

A CCSD(T)-based 4-body Potential for Water

Apurba Nandi,^{*,†} Chen Qu,^{*,‡} Paul L. Houston,^{*,¶} Riccardo Conte,^{*,§} Qi Yu,^{*,||}
and Joel M. Bowman^{*,†}

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

[‡]*Department of Chemistry & Biochemistry, University of Maryland, College Park,
Maryland 20742, 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*

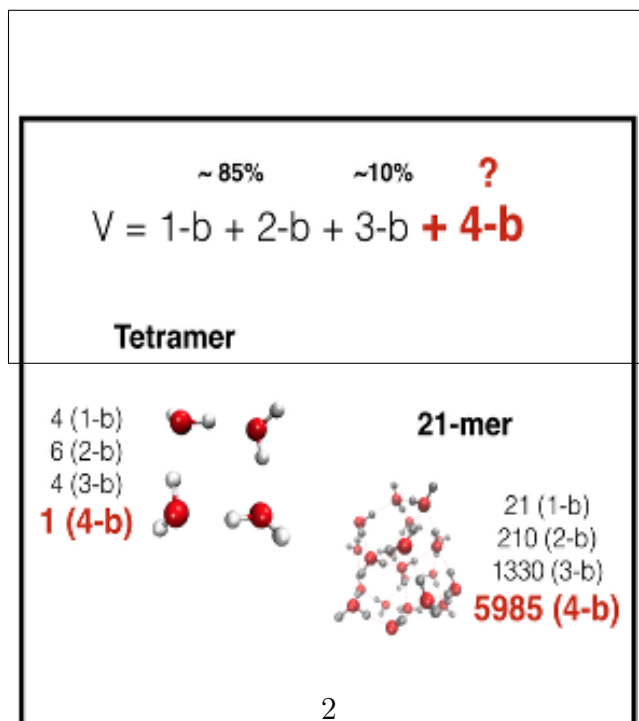
^{||}*Department of Chemistry, Yale University, New Haven, Connecticut 06520, U.S.A.*

E-mail: apurba.nandi@emory.edu; cqu12@umd.edu; plh2@cornell.edu; riccardo.conte1@unimi.it;
q.yu@yale.edu; jmbowma@emory.edu

Abstract

High-level, *ab initio* calculations find that the 4-body interaction is needed to account for near-100 percent of the total interaction energy for water clusters as large as the 21-mer. Motivated by this, we report a permutationally invariant polynomial, machine-learned, potential energy surface (PES) for the 4-body interaction. This PES is a fit to 2119, symmetry-unique, CCSD(T)-F12a/haTZ 4-b interaction energies. Configurations for these come from tetramer direct-dynamics calculations, fragments from an MD water simulation at 300 K, and from tetramer fragments in a variety of water clusters. The PIP basis is purified to ensure that the PES goes rigorously to zero in monomer+trimer and dimer+dimer dissociations. The 4-b energies of isomers of the hexamer calculated with the new PES are shown to be in better agreement with benchmark CCSD(T) results than those from the MB-pol potential. Tests on larger clusters further validate the high-fidelity of the PES. The PES is shown to be fast to evaluate, taking 2.4 seconds for 10^5 evaluations on a single core of 2.4 GHz Intel Xeon processor, and significantly faster using a parallel version of the PES.

Graphical TOC Entry



The many-body expansion (MBE) for the non-covalent interactions is pervasive in computational chemistry and has been applied in many guises for water. There are numerous studies of this expansion for water, and this large literature has been reviewed in a recent article.¹ There are many aspects of this expansion, but the one of particular interest here is the convergence of this expansion for water. This is not a trivial matter to examine since one has to ask, convergence with respect to what property? Certainly the most common property that has been examined in this respect is the complete dissociation energy of water clusters to all monomers or, almost equivalently, the interaction energy of clusters. The former includes relaxation of the monomer geometry in the full dissociation limit whereas the latter does not.

The general conclusion from numerical studies on moderate sized clusters is that an accurate, i.e., to within a few percent, description can be obtained with the MBE truncated after 3-body interactions. However, it has been known for many years that the 4-b interaction, while small, may be not negligible. Theoretical work that came to this conclusion dates back more than 35 years with the pioneering work investigating many-body aspects of the classical induced interaction of rigid water.^{2,3} Gregory and Clary investigated the effects of 4-b induction interactions on the zero-point energy of the water hexamer in 1996.⁴ (These interactions are formally N -body, where N is the number of monomers and are referred to as N -body induction energies.) It was assumed in this early work that 3 and 4-b interactions were negligible at short range and thus only the long-range classical induced energies were focused on. In 2000, Xantheas reported calculations of the total interaction energies of isomers of the water hexamer at the MP2 level and concluded that the 4-b contribution to the total interaction energy varies from 1.4 to 4.4 percent, depending on the isomer.⁵ In 2011, a many-body analysis of the lowest energy isomers for the hexamer, the prism and the cage, determined from CCSD(T)/aug-cc-pVTZ calculations that 4-b contributions to the interaction energy were -0.57 and -0.43 kcal/mol, respectively.⁶ A more recent examination of the 4-b interaction for the hexamer isomers confirms these values.⁷ This lowers the prism

energy by about 49 cm^{-1} more than the cage. Recently, Heindel and Xantheas reported a careful numerical examination (focusing on elimination of basis set superposition errors) of the MBE for interaction energies of water clusters, consisting of 7, 10, 13, 16 and 21 monomers.⁸ Based on this and previous analyses, they concluded that the MBE can thus be safely (i.e., converged to less than 1 percent) truncated after the 4-body term and not after the 3-body level.

The above conclusion notwithstanding, it is the case that the MBE truncated at the 3-body term does capture roughly 95 percent of the interaction energy for water clusters. Thus, in the past 10 or so years several *ab initio* water potentials have been developed based on 1-, 2- and 3-body *ab initio* PESs or correction PESs. These are known by the acronyms WHBB,⁹ HBB2-pol, MB-pol,¹⁰ and CC-pol.¹¹ The WHBB, HBB2-pol, and MB-pol potentials use a spectroscopically accurate 1-b potential,¹² and for the 2-b numerical fits, they are calculated using permutationally invariant polynomials (PIPs)¹³ and thousands of CCSD(T) energies.^{14–16} The 3-b potential in WHBB is a PIP fit to thousands of MP2 energies, whereas the one in the HBB2-pol and MB-pol potentials is based on fits to thousands of CCSD(T) energies. To be more precise, MB-pol uses PIP fits to the difference in 2-b and 3-b energies obtained from the TTM4-F PES¹⁷ and CCSD(T) energies. The CC-pol potential uses elaborate functional forms to represent the 2 and 3-b interactions and to fit them to many CCSD(T) energies.¹¹

The WHBB potentials account for 4 and higher-body interactions by switching to TTM3-F potential¹⁸ in the long-range. This potential provides a sophisticated treatment of long-range electrostatics for an arbitrary number of monomers. The MB-pol potential uses a different strategy, since it is based on correction PESs at the 2 and 3-body level to the TTM4-F potential¹⁷ plus additional electrostatic terms. A critical assessment of the accuracy of the TTM3-F and TTM4-F potentials for 4-b energies of isomers of the water hexamer against direct CCSD(T)-F12b/VTZ was reported in 2015.⁷ While the TTM4-F 4-b potential is generally more accurate than the TTM3-F one, it has errors between around 0.1 and 0.35

kcal/mol. These are fairly large fractions of the 4-b energy itself. Indeed these errors account substantially for the overall errors of the WHBB and MB-pol potentials for the interaction energies of hexamer isomers. In addition, these errors are potentially significant for rigorous studies of the relative energies of the prism and cage, which have been done with both the WHBB and MB-pol potentials and using rigorous treatments of the vibrational motion.¹⁹ In the latest study using MB-pol, the cage was reported to be more stable at 0 K than the prism by roughly 0.1 kcal/mol, including a rigorous treatment of zero-point energy.²⁰ This difference is within the range of errors in the hexamer 4-b interaction in the MB-pol potential and so the conclusion probably should be viewed with some caution.

Based on this and previous work, there is strong motivation to develop a full-dimensional PES for the 4-b water interaction potential. This is a challenging 12-atom system that was beyond consideration during the time when WHBB, MB-pol, and CC-pol were developed. Attempts have been made by some of us to obtain analytical expressions of high dimensional 4-body interactions of molecular systems, i.e. hydronium-water-water-water 4-b interaction.²¹ In that work, inspired by Skinner and coworkers' E3B water 3-body model,^{22,23} we proposed empirical exponential functions for different types of 4-b geometries and fitted relevant parameters from limited *ab initio* data. However, these simple exponential functions can not meet the requirement of a robust and high dimensional 4-body potential.

Recently, we extended the PIP approach to describe PESs for 12-atom *N*-methyl acetamide,^{24,25} 15-atom tropolone,²⁶ and acetylacetone.²⁷ Also noteworthy is a recent paper reporting a method to generate a very compact PIP fitting basis.²⁸ Both our papers from 2019 and this one have clearly broken the oft-cited "10-atom" limit for PIP bases that appeared 12 years ago.¹³ Because the long-range behavior is important for the 4-b interaction, here we use a PIP basis that is purified²⁹⁻³¹ so that the 4-b PES goes rigorously to zero as any monomer (or dimer) separates from the other group of monomers. The CCSD(T)-F12a/haTZ (aug-cc-pVTZ basis for O and cc-pVTZ for H) method was selected for the calculations of electronic energies as it provides accurate results at an acceptable computational cost.

Below we give details of the generation of the database of electronic energies and standard tests of the precision of the 4-b PES are given. Then the new PES is tested for the 4-b energies of 8 isomers of the water hexamer, and the total 4-b energies of a number of larger water clusters. The long-range behavior is examined for two dissociation 1d-profiles of the water tetramer to two dimers and to a monomer plus trimer.

It is instructive to investigate the contributions of the various n-body ($n=2, 3, 4$) interaction energies for the water tetramer. (Recall that these are obtained in a series of calculations where, for example, the 2-b energies correspond to the energies of all pairs of monomers, etc. Of course finally, the sum of these n-body interactions, $n=2-4$, must equal the total interaction energy of the tetramer.) This is shown in Fig. 1 for the the movement of two dimers starting at the global minimum (GM) structure where just the distance between two equivalent dimers, denoted R , is varied. At large R we have two equivalent non-interacting water dimers. The total interaction is repulsive for R less than 2.5 Å. But, as seen, the addition of the 3-b interaction to the 2-b ones moves the minimum to smaller R and likewise shifts the repulsive wall to smaller R . The same occurs, albeit to a smaller extent, by the addition of the 4-b interaction. Thus, over the range of the plot the 3-b and 4-b interactions are attractive. As seen, the 2-b interaction is the largest, followed by the 3-b, followed by the 4-b, as expected. The 4-b is very small in this case; however, it should be noted that while there are 6 2-b interactions and 4 3-b interactions there is only a single 4-b interaction for the tetramer. Of course, as the number of monomers increases the number of 4-body interactions grows faster than the number of 2 and 3-b interactions.

The above figure provides important evidence that the lower-limit of the range R for the 4-b PES should be roughly 2 Å. With that in mind a variety of approaches were used to generate a diverse database of 4-b configurations and energies. First, configurations were obtained from direct-dynamics calculations for the water tetramer. Second, tetramer fragments were selected from a classical dynamics simulation of water using the MB-pol potential. This source of data anticipates the eventual use of the 4-b potential for condensed phase

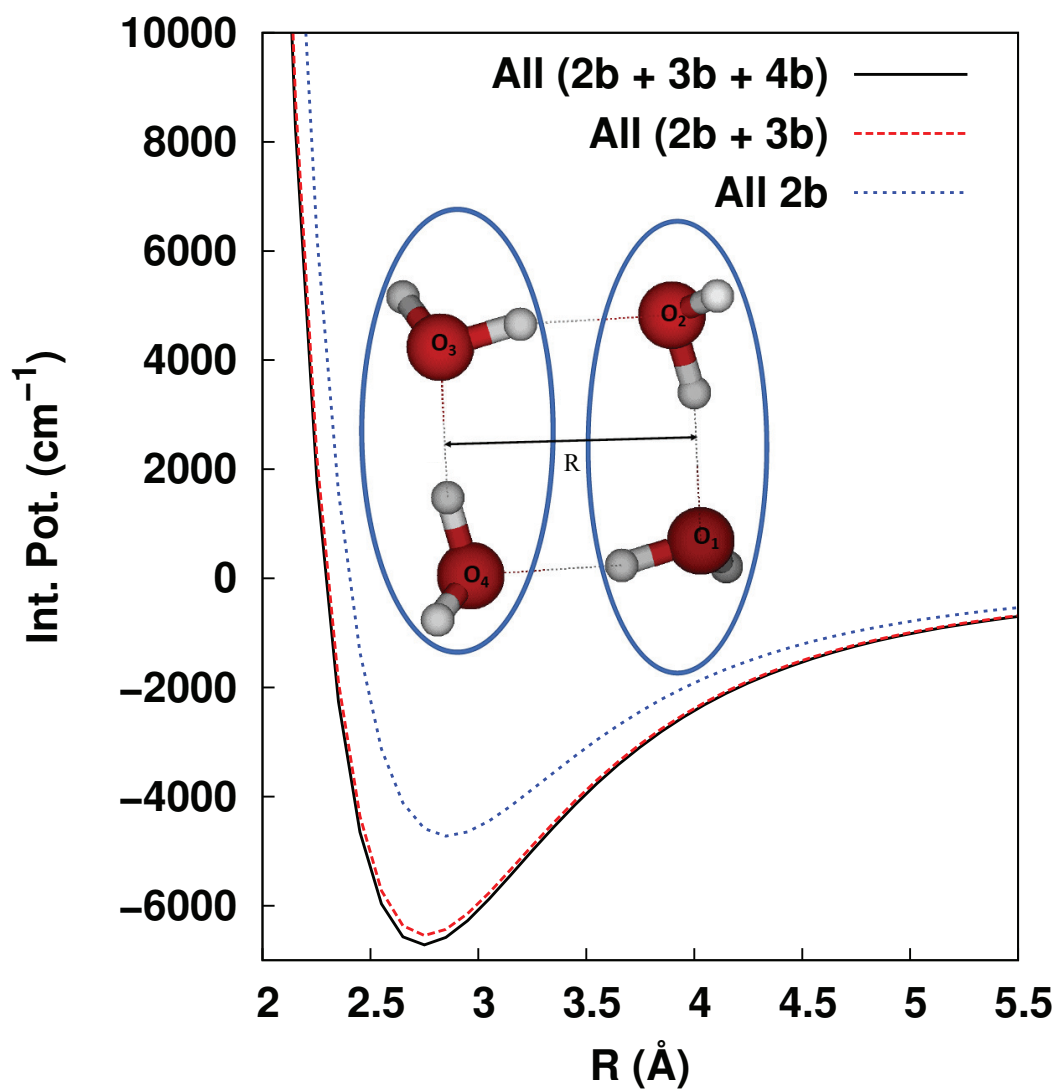


Figure 1: Interaction potential for the water tetramer including the n-body terms as indicated. See text for details of the 1d path defining this energy profile.

simulations. Third, tetramer fragments were selected from the equilibrium configuration of isomers of the water hexamer, heptamer, decamer, and 13-mer. Lastly, configurations along the dimer+dimer dissociation profile shown in Fig. 1. In total, 2119 configurations were obtained from this sampling, details of which are given next.

The NVE direct-dynamics calculations were done using the efficient B3LYP/6-31+G(d,p) level of theory. Trajectories were initiated from the two high-energy planar ring structures of the tetramer. These were generated by simply bringing the up-down-up-down ring structure of the global minimum to a planar structure by flipping the out-of-plane H atoms to the plane of the four O atoms. And for the first planar structure the distances between two equivalent dimers are 2.7515 and 2.7575 Å, respectively, and for the second one these are 2.3533 and 2.7575 Å, respectively. The energies of these structures are respectively 8,100 and 11,243 cm⁻¹ above the GM. These planar structures along with the GM one are given in the Supporting Information (SI). Two trajectories were calculated from the first planar structure and three trajectories were calculated from the second planar structure. An initial kinetic energy of 1000 cm⁻¹ was distributed randomly in these five independent trajectories to all atoms at these initial configurations. The total energies of tetramers in the direct-dynamics trajectories are thus 9,100 and 12,243 cm⁻¹. These trajectories were propagated for 30,000 time steps with the step size of 5.0 a.u.(about 0.12 fs). A histogram distribution of tetramer potential energies, relative to the global minimum, is shown in Fig. S2. The data generated from these trajectories was pruned by simply saving every 10th configuration (in time) to obtain 468 configurations for the final database. However, since dissociation of the tetramer did not occur in these direct-dynamics simulations, additional configurations are needed. These are describe next.

Additional 4-b configurations were obtained from tetramer clusters selected from a 300 K MD simulation of water using 256 monomers and the MB-pol potential. These are defined as four monomers with the third largest O-O distance smaller than 4.5 Å. These we selected quasi-randomly from the large database of monomer configurations, and 540 configurations

were selected.

Sixty one energies were obtained from the cut dissociating to the two dimers over the range 1.7 to 7.7 Å. Finally, tetramer configurations were obtained from 4 isomers of the water hexamer, namely the prism, cage, book-1, and cyclic-chair, the water heptamer, decamer, and 13-mer water clusters. All 4-body configurations were selected at the equilibrium structures of these clusters. For the hexamer isomers the structures are from ref. 32 and for the heptamer, decamer and 13-mer from ref. 8. In total 1050 configurations from these clusters were calculated.

To construct the new PES, we calculated the 4-b energies at the CCSD(T)-F12a/haTZ level of theory for the above 2119 configurations. These are not corrected for basis set superposition error, as this is negligible at the tetramer minimum and two other geometries, as shown in Table S1 of the SI. The entire energy range of these 4-b energies is -1063.1 to 105.2 cm^{-1} and the distribution of most of these energies in a smaller range is shown in Fig. 2. There are just a small number of 4-b energies outside this range.

Histograms of the OO distances in the tetramer configurations from the entire dataset are given in Fig. S4 of the SI. As seen there they span a range from roughly 2 to 9 Å, indicating good coverage of this important distance. Histograms of the 4-b energies for the heptamer and decamer are given in Fig. S5.

The PES is a linear least-squares fit to the data base of energies using a representation in permutationally invariant polynomials (PIPs).^{13,33,34} The permutational symmetry that describes all the permutations of like atoms has the designation 84, meaning that the eight hydrogens permute with one another and the four oxygens permute with one another. However, it is also possible to use reduced symmetries if one replicates the data set with the proper number of permutations of the water molecules. Here we use the symmetry 22221111, meaning that the hydrogens within each water monomer permute with one another but that they do not permute between water molecules, and that the oxygens do not permute among themselves. In order to describe the invariance with respect to the $4!$ permutations of water

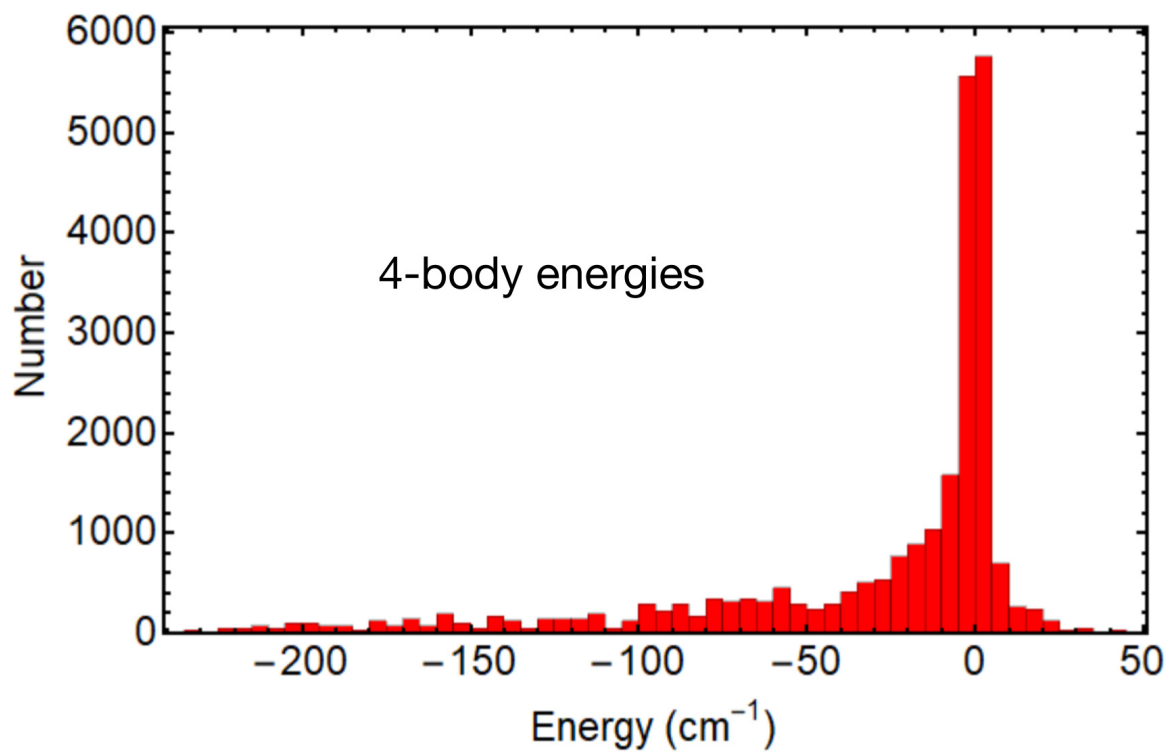


Figure 2: Histogram of the CCSD(T)-F12a/haTZ points used in the PES fit to the 4-body potential energy surface.

monomers we replicated the symmetry-unique configuration $4!$ times. The reason for using this reduced symmetry is given below, where we discuss the total polynomial order for the fit.

The usual method for fitting the energy data set with the PIP polynomials is not to use the internuclear distances directly in the PIPs but rather to transform them into functions such as Morse variables ($\exp(-r_{ij}/a)$) or $1/r_{ij}$ variables. For the current application, we used a mixture of these functions, namely Morse variables for the 12 intramolecular coordinates with the range parameter $a = 2$ Bohr, and $1/r_{ij}$ variables for the 54 intermolecular coordinates.

Another consideration for the basis set choice is the behavior of the PIPs as a monomer or dimer of the tetramer is removed to a great distance with respect to the remaining waters. By definition, the 4-body interaction energy must go to zero as any water monomer or water dimer is distanced. Many of the PIPs do not have this property.¹³ The process of eliminating these PIPs²⁹ is what we term “purification”.^{30,35,36} Our method of purification starts by assigning random numbers to all the internuclear distances for the four water molecules, 66 in this case, and evaluating the polynomials using Morse variables. One then adds a large number to the Cartesian coordinates of the water monomers, one at a time, and water dimers, two at a time, and then calculates the new values of the polynomials. If for any of these distancing operations the value of a polynomial does not become less than a small cutoff number (we used 10^{-6}), the polynomial is eliminated from the set.

For the maximum 84 symmetry and using a maximum polynomial order of 3, there are 86 PIPs, but only 2 remain after purification. Clearly this is not a usable basis. In the 22221111 symmetry with polynomial order 3, there are 10 737 PIPs, and 1649 remain after purification. Of course, we want the number of PIPs, equal to the number of unknown coefficients, to be less than the size of energy database but not to be so much less that we do not obtain good fit precision. Using 22221111 symmetry and replication the data set of 2119×24 we have 50 856 energies, and so 1649 coefficients should be sufficient and precise. Although we considered other possibilities, as listed in Table S2, many were infeasible and,

for some, the permutational replication of the data set was not effective because it mixed intra- and inter-monomer coordinates. In the end, we used the 22221111 symmetry with a data base size of 50 856 and 1649 coefficients. For this basis the fitting error for the energies, shown in detail in Fig. S3 of the SI, is small with an RMS error for the entire dataset of 6.2 cm^{-1} .

We now present several tests of the new 4-b PES, ranging from 1d energy profiles to tests for variety of water clusters. We compare to CCSD(T)-F12a calculations as well as results from the MB-pol and TTM4-F potentials. The first test is a comparison of 4-b energies from the PES, TTM4-F, and CCSD(T)-F12a for a 1-d profile showing the 4-b energy along the separation of the water tetramer to two rigid dimers passing through the cyclic minimum (OO equal to 2.75 \AA). These are shown in Figure 3 from the current PES, direct CCSD(T)-F12a/haTZ, and TTM4-F calculations. Note, CCSD(T) energies for OO greater than 8 \AA are not included in the training data set. As seen, over the range 1.7 to nearly 14 \AA , the PES is in very good agreement with direct CCSD(T)-F12a/haTZ energies. The 4-b energies from the TTM4-F potential are also in good agreement with the benchmarks for OO distance greater than 2 \AA . However, a major breakdown of the accuracy of that potential is seen for shorter distances. Evidently, this is where the classical electrostatic description of the two water dimer interaction fails, presumably due to significant chemical interaction between the electronic orbitals of the dimers. We focus here on the TTM4-F potential as it plays a central role in the MB-pol water potential.³⁷ Namely, MB-pol uses high-level, *ab initio* correction PESs to TTM4-F at the 2 and 3-b level in the short range and relies on that potential for all 4-b and higher body water interactions and also for the interactions beyond the range of the corrections. Finally note the gradual approach to zero in this profile. At 11 \AA the energy is less than -1.0 cm^{-1} .

A second profile showing PES and CCSD(T) energies is given in Fig. 4. Here the tetramer dissociates to a monomer plus trimer over an OO range of 2.7 to 15 \AA , and corresponding energy range of -175 cm^{-1} to 0 . This is a significant test of the PES as no CCSD(T) data

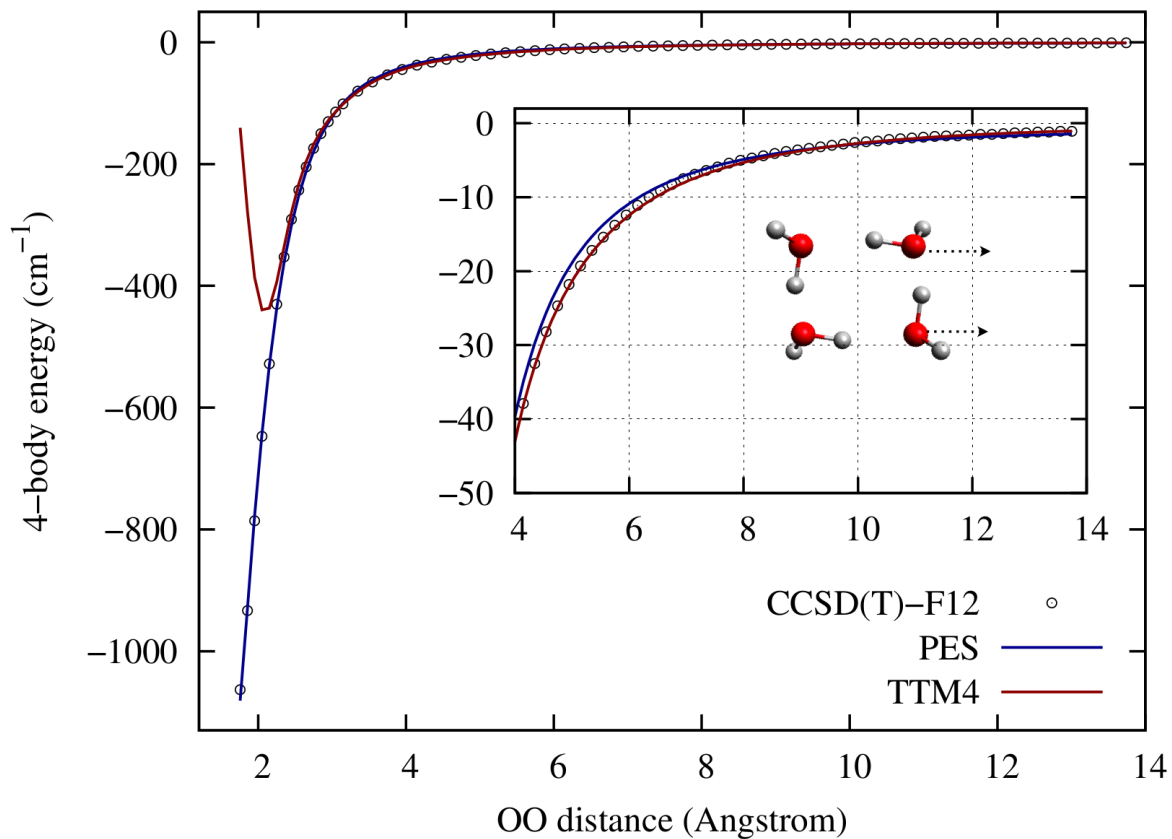


Figure 3: 4-b energies from indicated sources (see text for more details) as a function of the oxygen-oxygen distance between pairs of water dimers in the tetramer. The dashed arrows indicate the dimer pair that separates from the rigid tetramer. The equilibrium value of this distance is 2.7 Å.

for this profile were included in the fitting database. As seen, the PES agrees well with the benchmark results and the range of the interaction for this profile is roughly 7 Å.

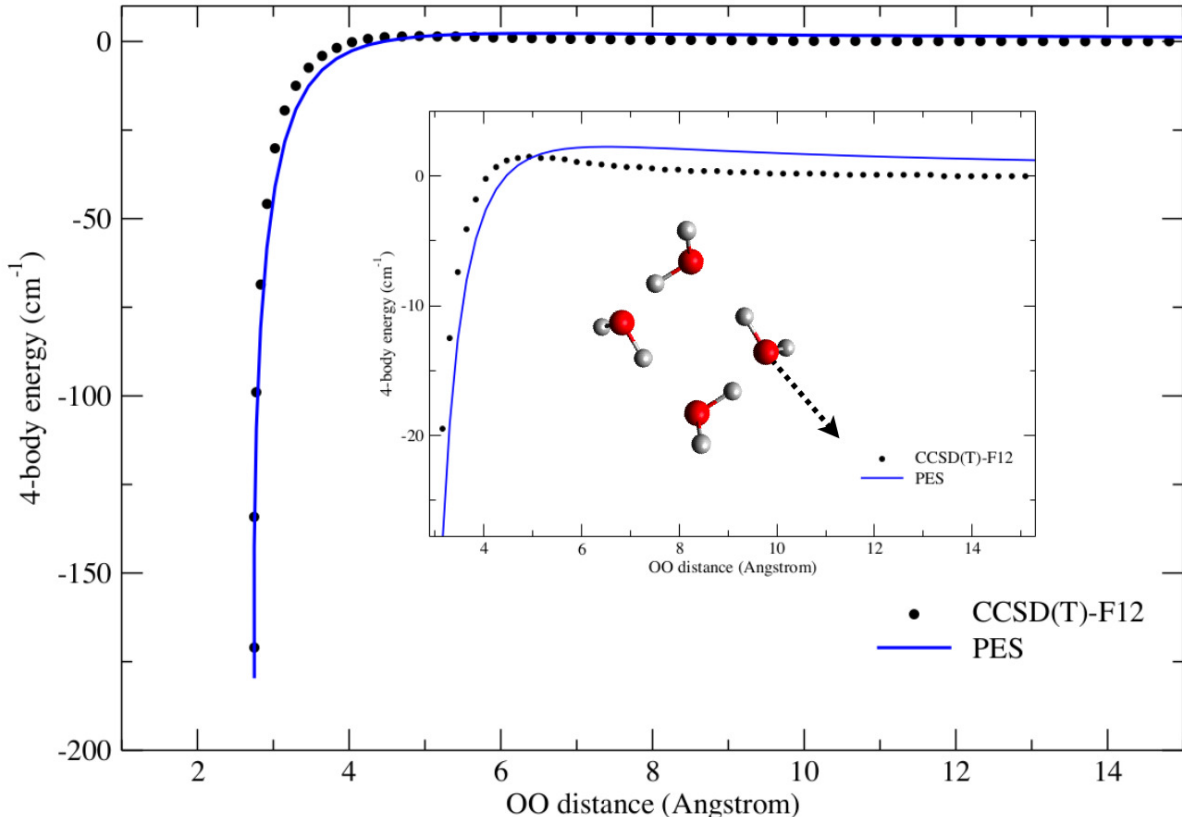


Figure 4: Test of the 4-b PES for a cut with a single monomer separating from the tetramer against CCSD(T)-F12a/haTZ energies. OO is the distance between the O atoms on the two monomers on the axis inferred from the arrow.

Both profiles show that the 4-b energy is rapidly becoming more negative at OO distances near and especially less than the equilibrium value of 2.7 Å. (This was already inferred in the discussion about Fig. 1, where the many-body decomposition of the tetramer potential was shown.) In general, this indicates the possible importance of the 4-b in softening the repulsive interaction in the condensed phase, especially for high pressure conditions. At long range the two 1d-potentials show different behavior with trimer+monomer interaction at somewhat shorter range than the dimer+dimer interaction. Also, the 1-d potential for the former decays to zero slightly more slowly (a difference of 1–2 cm⁻¹) than the CCSD(T) reference

energies. Below we comment on practical approaches to treat the long-range behavior of the 4-b interaction for uses in applications to the condensed phase with many monomers. Those comments will address in more detail this small difference between the PES and CCSD(T) energies for the trimer+monomer interaction.

Another test of the PES is the 4-b energies of the isomers of the water hexamer, prism, cage, book-1 (bk-1), book-2 (bk-2), bag, cyclic-chair (c-chair), cyclic-boat-1 (c-bt-1) and cyclic-boat-2 (c-bt-2). These isomers have been the focus of a number of papers, and the one by Medders et al. is of particular interest.⁷ There the 4-body energy from several water potentials, including WHBB5, MB-pol, TTM3-F and TTM4-F were compared with CCSD(T)-F12/VTZ energies for eight isomers. Here we focus on the errors in the MB-pol and the new 4-b PES. These are shown in Fig. 5. As seen, the 4-b PES errors are smaller than those from MB-pol. A numerical comparison of 4-b energies, the present CCSD(T) calculations and the previous ones from Medders et al are given in Table S3 in the SI. The two sets of CCSD(T) calculations are within 0.1 kcal/mol and generally less than that for the eight isomers. Given that MB-pol and TTM4-F 4-b energies differ by no more than 0.04 kcal/mol, and given the limited accuracy of TTM4-F shown in Fig. 3, for OO distances relevant to the hexamer, the lower level of accuracy of MB-pol compared to the PES is not unexpected.

One aspect of an accurate 4-b PES is that it can be used for arbitrarily many monomers with virtually zero BSSE. Thus, it is possible to test the prediction of 4-b energies of larger water clusters using the 4-b PES and against benchmark calculations. A convenient set of benchmark calculations has been published by Heindel and Xantheas,⁸ who provided geometries (in their Supporting Information) and reported many-body energies calculated with and without BSSE-correction at several levels of theory. Table 1 shows their results in columns 4 and 5 for MP2 and the basis set listed in column 3. The results using the PES to calculate all 4-b energy in each n-mer are shown in the last column. The agreement is good through $n = 16$. Remarkably, even better agreement is obtained between the PES

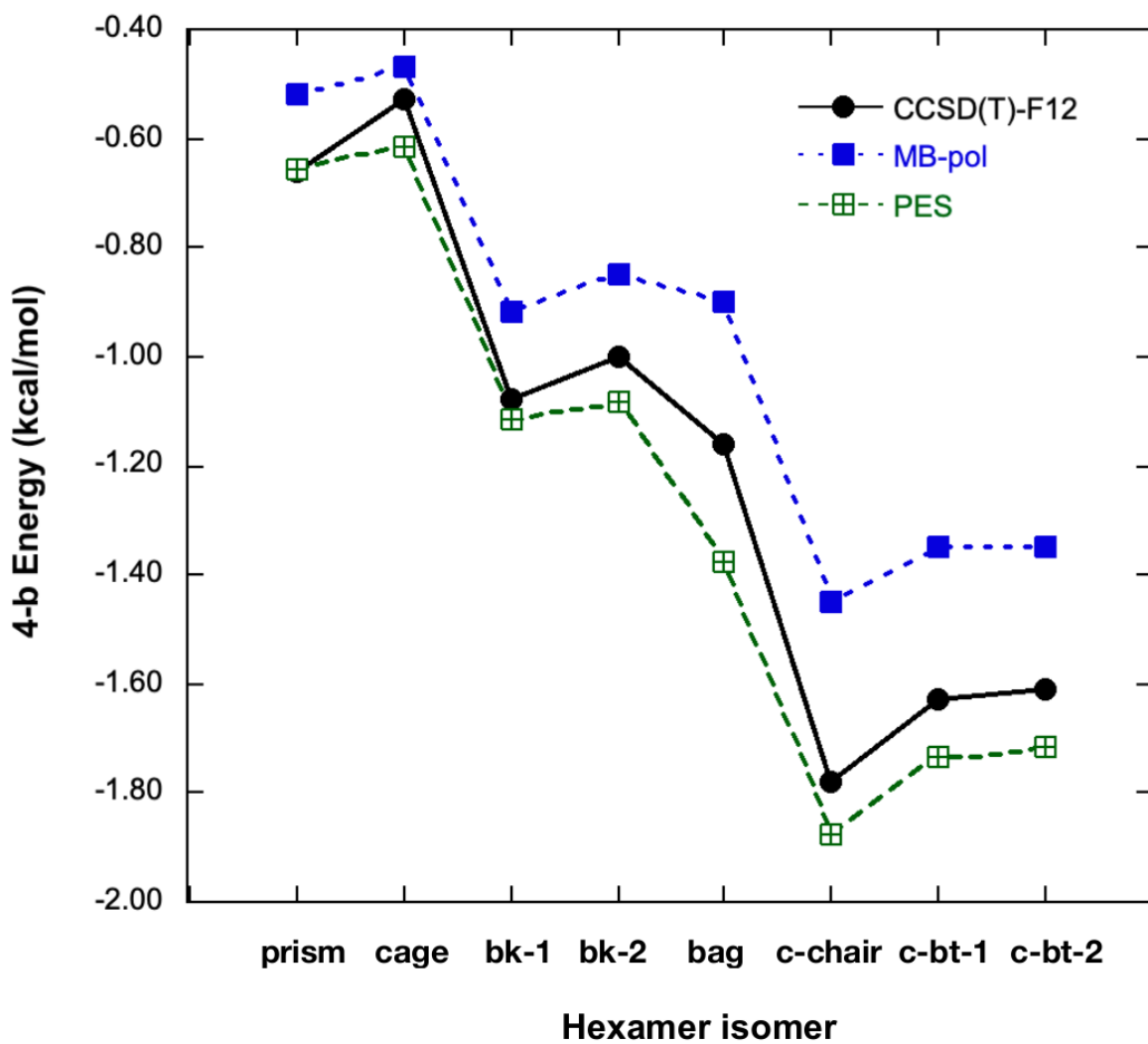


Figure 5: 4-b CCSD(T) energies for indicated isomers of the water hexamer from the 4-b PES, MB-pol, and previous CCSD(T)-F12/VTZ calculations.⁷

and benchmark CCSD(T)-F12a/haTZ calculations performed up to the 13-mer. For the two isomers of $n = 21$ the BSSE correction to MP2 calculations was not applied and coupled clusters energies were not obtained as both computations were evidently too intensive. As seen, the PES 4-b results are much less than the uncorrected MP2/aVDZ results. While this is almost certainly qualitatively correct, the quantitative accuracy of the 4-b results would require new benchmark calculations, which for these larger clusters are currently not easily done.

Table 1: 4-b interaction energies (kcal/mol) for various water clusters, where the information in columns 3–5 is taken from the SI in ref. 8, numbers in column 6 are from our CCSD(T)-F12/haTZ calculations, and the numbers in column 7 are calculated using our 4-b PES.

$(\text{H}_2\text{O})_n$ n=	No. of 4mers in nmer	H-X basis	MP2	MP2 BSSE- corrected	CCSD(T)- F12/haTZ	4-b PES
7 ^a	35	aVTZ	-1.106	-0.874	-0.987	-1.012
10 ^a	210	aVTZ	-3.028	-1.978	-2.576	-2.341
13 ^a	715	aVDZ	-6.352	-1.499	-1.539	-1.571
16 ^a	1820	aVDZ	-9.463	-2.179	-	-2.983
21 ^b	5985	aVDZ	-20.976	-	-	-9.888
21 ^c	5985	aVDZ	-19.892	-	-	-5.920

^a These are global minimum structures;

^b Fully solvated structure;

^c All surface structure.

For most of the large clusters, many of the 4-mer configurations involve at least one water that is quite distant from the others, so that the 4-b energy is quite small. For example, Fig. 6 shows a histogram of the PES-determined 4-b energies for the “fully-solvated” 21-mer. There are large peaks near zero, although the distribution actually stretches from -95 to 54 cm^{-1} , with only a minor number of energies outside the range depicted in the figure. Since the total 4-b energy is the sum of all the 4-body contributions (5985 in the case of $n = 21$), there is evidently a large degree of cancellation. The issue of the many very small 4-b interactions is of course an important one and we comment on it after we discuss the computational effort to evaluate the 4-b PES, which we do next.

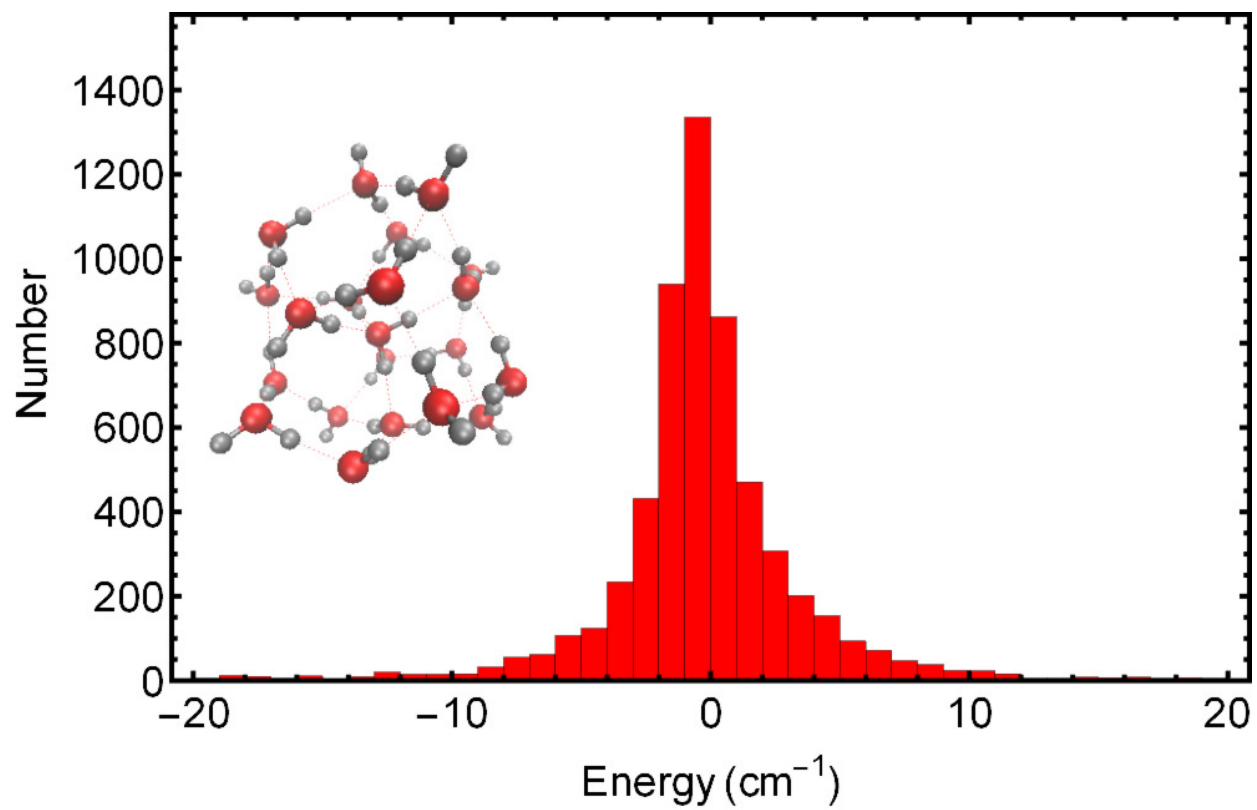


Figure 6: Distribution of PES 4-b energies for the fully solvated 21-mer

Finally, an important aspect of the new 4-b PES is the computational effort to use it. One hundred thousand evaluations of the 4-b PES takes 2.45 seconds on a single core of the 2.4 GHz Intel Xeon processor. This is significantly smaller than the computational effort to evaluate our WHBB PES for the water trimer.

While this computational effort is not large, it should be recalled that the number of 4-b interactions goes formally as N^4 for N water monomers, Thus, the cost would be high if all the tetramers were to be evaluated for say a condensed phase simulations which require hundreds of monomers. A standard approach to deal with this is to use a distance-based cutoff via a standard switching function. This would greatly reduce the number of tetramers to be evaluated because in large clusters many tetramers have large inter-monomer distances and the 4-b energy for such tetramers is essentially zero and thus not necessary to evaluate. We demonstrate this below with results shown in Table 2, for clusters with 64, 128 and 256 monomers. In this case we evaluate the 4-b potential of tetramers whose largest O-O distance is smaller than 7 Å. (We don't necessarily recommend this for all applications of course, it is used here just for illustration.) With this choice more than 99% of the tetramers are not evaluated, and calculating the total 4-b energy takes seconds, as seen, rather than more than an hour. Parallelization is another possible strategy since the calculation of one tetramer is independent of the others. This has been done using OpenMP and as seen there is a substantial speed-up in the calculations.

Table 2: Computation cost of 4-b PES calculation with cut-off O-O distance of 7 Å

$(\text{H}_2\text{O})_n$ n=	No. of total 4b	No. of calculated 4b	Time (s) single core	Time (s) 8 cores
64	635,376	5,078	0.18	0.03
128	10,668,000	24,460	0.79	0.12
256	174,792,640	58,804	1.80	0.25

Limiting the range of the 4-b interaction through the use of a multiplicative damping factor (switching function) that brings the 4-b to zero in a finite range is the well-established way to deal with the huge number of interactions. The parameters of such a function

of course are varied, within practical limits, to establish systematic uncertainties of any property with respect to this switching. An example from our group used damping of the 3-b interaction in $\text{H}_2\text{-H}_2\text{O-H}_2\text{O}$ interaction in a diffusion Monte Carlo quantum calculation of vibrational energies of H_2 in a hydrate clathrate.³⁸ In this calculation the WHBB water potential was used to describe the water interaction and a separate 1,2,3-body potential was used to describe the H_2 -water interactions.³⁵ It is worth noting that this is a general many-body approach to describe the interaction of molecules and ions with water.

In summary, we reported the first *ab initio*, full-dimensional potential energy surface for the 4-body interaction of water. The potential is a purified PIP fit to a diverse set of 2119 unique CCSD(T)-F12a/haTZ energies. This representation ensures that the PES rigorously goes to zero as a monomer or dimer is separated from a tetramer cluster. Tests of the 4-b PES demonstrate the high-fidelity compared to benchmark calculations. In addition, the 4-b PES was shown to significantly reduce errors in 4-b interaction energies for isomers of the water hexamer obtained from a high-level water potential that describes water interactions to the 3-b level. This new 4-b PES was shown to be fast to evaluate, and for applications to large numbers of water monomers a large fraction of 4-b configurations are beyond the range of the 4-b PES and so do not have to be evaluated. This was demonstrated for the 21-mer. The PES is also robust, at least for the present applications. In total roughly 100,000 4-b energies have been obtained using the new PES and none show huge negative values, i.e., “holes”. The current 4-b PES could be an add-on to a water potential that either does not contain 4-b terms or has an approximate treatment of the 4-b, e.g., the TTMn family of potentials. Specifically, as seen in Fig. 4, TTM4-F provides an accurate description of the 4-b interaction, except at short range where it fails. This suggests that a correction PES to the TTM4-F 4-b interaction might be a reasonable future project. However, the unphysical behavior of TTM4-F shown at short OO distances could be problematic for a correction PES. In any case, we plan to investigate this in the future. Finally, it should be clear that the present 4-b PES is “version 1.0”. And while it appears to be both fast to evaluate and

accurate, we anticipate being able to make significant speed-ups, the easiest one of which is to make use of multi-core architecture of all modern workstations and computer nodes.

Supporting Information

The Supporting Information contains plots mentioned in the main text, basis-set-superposition error analysis, an analysis of the PES fitting RMS error and a table showing 4-b energies for isomers of the water hexamer from the present CCSD(T)-F12a/haTZ calculations, the present PES and previously reported MB-pol and CCSD(T) energies.

Acknowledgment

JMB thanks NASA (80NSSC20K0360) for financial support. We thank Francesco Paesani for sending the 300 K classical trajectory data for 256 waters using the MB-pol potential.

Data Availability

The data and 4-b PES fit that support the findings of this study are available from the corresponding authors upon reasonable request.

References

- (1) 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.

- (2) Detrich, J.; Corongiu, G.; Clementi, E. Monte Carlo Liquid Water Simulation with Four-Body Interactions Included. *Chem. Phys. Lett.* **1984**, *112*, 426–430.
- (3) Lybrand, T. P.; Kollman, P. A. Water–Water and Water–Ion Potential Functions Including Terms for Many Body Effects. *J. Chem. Phys.* **1985**, *83*, 2923–2933.
- (4) Gregory, J. K.; Clary, D. C. Structure of Water Clusters. The Contribution of Many-Body Forces, Monomer Relaxation, and Vibrational Zero-Point Energy. *J. Phys. Chem.* **1996**, *100*, 18014–18022.
- (5) Xantheas, S. S. Cooperativity and Hydrogen Bonding Network in Water Clusters. *Chem. Phys.* **2000**, *258*, 225–231.
- (6) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction Energies of Large Clusters from Many-Body Expansion. *J. Chem. Phys.* **2011**, *135*, 224102.
- (7) Medders, G. R.; Götz, A. W.; Morales, M. A.; Bajaj, P.; Paesani, F. On the Representation of Many-Body Interactions in Water. *J. Chem. Phys.* **2015**, *143*, 104102.
- (8) Heindel, J. P.; Xantheas, S. S. The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 6843–6855.
- (9) 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.
- (10) 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.
- (11) Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. Predictions for Water Clusters from a First-Principles Two- and Three-Body Force Field. *J. Chem. Phys.* **2014**, *140*, 194101.

- (12) 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.
- (13) Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577.
- (14) Huang, X. C.; Braams, B. J.; Bowman, J. M. Ab Initio Potential Energy and Dipole Moment Surfaces of $(\text{H}_2\text{O})_2$. *J. Phys. Chem. A* **2006**, *110*, 445–451.
- (15) 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.
- (16) 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.
- (17) 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.
- (18) 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.
- (19) 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*, 11116–11119.
- (20) Mallory, J. D.; Mandelshtam, V. A. Diffusion Monte Carlo Studies of MB-pol $(\text{H}_2\text{O})_{2-6}$

- and $(\text{D}_2\text{O})_{2-6}$ Clusters: Structures and Binding Energies. *J. Chem. Phys.* **2016**, *145*, 064308.
- (21) Heindel, J. P.; Yu, Q.; Bowman, J. M.; Xantheas, S. S. Benchmark Electronic Structure Calculations for $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, $n = 0-5$, Clusters and Tests of an Existing 1,2,3-Body Potential Energy Surface with a New 4-Body Correction. *J. Chem. Theory. Comput.* **2018**, *14*, 4553–4566.
- (22) Kumar, R.; Skinner, J. L. Water Simulation Model with Explicit Three-Molecule Interactions. *J. Phys. Chem. B* **2008**, *112*, 8311–8318.
- (23) Tainter, C. J.; Pieniazek, P. A.; Lin, Y. S.; Skinner, J. L. Robust Three-Body Water Simulation Model. *J. Chem. Phys.* **2011**, *134*, 184501.
- (24) Qu, C.; Bowman, J. M. Communication: A Fragmented, Permutationally Invariant Polynomial Approach for Potential Energy Surfaces of Large Molecules: Application to N-methyl acetamide. *J. Chem. Phys.* **2019**, *150*, 141101.
- (25) Nandi, A.; Qu, C.; Bowman, J. M. Using Gradients in Permutationally Invariant Polynomial Potential Fitting: A Demonstration for CH_4 Using as Few as 100 Configurations. *J. Chem. Theor. Comp.* **2019**, *15*, 2826–2835.
- (26) Houston, P. L.; Conte, R.; Qu, C.; Bowman, J. M. Permutationally Invariant Polynomial Potential Energy Surfaces for Tropolone and H and D atom Tunneling Dynamics. *J. Chem. Phys.* **2020**, *153*, 024107.
- (27) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. Full-dimensional Potential Energy Surface for Acetylacetone and Tunneling Splittings. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7758–7767.
- (28) Moberg, D. R.; Jasper, A. W.; Davis, M. J. Parsimonious Potential Energy Surface Ex-

- pansions Using Dictionary Learning with Multipass Greedy Selection. *J. Phys. Chem. Lett.* **2021**, *12*, 9169–9174.
- (29) Paukku, Y.; Yang, K. R.; Varga, Z.; Truhlar, D. G. Global Ab Initio Ground-state Potential Energy Surface of N_4 . *J. Chem. Phys.* **2013**, *139*, 044309.
- (30) Conte, R.; Houston, P. L.; Bowman, J. M. Communication: A Benchmark-quality, Full-dimensional Ab Initio Potential Energy Surface for Ar–HOCO. *J. Chem. Phys.* **2014**, *140*, 151101.
- (31) Conte, R.; Qu, C.; Bowman, J. M. Permutationally Invariant Fitting of Many-Body, Non-covalent Interactions with Application to Three-Body Methane–Water–Water. *J. Chem. Theory Comput.* **2015**, *11*, 1631–1638.
- (32) Howard, J. C.; Tschumper, G. S. Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: $(H_2O)_{n=2,3,4,5,6}$. *J. Chem. Theory Comput.* **2015**, *11*, 2126–2136.
- (33) Xie, Z.; Bowman, J. M. Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting via Monomial Symmetrization. *J. Chem. Theory Comput.* **2010**, *6*, 26–34.
- (34) Conte, R.; Qu, C.; Houston, P. L.; Bowman, J. M. Efficient Generation of Permutationally Invariant Potential Energy Surfaces for Large Molecules. *J. Chem. Theory Comput.* **2020**, *16*, 3264–3272.
- (35) Homayoon, Z.; Conte, R.; Qu, C.; Bowman, J. M. Full-Dimensional, High-Level Ab Initio Potential Energy Surfaces for $H_2(H_2O)$ and $H_2(H_2O)_2$ with Application to Hydrogen Clathrate Hydrates. *J. Chem. Phys.* **2015**, *143*, 084302.
- (36) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. “Plug and Play” Full-Dimensional

Ab Initio Potential Energy and Dipole Moment Surfaces and Anharmonic Vibrational Analysis for $\text{CH}_4\text{-H}_2\text{O}$. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8172–8181.

- (37) 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.
- (38) Qu, C.; Bowman, J. M. Assessing the Importance of the $\text{H}_2\text{-H}_2\text{O-H}_2\text{O}$ Three-Body Interaction on the Vibrational Frequency Shift of H_2 in the sII Clathrate Hydrate and Comparison with Experiment. *J. Phys. Chem. A* **2019**, *123*, 329–335.