Interfacing q-AQUA with a Polarizable Force Field: The Best of Both Worlds

Chen Qu,[†,‖] Qi Yu,[*,‡,‖] Paul L. Houston,[¶] Riccardo Conte,[§] Apurba Nandi,[‡] and Joel M. Bowman[*,‡]

†*Independent researcher, Toronto, Ontario, Canada*

‡*Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, U.S.A.*

¶*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, U.S.A. and Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A*

§*Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*

‖*Contributed equally to this work*

E-mail: qyu28@emory.edu; jmbowma@emory.edu

Abstract

Polarizable force fields are pervasive in the fields of computational chemistry and biochemistry; however, their empirical or semi-empirical nature gives them both weaknesses and strengths. Here, we have developed a hybrid water potential, named q-AQUA-pol, by combining our recent *ab initio* q-AQUA potential with the TTM3-F water potential. The new potential demonstrates unprecedented accuracy ranging from gas-phase clusters, e.g., the eight low-lying isomers of the water hexamer, to the condensed phase, e.g., radial distribution functions, the self-diffusion coefficient, triplet OOO distribution, and the temperature dependence of the density. This represents a significant advancement in the field of polarizable machine learning potential and computational modeling.

Introduction

Force fields (FFs) for large molecules, many-atom or many-molecule non-reactive interactions, are pervasive in computational chemistry, biology, and materials. Among these are polarizable force fields with efficient parameterizations, as reviewed recently.[1–3] Water, in particular, has been the subject of an astonishingly large number (ca. 50) of FFs, including non-polarizable fixed-charge models and polarizable models like TTM3-F, AMOEBA, and MB-UCB.[4–6] A large set of these FFs has been critically reviewed for their performance across various properties, ranging from cluster binding energies to condensed phase thermodynamic properties.[7,8] Very recently, 15 water potentials were assessed against high-level *ab initio* results for the binding energies of water clusters by Herman and Xantheas.[9] Of these, seven are fixed-charge pairwise models, while eight are many-body potentials. Quoting from that paper "The potentials that perform most consistently across the entire $n = 2 - 25$ [cluster size] range include q-AQUA and TTM2.1-F which all give binding energies within 2.5 % of the reference ab initio values." q-AQUA[10] is a recent many-body CCSD(T)-based potential up to 4-body interactions. However, it was also noted that the polarizable TTM2.1-F potential relies on a cancellation of errors of the many-body contributions to the total binding energy. It should be noted that TTM2.1-F is semi-empirical polarizable potential[11] and a member of the TTMn family of sophisticated polarizable water potentials.[4,12]

Of particular interest here are what we term hybrid *ab initio*-polarizable water potentials. The first example may be the TTM3-F.[4] In this force field, the 2-b interaction was described by an exp-6 functional form in the OO distance, and fit to MP2 calculations. This was followed by the CC-pol potential which, in the original version, was a full dimensional fit to CCSD(T) 2-body energies for rigid monomers and correct long-range electrostatic interactions for higher body terms.[13] The HBB1 potential[14] was the first flexible potential that was a fit to thousands of CCSD(T) energies, using permutationally invariant polynomials (PIPs).[15] Extending these 2-b potentials to *ab initio* flexible 3-b interaction came soon after that, and the FFs including that interaction are known by the acronyms WHBB[16,17] and

MB-pol.[18,19] The WHBB and MB-pol potentials were "hybridized" by using the TTM3-F[4] and TTM4-F[12] potentials, respectively, to describe polarization interactions beyond the 3-b level. The details of how these potentials were interfaced to WHBB and MB-pol are different but not of relevance here. These hybrid potentials, which are approaching the first decade of use, have been applied with quantitative accuracy over the entire range of interest from isolated clusters and hydrate clathrates to the condensed phase[20,21] up to the phase diagram of water.[22] This statement underscores the major goal of such potentials, which is to be general and with transferable accuracy.

We recently reported a new water potential, q-AQUA,[10] which includes for the first time a PIP fit to CCSD(T) 4-b energies. This truncation of the many-body interactions appears to be sufficient for many challenging applications from clusters to the condensed phase (radial distribution functions and self-diffusion constants), as described in that paper and references therein. This potential performed very well in the recent assessment for clusters mentioned above and also for radial distribution functions and the self-diffusion coefficient. In order to enhance the treatment of polarization, beyond 4-b interactions, we follow the general strategy used in WHBB to interface q-AQUA with the TTM3-F potential and present the results of doing so here.

The approach we take can be considered as Δ-Machine Learning.[23–29] We recently proposed and demonstrated this approach to a many-body force field.[30] The general expression for this approach is

$$V_{\Delta\text{-ML+MB-FF}} = V_{\text{MB-FF}} + \sum_{i>j}^{N} \Delta V_{2\text{-b}}(i,j) + \sum_{i>j>k}^{N} \Delta V_{3\text{-b}}(i,j,k) + \sum_{i>j>k>l}^{N} \Delta V_{4\text{-b}}(i,j,k,l) + \cdots , \tag{1}$$

Ideally, and in our experience, the correction terms, $\Delta V_{n-\text{b}}$, are short-range. This will be the case provided the MB-FF does provide a quantitatively accurate description of the long-range interactions. Polarizable FFs do in principle provide this level of accuracy. As an aside, it is worth noting that this expression can be re-written in the more canonical MB

form seen in papers describing the MB-pol potential of Paesani and co-workers.[31] Namely,

$$E_N(1,\cdots,N) = \sum_{i}^{N} \epsilon_{\text{1-b}}(i) + \sum_{i>j}^{N} \epsilon_{\text{2-b}}(i,j) + \sum_{i>j>k}^{N} \epsilon_{\text{3-b}}(i,j,k) + E_{POL}, \tag{2}$$

where the 1-b term is the monomer potential and the 2-b and 3-b terms are the corresponding "short-range" 2 and 3-b interactions. E_{POL} is the TTM4-F potential (without monomer terms) which describes the full N-monomer polarization interactions using standard long-range expressions.[12] The terminology "short-range" is equivalent to the difference potentials $\Delta V_{\text{2-b}}$ and $\Delta V_{\text{3-b}}$, and E_{POL} is the same as $V_{\text{MB-FF}}$, which for MB-pol is TTM4-F. Note that the CCSD(T) corrections in MB-pol potential extend to 3-b terms.

Here we use the extensive datasets of CCSD(T) energies for 2, 3 and 4-b interactions that we recently reported and used to develop the q-AQUA potential,[10] which we remind the reader is truncated at the 4-b level. Thus, q-AQUA is a fully *ab initio* MB water potential. These datasets are used to fit the difference potentials $\Delta V_{n-\text{b}}$, $n = 2$, 3, 4, where we use the well-known TTM3-F as the polarizable FF for water.[4] Details of these fits and fast gradient evaluation are given below. Standard tests of the new potential, denoted q-AQUA-pol, are also provided. These are diffusion Monte Carlo calculations of the dissociation energies of the water dimer and trimer, the energies of the hexamer isomers, the water radial distribution functions, the temperature dependence of the density at 1 atm, the tetrahedral order parameter and the triplet OOO angular distribution. The new potential is shown to be slight improvement over the already highly accurate q-AQUA potential for the hexamer energies. Significant improvements are seen for these energies as well as for the temperature dependence of the density compared to the MB-pol potential.

Methods and Computational Details

Electronic Energies and Fits

The data set to construct the $\Delta V_{2\text{-b}}$ PES consists of 71 892 energies at widely dispersed geometries of the water dimer, taken from the 2-b data in q-AQUA.[10] For each dimer structure, the differences between CCSD(T)/CBS 2-b energies and TTM3-F 2-b were computed and used as the target of the fit. We used permutationally invariant polynomials (PIPs) of Morse variables ($\exp(-r_{ij}/\lambda)$, where r_{ij} is the internuclear distance and λ is a range parameter which is 3 bohr in this fit) as the fitting basis. We used the polynomial basis with 2211 permutational symmetry and a maximum polynomial order of 6. The notation "2211" indicates that permutational invariance of the two H atoms in each monomer. The basis was purified and therefore approaches zero asymptotically.[32] the RMS fitting error is 9.4 cm^{-1}. Because the correction is short-ranged, we applied a switching function when the OO distance in the dimer is between [6.5, 7.0] Å to make the correction decay to zero; thus the 2-b interaction beyond 7 Å is completely from the TTM3-F force field.

Two separate databases were used to construct the $\Delta V_{3\text{-b}}$ correction potential. The first database includes 43 214 reference energies of water trimer structures with maximum OO distance in the range of [2.0, 7.5] Å. The second database has 25 191 reference energies with maximum OO distance in the range of [5.0, 9.0] Å. These trimer structures were selected from the 3-b database in Ref. 10. For each trimer structure, the $\Delta V_{3-\text{b}}$ correction energy was computed as the difference between BSSE-corrected CCSD(T)-F12a/aVTZ and TTM3-F 3-b energies. The two data sets were fit separately using 4th-order 222111-symmetry PIPs which are functions of Morse variables with a range parameter of 2.5 bohr. Further symmetrization was done, as described in detail in Ref. 10, so that the fit is also invariant with respect to all permutations of the three monomers. The fitting RMS error is 9 cm^{-1} for the first data set and 2 cm^{-1} for the second one. The two fits were smoothly connected through a switching function when the maximum OO distance is in [5.0, 5.5] Å. Finally, for the consideration

of both computationally efficiency and accuracy, we applied an additional smooth switching function when the maximum OO distance is within [6.0, 7.0] Å for the $\Delta V_{3-\mathrm{b}}$ correction potential to make it decay to 0 smoothly in the long-range region.

The 4-b correction potential, $\Delta V_{4-\mathrm{b}}$, is a fit to a dataset of 3692 tetramer structures and corresponding reference energies computed at the CCSD(T)-F12/haTZ (aug-cc-pVTZ basis for O atoms and cc-pVTZ for H atoms) level of theory. This is the same data set we used to develop the 4-b component of the q-AQUA potential,[10] except that in this work we fit to the difference between CCSD(T) and TTM3-F 4-b energies rather than to direct 4-b energies. The fitting basis is the same as described in q-AQUA 4-b and the Δ-ML 4-b to MB-pol;[33] briefly, it consists of 200 "super PIPs" of Morse variables; these super PIPS are formed from sums of regular PIPs so as to ensure invariance with respect to permutation of monomers.[32] The fitting RMS error for the whole data set is 6.3 cm^{-1}. Similar to the $\Delta V_{2-\mathrm{b}}$ and $\Delta V_{3-\mathrm{b}}$, a switching function is applied to make $\Delta V_{4-\mathrm{b}}$ decay to 0 smoothly when the maximum OO distance in a tetramer is within [6.0, 6.5] Å.

Diffusion Monte Carlo Calculations

The diffusion Monte Carlo (DMC) method is based on the similarity between the diffusion equation and the imaginary-time Schrödinger equation with an energy shift E_{ref}

$$\frac{\partial \psi(\boldsymbol{x},\tau)}{\partial \tau} = \sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 \psi(\boldsymbol{x},\tau) - \left[V(\boldsymbol{x}) - E_{\mathrm{ref}}\right] \psi(\boldsymbol{x},\tau) \tag{3}$$

The reference energy E_{ref} in the above equation is used to stabilize the diffusion system in its ground state and thus is the estimator of the zero-point energy.[34] We employed the unbiased, unconstrained implementation of DMC,[35] in which the DMC calculation starts from an initial guess of the ground-state wave function, represented by a population of $N(0)$ equally weighted Gaussian random walkers. These walkers then diffuse randomly in imaginary time according to a Gaussian distribution. The population is controlled by a

birth-death processes, described elsewhere.[35] To maintain the number of random walkers at about the initial value $N(0)$, E_{ref} is adjusted at the end of each time step according to

$$E_{\text{ref}}(\tau) = \langle V(\tau) \rangle - \alpha \frac{N(\tau) - N(0)}{N(0)} \tag{4}$$

where $N(\tau)$ is the number of walkers at the time step τ, α is a feedback parameter, typically around 0.1, and $\langle V(\tau) \rangle$ represents the average potential energy of all of the walkers at that step. Finally the average of the E_{ref} provides an estimate of the ZPE.

In this study, the DMC calculations were carried out for water monomer, dimer, trimer, and three isomers of the water hexamer (prism, book, and cage). The ZPEs of the dimer and trimer are subsequently used to compute the dissociation energies of the dimer and trimer, while the calculations for the hexamer are mainly exploratory. In these DMC calculations, the imaginary time step $\Delta\tau = 5$ a.u. and $\alpha = 0.1$ are used. For monomer, dimer and trimer, five DMC calculations were performed for each system. In each DMC calculation, the number of walkers is 50 000, and these walkers are equilibrated for 5000 time steps followed by 50 000 propagation steps. The statistical uncertainty is estimated as the standard deviation of the 5 DMC runs for the same system.

Classical and path integral-based MD simulations

The q-AQUA-pol water potential was interfaced with the i-PI software[36] to enable classical, path integral and the ring polymer MD simulations (MD, PIMD, and TRPMD) to obtain the theomodynamical and dynamical properties of liquid water at various temperatures and atmosphere pressure ($P = 1$ atm). The system consists of 256 water molecules in a periodic cubic box. All simulations were performed in the isothermal-isobaric ensemble (constant NPT) which was controlled through a isotropic barostat as described in Ref. 37. For classical MD simulations, an optimal-sampling GLE thermostat[38] was applied. For PIMD and TRPMD simulations, the PIGLET thermostat[39] was employed and 8 beads were

used. The trajectories were propagated for 2 ns with time step of 0.5 fs in classical MD simulations. In PIMD simulations, we applied 8 beads and propagated the trajectories for 1 ns with time step of 0.5 fs. The first 200 ps of the trajectories was used for equilibrium and the reminder used for analysis. To obtain dynamical property, such as self-diffusion constants, additional NVE trajectories of 200 ps were propagated for classical MD and TRPMD using the calculated densities from NPT simulations. Finally, a radial cutoff distance of 9.0 Å was used in the q-AQUA-pol potential for the components of TTM3-F force field. The long-range electrostatic interactions were treated using an extension of Wolf's method which was shown to efficiently and accurately treat the electrostatic interactions for liquid water the aqueous solutions.[40]

Results

$\Delta V_{\text{2-b}}$, $\Delta V_{\text{3-b}}$ and $\Delta V_{\text{4-b}}$ PIP Fits

The first results focus on the large 2-b interaction, which is the interaction of highest magnitude. Fig. 1 is a scatter plot of all the $\Delta V_{\text{2-b}}$ data vs the OO distance in panel A. As seen there are large positive and negative differences in the region of the water dimer minimum, i.e., around 2.7 Å as surmised from panel C. Also $\Delta V_{\text{2-b}}$ goes to zero at OO distances greater than 6.0 Å. Note that a direct fit to the CCSD(T) 2-b interaction is roughly 0.4 kcal/mol at 6.5 Å.[10] Thus, as expected, $\Delta V_{\text{2-b}}$ is shorter range than the 2-b interaction. From the correlation plot shown in panel B it is clear that the PIP fit to the data is highly precise over the entire large range; From panels C and D it is clear that TTM3-F becomes rapidly inaccurate as the OO distance decreases for both geometries shown and thus the correction to it is large. Also it is clear that the corrected 2-b is in excellent agreement with direct CCSD(T)/CBS energies.

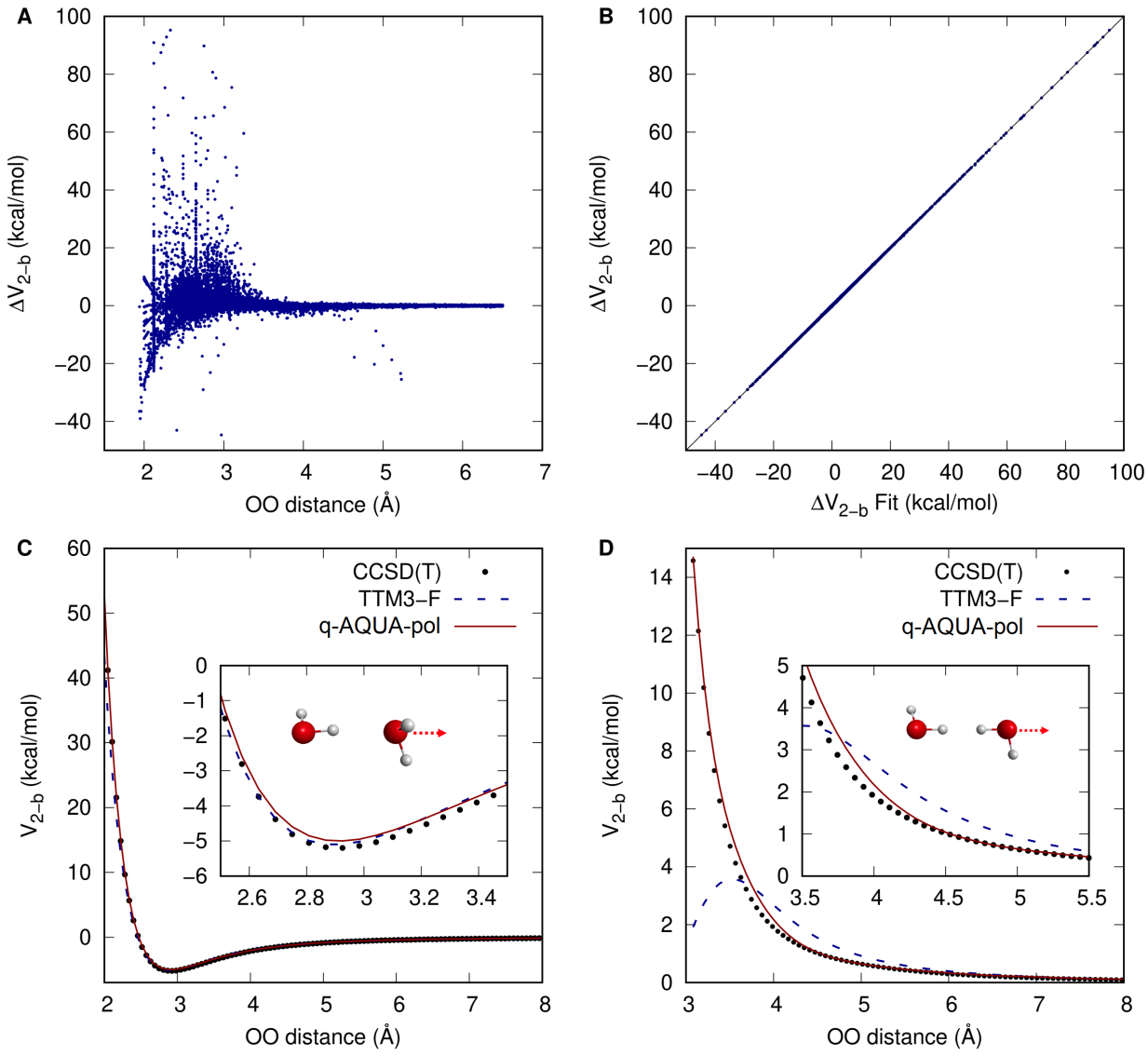

Figure 1: (A) Distribution of $\Delta V_{2-\mathrm{b}}$ correction energies ($V_{2-\mathrm{b}}^{\mathrm{CCSD(T)}} - V_{2-\mathrm{b}}^{\mathrm{TTM3\text{-}F}}$) and OO distance. (B) Correlation between fitted $\Delta V_{2-\mathrm{b}}$ correction energies and reference data. (C) Comparison of the q-AQUA-pol 2-b potential, the TTM3-F 2-b potential, and the direct CCSD(T)/CBS 2-b energies for an attractive cut. (D) Comparison of the q-AQUA-pol 2-b potential, the TTM3-F 2-b potential, and the direct CCSD(T)/CBS 2-b energies for a repulsive cut.

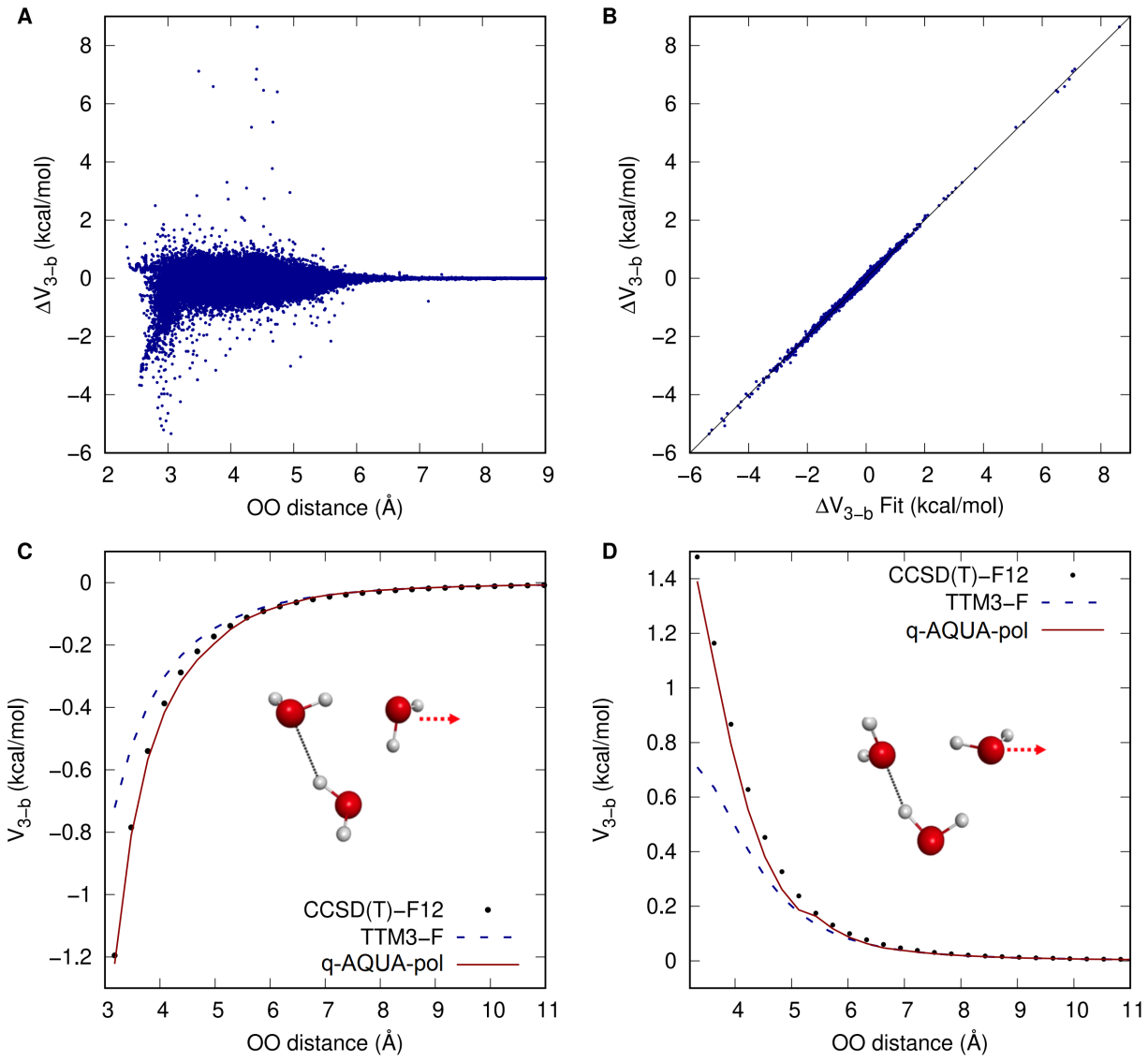

Figure 2: (A) Distribution of $\Delta V_{3-\mathrm{b}}$ correction energies ($V_{3-\mathrm{b}}^{\mathrm{CCSD(T)}} - V_{3-\mathrm{b}}^{\mathrm{TTM3\text{-}F}}$) and OO distance. (B) Correlation between fitted $\Delta V_{3-\mathrm{b}}$ correction energies and reference data. (C) Comparison of the q-AQUA-pol 3-b potential, the TTM3-F 3-b potential, and the direct CCSD(T) 3-b energies for an attractive cut. (D) Comparison of the q-AQUA-pol 3-b potential, the TTM3-F 3-b potential, and the direct CCSD(T) 3-b energies for a repulsive cut. All the CCSD(T) energies are calculated at CCSD(T)-F12/aVTZ level of theory

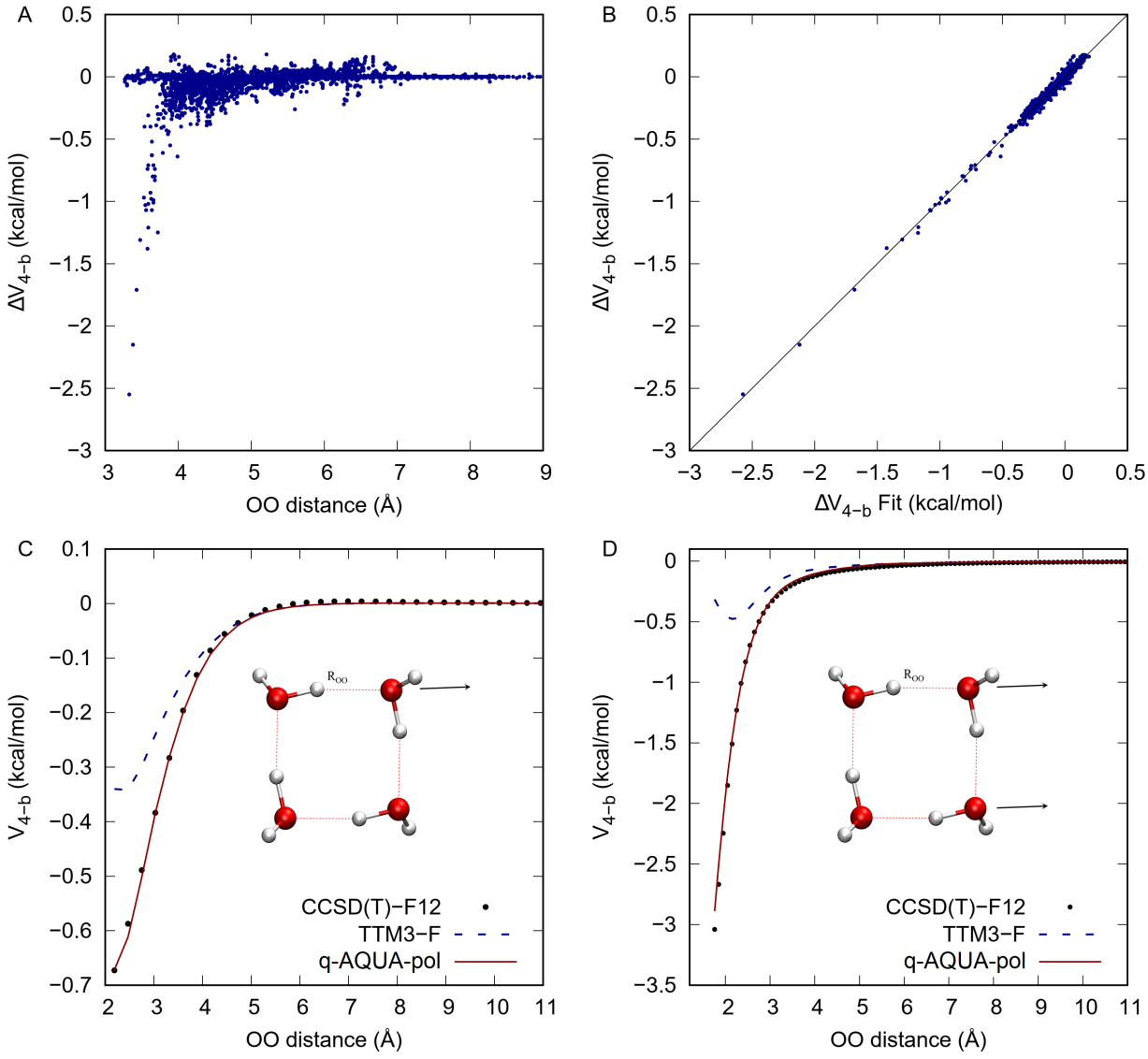

Figure 3: (A) Distribution of $\Delta V_{\text{4-b}}$ correction energies ($V_{\text{4-b}}^{\text{CCSD(T)}} - V_{\text{4-b}}^{\text{TTM3-F}}$) as a function of the maximum OO distance in a tetramer. (B) Correlation between fitted $\Delta V_{\text{4-b}}$ energies and reference data. (C) Comparison of the q-AQUA-pol 4-b potential, the TTM3-F 4-b potential, and the direct CCSD(T)-F12 4-b energies for a monomer-trimer cut. (D) Comparison of the q-AQUA-pol 4-b potential, the TTM3-F 4-b potential, and the direct CCSD(T) 4-b energies for a dimer-dimer cut.

As noted already, the 3-b and 4-b corrections are done using fits to $\Delta V_{3-\mathrm{b}}$ and $\Delta V_{4-\mathrm{b}}$. Figure 2 panel A shows a scatter plot of the $\Delta V_{3-\mathrm{b}}$ data (not a fit) vs the maximum OO distance in a trimer. As seen, many of these are less than 1 kcal/mol; however, a large number are several kcal/mol and as large as 10 kcal/mol. These are clearly significant errors in the TTM3-F 3-b interaction. Panel B shows the correlation plot of the precise $\Delta V_{3\text{-b}}$ PIP fit (described in detail above). Panels C and D show two 1-d cuts when one of the monomer is moved away from the dimer. As one can see, the ΔV_{3-b} energy can be very large at short range, indicating the inaccuracy of the TTM3-F force field in this range; this is also clearly shown in the two potential cuts. The fitting brings the 3-b interaction into much better agreement with CCSD(T) energies.

Similarly, Figure 3 panel A shows a scatter plot of the $\Delta V_{4-\mathrm{b}}$ data (not a fit) vs the maximum OO distance in a tetramer and again significant errors in the TTM3-F 4-b is seen. Panel B shows the correlation plot of the $\Delta V_{4\text{-b}}$ PIP fit. Panels C and D show two 1-d cuts when one of the monomer and a dimer are moved away from the remaining water molecules; again the $\Delta V_{4\text{-b}}$ correction brings the TTM3-F into better agreement with CCSD(T)-F12 energies, especially at the short range.

Tests for water clusters

Diffusion Monte Carlo calculations of D_0 for the dimer and trimer

Unconstrained diffusion Monte Carlo (DMC) calculations were preformed for the zero-point energies of the water dimer and trimer using the q-AQUA-pol. Recall that these are rigorous "exact" quantum calculations and give the exact dissociation energy, D_0, when combined with the electronic dissociation energy (D_e) and exact ZPE of the water monomer. Table 1 shows the electronic dissociation energy (D_e), ZPEs of the reactants and products, and ZPE-corrected dissociation energy (D_0) for the following dissociations: $(H_2O)_2 \rightarrow 2H_2O$, $(H_2O)_3 \rightarrow 2H_2O + H_2O$, and $(H_2O)_3 \rightarrow 3H_2O$. As seen, agreement with experiment is excellent. Agreement is also excellent with DMC calculations of D_0 for the dimer using the HBB2[17]

and MB-pol[41] potentials, both of which potentials predict 1101 cm^{-1}. These values are also in excellent agreement with D_0 reported for the dimer[42] using the recent flexible CCpol-8sf potential.[43] For the water trimer the present results are in excellent agreement with previous DMC calculations using WHBB,[44] 2724 cm^{-1}, and MB-pol,[41] 2693 cm^{-1}. Also there is very good agreement for dissociation to three monomers, namely 3854 cm^{-1} for WHBB and 3794 cm^{-1} for MB-pol.

Table 1: Electronic dissociation energy (D_e), ZPEs of the reactants and products, and ZPE-corrected dissociation energy (D_0) for water dimer and trimer. Energies are in cm^{-1}

.

Dissociation	D_e	ZPE (react.)	ZPE (prod.)	D_0	D_0 (expt.)
$(H_2O)_2 \rightarrow 2H_2O$	1739	9913±1	9272±1	1098±2	1105± 10 [a]
$(H_2O)_3 \rightarrow 2H_2O + H_2O$	3764	15616±2	14549±2	2697±4	2650±150 [b]
$(H_2O)_3 \rightarrow 3H_2O$	5503	15616±2	13908±2	3795±4	NA

[a] From Ref. 45
[b] From Ref. 46

Finally, note that we performed exploratory DMC calculations using 20 000 walkers and 25 000 steps for the hexamer using the new PES and did not find "holes", i.e., regions of spuriously large negative energies. The ZPE of the cage isomer is roughly 26 cm^{-1} lower than that of the prism isomer (both ZPEs are referenced to the electronic energy of the prism isomer), in semi-quantitative agreement with the results using q-AQUA,[10] though the uncertainties in this work are much larger due to the smaller number of walkers and steps.

Energy analysis and harmonic frequencies of the water hexamer isomers

The isomers of the water hexamer play a major role in both experimental and theoretical studies of water clusters. The lowest energy isomers are non-cyclic.[47,48] The energies of eight isomers have become a standard test of the fidelity of a water FF.[49] A detailed analysis of the electronic energies of the eight isomers is given Table 2. The electronic dissociation energies, D_e, for q-AQUA-pol are in better agreement the benchmark CCSD(T) result than MB-pol, which is in good agreement with the benchmarks for D_e and the many-body interactions.

This analysis is presented graphically in Fig. 4 which also shows results from TTM3-F. The results shown there and in the Table 2 are even more accurate those from our q-AQUA potential, which is truncated at the 4-b level[10] and also more accurate than the MB-pol potential.[31]

Figure 5 shows the deviations in the harmonic frequencies of various water potential on 4 hexamer isomers, compared to the benchmark calculations.[50] The TTM3-F model significantly underestimates the frequencies of OH stretches, as seen in the figure. Both MB-pol and q-AQUA are quite accurate, and can achieve a mean absolute deviation of ~ 10 cm^{-1}. The new q-AQUA-pol achieves an unprecedented accuracy on the harmonic frequencies of hexamers, with a mean absolute deviation of $\sim$5 cm^{-1}, and a maximum deviation of only 15 cm^{-1}.

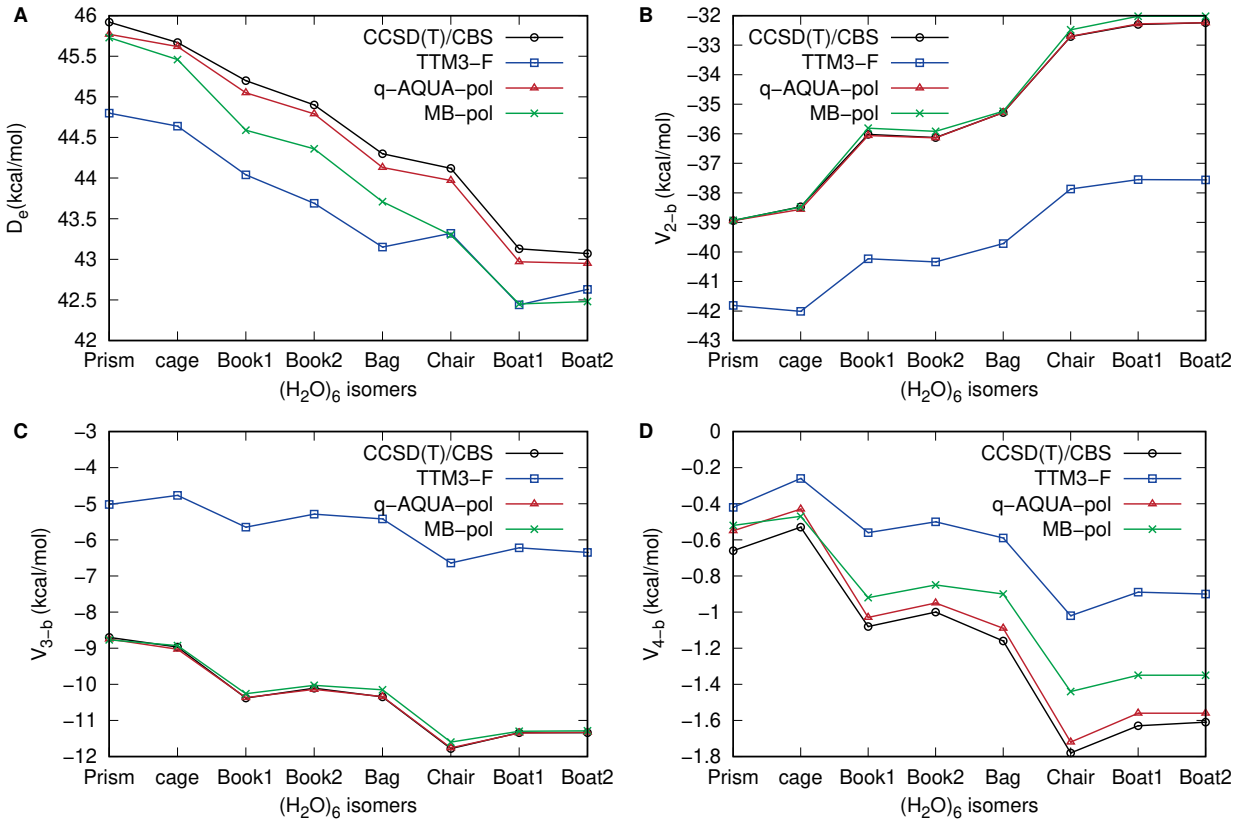

Figure 4: Binding energies (A), 2-body energies (B), 3-body energies (C) and 4-body energies (D) for water hexamer isomers from TTM3-F, q-AQUA-pol, MB-pol and benchmark CCSD(T) calculations (taken from refs. 51 and 31).

Table 2: 2-b, 3-b, 4-b and total dissociation energies (kcal/mol) for water hexamer isomers. The geometries are taken from Ref. 51 without further optimization.

	De			2-b energy		
Isomer	CCSD(T)[a]	q-AQUA-pol	MB-pol	CCSD(T)[b]	q-AQUA-pol	MB-pol
Prism	45.92	45.77	45.73	-38.94	-38.94	-38.93
Cage	45.67	45.62	45.46	-38.47	-38.56	-38.48
Book 1	45.20	45.05	44.59	-36.02	-36.06	-35.81
Book 2	44.90	44.79	44.36	-36.13	-36.14	-35.92
Bag	44.30	44.13	43.71	-35.28	-35.28	-35.24
Chair	44.12	43.97	43.30	-32.71	-32.70	-32.48
Boat 1	43.13	42.97	42.45	-32.30	-32.28	-32.02
Boat 2	43.07	42.95	42.48	-32.24	-32.24	-32.02
MAE	/	0.13	0.53	/	0.02	0.15
	3-b energy			4-b energy		
Isomer	CCSD(T)[a]	q-AQUA-pol	MB-pol	CCSD(T)[b]	q-AQUA-pol	MB-pol
Prism	-8.70	-8.75	-8.77	-0.66	-0.55	-0.52
Cage	-8.97	-9.03	-8.93	-0.53	-0.43	-0.47
Book 1	-10.38	-10.37	-10.26	-1.08	-1.03	-0.92
Book 2	-10.11	-10.14	-10.03	-1.00	-0.95	-0.85
Bag	-10.35	-10.34	-10.15	-1.16	-1.09	-0.90
Chair	-11.78	-11.76	-11.60	-1.78	-1.72	-1.44
Boat 1	-11.34	-11.35	-11.30	-1.63	-1.56	-1.35
Boat 2	-11.34	-11.33	-11.29	-1.61	-1.56	-1.35
MAE	/	0.03	0.10	/	0.07	0.21
	higher-body (> 4-b) energy					
Isomer	CCSD(T)[a]	q-AQUA-pol	MB-pol			
Prism	0.06	0.03	0.05			
Cage	0.01	0.01	0.03			
Book 1	-0.04	-0.02	-0.03			
Book 2	-0.02	-0.01	-0.01			
Bag	-0.01	-0.02	-0.02			
Chair	-0.20	-0.12	-0.11			
Boat 1	-0.17	-0.10	-0.10			
Boat 2	-0.17	-0.10	-0.10			
MAE	/	0.04	0.04			

[a] CCSD(T)/CBS data from Ref. 51
[b] CCSD(T)/CBS data from Ref. 31

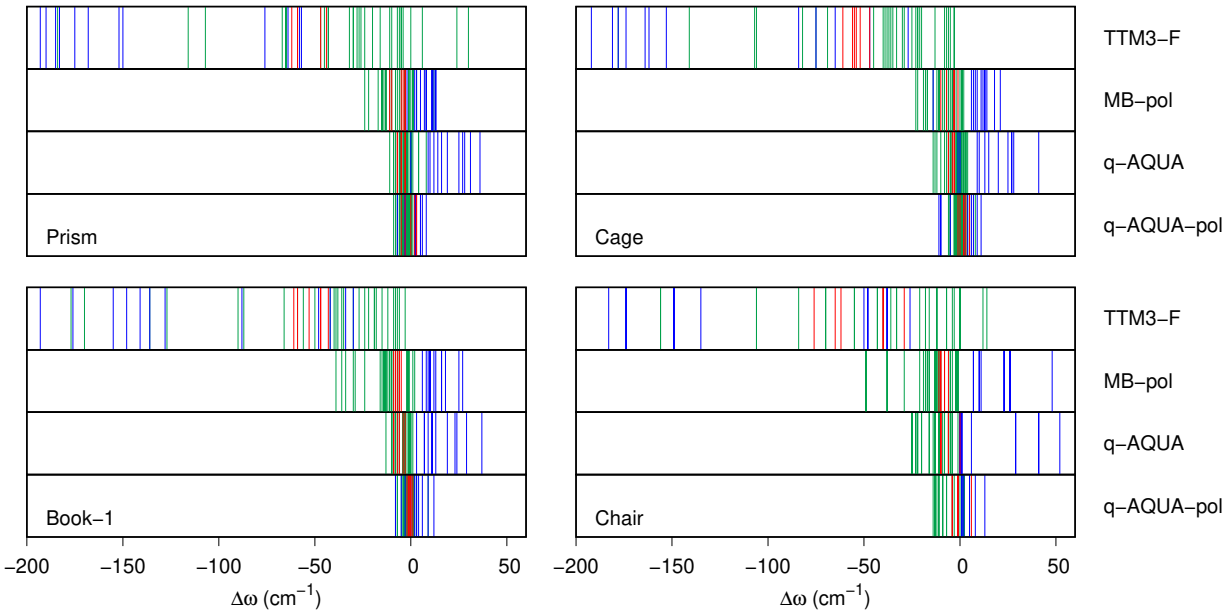

Figure 5: Deviations from benchmark harmonic frequencies of water hexamers,[50] calculated with different water potentials. Each normal mode, indicated by a stick line, is colored based on the type of the vibration (*blue*: OH stretches; *red*: bending; *green*: inter-molecular vibrations).

Condensed phase properties

Properties at T = 298 K, P = 1 atm

We performed both classical and path integral simulations of liquid water under ambient conditions for its thermodynamic properties. The simulations were carried out in constant temperature-constant pressure (NPT) ensembles. The calculated liquid water density, tetrahedral order parameter, and self-diffusion coefficient are given in Table 3. At 298 K and 1 atm, the calculated water density from classical MD is 1.002 g/cm^3 which is slightly higher than the experimental data. After nuclear quantum effects are considered in PIMD simulations, the density decreases to 0.997 g/cm^{-1} which reaches excellent agreement with experiment. This indicates the importance of zero-point energy and quantum delocalization in affecting the water structures and dimensions.

Table 3: Properties of liquid water (tetrahedral order parameter q, density ρ, and self-diffusion coefficient D) at 298 K from classical and PIMD or TRPMD simulations with q-AQUA-pol potential

	Classical	PIMD	Expt.
q	0.684	0.596	0.576[a]
ρ (g/cm^3)	1.002 ± 0.002	0.997 ± 0.002	0.997[b]
D (Å^2/ps)	0.185 ± 0.004	0.233[d] ± 0.027	0.230[c]

[a] from Ref. 52, [b] from Ref. 53, [c] from Ref. 54 and 55. [d] TRPMD

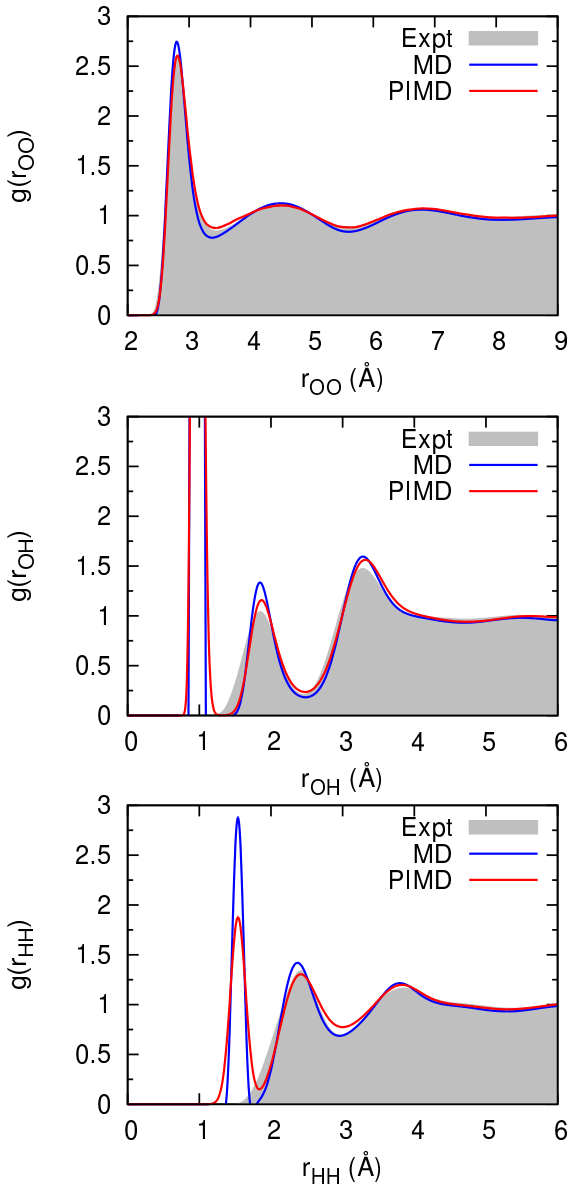

Figure 6: OO, OH, and HH radial distribution function for liquid water at 298 K from classical MD and PIMD NPT simulations. The experimental data are taken from Ref. 56,57.

Figure 6 shows the OO, OH, and HH radial distribution functions (RDF) from both classical MD and PIMD NPT simulations. As seen, the RDFs from q-AQUA-pol are in good agreement with experimental measurement. For the OO RDF from classical MD simulation, the first and second peak appear at 2.80 and 4.47 Å, respectively, while the experimental peaks are at around 2.81 and 4.50 Å. The classical MD simulations with q-AQUA-pol potential correctly describe both H-bonded neighbors and non–H-bonded water molecules that occupy the interstitial space between H-bonded neighbors. Such agreement is further improved by the PIMD simulations, where the first and second peaks locate at 2.82 and 4.50 Å. In addition to the peak positions, other regions of the OO RDF also show better agreement with experiment, especially the fact that the distribution becomes more delocalized in the region of [3,4] Å. This is consistent with our previous finding in the NVT simulations with original q-AQUA potential.[10] Similar behavior is seen in the OH and HH RDFs which further verifies the accuracy of q-AQUA-pol potential and the importance of nuclear quantum effects in describing the covalent OH bonds in liquid water, as well as the H-bonded and non-H-bonded solvation shells.

The oxygen-oxygen-oxygen triplet angular distribution $P_{\mathrm{OOO}}(\theta)$ and tetrahedral order parameter[58,59] are convenient properties of detecting the tetrahedral orientational ordering of the liquid water induced by the H-bonded network. To compute $P_{\mathrm{OOO}}(\theta)$, three oxygen atoms are considered as a triplet combination where two of the oxygen atoms are within a cut-off distance from the third oxygen. The cut-off distance was selected to yield an average oxygen-oxygen coordination number of around 4. The cut-off distances were chosen as 3.27 and 3.45 Å for classical MD and PIMD simulations respectatively. The larger cut-off distance in PIMD simulation indicates a less compact first solvation shell and a lower liquid density after nuclear quantum effects are considered. This agrees with the findings in radial distribution functions above. The tetrahedral order parameter, q, was calculated as:[59]

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos(\theta_{ijk}) + \frac{1}{3}\right)^2 \tag{5}$$

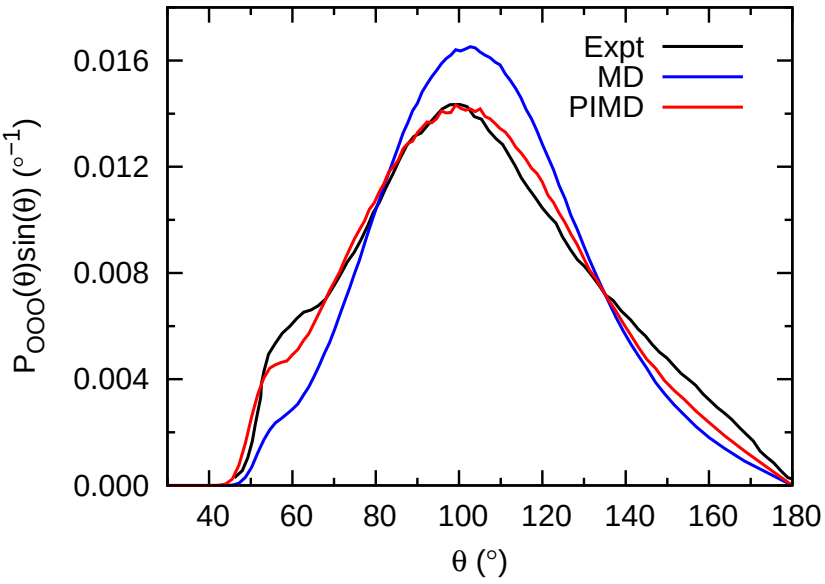

Figure 7: The oxygen-oxygen-oxygen triplet angular distribution functions of liquid water at 298 K from classical MD and PIMD simulations. The experimental data are taken from Ref. 52. The triplet angular distribution functions shown here were normalized to $\int_0^\pi \mathrm{P_{OOO}}(\theta)\sin(\theta)d\theta$. The calculated tetrahedral order parameters at T = 298K and P = 1 atm are 0.684 for classical MD, 0.596 for PIMD, and the experimental value is 0.576.[52]

As seen from Eq. 5, the tetrahedral order parameter, q, varies between 0 and 1. When the system has perfect tetrahedral structure, q is exactly equal to 1. From classical MD simulation at 298 K and 1 atm, we obtained $q = 0.684$, shown in Table 3, which is higher than the experimental value of 0.576.[52] This is confirmed in the distribution plot of $P_{\mathrm{OOO}}(\theta)$ in Figure 7, where the classical MD simulation presents a peak at 102.7° and a narrower distribution than that from the experiment. This indicates that the classical MD simulation with q-AQUA-pol potential overestimates the tetrahedral structural of the liquid water with overly strong H-bond network. However, PIMD simulation generates tetrahedral order parameter as $q = 0.596$ which is close to the experimental value. As for the distribution of $P_{\mathrm{OOO}}(\theta)$, results from the PIMD simulation with q-AQUA-pol potential are in excellent agreement with experiment in terms of both peak position, width, and intensity. Again, these observations confirm the effectiveness of combining PIMD and q-AQUA-pol potential in describing the structural properties of liquid water.

The self-diffusion coefficient, D, was also calculated to investigate the dynamical property of liquid water at ambient condition. The results are listed in Table 3. With classical MD, D is predicted as 0.185 $\pm$ 0.004 $\text{Å}^2/ps$ which is lower than the experimental value of 0.230 $\text{Å}^2/ps$. With PIMD, a higher diffusion coefficient (0.233 $\pm$ 0.027) is observed and agrees better with experiment. The significant change of diffusion coefficients in PIMD simulation is due to nuclear quantum effects. As indicated in Ref. 60, there exist two competing nuclear quantum effects in liquid water, intramolecular and intermolecular quantum effects. The former weakens the covalent bond and induces a larger molecular dipole moment. The later weakens the hydrogen bond and results in a less structured H-bond network. From our work with q-AQUA-pol potential, the latter (intermolecular) effect dominates over the former one (intramolecular ZPE). Liquid water has a destabilized hydrogen-bonding network and larger diffusion coefficients. This observation also agrees with our previous NVT simulation with the original q-AQUA potential.

Properties at various temperatures

Next, we explore the properties of condensed water at various temperatures. We performed NPT classical MD simulations at P = 1 atm with temperatures from 238 K to 340 K, which spans the region from supercooled to room-temperature water liquid water. The calculated densities of liquid water are shown in Figure 8 in comparison with MB-pol data and experimental measurements. As seen, the q-AQUA-pol results agree excellently with the corresponding experimental values in high temperature region. When the temperature decreases, the exists differences between q-AQUA-pol and experiment. However, the overall trend agrees well with experiment and the estimated temperature of maximum density is between 268-278 K, closer to the experimental maximum at 277 K. Such behavior outperforms MB-Pol which has larger disagreement with experiment and lower temperature of maximum density.[31,61] The difference between q-AQUA-pol and experimental densities can be partly attributed to the neglect of nuclear quantum effects (NQE) in classical MD sim-

ulations, which has been shown to be important.[62] We conducted PIMD NPT simulations at 288 K, 298 K, and 320 K to investigate the effect of NQE on densities. As shown in the red diamond data in Figure 8, when the nuclear quantum effects are included, the predicted densities decreases and excellent agreement with experiment is achieved for all these three temperatures. It is satisfying to see that the two sets of results appear to merge at higher temperatures. The error bars shown for the q-AQUA-pol results come from the density vs time plots. These are shown in Fig. S3 of the Supporting Information, which extends to 1 ns of the trajectory. It can be reasonably inferred that liquid water densities from q-AQUA-pol can be improved with NQE included, resulting in better agreement with experiment.

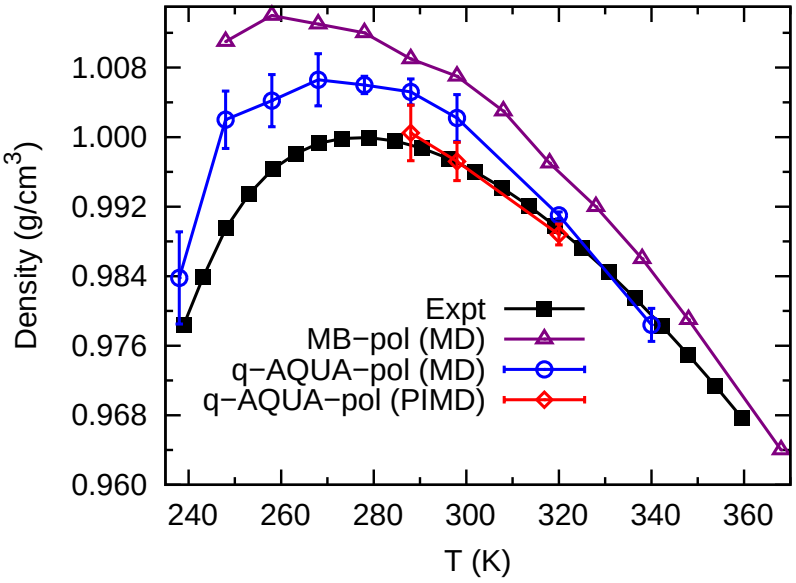

Figure 8: Temperature-dependence of the density of liquid water at 1 atm. Densities calculated from PIMD simulations with q-AQUA-pol potential are indicated in red diamond symbols. The MB-pol data are taken from Ref. 31. The experimental data are taken from Ref. 53 and 63.

The temperature dependence of the order parameter[58] is shown in Figure 9. As seen, the distribution narrows as the temperature decreases into the supercooled liquid regime. That cold water, i.e., ice, assumes a tetrahedral H-bonding motif can be traced at least to Bernal and Fowler.[64] The ridge at ≈ 0.5 is interesting; however, it is beyond the scope of this paper

to investigate this in detail. Some discussion of it can be found in a recent paper where the order parameter obtained with MB-pol was reported.[61]

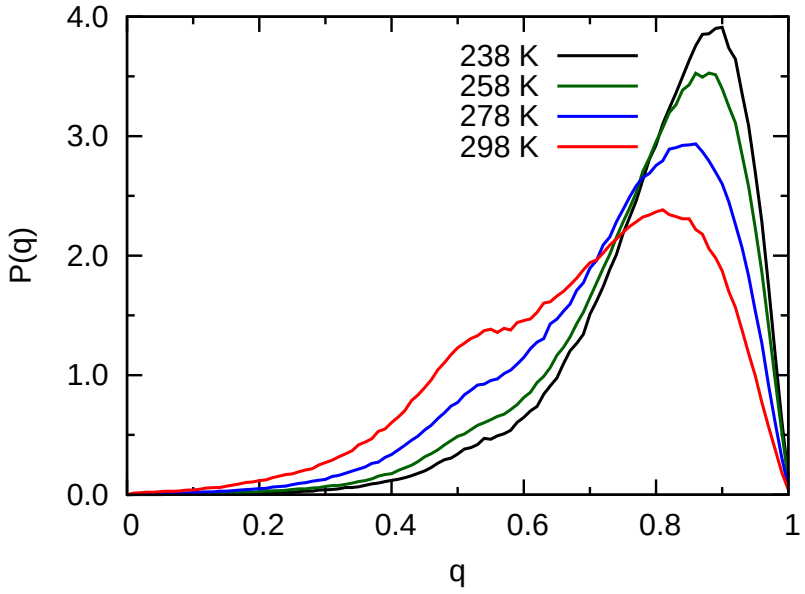

Figure 9: Probability distribution of the tetrahedral order parameter q at different temperatures from classical MD simulations.

Timing

Finally, we examine the computational cost in q-AQUA-pol for both energy and gradient calculations for a 256-water system, using one single core or multiple cores of 2.4 GHz Intel Xeon processor, where the cost for each correction term $\Delta V_{n-\mathrm{b}}$ is also listed. The 256-water structure is chosen from MD simulation of liquid water using our recently developed water potential, q-AQUA.[10] Note, the number of terms given is less than the factorial result owing to the use of the finite range switching function. For example, the total number of 3-b interaction terms in this 256 mer water system is $\binom{256}{3} = 2\ 763\ 520$, however, only 3-b interactions with maximum O-O distance smaller than 7.0 Å are considered for correction, which resulted in only 20 790 $\Delta V_{3-\mathrm{b}}$ terms. It is also straightforward to do multi-core

processing (using OpenMP) of the potential as shown in Table 4. Overall the speed up is a factor of 5.9.

It is also clear from this table that the evaluation of the 20 790 3-b interactions is the largest single contribution to the total time. Clearly if the cutoff range is smaller the number of 3-b evaluations will decrease substantially; however, at some cost in accuracy. We plan to investigate this in the future.

Table 4: The computation cost of the q-AQUA-pol potential for energy and gradient calculations of a 256 water system

		Time for energy (s)		Time for energy+gradient (s)	
Component	Number	1 core	8 core	1 core	8 core
TTM3-F	/	0.27	0.07[a]	0.27	0.07[a]
$\Delta V_{\text{2-b}}$	3816	0.11	0.02	0.36	0.06
$\Delta V_{\text{3-b}}$	20790	0.65	0.23	2.20	0.39
$\Delta V_{\text{4-b}}$	28786	0.30	0.04	1.05	0.14
Total		1.33	0.36	3.88	0.66

[a] Current TTM3-F force filed code calculates the gradients by default.

Discussion

The current state-of-the-art, machine-learned, CCSD(T) based force-fields for condensed-phase water are substantially beyond what it was even 5 or so years ago. At that time there were two machine-learned CCSD(T)-based FFs using the many-body approach, WHBB and MB-pol. In these FFs, 2 and 3-b interactions for flexible monomers were obtained using permutationally invariant polynomial regression[15] fits to thousands of *ab initio* 2 and 3-body interactions. These were rigorously tested against direct CCSD(T)/CBS results for the electronic dissociation energy (atomization energies) of isomers of the water hexamer. Such calculations[51] were basically at the limit of feasibility for hexamers. Since these FFs were trained peforce on small clusters, i.e., the dimer and trimer, the CCSD(T) calculations were feasible. It should be noted that both WHBB and MB-pol made use of polarizable FFs (TTM3 for WHBB and TTM4 for MB-pol) to account approximately for 4- and higher-

body interactions. Numerous tests of the accuracy of MB-pol by Reddy et al[31] have shown that this potential is robust for many applications. In particular for the eight hexamer isomers, a mean absolute error of 0.53 kcal/mol was reported. (Also, see Table 2.) Errors are larger using WHBB, however, both potentials obtained the correct energy ordering of the isomers, with the prism being the lowest energy isomer in electronic energy. In 2022, we reported q-AQUA which is a strict machine-learned, CCSD(T)-based potential including 4-b interactions.[10] The MAE for these hexamer isomers is 0.25 kcal/mol, which is roughly half the MB-pol MAE. As shown here, q-AQUA-pol yields even more accurate electronic dissociation energies for the benchmark water hexamer isomers than q-AQUA.

Coarser properties for the condensed phase liquid such as the OO, OH and HH radial distribution functions and self-diffusion coefficient were accurately obtained using q-AQUA (in NVT calculations) as well as MB-pol (in NPT calculations). These properties are also accurately described by q-AQUA-pol in NPT calculations. Nuclear quantum effects, described using PIMD calculations, using q-AQUA and here using q-AQUA-pol, were found to be noticeable and generally to bring theory closer to experiment. The current results using q-AQUA-pol for the temperature dependence of the density of H_2O are especially encouraging. They (both classical and PIMD) are significantly closer to experiment than those from MB-pol.

Atom-centered neural network approaches have been reported for water for a number of years, of which the "Deep Neural Network" (DNN) architecture, which is trained on snapshots of AIMD trajectories using efficient DFT approach for the potential and forces, is a prominent and versatile example.[65,66] But even a perfect fit provides a level of accuracy that is below the gold standard CCSD(T) method of electronic structure theory used in the many-body approaches just discussed. Very recent reports using transfer-learning[67] have led to CCSD(T)-transfer-learned FFs for water.[68,69] These are both atom-centered NN fits, and in both cases were trained on DFT-based samples of condensed phase water consisting of 32 or 64 monomers. Several condensed phase quantities, i.e., radial distribution functions

and diffusion constants, were reported using NVT simulations and shown to in very good agreement with experiment. So this recent work perhaps marks a convergence of the MB and atom-centered NN approaches. One graphical example of this is Fig. 10, which shows the tetrahedral order parameter at room temperature calculated from q-AQUA-pol and the recent NN-transfer-learned potential,[69] as well as from q-AQUA in NVT simlulations. The q-AQUA distributions were smoothed with a gaussian of FWHM=0.05. (The unsmoothed q-AQUA-pol distributions are given in Fig. 9) Note all distributions are normalized to one. As seen there is very good agreement, and we find this interesting and significant as the three potentials, albeit all at a CCSD(T) level, are different in architecture. Also note the self-diffusion constant obtained from this NN-TL force field (from NVT TRPMD), 0.221 $\pm$ 0.06 Å^2/ps , is in good agreement with the present one of 0.233 $\pm$ 0.027 Å^2/ps (from NPT TRPMD).

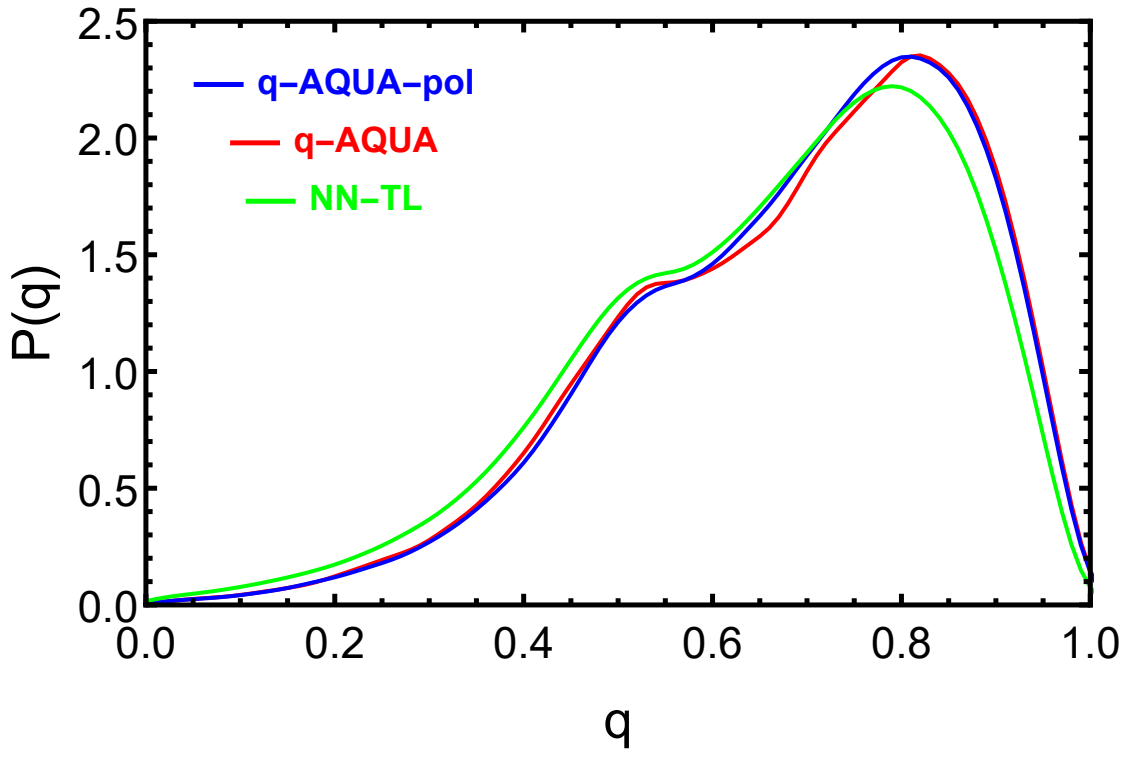

Figure 10: Probability distribution of the tetrahedral order parameter q at room temperature from q-AQUA-pol, q-AQUA, and recent NN-transfer-learned potential with CCSD(T) accuracy.[69] See text for more details.

A similar, at least in spirit, NN-TL water potential was reported by Marx and co-workers.[68] That potential made use of long-range fixed-charged electrostatic interactions, that are damped as usual to avoid the Coulomb singularity, as well as short range repulsive Yukawa potentials. The charges are -0.8 for O and 0.4 for H, exactly the ones in the first generation (1981) fixed charged model, TIP.[70] The properties reported were radial distribution functions and the self-diffusion constant, which are in good agreement with experiment.

We have already described the basic difference between q-AQUA and q-AQUA-pol and also the difference in treating the long-range for NN-TL potential of Marx and co-workers. We now make some general comments on the differences between the q-AQUA or q-AQUA-pol and the NN-TL potentials. One is the training of the models. For the MB approaches, training is done on small water clusters, as these are all that is needed for n-b interactions. For 2-b, 3-b and 4-b interactions included in q-AQUA and q-AQUA-pol, these are respectively the dimer, trimer and tetramer. CCSD(T) calculations are quite feasible for these and so a large range of geometries and energies can be explored for them. PIP fitting for the dimer and trimer was certainly feasible a decade ago as these were just 6 and 9-atom systems. The tetramer was a challenge a decade ago as it is a 12-atom system; however, given the recent progress in doing PIP fitting,[32] the fit for the 4-b became quite feasible. In any case, one consequence of training on these small clusters is the high accuracy in describing the energies of other clusters such as the hexamer, provided the MB representation up to the 4-b is accurate for these clusters. Indeed that was verified even for clusters as large as the 20-mer[71] and also for clusters up to the 25-mer.[9] The question that if a 1-4 body MBE would also provide an accurate description of condensed phase water, as demonstrated by q-AQUA, was certainly not obvious; however, as we showed elsewhere[10] and here for the order parameter it is. In one respect, however, we were not satisfied with accuracy of q-AQUA for the density. By interfacing q-AQUA to TTM3 we have "hedged our bets" and hopefully provided a somewhat more solid footing for q-AQUA. By contrast, training for the atom-centered NN force fields is done on samples of ca 32 or 64 monomers from direct

DFT-MD simulations. Transfer learning is then done on even smaller samples using approximate but feasible CCSD(T) calculations. So, it would be interesting to see if these NN-TL force fields produced the level of accuracy seen here or in MB-pol for the hexamer isomer energies. These were not reported in the two recent NN-TL force fields,[68,69] and so we are only making an educated speculation here.

Finally, we note an interesting recent study by Paesani and co-workers in which a NN FF was "transfer-learned" using MB-pol as the source of data.[72] The success in doing this was limited with respect to training on one property and transferability to other properties. This was especially extreme in the case of the many-body contributions to the water hexamer isomer energies. On the other hand if training was done on the n-body contributions to these energies, transferability to other properties was more successful, i.e., in agreement with MB-pol was good. Given that MB-pol is a many-body based potential, this could perhaps be seen as somewhat circular. However, it may also indicate that an accurate treatment of many-body components of the water potential is indeed an effective way to train a force field. The excellent results for the many-body q-AQUA and q-AQUA-pol for the properties of water investigated to date support this latter possibility.

Summary and Conclusions

The q-AQUA water potential has been enhanced to describe higher (greater than 4)-body interactions by interfacing it to the TTM3-F potential. In the spirit of Δ-machine learning, short-range 2- 3- and 4-b corrections were connected to the TTM3-F force field. These are permutationally invariant polynomial fits to 2b-, 3- and 4-b energy differences between corresponding CCSD(T) and TTM3-F energies. The resulting water potential, denoted as q-AQUA-pol, was shown, as expected, to be even more accurate than q-AQUA for benchmark energies and harmonic frequencies of eight isomers of the hexamer. Both q-AQUA-pol and q-AQUA are more accurate than the excellent MB-pol potential for this benchmark. Unre-

stricted diffusion Monte Carlo calculations of the water dimer and trimer using q-AQUA-pol, demonstrate the robustness of the potential for these demanding quantum calculations. Condensed phase properties including radial distribution functions, the self-diffusion constant, the liquid and supercooled liquid densities, the triplet OOO and tetrahedral order distribution were reported in NPT calculations for MD and PIMD or TRPMD calculaions. Agreement with available experimental data is uniformly excellent, especially for the temperature dependence of the density, which has been a major challenge for first-principles theory and to some extent for the MB-pol potential. The tetrahedral order parameter distributions from q-AQUA-pol and q-AQUA (shown here for the first time) are in very good agreement with one from a recent transfer-learned Neural Network potential[69] at 298 and 300 K, respectively. The Δ-machine learning approach taken here and described earlier[30] appears to be general, and it would be straightforward to use in any polarizable water potential.

Supplementary information

Effects of many-body corrections on the OO radial distribution function, representative plots of the time dependence of the the density of liquid water in NPT simulations and the mean square displacement for classical MD and PIMD simulations

Data Availability

The data generated and used in this study are available at upon request to the authors.

References

(1) McDaniel, J. G.; Schmidt, J. Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory. *Ann. Rev. Phys. Chem.* **2016**, *67*, 467–488.

(2) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Ann. Rev. Biophys.* **2019**, *48*, 371–394.

(3) Inakollu, V. S.; Geerke, D. P.; Rowley, C. N.; Yu, H. Polarisable Force Fields: What Do They Add in Biomolecular Simulations? *Curr. Opin. Struct. Biol.* **2020**, *61*, 182–190.

(4) 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.

(5) 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.

(6) 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**, 5001–5013.

(7) 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.

(8) Blazquez, S.; Vega, C. Melting Points of Water Models: Current Situation. *J. Chem. Phys.* **2022**, *156*, 216101.

(9) Herman, K. M.; Xantheas, S. S. An Extensive Assessment of the Performance of Pairwise and Many-body Interaction Potentials in Reproducing Ab Initio Benchmark Binding Energies for Water Clusters n = 2 – 25. *Phys. Chem. Chem. Phys.* **2023**, xx.

(10) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. q-AQUA: a Many-body CCSD(T) Water Potential, Including 4-body Interactions, Demonstrates the Quantum Nature of Water from Clusters to the Liquid Phase. *J. Phys. Chem. Letts.* **2022**, *13*, 5068–5074.

(11) 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.

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

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

(14) Huang, X. C.; 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.

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

(16) Wang, Y. M.; 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.

(17) Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M. Accurate *Ab initio* and "Hybrid" Potential Energy Surfaces, Intramolecular Vibrational Energies, and Classical IR Spectrum of the Water Dimer. *J. Chem. Phys.* **2009**, *130*, 144314.

(18) 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.

(19) 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.

(20) 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.

(21) Bowman, J. M.; Wang, Y. M.; Liu, H. C.; Mancini, J. S. Ab Initio Quantum Approaches to the IR Spectroscopy of Water and Hydrates. *J. Phys. Chem. Lett.* **2015**, *6*, 366.

(22) Bore, S.; Paesani, F. Quantum Phase Diagram of Water. *ChemRxiv* **2023**, *10.26434/chemrxiv-2023-kmmmz*.

(23) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(24) Zaspel, P.; Huang, B.; Harbrecht, H.; von Lilienfeld, O. A. Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited. *J. Chem. Theory and Comput.* **2019**, *15*, 1546–1559.

(25) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.

(26) Zhu, J.; Vuong, V. Q.; Sumpter, B. G.; Irle, S. Artificial Neural Network Correction for Density-functional Tight-binding Molecular Dynamics Simulations. *MRS Communications* **2019**, *9*, 867–873.

(27) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. 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.* **2021**, *154*, 051102.

(28) Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. Breaking the Coupled Cluster Barrier for Machine-Learned Potentials of Large Molecules: The Case of 15-Atom Acetylacetone. *J. Phys. Chem. Letts.* **2021**, *12*, 4902–4909.

(29) Liu, Y.; Li, J. Permutation-Invariant-Polynomial Neural-Network-Based Δ-Machine Learning Approach: A Case for the HO_2 Self-Reaction and Its Dynamics Study. *J. Phys. Chem. Letts.* **2022**, *13*, 4729–4738.

(30) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. Δ-Machine Learned Potential Energy Surfaces and Force Fields. *J. Chem. Theory and Comput.* **2023**, *19*, 1–17.

(31) 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.

(32) Houston, P. L.; Qu, C.; Yu, Q.; Conte, R.; Nandi, A.; Li, J. K.; Bowman, J. M. PESPIP: Software to Fit Complex Molecular and Many-body Potential Energy Surfaces with Permutationally Invariant Polynomials. *J. Chem. Phys.* **2023**, *158*, 044109.

(33) 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.

(34) Anderson, J. B. A Random-walk Simulation of the Schrödinger Equation: H_3^+. *J. Chem. Phys.* **1975**, *63*, 1499–1503.

(35) Kosztin, I.; Faber, B.; Schulten, K. Introduction to the Diffusion Monte Carlo Method. *Am. J. Phys.* **1996**, *64*, 633–644.

(36) Kapil, V. et al. i-PI 2.0: A Universal Force Engine for Advanced Molecular Simulations. *Comput. Phys. Commun.* **2019**, *236*, 214–223.

(37) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python Interface for Ab Initio Path Integral Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2014**, *185*, 1019–1026.

(38) Ceriotti, M.; Bussi, G.; Parrinello, M. Colored-Noise Thermostats à la Carte. *J. Chem. Theory Comput.* **2010**, *6*, 1170–1180.

(39) Ceriotti, M.; Manolopoulos, D. E. Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei. *Phys. Rev. Lett.* **2012**, *109*, 100604.

(40) Fanourgakis, G. S. An Extension of Wolf's Method for the Treatment of Electrostatic Interactions: Application to Liquid Water and Aqueous Solutions. *J. Phys. Chem. B* **2015**, *119*, 1974–1985.

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

(42) Wang, X.-G.; Carrington, T. Using Monomer Vibrational Wavefunctions to Compute Numerically Exact (12D) Rovibrational Levels of Water Dimer. *J. Chem. Phys.* **2018**, *148*, 074108.

(43) Leforestier, C.; Szalewicz, K.; van der Avoird, A. Spectra of Water Dimer from a New Ab Initio Potential with Flexible Monomers. *J. Chem. Phys.* **2012**, *137*, 014305.

(44) Wang, Y.; Bowman, J. M. Communication: Rigorous Calculation of Dissociation Energies (D_0) of the Water Trimer $(H_2O)_3$ and $(D_2O)_3$. *J. Chem. Phys* **2011**, *135*, 131101.

(45) Rocher-Casterline, B. E.; Ch'ng, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the Bond Dissociation Energy (D_0) of the Water Dimer, $(H_2O)_2$, by Velocity Map Imaging. *J. Chem. Phys.* **2011**, *134*, 211101.

(46) Ch'ng, L. C.; Samanta, A. K.; Wang, Y.; Bowman, J. M.; Reisler, H. Experimental and Theoretical Investigations of the Dissociation Energy (D_0) and Dynamics of the Water Trimer, $(H_2O)_3$. *J. Phys. Chem. A* **2013**, *117*, 7207–7216.

(47) 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.

(48) Wang, Y. M.; 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.

(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) Howard, J. C.; Tschumper, G. S. Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: (H2O)n=2, 3, 4, 5, 6. *J. Chem. Theory Comput.* **2015**, *11*, 2126.

(51) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(52) Soper, A. K.; Benmore, C. J. Quantum Differences between Heavy and Light Water. *Phys. Rev. Lett.* **2008**, *101*, 065502.

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

(54) Mills, R. Self-diffusion in Normal and Heavy Water in the Range 1-45°. *J. Phys. Chem.* **1973**, *77*, 685–688.

(55) Holz, M.; Heil, S. R.; Sacco, A. Temperature-dependent Self-diffusion Coefficients of Water and Six Selected Molecular Liquids for Calibration in Accurate ^{1}H NMR PFG Measurements. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4740–4742.

(56) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; 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.

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

(58) Chau, P.-L.; Hardwick, A. J. A New Order Parameter for Tetrahedral Configurations. *Mol. Phys.* **1998**, *93*, 511–518.

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

(60) 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.

(61) 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.

(62) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear Quantum Effects in Water and Aqueous

Systems: Experiment, Theory, and Current Challenges. *Chem. Rev.* **2016**, *116*, 7529–7550.

(63) Harvey, A. *Properties of Ice and Supercooled Water*; CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, FL, 2019.

(64) Bernal, J. D.; Fowler, R. H. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* **1933**, *1*, 515–548.

(65) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A Deep Learning Package for Many-body Potential Energy Representation and Molecular Dynamics. *Comp. Phys. Comm.* **2018**, *228*, 178–184.

(66) Zhang, L.; Wang, H.; Car, R.; E, W. Phase Diagram of a Deep Potential Water Model. *Phys. Rev. Lett.* **2021**, *126*, 236001.

(67) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359.

(68) Daru, J.; Forbert, H.; Behler, J.; Marx, D. Coupled Cluster Molecular Dynamics of Condensed Phase Systems Enabled by Machine Learning Potentials: Liquid Water Benchmark. *Phys. Rev. Lett.* **2022**, *129*, 226001.

(69) Chen, M. S.; Lee, J.; Ye, H.-Z.; Berkelbach, T. C.; Reichman, D. R.; Markland, T. E. Data-Efficient Machine Learning Potentials from Transfer Learning of Periodic Correlated Electronic Structure Methods: Liquid Water at AFQMC, CCSD, and CCSD(T) Accuracy. *J. Chem. Theory Comput.* **2023**, *x*, xxx.

(70) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(71) Heindel, J. P.; Herman, K. M.; Aprà, E.; Xantheas, S. S. Guest–Host Interactions in Clathrate Hydrates: Benchmark MP2 and CCSD(T)/CBS Binding Energies of CH_4, CO_2, and H_2S in $(H_2O)_{20}$ Cages. *J. Phys. Chem. Lett.* **2021**, *12*, 7574–7582.

(72) 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.

Notes

The authors declare no competing financial interests.

Acknowledgment

JMB thanks the ARO, DURIP grant (W911NF-14-1-0471), for funding a computer cluster where most of the calculations were performed and current financial support from NASA (80NSSC20K0360). We also acknowledge discussions with George Fanourgakis and for sending the source code for TTM2.1-F and discussions with Tom Markland and Tim Berkelbach and finally thank Michael Chen for sending the NN-TL data for the tetrahedral order parameter.