MB-pol(2023): Sub-Chemical Accuracy for Water Simulations from the Gas to the Liquid Phase

Xuanyu Zhu,[*,†] Marc Riera,[†] Ethan F. Bull-Vulpe,[†] and Francesco Paesani[*,†,‡,¶,§]

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

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

[¶]*Halicioğlu Data Science Institute, University of California San Diego,*
La Jolla, California 92093, United States

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

E-mail: xuz144@ucsd.edu; fpaesani@ucsd.edu

Abstract

We use the MB-pol theoretical/computational framework to introduce a new family of data-driven many-body potential energy functions (PEFs) for water, named MB-pol(2023). By employing larger 2-body and 3-body training sets, including an explicit machine-learned representation of 4-body energies, and adopting more sophisticated machine-learned representations of 2-body and 3-body energies, we demonstrate that the MB-pol(2023) PEFs achieve sub-chemical accuracy in modeling the energetics of the hexamer isomers, outperforming both the original MB-pol and q-AQUA PEFs, which currently provide the most accurate description of water clusters in the gas phase. Importantly, the MB-pol(2023) PEFs provide remarkable agreement with the experimental results for various properties of liquid water, improving upon the original MB-pol PEF and effectively closing the gap with experimental measurements.

INTRODUCTION

A realistic representation of the properties of water across different phases has been a long-standing goal in computational molecular sciences since the first Monte Carlo (MC)[1] and molecular dynamics (MD)[2] simulations carried out more than 50 years ago. It is thus not surprising that many computer models of water have been reported in the literature over the past five decades.[3–6] The most common models maintain rigid water molecules and describe the underlying interactions in a pairwise additive manner, attempting to capture many-body electrostatic interactions through a sum of effective pairwise Coulomb interactions between atomic point charges. These models are generally parameterized to reproduce a subset of experimental properties (e.g., density, freezing point, enthalpy of vaporization, etc.).[4] Examples of empirical pairwise-additive water models include RWK,[7] SPC,[8] SPC/E,[9] TIP4P,[10] TIP4P-Ew,[11] TIP4P/2005,[12] TIP4P/Ice,[13] and TIP5P.[14] The reader is referred to refs 3 and 4 for a more detailed assessment of the performance of pairwise-additive water models with rigid monomers. Empirical pairwise-additive models with flexible monomers (e.g., TIP4P/2005f,[15] q-TIP4P/F,[16] SPC/Fw,[17] and q-SPC/Fw[18]) were also developed to investigate vibrational dynamics and nuclear quantum effects in liquid water. Although empirical pairwise-additive models have been the workhorse of computer simulations of water, aqueous solutions, and hydration phenomena, they suffer from intrinsic limitations that prevent them from being fully transferable across different phases.[6]

In an attempt to overcome these limitations, several water models have been developed with the goal of accounting for many-body effects. Examples of these water models are WAIL,[19] which was developed using adaptive force matching, and E3B,[20–22] which includes an empirical parameterization of 3-body interactions, as well as several polarizable models, such as BK3,[23] SWM4-DP,[24] SWM4-NDP,[25] SWM6,[26] COS,[27,28] TTMx-F,[29–34] AMOEBA,[35–41] GEM*,[42,43] POLIR,[44] POLI2VS,[45] MB-UCB,[46] and HIPPO,[47] which implicitly represent many-body effects through a classical polarization. Although they correctly reproduce the interactions between water molecules for minimum-energy hydrogen-bonding arrangements, polarizable models become less accurate in representing many-body effects for distorted configurations, which prevents them from achieving chemical accuracy (1 kcal/mol)[48] in the representation of the (free-)energy landscape of water across different thermodynamic state points.[49]

Another class of models aims to explicitly capture many-body effects by reproducing each term in the many-body expansion (MBE),[50]

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

where each 1-body energy, $\varepsilon^{1\mathrm{B}}(i)$, is the energy of the isolated ith water molecule, $E_1(i)$. For $n \geq 2$, the n-body (nB) energies are defined recursively by a rearrangement of eq 1:

$$\begin{aligned} \varepsilon^{n\mathrm{B}}(1,\ldots,n) = E_n(1,\ldots,n) - \sum_{i=1}^{N} \varepsilon^{1\mathrm{B}}(i) - \sum_{i<j}^{N} \varepsilon^{2\mathrm{B}}(i,j) \\ - \sum_{i<j<k}^{N} \varepsilon^{3\mathrm{B}}(i,j,k) - \ldots - \varepsilon^{(n-1)\mathrm{B}}(1,\ldots,n-1) \end{aligned} \tag{2}$$

Since the MBE converges quickly for nonmetallic systems,[50–55] eq 1 provides a rigorous and efficient theoretical/computational framework for the development of potential energy functions

(PEFs) for water that reproduce the fully-dimensional potential energy surface by explicitly representing each n-body contribution to the interaction energy.[56]

The first many-body PEFs for water derived from eq 1 were developed by Clementi and coworkers, assuming rigid molecules and explicitly including 2-body and 3-body terms, along with a classical description of many-body polarization.[57–60] Building upon the pioneering work by Clementi and coworkers, several many-body PEFs have been reported in the literature, including CC-pol,[61–63] WHBB,[64–67] HBB2-pol,[68,69] and MB-pol,[70–72] which represent all individual n-body terms of eq 1. More recently, the q-AQUA PEF was introduced which, contrary to CC-pol, WHBB, HBB2-pol, and MB-pol, neglects all n-body terms with $n > 4$.[73] Among these many-body PEFs, MB-pol integrates physics-based many-body representations with data-driven machine-learning (ML) many-body permutationally invariant polynomials (PIPs)[74] trained on reference data calculated at the coupled cluster level of theory, including single, double, and perturbative triple excitations [CCSD(T)], which is the current “gold standard” for chemical accuracy.[75] MB-pol was shown to accurately predict the properties of water from the gas to the condensed phases.[76] In particular, MB-pol quantitatively reproduces the vibration–rotation tunneling spectrum of the water dimer,[70] the energetics, quantum isomeric equilibria, tunneling splittings, and vibrational spectra of small water clusters;[77–88] the structural, thermodynamic, and dynamical properties[89,90] as well as the infrared, Raman, and X-ray spectra of liquid water;[91–97] the sum-frequency generation spectra of the air/water interface;[98–101] the vapor-liquid equilibrium properties;[102] and the energetics as well as the infrared and Raman spectra of various ice phases.[103–106] It has recently been demonstrated that MB-pol is, to date, the only water model that correctly reproduce the phase diagram of water over a wide range of temperature and pressure conditions.[107] Given its accuracy, MB-pol was used as a reference in the development of an optimized exchange–correlation density functional for water.[108]

It should be noted that MB-pol also represented the first step towards the development of the many-body energy (MB-nrg) theoretical/computational framework,[109] which exploits the “near-sightedness” of electronic matter[110] to rigorously represent the energy of a given molecular system in terms of individual many-body contributions.[50] In this context, MB-pol was used to represent water in MB-nrg PEFs of various aqueous systems, including halide and alkali-metal ions in water,,[111,112] CH_4/H_2O mixtures,[113,114] and CO_2/H_2O mixtures.[115,116] In particular, MB-nrg PEFs developed for alkali-metal and halide ions were shown to accurately predict the structures, binding and interaction energies, and vibrational spectra of small $X^-(H_2O)_N$ (with X = F, Cl, Br, and I)[117–120] and $M^+(H_2O)_N$ (with M = Li, Na, K, Rb, and Cs) clusters as well as the hydration structures of Cl^-, Br^-, I^-, Na^+, K^+, and Cs^+ in solution.[121–124]

The MB-pol theoretical/computational framework is fully transferable and highly flexible, which implies that, as discussed in ref 76, MB-pol can be systematically improved by 1) training the machine-learned PIPs on larger datasets, 2) including machine-learned PIPs explicitly representing higher n-body interactions, and 3) adopting higher-order PIPs with more terms to represent the n-body energies. Since larger training sets for 2-body and 3-body energies along with a new training set of 4-body energies have recently become available through the q-AQUA model[73] and considering that 10 years have passed since the first release of MB-pol,[70–72] in this study we introduce a new family of MB-pol PEFs, hereafter referred to as MB-pol(2023), developed by implementing all three points listed above and assess their impact on the overall performance of each of the MB-pol(2023) PEFs relative to the original MB-pol PEF.

THEORY AND METHODS

Models

Building upon the original MB-pol PEF,[70–72] MB-pol(2023) approximates the MBE (eq 1) as

$$E_N(1,\ldots,N) = \sum_{i=1}^{N} V^{1B}(i) + \sum_{i<j}^{N} V^{2B}(i,j) + \sum_{i<j<k}^{N} V^{3B}(i,j,k) + \sum_{i<j<k<l}^{N} V^{4B}(i,j,k,l) + V_{\mathrm{pol}}(1,\ldots,N) \tag{3}$$

where V^{1B}, V^{2B}, V^{3B}, and V^{4B} are explicit 1-body, 2-body, 3-body, and 4-body terms fitted to reproduce the corresponding reference 1-body, 2-body, 3-body, and 4-body energies, and V_{pol} implicitly represents the polarization contribution at all many-body levels. By convention, $E_N(1,\ldots,N) = 0$ corresponds to the configuration in which all N water molecules are separated by infinite distances and each molecule is in its minimum-energy geometry.

Each of the $V^{2\mathrm{B}}$, $V^{3\mathrm{B}}$, and $V^{4\mathrm{B}}$ terms contains a corresponding 2-body, 3-body, and 4-body machine-learned term ($V_{\mathrm{ML}}^{n\mathrm{B}}$) which is a product of a switching function and a PIP (i.e., $V_{\mathrm{ML}}^{n\mathrm{B}} = s_n V_{\mathrm{PIP}}^{n\mathrm{B}}$). As in the original MB-pol PEF,[70,71] the PIPs adopted by MB-pol(2023) take the following form:

$$V_{\mathrm{PIP}}^{\mathrm{nB}}(\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_\mathrm{n}) = \sum_{l=1}^{L} c_l \cdot \eta_l(\xi_1,\xi_2,\ldots,\xi_N) \tag{4}$$

Here, $\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_\mathrm{n}$ are n water molecules, L is the number of linear parameters, c_l are the linear parameters, η_l are the symmetrized monomials built from the variables, ξ_{1-N}, each of which is an exponential of an interatomic distance with one of the following forms:

$$\xi^{\mathrm{exp}}(R_{mn}) = e^{-k_{\tau(mn)}R_{mn}} \tag{5a}$$

$$\xi^{\mathrm{exp0}}(R_{mn}) = e^{-k_{\tau(mn)}(R_{mn}-d_{0,\tau(mn)})} \tag{5b}$$

$$\xi^{\mathrm{coul}}(R_{mn}) = e^{-k_{\tau(mn)}R_{mn}}/R_{mn} \tag{5c}$$

$$\xi^{\mathrm{coul0}}(R_{mn}) = e^{-k_{\tau(mn)}(R_{mn}-d_{0,\tau(mn)})}/R_{mn} \tag{5d}$$

where m, n are the indices for the atoms (O and H) or the lone-pair sites (L_1 and L_2) defined in Fig. 1, and R_{mn} is the distance between two atoms/sites. $\tau(mn)$ maps the pair of atoms/sites into distinct classes, such that all atom/site pairs within the same class share the same nonlinear fitting parameters $k_{\tau(mn)}$ and $d_{0,\tau(mn)}$. The switching functions (s_n) ensure that the contribution from the $V_{\mathrm{ML}}^{n\mathrm{B}}$ terms goes to zero as the monomers in the dimer, trimer, or tetramer are separated. Specific details about the functional form of the 2-body, 3-body, and 4-body PIPs are discussed in the Supporting Information.

1-body term. As in the original MB-pol PEF,[70] the 1-body term of the MB-pol(2023) PEFs is represented by the analytical potential developed by Partridge and Schwenke,[125] which was derived from high-level electronic structure calculations and further refined to reproduce the rovi-

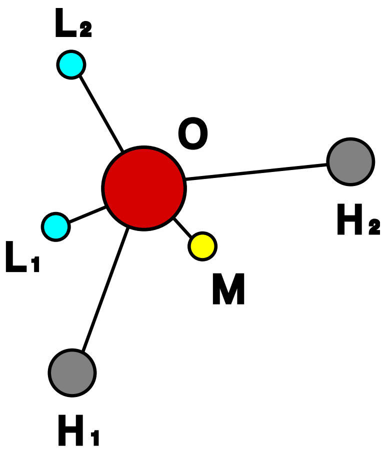


Figure 1: Schematic representation of a MB-pol water molecule, showing the location of the fictitious site (M) and lone-pair sites (L_1 and L_2) relative to the O and H atoms. Note that the L_1 and L_2 sites occupy opposite positions on a plane perpendicular to the molecular plane containing the O and H atoms (i.e., they point below and above the molecular plane, respectively.)

brational transitions of an isolated water molecule in the gas phase. Each water molecule is defined by 6 sites (Fig. 1): three physical atoms (O and H), one fictitious site (M) along the bisector of the HOH angle whose position is defined as

$$\mathbf{r}_\mathrm{M} = \gamma_1 \mathbf{r}_\mathrm{O} + \gamma_2 (\mathbf{r}_{\mathrm{H}_1} + \mathbf{r}_{\mathrm{H}_2}) \tag{6}$$

with $\gamma_\mathrm{M} = 0.426706882$, $\gamma_1 = 1 - \gamma_\mathrm{M}$, and $\gamma_2 = 0.5 * \gamma_\mathrm{M}$, and two lone-pairs ($L_1$ and L_2) whose positions along the oxygen–lone-pair directions are defined as

$$\mathbf{r}_{\mathrm{L}_1,\mathrm{L}_2} = \mathbf{r}_\mathrm{O} + \frac{1}{2}\gamma_\parallel (\mathbf{r}_{\mathrm{OH}_1} + \mathbf{r}_{\mathrm{OH}_2}) \pm \gamma_\perp (\mathbf{r}_{\mathrm{OH}_1} \times \mathbf{r}_{\mathrm{OH}_2}) \tag{7}$$

Here, $\mathbf{r}_\mathrm{O}$ is the position of the oxygen atom, $\mathbf{r}_{\mathrm{OH}_1}$ and $\mathbf{r}_{\mathrm{OH}_2}$ are the two OH bonds, and $\gamma_\parallel = -9.721486914088159 \times 10^{-2}$ and $\gamma_\perp = 9.859272078406150 \times 10^{-2}$ Å $^{-1}$ as in the original MB-pol PEF.[70]

Geometry-dependent partial charges derived from the Partridge-Schwenke dipole moment surface[125] are located on the M site and two H atoms, while atomic polarizabilities are located on the three physical atoms (O and two H atoms).

2-body term. The explicit 2-body term of the MB-pol(2023) PEFs is expressed as

$$V^{\text{2B}} = V^{\text{2B}}_{\text{ML}} + V^{\text{2B}}_{\text{perm}} + V^{\text{2B}}_{\text{disp}} \tag{8}$$

Here, $V^{\text{2B}}_{\text{perm}}$ represents permanent electrostatics as Coulomb interactions between the geometry-dependent partial charges on the M sites and H atoms. $V^{\text{2B}}_{\text{disp}}$ represents the 2-body dispersion energy as a sum of pairwise additive contributions,

$$V^{\text{2B}}_{\text{disp}} = \sum_{\substack{i \in M_1 \\ j \in M_2}} -f(\delta_{ij} R_{ij}) \frac{C_{6,ij}}{R^6_{ij}} \tag{9}$$

where R_{ij} is the distance between atoms i and j respectively on monomers M_1 and M_2, $\text{C}_{6,ij}$ is the corresponding dispersion coefficient, and $f(\delta_{ij} R_{ij})$ is the Tang-Toennies damping function,[126]

$$f(\delta_{ij}, R_{ij}) = 1 - \exp(-\delta_{ij} R_{ij}) \sum_{n=0}^{6} \frac{(\delta_{ij} R_{ij})^n}{n!} \tag{10}$$

All parameters (i.e., $\text{C}_{6,ij}$ and δ_{ij}) entering the expression of $V^{\text{2B}}_{\text{disp}}$ in eq 9 are taken from the original MB-pol PEF.[70,71]

The first term in eq 8, $V^{\text{2B}}_{\text{ML}}$, is a machine-learned PIP (as in eq 4) built from the variables (ξ) listed in eqs. 5a-5c, which are functions of the distances between the atoms (O and H) and lone-pair sites (L_1 and L_2) of the two water molecules within a dimer. It was demonstrated that $V^{\text{2B}}_{\text{ML}}$ in the original MB-pol PEF effectively accounts for short-range quantum-mechanical interactions (e.g., exchange-repulsion, charge transfer, and charge penetration) that arise when the electron densities of two water molecules overlap.[127] $V^{\text{2B}}_{\text{ML}}$ smoothly switches to zero as the distance between the two molecules (M_1 and M_2) becomes larger than a predefined cutoff value,

$$V^{\text{2B}}_{\text{ML}} = s_2 \left(\frac{R_{\text{OO}} - R_{\text{in}}}{R_{\text{out}} - R_{\text{in}}} \right) V^{\text{2B}}_{\text{PIP}}(\text{M}_1, \text{M}_2) \tag{11}$$

where R_{OO} is the distance between the oxygen atoms of the two molecules (M_1 and M_2). In eq 11,

$s_2(t)$ is the same switching function adopted by MB-pol,[70,71] which is defined as

$$s_2(t) = \begin{cases} 1 & \text{if } t < 0 \\ \frac{1}{2}\left[1 + \cos(\pi \mathrm{t})\right] & \text{if } 0 \leq t < 1 \\ 0 & \text{if } 1 \leq t \end{cases} \tag{12}$$

By construction, $s_2(t) = 1$ when $R_{\mathrm{OO}} \leq R_{\mathrm{in}}$ and $s_2(t) = 0$ when $R_{\mathrm{OO}} \geq R_{\mathrm{out}}$. Therefore, the values of the inner (R_{in}) and outer (R_{out}) cutoffs in eq 11 define the region over which $V_{\mathrm{ML}}^{\mathrm{2B}}$ is continuously switched off. The inner and outer cutoffs are kept as in the original MB-pol PEF,[70] with $R_{\mathrm{in}} = 4.5$ Å and $R_{\mathrm{out}} = 6.5$ Å.

By systematically increasing the highest degree of the polynomial, we developed three distinct versions of $V_{\mathrm{PIP}}^{\mathrm{2B}}(\mathrm{M}_1, \mathrm{M}_2)$, hereafter referred to as 2a, 2b, and 2c, whose highest degree is equal to 4, 5, and 6, respectively. All three versions of $V_{\mathrm{PIP}}^{\mathrm{2B}}(\mathrm{M}_1, \mathrm{M}_2)$ are functions of 31 distances between the O and H atoms and the L_1 and L_2 sites. Model 2a adopts the same functional form as the 2-body PIP in the original MB-pol PEF,[70] including 6 1st-degree, 63 2nd-degree, 491 3rd-degree, and 593 4th-degree symmetrized monomials, resulting in a total of 1153 linear parameters. Model 2b includes 6 1st-degree, 63 2nd-degree, 491 3rd-degree, 593 4th-degree, and 1022 5th-degree symmetrized monomials, resulting in a total of 2175 linear parameters. Model 2c includes 6 1st-degree, 63 2nd-degree, 491 3rd-degree, 593 4th-degree, 1022 5th-degree, and 1653 6th-degree symmetrized monomials, resulting in a total of 3828 linear parameters. Similar to the original MB-pol PEF,[70] all polynomial terms approach zero when the distance between the two water molecules increases. All three models also include 8 nonlinear fitting parameters. Specific details about the functional form of $V_{\mathrm{PIP}}^{\mathrm{2B}}(\mathrm{M}_1, \mathrm{M}_2)$ are discussed in the Supporting Information.

3-body term. As in the original MB-pol PEF,[71] the explicit 3-body term of the MB-pol(2023) PEFs is represented by a short-range term:

$$V^{\mathrm{3B}} = V_{\mathrm{ML}}^{\mathrm{3B}} \tag{13}$$

Similar to the 2-body term, $V_{\mathrm{ML}}^{3\mathrm{B}}$ is represented by a machine-learned PIP[74] built from the variables (ξ) listed in eqs. 5a-5c, which are functions of the distances between the atoms and lone pair sites of the three water molecules within a trimer,

$$V_{\mathrm{ML}}^{3\mathrm{B}} = [s_2(t_{12})s_2(t_{13}) + s_2(t_{12})s_2(t_{23}) + s_2(t_{13})s_2(t_{23})]V_{\mathrm{PIP}}^{3\mathrm{B}}(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3) \tag{14}$$

Here, the sum of the three terms in the square bracket represents a compound switching function that smoothly goes to zero as any of the molecules moves apart from the other two. Specifically, $s_2(t_{mn})$ is defined in eq 12 and t_{mn} is given by

$$t_{mn} = \frac{R_{mn}}{R_{\mathrm{cut}}} \tag{15}$$

where R_{mn} is the distance between the oxygen atoms of monomers m and n, and R_{cut} is a 3-body cutoff chosen to disable the 3-body short-range term at distances where its contribution is negligible. As in the original MB-pol PEF,[71] $R_{\mathrm{cut}} = 4.5$ Å.

We developed three distinct versions of $V_{\mathrm{PIP}}^{3\mathrm{B}}(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3)$, hereafter referred to as 3a, 3b, and 3c, keeping the highest degree of the polynomial equal to 4 but systematically increasing the number of polynomial terms.

Model 3a adopts the same form as the 3-body PIP in the original MB-pol PEF,[71] which implies that it is a function of 36 distances between the O and H atoms of the three water molecules within a trimer and includes 13 2$^{\mathrm{nd}}$-degree, 202 3$^{\mathrm{rd}}$-degree, and 948 4$^{\mathrm{th}}$-degree symmetrized monomials, resulting in a total of 1163 linear parameters. Model 3a also include 10 nonlinear fitting parameters. Models 3b and 3c are instead functions of 84 distances between the O and H atoms as well as the L_1 and L_2 sites of the three water molecules within a trimer. Model 3b includes 57 2$^{\mathrm{nd}}$-degree, 148 3$^{\mathrm{rd}}$-degree, and 1599 4$^{\mathrm{th}}$-degree symmetrized monomials, resulting in a total of 1804 linear parameters. Model 3c includes 14 2$^{\mathrm{nd}}$-degree, 182 3$^{\mathrm{rd}}$-degree, and 2011 4$^{\mathrm{th}}$-degree symmetrized monomials, resulting in a total of 2207 linear parameters. Model 3b and 3c also include 13 nonlinear fitting parameters. Similar to the original MB-pol PEF,[71] all polynomial terms approach zero when at least one of the water molecules moves away from the others. Specific details about the functional form of $V^{3B}_{PIP}(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3)$ are discussed in the Supporting Information.

4-body term. The explicit 4-body term of the MB-pol(2023) PEFs is represented by a short-range term:

$$V^{4B} = V^{4B}_{ML} \tag{16}$$

Similar to the 2-body and 3-body terms, V^{4B}_{ML} is represented by a machine-learned PIP[74] built from the variables (ξ) listed in eqs. 5a-5c, which are functions of the distances between all O and H atoms of the four water molecules within a tetramer,

$$V^{4B}_{ML} = s_4(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4)V^{4B}_{PIP}(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4) \tag{17}$$

The 4-body switching function is defined as:

$$s_4(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4) = s_2(t_{12})s_2(t_{13})s_2(t_{14})s_2(t_{23})s_2(t_{24})s_2(t_{34}) \tag{18}$$

where

$$t_{mn} = \left(\frac{R_{mn} - R_{\mathrm{in}}}{R_{\mathrm{out}} - R_{\mathrm{in}}}\right) \tag{19}$$

s_4 smoothly varies from one to zero as the largest O-O distance within the tetramer goes from R_{in} to R_{out}.

To be consistent with the notation used for the 2-body and 3-body PIP models discussed above, we define model 4a as $V^{4B}_{ML} = 0$, because the original MB-pol PEF did not include a V^{4B}_{ML} term, instead letting the many-body polarization term (V_{pol}) model all 4-body interactions.[70,71] We also developed two (non-zero) versions of V^{4B}_{ML}, hereafter referred to as 4b and 4c, by systematically increasing the cutoff distance of the switching function $s_4(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4)$ in eq 18. Since the available 4-body training set is relatively small,[73] we only developed one version of $V^{4B}_{PIP}(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4)$ which is adopted by both models 4b and 4c, which implies that the two versions of V^{4B}_{ML} only differs due to the different range of action of the switching functions. The

4-body PIP is a function of 66 distances between all O and H atoms within a tetramer. Consistently with the 2-body and 3-body PIPs, the selection of polynomial terms ensures that all accepted terms approach zero as any subset of the monomers in the tetramer move apart from the others. Model 4b adopts $(R_{\mathrm{in}}, R_{\mathrm{out}}) = (4.5, 5.5)$ and model 4c adopts $(R_{\mathrm{in}}, R_{\mathrm{out}}) = (5.5, 6.5)$. Specific details about the functional form of $V^{\mathrm{4B}}_{\mathrm{PIP}}(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3, \mathrm{M}_4)$ are discussed in the Supporting Information.

Many-body polarization. Finally, V_{pol} in eq 3 describes the induction energy and is represented by a classical many-body polarization term built upon a modified version of the Thole-type model[128] originally introduced in ref 34. This implies that the MB-pol(2023) PEFs represent the n-body energies, with $n \geq 2$ as follows:

$$\varepsilon^{\mathrm{2B}}_{\mathrm{MB-pol}} = V^{\mathrm{2B}} + V^{\mathrm{2B}}_{\mathrm{pol}} \tag{20a}$$

$$\varepsilon^{\mathrm{3B}}_{\mathrm{MB-pol}} = V^{\mathrm{3B}} + V^{\mathrm{3B}}_{\mathrm{pol}} \tag{20b}$$

$$\varepsilon^{\mathrm{4B}}_{\mathrm{MB-pol}} = V^{\mathrm{4B}} + V^{\mathrm{4B}}_{\mathrm{pol}} \tag{20c}$$

$$\varepsilon^{>\mathrm{4B}}_{\mathrm{MB-pol}} = V^{>\mathrm{4B}}_{\mathrm{pol}} \tag{20d}$$

with

$$V_{\mathrm{pol}} = V^{\mathrm{2B}}_{\mathrm{pol}} + V^{\mathrm{3B}}_{\mathrm{pol}} + V^{\mathrm{4B}}_{\mathrm{pol}} + V^{>\mathrm{4B}}_{\mathrm{pol}} \tag{21}$$

As the original MB-pol PEF,[70–72] all MB-pol(2023) PEFs, by construction, thus fully represent both short- and long-range many-body interactions at all n-body levels, explicitly up to the 4-body term and implicitly for all n-body terms with $n > 4$.[76,127]

Training

The linear and nonlinear parameters in each of the n-body PIPs, i.e., $V^{\mathrm{2B}}_{\mathrm{PIP}}(\mathrm{M}_1, \mathrm{M}_2)$ in eq 11, $V^{\mathrm{3B}}_{\mathrm{PIP}}(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3)$ in eq 14, and $V^{\mathrm{4B}}_{\mathrm{PIP}}(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3, \mathrm{M}_4)$ in eq 17, were determined by minimizing the following loss function as in the original MB-pol PEF:[70,71]

$$\chi^2 = \sum_{k\in\Omega} w_k[\varepsilon^{n\mathrm{B}}_{\mathrm{MB-pol}}(k) - \varepsilon^{n\mathrm{B}}_{\mathrm{ref}}(k)]^2 + \Gamma^2 \sum_{l=1}^{v} c_l^2 \tag{22}$$

Here, Ω is the n-body training set, $\varepsilon^{n\mathrm{B}}_{\mathrm{MB-pol}}(k)$ and $\varepsilon^{n\mathrm{B}}_{\mathrm{ref}}(k)$ are the predicted and reference n-body energies, respectively, for the k-th configuration in the corresponding n-body training set; v is the number of terms in the corresponding n-body PIP; c_l is the linear parameter associated with the lth term; and Γ is a regularization parameter set to 5×10^{-4} for the 2-body energies and 1×10^{-4} for the 3-body and 4-body energies. The weights w_k were calculated to bias the fit in favor of low-energy n-body configurations:

$$w_k = \left(\frac{\Delta E}{E_k - E_{\mathrm{min}} + \Delta E}\right)^2 \tag{23}$$

where E_k is the binding energy of the k-th n-body configuration and E_{min} is the minimum binding energy in the corresponding n-body training set; and ΔE is a parameter set to 25 kcal/mol for the 2-body energies, and 37.5 kcal/mol for the 3-body and 4-body energies. By construction, w_k effectively reduces the impact unphysically-distorted configurations have on the training set. The polynomial generation and training process of the 2-body, 3-body, and 4-body terms of the MB-pol(2023) PEFs was carried out with the MB-Fit software.[129] Briefly, following the procedure described in ref 130, the non-linear fitting parameters were optimized using the Nelder-Mead simplex algorithm.[131] At each step of the simplex algorithm, the linear parameters were determined using ridge regression.[132]

The training sets and reference energies used to fit the 2-body, 3-body, and 4-body terms of the MB-pol(2023) PEFs were obtained from ref 73, and contain 71892 dimer configurations, 45332 trimer configurations, and 3692 tetramer configurations, respectively. As discussed in detail in ref 73, the reference 2-body energies were calculated at the CCSD(T) level of theory using a two-point extrapolation between the values obtained with the aug-cc-pVTZ and aug-cc-pVQZ basis sets, which were corrected for the basis set superposition error (BSSE) using the counterpoise method.[133] The reference 3-body energies were calculated at explicitly correlated CCSD(T) level of theory, i.e., CCSD(T)-F12a, with the the aug-cc-pVTZ basis set and corrected for the BSSE using the counterpoise method.[133] The reference 4-body energies were calculated at CCSD(T)-F12 level of theory with the heavy-aug-cc-pvtz basis set. For more details about the training set configurations and energy calculations, we refer the reader to ref 73. It should be noted that, based on eq 22 and our fitting process, 2-body and 3-body configurations that lie outside the range of action of the corresponding 2-body (eq 12) and 3-body (eq 14) switching functions have no effect on the loss function. As a result, the actual 2-body and 3-body training sets used in the development of the MB-pol(2023) PEFs contain 71117 dimers and 38631 trimers. Since the original 4-body training set is relatively small,[73] we retained all tetramer configurations and set the 4-body switching function equal to 1 during the fitting process (i.e., s_4 was only applied after $V_{\mathrm{PIP}}^{\mathrm{4B}}(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4)$ was optimized). Since the dimer, trimer, and tetramer binding energies calculated at the same level of theory used in the development of the original 2-body, 3-body, and 4-body training sets[73] were not available to us,[134] each E_k in eq 23 was calculated with the q-AQUA PEF that has been shown to accurately reproduce the the 2-body, 3-body, and 4-body energies of the training sets.[73] Root-mean-square errors (RMSEs) associated with 2-body, 3-body, and 4-body energies calculated with the MB-pol(2023) PEFs on the corresponding training sets discussed above are summarized in Table 1.

Benchmarks

Following our previous studies, the accuracy of the MB-pol(2023) PEFs was assessed through a systematic analysis of the energetics of the water hexamer as well as various structural and thermodynamic properties of liquid water studied as a function of temperature. Specifically, we performed a many-body decomposition analysis for the first eight low-lying energy isomers of the water hexamer shown in Fig. 2 whose geometries were taken from ref 89. As discussed in the literature,[135–137] the hexamer cluster holds a special place in the development of water models because it is the smallest water cluster for which the low-lying isomers display three-dimensional hydrogen-bonded structures similar to those found in liquid water and ice. Importantly, it has been shown that the ability of a water model to accurately reproduce each individual *n*-body contribution to the interaction energies of the hexamer isomers is directly correlated to the ability of the model to correctly predict the properties of water across different phases and thermodynamic state points.[76,138,139]

To assess the ability of the MB-pol(2023) PEFs to predict the properties of liquid water, MD simulations were performed in the canonical (NVT: constant number of molecules, volume, and temperature) and isothermal-isobaric (NPT: constant number of molecules, pressure, and temperature) ensembles for a cubic box of $N = 256$ water molecules in periodic boundary conditions. The NVT simulations were carried out at a temperature of 298 K and corresponding experimen-

Table 1: Summary of the 2-body, 3-body, and 4-body models reported in this study. Root-mean-square error (RMSE) is in kcal/mol. RMSE(BE<25) corresponds to the RMSE calculated for configurations with binding energies lower than 25 kcal/mol, with the binding energies calculated with the q-AQUA PEF.[73]

2-body energy			
model	2a	2b	2c
degree of polynomial	4	5	6
number of parameters	1153	2175	3828
RMSE	0.1542	0.0875	0.0492
RMSE (BE<25)	0.0431	0.0296	0.0219
3-body energy			
model	3a	3b	3c
degree of polynomial	4	4	4
number of parameters	1163	1804	2207
RMSE	0.0630	0.0556	0.0509
RMSE (BE<25)	0.0304	0.0231	0.0214
4-body energy			
model		4b	4c
degree of polynomial		4	4
number of parameters		266	266
RMSE		0.0430	0.0273
RMSE (BE<25)		0.0425	0.0264

tal density of 0.997 g·cm^{-3}. The NPT simulations were carried out at a pressure 1 atm over the temperature range from 238 K to 338 K. The velocity-Verlet algorithm was used to propagate the equations of motion with a time step of 0.5 fs according to ref 140. In both NVT and NPT simulations the temperature was controlled by a global Nosé–Hoover chain of 3 thermostats with a relaxation time of 0.25 ps. In the NPT simulations the pressure was controlled by a global Nosé–Hoover barostat with a relaxation time of 2.5 ps whose temperature was controlled by a Nosé–Hoover chain of three thermostats. After an equilibration time of 0.1 ns, both NVT and NPT simulations were run for production with durations ranging from 0.35 ns to 2.0 ns, depending on the temperature. A real-space cutoff of 9 Å was applied to evaluate all short-range nonbonded interactions, while all long-range interactions (including electrostatic, dispersion, and polarization contributions) were calculated in reciprocal space using the particle mesh Ewald (PME) solver as implemented in the helPME library.[141,142] All MD simulations were carried out with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[143] package through “fix_mbx” and “pair_mbx” styles that enable the interface with the MBX software for the calculation of many-

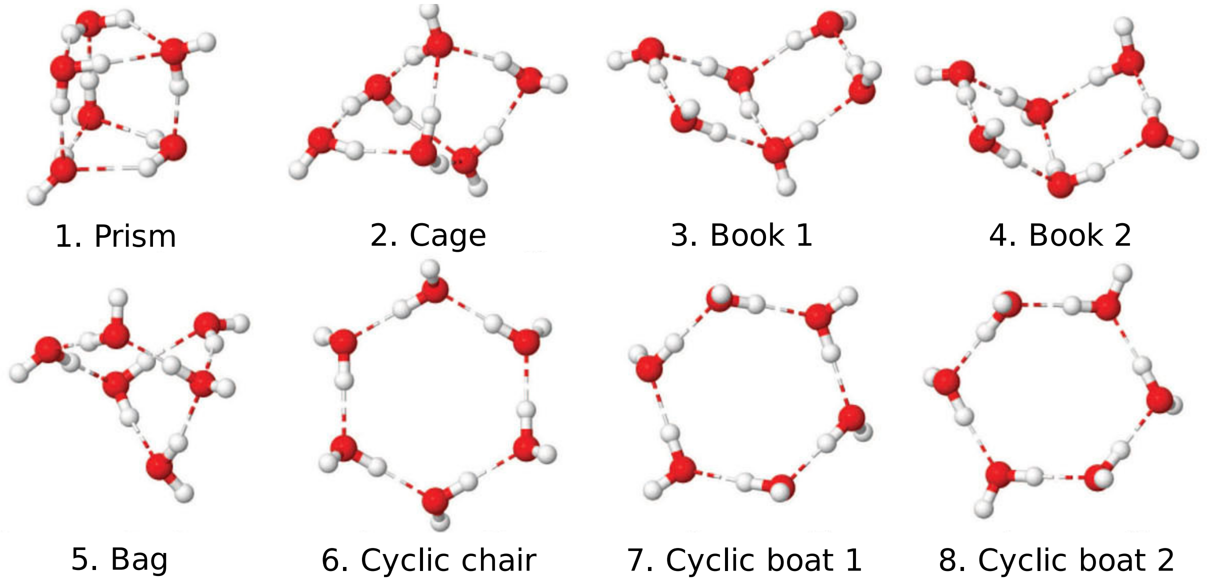


Figure 2: Structures of the first eight low-lying energy isomers of the water hexamer used in the analysis of interaction and many-body energies. The structure of each isomer was taken from ref 89.

body energies and forces in both periodic and non-periodic boundary conditions.[144]

Besides calculating the radial distribution functions (RDFs), the structure of liquid water predicted by the MD simulations carried out with the different MB-pol(2023) PEFs was also characterized by analyzing the tetrahedral order parameter q_{tet} defined by:[145]

$$q_{\text{tet}} = 1 - \frac{3}{8} \cdot \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\psi_{jk}) + \frac{1}{3} \right) \tag{24}$$

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

RESULTS

Interaction and many-body energies

Following from eq 3 and as discussed in ref 76, the accuracy of a many-body PEF developed within the MB-pol theoretical/computational framework depends on both the underlying classical many-body model (i.e., $V_{\text{perm}}^{\text{2B}}$, $V_{\text{disp}}^{\text{2B}}$, and V_{pol}) and the complexity of the n-body PIPs used to correct for the short-range quantum-mechanical nature of the n-body energies. In particular, since the larger PIPs in the models b and c have greater flexibility, they are, consequently, able to more accurately reproduce quantum-mechanical n-body energies. We will begin by analyzing the errors made at the different many-body levels by the different n-body models adopted by the MB-pol(2023) PEFs.

Fig. 3a shows that the 2-body error decreases significantly when the number of terms approximately doubles from model 2a (1153 terms) to the model 2b (2175 terms). However, doubling the number of terms again from model 2b to model 2c (3828 terms) does not show the same drastic improvement. While model 2c exhibits smaller errors for most of the hexamer isomers, it is unable to improve the description of 2-body energies for the lowest isomers (i.e., the prism and cage isomers). Nonetheless, it is worth mentioning that the order of the error is lower than 0.1 kcal/mol,

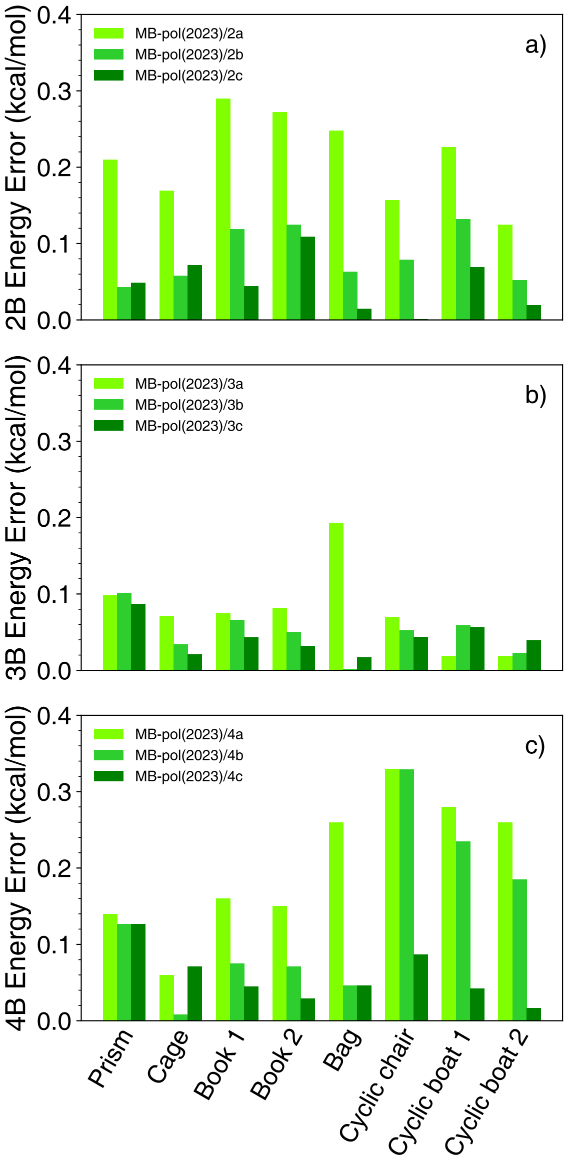


Figure 3: Unsigned errors associated with the different MB-pol(2003) models of a) 2-body, b) 3-body, and c) 4-body energies relative to the CCSD(T) reference energies reported in ref 73.

which is close to the intrinsic limit of accuracy of CCSD(T) calculations.[75]

While the errors in 3-body energies reported in Fig. 3b do improve to some extent as the number of terms in the 3-body PIP is increased from model 3a (1163) to model 3b (1804), and again to model 3c (2207), the change is not significant (except for the bag isomer). This is likely because the 3-body errors for all the isomers, except for the bag isomer, are already very small, with their absolute values being less than 0.1 kcal/mol.

As discussed in the Theory and Methods section, due to the limited number of configurations in the training set, only 266 terms were used for the 4-body PIP which were combined with two different switching functions (eq. 18) to build the 4b and 4c models. Due to a shorter cutoff distance (R_{out} in eq. 19), model 4b exhibits some deficiencies in describing configurations where the water molecules are distant from each other since some relevant tetramers are outside the switching range and thus not included in $V_{\mathrm{ML}}^{\mathrm{4B}}$. On the other hand, Fig. 3c shows that model 4c, which adopts a larger cutoff distance in the switching function (eq. 18), is able to accurately describe all hexamer isomers.

The magnitude of the errors associated with the different n-body models shown in Fig. 3 can be put into perspective by analyzing individual n-body contributions to the interaction energies of each hexamer isomer, as shown in Fig. 4. As expected,[76] compared to models 2a and 3a that adopt the same 2-body and 3-body PIPs as the original MB-pol PEF, respectively, models 2b and 2c, and models 3b and 3c predict 2-body and 3-body energies that get progressively closer to the CCSD(T) reference values. The inclusion of $V_{\mathrm{PIP}}^{\mathrm{4B}}$ in models 4b and 4c significantly improves the descriptions of 4-body energies when compared to the original MB-pol PEF, which does not include a PIP 4-body term (model a). This behavior is consistent with ref 146.

Fig. 5 reports a systematic analysis of the interaction energies calculated for the hexamer isomers using the different MB-pol(2023) PEFs obtained by combining the various 2-body, 3-body, and 4-body models presented in Fig. 4. For comparison, we report in Fig. 5 the CCSD(T) reference interaction energies[89] as well as the corresponding values calculated with the q-AQUA PEF in ref 73 and the original MB-pol PEF in ref 89. To provide further insights into the performance of the MB-pol(2023) PEFs, also shown in Fig. 5 are the interaction energies obtained by using the reference CCSD(T) values for the low-order n-body energies and the implicit many-body polarization term (V_{pol}) adopted by the MB-pol PEFs for the higher-order n-body energies, which are labeled as CCSD(T)+V_{pol}. For example, CCSD(T)(2B+3B)+$V_{\mathrm{pol}}^{>3\mathrm{B}}$ corresponds to a hybrid model that uses the CCSD(T) 2-body and 3-body reference energies, and the many-body polarization $V_{\mathrm{pol}}^{>3\mathrm{B}}$ term for all n-body energies with $n > 3$. The differences between the CCSD(T) reference

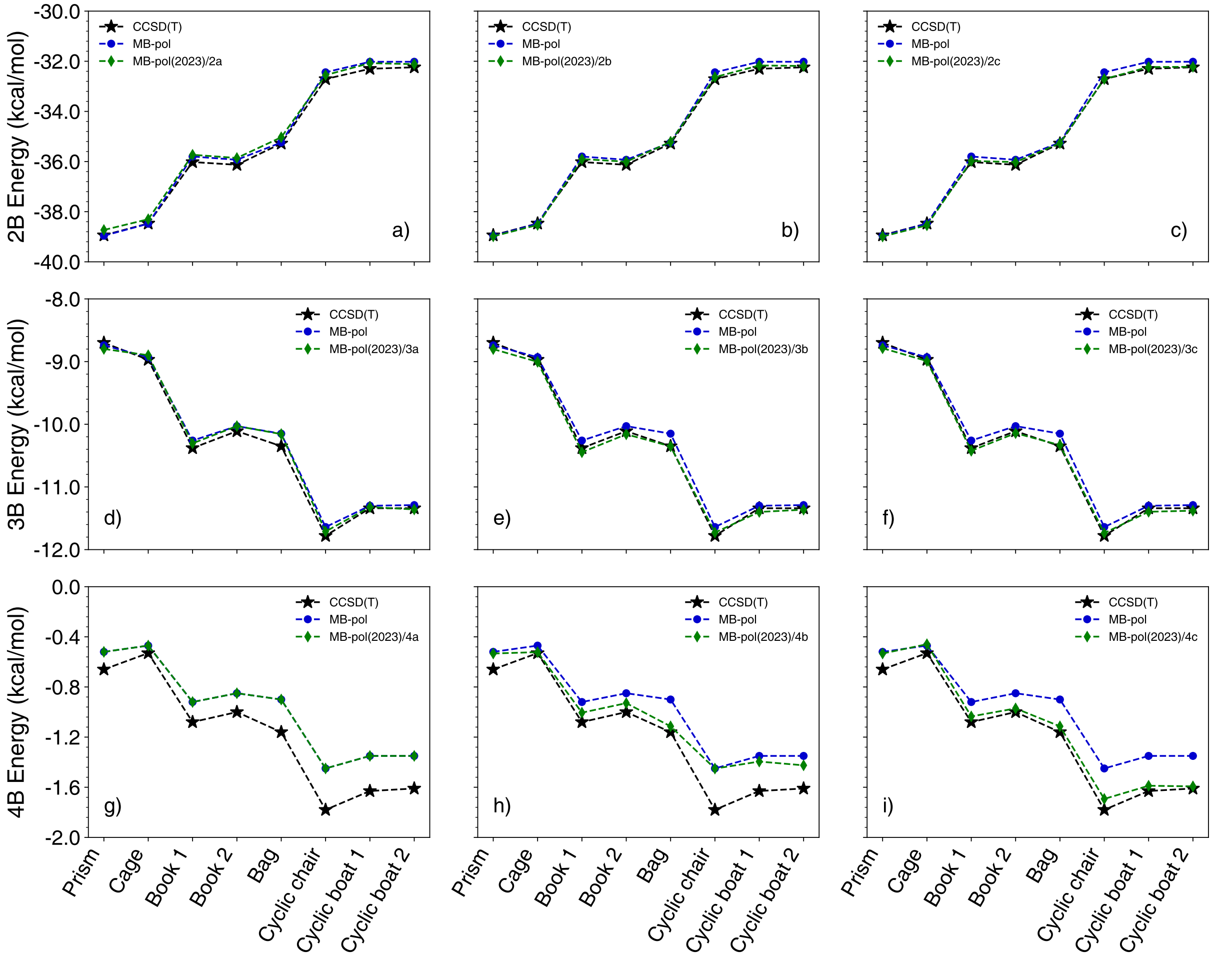


Figure 4: Many-body energies of the first eight isomers of the water hexamer calculated with the different MB-pol(2003) models. Panels a-c: 2-body energies calculated with the 2a, 2b, and 2c models. Panels d-f: 3-body energies calculated with the 3a, 3b, and 3c models. Panels g-i: 4-body energies calculated with the 4a, 4b, and 4c models. For comparison also shown in each panel are the MB-pol and CCSD(T) reference values reported in ref 89.

interaction energies and the corresponding CCSD(T)+V_{pol} values are due to the truncation of the explicit many-body terms, while the differences between the MB-pol(2023) interaction energies and the corresponding values calculated with the CCSD(T)+V_{pol} model are effectively a measure of how accurately the explicit n-body terms of the MB-pol(2023) PEFs reproduce the corresponding CCSD(T) reference values. Additional analyses of the interaction energies of the hexamer isomers calculated with the different MB-pol(2023) PEFs that do not include an explicit 4-body term are reported in the Supporting Information.

Fig. 5 clearly demonstrates that as the number of explicit *n*-body terms and the size of each *n*-body PIP increase, so does the agreement between the interaction energies predicted by the corresponding MB-pol(2023) PEFs and the CCSD(T) reference values. Specifically, Fig. 5a-c show the interaction energies calculated with the MB-pol(2023) PEFs that adopts models 2a, 2b, and 2c for the 2-body term, respectively, in combination with model 3c for the 3-body term, without an explicit 4-body term. While the 3-body model in the corresponding MB-pol(2023) PEFs does not change, the 2-body model increases in accuracy from model a to model c. Importantly, Fig. 5c

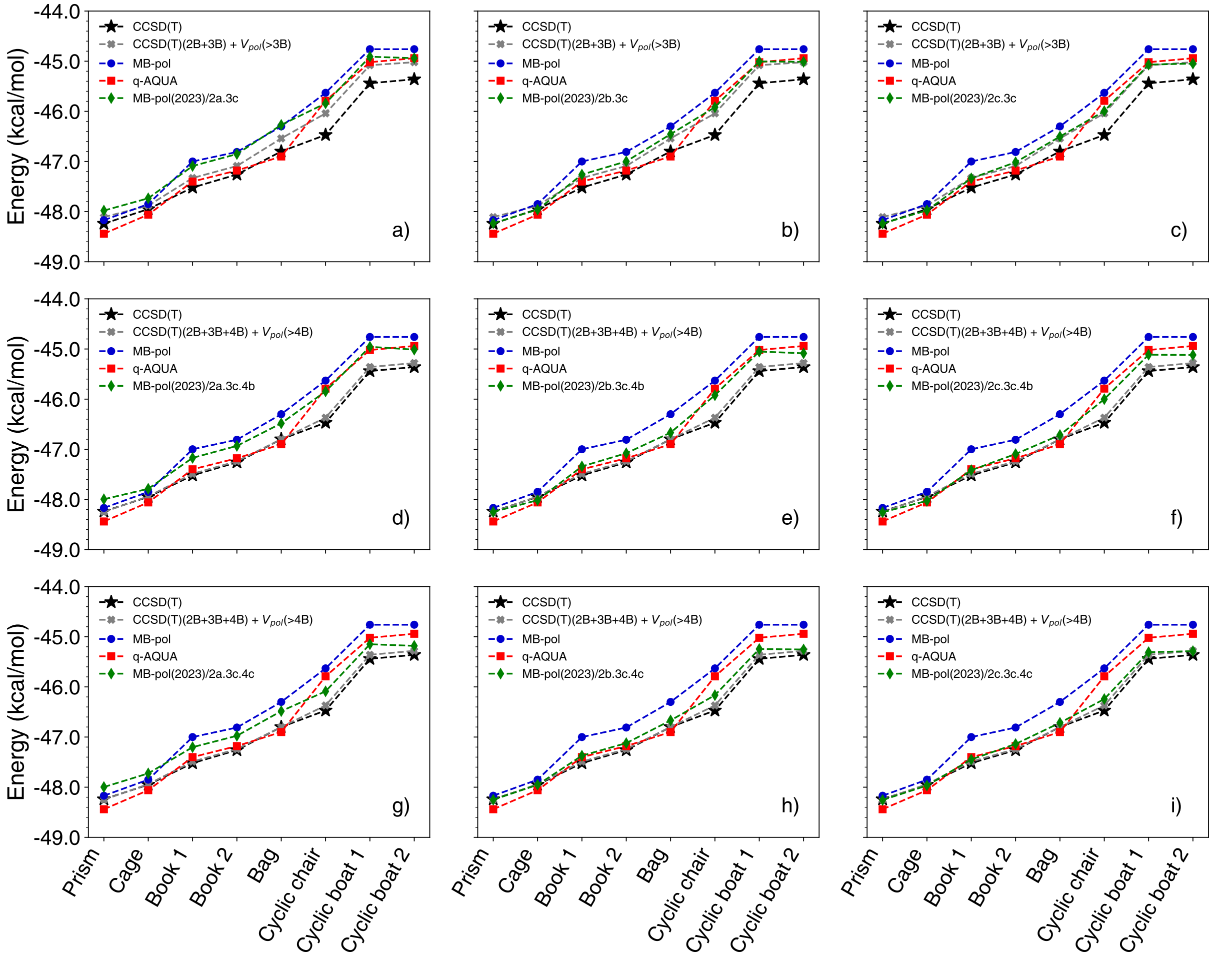


Figure 5: Interaction energies of the first eight isomers of the water hexamer calculated with the different MB-pol(2023) PEFs. Also shown for comparison are the MB-pol and q-AQUA values reported in refs. 89 and 73, respectively. The CCSD(T) reference values are from eq 89. CCSD(T)+V_{pol} refers to a hybrid model where the low-order *n*-body energies are represented by the CCSD(T) values and all higher-order *n*-body energies are represented by the implicit many-body polarization term (V_{pol}) adopted by the MB-pol PEFs. See main text for details.

shows that the interaction energies predicted by the MB-pol(2023)/2c.3c PEF effectively overlap with the corresponding values calculated with the hybrid CCSD(T)+V_{pol} model, indicating that models 2c and 3c closely reproduce the 2-body and 3-body CCSD(T) reference energies. The increased accuracy achieved by going from model 2a to model 2c provides support to the conclusions drawn in ref 76 that increasing the size of both PIPs and associated training sets within the MB-pol theoretical/computational framework leads to higher accuracy in the representations of n-body energies.

The effect on the interaction energies predicted by the MB-pol(2023) PEFs due to the addition of an explicit 4-body term described by models 4b and 4c is analyzed in Fig. 5d-f and Fig. 5g-i, respectively. All three MB-pol(2023) PEFs that include an explicit 4-body term represented by model 4b, i.e., MB-pol(2023)/2a.3c.4b, MB-pol(2023)/2b.3c.4b, and MB-pol(2023)/2c.3c.4b in Figs. 5d-f, respectively, improve upon the original MB-pol PEF. Importantly, the most accurate member of this family of PEFs, i.e., MB-pol(2023)/2c.3c.4b also improves upon the q-AQUA PEF,[73] but still exhibits some deviations from the CCSD(T) reference values for the planar hexamer isomers (cyclic chair, cyclic boat 1, and cyclic boat 2). The differences with the the CCSD(T) reference values can be traced back to the limitations of model 4b to correctly describe tetramers where the water molecules are distant from each other since these configurations lie outside the switching range as shown in Fig. 4.

Fig. 5g-i shows that adding an explicit 4-body term represented by model 4c significantly elevates the accuracy of the corresponding MB-pol(2023) PEFs, with MB-pol(2023)/2c.3c.4c quantitatively reproducing the CCSD(T) reference interaction energies of all hexamer isomers, outperforming both the original MB-pol PEF[70,71] and the more recent q-AQUA PEF.[73] In this regard, it should be noted that MB-pol(2023)/2b.3c.4c also improves upon q-AQUA, which is consistent with the analysis presented in Fig. 3 showing that the improvement in the representation of 2-body energies achieved by model 2c relative to model 2b is only marginal.

The analyses of many-body and interaction energies presented here provide a clear demonstration of how the accuracy of data-driven many-body PEFs of water developed within the MB-pol theoretical/computational framework, such as the MB-pol(2023) PEFs introduced in this study, can be systematically improved by 1) training the n-body PIPs on larger training sets, 2) including PIPs that explicitly represent higher n-body interactions, and 3) adopting higher-order PIPs with more terms to represent the n-body energies, as originally proposed in ref.[76] Importantly, the higher accuracy exhibited by the most sophisticated MB-pol(2023) PEFs introduced in this study (MB-pol(2023)/2b.3c.4c and MB-pol(2023)/2c.3c.4c), which include all many-body terms through the combination of machine-learned PIPs and implicit many-body polarization, when compared to q-AQUA,[73] which instead only includes up to 4-body effects, provides further support for the importance of including all n-body terms of the MBE when representing the molecular interactions in water.[76]

Liquid water

The next step in assessing the accuracy of the MB-pol(2023) PEFs involves analyzing the structural and thermodynamic properties of liquid water. To this end, we performed MD simulations in both NVT and NPT ensemble and calculated the RDFs and q_{tet} of liquid water at different temperatures. In the NVT simulations, the density was fixed at the experimental value (0.997 g/cm^3 at 298 K).

We begin our analysis focusing on the MB-pol(2023) PEFs that, as the original MB-pol PEF,[70,71] do not include an explicit 4-body term. Fig. 6 show that these MB-pol(2023) PEFs predict oxygen-oxygen RDF at 298 K in quantitative agreement with the experimental RDFs of ref 147. It should be noted that similar agreement with the experimental data was also obtained from MD simulations carried out with the original MB-pol PEF.[72,89,90] The quantitative agreement with the experimental RDFs provided by the different MB-pol(2023) PEFs is not surprising because, despite each of these PEF adopting different 2-body and 3-body PIPs, they are all able to reproduce 2-body and 3-body energies with sub-chemical accuracy (Fig. 3), which is significantly smaller than the thermal fluctuations occurring in liquid water (k$_B T$ = 0.59 kcal/mol at 298 K, with k$_B$ being Boltzmann's constant). The high accuracy of the MB-pol(2023) PEFs is further demonstrated by the close agreement between the RDFs calculated in the NVT and NPT ensembles which indicates that all MB-pol(2023) PEFs correctly represent the free-energy landscape of liquid water. Comparisons between the oxygen-oxygen RDFs calculated from NVT and NPT simulations carried out as a function of temperature with the different MB-pol(2023) PEFs that do not include an explicit 4-body term are reported in the Supporting Information.

Fig. 7 shows that, by consistently improving the representation of 2-body and 3-body energies, all MB-pol(2023) PEFs that do not include an explicit 4-body term also predict the density of water over the temperature range between 238 K and 338 K in closer agreement with the experimental values than the original MB-pol PEF . In particular, MB-pol(2023)/2c.3c predicts a density maximum of $\sim$0.999 g/cm^3 at $\sim$268 K, which is in good agreement with the experimental value of 1.000 g/cm^3 at 277 K.

Additional insights into the performance of the different MB-pol(2023) PEFs, which do not include an explicit 4-body term, can be gained from the analysis of the tetrahedral order parameter distributions $P(q_{\text{tet}})$. As discussed in the literature,[145] $P(q_{\text{tet}})$ is a direct probe of the local structure of liquid water. Fig. 8 shows that all different MB-pol(2023) PEFs predict similar trends for $P(q_{\text{tet}})$ calculated from NPT simulations carried out over the temperature range between 238 K and 338 K. In particular, all different MB-pol(2023) PEFs predict $P(q_{\text{tet}})$ to be bimodal at high temperatures, with two peaks at $q_{\text{tet}} \sim 0.5$ and $q_{\text{tet}} \sim 0.8$. As the temperature decreases, the peak at $q_{\text{tet}} \sim 0.8$ grows in intensity and shifts to higher q_{tet} values, while the peak at $q_{\text{tet}} \sim 0.5$ disappears, indicating the progressive development of a more tetrahedral liquid structure. An analogous evolution of $P(q_{\text{tet}})$ was obtained from NPT simulations carried out with MB-pol in ref.[148] It should be noted that, while all different MB-pol(2023) PEFs predict effectively indistinguishable $P(q_{\text{tet}})$ at 338 K, some noticeable differences are evident in $P(q_{\text{tet}})$ calculated at lower temperatures. In particular, all MB-pol(2023) PEFs that adopt model 3c to represent 3-body energies systematically predict relatively sharper $P(q_{\text{tet}})$ as the temperature decreases. Since the hydrogen-bond network in supercooled water becomes more extended, $P(q_{\text{tet}})$ at low temperature is expected to be more sensitive to many-body effects. The differences between $P(q_{\text{tet}})$ calculated with MB-pol(2023) PEFs that adopt models 3b and 3c can thus be traced back to the differences between models 3b

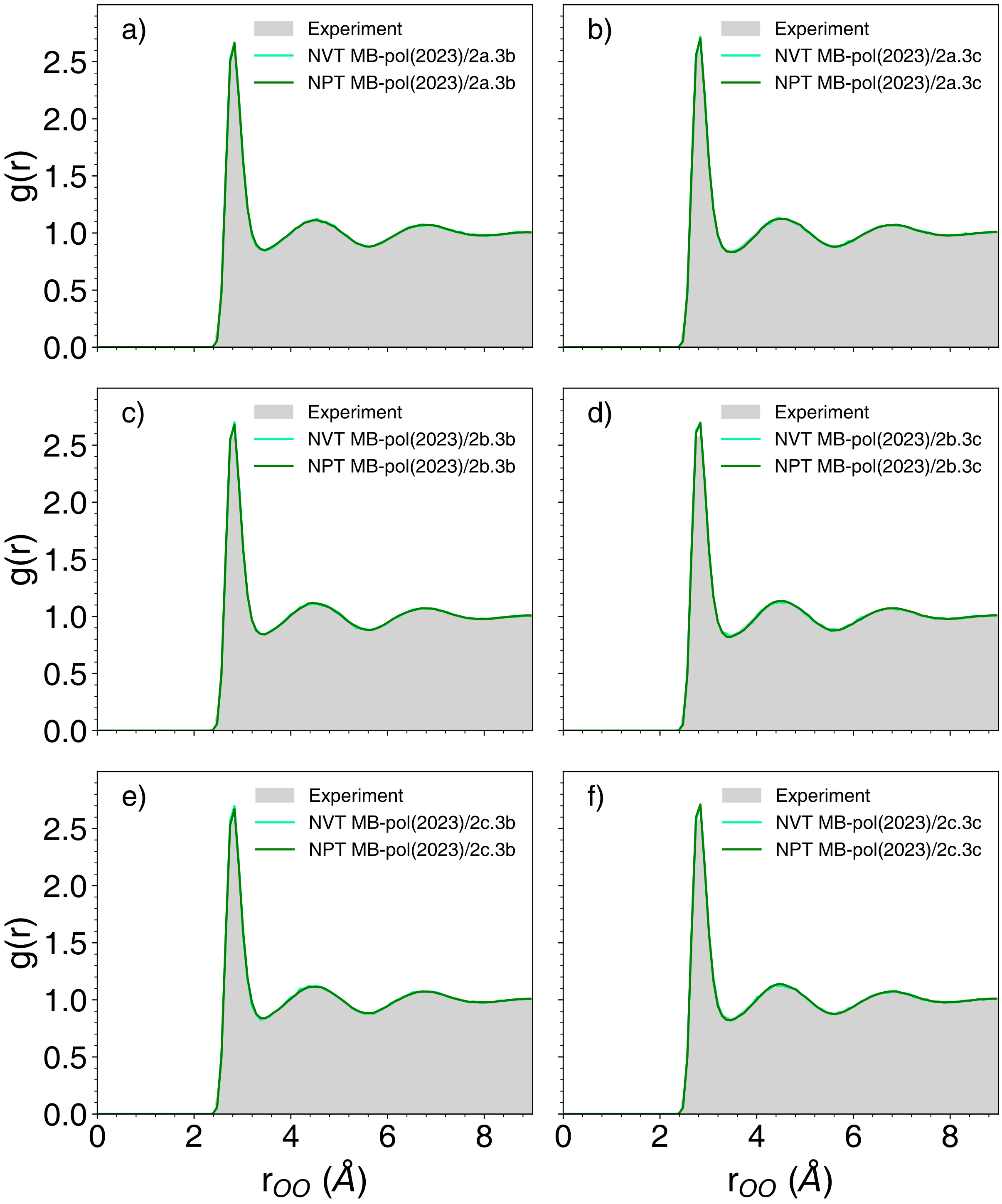


Figure 6: Oxygen–oxygen radial distribution function, $g(r)$, calculated from NPT simulations carried out at 1 atm and 298 K with the different MB-pol(2023) PEFs that do not include an explicit 4-body term. See main text for details about the different MB-pol(2023) PEFs.

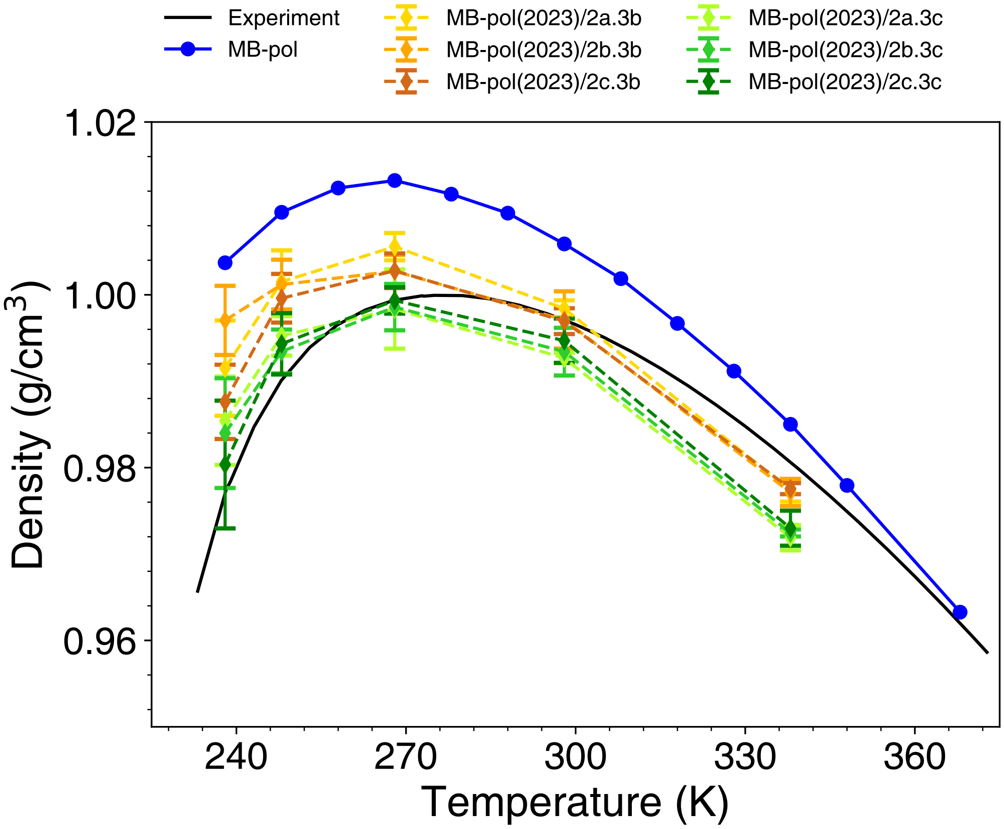


Figure 7: Density of liquid water calculated from NPT simulations carried out as a function of temperature at 1 atm with the different MB-pol(2023) PEFs that do not include an explicit 4-body term. Error bars represent 95% confidence intervals. See main text for details about the different MB-pol(2023) PEFs.

and 3c in representing 3-body energies, with model 3c providing overall closer agreement with the CCSD(T) reference values as shown in Fig. 3. This is also consistent with all MB-pol(2023) PEFs that use mode 3c predicting liquid densities in closer agreement with the experimental values at low temperature (Fig. 7).

Finally, we examine the impact that including an explicit 4-body term has on the performance of the MB-pol(2023) PEFs when used to simulate liquid water. To this end, we performed MD simulations at 298 K in both NVT and NPT ensembles using the MB-pol(2023)/2c.3c.4b and MB-pol(2023)/2c.3c.4c PEFs, which, as discussed in the Theory and Methods section, only differ in the switching function adopted by the corresponding $V_{\text{ML}}^{\text{4B}}$ terms. Fig. 9a shows that the oxygen-oxygen RDFs calculated from NVT and NPT simulations carried out with MB-pol(2023)/2c.3c.4b are very similar. Such agreement is also found for $P(q_{\rm tet})$ calculated from NVT and NPT simulations (Fig. 9b). Small differences, however, exist between the RDF calculated in the NPT ensemble and the experimental RDF which manifest in MB-pol(2023)/2c.3c.4b predicting a density of

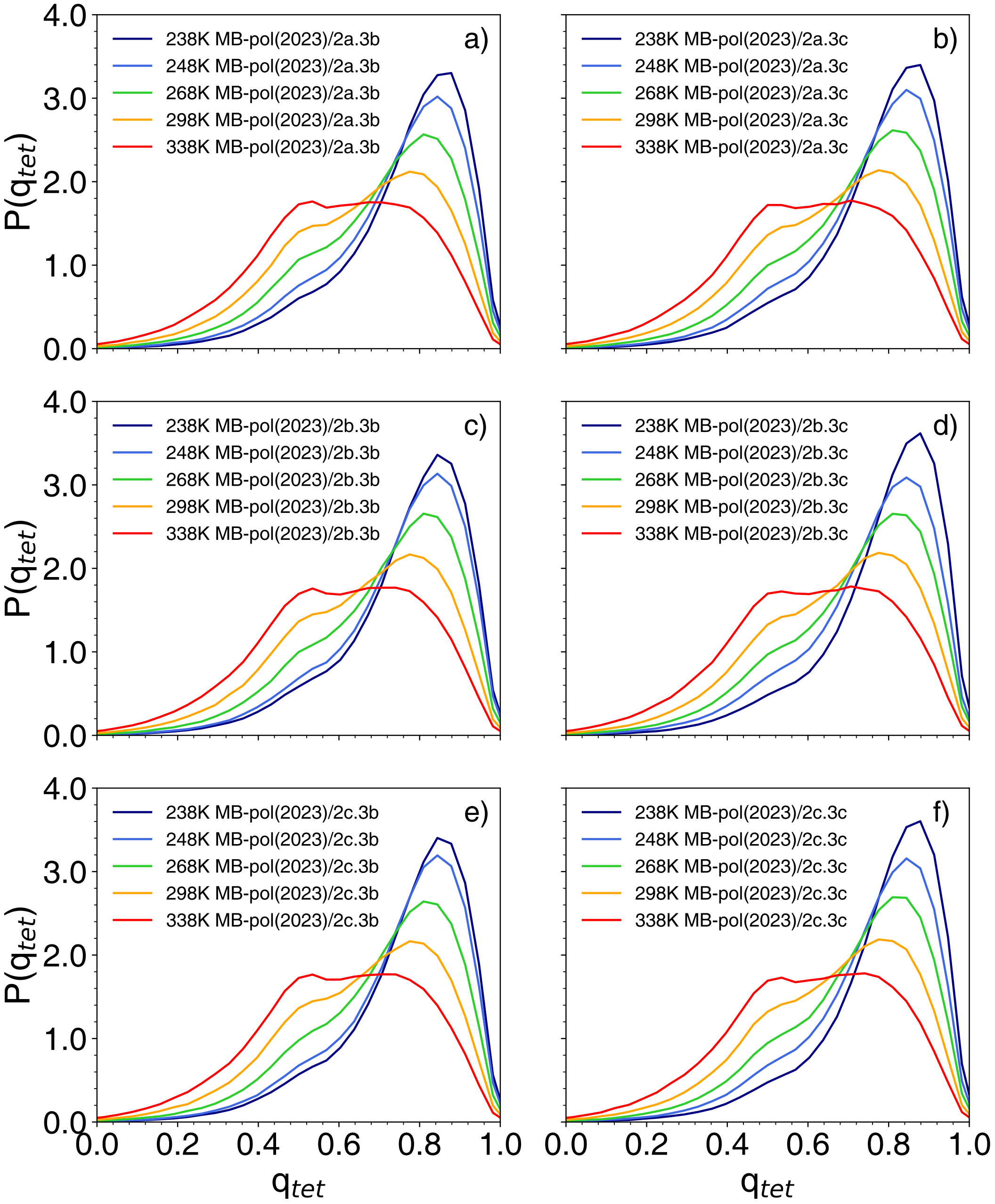


Figure 8: Probability distribution of the tetrahedral order parameter, $P(q_{\rm tet})$, calculated from NPT simulations carried out as a function of temperature at 1 atm with the MB-pol(2023) PEFs that do not include an explicit 4-body term. See main text for details about the different MB-pol(2023) PEFs.

1.015±0.002 g/cm^3 at 298 K compared to the experimental density of 0.997 g/cm^3.

In contrast, noticeable differences between the NVT and NPT RDFs and $P(q_{\mathrm{tet}})$ calculated with MB-pol(2023)/2c.3c.4c are evident in Fig. 9b and Fig. 9d, respectively. In particular, MB-pol(2023)/2c.3c.4c predicts a more disordered liquid structure as demonstrated by the relatively higher amplitude of the RDF between ∼2.6 Å and ∼2.6 Å, which is indicative of the presence of

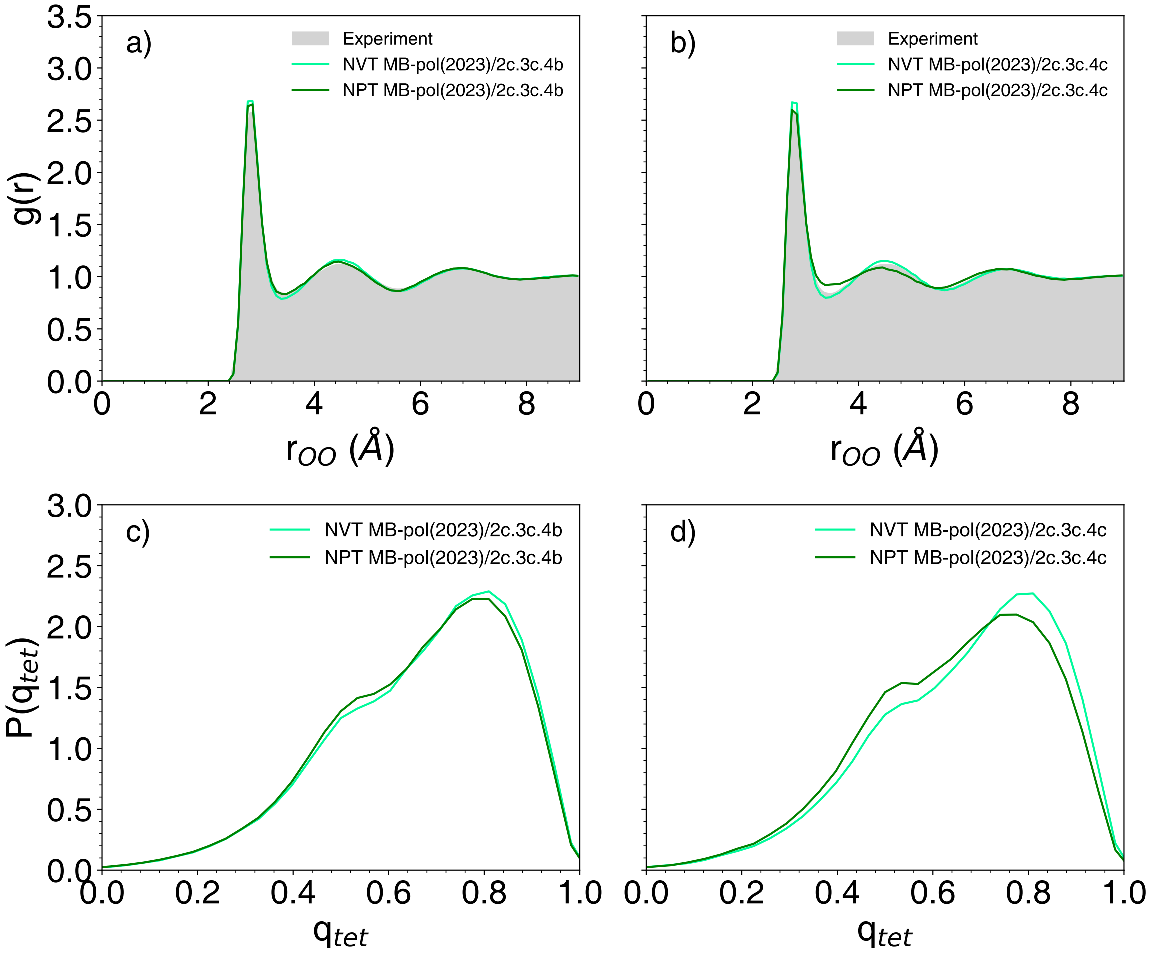


Figure 9: a-b) Oxygen–oxygen radial distribution function, $g(r)$, calculated from NVT (light green) and NPT (dark green) simulations carried out at 298 K with the MB-pol(2023)/2c.3c.4b and MB-pol(2023)/2c.3c.4c PEFs, respectively. The experimental radial distribution function from ref 147 is shown as a grade shade. c-d) Probability distribution of the tetrahedral order parameter, $P(q_{\mathrm{tet}})$, calculated from NVT (light green) and NPT (dark green) simulations carried out at 298 K with the MB-pol(2023)/2c.3c.4b and MB-pol(2023)/2c.3c.4c PEFs, respectively. In the NVT simulations, the density of liquid water was fixed at the experimental value of 0.997 g/cm^3. In the NPT simulations, the pressure was fixed at 1 atm. See main text for details about the different MB-pol(2023) PEFs.

interstitial water molecules between the first and second solvation shells, and a relatively wider $P(q_{\text{tet}})$. As a result, MB-pol(2023)/2c.3c.4c predicts a density of 1.060±0.003 g/cm^3 at 298 K, which is significantly larger than the experimental value of 0.997 g/cm^3.

The unsatisfactory performance of MB-pol(2023)/2c.3c.4b and MB-pol(2023)/2c.3c.4c on modeling liquid water when compared to the performance of the MB-pol(2023) PEFs that do not include an explicit 4-body term can be rationalized by analyzing the size and composition of the 4-body training set introduced in ref 73. As discussed in the Theory and Methods section, the 4-body training set only contains 3692 tetramers, which should be compared with 2-body and 3-body containing 71892 dimers and 45332 trimers, respectively. It follows that, because of its relatively small size, the available 4-body training set is thus unlikely to provide a "complete" representation of the tetramer configuration space, which is needed to correctly describe 4-body energies across different phases and thermodynamic states. While MB-pol(2023)/2c.3c.4c was able to effectively provide "CCSD(T) accuracy" for the interaction energies of the low-lying hexamer isomers, it is less accurate at modeling the properties of liquid water, because doing so requires visiting regions of the tetramer configuration space that are not fully represented in the training set.

The different performance of the MB-pol(2023)/2c.3c.4b and MB-pol(2023)/2c.3c.4c PEFs on modeling liquid water can, instead, be explained by considering the different ranges of action of the switching function adopted by models 4b and 4c. As discussed in the Theory and Methods section, model 4c employs a larger cutoff distance that allows MB-pol(2023)/2c.3c.4c to quantitatively reproduce the CCSD(T) reference energies of the hexamer isomers. However, since this larger cutoff distance extends beyond well-defined tetramers found in the gas phase, it enables the 4-body term on tetramers in the liquid phase that are not represented in the relatively small 4-body training set, resulting in the somewhat ill-behavior of MB-pol(2023)/2c.3c.4c in simulations of liquid water. On the other hand, the smaller cutoff distance adopted by model 4b effectively limits the impact of $V_{\text{ML}}^{\text{4B}}$ on the performance of MB-pol(2023)/2c.3c.4b, reducing the negative impact on the predicted water properties but also having a lesser positive impact on the hexamer interaction energies.

These analyses suggest that, while the size and composition of the current 4-body dataset are sufficient for modeling 4-body energies in gas-phase clusters, they appear to be inadequate to correctly represent the diversity of tetramer configurations found in the liquid phase. This implies that, within the MB-pol theoretical/computational framework, the explicit 4-body term must be trained on larger and more diverse 4-body datasets than currently available in order to guarantee its full transferability from the gas to the liquid phase. It should, however, be noted that the MB-pol(2023) PEFs that do not include an explicit 4-body term already improve upon the original MB-pol PEF, providing closer agreement with experimental data for various properties of liquid water. Importantly, while MB-pol(2023)/2c.3b and MB-pol(2023)/2c.3c appear to provide the closest agreement to reference data, given the magnitude of thermal fluctuations at finite temperature, which are significantly larger than the differences in 2-body and 3-body errors shown in Fig. 3, all MB-pol(2023) PEFs are expected to perform similarly in simulations of liquid water across different temperatures and pressures.

Conclusions

Building on the success of the MB-pol data-driven many-body PEF of water, which was introduced by our group ten years ago,[70–72] in this study we have developed a family of MB-pol(2023) PEFs that improve upon the original MB-pol PEF by 1) training the machine-learned n-body PIPs on larger n-body datasets for $n = 2$, 3, and 4, which have recently become available,[73] 2) including an explicit representation of 4-body energies through a corresponding machine-learned 4-body PIP, and 3) adopting higher-order PIPs with more terms than the original MB-pol PEF to represent the 2-body and 3-body energies.

Through systematic analyses of many-body and interaction energies of the hexamer clusters, we demonstrated that, as the number of explicit n-body terms and the size of each n-body PIP increase, all MB-pol(2023) PEFs improve upon the original MB-pol PEF and progressively approach CCSD(T) accuracy. In particular, the most sophisticated MB-pol(2023) PEF, corresponding to MB-pol(2023)/2c.3c.4c, quantitatively reproduces CCSD(T) 2-body, 3-body, and 4-body energies as well as interaction energies of the hexamer isomers, outperforming the q-AQUA PEF that currently provides the most accurate description of water clusters.[73]

MD simulations of liquid water carried out with the MB-pol(2023) PEFs in both NVT and NPT ensembles show that including an explicit 4-body term does not necessarily improve the performance of data-driven many-body PEFs developed within the MB-pol theoretical/computational framework unless the size of the corresponding 4-body training set is sufficiently large to properly represent the diversity of tetramers found in the liquid phase.

Considering that the original MB-pol PEF has already demonstrated outstanding accuracy,[76] we believe that the MB-pol(2023) PEFs that do not include an explicit 4-body term will enable simulations of water across different phases with even higher accuracy, effectively closing the gap with experimental measurements. Importantly, as already discussed in ref 76, our study demonstrates that the MB-pol theoretical/computational framework provides a rigorous and efficient platform for the development of data-driven many-body PEFs for water that can be systematically improved as larger and more diverse training sets of n-body energies become available. This implies that the MB-pol(2023) PEFs introduced in our study can be trivially improved, without changing the underlying functional form, by training the current 2-body, 3-body, and 4-body terms on more "complete" datasets and/or adding explicit n-body terms with $n > 4$. In this regard, it should, however, be noted that, as discussed in our study, the original MB-pol PEF as well as the new MB-pol(2023) PEFs (with explicit terms up to the 3-body term) already achieve sub-chemical accuracy, which implies that the additional accuracy gained by increasing the complexity of the MB-pol(2023) PEFs will increase the associated computational cost and only result in a marginal improvement in the description of the properties of water, especially at finite temperature where thermal fluctuations are significantly larger than the intrinsic errors associated with the MB-pol(2023) representations of n-body energies.

ASSOCIATED CONTENT

Supporting Information

Technical details about the n-body PIPs adopted by the MB-pol(2023) PEFs. Correlation plots for 2-body, 3-body, and 4-body energies. Additional analyses of the interaction energies of the hexamer isomers and radial distribution functions calculated with the MB-pol(2023) PEFs that do not include an explicit 4-body term.

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation through grants nos. 1954895 and 2102309. The simulations used resources of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation through Grant No. 1548562, and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center.

Data availability

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

References

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

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

(3) Guillot, B. A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water. *J. Mol. Liq.* **2002**, *101*, 219–260.

(4) Vega, C.; Abascal, J. L. Simulating Water with Rigid Non-Polarizable Models: A General Perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663–19688.

(5) Shvab, I.; Sadus, R. J. Atomistic Water Models: Aqueous Thermodynamic Properties from Ambient to Supercritical Conditions. *Fluid Phase Equilib.* **2016**, *407*, 7–30.

(6) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501–7528.

(7) Reimers, J.; Watts, R.; Klein, M. Intermolecular Potential Functions and the Properties of Water. *Chem. Phys.* **1982**, *64*, 95–114.

(8) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981. 1981; pp 331–342.

(9) Berendsen, H.; Grigera, J.; Straatsma, T. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(11) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(12) Abascal, J. L.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(13) Abascal, J.; Sanz, E.; García Fernández, R.; Vega, C. A Potential Model for the Study of Ices and Amorphous Water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511.

(14) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(15) González, M. A.; Abascal, J. L. A Flexible Model for Water Based on TIP4P/2005. *J. Chem. Phys.* **2011**, *135*, 224516.

(16) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing Quantum Effects in the Dynamics of a Flexible Water Model. *J. Chem. Phys.* **2009**, *131*, 024501.

(17) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006**, *124*, 024503.

(18) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham III, T. E.; Voth, G. A. An Accurate and Simple Quantum Model for Liquid Water. *J. Chem. Phys.* **2006**, *125*, 184507.

(19) Akin-Ojo, O.; Song, Y.; Wang, F. Developing Ab Initio Quality Force Fields from Condensed Phase Quantum-Mechanics/Molecular-Mechanics Calculations through the Adaptive Force Matching Method. *J. Chem. Phys.* **2008**, *129*, 064108.

(20) Kumar, R.; Skinner, J. L. Water Simulation Model with Explicit Three-Molecule Interactions. *J. Phys. Chem. B* **2008**, *112*, 8311–8318.

(21) Tainter, C.; Pieniazek, P. A.; Lin, Y.-S.; Skinner, J. L. Robust Three-Body Water Simulation Model. *J. Chem. Phys.* **2011**, *134*, 184501.

(22) Tainter, C. J.; Shi, L.; Skinner, J. L. Reparametrized E3B (Explicit Three-Body) Water Model Using the TIP4P/2005 Model as a Reference. *J. Chem. Theory Comput.* **2015**, *11*, 2268–2277.

(23) Kiss, P. T.; Baranyai, A. A Systematic Development of a Polarizable Potential of Water. *J. Chem. Phys.* **2013**, *138*, 204507.

(24) Lamoureux, G.; MacKerell Jr, A. D.; Roux, B. A Simple Polarizable Model of Water Based on Classical Drude Oscillators. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(25) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell Jr, A. D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249.

(26) Yu, W.; Lopes, P. E.; Roux, B.; MacKerell Jr, A. D. Six-Site Polarizable Model of Water Based on the Classical Drude Oscillator. *J. Chem. Phys.* **2013**, *138*, 034508.

(27) Yu, H.; Hansson, T.; van Gunsteren, W. F. Development of a Simple, Self-Consistent Polarizable Model for Liquid Water. *J. Chem. Phys.* **2003**, *118*, 221–234.

(28) Yu, H.; van Gunsteren, W. F. Charge-on-Spring Polarizable Water Models Revisited: From Water Clusters to Liquid Water to Ice. *J. Chem. Phys.* **2004**, *121*, 9549–9564.

(29) Burnham, C. J.; Xantheas, S. S. Development of Transferable Interaction Models for Water. III. Reparametrization of an All-Atom Polarizable Rigid Model (TTM2-R) from First Principles. *J. Chem. Phys.* **2002**, *116*, 1500–1510.

(30) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. Development of Transferable Interaction Models for Water. II. Accurate Energetics of the First Few Water Clusters from First Principles. *J. Chem. Phys.* **2002**, *116*, 1493–1499.

(31) Burnham, C. J.; Xantheas, S. S. Development of Transferable Interaction Models for Water. IV. A Flexible, All-Atom Polarizable Potential (TTM2-F) Based on Geometry Dependent

Charges Derived from an Ab Initio Monomer Dipole Moment Surface. *J. Chem. Phys.* **2002**, *116*, 5115–5124.

(32) Fanourgakis, G. S.; Xantheas, S. S. The Flexible, Polarizable, Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.

(33) Fanourgakis, G. S.; Xantheas, S. S. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-Type Model Potential (TTM3-F, v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.* **2008**, *128*, 074506.

(34) Burnham, C.; Anick, D.; Mankoo, P.; Reiter, G. The Vibrational Proton Potential in Bulk Liquid Water and Ice. *J. Chem. Phys.* **2008**, *128*, 154519.

(35) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(36) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *J. Phys. Chem. B* **2004**, *108*, 13427–13437.

(37) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(38) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117*, 9956–9972.

(39) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122–4139.

(40) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of Geometry Dependent Charge Flux into Polarizable AMOEBA+ Potential. *J. Phys. Chem. Lett.* **2020**, *11*, 419–426.

(41) Mauger, N.; Plé, T.; Lagardère, L.; Huppert, S.; Piquemal, J.-P. Improving Condensed-Phase Water Dynamics with Explicit Nuclear Quantum Effects: The Polarizable Q-AMOEBA Force Field. *J. Phys. Chem. B* **2022**, *126*, 8813–8826.

(42) Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-P.; Cisneros, G. A. GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1361–1365.

(43) Duke, R. E.; Cisneros, G. A. Ewald-Based Methods for Gaussian Integral Evaluation: Application to a New Parameterization of GEM. *J. Mol. Model.* **2019**, *25*, 307.

(44) Mankoo, P. K.; Keyes, T. POLIR: Polarizable, Flexible, Transferable Water Potential Optimized for IR Spectroscopy. *J. Chem. Phys.* **2008**, *129*, 034504.

(45) Hasegawa, T.; Tanimura, Y. A Polarizable Water Model for Intramolecular and Intermolecular Vibrational Spectroscopies. *J. Phys. Chem. B* **2011**, *115*, 5545–5553.

(46) Das, A. K.; Urban, L.; Leven, I.; Loipersberger, M.; Aldossary, A.; Head-Gordon, M.; Head-Gordon, T. Development of an Advanced Force Field for Water Using Variational Energy Decomposition Analysis. *J. Chem. Theory Comput.* **2019**, *15*, 5001–5013.

(47) Rackers, J. A.; Silva, R. R.; Wang, Z.; Ponder, J. W. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.* **2021**, *17*, 7056–7084.

(48) Pople, J. A. Nobel Lecture: Quantum Chemical Models. *Rev. Mod. Phys.* **1999**, *71*, 1267.

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

(50) Nesbet, R. K. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 1969; pp 1–34.

(51) Stoll, H. Correlation Energy of Diamond. *Phys. Rev. B* **1992**, *46*, 6700.

(52) Stoll, H. On the Correlation Energy of Graphite. *J. Chem. Phys.* **1992**, *97*, 8449–8454.

(53) Stoll, H. The Correlation Energy of Crystalline Silicon. *Chem. Phys. Lett.* **1992**, *191*, 548–552.

(54) Paulus, B.; Rosciszewski, K.; Gaston, N.; Schwerdtfeger, P.; Stoll, H. Convergence of the Ab Initio Many-Body Expansion for the Cohesive Energy of Solid Mercury. *Phys. Rev. B* **2004**, *70*, 165106.

(55) Stoll, H.; Paulus, B.; Fulde, P. On the Accuracy of Correlation-Energy Expansions in Terms of Local Increments. *J. Chem. Phys.* **2005**, *123*, 144108.

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

(57) Matsuoka, O.; Clementi, E.; Yoshimine, M. CI Study of the Water Dimer Potential Surface. *J. Chem. Phys.* **1976**, *64*, 1351–1361.

(58) Lie, G.; Clementi, E. Molecular-Dynamics Simulation of Liquid Water with an Ab Initio Flexible Water–Water Interaction Potential. *Phys. Rev. A* **1986**, *33*, 2679.

(59) Evans, M.; Refson, K.; Swamy, K.; Lie, G.; Clementi, E. Molecular-Dynamics Simulation of Liquid Water with an Ab Initio Flexible Water–Water Interaction Potential. II. The Effect of Internal Vibrations on the Time Correlation Functions. *Phys. Rev. A* **1987**, *36*, 3935.

(60) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G.; Bhattacharya, D. Molecular Dynamics Simulations of Liquid Water Using the NCC Ab Initio Potential. *J. Phys. Chem.* **1990**, *94*, 7949–7956.

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

(62) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. Polarizable Interaction Potential for Water from Coupled Cluster Calculations. I. Analysis of Dimer Potential Energy Surface. *J. Chem. Phys.* **2008**, *128*, 094313.

(63) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. Polarizable Interaction Potential for Water from Coupled Cluster Calculations. II. Applications to dimer spectra, virial coefficients, and simulations of liquid water. *J. Chem. Phys.* **2008**, *128*, 094314.

(64) Huang, X.; Braams, B. J.; Bowman, J. M. Ab Initio Potential Energy and Dipole Moment Surfaces of $(H_2O)_2$. *J. Phys. Chem. A* **2006**, *110*, 445–451.

(65) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Flexible, Ab Initio Potential, and Dipole Moment Surfaces for Water. I. Tests and Applications for Clusters up to the 22-mer. *J. Chem. Phys.* **2011**, *134*, 094509.

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

(67) Wang, Y.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Full-Dimensional, Ab Initio Potential Energy and Dipole Moment Surfaces for Water. *J. Chem. Phys.* **2009**, *131*, 054511.

(68) Medders, G. R.; Babin, V.; Paesani, F. A Critical Assessment of Two-Body and Three-Body Interactions in Water. *J. Chem. Theory Comput.* **2013**, *9*, 1103–1114.

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

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

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

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

(73) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. 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.* **2022**, *13*, 5068–5074.

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

(75) Rezac, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038–5071.

(76) Paesani, F. Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions. *Acc. Chem. Res.* **2016**, *49*, 1844–1851.

(77) Richardson, J. O.; Pérez, C.; Lobsiger, S.; Reid, A. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Wales, D. J.; Pate, B. H.; Althorpe, S. C. Concerted Hydrogen-Bond Breaking by Quantum Tunneling in the Water Hexamer Prism. *Science* **2016**, *351*, 1310–1313.

(78) Cole, W. T.; Farrell, J. D.; Wales, D. J.; Saykally, R. J. Structure and Torsional Dynamics of the Water Octamer from THz Laser Spectroscopy Near 215 μm. *Science* **2016**, *352*, 1194–1197.

(79) Mallory, J. D.; Mandelshtam, V. A. Diffusion Monte Carlo Studies of MB-pol $(H_2O)_{2-6}$ and $(D_2O)_{2-6}$ clusters: Structures and Binding Energies. *J. Chem. Phys.* **2016**, *145*, 064308.

(80) Videla, P. E.; Rossky, P. J.; Laria, D. Communication: Isotopic Effects on Tunneling Motions in the Water Trimer. *J. Chem. Phys.* **2016**, *144*, 061101.

(81) Brown, S. E.; Götz, A. W.; Cheng, X.; Steele, R. P.; Mandelshtam, V. A.; Paesani, F. Monitoring Water Clusters "melt" through Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 7082–7088.

(82) Vaillant, C. L.; Cvitaš, M. T. Rotation-Tunneling Spectrum of the Water Dimer from Instanton Theory. *Phys. Chem. Chem. Phys.* **2018**, *20*, 26809–26813.

(83) Vaillant, C.; Wales, D.; Althorpe, S. Tunneling Splittings from Path-Integral Molecular Dynamics Using a Langevin Thermostat. *J. Chem. Phys.* **2018**, *148*, 234102.

(84) Schmidt, M.; Roy, P.-N. Path Integral Molecular Dynamic Simulation of Flexible Molecular Systems in Their Ground State: Application to the Water Dimer. *J. Chem. Phys.* **2018**, *148*, 124116.

(85) Bishop, K. P.; Roy, P.-N. Quantum Mechanical Free Energy Profiles with Post-Quantization Restraints: Binding Free Energy of the Water Dimer Over a Broad Range of Temperatures. *J. Chem. Phys.* **2018**, *148*, 102303.

(86) Videla, P. E.; Rossky, P. J.; Laria, D. Isotopic Equilibria in Aqueous Clusters at Low Temperatures: Insights from the MB-pol Many-Body Potential. *J. Chem. Phys.* **2018**, *148*, 084303.

(87) Samala, N. R.; Agmon, N. Temperature Dependence of Intramolecular Vibrational Bands in Small Water Clusters. *J. Phys. Chem. B* **2019**, *123*, 9428–9442.

(88) Cvitaš, M. T.; Richardson, J. O. Quantum Tunnelling Pathways of the Water Pentamer. *Phys. Chem. Chem. Phys.* **2020**, *22*, 1035–1044.

(89) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the Accuracy of the MB-pol Many-Body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid water and Ice. *J. Chem. Phys.* **2016**, *145*, 194504.

(90) Gartner III, T. E.; Hunter, K. M.; Lambros, E.; Caruso, A.; Riera, M.; Medders, G. R.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Paesani, F. 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.* **2022**, *13*, 3652–3658.

(91) Medders, G. R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through "First-Principles" Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145–1154.

(92) Straight, S. C.; Paesani, F. Exploring Electrostatic Effects on the Hydrogen Bond Network of Liquid Water through Many-Body Molecular Dynamics. *J. Phys. Chem. B* **2016**, *120*, 8539–8546.

(93) Reddy, S. K.; Moberg, D. R.; Straight, S. C.; Paesani, F. Temperature-Dependent Vibrational Spectra and Structure of Liquid Water from Classical and Quantum Simulations With the MB-pol Potential Energy Function. *J. Chem. Phys.* **2017**, *147*, 244504.

(94) Hunter, K. M.; Shakib, F. A.; Paesani, F. Disentangling Coupling Effects in the Infrared Spectra of Liquid Water. *J. Phys. Chem. B* **2018**, *122*, 10754–10761.

(95) Sun, Z.; Zheng, L.; Chen, M.; Klein, M. L.; Paesani, F.; Wu, X. Electron-Hole Theory of the Effect of Quantum Nuclei on the X-ray Absorption Spectra of Liquid Water. *Phys. Rev. Lett.* **2018**, *121*, 137401.

(96) Gaiduk, A. P.; Pham, T. A.; Govoni, M.; Paesani, F.; Galli, G. Electron Affinity of Liquid Water. *Nat. Commun.* **2018**, *9*, 1–6.

(97) Cruzeiro, V.; Wildman, A.; Li, X.; Paesani, F. Relationship Between Hydrogen-Bonding Motifs and the $1b_1$ Splitting in the X-ray Emission Spectrum of Liquid Water. *J. Phys. Chem. Lett.* **2021**, *12*, 3996–4002.

(98) Medders, G. R.; Paesani, F. Dissecting the Molecular Structure of the Air/Water Interface from Quantum Simulations of the Sum-Frequency Generation Spectrum. *J. Am. Chem. Soc.* **2016**, *138*, 3912–3919.

(99) Moberg, D. R.; Straight, S. C.; Paesani, F. Temperature Dependence of the Air/Water Interface Revealed by Polarization Sensitive Sum-Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2018**, *122*, 4356–4365.

(100) Sun, S.; Tang, F.; Imoto, S.; Moberg, D. R.; Ohto, T.; Paesani, F.; Bonn, M.; Backus, E. H.; Nagata, Y. Orientational Distribution of Free OH Groups of Interfacial Water is Exponential. *Phys. Rev. Lett.* **2018**, *121*, 246101.

(101) Sengupta, S.; Moberg, D. R.; Paesani, F.; Tyrode, E. Neat Water–Vapor Interface: Proton Continuum and the Nonresonant Background. *J. Phys. Chem. Lett.* **2018**, *9*, 6744–6749.

(102) Muniz, M. C.; Gartner III, T. E.; Riera, M.; Knight, C.; Yue, S.; Paesani, F.; Panagiotopoulos, A. Z. Vapor-Liquid Equilibrium of Water with the MB-pol Many-Body Potential. *J. Chem. Phys.* **2021**, *154*, 211103.

(103) Pham, C. H.; Reddy, S. K.; Chen, K.; Knight, C.; Paesani, F. Many-Body Interactions in Ice. *J. Chem. Theory Comput.* **2017**, *13*, 1778–1784.

(104) Moberg, D. R.; Straight, S. C.; Knight, C.; Paesani, F. Molecular Origin of the Vibrational Structure of Ice I_h. *J. Phys. Chem. Lett.* **2017**, *8*, 2579–2583.

(105) Moberg, D. R.; Sharp, P. J.; Paesani, F. Molecular-Level Interpretation of Vibrational Spectra of Ordered Ice Phases. *J. Phys. Chem. B* **2018**, *122*, 10572–10581.

(106) Moberg, D. R.; Becker, D.; Dierking, C. W.; Zurheide, F.; Bandow, B.; Buck, U.; Hudait, A.; Molinero, V.; Paesani, F.; Zeuch, T. The End of Ice I. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 24413–24419.

(107) Bore, S. L.; Paesani, F. Quantum Phase Diagram of Water. **2023**,

(108) Fritz, M.; Fernández-Serra, M.; Soler, J. M. Optimization of an Exchange-Correlation Density Functional for Water. *J. Chem. Phys.* **2016**, *144*, 224101.

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

(110) Prodan, E.; Kohn, W. Narsightedness of Electronic Matter. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11635–11638.

(111) Bajaj, P.; Götz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion–Water Interactions through Many-Body Representations. I. Halide–Water Dimer Potential Energy Surfaces. *J. Chem. Theory Comput.* **2016**, *12*, 2698–2705.

(112) Riera, M.; Mardirossian, N.; Bajaj, P.; Götz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion–Water Interactions through Many-Body Representations. Alkali-Water Dimer Potential Energy Surfaces. *J. Chem. Phys.* **2017**, *147*, 161715.

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

(114) Robinson, V. N.; Ghosh, R.; Egan, C. K.; Riera, M.; Knight, C.; Paesani, F.; Hassanali, A. The Behavior of Methane–Water Mixtures Under Elevated Pressures from Simulations Using Many-Body Potentials. *J. Chem. Phys.* **2022**, *156*, 194504.

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

(116) Yue, S.; Riera, M.; Ghosh, R.; Panagiotopoulos, A. Z.; Paesani, F. Transferability of Data-Driven, Many-Body Models for CO_2 Simulations in the Vapor and Liquid Phases. *J. Chem. Phys.* **2022**, *156*, 104503.

(117) Bajaj, P.; Wang, X.-G.; Carrington Jr., T.; Paesani, F. Vibrational spectra of halide-water dimers: Insights on Ion Hydration from Full-Dimensional Quantum Calculations on Many-Body Potential Energy Surfaces. *J. Chem. Phys.* **2017**, *148*, 102321.

(118) Bajaj, P.; Zhuang, D.; Paesani, F. Specific Ion Effects on Hydrogen-Bond Rearrangements in the Halide–Dihydrate Complexes. *J. Phys. Chem. Lett.* **2019**, *10*, 2823–2828.

(119) Bajaj, P.; Riera, M.; Lin, J. K.; Mendoza Montijo, Y. E.; Gazca, J.; Paesani, F. Halide Ion Microhydration: Structure, Energetics, and Spectroscopy of Small Halide–Water Clusters. *J. Phys. Chem. A* **2019**, *123*, 2843–2852.

(120) Paesani, F.; Bajaj, P.; Riera, M. Chemical Accuracy in Modeling Halide Ion Hydration from Many-Body Representations. *Adv. Phys. X* **2019**, *4*, 1631212.

(121) Zhuang, D.; Riera, M.; Schenter, G. K.; Fulton, J. L.; Paesani, F. Many-Body Effects Determine the Local Hydration Structure of Cs^+ in Solution. *J. Phys. Chem. Lett.* **2019**, *10*, 406–412.

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

(123) Caruso, A.; Zhu, X.; Fulton, J. L.; Paesani, F. Accurate Modeling of Bromide and Iodide Hydration with Data-Driven Many-Body Potentials. *J. Phys. Chem. B* **2022**, *126*, 8266–8278.

(124) Zhuang, D.; Riera, M.; Zhou, R.; Deary, A.; Paesani, F. Hydration Structure of Na^+ and K^+ Ions in Solution Predicted by Data-Driven Many-Body Potentials. *J. Phys. Chem. B* **2022**, *126*, 9349–9360.

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

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

(127) Paesani, F. Water: Many-Body Potential from First Principles (From the Gas to the Liquid Phase). *Handbook of Materials Modeling: Methods: Theory and Modeling* **2020**, 635–660.

(128) Thole, B. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(129) GitHub, MB-Fit: Software Infrastructure for Data-Driven Many-Body Potential Energy Functions. `https://github.com/paesanilab/MB-Fit`.

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

(131) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **1965**, *7*, 308–313.

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

(133) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

(134) Qu, C. Private communication.

(135) Liu, K.; Brown, M.; Carter, C.; Saykally, R.; Gregory, J.; Clary, D. Characterization of a Cage Form of the Water Hexamer. *Nature* **1996**, *381*, 501–503.

(136) Nauta, K.; Miller, R. Formation of Cyclic Water Hexamer in Liquid Helium: The Smallest Piece of Ice. *Science* **2000**, *287*, 293–295.

(137) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116–11119.

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

(139) Zhai, Y.; Caruso, A.; Bore, S. L.; Luo, Z.; Paesani, F. A “Short Blanket” Dilemma for a State-of-the-Art Neural Network Potential for Water: Reproducing Experimental Properties or the Physics of the Underlying Many-Body Interactions? *J. Chem. Phys.* **2023**, *158*, 084111.

(140) Shinoda, W.; Shiga, M.; Mikami, M. Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation under Constant Stress. *Phys. Rev. B* **2004**, *69*, 134103.

(141) Simmonett, A. C.; Brooks, B. R. Analytical Hessians for Ewald and Particle Mesh Ewald Electrostatics. *J. Chem. Phys.* **2021**, *154*, 104101.

(142) Simmonett, A. C.; Brooks, B. R. A Compression Strategy for Particle Mesh Ewald Theory. *J. Chem. Phys.* **2021**, *154*, 054112.

(143) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in ’t Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS – A Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(144) MBX: A Many-Body Energy and Force Calculator for Data-Driven Many-Body Simulations. `https://paesanigroup.ucsd.edu/software/mbx.html`.

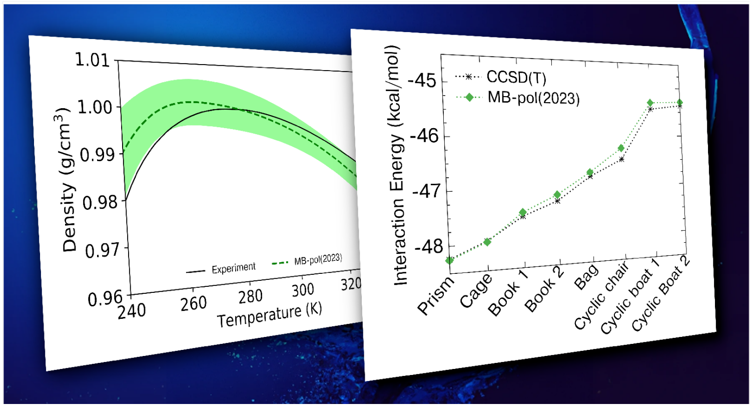
(145) Errington, J. R.; Debenedetti, P. G. Relationship between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, *409*, 318–321.

(146) Qu, C.; Yu, Q.; Conte, R.; Houston, P. L.; Nandi, A.; Bomwan, J. M. A Δ-Machine Learning Approach for Force Fields, Illustrated by a CCSD(T) 4-Body Correction to the MB-pol Water Potential. *Digital Discovery* **2022**, *1*, 658–664.

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

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

For use in table of contents only.


Density (g/cm³)
Temperature (K)
Experiment
MB-pol(2023)
Interaction Energy (kcal/mol)
CCSD(T)
MB-pol(2023)
Prism
Cage
Book 1
Book 2
Bag
Cyclic chair
Cyclic boat 1
Cyclic Boat 2