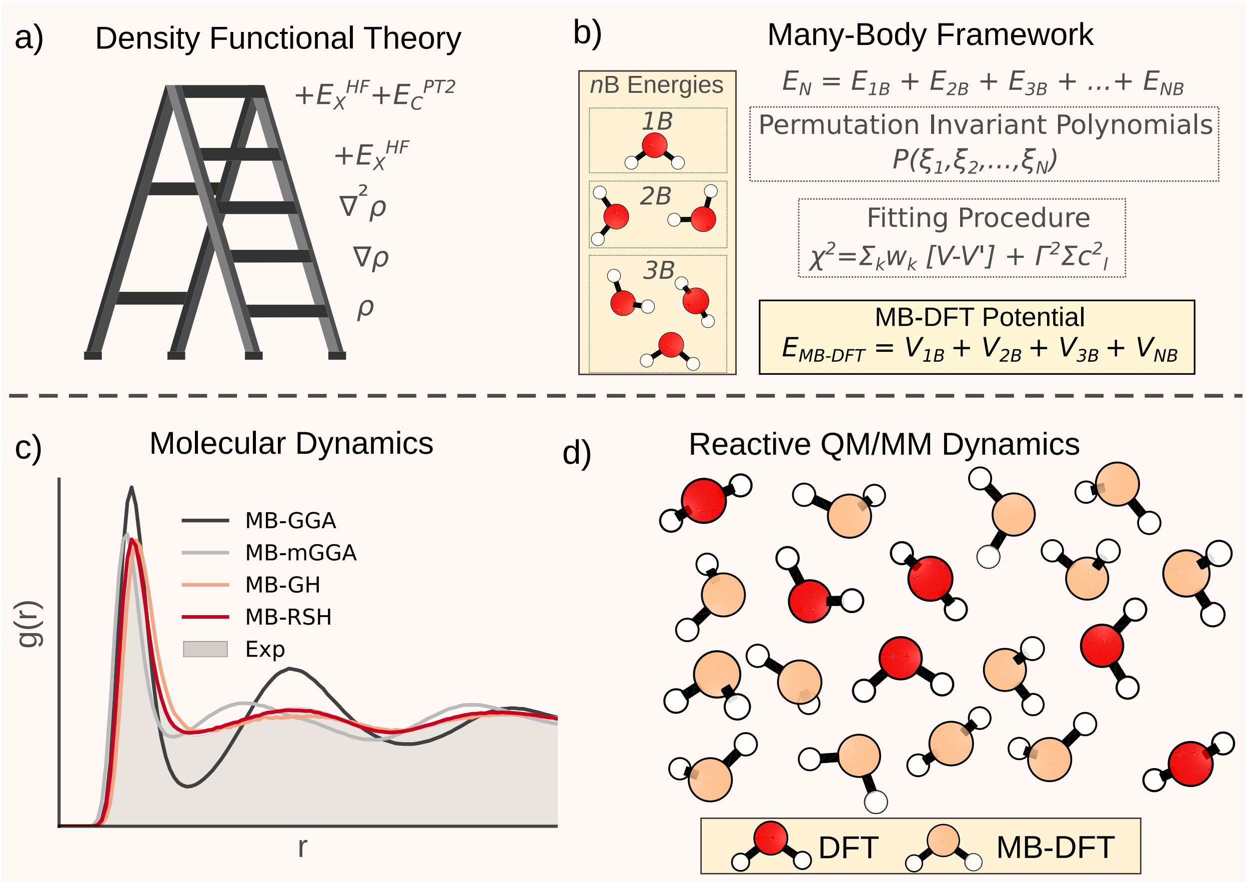

FIG. 1. This schematically depicts the main features of MB-DFT, integrating (a) electronic structure theory (QM) with (b) the many-body framework[144], enabling (c) non-reactive MB molecular dynamics (MB-MD) and (d) reactive hybrid quantum mechanics/many-body molecular mechanics (QM/MB-MM) simulations.

representation of the underlying multidimensional potential energy surface.[263,264] These developments along with the corresponding applications are discussed in the following sections. The latest versions of MB-Fit and MBX can be downloaded from the GitHub repositories in Refs. 265 and 266, respectively.

IV. DENSITY FUNCTIONAL THEORY

A. Kohn-Sham DFT

The form of the exact density functional $E[\rho(\mathbf{r})]$ that determines the ground-state electronic energy of an arbitrary N-electron system is unknown. To this end, the energy expression for a given DFA can be written as

$$\tilde{E}[\rho(\mathbf{r})] = \tilde{F}[\rho(\mathbf{r})] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} \tag{12}$$

where $\tilde{F}[\rho]$ is the approximate one-electron density functional, which is independent of the external potential acting on the N-body system, $v_{\text{ext}}(\mathbf{r})$. The approximate functional $\tilde{F}[\rho]$ is defined as

$$\tilde{F}[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + \tilde{V}_{\text{XC}}[\rho(\mathbf{r})] \tag{13}$$

where $T_s[\rho(\mathbf{r})]$ is the non-interacting Kohn-Sham kinetic energy, and $J[\rho(\mathbf{r})]$ defines the Coulomb interaction between the electrons. The last term of Eq. 13, $\tilde{V}_{\text{XC}}[\rho(\mathbf{r})]$, is known as the exchange-correlation (XC) potential and it is the only term in Eq. 13 that is be approximated to describe the non-classical electron-electron interactions, as the development of the exact functional for molecular systems currently is an unattainable objective. As a consequence, the accuracy of any physical properties derived from $\tilde{F}[\rho]$ is determined by the accuracy of $\tilde{V}_{\text{XC}}[\rho(\mathbf{r})]$. The complexity of $\tilde{V}_{\text{XC}}[\rho(\mathbf{r})]$ can be understood in terms of a sum of semi-local (sl) and nonlocal (nl) XC terms,

$$\tilde{V}_{XC}[\rho(\mathbf{r})] = V_{xc}^{sl}[\rho(\mathbf{r})] + V_{xc}^{nl}[\rho(\mathbf{r})] \tag{14}$$

For some functionals, a fraction of Hartree-Fock exchange as well as an amount of correlation energy from post-Hartree-Fock wavefunction theories are added to improve the accuracy.[87] Perhaps, the most intuitive way to understand the complexity of DFAs is through "Jacob's ladder"[53] of DFAs which locates DFAs in rungs based on their complexity and accuracy. Pure DFT is defined by all XC potentials strictly of the form given by Eq. 14 (rungs 1-3), while the introduction of Hartree-Fock exchange defines hybrid DFT (rung 4), and the incorporation of post Hartree-Fock correlation defines double-hybrid DFT (rung 5).

B. Density-corrected DFT

In the Kohn-Sham framework, the energy error of any DFT calculation can be decomposed as

$$\Delta E = \underbrace{\tilde{F}[\rho] - F[\rho]}_{\Delta E_{FD}} + \underbrace{\tilde{E}[\tilde{\rho}] - \tilde{E}[\rho]}_{\Delta E_{DD}} \tag{15}$$

where ΔE_{FD} is the error due to the density functional approximation (FD error), and ΔE_{DD} is the error due to the approximated density (DD error). The significance of these errors are related to how well a DFA describes properties such as electron density and static correlation energy, which hinders most DFAs from accurately describing systems with fractional charge (FC)[267] as well as systems with fractional spin (FS),[268] giving rise to the delocalization error (DE) and static correlation errors (SCE), respectively. The SCE originates entirely in the XC potential and is known to be significant in local and semi-local DFAs that, while describing dynamic correlation fairly well, are not able to correctly describe static correlation due to its inherent mid- to long-range nature.[268] Therefore, understanding the SCE, a component of ΔE_{FD}, is crucial for the study of strongly-correlated systems as well as for an accurate description of molecular dissociation.[268,269]

Although ΔE_{FD} is the principal contributor to the total error in most cases,[270–273] all XC functionals deviate from the piecewise-linearity[274,275] of the exact functional for fractional charges, causing excess charge delocalization and resulting in incorrect densities. In these cases where the DE is significant, ΔE_{DD} becomes the dominant contributor to the total error.[273,276] The cause of ΔE_{DD} is the manifestation of the one-electron and many-electron SIE,[57] which is thought to be the biggest contributor to ΔE_{DD} and thus responsible for the observed over-delocalization of the electron density.[140,277,278]

Consequently, most DFAs are unable to satisfy the following three conditions for the one-electron system,[279]

$$T_{\mathrm{s}}[\rho(\mathbf{r})] = \int \mathrm{d}\mathbf{r} \frac{|\nabla \rho(\mathbf{r})|^2}{8\rho(\mathbf{r})}, \quad J + E_{\mathrm{X}} = 0, \quad E_{\mathrm{C}} = 0 \tag{16}$$

where E_{X} and E_{C} are the exchange and correlation energies, respectively. The second condition of Eq. 16 is the vanishing of self-interaction. The inability to satisfy this requirement gives rise to the SIE, since $J \neq 0$ is clearly unphysical. Correcting for SIE (to whatever extent) reduces the ΔE_{DD} and thus raises the accuracy of DFAs in both predicting energies as well as densities.[280–282]

Density-corrected DFT (DC-DFT) provides a practical manner of improving the accuracy of DFT *via* the non-self-consistent evaluation of a DFA on an accurate electron density.[283–286] Eq. 15 is equivalent to the energy difference between a DFA evaluated on its predicted density and the exact functional evaluated on the exact density, $\tilde{E}[\tilde{\rho}] - E[\rho]$. For systems for which DFT notoriously overdelocalizes the density,[278] the Hartree-Fock density is known to be a reasonable choice for approximating the density-corrected energy, for which $\Delta E \simeq \Delta E_{FD}$.[283–286] In this scheme of DC-DFT, which is also referred to as HF-DFT, the density-corrected DFT energy $E_{\mathrm{DC-DFT}}$ is approximated as[279]

$$E_{\mathrm{DC-DFT}} \simeq E^{\mathrm{HF}} + \left\{ \tilde{E}_{\mathrm{XC}}[\rho^{\mathrm{HF}}] - E_X^{\mathrm{HF}} \right\}. \tag{17}$$

This approximation has been shown to be accurate to the second order in density difference, in line with the variational principle.[279] For further details regarding the DC-DFT framework, we refer the reader to a recent review.[287]

V. MANY-BODY PEFS FOR WATER FROM DFT

A. First-generation MB-DFT

The MB-DFT framework aims to exploit the strengths of DFT in treating quantum-mechanical interactions, while simultaneously enabling efficient simulations of condensed-phase systems. The first generation of MB-DFT PEFs for water were derived analogously to the MB-pol PEF, which predicts the properties of gas-phase clusters, bulk water, vapor/liquid interface, and ice within sub-chemical accuracy.[213] The first-generation MB-DFT PEFs represent the 1-body energy through the Partridge-Schwenke PEF,[252] while the second and third terms of Eq. 5 were fitted to 2-body and 3-body energies calculated at a given DFT level. A note on notation: an arbitrary MB-DFT potential labeled as (2B+3B)-DFA treats the first and fourth terms of Eq. 5 as in MB-pol, and the 2-body and 3-body terms according to the specified DFA. Analogously, a MB-DFT potential labeled as 2B-DFA represents all terms of the PEF as in MB-pol *except* the 2-body term, and in a similar fashion for 3B-DFA, *only* the 3-body term is fitted to reproduce the underlying functional. By construction, the MB-DFT PEFs thus enable a systematic assessment of the interplay between 2-body and 3-body interactions in determining the structural properties of water.

A schematic representation of the many-body decomposition (MBD) of the interaction energy of a water hexamer cluster is shown in Figure 2(a), suggesting that the 2B+3B energies make up $\sim$ 95% of the interaction energy. The eight low-energy isomers of the water hexamer are shown Figure 2(b). Figure 2(c) shows the correlations between 2-body and 3-body energies calculated with four DFAs belonging to rung 2 (revPBE-D3),[288] rung 3 (B97M-rV),[52] and rung 4 (revPBE0-D3 and ωB97M-V)[289] across Jacob's ladder of DFAs, and the corresponding CCSD(T) reference energies calculated in the complete basis set limit (CBS).[169] Specifically, revPBE-D3 is a GGA dispersion-corrected functional, B97M-rV

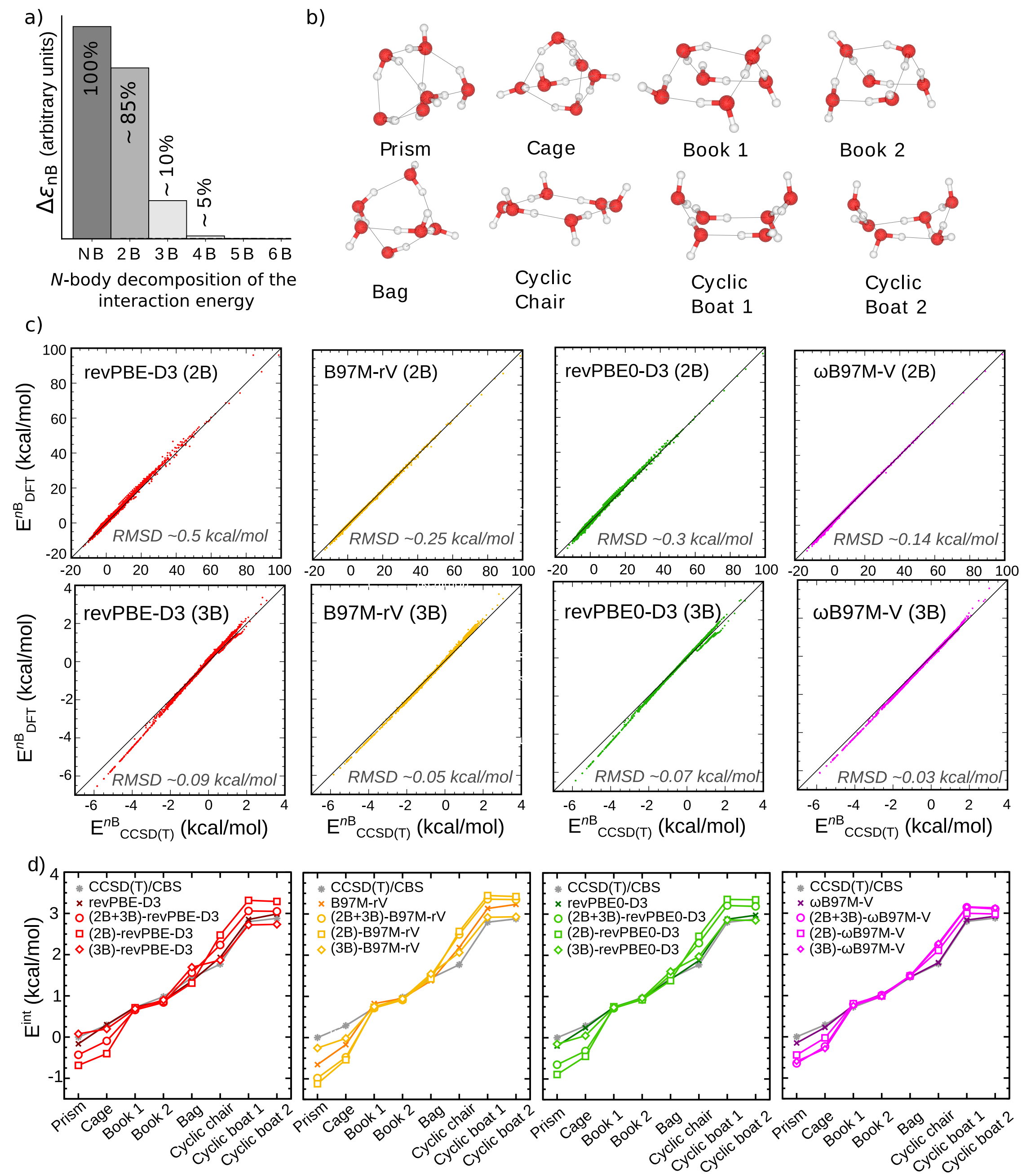

FIG. 2. Importance of 2B and 3B energies in water. (a) schematic of many-body decomposition (MBD) of the interaction energy of the $(H_2O)_6$ prism isomer into individual nB contributions. (b) the eight low-energy isomers of $(H_2O)_6$. (c) shows the correlation plots for 2B (top) and 3B (bottom) PEFs evaluated against CCSD(T). Correlation plots are shown for revPBE-D3 (red), B97M-rV (gold), MB-revPBE0-D3 (green), ωB97M-V (purple). Panel (d) shows the difference between MB-DFT models in predicting the interaction energy relative to CCSD(T). Figure adapted with permission from Chem. Sci. **10**, 8211–8218 (2019). Copyright 2019 Royal Society of Chemistry.

is a meta-GGA functional with non-local correlation, rev-PBE0-D3 is a dispersion-corrected hybrid functional, and ωB97M-V is a range-separated, meta-GGA, hybrid functional.

A common, but not systematic, "rule of thumb" in DFT is that, as one climbs *up* Jacob's ladder, the accuracy increases. This is generally true, as seen in the 2-body and 3-body correlation plots showing significant improvement as one goes from revPBE-D3 towards ωB97M-V where a lower RMSD relative to CCSD(T) is displayed. For instance, in the 2B PES, revPBE-D3 has an RMSD of -0.5 kcal/mol (rung 2), B97M-rV -0.25 kcal/mol (rung 3), and finally 0.14 kcal/mol for ωB97M-V. These findings are in line with AIMD simulations reported in Refs. 50,52. In general, the four DFAs analyzed in Figure 2(c) provide better agreement with the CCSD(T) reference values for 2-body than 3-body energies. As suggested by Figure 2(a), this directly affects how the corresponding four MB-DFT PEFs predict the properties of larger water systems, from gas-phase clusters to liquid water.

The interaction energies (E_{int}) of the first eight low-energy isomers of the water hexamer calculated with the revPBE-D3, B97M-rV, revPBE0-D3, and ωB97M-V functionals and the corresponding MB-DFT PEFs are shown in Figure 2(d) along with the CCSD(T)/CBS reference values. In this analysis, E_{int} is defined as the difference between the total energy of the $(H_2O)_n$ cluster ($E_{n-\text{mer}}$) and the sum of the energies of the individual n water molecules in the same distorted geometries as in the cluster (E^{H_2O}),

$$E_{\text{int}} = E_{n\text{-mer}} - \sum_{i=1}^{n} E_i^{H_2O} \tag{18}$$

To emphasize the importance of a rigorous treatment of both 2-body and 3-body energies, E_{int} is computed using: (i) DFAs, (ii) full (2B+3B)-DFA PEFs, (iii) 2B-DFA PEFs, and (iv) 3B-DFA PEFs. Figure 2(d) shows that ωB97M-V predicts E_{int} in fair agreement with the CCSD(T)/CBS reference values. The dominance of the 2-body energies over the 3-body energies is evidenced by the fact that the 2B-DFA PEFs display the highest errors in the case of all DFAs, while the 3B-DFA PEFs, which use the same 2-body term as MB-pol, display the smallest errors. Morever, the three different MB-DFT PEFs derived from each of the four DFAs predict E_{int} within ± 1 kcal/mol of the bare DFA, attesting to the accuracy of the MB-DFT PEFs in faithfully reproducing the corresponding DFA n-body energies.

The delicate interplay among many-body effects in gas-phase clusters transfers over to the condensed phase, directly affecting the structural properties of liquid water. For all MB-DFT PEFs, path-integral molecular dynamics (PIMD) simulations,[290,291] which explic-

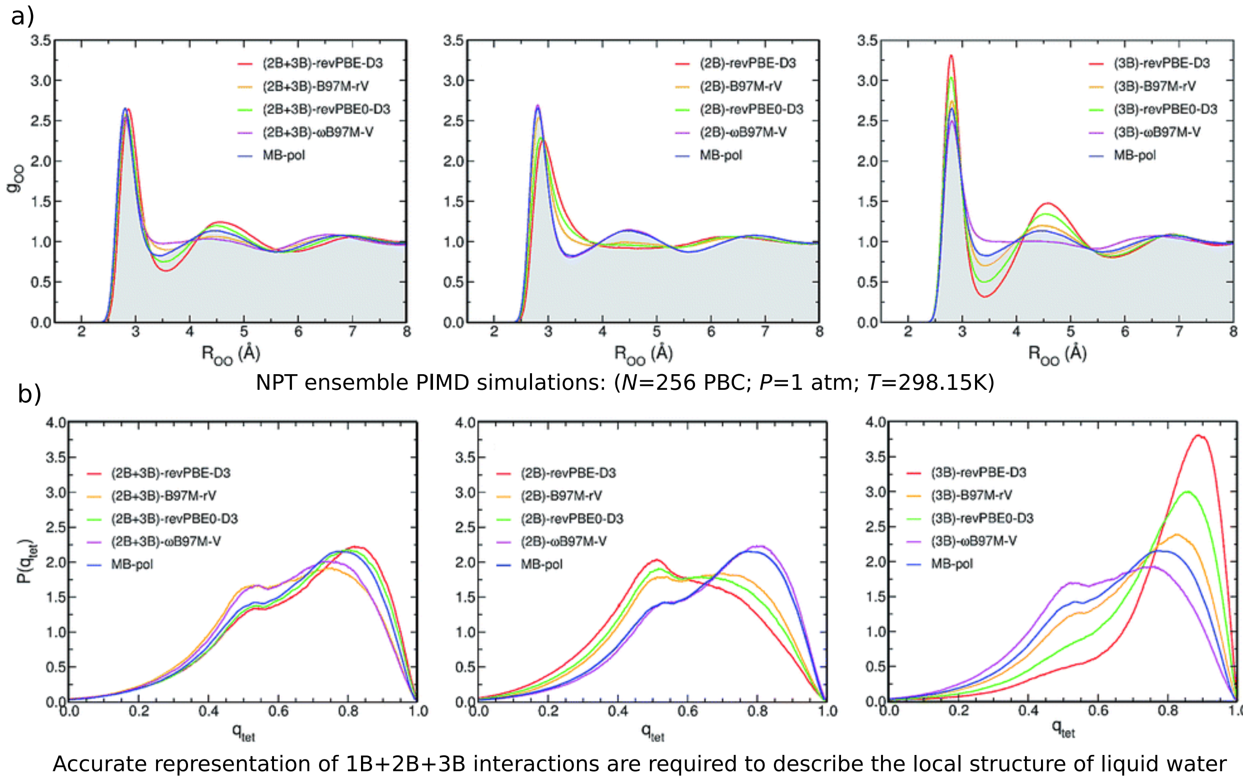

FIG. 3. Structural properties of liquid water as predicted by first-generation MB-DFT models. From left to right, (2B+3B)-, 2B- and 3B- MB-DFT potentials are shown. Panel (a) shows the comparison between oxygen–oxygen radial distribution functions (RDFs), $g_{\mathrm{OO}}(r)$, of liquid water at ambient conditions derived from X-ray diffraction measurements (gray area) and calculated from path-integral molecular dynamics (PIMD) simulations in the isothermal-isobaric (NPT) ensemble at ambient conditions (T=298.15 K and P=1 atm). Panel (b) shows the corresponding normalized probability distribution of the tetrahedral order parameter $P(q_{\mathrm{tet}})$ under the same thermodynamic conditions. Here, MB-revPBE-D3 (red), B97M-rV (yellow), revPBE0-D3 (green), and ωB97M-V (magenta) are shown with MB-pol (blue) for comparison. Figure adapted with permission from Chem. Sci. **10**, 8211–8218 (2019). Copyright 2019 Royal Society of Chemistry.

itly accounts for nuclear quantum effects, were carried out in the canonical (NVT) ensemble at 298.15 K and experimental density as well as in the isothermal-isobaric (NPT) ensemble at 298.15 K and 1 atm.[169] The structural properties of liquid water predicted by the (2B+3B)-DFA, (2B)-DFA, (3B)-DFA, and MB-pol PEFs are shown in Figure 3. Specifically, Figure 3(a) shows the oxygen-oxygen RDFs calculated from NPT PIMD simulations carried out with the four (2B+3B)-DFA PEFs described above as well as MB-pol that is used as a reference. For those DFAs for which AIMD simulations were reported in the literature, the corresponding (2B+3B)-DFA PEFs closely reproduces the AIMD results.[169] This attests to the ability of the (2B+3B)-DFA PEFs to predict the properties of water with full *ab initio* accuracy. This preservation of accuracy holds for all functionals with the exception of revPBE-D3. The nature of this discrepancy will be ascertained in Section VII.

Unlike simulations performed in the NVT ensemble, the volume and, thus, the density are allowed to fluctuate in the NPT simulations, resulting in a local structure that is more sensitive to the "realism" and predictive capabilities of a given PEF. To this end, Ref.[169] demonstrated that all four MB-DFT PEFs described above provide RDFs in fair agreement to experiment when calculated from NVT simulations. However, when the simulations were carried out in the NPT ensemble, the same four MB-DFT PEFs predict RDFs that deviate significantly from the experimental results which correlates with the ability of a given MB-DFT to reproduce low-order n-body energies. As it may be expected from the MBD analysis, the GGA revPBE-D3 and hybrid revPBE0-D3 functionals predict an ice-like local hydrogen-bonding environment, which manifests in *over*-structured oxygen-oxygen RDFs calculated with the corresponding MB-revPBE-D3 and MB-revPBE0-D3 PEFs. This is in contrast to the performance of the MB-B97M-rV and MB-ωB97M-V PEFs that predict a slightly *under*-structured liquid, which can been traced back to the more repulsive nature of the non-local rVV10 correlation functional.

Dissecting the full (2B+3B)-DFA PEFs into their underlying (2B)-DFA and (3B)-DFA components highlights the cooperative nature of many-body interactions in determining the local structure of water. This is manifested in the differences in the oxygen-oxygen RDFs, which translate into differences in the local structure of liquid water as described by the distribution of the tetrahedral order parameter, $P(q_{\text{tet}})$, shown in Figure 3(b) for the different MB-DFT PEFs. The tetrahedral order parameter, q_{tet}, is given by[292]

$$q_{\text{tet}} = 1 - \frac{3}{8} \cdot \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\psi_{jk}) + \frac{1}{3} \right)^2 \tag{19}$$

where ψ_{jk} is the angle between the oxygen of the central water molecule and the oxygen atoms of the two neighboring water molecules. The tetrahedral order parameter has values of $0 \leq q_{\text{tet}} \leq 1$, for which the limiting cases are the ideal gas ($q_{\text{tet}} = 0$) and perfect tetrahedral coordination ($q_{\text{tet}} = 1$).

These analyses demonstrate that the ability of a given DFA to accurately describe individual n-body energies effectively determines its ability to correctly predict the structure of liquid water. In this context, comparing the (2B+3B)-DFA PEFs with the (2B)-DFA and (3B)-DFA PEFs demonstrates that most of the DFAs rely on error compensation in the representation of 2-body and 3-body energies. To summarize, the original MB-DFT framework provides a platform for (i) the construction of many-body PEFs rigorously derived from a given DFA, and (ii) the systematic assessment of the reliability of a given DFA in representing water from the gas to the condensed phase.

B. Generalized MB-DFT framework

The generalized MB-DFT framework goes beyond the first-generation MB-DFT PEFs described above by determining all parameters used in the classical components of the PEF (i.e., atomic charges and polarizabilities entering V_{elec} and V_{pol}, respectively, and dispersion coefficients entering V_{disp}) from *ab initio* calculations carried out using the same DFA. In the generalized MB-DFT framework, the 1-body, 2-body, and 3-body terms of Eq. 5 contain short-range PIPs that represent the corresponding 1-body, 2-body, and 3-body energies calculated with a given DFA. Hence, by construction, the generalized MB-DFT framework expands the applicability of MB-DFT PEFs to generic molecular systems beyond water. Importantly, building upon the predictive power of prior families of data-driven many-body PEFs,[163–165,167–169,171,172] the generalized MB-DFT framework[140] was developed for application in fully polarizable QM/MB-MM simulations of chemical transformations in solution which correctly account for many-body interactions in both QM and MB-MM regions. The necessity of a QM/MB-MM scheme arises from the fact that a rigorous description of the interactions between the reactive species and the solvent molecules is warranted as many reaction mechanisms and associated rates are solvent-supported.[293]

Since the generalized MB-DFT PEFs are physics-based DDPs, they lend themselves well to QM/MM simulations where pure ML models come short as they cannot provide a physics-based description of many-body electrostatics and, consequently, the coupling between the MM region and the QM density.[294] In this section, we discuss the properties and limitations of the generalized MB-DFT potentials for water, while the application of MB-DFT in QM/MB-MM is discussed in Section VIII.

For the MB-DFT PEFs, the dipole polarizabilities of the free atoms are computed at

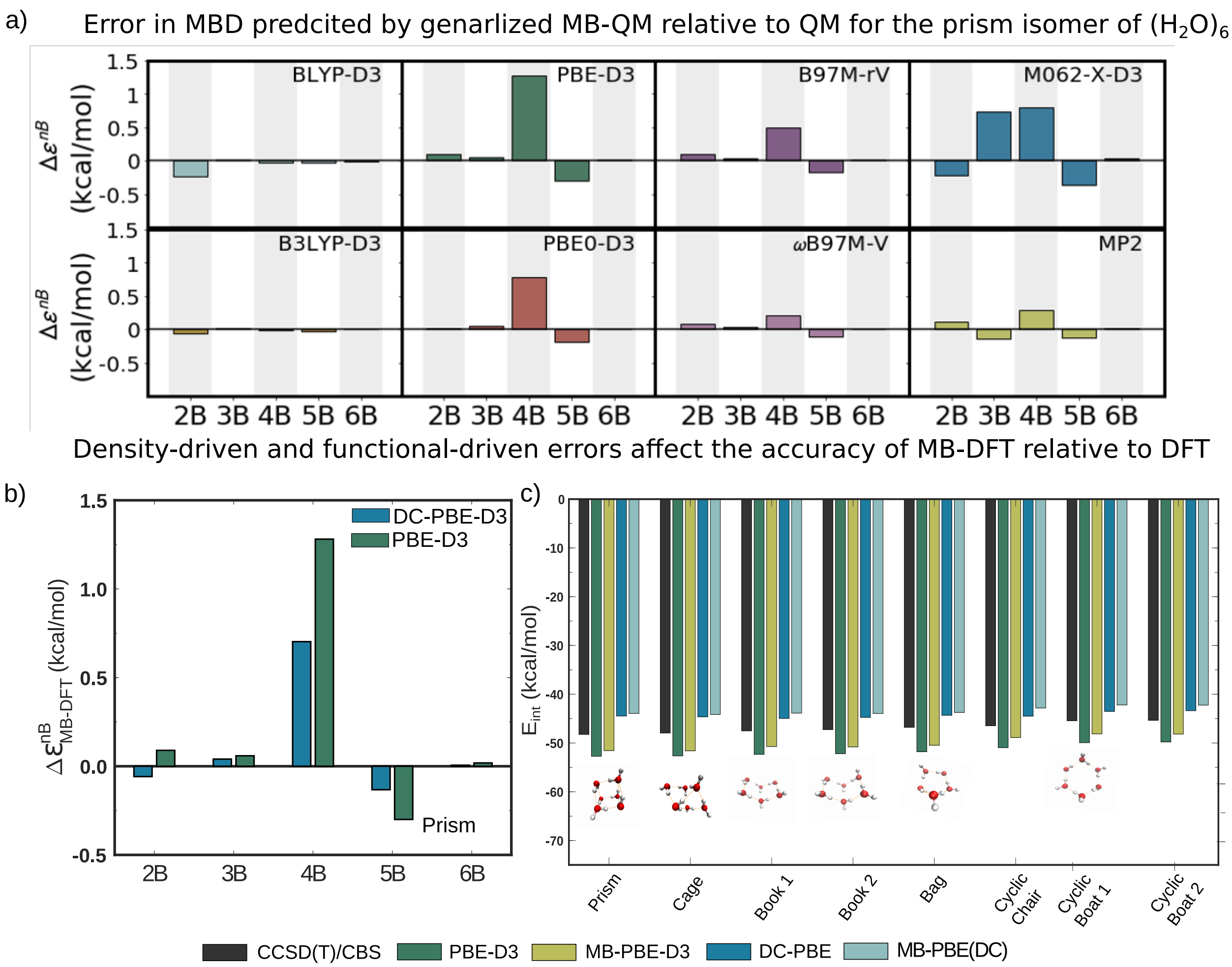

FIG. 4. Panel (a) shows the errors (in kcal/mol) associated with the MBD of the interaction energy of the prism isomer of the water hexamer, calculated with MB-QM models relative to their *ab-initio* reference values, for density functionals belonging to rungs 2,3, and 4 of Jacob's ladder, in addition to MP2. Panel (b) shows compares the nB-decomposition of PBE-D3 with its density-corrected functional, DC-PBE-D3. Panel (c) displays the interaction energies for the eight water isomers of the hexamer computed with PBE-D3, MB-PBE-D3, DC-PBE MB-PBE(DC), relative to CCSD(T). Figure adapted with permission from J. Chem. Theory Comput. **17**, 5635–5650 (2021). Copyright 2021 American Chemical Society.

the desired DFA level using the XDM model,[60,254,255] and the atomic charges are computed using the charge-model-5 (CM5) scheme.[295] The first concern is how well the MB-DFT PEF derived from a given DFA reproduces the corresponding 1-body, 2-body, and 3-body energies. Generally speaking, the MB-DFT framework is robust enough so that a MB-

DFT PEF accurately reproduces the corresponding 1-body, 2-body, and 3-body energies calculated with the corresponding DFA, with RMSDs of ∼0.08, ∼0.12, and ∼0.03 kcal/mol, respectively.[140]. This high correlation between the MB-DFT and DFA 2-body and 3-body energies supports that the notion that the associated 2-body and 3-body PIPs effectively account for genuinely short-range quantum-mechanical 2-body and 3-body contributions to the interaction energy, such as exchange-repulsion, charge penetration, and charge transfer, which, by definition, cannot be represented by classical expressions adopted by conventional force fields.[140] This is an important point, as it directly relates to the ability of a MB-DFT PEF to correctly represent individual n-body energies in both gas-phase and condensed-phase systems.

In Figure 4(a), the MBD of the interaction energy of the prism isomer of the water hexamer is shown for several MB-DFT PEFs with respect to their parent DFA, as well as for MB-MP2 relative to MP2. The general trend is that the MBDs calculated with the different MB-DFT PEFs are in qualitative agreement with the corresponding DFA reference values. This analysis shows that the 4-body energy is the principal contributor of the total error in the MBD calculated with all the MB-DFT PEFs, albeit the 4-body energy only accounts for $\sim 5\%$ of E_{int}, as shown in Figure 2. For example, Figure 4(a) shows that MB-PBE-D3 exhibits an average error (over the eight isomers of the water hexamer) in the 4-body energy, ΔE^{4B}_{avg}, of 1.07 kcal/mol. The inclusion of 25% Hartree-Fock exchange does not significantly improve the 4-body energy, since $\Delta E^{4B}_{avg} = 0.72$ kcal/mol for MB-PBE0-D3. Since the MB-DFT PEFs represent all n-body interactions with $n \geq 4$ using a classical many-body polarization term, the relatively large 4-body errors associated with PBE-D3, and PBE0-D3, as well as B97M-rV and M06-2X-D3, indicate that this classical representation, which was shown to accurately represent 4-body CCSD(T) energies in MB-pol,[164] appears to be not sufficient to fully recover 4-body energies calculated with these DFAs.

Figure 4(b) shows the MBD of the prism isomer calculated with PBE-D3 and its density-corrected analog, DC-PBE-D3, relative to CCSD(T)/CBS. The errors in 4-body and 5-body energies decrease by a factor of ∼2 when going from PBE-D3 to DC-PBE-D3, which implies that PBE-D3 suffers from significant density-driven errors. Importantly, Figure 4(c) shows that, beyond reducing the errors with respect to CCSD(T)/CBS, the discrepancy between MB-PBE-D3(DC) and DC-PBE-D3 is also reduced significantly compared to that found between MB-PBE-D3 and PBE-D3.

VI. MANY-BODY POTENTIALS FROM SCAN AND RELATED FUNCTIONALS

The SCAN functional has been used in a variety of studies that examine various structural and thermodynamic properties of water across the phase diagram.[99,102,157,158,296] In the context of these studies, it is relevant to note that SCAN exhibits a notably high error in the 2-body energies, while broadly predicting very accurate higher order terms of the MBE.[98] This can be traced back to over-delocalization errors arising from the approximate description of the exchange-correlation term in the different DFAs.[98]

A. The effect of Hartree-Fock exchange: SCAN and SCANα functionals

The adiabatic connection formula enables the effective formulation of hybrid functionals that incorporate a certain fraction of Hartree-Fock exchange along with the approximated semi-local exchange.[297–300] Here, SCAN0 represents a hybrid functional with 25% of Hartree-Fock exchange, which formally cancels out the self interaction error introduced in the Coulomb term, up to the percentage of Hartree-Fock exchange.[60,278] Building from the adiabatic connection formula, the MB-SCANα PEFs were developed to systematically explore the effect of adding a fraction (α) of Hartree-Fock exchange on the structure and dynamics of liquid water, and connect the smaller energy differences between SCAN and SCAN0 in the gas phase to the more substantial structural and thermodynamic differences between these two DFAs in the liquid phase.

The MBEs calculated for the water hexamers using the SCANα DFAs display larger errors in the 2-body energies. Interestingly, the description of 2-body energies of the 3-dimensional hexamer isomers, such as the prism and cage isomers, benefits from larger amounts of Hartree-Fock exchange, whereas strictly planar isomers, such as the cyclic ring, are better represented by the pure SCAN functional.[101] It should be noted that isomers with nearly planar geometries, such as the book isomers, minimize their 2-body errors when a fraction of 15-20% Hartree-Fock exchange is added to SCAN. It was demonstrated that the variation in the 2-body energies between SCAN and SCAN0 (which is on the order 1

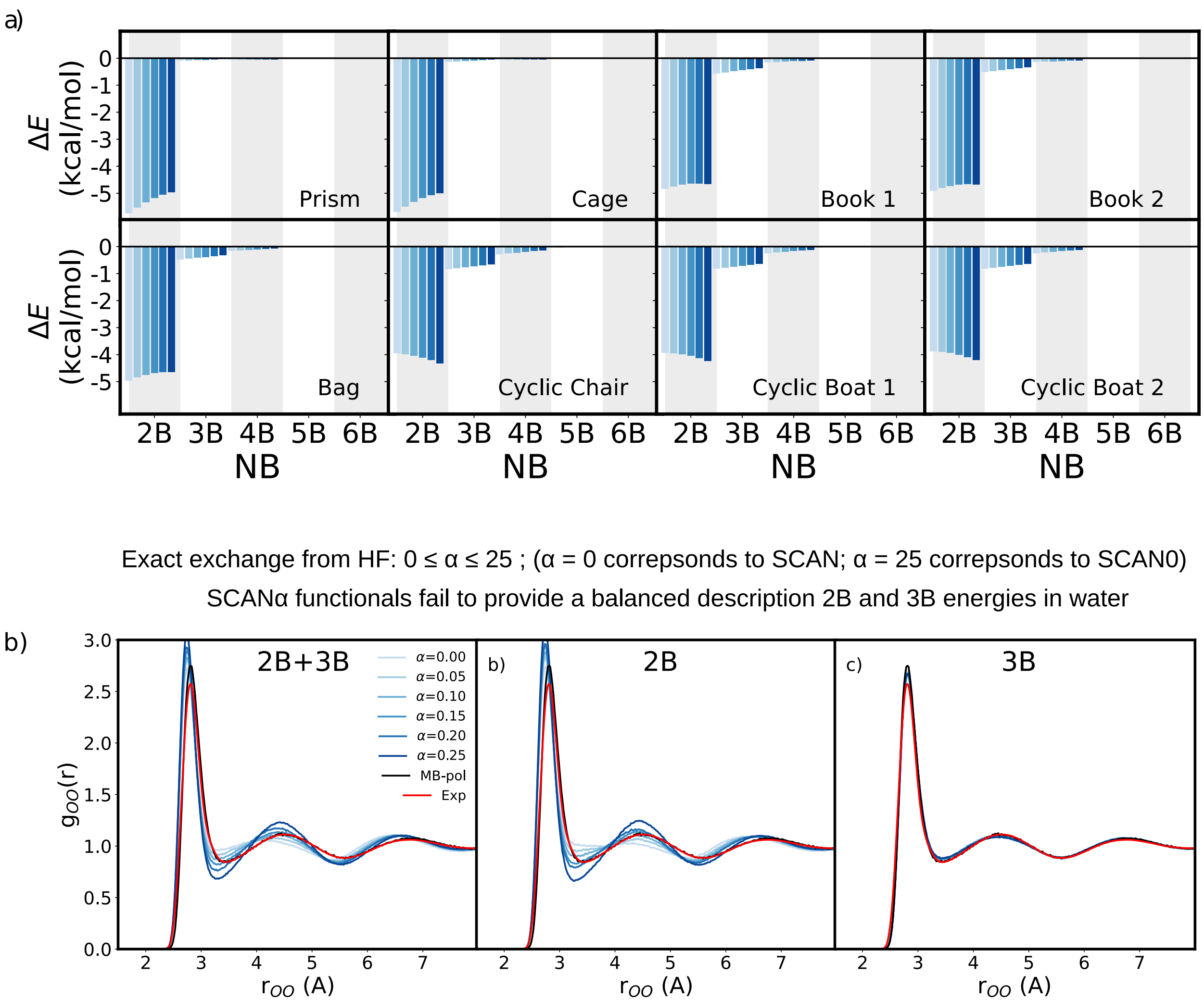

FIG. 5. Panel (a) shows the errors (in kcal/mol) associated with the MBD of the interaction energy of the the first eight low-energy isomers of the water hexamer computed with SCANα functionals, where α is the fraction of Hartree-Fock exchange, ranging from $0.00 \leq \alpha \leq 0.25$. Panel (b) systematically relates the 2B energy error in the MBD with an over-structured $g_{\mathrm{OO}}(r)$ caculated from NPT MD simulatons using the MB-SCANα PEFs. Figure adapted with permission from J. Chem. Theory Comput. **17**, 3739–3749 (2021). Copyright 2021 American Chemical Society.

kcal/mol) is largely responsible for the qualitivative differences observed in the description of liquid water as predicted by the MB-SCAN and MB-SCAN0 PEFs.[101]

Varying the fraction of Hartree-Fock exchange can shift the structure of liquid water from an overly disordered liquid state at $\alpha = 0$ to a more ice-like liquid state at $\alpha = 0.25$. While the effect of modulating α is consistent with AIMD simulations of liquid water described using SCAN and SCAN0, MB-SCANα simulations caried out in the NPT ensemble were unable to exactly reproduce the oxygen-oxygen RDF and the liquid density of the corresponding AIMD simulations. DFAs that suffer from delocalization errors are particularly sensitive to size-dependent density-driven artifacts between gas and condensed phases, wherein the short-range physics is fundamentally different between the two phases. As described in Section III, the MB-DFT PEFs are trained on gas-phase data and rely on a description of classical polarization to account for long-range many-body interactions. Therefore, within the current MB-DFT framework, they cannot capture quantum-mechanical effects that extend into n-body terms with $n \geq 4$. While the current MB-DFT framework is consistent with the MBE calculated at the CCSD(T)/CBS level of theory,[213] the analyses reported in Ref. 71 indicate that the MBEs calculated with DFAs that severely suffer from delocalization errors display a different convergence that cannot quantitatively be reproduced by the current MB-DFT framework.

B. The density-corrected SCAN functional with chemical accuracy

As discussed in the previous section, the SCAN functional deviates significantly from chemical accuracy in predicting the energetics of various water systems. Figure 5 illustrates that the delocalization error in SCAN is manifested primarily in the 2-body energies, which is in line with other studies.[70,98,99] By adding a fraction of Hartree-Fock exchange through the adiabatic connection formula, the delocalization error is reduced to a minimum for of $0.10 \leq \alpha \leq 0.15$ Hartree-Fock exchange.[99,101] Adding more than 15% Hartree-Fock exchange leads to an increase in the functional-driven error that results in deteriorated accuracy.[101] Despite providing better agreement with experimental data than SCAN for liquid water, MB-SCANα (α=0.15) still produces a more compact water structure at ambient conditions.[101]

The density-corrected DFT formalism described in Section IVB yields remarkable agreement for the SCAN functional, demonstrating that minimizing the density-driven error effectively elevates the accuracy of SCAN towards CCSD(T) accuracy.[70,277] This further supports the overall robustness of SCAN which is associated with its ability to satisfy all the 17 exact constraints known for a meta-GGA DFA.[91] Figure 6(a) depicts the errors in binding energies of the 38 low-energy isomers of the $(H_2O)_{n=2-10}$ clusters included in the BEGDB dataset.[301] Relative to the CCSD(T)/CBS reference values, the SCAN functional displays a mean unsigned error (MUE) of 5.69 kcal/mol compared to 0.54 kcal/mol associated with DC-SCAN.[277] Interestingly, adding a dispersion correction to both SCAN and DC-SCAN deteriorates their accuracy. This can be explained by considering that the SCAN functional form contains terms representing short- and mid-range components of the dispersion energy,[51,91] which appears to be sufficient for accurately describing the interactions between water molecules.[70,277]

It is well known that self-consistent densities of GGA functionals can result in significant overbinding and lead to spurious fractional charges on separated fragments.[274,303] Such fractional charge errors, a byproduct of electron over-delocalization, are strongly reduced by the Hartree-Fock density. Note that, even for overlapped and interacting fragments, the Hartree-Fock density partitions the density among the fragments in a more correct manner than the self-consistent density.[277] In fact, beyond neutral water, DC-SCAN improves the accuracy of the protonated and deprotonated water clusters contained in the WATER27 dataset, reducing the MUE from 9.89 kcal/mol displayed by SCAN to 1.43 kcal/mol displayed by DC-SCAN as shown in Figure 6(b).[304] For reference, the WATER27 dataset includes 14 neutral water clusters [$(H_2O)_n$, with $n = 2-6, 8, 20$], 5 protonated water clusters [$(H_3O^+(H_2O)_n$, with $n = 1-3, 6$], 7 deprotonated water clusters [$OH^-(H_2O)_n$, with $n = 1-6$], and 1 autoionized water cluster [$H_3O^+(H_2O)_4OH^-$]. It should be noted that for compact systems, such as the water monomer, the self-consistent SCAN density provides better energetics compared to the Hartree-Fock density. This is illustrated in Figs. 6(c,d) that depict the error in the water monomer's distortion energies for SCAN and DC-SCAN, respectively. On the other hand, using the Hartree-Fock density in DC-SCAN effectively removes the delocalization error from the 2-body energies of neutral and protonated water clusters as shown in Figure 6(e,f), which brings both binding and interaction energies of different water clusters very close to CCSD(T) reference values. Due to large density-driven errors, deprotonated water clusters display large 2-body and 3-body errors, which can only be mitigated by DC-SCAN as shown in Figure 6(g).

Figure 7(a,b) shows that the MB-DFT PEF based on DC-SCAN, MB-SCAN(DC), is able to accurately reproduce the experimental oxygen-oxygen RDF as well as the MB-pol tetrahedral-order parameter of liquid water at 298 K. As shown in Figure7(c), MB-SCAN(DC) also correctly predicts the temperature-dependence of the density of liquid water, with a deviation of $\sim$0.1 g/cm^3 across the entire tempreature range. Additionally, Fig-

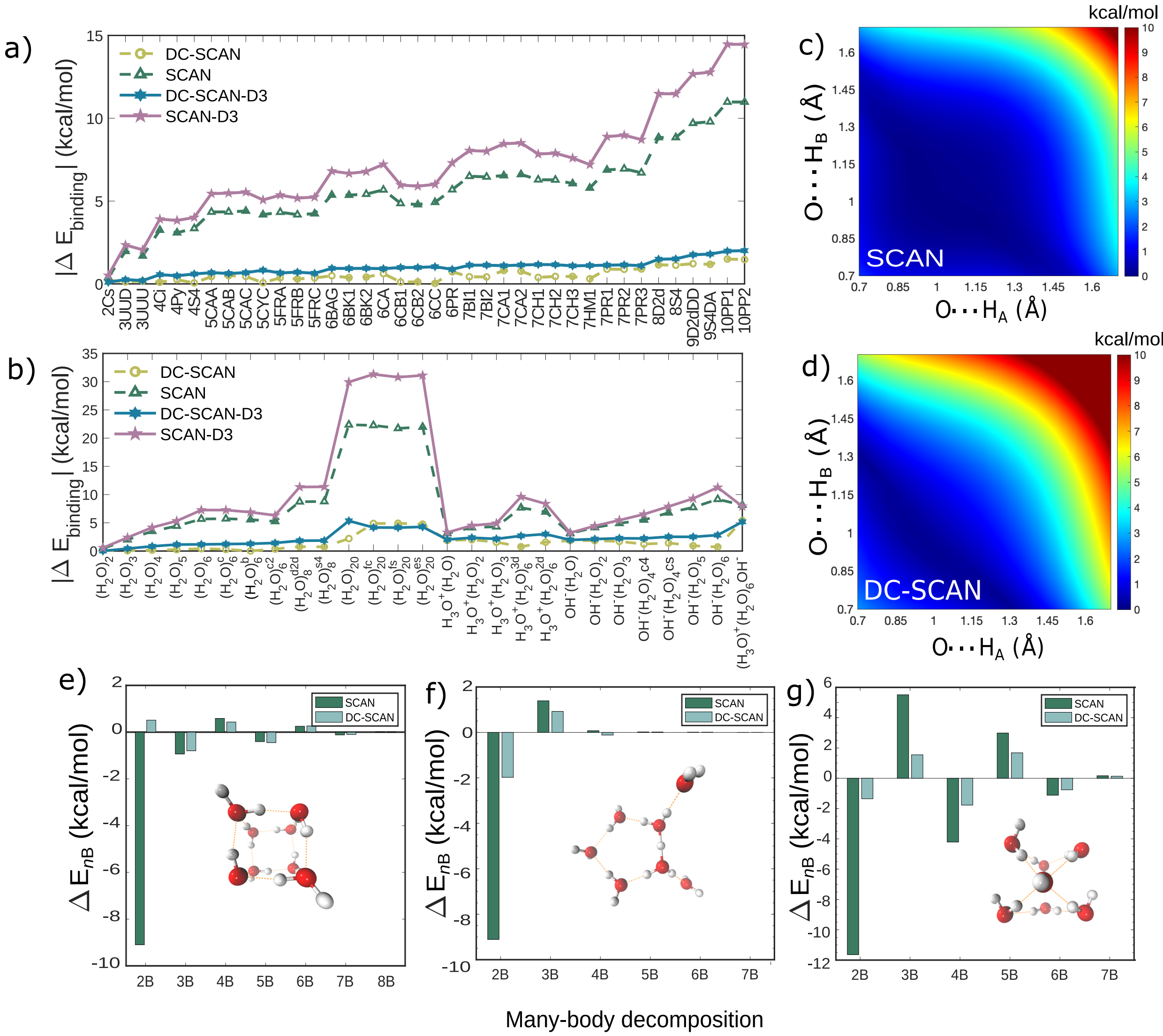

FIG. 6. Panel (a) shows absolute errors binding energies calculated for the neutral water cluster subset of the BEGDB dataset using SCAN, DC-SCAN, SCAN-D3, and DC-SCAN-D3 with respect to the CCSD(T)/CBS reference value. Panel (b) shows absolute errors in binding energies calculated for the WATER27 dataset structures using SCAN, DC-SCAN, SCAN-D3, and DC-SCAN-D3, with respect to the CCSD(T)/CBS benchmark.[302]. Panel (c) and (d) describe the errors in the distortion energies of a free water monomer from a) SCAN and b) DC-SCAN for $O \cdots H_A$ and $O \cdots H_B$ distortions, relative to CCSD(T). The color scale indicates the absolute error. Adapted with permission from J. Chem. Theory Comput. **18**, 4745–4761 (2022). Copyright 2021 American Chemical Society. Errors relative to CCSD(T)-F12b reference values for each nB energy contribution to the interaction energies calculated with SCAN and DC-SCAN for the (e) $(H_2O)_8^{(D2d)}$ (f) $H_3O^+(H_2O)_6^{(2d)}$ (g) $OH^-(H_2O)_6$. Reprinted with permission from S. Dasgupta, *et al.*, Nat. Commun. **12**, 1–12 (2021); licensed under a Creative Commons Attribution (CC BY) license.

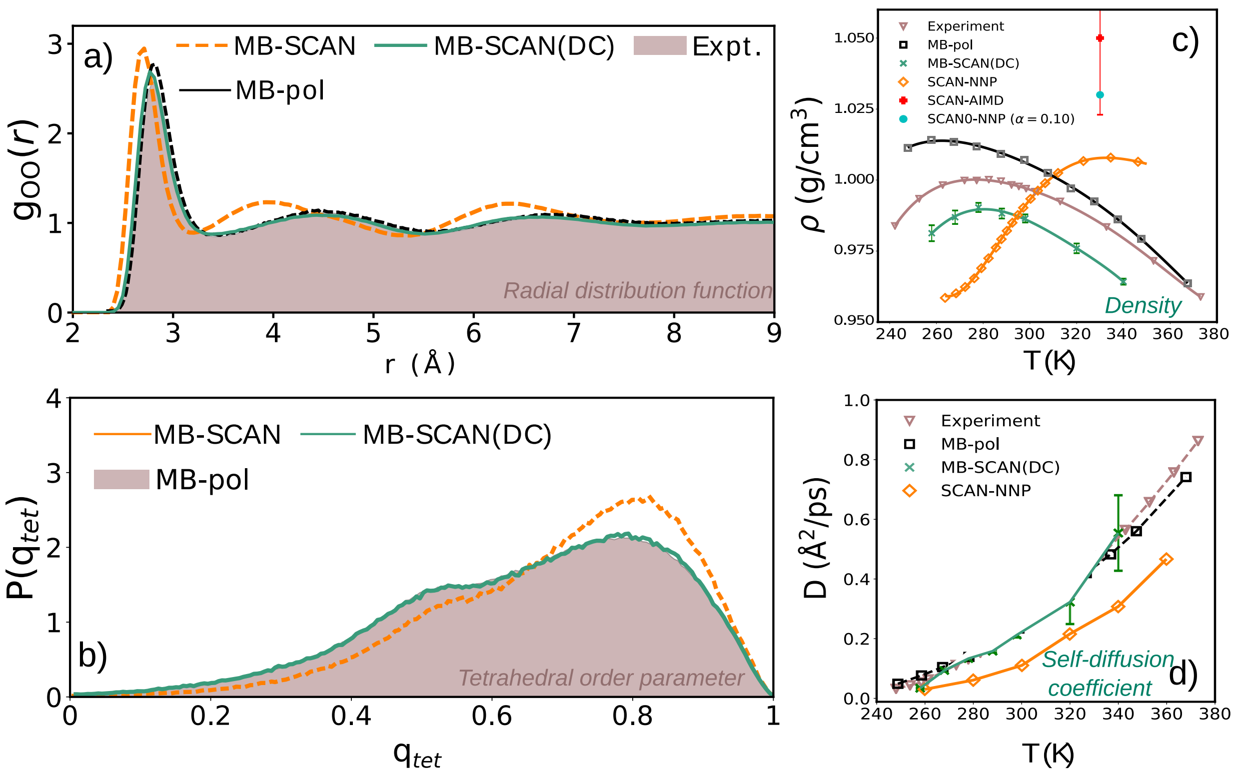

FIG. 7. (a) Oxygen-oxygen (g_{OO}) radial distribution function (RDF) calculated from NPT simulations carried out with the MB-SCAN(DC) PEF at 298 K and 1 atm. (b) Distributions of the tetrahedral order parameter, q_{tet}, calculated from MD simulations in the NPT ensemble at 298 K and 1 atm with with the MB-SCAN(DC) PEF (c) 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,[51] SCAN-NNP,[102] and SCAN0-NNP (with 10% Hartree-Fock exchange)[99] simulations. The MB-pol results are from ref. 213, while the experimental data are from the NIST Chemistry WebBook.[305] (d) Temperature-dependence of the self-diffusion coefficient of liquid water calculated from NVE simulations carried out with the MB-SCAN(DC) PEF. Figure adapted from S. Dasgupta, *et al*, Nat. Commun. **12**, 1–12 (2021); licensed under a Creative Commons Attribution (CC BY) license.

ure 7(d) demonstrates that MB-SCAN(DC) predicts the self-diffusion coefficient of liquid water between 250 K and 340 K in excellent agreement with the corresponding experimental values.

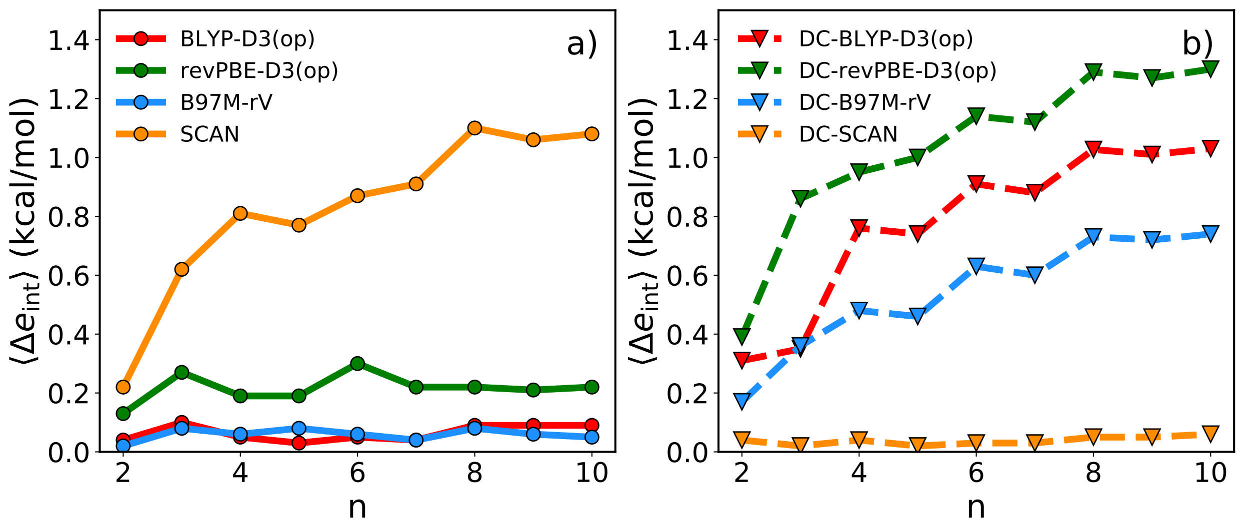

FIG. 8. Mean absolute error per molecule, $\langle \Delta e_{\text{int}} \rangle$, for the interaction energy as a function of cluster size for $(H_2O)_n$ for $n = 2 - 10$ predicted by (a) DFT and (b) DC-DFT. $\langle \Delta e_{\text{int}} \rangle$ is shown relative relative to the CCSD(T)/CBS reference values. Figure reprinted with permission from J. Chem. Theory Comput. **18**, 3410–3426 (2022). Copyright 2022 American Chemical Society.

VII. MANY-BODY POTENTIALS FROM DC-DFT FUNCTIONALS

The interplay between functional-driven and density-driven errors effectively determines the accuracy of a DFA.[71,273,287] Understanding the roles played these errors in DFT models of water is therefore important for: (1) rationalizing the predictive capabilities of DFT-based simulations of water, (2) defining theoretical considerations for future development of DFT and DFT-based models of water with improved accuracy, and (3) identifying DFAs suitable for simulations of reactive processes in water within a hybrid DFT/MB-DFT scheme.

A. Functional- and density-driven errors in water clusters: DC-DFT

It is necessary to first understand the role of ΔE_{FD} and ΔE_{DD} in predicting molecular interactions in gas-phase water clusters. In this section, we revisit some of the representative GGA DFAs discussed in Section VA, namely, BLYP-D3(op) and revPBE-D3(op), and meta-GGA functionals, namely B97M-rV and SCAN. On a technical note, the optimized-power D3(op)[306] empirical dispersion correction is used for revPBE and BLYP, rather then the original D3 (zero damping)[307] parameters, as the latter was shown to be erroneous for revPBE, particularly in predicting the 2-body energies.[71] Figure 8(a) shows the mean absolute error (MAE) in the interaction energy per water molecule for all of the 38 clusters included in the BEGDB dataset relative to the CCSD(T)/CBS reference values.[301,308]. The MAE per molecule, $\langle \Delta e_{\mathrm{int}} \rangle$, is given by[71]

$$\langle \Delta e_{\mathrm{int}} \rangle = \frac{1}{N} \sum_{i=1}^{N} \Delta e_{\mathrm{int},i} \tag{20}$$

where $\Delta e_{\mathrm{int}} = \left| E_{\mathrm{int}}^{\mathrm{model}} - E_{\mathrm{int}}^{\mathrm{ref}} \right| / n$. Here n is the number of water molecules in the $(H_2O)_n$ cluster, and N is the number of isomers of a given $(H_2O)_n$ cluster in the data set. Comparing panels (a) and (b) of Figure 8, it is apparent that many DFAs benefit from error compensation between ΔE_{FD} and ΔE_{DD}. This is clearly the case of BLYP-D3(op) and revPBE-D3(op), for which $\langle \Delta e_{\mathrm{int}} \rangle$ increases with n when the energy functional is evaluated on the Hartree-Fock density.

An alternative but useful interpretation of the DC-DFT energies is that *not only* are the density-driven errors "minimized", but the physical robustness of an approximate functional is *exposed*. Keeping in mind that the Hartree-Fock density is overlocalized compared to the exact density, the magnitude of the density-driven errors introduced by the Hartree-Fock density is still appreciably smaller than the delocalization error displayed by semi-local DFAs. Therefore, an analysis of Eq. 15 suggests that a density-corrected energy can be thought of as $E^{\mathrm{DC}} \sim \Delta E_{\mathrm{FD}}$ for DFAs prone to large delocalization error. By approximating the functional-driven errors, Figure 8(b) reveals that the capability of certain DFAs (e.g. BLYP-D3(op), revPBE-D3(op),and B97M-rV) in representing interactions in water may be sensitive to the size of the system. The opposite trend is observed for SCAN. The MAE increases as n for self-consistent SCAN, while this size-sensitivity is suppressed by DC-SCAN. This suggests that SCAN is a special case for which $|\Delta E_{\mathrm{FD}}| \ll |\Delta E_{\mathrm{DD}}|$. As a result, $\langle \Delta e_{\mathrm{int}} \rangle$ is significantly smaller for DC-SCAN than SCAN calculations. Using the water hexamer as a benchmark system, $\langle \Delta e_{\mathrm{int}} \rangle$ drops from 0.87 kcal/mol for SCAN to 0.03 kcal/mol for DC-SCAN, with the latter effectively reproducing the CCSD(T)/CBS reference energies.

Generally speaking, when functional-driven errors dominate in a given DFA, Δe_{int} increases with system's size. For revPBE-D3(op) it is clear that functional-driven errors dominate, while ΔE_{FD} and E_{DD} are similar in magnitude for BLYP-D3(op) and B97M-rV, leading to nearly complete error cancellation. Even within DC-DFT, it should be noted that DC-B97M-rV is in fact still closer to the CCSD(T)/CBS reference values for all the clusters included in the analyses of Ref. 71, suggesting that (1) B97M-rV is not as sensitive to density-driven errors as the other DFAs and (2) B97M-rV can serve as reliable DFA for hybrid DFT/MB-DFT simulations of aqueous environments.

The important lesson in Figure 8 is that the size sensitivity of functional-driven and density-driven errors has direct implications for DFT-based and AIMD simulations of extended systems (e.g., liquid water, ice). A few general rules can be extracted from the analyses discussed in Ref. 71: (1) ΔE_{FD} depends on the system's size for DFAs where $\Delta E_{\mathrm{FD}} > E_{\mathrm{DD}}$, (2) for DFAs where $\Delta E_{\mathrm{DD}} > E_{\mathrm{FD}}$ significantly, DC-DFT *elevates* the accuracy of said DFAs, and (3) the size-dependence of functional-driven and density-driven errors in semi-local DFAs shines light onto the error compensation that allows certain DFAs to be successful in representing some properties of liquid water while, at the same time, failing to predict the properties of water clusters.[4,309–313].

B. Functional- and density-driven errors in liquid water: MB-DFT(DC)

The MB-DFT framework enables the acceleration of DC-DFT simulations, for which an efficient "on the fly" DC-DFT simulation scheme is currently unavailable. As seen in Section VI, the chemical accuracy found in DC-SCAN for water clusters translates into a correct description of various properties of liquid water calculated from MB-SCAN(DC) simulations. Figure 9 shows the oxygen-oxygen RDFs for the MB-revPBE-D3(op) and MB-B97M-rV PEFs, along with their DC-DFT analogs, MB-revPBE-D3(op)(DC) and MB-B97M-rV(DC). In general, the systematic raising of E_{int} in the gas phase by the density correction translates into a more "disordered" local structure of liquid water. This is apparent in the case of MB-revPBE-D3(op)(DC), where the density correction breaks the solvation structure up to the first solvation shell, as suggested by the flattening of the oxygen-oxygen RDF beyond the first peak. As revPBE-D3(op) includes semi-local correlation, the minimization of E_{DD} allows the D3(op) correction to improve the description of liquid water predicted by DC-revPBE-D3(op), which was shown to predict correct electrostatics.[71] Interestingly, the oxygen-oxygen RDF predicted by MB-B97M-rV(DC) is slightly more structured than that of MB-B97M-rV, an effect opposite to that seen for revPBE-D3(op), which results from the overlocalized nature of the Hartree-Fock density. Overall, B97M-rV is found to be a robust density functional, with low sensitivity to density-driven errors. Since DC-B97M-rV and MB-B97M-

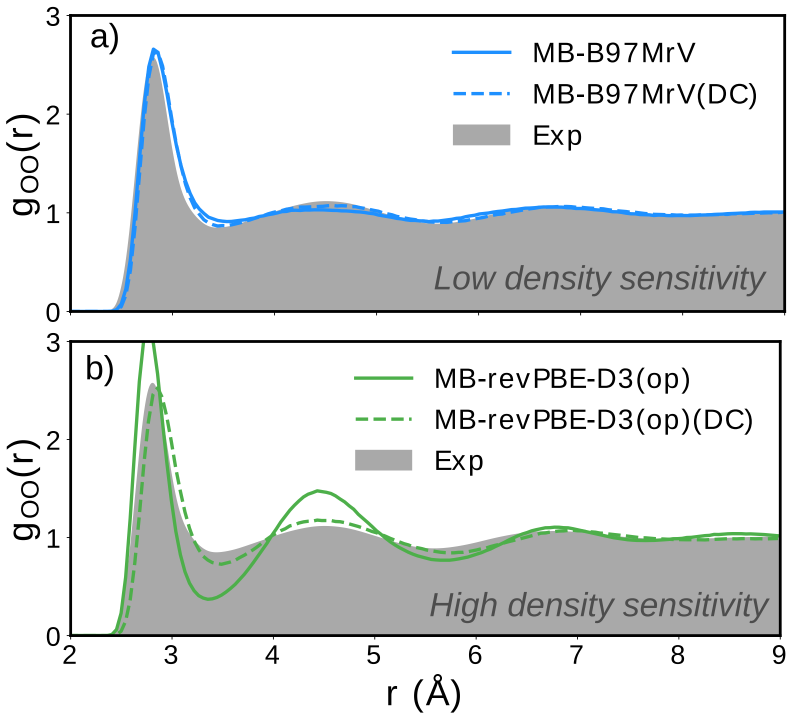

FIG. 9. Comparison between MB-DFT and density-corrected MB-DFT(DC) models in predicting the structure of liquid water. The oxygen-oxygen RDF, $g_{OO}(r)$, calculated from MD simulations carried out in the NPT ensemble (T=298 K ; P=1 atm) with (a) MB-BLYP-D3(op),(DC), (b) MB-revPBE-D3(op)(DC), (c) MB-B97M-rV(DC), and (d) MB-SCAN(DC). Figure adapted with permission from J. Chem. Theory Comput. **18**, 3410 18, 3410–3426 (2022). Copyright 2022 American Chemical Society.

rV(DC) exhibit similar accuracy to B97M-rV and MB-B97M-rV, respectively, B97M-rV can be considered as a balanced compromise between accuracy and efficiency for MB-DFT and DFT/MB-DFT simulations for aqueous phase chemistry.

VIII. MANY-BODY POTENTIALS FROM MACHINE-LEARNED FUNCTIONALS

As machine learning approaches continue to solidify their place in DFT,[117,314–317] the DM21 functional represents a significant milestone in the field, as it was trained on exact constraints including fractional charge (FC) and fractional spin (FS) constraints, in combination with atomic and molecular data.[115] Recent work highlights that machine-learning DFAs

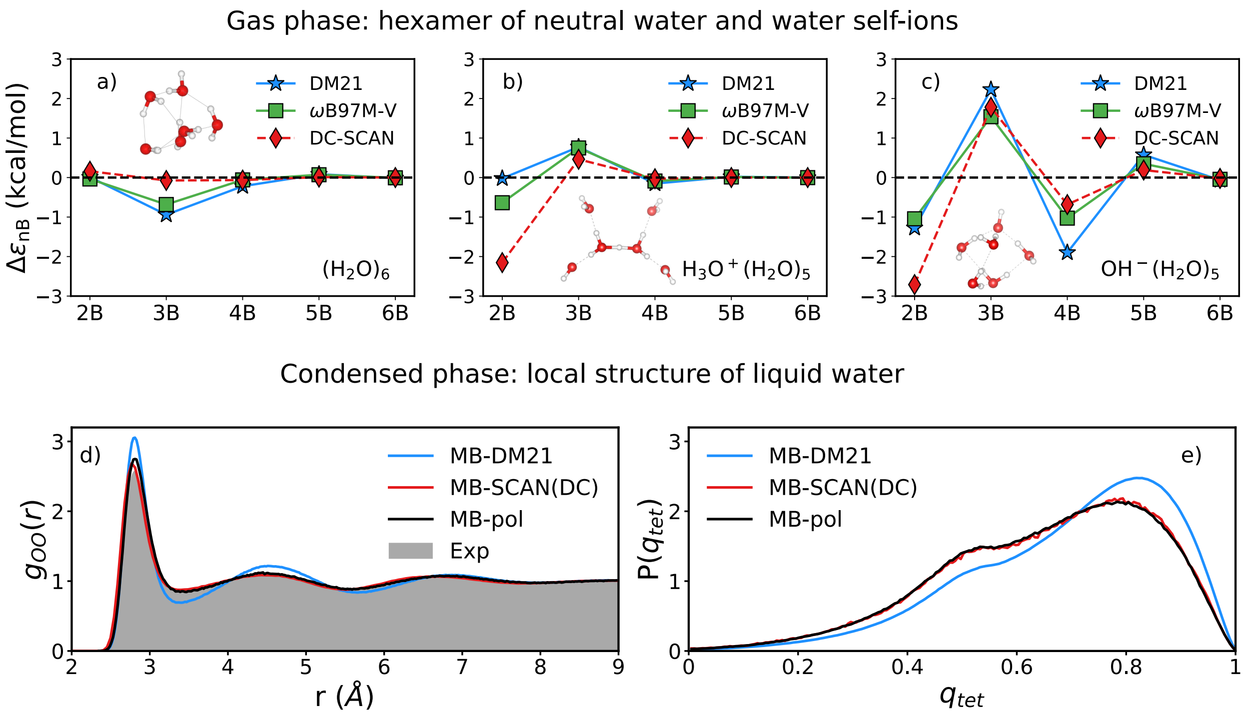

FIG. 10. Top: Errors (in kcal/mol) associated with individual n-body contributions to the interaction energies of the lowest-energy isomers of the $(H_2O)_6$ (a), $H_3O^+(H_2O)_5$ (b), and $OH^-(H_2O)_5$ (c) clusters relative to the corresponding CCSD(T) reference values. Bottom: Oxygen-oxygen radial distribution function (d), and tetrahedral order parameter distribution (e) calculated from MD simulations carried out with MB-DM21 at 298 K and 1 atm. Figure adapted from E. Palos, E. Lambros, S. Dasgupta, and F. Paesani, "Density functional theory of water with the machine-learned DM21 functional," J. Chem. Phys. **156**, 161103 (2022), with the permission of AIP Publishing.

capable of quantitatively describing the properties of water can serve as an important step toward a universal DFA with transferability across phases.[114,318] The deep-learned local-hybrid DM21 functional was shown to outperform conventional DFAs such as SCAN and ωB97X-V in predicting the energetics of a large dataset of diversely-bonded compounds.[115,319] While DM21 succeeds in modeling some systems that typically represent a challenge to most DFAs, it is significantly expensive relative to any other DFA within the fourth rung of Jacob's ladder. For this reason, condensed-phase AIMD simulations using DM21 are not currently feasible.

A. DM21 as a case-study: Neutral, protonated and deprotonated water clusters

The acceleration of simulations based on high-accuracy electronic structure methods is the forte of the MB-DFT framework,[70,71,101,104,169] readily enabling a description of water from the gas to the liquid phase as predicted by DM21. In a recent study,[212] the accuracy of DM21 was assessed for neutral, protonated and deprotonated water clusters, and the properties of liquid water were investigated using the newly developed MB-DM21 PEF. As shown in Figure 10(a), in the gas phase, DM21 predicts the MBD of the interaction energy of the prism isomer of the water hexamer with similar accuracy to ωB97M-V. Interestingly, DM21 predicts the 2-body energy virtually at the level of CCSD(T)/CBS. However, the 3-body energy displays an error of ~ 1 kcal/mol. This is an interesting result, as failure to accurately represent n-body energies in DFT models is usually attributed to the delocalization error.[278] As a result, DM21 predicts the MBD of the interaction energies of the water hexamer isomers with lower accuracy than DC-SCAN. Since DM21 is, in principle, free of delocalization error, this discrepancy in the 3-body energies is attributed to the size-dependence of functional-driven errors arising potentially from the training-set of the ML potential[115] that defines the semi-local portion of the exchange-correlation functional, which accounts for the description of static correlation.[212,268,320] This is opposite to what was observed for DC-SCAN, where the error decreases as n increases.

In the case of protonated water clusters, DM21 outperforms both ωB97M-V and DC-SCAN in predicting the 2-body energies, quantitatively reproducing the CCSD(T) reference energies, while all three DFAs predict 3-body energies with similar accuracy, with DM21 displaying a larger error than DC-SCAN by only ~ 0.3 kcal/mol. The notoriously challenging deprotonated water clusters represent a different story, where DM21 provides the smallest 2-body error but relatively larger errors than ωB97M-V and DC-SCAN for the 3-body, 4-body, and 5-body energies. Figure 10(c) shows that the n-body errors oscillate between positive and negative values for all three functionals, with higher absolute errors in individual n-body energies associated with DM21. Since DM21 is trained on both FC and FS constraints, the better agreement between DM21 and the reference CCSD(T) 2-body energies may indicate that DM21 is capable of providing a better description of the static correlation of the ionized clusters, relative to DC-SCAN and ωB97M-V. In general, the three functionals are less accurate for the deprotonated water clusters than the corresponding neutral and protonated clusters. As the density of deprotonated water is more diffuse than in neutral water, this suggests that all three functionals may systematically overestimate the energy of the highest occupied molecular orbital (HOMO) of deprotonated water clusters, which is likely due to incorrect features of the functionals in the asymptotic limit.[321–323] As DM21 is a local-range-separated hybrid functional, further investigation on its ability to describe static correlation in processes involving proton transfer, as in the Grotthuss mechanism, is warranted.

B. The MB-DM21 potential

The structural properties of liquid water as predicted by NPT simulations carried out at 298 K and 1 atm with the MB-DM21 PEF are shown in Figure 10(d-e). Unexpectedly, MB-DM21 predicts a slightly overstructured and more tetrahedral liquid phase relative to MB-SCAN(DC)[70,71] and MB-pol[163–165,213], which, on the other hand, are in good agreement with experiment. This finding supports the possibility of relatively large functional-driven errors in the MLP term of DM21, as it is trained on atomic and molecular data, leading to incomplete error cancellation in the condensed phase.[71,320] From Figure 10(d-e), it appears that DM21 has limited ability in predicting the thermodynamic properties of liquid water. In this regard, recent work showed that DM21 systematically underestimates the liquid density, overestimates the heat of vaporization, and provides a poor description of the isothermal compressibility of liquid water in the supercooled regime.[71]

In summary, functional-driven errors in DM21 are non-negligible and, due to their sensitivity to system size, affect the overall accuracy of DM21 when applied to aqueous systems. These functional-driven errors have direct effects on the ability of DM21 to describe liquid water. Furthermore, while DM21 represents only one ML-DFA, this case study suggests that improving the functional form and physical content of machine-learned DFAs should allow for further reducing functional-driven errors and, in turn, enable accurate representations of aqueous systems from gas-phase clusters to the liquid phase.

IX. MANY-BODY POTENTIALS IN POLARIZABLE QM/MM

The MB-DFT PEFs may be used to elevate the accuracy of multiscale modeling of diverse chemical processes in solution.[263,264,324] As chemical reactions in solution pose a challenge to both purely quantum mechanical (QM) and classical molecular mechanics (MM) techniques, a hybrid QM/MM approach may be used to represent the system of interest as partitioned into a (QM) subsystem, $\mathcal{S}$, that contains the reactive species and is commonly treated at the DFT level, and the environment, $\mathcal{E}$, that modulates the reaction which is usually described by a (MM) force field.[325–327] QM/MM has been successfully used in simulations of chemical systems of varying complexity, ranging in applications from modeling enzymatic reactions,[328,329] to chemical reactions in solution,[330,331] to spectroscopy of biomolecular complexes.[330,332–338]

The total QM/MM Hamiltonian of the molecular system of interest is given by

$$\hat{H} = \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} + \hat{H}_{\mathrm{QM/MM}} \tag{21}$$

where $\hat{H}_{\mathrm{QM}}$ stores all information pertaining to the electronic density of $\mathcal{S}$, $\hat{H}_{\mathrm{MM}}$ describes the MM atoms that constitute $\mathcal{E}$, and $\hat{H}_{\mathrm{QM/MM}}$ represents the interaction between $\mathcal{S}$ and $\mathcal{E}$. In the polarizable QM/MM formulation,[263,324,338,339] $\hat{H}_{\mathrm{QM/MM}}$ couples the properties of the QM electronic density with the MM induced point dipoles (IPDs). Within this scheme, the AMOEBA force field has received recent attention,[340,341] as QM/AMOEBA has been shown to achieve higher accuracy compared to conventional electrostatic-embedding formulations by achieving full mutual polarization between $\mathcal{S}$ and $\mathcal{E}$.[342] However, since AMOEBA is a fully classical PEF, QM/AMOEBA cannot overcome the energy discontinuity at the QM/MM boundary.[264] Energy discontinuities at the QM/MM boundary emerge from the inaccuracy of the force field used to represent the MM region relative to the *ab initio* method used in the QM region.[343,344] Since chemical reactions in solution typically occur within the diffusion limit, the QM/MM system must be adaptively repartitioned at regular intervals to prevent the diffusive breakup of the QM region. One possible way to minimize energy discontinuities when molecules transition between QM and MM representations consists in adding computationally expensive transition layers that smoothly average between the QM and MM energies as molecules transition between layers.[345–347]

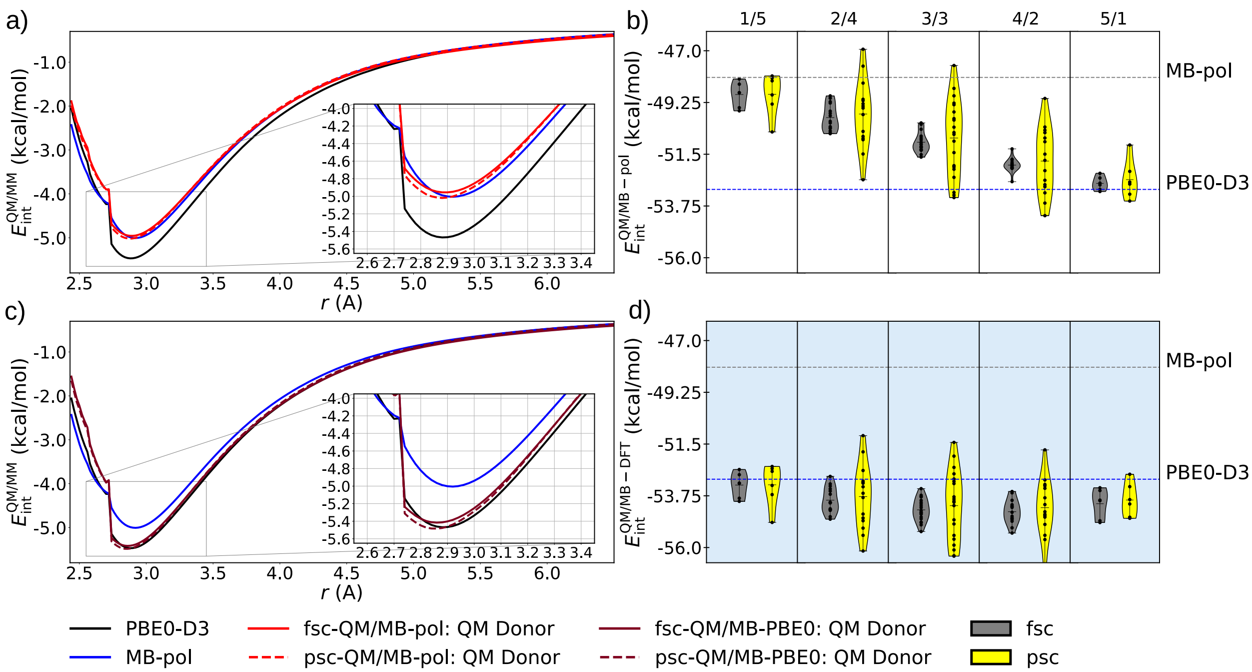

FIG. 11. Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations using DFT and MB-DFT potentials in fully-self-consistent (fsc) and partially-self-consistent (psc) QM/MM. Comparison of the QM/MM interaction energy for PBE0-D3, MB-PBE0-D3 and MB-pol over a reference $(H2O)_2$ potential energy surface scan. Panel (a) displays QM/MB-pol results with the QM water molecule as the hydrogen-bond donor. Analogously, panel (c) and QM/MB-DFT (PBE0-D3/MB-PBE0-D3) results with the QM water molecule as the hydrogen-bond donor. Panels (b) and (d) respectively show the distributions of QM/MM interaction energies of a water hexamer as waters are successively added into the QM region for QM/MB-pol and QM/MB-PBE0.[263] Figure reprinted with permission from J. Chem. Theory Comput. **16**, 7462–7472 (2020). Copyright 2020 American Chemical Society.

A. Quantum Mechanics/Many-Body Molecular Mechanics

Since, by construction, the MB-DFT PEFs are many-body in nature and polarizable, they provide a robust physics-based representation of $\hat{H}_{\mathrm{QM/MM}}$ without unphysical discontinuities at the quantum/classical boundary, effectively removing the need for transition layers. Within DFT/MB-DFT, the accuracy of the DFT level of theory chosen to represent the QM region is effectively extended to the MM region by coupling it to a corresponding MB-DFT representation of the environment. Therefore, in the DFT/MB-DFT scheme the entire system ($\mathcal{S} + \mathcal{E}$) is effectively described by the general Hamiltonian:

$$\hat{H} = \underbrace{\hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{MB-QM}} + \hat{H}_{\mathrm{QM/MB-QM}}}_{\approx \hat{H}_{QM}}. \tag{22}$$

As the MB-DFT polarization energy is variational with respect to the polarization degrees of freedom, the coupled QM/MM equations and energy can be obtained self-consistently, starting from a variational energy functional of the QM density and the MB-DFT IDPs. From this, a fully self consistent (fsc) representation of the QM/MM polarization energy is realized by equilibrating the IDPs at each SCF cycle[348,349], though it is possible, for purposes of efficiency, to use a partially self-consistent scheme (psc) where the IDPs respond only the QM density polarized by the permanent external field.[350] For a detailed derivation of the QM/MB-MM working equations, we refer the reader to Ref. 263

Understanding chemical reactions in solution requires a full characterization of the solvent role.[293] Within a QM/MM scheme, this requires a seamless and accurate representation of the interactions between QM and MM molecules. As an example, a comparison among QM, MB-MM, and QM/MB-MM interaction energies for a water dimer is shown over a reference radial scan in Figure 11(a). It should be noted that since MB-pol was derived from CCSD(T)/CBS data, the nature of MB-pol water differs substantially from that provided by PBE0-D3, which is thus responsible for the energy differences shown in Figure 11(a). The MB-DFT potentials serve as a solution to this problem. Figure 11(c) shows that DFT/MB-DFT calculation carried out using the hybrid PBE0-D3 functional in the QM region and the corresponding MB-PBE0-D3 PEF in the MM region, allows for nearly complete removal of the energy differences between the QM and MM regions when the two water molecules are swapped.

Figure 11(b,d) shows the interaction energies of the prism isomer of the water hexamer calculated for different DFT/MB-DFT partitions. Starting with the cluster containing 1 water molecule in the QM region and 5 in the MM region (labeled as the 1/5 configuration), water molecules are successively included in the QM region until a total of five water molecules are placed in the QM region (5/1). Analogous to panels (a) and (c), panels (b) and (d) represent PBE0-D3/MB-pol and PBE0-D3/MB-PBE0-D3 energies, respectively, examining the effects of re-partitioning of the DFT region (adding/removing solvent molecules to/from the DFT region) as it occurs during DFT/MB-DFT simulations in solution. Notably, the PBE0-D3/MB-PBE0-D3 interaction energies calculated using either a partially self-consistent or a fully self-consistent scheme remain fairly close to the PBE0-D3 reference values, independently of how the water molecules are partitioned between the PBE0-D3 and MB-PBE0-D3 regions.

In summary, the coupling between a DFT subsystem with a matching-accuracy MB-DFT solvent gives rise to a physically robust formulation of DFT/MB-DFT simulations. Since this approach is general, it holds for arbitrary levels of QM theory and MB-QM potentials. In QM/MB-MM, the potential energy surface of the overall (reactive + non-reactive) system is effectively free of unphysical discontinuities as it is described by one QM-level Hamiltonian that is only represented in different ways in the QM, MM, and QM/MM regions. From a practical standpoint, the QM/MB-MM scheme enables QM representations of large systems at the cost of a -reduced- QM/MM calculation.[263,264]. Since QM/MB-MM represents, in principle, the reactive subsystem and its environment on equal footing, it is expected that QM/MB-MM will find applications in the modeling of light-driven chemical processes and spectroscopy where environmental effects have a significant role[334,337]. As of today, the fully-self-consistent (fsc) QM/MB-MM scheme has been implemented in a development version of Gaussian[351] and the partially self-consistent (psc) QM/MB-MM scheme is implemented in the open-source Layered Interacting CHEmical Models (LICHEM) package.[264,350,352] Further developments of the QM/MB-MM formulation, and its implementation in different software are the subject of ongoing work.

X. SUMMARY AND OUTLOOK

Computational condensed-phase chemistry has advanced significantly in recent years in the context of data-driven modeling. Developments in pure and physically-motivated machine-learning models of aqueous systems have found applications in accelerating molecular simulations within both non-reactive and reactive computational frameworks with quantum-mechanical accuracy. In this work, we presented a concise overview of representative data-driven models of water and aqueous systems, followed by a comprehensive description of the theory, development, and applications of a generalized class of data-driven many-body potentials for aqueous-phase simulations. As density functional theory is currently the most widely used method for quantum-chemical modeling of molecular and extended systems, the general purpose of the MB-DFT class of PEFs is to accelerate non-reactive and reactive simulations with DFT accuracy.

The generalized MB-DFT framework enables the development and use of full-dimensional data-driven PEFs that are derived from different DFAs and explicitly account for one-, two-, and three-body interactions through permutationally invariant polynomial representations of short-range interactions in the density-overlap regime, while rigorously treating long-range interactions through many-body classical polarization. Since the MB-DFT PEFs are physically motivated, they provide both a basis "grounded in reality" for gaining fundamental insights into complex many-body molecular systems and a "theoretical playground" that enables systematic analyses of the accuracy and predictive power as well as the limitations, and challenges of different density functional approximations.

The MB-DFT PEFs distinguish themselves from other classes of recently developed polarizable PEFs in that they are derived *à la carte*, meaning that all parameters are obtained from the desired DFA. In general, the MB-DFT PEFs retain the accuracy of their underlying DFA, with the exception of DFAs that are heavily subject to systematic errors. Importantly, the MB-DFT framework provides an efficient platform for the development of many-body PEFs for more complex systems than those currently accessible to PEFs based on wave-function theories, thus expanding the applicability of data-driven many-body PEFs. At the same time, the MB-DFT framework accelerates (and in certain cases *enables*) simulations based on high-level and unorthodox flavors of DFT for which the corresponding *ab initio* simulations are not feasible, such as computationally intensive levels of theory, e.g., the DM21 machine-learned functional, and non-variational methods such as DC-DFT.

Beyond providing a framework for quantitative non-reactive molecular simulations of aqueous systems, the MB-DFT framework also enables robust computational modeling of chemical reactions in solution. Specifically, when a DFT subsystem is coupled to its analogous MB-DFT environment, the entire system is effectively characterized by a seamless potential energy surface, achieved without introducing transition layers for repartitioning the DFT and MB-DFT subsystems within a hybrid DFT/MB-DFT scheme. In general, DFT/MB-DFT enables the modeling of chemical transformations in solution with DFT accuracy at the cost of conventional polarizable embedding simulations.

We believe that the MB-DFT framework described in this review will enable predictive molecular simulations of aqueous systems, both in the bulk and at interfaces. Although this review focuses on aqueous systems, the MB-DFT framework is broader in scope and applicable to generic molecules. In this regard, the software infrastructure for the development (MB-Fit)[265] and application (MBX)[261] of MB-DFT PEFs in molecular simulations (MBX) is freely available to the community. We envision future applications of MB-DFT and DFT/MB-DFT to the modeling of various molecular systems in complex environments, such as molecular fluid mixtures, electrolytes, and multi-phase interfaces where chemical reactions can take place.

ACKNOWLEDGEMENTS

We are grateful to Ethan Bull-Vulpe and Alessandro Caruso for useful discussions. We thank Chris Mundy, Greg Schenter, John Perdew, and Kieron Burke for many stimulating discussions about density functional theory and related applications to aqueous systems throughout the years. This research was supported by the National Science Foundation under Grant No. CHE-1453204. E.P. acknowledges support from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2038238, as well as the Alfred P. Sloan Foundation Ph.D. Fellowship Program under Grant No. G-2020-14067.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

Any data generated and analyzed for this study are available from the authors upon request.

REFERENCES

[1]F. Franks, *Water: A Matrix of Life*, Vol. 21 (Royal Society of Chemistry, 2000).

[2]J. Dong, R. A. Fischer, L. P. Stixrude, and C. R. Lithgow-Bertelloni, "Constraining the volume of earth's early oceans with a temperature-dependent mantle water storage capacity model," AGU Adv. **2**, e2020AV000323 (2021).

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

[4]G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many-body potential energy functions," Chem. Rev. **116**, 7501–7528 (2016).

[5]M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, "Nuclear quantum effects in water and aqueous systems: Experiment theory and current challenges," Chem. Rev. **116**, 7529–7550 (2016).

[6]T. Fransson, Y. Harada, N. Kosugi, N. A. Besley, B. Winter, J. J. Rehr, L. G. Pettersson, and A. Nilsson, "X-ray and electron spectroscopy of water," Chem. Rev. **116**, 7551–7569 (2016).

[7]K. Amann-Winkel, M.-C. Bellissent-Funel, L. E. Bove, T. Loerting, A. Nilsson, A. Paciaroni, D. Schlesinger, and L. Skinner, "X-ray and neutron scattering of water," Chem. Rev. **116**, 7570–7589 (2016).

[8]F. Perakis, L. De Marco, A. Shalit, F. Tang, Z. R. Kann, T. D. Kühne, R. Torre, M. Bonn, and Y. Nagata, "Vibrational spectroscopy and dynamics of water," Chem. Rev. **116**, 7590–7607 (2016).

[9]S. Cerveny, F. Mallamace, J. Swenson, M. Vogel, and L. Xu, "Confined water as model of supercooled water," Chem. Rev. **116**, 7608–7625 (2016).

[10]N. F. Van Der Vegt, K. Haldrup, S. Roke, J. Zheng, M. Lund, and H. J. Bakker, "Water-mediated ion pairing: Occurrence and relevance," Chem. Rev. **116**, 7626–7641 (2016).

[11]N. Agmon, H. J. Bakker, R. K. Campen, R. H. Henchman, P. Pohl, S. Roke, M. Thämer, and A. Hassanali, "Protons and hydroxide ions in aqueous systems," Chem. Rev. **116**, 7642–7672 (2016).

[12]M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. Van Der Spoel, Y. Xu, and A. E. Garcia, "Water determines the structure and dynamics of proteins," Chem. Rev. **116**, 7673–7697 (2016).

[13]O. Björneholm, M. H. Hansen, A. Hodgson, L.-M. Liu, D. T. Limmer, A. Michaelides, P. Pedevilla, J. Rossmeisl, H. Shen, G. Tocci, E. Tyrode, M.-M. Walz, J. Werner, and H. Bluhm, "Water at interfaces," Chem. Rev. **116**, 7698–7726 (2016).

[14]A. Rahman and F. H. Stillinger, "Molecular dynamics study of liquid water," J. Chem. Phys. **55**, 3336–3359 (1971).

[15]F. H. Stillinger and A. Rahman, "Improved simulation of liquid water by molecular dynamics," J. Chem. Phys. **60**, 1545–1557 (1974).

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

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

[18]K. Laasonen, M. Sprik, M. Parrinello, and R. Car, ""Ab initio" liquid water," J. Chem. Phys. **99**, 9080–9089 (1993).

[19]N. Bernstein, C. Várnai, I. Solt, S. A. Winfield, M. C. Payne, I. Simon, M. Fuxreiter, and G. Csányi, "QM/MM simulation of liquid water with an adaptive quantum region," Phys. Chem. Chem. Phys. **14**, 646–656 (2012).

[20]J. A. Barker and R. Watts, "Structure of water; a Monte Carlo calculation," Chem. Phys. Lett. **3**, 144–145 (1969).

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

[22]V. F. Petrenko and R. W. Whitworth, *Physics of Ice* (OUP Oxford, 1999).

[23]A. Luzar and D. Chandler, "Effect of environment on hydrogen bond dynamics in liquid water," Phys. Rev. Lett. **76**, 928 (1996).

[24]C. Fecko, J. Eaves, J. Loparo, A. Tokmakoff, and P. Geissler, "Ultrafast hydrogen-bond dynamics in the infrared spectroscopy of water," Science **301**, 1698–1702 (2003).

[25]J. Eaves, J. Loparo, C. J. Fecko, S. Roberts, A. Tokmakoff, and P. Geissler, "Hydrogen bonds in liquid water are broken only fleetingly," Proc. Natl. Acad. Sci. U.S.A. **102**, 13019–13022 (2005).

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

[27] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, 2012).

[28] K. Burke, "Perspective on density functional theory," J. Chem. Phys. **136**, 150901 (2012).

[29] A. M. Teale, T. Helgaker, A. Savin, C. Adano, B. Aradi, A. V. Arbuznikov, P. Ayers, E. J. Baerends, V. Barone, P. Calaminici, E. Cances, E. A. Carter, P. K. Chattaraj, H. Chermette, I. Ciofini, T. D. Crawford, F. D. Proft, J. Dobson, C. Draxl, T. Frauenheim, E. Fromager, P. Fuentealba, L. Gagliardi, G. Galli, J. Gao, P. Geerlings, N. Gidopoulos, P. M. W. Gill, P. Gori-Giorgi, A. Görling, T. Gould, S. Grimme, O. Gritsenko, H. J. A. Jensen, E. R. Johnson, R. O. Jones, M. Kaupp, A. Koster, L. Kronik, A. I. Krylov, S. Kvaal, A. Laestadius, M. P. Levy, M. Lewin, S. Liu, P.-F. Loos, N. T. Maitra, F. Neese, J. Perdew, K. Pernal, P. Pernot, P. Piecuch, E. Rebolini, L. Reining, P. Romaniello, A. Ruzsinszky, D. Salahub, M. Scheffler, P. Schwerdtfeger, V. N. Staroverov, J. Sun, E. Tellgren, D. J. Tozer, S. Trickey, C. A. Ullrich, A. Vela, G. Vignale, T. A. Wesolowski, X. Xu, and W. Yang, "DFT exchange: Sharing perspectives on the workhorse of quantum chemistry and materials science," Phys. Chem. Chem. Phys. (2022).

[30] E. Fois, M. Sprik, and M. Parrinello, "Properties of supercritical water: An ab initio simulation," Chem. Phys. Lett. **223**, 411–415 (1994).

[31] M. Sprik, J. Hutter, and M. Parrinello, "Ab initio molecular dynamics simulation of liquid water: Comparison of three gradient-corrected density functionals," J. Chem. Phys. **105**, 1142–1152 (1996).

[32] J. Ortega, J. P. Lewis, and O. F. Sankey, "First principles simulations of fluid water: The radial distribution functions," J. Chem. Phys. **106**, 3696–3702 (1997).

[33] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, "Ab initio infrared spectrum of liquid water," Chem. Phys. Lett. **277**, 478–482 (1997).

[34] P. L. Silvestrelli and M. Parrinello, "Water molecule dipole in the gas and in the liquid phase," Phys. Rev. Lett. **82**, 3308–3311 (1999).

[35] B. L. Trout and M. Parrinello, "The dissociation mechanism of H_2O in water studied by first-principles molecular dynamics," Chem. Phys. Lett. **288**, 343–347 (1998).

[36] B. L. Trout and M. Parrinello, "Analysis of the dissociation of H_2O in water using first-principles molecular dynamics," J. Phys. Chem. B **103**, 7340–7345 (1999).

[37] D. Marx, M. E. Tuckerman, and M. Parrinello, "Solvated excess protons in water: Quantum effects on the hydration structure," J. Condens. Matter Phys. **12**, A153–A159 (2000).

[38]M. Boero, K. Terakura, T. Ikeshoji, C. C. Liew, and M. Parrinello, "Hydrogen bonding and dipole moment of water at supercritical conditions: A first-principles molecular dynamics study," Phys. Rev. Lett. **85**, 3245–3248 (2000).

[39]M. Boero, K. Terakura, T. Ikeshoji, C. C. Liew, and M. Parrinello, "Water at supercritical conditions: A first principles study," J. Chem. Phys. **115**, 2219–2227 (2001).

[40]S. Izvekov and G. A. Voth, "Car-Parrinello molecular dynamics simulation of liquid water: New results," J. Chem. Phys. **116**, 10372–10376 (2002).

[41]M. Boero, M. Parrinello, K. Terakura, T. Ikeshoji, and C. C. Liew, "First-principles molecular-dynamics simulations of a hydrated electron in normal and supercritical water," Phys. Rev. Lett. **90**, 226403 (2003).

[42]G. Hura, D. Russo, R. M. Glaeser, T. Head-Gordon, M. Krack, and M. Parrinello, "Water structure as a function of temperature X-ray scattering experiments and ab initio molecular dynamics," Phys. Chem. Chem. Phys. **5**, 1981–1991 (2003).

[43]I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello, "Liquid water from first principles: Investigation of different sampling approaches," J. Phys. Chem. B **108**, 12990–12998 (2004).

[44]J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water," J. Chem. Phys. **122**, 014515 (2005).

[45]D. Prendergast and G. Galli, "X-ray absorption spectra of water from first principles calculations," Phys. Rev. Lett. **96**, 215502 (2006).

[46]J. A. Morrone and R. Car, "Nuclear quantum effects in water," Phys. Rev. Lett. **101**, 017801 (2008).

[47]T. D. Kühne, M. Krack, and M. Parrinello, "Static and dynamical properties of liquid water from first principles by a novel Car-Parrinello-like approach," J. Chem. Theory Comput. **5**, 235–241 (2009).

[48]R. A. DiStasio Jr, B. Santra, Z. Li, X. Wu, and R. Car, "The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water," J. Chem. Phys. **141**, 084502 (2014).

[49]I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, "Structure and dynamics of liquid water from ab initio molecular dynamics—comparison of BLYP, PBE and revPBE

density functionals with and without van der Waals corrections," J. Chem. Theory Comput. **8**, 3902–3910 (2012).

[50]L. Ruiz Pestana, N. Mardirossian, M. Head-Gordon, and T. Head-Gordon, "Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals," Chem. Sci. **8**, 3554–3565 (2017).

[51]M. Chen, H.-Y. Ko, R. C. Remsing, M. F. Calegari Andrade, B. Santra, Z. Sun, A. Selloni, R. Car, M. L. Klein, J. P. Perdew, and X. Wu, "Ab initio theory and modeling of water," Proc. Natl. Acad. Sci. U.S.A. **114**, 10846–10851 (2017).

[52]L. Ruiz Pestana, O. Marsalek, T. E. Markland, and T. Head-Gordon, "The quest for accurate liquid water properties from first principles," J. Phys. Chem. Lett. **9**, 5009–5016 (2018).

[53]J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," AIP Conf. Proc. **577**, 1–20 (2001).

[54]D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," Phys. Rev. Lett. **45**, 566–569 (1980).

[55]J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," Phys. Rev. B **45**, 13244–13249 (1992).

[56]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximations made simple," Phys. Rev. Lett. **77**, 3865–3868 (1996), erratum: *ibid.* **78**, 1396 (1997).

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

[58]F. Gianturco, F. Paesani, M. Laranjeira, V. Vassilenko, and M. Cunha, "Intermolecular forces from density functional theory. III. A multiproperty analysis for the Ar(^{1}S)-CO($^1\Sigma$) interaction," J. Chem. Phys. **110**, 7832–7845 (1999).

[59]F. Gianturco and F. Paesani, "Van der Waals interactions from density functional theories," in *Conceptual Perspectives in Quantum Chemistry* (Springer, 1997) pp. 337–382.

[60]A. D. Becke and E. R. Johnson, "Exchange-hole dipole moment and the dispersion interaction," J. Chem. Phys. **122**, 154104 (2005).

[61]S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," J. Comput. Chem. **27**, 1787–1799 (2006).

[62]A. Tkatchenko and M. Scheffler, "Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data," Phys. Rev. Lett. **102**, 073005

(2009).

[63]O. A. Vydrov and T. Van Voorhis, "Nonlocal van der Waals density functional made simple," Phys. Rev. Lett. **103**, 063004 (2009).

[64]O. A. Vydrov and T. Van Voorhis, "Nonlocal van der Waals density functional: The simpler the better," J. Chem. Phys. **133**, 244103 (2010).

[65]S. Grimme, "Density functional theory with london dispersion corrections," Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 211–228 (2011).

[66]A. Tkatchenko, R. A. DiStasio Jr, R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions," Phys. Rev. Lett. **108**, 236402 (2012).

[67]R. Sabatini, T. Gorni, and S. De Gironcoli, "Nonlocal van der Waals density functional made simple and efficient," Phys. Rev. B **87**, 041108 (2013).

[68]E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-charge dependent london dispersion correction," J. Chem. Phys. **150**, 154122 (2019).

[69]S. Song, S. Vuckovic, E. Sim, and K. Burke, "Density sensitivity of empirical functionals," J. Phys. Chem. Lett. **12**, 800–807 (2021).

[70]S. Dasgupta, E. Lambros, J. Perdew, and F. Paesani, "Elevating density functional theory to chemical accuracy for water simulations through a density-corrected many-body formalism," Nat. Commun. **12**, 1–12 (2021).

[71]E. Palos, E. Lambros, S. Swee, J. Hu, S. Dasgupta, and F. Paesani, "Assessing the interplay between functional-driven and density-driven errors in DFT models of water," J. Chem. Theory Comput. 18, 3410 **18**, 3410–3426 (2022).

[72]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," Phys. Rev. Lett. **91**, 146401 (2003).

[73]N. Mardirossian, L. Ruiz Pestana, J. C. Womack, C.-K. Skylaris, T. Head-Gordon, and M. Head-Gordon, "Use of the rVV10 nonlocal correlation functional in the B97M-V density functional: Defining B97M-rV and related functionals," J. Phys. Chem. Lett. **8**, 35–40 (2017).

[74]N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals," Mol. Phys. **115**, 2315–2372 (2017).

[75] A. D. Becke, "Density-functional thermochemistry. I. The effect of the exchange-only gradient correction," J. Chem. Phys. **96**, 2155–2160 (1992).

[76] A. D. Becke, "A new mixing of Hartree–Fock and local density-functional theories," J. Chem. Phys. **98**, 1372–1377 (1993).

[77] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "General performance of density functionals," J. Phys. Chem. A **111**, 10439–10452 (2007).

[78] A. D. Becke, "Perspective: Fifty years of density-functional theory in chemical physics," J. Chem. Phys. **140**, 18A301 (2014).

[79] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II," J. Chem. Phys. **121**, 5400–5409 (2004).

[80] T. Todorova, A. P. Seitsonen, J. Hutter, I.-F. W. Kuo, and C. J. Mundy, "Molecular dynamics simulation of liquid water: Hybrid density functionals," J. Phys. Chem. B **110**, 3685–3691 (2006).

[81] J. A. Morrone and R. Car, "Nuclear quantum effects in water," Phys. Rev. Lett. **101**, 017801 (2008).

[82] C. Zhang, J. Wu, G. Galli, and F. Gygi, "Structural and vibrational properties of liquid water from van der Waals density functionals," J. Chem. Theory Comput. **7**, 3054–3061 (2011).

[83] A. P. Gaiduk, C. Zhang, F. Gygi, and G. Galli, "Structural and electronic properties of aqueous nacl solutions from ab initio molecular dynamics simulations with hybrid density functionals," Chem. Phys. Lett. **604**, 89–96 (2014).

[84] B. Santra, R. A. DiStasio Jr, F. Martelli, and R. Car, "Local structure analysis in ab initio liquid water," Mol. Phys. **113**, 2829–2841 (2015).

[85] A. P. Gaiduk, F. Gygi, and G. Galli, "Density and compressibility of liquid water and ice from first-principles simulations with hybrid functionals," J. Phys. Chem. Lett. **6**, 2902–2908 (2015).

[86] A. P. Gaiduk, J. Gustafson, F. Gygi, and G. Galli, "First-principles simulations of liquid water using a dielectric-dependent hybrid functional," J. Phys. Chem. Lett. **9**, 3068–3073 (2018).

[87] L. Goerigk and S. Grimme, "Double-hybrid density functionals," Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 576–600 (2014).

[88]M. Alipour, "How well can parametrized and parameter-free double-hybrid approximations predict response properties of hydrogen-bonded systems? Dipole polarizabilities of water nanoclusters as a working model," J. Phys. Chem. A **117**, 4506–4513 (2013).

[89]J. C. Howard, J. D. Enyard, and G. S. Tschumper, "Assessing the accuracy of some popular DFT methods for computing harmonic vibrational frequencies of water clusters," J. Chem. Phys. **143**, 214103 (2015).

[90]J. Cheng and J. VandeVondele, "Calculation of electrochemical energy levels in water using the random phase approximation and a double hybrid functional," Phys. Rev. Lett. **116**, 086402 (2016).

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

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

[93]A. P. Bartók and J. R. Yates, "Regularized SCAN functional," J. Chem. Phys. **150**, 161101 (2019).

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

[95]S. Ehlert, U. Huniar, J. Ning, J. W. Furness, J. Sun, A. D. Kaplan, J. P. Perdew, and J. G. Brandenburg, "r^2SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications," J. Chem. Phys. **154**, 061101 (2021).

[96]M. Bursch, H. Neugebauer, S. Ehlert, and S. Grimme, "Dispersion corrected r^2SCAN based global hybrid functionals: r^2SCANh, r^2SCAN0, and r^2SCAN50," J. Chem. Phys. **156**, 134105 (2022).

[97]J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, "Construction of meta-GGA functionals through restoration of exact constraint adherence to regularized SCAN functionals," J. Chem. Phys. **156**, 034109 (2022).

[98]K. Sharkas, K. Wagle, B. Santra, S. Akter, R. R. Zope, T. Baruah, K. A. Jackson, J. P. Perdew, and J. E. Peralta, "Self-interaction error overbinds water clusters but cancels in structural energy differences," Proc. Natl. Acad. Sci. U.S.A. **117**, 11283–11288 (2020).

[99]C. Zhang, F. Tang, M. Chen, J. Xu, L. Zhang, D. Y. Qiu, J. P. Perdew, M. L. Klein, and X. Wu, "Modeling liquid water by climbing up Jacob's ladder in density functional theory

facilitated by using deep neural network potentials," J. Phys. Chem. B **125**, 11444–11456 (2021).

[100]L. Zhang, H. Wang, R. Car, and W. E, "Phase diagram of a deep potential water model," Phys. Rev. Lett. **126**, 236001 (2021).

[101]E. Lambros, J. Hu, and F. Paesani, "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).

[102]P. M. Piaggi, A. Z. Panagiotopoulos, P. G. Debenedetti, and R. Car, "Phase equilibrium of water with hexagonal and cubic ice using the SCAN functional," J. Chem. Theory Comput. **17**, 3065–3077 (2021).

[103]R. Wang, V. Carnevale, M. L. Klein, and E. Borguet, "First-principles calculation of water pk$_a$ using the newly developed SCAN functional," J. Phys. Chem. Lett. **11**, 54–59 (2019).

[104]T. T. Duignan, S. M. Kathmann, G. K. Schenter, and C. J. Mundy, "Toward a first-principles framework for predicting collective properties of electrolytes," Acc. Chem. Res. **54**, 2833–2843 (2021).

[105]A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics without the electrons," Phys. Rev. Lett. **104**, 136403 (2010).

[106]A. P. Bartók and G. Csányi, "Gaussian approximation potentials: A brief tutorial introduction," Int. J. Quantum Chem. **115**, 1051–1057 (2015).

[107]J. Behler, "Perspective: Machine learning potentials for atomistic simulations," J. Chem. Phys. **145**, 170901 (2016).

[108]A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, "Machine learning unifies the modeling of materials and molecules," Sci. Adv. **3**, e1701816 (2017).

[109]P. Rowe, G. Csányi, D. Alfè, and A. Michaelides, "Development of a machine learning potential for graphene," Phys. Rev. B **97**, 054303 (2018).

[110]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," Nature **559**, 547–555 (2018).

[111]G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," Rev. Mod. Phys. **91**, 045002

(2019).

[112] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, "Machine learning for molecular simulation," Annu. Rev. Phys. Chem. **71**, 361–390 (2020).

[113] P. O. Dral, "Quantum chemistry in the age of machine learning," J. Phys. Chem. Lett. **11**, 2336–2347 (2020).

[114] S. Dick and M. Fernandez-Serra, "Machine learning accurate exchange and correlation functionals of the electronic density," Nat. Commun. **11**, 1–10 (2020).

[115] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen, "Pushing the frontiers of density functionals by solving the fractional electron problem," Science **374**, 1385–1389 (2021).

[116] J. T. Margraf and K. Reuter, "Pure non-local machine-learned density functional theory for electron correlation," Nat. Commun. **12**, 1–7 (2021).

[117] R. Pederson, B. Kalita, and K. Burke, "Machine learning and density functional theory," Nat. Rev. Phys. **4**, 357–358 (2022).

[118] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).

[119] J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," Phys. Chem. Chem. Phys. **13**, 17930–17955 (2011).

[120] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, "Machine-learning approach for one-and two-body corrections to density functional theory: Applications to molecular and condensed water," Phys. Rev. B **88**, 054104 (2013).

[121] J. Behler, "Representing potential energy surfaces by high-dimensional neural network potentials," J. Condens. Matter Phys. **26**, 183001 (2014).

[122] W. J. Szlachta, A. P. Bartók, and G. Csányi, "Accuracy and transferability of Gaussian approximation potential models for tungsten," Phys. Rev. B **90**, 104108 (2014).

[123] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," J. Phys. Chem. Lett. **6**, 2326–2331 (2015).

[124] J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem. Int. Ed. **56**, 12828–12840 (2017).

[125] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," Sci. Adv. **3**, e1603015 (2017).

[126] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost," Chem. Sci. **8**, 3192–3203 (2017).

[127] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1 a data set of 20 million calculated off-equilibrium conformations for organic molecules," Sci. Data **4**, 1–8 (2017).

[128] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, "Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics," Phys. Rev. Lett. **120**, 143001 (2018).

[129] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, "Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron," Phys. Rev. Materials **2**, 013808 (2018).

[130] K. Yao, J. E. Herr, D. W. Toth, R. Mckintyre, and J. Parkhill, "The TensorMol-0.1 model chemistry: A neural network augmented with long-range physics," Chem. Sci. **9**, 2261–2269 (2018).

[131] C. Qu, Q. Yu, and J. M. Bowman, "Permutationally invariant potential energy surfaces," Annu. Rev. Phys. Chem. **69**, 151–175 (2018).

[132] M. Haghighatlari and J. Hachmann, "Advances of machine learning in molecular modeling and simulation," Curr. Opin. Chem. Eng. **23**, 51–57 (2019).

[133] S. Manzhos and T. Carrington Jr., "Neural network potential energy surfaces for small molecules and reactions," Chem. Rev. **121**, 10187–10217 (2020).

[134] P. Gkeka, G. Stoltz, A. Barati Farimani, Z. Belkacemi, M. Ceriotti, J. D. Chodera, A. R. Dinner, A. L. Ferguson, J.-B. Maillet, H. Minoux, C. Peter, F. Pietrucci, A. Silveira, A. Tkatchenko, Z. Trstanova, R. Wiewiora, and T. Lelièvre, "Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and biological systems," J. Chem. Theory Comput. **16**, 4757–4775 (2020).

[135] Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, "Ap-net: An atomic-pairwise neural network for smooth and transferable interaction

potentials," J. Chem. Phys. **153**, 044112 (2020).

[136]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine learning force fields," Chem. Rev. **121**, 10142–10186 (2021).

[137]T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer," Nat. Commun. **12**, 1–11 (2021).

[138]J. Behler, "Four generations of high-dimensional neural network potentials," Chem. Rev. **121**, 10037–10072 (2021).

[139]T. Zubatiuk and O. Isayev, "Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence," Acc. Chem. Res. **54**, 1575–1585 (2021).

[140]E. Lambros, S. Dasgupta, E. Palos, S. Swee, J. Hu, and F. Paesani, "General many-body framework for data-driven potentials with arbitrary quantum mechanical accuracy: Water as a case study," J. Chem. Theory Comput. **17**, 5635–5650 (2021).

[141]V. Zaverkin, D. Holzmüller, R. Schuldt, and J. Kästner, "Predicting properties of periodic systems from cluster data: A case study of liquid water," J. Chem. Phys. **156**, 114103 (2022).

[142]L. D. Jacobson, J. M. Stevenson, F. Ramezanghorbani, D. Ghoreishi, K. Leswing, E. D. Harder, and R. Abel, "Transferable neural network potential energy surfaces for closed-shell organic molecules: Extension to ions," J. Chem. Theory Comput. **18**, 2354–2366 (2022).

[143]S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nat. Commun. **13**, 1–11 (2022).

[144]E. Bull-Vulpe, M. Riera, A. Götz, and F. Paesani, "MB-Fit: Software infrastructure for data-driven many-body potential energy functions," J. Chem. Phys. **155**, 124801 (2021).

[145]E. Kocer, T. W. Ko, and J. Behler, "Neural network potentials: A concise overview of methods," Annu. Rev. Phys. Chem. **73**, 163–186 (2022).

[146]K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, and H. A. Scheraga, "Description of the potential energy surface of the water dimer with an artificial neural network," Chem. Phys. Lett. **271**, 152–156 (1997).

[147]X. Huang, B. J. Braams, and J. M. Bowman, "Ab initio potential energy and dipole moment surfaces of $(H_2O)_2$," J. Phys. Chem. A **110**, 445–451 (2006).

[148]A. Shank, Y. Wang, A. Kaledin, B. J. Braams, and J. M. Bowman, "Accurate ab initio and "hybrid" potential energy surfaces intramolecular vibrational energies and classical IR spectrum of the water dimer," J. Chem. Phys. **130**, 144314 (2009).

[149]Y. Wang, X. Huang, B. C. Shepler, B. J. Braams, and J. M. Bowman, "Flexible ab initio potential and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer," J. Chem. Phys. **134**, 094509 (2011).

[150]Y. Wang and J. M. Bowman, "Ab initio potential and dipole moment surfaces for water. II. local-monomer calculations of the infrared spectra of water clusters," J. Chem. Phys. **134**, 154510 (2011).

[151]T. Morawietz, V. Sharma, and J. Behler, "A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges," J. Chem. Phys. **136**, 064103 (2012).

[152]T. Morawietz and J. Behler, "A density-functional theory-based neural network potential for water clusters including van der Waals corrections," J. Phys. Chem. A **117**, 7356–7366 (2013).

[153]S. K. Natarajan, T. Morawietz, and J. Behler, "Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials," Phys. Chem. Chem. Phys. **17**, 8356–8371 (2015).

[154]T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der Waals interactions determine the unique properties of water," Proc. Natl. Acad. Sci. USA **113**, 8368–8373 (2016).

[155]B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, "Ab initio thermodynamics of liquid and solid water," Proc. Natl. Acad. Sci. U.S.A. **116**, 1110–1115 (2019).

[156]C. Schran, J. Behler, and D. Marx, "Automated fitting of neural network potentials at coupled cluster accuracy: Protonated water clusters as testing ground," J. Chem. Theory Comput. **16**, 88–99 (2019).

[157]T. E. Gartner, L. Zhang, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, "Signatures of a liquid–liquid transition in an ab initio deep neural network model for water," Proc. Natl. Acad. Sci. USA **117**, 26040–26046 (2020).

[158] P. M. Piaggi and R. Car, "Enhancing the formation of ionic defects to study the ice Ih/XI transition with molecular dynamics simulations," Mol. Phys. **119**, e1916634 (2021).

[159] S. Yue, M. C. Muniz, M. F. Calegari Andrade, L. Zhang, R. Car, and A. Z. Panagiotopoulos, "When do short-range atomistic machine-learning models fall short?" J. Chem. Phys. **154**, 034111 (2021).

[160] Q. Yu, C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, "q-AQUA: A many-body CCSD(T) water potential including four-body interactions demonstrates the quantum nature of water from clusters to the liquid phase," J. Phys. Chem. Lett. **13**, 5068–5074 (2022).

[161] Y. Zhai, A. Caruso, S. L. Bore, and F. Paesani, "A "short blanket" dilemma for a state-of-the-art neural network potential for water: Reproducing properties or learning the underlying physics?" ChemRxiv, https://doi.org/10.26434/chemrxiv-2022-t92nd-v2 (2022).

[162] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. Van der Avoird, "Predictions of the properties of water from first principles," Science **315**, 1249–1252 (2007).

[163] V. Babin, C. Leforestier, and F. Paesani, "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).

[164] V. Babin, G. R. Medders, and F. Paesani, "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).

[165] G. R. Medders, V. Babin, and F. Paesani, "Development of a "first principles" water potential with flexible monomers. III. Liquid phase properties," J. Chem. Theory Comput. **10**, 2906–2910 (2014).

[166] T. T. Nguyen, E. Székely, G. Imbalzano, J. Behler, G. Csányi, M. Ceriotti, A. W. Götz, and F. Paesani, "Comparison of permutationally invariant polynomials neural networks and gaussian approximation potentials in representing water interactions through many-body expansions," J. Chem. Phys. **148**, 241725 (2018).

[167] P. Bajaj, A. W. Götz, and F. Paesani, "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).

[168] M. Riera, N. Mardirossian, P. Bajaj, A. W. Götz, and F. Paesani, "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).

[169]M. Riera, E. Lambros, T. T. Nguyen, A. W. Götz, and F. Paesani, "Low-order many-body interactions determine the local structure of liquid water," Chem. Sci. **10**, 8211–8218 (2019).

[170]F. Paesani, P. Bajaj, and M. Riera, "Chemical accuracy in modeling halide ion hydration from many-body representations," Adv. Phys. X **4**, 1631212 (2019).

[171]M. Riera, E. P. Yeh, and F. Paesani, "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).

[172]M. Riera, A. Hirales, R. Ghosh, and F. Paesani, "Data-driven many-body models with chemical accuracy for CH_4/H_2O mixtures," J. Phys. Chem. B **124**, 11207–11221 (2020).

[173]L. Zhang, J. Han, H. Wang, R. Car, and W. E, "Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics," Phys. Rev. Lett. **120**, 143001 (2018).

[174]O. Wohlfahrt, C. Dellago, and M. Sega, "Ab initio structure and thermodynamics of the RPBE-D3 water/vapor interface by neural-network molecular dynamics," J. Chem. Phys. **153**, 144710 (2020).

[175]C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek, and A. Michaelides, "Machine learning potentials for complex aqueous systems made simple," Proc. Natl. Acad. Sci. U.S.A. **118**, e2110077118 (2021).

[176]A. Torres, L. S. Pedroza, M. Fernandez-Serra, and A. R. Rocha, "Using neural network force fields to ascertain the quality of ab initio simulations of liquid water," J. Phys. Chem. B **125**, 10772–10778 (2021).

[177]B. J. Braams and J. M. Bowman, "Permutationally invariant potential energy surfaces in high dimensionality," Int. Rev. Phys. Chem. **28**, 577–606 (2009).

[178]Y. Wang and J. M. Bowman, "Towards an ab initio flexible potential for water and post-harmonic quantum vibrational analysis of water clusters," Chem. Phys. Lett. **491**, 1–10 (2010).

[179]G. R. Medders, V. Babin, and F. Paesani, "A critical assessment of two-body and three-body interactions in water," J. Chem. Theory Comput. **9**, 1103–1114 (2013).

[180]K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "Schnet–a deep learning architecture for molecules and materials," J. Chem. Phys. **148**, 241722 (2018).

[181]O. T. Unke and M. Meuwly, "A reactive scalable and transferable model for molecular energies from a neural network approach based on local information," J. Chem. Phys. **148**, 241708 (2018).

[182]O. T. Unke and M. Meuwly, "Physnet: A neural network for predicting energies forces dipole moments and partial charges," J. Chem. Phys. **15**, 3678–3693 (2019).

[183]S. Houlding, S. Liem, and P. Popelier, "A polarizable high-rank quantum topological electrostatic potential developed using neural networks: Molecular dynamics simulations on the hydrogen fluoride dimer," Int. J. Quantum Chem. **107**, 2817–2827 (2007).

[184]C. M. Handley and P. L. Popelier, "Dynamically polarizable water potential based on multipole moments trained by machine learning," J. Chem. Theory Comput. **5**, 1474–1489 (2009).

[185]N. Artrith, T. Morawietz, and J. Behler, "High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide," Phys. Rev. B **83**, 153101 (2011).

[186]D. P. Metcalf, A. Jiang, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, "Electron-passing neural networks for atomic charge prediction in systems with arbitrary molecular charge," J. Chem. Inf. Model **61**, 115–122 (2020).

[187]R. Zubatyuk, J. S. Smith, B. T. Nebgen, S. Tretiak, and O. Isayev, "Teaching a neural network to attach and detach electrons from molecules," Nat. Commun. **12**, 1–11 (2021).

[188]D. Rosenberger, J. S. Smith, and A. E. Garcia, "Modeling of peptides with classical and novel machine learning force fields: A comparison," J. Phys. Chem. B **125**, 3598–3612 (2021).

[189]S. P. Niblett, M. Galib, and D. T. Limmer, "Learning intermolecular forces at liquid–vapor interfaces," J. Chem. Phys. **155**, 164101 (2021).

[190]A. Gao and R. C. Remsing, "Self-consistent determination of long-range electrostatics in neural network potentials," Nat. Commun. **13**, 1–11 (2022).

[191]N. Rezajooei, T. N. T. Phuc, E. Johnson, and C. Rowley, "A neural network potential with rigorous treatment of long-range dispersion," ChemRxiv, https://doi.org/10.26434/chemrxiv-2022-mdz85 (2022).

[192]J. Zhang, J. Pagotto, and T. T. Duignan, "Towards predictive design of electrolyte solutions by accelerating ab initio simulation with neural networks," J. Mater. Chem. A **10**, 19560–19571 (2022).

[193]B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces," J. Chem. Phys. **139**, 054112 (2013).

[194]J. Li, B. Jiang, and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems," J. Chem. Phys. **139**, 204103 (2013).

[195]B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. III. Molecule-surface interactions," J. Chem. Phys. **141**, 034109 (2014).

[196]B. Jiang, J. Li, and H. Guo, "Potential energy surfaces from high fidelity fitting of ab initio points: The permutation invariant polynomial-neural network approach," Int. Rev. Phys. Chem. **35**, 479–506 (2016).

[197]C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, "Breaking the coupled cluster barrier for machine-learned potentials of large molecules: The case of 15-atom acetylacetone," J. Phys. Chem. Lett. **12**, 4902–4909 (2021).

[198]A. Nandi, C. Qu, P. L. Houston, R. Conte, and J. M. Bowman, "δ-machine learning for potential energy surfaces: A PIP approach to bring a DFT-based pes to CCSD(T) level of theory," J. Chem. Phys. **154**, 051102 (2021).

[199]C. Qu, Q. Yu, R. Conte, P. L. Houston, A. Nandi, and J. M. Bomwan, "A δ-machine learning approach for force fields illustrated by a CCSD(T) 4-body correction to the MB-pol water potential," Digital Discovery **1**, 658–664 (2022).

[200]C. Qu and J. M. Bowman, "A fragmented permutationally invariant polynomial approach for potential energy surfaces of large molecules: Application to N-methyl acetamide," J. Chem. Phys. **150**, 141101 (2019).

[201]A. Nandi, C. Qu, and J. M. Bowman, "Full and fragmented permutationally invariant polynomial potential energy surfaces for trans and cis N-methyl acetamide and isomerization saddle points," J. Chem. Phys. **151**, 084306 (2019).

[202]R. Conte, C. Qu, P. L. Houston, and J. M. Bowman, "Efficient generation of permutationally invariant potential energy surfaces for large molecules," J. Chem. Theory Comput. **16**, 3264–3272 (2020).

[203]F. Paesani, "Getting the right answers for the right reasons: Toward predictive molecular simulations of water with many-body potential energy functions," Acc. Chem. Res. **49**, 1844–1851 (2016).

[204]W. Cencek, K. Szalewicz, C. Leforestier, R. van Harrevelt, and A. van der Avoird, "An accurate analytic representation of the water pair potential," Phys. Chem. Chem. Phys. **10**, 4716–4731 (2008).

[205]A. van der Avoird and K. Szalewicz, "Water trimer torsional spectrum from accurate ab initio and semiempirical potentials," J. Chem. Phys. **128**, 014302 (2008).

[206]R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, "Polarizable interaction potential for water from coupled cluster calculations. I. Analysis of dimer potential energy surface," J. Chem. Phys. **128**, 094313 (2008).

[207]R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, "Polarizable interaction potential for water from coupled cluster calculations. II. Applications to dimer spectra virial coefficients and simulations of liquid water," J. Chem. Phys. **128**, 094314 (2008).

[208]R. Hellmann, E. Bich, E. Vogel, A. S. Dickinson, and V. Vesovic, "Calculation of the transport and relaxation properties of dilute water vapor," J. Chem. Phys. **131**, 014303 (2009).

[209]K. Szalewicz, C. Leforestier, and A. van der Avoird, "Towards the complete understanding of water by a first-principles computational approach," Chem. Phys. Lett. **482**, 1–14 (2009).

[210]C. Leforestier, K. Szalewicz, and A. van der Avoird, "Spectra of water dimer from a new ab initio potential with flexible monomers," J. Chem. Phys. **137**, 014305 (2012).

[211]U. Góra, W. Cencek, R. Podeszwa, A. van der Avoird, and K. Szalewicz, "Predictions for water clusters from a first-principles two-and three-body force field," J. Chem. Phys. **140**, 194101 (2014).

[212]E. Palos, E. Lambros, S. Dasgupta, and F. Paesani, "Density functional theory of water with the machine-learned DM21 functional," J. Chem. Phys. **156**, 161103 (2022).

[213]S. K. Reddy, S. C. Straight, P. Bajaj, C. Huy Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, "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).

[214]J. O. Richardson, C. Pérez, S. Lobsiger, A. A. Reid, B. Temelso, G. C. Shields, Z. Kisiel, D. J. Wales, B. H. Pate, and S. C. Althorpe, "Concerted hydrogen-bond breaking by

quantum tunneling in the water hexamer prism," Science **351**, 1310–1313 (2016).

[215]W. T. Cole, J. D. Farrell, D. J. Wales, and R. J. Saykally, "Structure and torsional dynamics of the water octamer from THz laser spectroscopy near 215 μm," Science **352**, 1194–1197 (2016).

[216]S. E. Brown, A. W. Götz, X. Cheng, R. P. Steele, V. A. Mandelshtam, and F. Paesani, "Monitoring water clusters "melt" through vibrational spectroscopy," J. Am. Chem. Soc. **139**, 7082–7088 (2017).

[217]S. K. Reddy, D. R. Moberg, S. C. Straight, and F. Paesani, "Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations with the MB-pol potential energy function," J. Chem. Phys. **147**, 244504 (2017).

[218]T. E. Gartner III, K. M. Hunter, E. Lambros, A. Caruso, M. Riera, G. R. Medders, A. Z. Panagiotopoulos, P. G. Debenedetti, and F. Paesani, "Anomalies and local structure of liquid water from boiling to the supercooled regime as predicted by the many-body MB-pol model," J. Phys. Chem. Lett. **13**, 3652–3658 (2022).

[219]G. R. Medders and F. Paesani, "Dissecting the molecular structure of the air/water interface from quantum simulations of the sum-frequency generation spectrum," J. Am. Chem. Soc. **138**, 3912–3919 (2016).

[220]D. R. Moberg, S. C. Straight, and F. Paesani, "Temperature dependence of the air/water interface revealed by polarization sensitive sum-frequency generation spectroscopy," J. Phys. Chem. B **122**, 4356–4365 (2018).

[221]M. C. Muniz, T. E. Gartner, M. Riera, C. Knight, S. Yue, F. Paesani, and A. Z. Panagiotopoulos, "Vapor–liquid equilibrium of water with the mb-pol many-body potential," J. Chem. Phys. **154**, 211103 (2021).

[222]C. H. Pham, S. K. Reddy, K. Chen, C. Knight, and F. Paesani, "Many-body interactions in ice," J. Chem. Theory Comput. **13**, 1778–1784 (2017).

[223]D. R. Moberg, S. C. Straight, C. Knight, and F. Paesani, "Molecular origin of the vibrational structure of ice I_h," J. Phys. Chem. Lett. **8**, 2579–2583 (2017).

[224]D. R. Moberg, P. J. Sharp, and F. Paesani, "Molecular-level interpretation of vibrational spectra of ordered ice phases," J. Phys. Chem. B **122**, 10572–10581 (2018).

[225]D. R. Moberg, D. Becker, C. W. Dierking, F. Zurheide, B. Bandow, U. Buck, A. Hudait, V. Molinero, F. Paesani, and T. Zeuch, "The end of ice I," Proc. Natl. Acad. Sci. U.S.A. **116**, 24413–24419 (2019).

[226]V. N. Robinson, R. Ghosh, C. K. Egan, M. Riera, C. Knight, F. Paesani, and A. Hassanali, "The behavior of methane–water mixtures under elevated pressures from simulations using many-body potentials," J. Chem. Phys. **156**, 194504 (2022).

[227]S. Yue, M. Riera, R. Ghosh, A. Z. Panagiotopoulos, and F. Paesani, "Transferability of data-driven, many-body models for CO_2 simulations in the vapor and liquid phases," J. Chem. Phys. **156**, 104503 (2022).

[228]P. Bajaj, X.-G. Wang, T. CarringtonJr, and F. Paesani, "Vibrational spectra of halide-water dimers: Insights on ion hydration from full-dimensional quantum calculations on many-body potential energy surfaces," J. Chem. Phys. **148**, 102321 (2017).

[229]P. Bajaj, J. O. Richardson, and F. Paesani, "Ion-mediated hydrogen-bond rearrangement through tunnelling in the iodide–dihydrate complex," Nat. Chem. **11**, 367–374 (2019).

[230]P. Bajaj, D. Zhuang, and F. Paesani, "Specific ion effects on hydrogen-bond rearrangements in the halide–dihydrate complexes," J. Phys. Chem. Lett. **10**, 2823–2828 (2019).

[231]P. Bajaj, M. Riera, J. K. Lin, Y. E. Mendoza Montijo, J. Gazca, and F. Paesani, "Halide ion microhydration: Structure energetics and spectroscopy of small halide–water clusters," J. Phys. Chem. Lett. **123**, 2843–2852 (2019).

[232]D. Zhuang, M. Riera, G. K. Schenter, J. L. Fulton, and F. Paesani, "Many-body effects determine the local hydration structure of Cs^+ in solution," J. Phys. Chem. Lett. **10**, 406–412 (2019).

[233]M. Riera, J. J. Talbot, R. P. Steele, and F. Paesani, "Infrared signatures of isomer selectivity and symmetry breaking in the $Cs^+(H_2O)_3$ complex using many-body potential energy functions," J. Chem. Phys. **153**, 044306 (2020).

[234]A. Caruso and F. Paesani, "Data-driven many-body models enable a quantitative description of chloride hydration from clusters to bulk," J. Chem. Phys. **155**, 064502 (2021).

[235]B. T. Thole, "Molecular polarizabilities calculated with a modified dipole interaction," Chem. Phys. **59**, 341–350 (1981).

[236]E. F. Bull-Vulpe, M. Riera, S. L. Bore, and F. Paesani, "Data-driven many-body potential energy functions for generic molecules: Linear alkanes as a proof-of-concept application," J. Chem. Theory Comput. , https://doi.org/10.1021/acs.jctc.2c00645 (2022).

[237]J.-F. Sadoc and R. Mosseri, *Geometrical Frustration* (Cambridge University Press, 1999).

[238]W. Kohn, "Theory of the insulating state," Phys. Rev. **133**, A171 (1964).

[239]N. Marzari and D. Vanderbilt, "Maximally localized generalized wannier functions for composite energy bands," Phys. Rev. B **56**, 12847 (1997).

[240]A. Marrazzo and R. Resta, "Local theory of the insulating state," Phys. Rev. Lett. **122**, 166602 (2019).

[241]D. Hankins, J. Moskowitz, and F. Stillinger, "Water molecule interactions," J. Chem. Phys. **53**, 4544–4554 (1970).

[242]J. Del Bene and J. Pople, "Intermolecular energies of small water polymers," Chem. Phys. Lett. **4**, 426–428 (1969).

[243]E. Clementi, W. Kołos, G. Lie, and G. Ranghino, "Nonadditivity of interaction in water trimers," Int. J. Quantum Chem. **17**, 377–398 (1980).

[244]K. Kim, M. Dupuis, G. Lie, and E. Clementi, "Revisiting small clusters of water molecules," Chem. Phys. Lett. **131**, 451–456 (1986).

[245]S. S. Xantheas and T. H. Dunning Jr., "The structure of the water trimer from ab initio calculations," J. Chem. Phys. **98**, 8037–8040 (1993).

[246]S. S. Xantheas and T. H. Dunning Jr., "Ab initio studies of cyclic water clusters $(H_2O)_n$, $n = 1 - 6$. I. Optimal structures and vibrational spectra," J. Chem. Phys. **99**, 8774–8792 (1993).

[247]S. S. Xantheas, "Ab initio studies of cyclic water clusters $(H_2O)_n$, $n = 1 - 6$. II. Analysis of many-body interactions," J. Chem. Phys. **100**, 7523–7534 (1994).

[248]S. S. Xantheas, "Ab initio studies of cyclic water clusters $(H_2O)_n$, $n = 1 - 6$. III. Comparison of density functional with MP2 results," J. Chem. Phys. **102**, 4505–4517 (1995).

[249]L. Ojamäe and K. Hermansson, "Ab initio study of cooperativity in water chains: Binding energies and anharmonic frequencies," J. Phys. Chem. **98**, 4271–4282 (1994).

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

[251]V. Babin, G. R. Medders, and F. Paesani, "Toward a universal water model: First principles simulations from the dimer to the liquid phase," J. Phys. Chem. Lett. **3**, 3765–3769 (2012).

[252]H. Partridge and D. W. Schwenke, "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).

[253]K. Tang and J. P. Toennies, "An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients," J. Chem. Phys. **80**, 3726–3741 (1984).

[254]E. R. Johnson and A. D. Becke, "A post-Hartree–Fock model of intermolecular interactions," J. Chem. Phys. **123**, 024101 (2005).

[255]E. R. Johnson and A. D. Becke, "A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections," J. Chem. Phys. **124**, 174104 (2006).

[256]F. Paesani, "Water: Many-body potential from first principles (from the gas to the liquid phase)," Handbook of Materials Modeling: Methods: Theory and Modeling , 635–660 (2020).

[257]E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski,

D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, "Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package," J. Chem. Phys. **155**, 084801 (2021).

[258] D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "PSI4 1.4: Open-source software for high-throughput quantum chemistry," J. Chem. Phys. **152**, 184108 (2020).

[259] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton", "LAMMPS - A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Comput. Phys. Commun. **271**, 108171 (2022).

[260] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. Van Speybroeck, and M. Ceriotti, "i-PI 2.0: A universal force engine for advanced molecular simulations," Comput. Phys. Commun. **236**, 214–223 (2019).

[261] "MBX v0.7," http://paesanigroup.ucsd.edu/software/mbx.html (2021).

[262] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining Inference and Prediction* (Springer Science & Business Media, 2009).

[263] E. Lambros, F. Lipparini, G. A. Cisneros, and F. Paesani, "A many-body fully polarizable approach to QM/MM simulations," J. Chem. Theory Comput. **16**, 7462–7472 (2020).

[264] J. Nochebuena, S. Naseem-Khan, and G. A. Cisneros, "Development and application of quantum mechanics/molecular mechanics methods with advanced polarizable potentials," Wiley Interdiscip. Rev. Comput. Mol. Sci. **11**, e1515 (2021).

[265] "MB-Fit: Software infrastructure for data-driven many-body potential energy functions," https://github.com/paesanilab/MB-Fit (2021).

[266] "MBX: A many-body energy and force calculator, version 0.7," https://github.com/paesanilab/MBX (2021).

[267] P. Mori-Sánchez, A. J. Cohen, and W. Yang, "Localization and delocalization errors in density functional theory and implications for band-gap prediction," Phys. Rev. Lett. **100**, 146401 (2008).

[268] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Fractional spins and static correlation error in density functional theory," J. Chem. Phys. **129**, 121104 (2008).

[269] N. Q. Su, C. Li, and W. Yang, "Describing strong correlation with fractional-spin correction in density functional theory," Proc. Natl. Acad. Sci. U.S.A. **115**, 9678–9683 (2018).

[270] P. M. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, "An investigation of the performance of a hybrid of Hartree-Fock and density functional theory," Int. J. Quantum Chem. **44**, 319–331 (1992).

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

[272] N. Oliphant and R. J. Bartlett, "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).

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

[274] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr., "Density-functional theory for fractional particle number: Derivative discontinuities of the energy," Phys. Rev. Lett. **49**, 1691 (1982).

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

[276]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, "Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes," J. Chem. Phys. **137**, 224113 (2012).

[277]S. Dasgupta, C. Shahi, P. Bhetwal, J. P. Perdew, and F. Paesani, "How good is the density-corrected SCAN functional for neutral and ionic aqueous systems and what is so right about the Hartree-Fock density?" J. Chem. Theory Comput. **18**, 4745–4761 (2022).

[278]K. R. Bryenton, A. A. Adeleke, S. G. Dale, and E. R. Johnson, "Delocalization error: The greatest outstanding challenge in density-functional theory," Wiley Interdiscip. Rev. Comput. Mol. Sci. , e1631 (2022).

[279]A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim, and K. Burke, "The importance of being inconsistent," Annu. Rev. Phys. Chem. **68**, 555–581 (2017).

[280]F. Liu, E. Proynov, J.-G. Yu, T. R. Furlani, and J. Kong, "Comparison of the performance of exact-exchange-based density functional methods," J. Chem. Phys. **137**, 114104 (2012).

[281]T. Tsuneda and K. Hirao, "Self-interaction corrections in density functional theory," J. Chem. Phys. **140**, 18A513 (2014).

[282]J. L. Bao, L. Gagliardi, and D. G. Truhlar, "Self-interaction error in density functional theory: An appraisal," J. Phys. Chem. Lett. **9**, 2353–2358 (2018).

[283]M.-C. Kim, E. Sim, and K. Burke, "Communication: Avoiding unbound anions in density functional calculations," J. Chem. Phys. **134**, 171103 (2011).

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

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

[286]S. Song, M.-C. Kim, E. Sim, A. Benali, O. Heinonen, and K. Burke, "Benchmarks and reliable DFT results for spin gaps of small ligand Fe(II) complexes," J. Chem. Theory Comput. **14**, 2304–2311 (2018).

[287]E. Sim, S. Song, S. Vuckovic, and K. Burke, "Improving results by improving densities: Density-corrected density functional theory," J. Am. Chem. Soc. **144**, 6625–6639 (2022).

[288] Y. Zhang and W. Yang, "Comment on "Generalized gradient approximation made simple"," Phys. Rev. Lett. **80**, 890–890 (1998).

[289] N. Mardirossian and M. Head-Gordon, "ωB97M-V: A combinatorially optimized range-separated hybrid meta-GGA density functional with VV10 nonlocal correlation," J. Chem. Phys. **144**, 214110:1–23 (2016).

[290] D. Chandler and P. G. Wolynes, "Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids," J. Chem. Phys. **74**, 4078–4095 (1981).

[291] B. J. Berne and D. Thirumalai, "On the simulation of quantum systems: Path integral methods," Annu. Rev. Phys. Chem. **37**, 401–424 (1986).

[292] J. R. Errington and P. G. Debenedetti, "Relationship between structural order and the anomalies of liquid water," Nature **409**, 318–321 (2001).

[293] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation Transfer and Reactions in Condensed Molecular Systems* (Oxford university press, 2006).

[294] Y.-J. Zhang, A. Khorshidi, G. Kastlunger, and A. A. Peterson, "The potential for machine learning in hybrid QM/MM calculations," J. Chem. Phys. **148**, 241740 (2018).

[295] A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge model 5: An extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases," J. Chem. Theory Comput. **8**, 527–541 (2012).

[296] Y. Yao and Y. Kanai, "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).

[297] J. Harris and R. Jones, "The surface energy of a bounded electron gas," J. Phys. F **4**, 1170 (1974).

[298] D. C. Langreth and J. P. Perdew, "The exchange-correlation energy of a metallic surface," Solid State Commun. **17**, 1425–1429 (1975).

[299] O. Gunnarsson and B. I. Lundqvist, "Exchange and correlation in atoms molecules and solids by the spin-density-functional formalism," Phys. Rev. B **13**, 4274 (1976).

[300] J. Harris, "Adiabatic-connection approach to kohn-sham theory," Phys. Rev. A **29**, 1648 (1984).

[301] J. Řezáč, P. Jurečka, K. E. Riley, J. Černỳ, H. Valdes, K. Pluháčková, K. Berka, T. Řezáč, M. Pitoňák, J. Vondrášek, and P. Hobza, "Quantum chemical benchmark energy and geometry database for molecular clusters and complex molecular systems (www.begdb.com): A users manual and examples," Collect. Czechoslov. Chem. Commun. **73**, 1261–1270 (2008).

[302] D. Manna, M. K. Kesharwani, N. Sylvetsky, and J. M. Martin, "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).

[303] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, "Spurious fractional charge on dissociated atoms: Pervasive and resilient self-interaction error of common density functionals," J. Chem. Phys. **125**, 194112 (2006).

[304] V. S. Bryantsev, M. S. Diallo, A. C. Van Duin, and W. A. Goddard III, "Evaluation of B3LYP, X3LYP, and M06-class density functionals for predicting the binding energies of neutral protonated and deprotonated water clusters," J. Chem. Theory Comput. **5**, 1016–1026 (2009).

[305] E. W. Lemmon, M. O. McLinden, and D. G. Friend, "Thermophysical properties of fluid systems," in *NIST Chemistry WebBook*, edited by P. Linstrom and W. Mallard (National Institute of Standards and Technology, Gaithersburg (MD), 2022).

[306] J. Witte, N. Mardirossian, J. B. Neaton, and M. Head-Gordon, "Assessing DFT-d3 damping functions across widely used density functionals: Can we do better?" J. Chem. Theory Comput. **13**, 2043–2052 (2017).

[307] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**, 154104 (2010).

[308] B. Temelso, K. A. Archer, and G. C. Shields, "Benchmark structures and binding energies of small water clusters with anharmonicity corrections," J. Phys. Chem. A **115**, 12034–12046 (2011).

[309] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, "Structure and dynamics of liquid water from ab initio molecular dynamics—comparison of BLYP, PBE and revPBE density functionals with and without van der Waals corrections," J. Chem. Theory Comput. **8**, 3902–3910 (2012).

[310]A. Bankura, A. Karmakar, V. Carnevale, A. Chandra, and M. L. Klein, "Structure dynamics and spectral diffusion of water from first-principles molecular dynamics," J. Phys. Chem. C **118**, 29401–29411 (2014).

[311]M. Galib, T. T. Duignan, Y. Misteli, M. D. Baer, G. K. Schenter, J. Hutter, and C. J. Mundy, "Mass density fluctuations in quantum and classical descriptions of liquid water," J. Chem. Phys. **146**, 244501 (2017).

[312]T. Ohto, M. Dodia, S. Imoto, and Y. Nagata, "Structure and dynamics of water at the water–air interface using first-principles molecular dynamics simulations within generalized gradient approximation," J. Chem. Theory Comput. **15**, 595–602 (2019).

[313]M. Dodia, T. Ohto, S. Imoto, and Y. Nagata, "Structure and dynamics of water at the water–air interface using first-principles molecular dynamics simulations. ii. nonlocal vs empirical van der Waals corrections," J. Chem. Theory Comput. **15**, 3836–3843 (2019).

[314]J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, "Finding density functionals with machine learning," Phys. Rev. Lett. **108**, 253002 (2012).

[315]L. Li, T. E. Baker, S. R. White, and K. Burke, "Pure density functional for strong correlation and the thermodynamic limit from machine learning," Phys. Rev. B **94**, 245129 (2016).

[316]O. A. von Lilienfeld and K. Burke, "Retrospective on a decade of machine learning for chemical discovery," Nat. Commun. **11**, 1–4 (2020).

[317]H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. L. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. T. Schütt, J. Westermayr, M. Gastegger, R. J. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. V. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger, and L. M. Ghiringhelli, "Roadmap on machine learning in electronic structure," Electronic Structure **4**, 023004 (2022).

[318]S. Dick and M. Fernandez-Serra, "Highly accurate and constrained density functional obtained with differentiable programming," Phys. Rev. B **104**, L161109 (2021).

[319]T. Gould and S. Dale, "Poisoning density functional theory with benchmark sets of difficult systems," Phys. Chem. Chem. Phys. **24**, 6398–6403 (2022).

[320] I. S. Gerasimov, T. V. Losev, E. Y. Epifanov, I. Rudenko, I. S. Bushmarinov, A. A. Ryabov, P. A. Zhilyaev, and M. G. Medvedev, "Comment on "Pushing the frontiers of density functionals by solving the fractional electron problem"," Science **377**, eabq3385 (2022).

[321] D. J. Tozer and N. C. Handy, "The importance of the asymptotic exchange-correlation potential in density functional theory," Mol. Phys. **101**, 2669–2675 (2003).

[322] C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. I: $OH^{-}(H_2O)_n$ clusters," J. Chem. Theory Comput. **14**, 1982–1997 (2018).

[323] C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. II: $H_3O^{+}(H_2O)_n$ clusters," J. Chem. Theory Comput. **15**, 4816–4833 (2019).

[324] F. Lipparini and B. Mennucci, "Hybrid QM/classical models: Methodological advances and new applications," Chem. Phys. Rev. **2**, 041303 (2021).

[325] A. Warshel and M. Karplus, "Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization," J. Am. Chem. Soc. **94**, 5612–5625 (1972).

[326] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol. **103**, 227–249 (1976).

[327] A. Warshel, M. Kato, and A. V. Pisliakov, "Polarizable force fields: History test cases and prospects," J. Chem. Theory Comput. **3**, 2034–2045 (2007).

[328] R. B. Murphy, D. M. Philipp, and R. A. Friesner, "A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments," J. Comp. Chem. **21**, 1442–1457 (2000).

[329] H. Hu and W. Yang, "Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods," Annu. Rev. Phys. Chem. **59**, 573–601 (2008).

[330] O. Acevedo and W. L. Jorgensen, "Understanding rate accelerations for Diels–Alder reactions in solution using enhanced QM/MM methodology," J. Chem. Theory Comput. **3**, 1412–1419 (2007).

[331] O. Acevedo and W. L. Jorgensen, "Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions," Acc. Chem. Res. **43**, 142–151 (2010).

[332]M. A. Marques, X. López, D. Varsano, A. Castro, and A. Rubio, "Time-dependent density-functional approach for biological chromophores: The case of the green fluorescent protein," Phys. Rev. Lett. **90**, 258101 (2003).

[333]J. M. Olsen, K. Aidas, and J. Kongsted, "Excited states in solution through polarizable embedding," J. Chem. Theory Comput. **6**, 3721–3734 (2010).

[334]C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker, and T. J. Martínez, "Electronic absorption spectra from mm and ab initio QM/MM molecular dynamics: Environmental effects on the absorption spectrum of photoactive yellow protein," J. Chem. Theory Comput. **8**, 5092–5106 (2012).

[335]U. N. Morzan, D. J. Alonso de Armino, N. O. Foglia, F. Ramirez, M. C. Gonzalez Lebrero, D. A. Scherlis, and D. A. Estrin, "Spectroscopy in complex environments from QM–MM simulations," Chem. Rev. **118**, 4071–4113 (2018).

[336]M. Bondanza, L. Cupellini, F. Lipparini, and B. Mennucci, "The multiple roles of the protein in the photoactivation of orange carotenoid protein," Chem **6**, 187–203 (2020).

[337]M. Nottoli, B. Mennucci, and F. Lipparini, "Excited state born–oppenheimer molecular dynamics through coupling between time dependent DFT and AMOEBA," Phys. Chem. Chem. Phys. **22**, 19532–19541 (2020).

[338]M. Nottoli, L. Cupellini, F. Lipparini, G. Granucci, and B. Mennucci, "Multiscale models for light-driven processes," Annu. Rev. Phys. Chem. **72**, 489–513 (2021).

[339]D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci, and J.-P. Piquemal, "Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings," Chem. Sci. **10**, 7200–7211 (2019).

[340]J. Dziedzic, Y. Mao, Y. Shao, J. Ponder, T. Head-Gordon, M. Head-Gordon, and C.-K. Skylaris, "TINKTEP: A fully self-consistent mutually polarizable QM/MM approach based on the AMOEBA force field," J. Chem. Phys. **145**, 124106 (2016).

[341]J. Dziedzic, T. Head-Gordon, M. Head-Gordon, and C.-K. Skylaris, "Mutually polarizable QM/MM model with in situ optimized localized basis functions," J. Chem. Phys. **150**, 074103 (2019).

[342]A. Albaugh, H. A. Boateng, R. T. Bradshaw, O. N. Demerdash, J. Dziedzic, Y. Mao, D. T. Margul, J. Swails, Q. Zeng, D. A. Case, P. Eastman, L.-P. Wang, J. W. Essex, M. Head-Gordon, V. S. Pande, J. W. Ponder, Y. Shao, C.-K. Skylaris, I. T. Todorov, M. E. Tuckerman, and T. Head-Gordon, "Advanced potential energy surfaces for molecular

simulation," J. Phys. Chem. B **120**, 9811–9832 (2016).

[343]A. Heyden, H. Lin, and D. G. Truhlar, "Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations," J. Phys. Chem. B. **111**, 2231–2241 (2007).

[344]S. Pezeshki and H. Lin, "Adaptive-partitioning redistributed charge and dipole schemes for QM/MM dynamics simulations: On-the-fly relocation of boundaries that pass through covalent bonds," J. Chem. Theory Comput. **7**, 3625–3634 (2011).

[345]R. E. Bulo, B. Ensing, J. Sikkema, and L. Visscher, "Toward a practical method for adaptive QM/MM simulations," J. Chem. Theory Comput. **5**, 2212–2221 (2009).

[346]J. M. Boereboom, R. Potestio, D. Donadio, and R. E. Bulo, "Toward hamiltonian adaptive QM/MM: Accurate solvent structures using many-body potentials," J. Chem. Theory Comput. **12**, 3441–3448 (2016).

[347]J. M. Boereboom, P. Fleurat-Lessard, and R. E. Bulo, "Explicit solvation matters: Performance of QM/MM solvation models in nucleophilic addition," J. Chem. Theory Comput. **14**, 1841–1852 (2018).

[348]D. Loco, E. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal, and B. Mennucci, "A QM/MM approach using the AMOEBA polarizable embedding: From ground state energies to electronic excitations," J. Chem. Theory Comput. **12**, 3654–3661 (2016).

[349]D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci, and J.-P. Piquemal, "Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding," J. Chem. Theory Comput. **13**, 4025–4033 (2017).

[350]E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, and G. Andrés Cisneros, "LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields," J. Comput. Chem. **37**, 1019–1029 (2016).

[351]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A.

Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian˜16 revision c.01," (2016), gaussian Inc. Wallingford CT.

[352]H. Gökcan, E. A. Vázquez-Montelongo, and G. A. Cisneros, "Lichem 1.1: Recent improvements and new capabilities," J. Chem. Theory Comput. **15**, 3056–3065 (2019).