JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

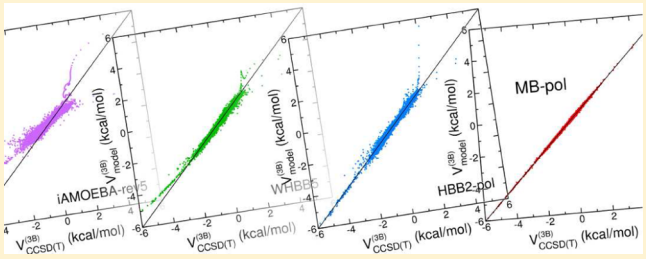
Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters

Volodymyr Babin, Gregory R. Medders, and Francesco Paesani*

Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, United States

Supporting Information

ABSTRACT: A full-dimensional potential energy function (MB-pol) for simulations of water from the dimer to bulk phases is developed entirely from "first principles" by building upon the many-body expansion of the interaction energy. Specifically, the MB-pol potential is constructed by combining a highly accurate dimer potential energy surface [*J. Chem. Theory Comput.* **2013**, *9*, 5395] with explicit three-body and many-body polarization terms. The three-body contribution, expressed as a combination of permutationally invariant polynomials and classical polarizability, is derived from a fit to more than 12000 three-body energies calculated at the CCSD(T)/aug-cc-pVTZ level of theory, imposing the correct asymptotic behavior as predicted from "first principles". Here, the accuracy of MB-pol is demonstrated through comparison of the calculated third virial coefficient with the corresponding experimental data as well as through analysis of the relative energy differences of small clusters.



1. INTRODUCTION

Despite recent progress in the development of electronic structure approaches to modeling molecular interactions in increasingly larger systems, chemical accuracy remains out of reach for most applications to condensed phase systems.[1−10] In principle, accurate multidimensional potential energy surfaces (PESs) can be obtained at the coupled cluster level of theory including single, double, and perturbative triple excitations, CCSD(T), which represents the current "gold standard" of chemical accuracy. Unfortunately, the computational cost associated with CCSD(T) makes these calculations prohibitively expensive even for small molecular clusters.[11] To overcome this computational barrier while still providing an *ab initio* representation of the underlying PES, different models based on density functional theory (DFT) have been developed. However, the choice of the most appropriate functional for studying (weak) intermolecular interactions, and, particularly, hydrogen bonds in condensed phases remains the subject of ongoing research.[12−15]

The many-body expansion of interactions[17] provides a convenient framework for the analysis and development of multidimensional PESs. Assuming a single-component system, the total energy of N interacting molecules can be expressed as a sum over n-body terms with $1 \le n \le N$,

$$E_N(x_1, ..., x_N) = \sum_a V^{(1B)}(x_a) + \sum_{a>b} V^{(2B)}(x_a, x_b) + \sum_{a>b>c} V^{(3B)}(x_a, x_b, x_c) + ... + V^{(NB)}(x_1, ..., x_N) \quad (1)$$

Here, x_a denotes the coordinates of all atoms in the a-th molecule, and $V^{(nB)}$ are the n-body interactions defined recursively as

$$V^{(nB)}(x_1, ..., x_n) = E_n(x_1, ..., x_n) - \sum_a V^{(1B)}(x_a) - \sum_{a>b} V^{(2B)}(x_a, x_b) - \cdots - \sum_{a_1>a_2>\cdots>a_{n-1}} V^{((n-1)B)}(x_{a_1}, x_{a_2}, ..., x_{a_{n-1}}) \quad (2)$$

For water, which is the focus of the present study, the many-body expansion has been shown to converge rapidly.[11,18−24] Since the low-order interaction terms can be accurately calculated at the CCSD(T) level, eq 1 effectively enables the development of water PESs with chemical accuracy.[7,25−28]

This paper describes the development of the many-body MB-pol water potential whose 2-body term was introduced in ref 29. The MB-pol potential, which represents an iterative improvement over the HBB2-pol model described in refs 27 and 28, takes advantage of a Thole-type model (TTM) for polarization[30,31] along with a pair-additive approximation of the dispersion energy to represent the intermolecular interactions in regions where the overlap between the electron densities of the individual water molecules is negligible. The complexity of the interactions at short-range is instead modeled via permutationally invariant polynomials[32,33] in quickly decaying functions of the interatomic distances. The coefficients of these polynomials are fitted to CCSD(T) reference energies. The

Received: February 1, 2014
Published: March 11, 2014

   dx.doi.org/10.1021/ct500079y | *J. Chem. Theory Comput.* 2014, 10, 1599−1607

functional form of the MB-pol potential is thus given by the following expression

$$E_N(x_1, ..., x_N) = \sum_a V^{(1B)}(x_a) + \sum_{a>b} V^{(2B)}_{short}(x_a, x_b) + \sum_{a>b>c} V^{(3B)}_{short}(x_a, x_b, x_c) + V_{TTM} + V_{disp} \quad (3)$$

Here, $V^{(1B)}$ is the 1-body term associated with intramolecular distortion described by the spectroscopically accurate monomer PES developed by Partridge and Schwenke,[34] V_{TTM} represents the (classical) electrostatic contributions described in detail in ref 29, V_{disp} represents the dispersion energy, and $V^{(2B)}_{short}$ and $V^{(3B)}_{short}$ are the short-range two-body and three-body terms. Both $V^{(2B)}_{short}$ and V_{disp} terms were developed and validated in ref 29. During the development of the complete MB-pol potential, slight modifications of $V^{(2B)}_{short}$ and V_{disp} (see Appendix A) were required to guarantee full integration with the three-body term, $V^{(3B)}_{short}$, which is the primary subject of this study. The article is organized as follows: Section 2 describes the CCSD(T) electronic structure calculations involved in the development of a representative training set of water trimers. The *ab initio* reference energies are used in section 3 to analyze the accuracy of the three-body energies predicted by the iAMOEBA-rev5,[35] WHBB,[26] HBB2-pol,[27] and, when possible, (rigid) CCpol[36] potentials. The functional form of the MB-pol $V^{(3B)}_{short}$ term is described in section 4 along with details related to the fitting procedure, and section 5 is devoted to the validation of the MB-pol potential through comparison of the calculated third virial coefficient and relative energies of small water clusters with the corresponding (experimental and *ab initio*) data available in the literature. A brief summary and outlook are given in section 6.

2. TRAINING SET AND REFERENCE ENERGIES

2.1. "First Principles" Energies. The trimer reference energies were calculated at the CCSD(T)[37] level using the augmented correlation-consistent polarized-valence triple (aug-cc-pVTZ) basis set[38,39] supplemented by an additional set of (3s,3p,2d,1f) midbond functions[40] with exponents equal to (0.9, 0.3, 0.1) for s and p orbitals, (0.6, 0.2) for d, and 0.3 for f, placed at the center of mass of each trimer configuration. The counterpoise method[41] was applied to remove the basis set superposition error (BSSE); specifically, the "site−site" variation of the counterpoise method was used, with the energies of each trimer and its constituent fragments (dimers and monomers) computed in the trimer basis.[42] Since the calculated three-body interaction energies were found to be very close to the complete basis set (CBS) limit, no explicit extrapolation was performed. All CCSD(T) calculations were carried out with Molpro[43] without assuming any symmetry constraints on the trimer structures.

2.2. Short-Range Three-Body Training Set. The short-range three-body training set contains 12347 trimers extracted from (a) the low-energy subset of the HBB2-pol training set described in ref 28; (b) path-integral molecular dynamics (PIMD) simulations of liquid water carried out at ambient conditions (T = 298.15 K and ρ = 0.997 g/cm^3 with the HBB2-pol potential;[27] (c) PIMD/HBB2-pol simulations of small $(H_2O)_N$ clusters with $N \leq 6$ carried out in the temperature range between 30 and 100 K;[44] and (d) constant pressure−constant temperature (NPT) molecular dynamics simulations of liquid water carried out at ambient conditions with intermediate versions of the present MB-pol potential. A large ensemble of configurations from the (b) and (c) sources described above was first formed, containing approximately an order of magnitude more trimers than were actually used in the final fitting process. These initial configurations were then filtered to remove redundant configurations, and only dissimilar geometries as determined from a root-mean squared deviation (RMSD) criterion were kept in the final ensemble.

Ideally, the perfect training set should comprise trimer configurations whose internal coordinates are defined on a multidimensional grid that uniformly covers all relevant regions of the configurational space. Unfortunately, given the dimensionality of the problem, this procedure cannot be performed for trimer configurations with flexible monomers. However, we expect that the structures extracted from PIMD trajectories and within neighborhoods of the stationary points of the trimer PES are sufficiently representative of molecular configurations with energies in the range relevant to molecular simulations of water under conditions of moderate temperature and pressure. In addition, within the protocol described above, the geometric uniformity of the training set is approximately achieved via the filtering step. The trimer geometries included in the final training set are reported in the Supporting Information along with the corresponding CCSD(T)/aug-cc-pVTZ reference energies.

3. DESCRIPTION OF THE MB-POL POTENTIAL

The rationale behind the MB-pol functional form has roots in the observation that, due to the different physical character of the interactions between water molecules at different separations, the intermolecular potential energy can be effectively split into short-range and long-range contributions.[45] The latter arise from electrostatic interactions between permanent and induced moments of the molecular charge distributions as well as from dispersion forces. The MB-pol representation of these (long-range) contributions to the total interaction energy is described in ref 29. Within the many-body expansion of the interaction energy, the short-range contributions can be introduced systematically at the n-body levels ($n \geq 2$) to account for the complexity of the multidimensional potential energy surface in regions where the monomer electron densities display significant overlap. The specific functional form of the MB-pol potential at small intermolecular separations is derived by noting that, for water systems, short-range corrections to the induction energy beyond the three-body level are not necessary for achieving chemical accuracy (see section 5.3). The two-body short-range term of the MB-pol potential is described in detail in ref 29. Minor modifications to the original $V^{(2B)}_{short}$ term were required to further improve both accuracy and performance of the full many-body MB-pol potential in applications to water systems larger than the dimer (see Appendix A). Similarly to HBB2-pol,[28] the three-body short-range term of the MB-pol potential is expressed as

$$V^{(3B)}(x_a, x_b, x_c) = V^{(3B)}_{short}(x_a, x_b, x_c) + V^{(3B)}_{TTM,ind}(x_a, x_b, x_c) \quad (4)$$

where $V^{(3B)}_{TTM,ind}$ is the TTM 3-body induction energy, and $V^{(3B)}_{short}$ is given by

$$V^{(3B)}_{short}(x_a, x_b, x_c) = [s(t_{ab})s(t_{ac}) + s(t_{ab})s(t_{bc}) + s(t_{ac})s(t_{bc})] V^{(3B)}_{poly}(x_a, x_b, x_c) \quad (5)$$

where the sum of the three terms in the square bracket represents a switching function that smoothly goes to zero as one of the water molecules moves apart from the other two. In eq 5,

$$s(t) = \begin{cases} 1 & \text{if} \quad t < 0 \\ \cos^2 \dfrac{\pi}{2} t & \text{if} \quad 0 \le t < 1 \\ 0 & \text{if} \quad 1 \le t \end{cases} \tag{6}$$

and

$$t_{mn} = \frac{R_{mn}}{R_{\text{cut}}^{(3B)}} \tag{7}$$

where R_{mn} is the distance between the oxygen atoms of water molecules m and n, and $R_{\text{cut}}^{(3B)} = 4.5$ Å is the three-body cutoff distance. This specific value for $R_{\text{cut}}^{(3B)}$ was found to provide the optimal compromise between accuracy and computational cost.

The $V_{\text{poly}}^{(3B)}$ term in eq 5 is a permutationally invariant polynomial in exponentials of the interatomic distances, $d_{m=1-36}$, between all possible pairs of atoms as defined in Table 1. From the definition of $d_{m=1-36}$, the following variables are formed

$$\xi_1 = e^{-k_{\text{HH,intra}}(d_1 - d_{\text{HH,intra}}^{(0)})},$$
$$\cdots$$
$$\xi_9 = e^{-k_{\text{OH,intra}}(d_9 - d_{\text{OH,intra}}^{(0)})},$$
$$\xi_{10} = e^{-k_{\text{HH}}(d_{10} - d_{\text{HH}}^{(0)})},$$
$$\cdots$$
$$\xi_{36} = e^{-k_{\text{OO}}(d_{36} - d_{\text{OO}}^{(0)})}.$$

$V_{\text{poly}}^{(3B)}$ is then constructed as a polynomial in ξ_i imposing the permutational invariance with respect to permutations of the water molecules as well as to permutations of the hydrogen atoms within each molecule. $V_{\text{poly}}^{(3B)}$ contains the following symmetrized monomials: (a) 13 s-degree monomials formed from all intermolecular ($\xi_{10},\ldots,\xi_{36}$) variables; (b) 202 third-degree symmetrized monomials with at most linear intramolecular terms; and (c) 948 symmetrized fourth-degree monomials with at most linear dependence on the intramolecular variables, as well as intermolecular variables involving oxygen−oxygen and hydrogen−hydrogen distances. The complete list of all 1163 symmetrized monomials is available in the Supporting Information. The coefficients ($c_{l=1-1163}$) of these terms are linear fitting parameters. In addition, $V_{\text{poly}}^{(3B)}$ also contains the following 10 nonlinear parameters: $d_{\text{HH,intra}}^{(0)}$, $d_{\text{OH,intra}}^{(0)}$, $d_{\text{HH}}^{(0)}$, $d_{\text{OH}}^{(0)}$, $d_{\text{OO}}^{(0)}$, $k_{\text{HH,intra}}$, $k_{\text{OH,intra}}$, k_{HH}, k_{OH}, and k_{OO}.

Both linear and nonlinear parameters were obtained by minimizing the (regularized) weighted sum of squared residuals calculated for the short-range training set $\mathcal{J}$ described in section 2.2:

$$\chi^2 = \sum_{n \in \mathcal{J}} w_n [V_{\text{model}}^{(3B)}(n) - V_{\text{ref}}^{(3B)}(n)]^2 + \Gamma^2 \sum_{l=1}^{1163} c_l^2$$

The weights, w_n, were set to emphasize trimers with low total energies

Table 1. Distances Entering the Short-Range Three-Body Part of the Potential[a]

d_1	Ha1	Ha2
d_2	Hb1	Hb2
d_3	Hc1	Hc2
d_4	Oa	Ha1
d_5	Oa	Ha2
d_6	Ob	Hb1
d_7	Ob	Hb2
d_8	Oc	Hc1
d_9	Oc	Hc2
d_{10}	Ha1	Hb1
d_{11}	Ha1	Hb2
d_{12}	Ha1	Hc1
d_{13}	Ha1	Hc2
d_{14}	Ha2	Hb1
d_{15}	Ha2	Hb2
d_{16}	Ha2	Hc1
d_{17}	Ha2	Hc2
d_{18}	Hb1	Hc1
d_{19}	Hb1	Hc2
d_{20}	Hb2	Hc1
d_{21}	Hb2	Hc2
d_{22}	Oa	Hb1
d_{23}	Oa	Hb2
d_{24}	Oa	Hc1
d_{25}	Oa	Hc2
d_{26}	Ob	Ha1
d_{27}	Ob	Ha2
d_{28}	Ob	Hc1
d_{29}	Ob	Hc2
d_{30}	Oc	Ha1
d_{31}	Oc	Ha2
d_{32}	Oc	Hb1
d_{33}	Oc	Hb2
d_{34}	Oa	Ob
d_{35}	Oa	Oc
d_{36}	Ob	Oc

[a]The first letter of the site label denotes the atom name (O, H), the second letter distinguishes the molecules (a, b, or c), the trailing digit indexes equivalent atoms within the molecule.

$$w(E) = \left(\frac{\Delta E}{E - E_{\text{min}} + \Delta E} \right)^2 \tag{8}$$

Here, E_{min} denotes the lowest energy in the training set (i.e., trimer global minimum energy) and ΔE defines the range of the favorably weighted energies. $\Delta E = 37.5$ kcal/mol was used in eq 8, which is consistent with the value used in the fit of the two-body term (25 kcal/mol).[29] The regularization parameter, Γ, was set to 1×10^{-4} in order to reduce the variation of the linear fitting parameters (larger Γ values effectively suppress any variation of the parameters) without spoiling the overall accuracy of the fit (favored by smaller Γ values). The choice of the regularization weight was further constrained by the requirement that the regularization term contributed to no more than 1% to χ^2. The linear parameters ($c_{l=1,\ldots,1163}$) were obtained through singular value decomposition while the simplex algorithm was used to iteratively optimize the nonlinear parameters. The optimization procedure results in a χ^2 value of 7.83 (kcal/mol)^2, corresponding to a RMSD of 0.028 kcal/mol per trimer. The RMSD for trimers with total energy within 37.5

dx.doi.org/10.1021/ct500079y | J. Chem. Theory Comput. 2014, 10, 1599−1607

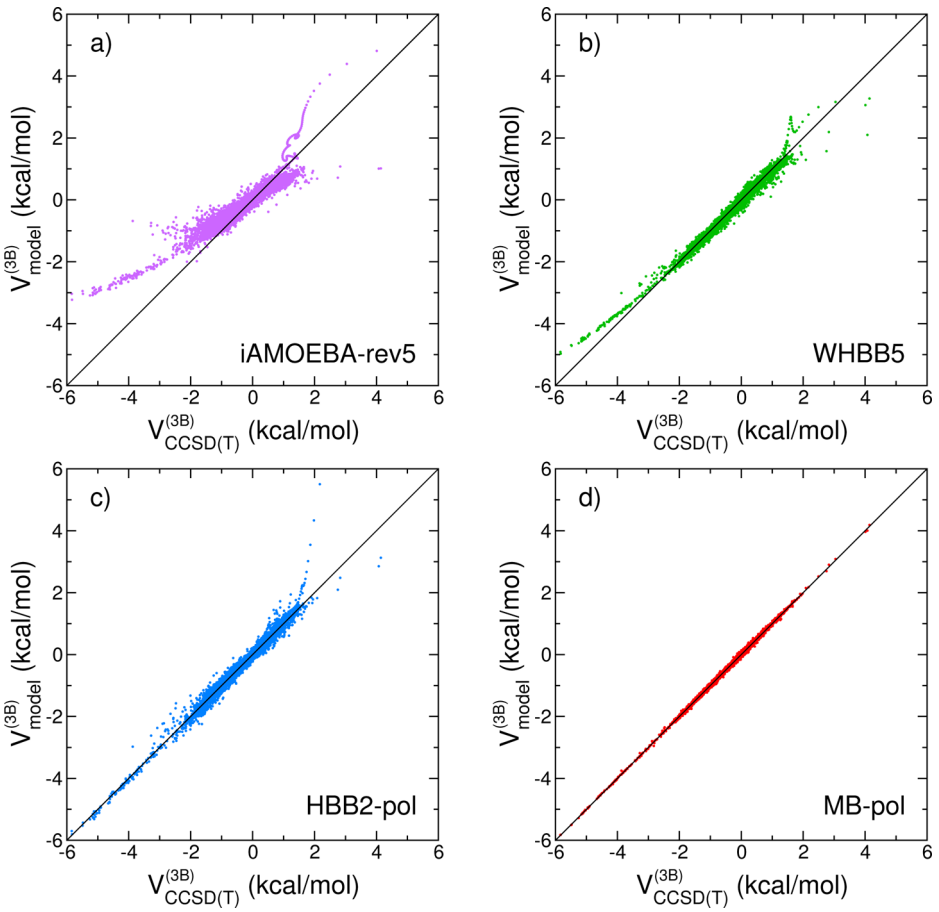


Figure 1. Correlation plots showing three-body energies obtained from the analysis of 12347 water trimers. Plotted on the *x*-axis are the BSSE-corrected CCSD/aug-cc-pVTZ+mb values while on the *y*-axis are the corresponding values calculated with (a) iAMOEBA-rev5, (b) WHBB5, (c) HBB2-pol, and (d) MB-pol. See main text for details.

kcal/mol of the trimer minimum energy is 0.00068 kcal/mol per trimer and the largest deviation for the same trimers is 0.16 kcal/mol. All optimized parameters are included in the reference implementation of the MB-pol potential available in the Supporting Information.

4. ASSESSMENT OF THE MB-POL POTENTIAL

The CCSD(T) reference energies described in section 2 are used in Figure 1 to investigate the accuracy of MB-pol three-body energies compared to the corresponding results obtained with the iAMOEBA-rev5,[35] WHBB,[26] and HBB2-pol[27,28] potentials, all of which have recently been shown to correctly reproduce some of the water properties under different conditions. The CC-pol model is omitted from this comparison since a flexible version of it has only been developed at the two-body level,[46] which prevents a direct comparison with the present CCSD(T) data calculated for flexible trimers. Simpler water potentials, parametrized to experimental and/or to relatively small sets of *ab initio* data (e.g., q-TIP4P/F,[47] qSPC/Fw,[48] TTM3-F,[49] and E3B[50]), are also not considered in the present study because it was previously shown that these models are unable to correctly reproduce both two-body and three-body water energies.[27]

Among the four water potentials examined in Figure 1, MB-pol provides the closest agreement with the CCSD(T) reference data. Noticeable differences exist between iAMOEBA-rev5 (RMSD = 0.313 kcal/mol) and *ab initio* results, especially for three-body energies below −2 kcal mol^{-1}. Although the HBB2-pol potential (RMSD = 0.158 kcal/mol) appears to perform slightly better than WHBB (RMSD = 0.104 kcal/mol) for low-energy trimers, some important deficiencies are identified, which, based on a preliminary analysis of the liquid properties, can be attributed to trimer geometries that are associated with the low-density phase of water. For these configurations, which correspond to the "tail" deviating from the diagonal in the upper-right corner of Figure 1c, HBB2-pol predicts considerably more repulsive three-body energies than CCSD(T). The energetics of these structures is analyzed in more detail in section 5.

5. VALIDATION

5.1. Stationary Points. The total energies and corresponding gradients calculated at the six stationary points of the *ab initio* water trimer PES reported in ref 51 are listed in Table 2. The same quantities for trimer geometries optimized on the MB-pol PES are given in Table 3 along with the RMSD values calculated relative to the *ab initio* configurations from ref 51. Both sets of results demonstrate the high accuracy of the MB-pol trimer energies. Importantly, the MB-pol gradients associated with the reference geometries are relatively small, providing further evidence for the closeness between the MB-pol and *ab initio* PESs. The remaining disagreement between the MB-pol and *ab initio* PESs can be attributed, at least partially, to the differences in the corresponding one-body

dx.doi.org/10.1021/ct500079y | J. Chem. Theory Comput. 2014, 10, 1599−1607

Table 2. MB-pol Binding Energies (E), Gradient Norms ($|\nabla E|_2$), and Relative Energies (ΔE) Calculated for the Reference Stationary Points of the Water Trimer PES Reported in Ref 51[a]

no.	E (cm^{-1})	$\|\nabla E\|_2$ (cm^{-1}/Å)	ΔE (cm^{-1})	ΔE_{ref} (cm^{-1})
1	−5460.57	1.07×10^3	0.00	0.00
2	−5210.91	1.12×10^3	249.67	268.34
3	−5016.46	1.08×10^3	444.12	436.36
4	−5378.82	1.05×10^3	81.75	81.09
5	−5201.36	1.09×10^3	259.22	274.19
6	−4701.06	1.10×10^3	759.51	752.34

[a]Also listed are the corresponding *ab initio* relative energies (ΔE_{ref}).[51] The RMSD of the MB-pol ΔE values relative to ΔE_{ref} is 10.68 cm^{-1}.

Table 3. MB-pol Binding Energies (E), Gradient Norms ($|\nabla E|_2$), RMSD with Respect to the Reference *Ab Initio* Geometries,[51] and Relative Energies (ΔE) Calculated for the Stationary Points of the MB-pol Water Trimer PES[a]

no.	E (cm^{-1})	$\|\nabla E\|_2$ (cm^{-1}/Å)	RMSD (Å)	ΔE (cm^{-1})	ΔE_{ref} (cm^{-1})
1	−5488.26	4.67×10^{-5}	8.96×10^{-3}	0.00	0.00
2	−5239.88	5.16×10^{-5}	8.73×10^{-3}	248.39	268.34
3	−5040.52	9.14×10^{-5}	5.49×10^{-3}	447.75	436.36
4	−5405.15	9.75×10^{-5}	8.71×10^{-3}	83.11	81.09
5	−5228.84	5.01×10^{-5}	9.01×10^{-3}	259.43	274.19
6	−4729.79	8.29×10^{-5}	9.35×10^{-3}	758.47	752.34

[a]Also listed are the reference *ab initio* relative energies (ΔE_{ref}) for the reference stationary points of the water trimer PES reported in ref 51. The RMSD of the MB-pol ΔE values relative to ΔE_{ref} is 9.35 cm^{-1}.

energies. As described in ref 29, the monomer PES employed in MB-pol includes corrections aimed at representing mass-polarization and relativistic effects,[34] which are absent in the CCSD(T) calculations reported in ref 51.

In an attempt to eliminate the "bias" associated with the different treatment of the one-body term, the comparison between MB-pol and CCSD(T) trimer interaction energies is repeated including only two- and three-body contributions $[E_3(x_1,x_2,x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)]$. For this purpose, the CCSD(T) reference values were obtained in the complete basis set limit using a two-point extrapolation of the CCSD(T) energies calculated with aug-cc-pVTZ and aug-cc-pVQZ basis sets supplemented by either one midbond function placed at the center of mass of each trimer (Table 4) or four midbond functions placed at the centers of mass of the trimer and all

Table 4. Interaction Energies ($E_3(x_1, x_2, x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)$, in cm^{-1}) for the Reference Trimer Geometries (Ref 51)[a]

no.	CCSD(T)/ aVTZ+mb (cm^{-1})	CCSD(T)/ aVQZ+mb (cm^{-1})	CCSD(T)/ CBS (cm^{-1})	MB-pol (cm^{-1})	Δ (cm^{-1})
1	−5436	−5596	−5713	−5662	51
2	−5167	−5322	−5435	−5402	33
3	−4965	−5113	−5222	−5185	37
4	−5349	−5505	−5620	−5574	46
5	−5158	−5312	−5425	−5390	35
6	−4715	−4841	−4933	−4883	50

[a]Here, "mb" corresponds to one set of midbond functions placed at the center of mass of the trimer. The RMSD of the MB-pol energies with respect to the CCSD(T)/CBS results is 42.59 cm^{-1}.

constituent dimers (Table 5). After removal of the differences associated with the one-body term, the MB-pol energies for the

Table 5. Interaction Energies ($E_3(x_1, x_2, x_3) - E_1(x_1) - E_1(x_2) - E_1(x_3)$, in cm^{-1}) for the Reference Trimer Geometries (Ref 51)[a]

no.	CCSD(T)/ aVTZ+mb[4] (cm^{-1})	CCSD(T)/ aVQZ+mb[4] (cm^{-1})	CCSD(T)/ CBS (cm^{-1})	MB-pol (cm^{-1})	Δ (cm^{-1})
1	−5514	−5608	−5677	−5662	15
2	−5259	−5349	−5415	−5402	13
3	−5073	−5159	−5221	−5185	36
4	−5431	−5523	−5591	−5574	17
5	−5251	−5340	−5405	−5390	15
6	−4724	−4801	−4857	−4883	−26

[a]Here, "mb[4]" corresponds to four sets of midbond functions placed at the centers of mass of the trimer and its three constituent dimers. The RMSD of the MB-pol energies with respect to the CCSD(T)/CBS results is 21.92 cm^{-1}.

six reference trimers lie at most within 50 cm^{-1} of the corresponding CCSD(T) energies. The agreement of MB-pol with the CCSD(T)/CBS energies is improved when larger basis sets are used in the CBS extrapolation; the RMS of the MB-pol energies with respect to the CCSD(T)/CBS energies for the trimer stationary points decreases from 42.59 cm^{-1} to 21.92 cm^{-1} when midbond functions are added to the centers of mass of both the trimer and its three constituent dimers. It is also interesting to note that the CCSD(T)/aug-cc-pVQZ+mb energy computed for the optimized trimer structure obtained on the MB-pol PES is lower than the energy associated with the optimized structure of ref 51.

The accuracy of the potential is further illustrated in Figure 3 showing the comparison between the MB-pol and *ab initio* three-body interaction energies for a representative scan through the water trimer PES. The scan shown in Figure 2

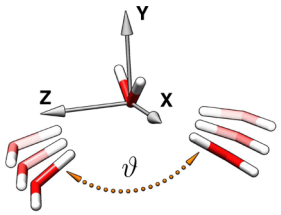


Figure 2. Schematic representation of the water trimer configurations used in the scan shown in Figure 3.

was investigated because it involves trimer configurations resembling those extracted from simulations of liquid water, which were found to be critical for reproducing the temperature dependence of the density at ambient pressure. Both coordinates and energies of the trimer configurations used in the scan of Figure 2 are available in the Supporting Information. In the scan calculations of Figure 2, all three water molecules were kept in the vibrationally averaged geometry (r_{OH} = 1.836106337 Bohr and ϑ_{HOH} = 104.69° as defined in ref 25) so that the CCpol[36] potential could also be included in the comparison. Also shown in Figure 2 is a modified version (TTM4x-F) of the original TTM4-F polarizable water model,[31] which employs the same many-body polarization scheme as MB-pol.[29] In all trimer geometries used in the scan, the "central" molecule is placed on the XY plane

dx.doi.org/10.1021/ct500079y | J. Chem. Theory Comput. 2014, 10, 1599−1607

with its oxygen atom at the origin and the HOH bisector lie along the *y*-axis; the "left" and "right" water molecules lie on the XZ plane with their oxygen atoms located on the *z*-axis at a distance of 2.7 Å from the central water oxygen. Both "left" and "right" molecules have one hydrogen atom on the *z*-axis so that the corresponding OH bond is oriented toward the central oxygen atom. The potential energy scan was performed by rotating the "left" and "right" molecules about the *x*-axis by 180 − $\vartheta/2$ where ϑ is the angle between the oxygen atoms of the "left", "central", and "right" water molecules. The results shown in Figure 3 indicate that both TTM4x-F and HBB2-pol predict

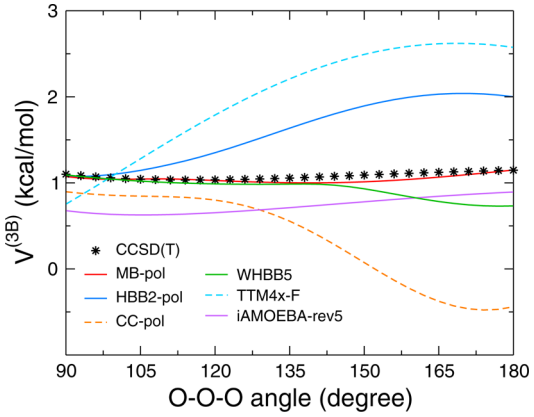


Figure 3. Three-body interaction energies for the scan through the water trimer PES shown in Figure 2. The reference CCSD(T)/aug-cc-pVTZ values (black stars) are shown along with the MB-pol (red), HBB2-pol (blue), CC-pol (dashed orange), WHBB5 (green), TTM4x-F (dashed light blue), and iAMOEBA (violet) results.

more repulsive three-body energies, while CC-pol underestimates the *ab initio* values, predicting significantly more attractive three-body energies for wider O−O−O angles. Interestingly, iAMOEBA-rev5 correctly reproduces the angular dependence of the three-body energy along the entire scan, although it underestimates the magnitude of this contribution. Good agreement with the reference CCSD(T) data is provided by both WHBB and MB-pol, which is effectively quantitative for MB-pol over the entire scan.

5.2. Third Virial Coefficient. By accounting for deviations from ideal gas behavior,[53] the third virial coefficient provides an indirect measure of the three-body interactions. To assess the overall accuracy of the MB-pol three-body term, the third virial coefficient was calculated at the classical level with each water monomer held fixed at the ground-state vibrationally averaged geometry (r_{OH} = 1.836106337 Bohr and ϑ_{HOH} = 104.69°).[25] This classical approach is justified by the fact that the experimental data for the third virial coefficient are only available at relatively high temperatures, where quantum effects are less important. All calculations were carried out using the scheme summarized in ref 28. The close agreement between the MB-pol results and the available experimental data[54] shown in Figure 4 provides further support for the accuracy of the MB-pol three-body term.

5.3. Higher-Order Contributions to the Interaction Energy. It has been shown that *n*-body interactions with $n \geq 4$ are relatively small for water systems.[11,18−24] By construction, these nonpair-additive interactions are described in the MB-pol potential through the $V_{TTM,ind}$ induction energy term in eq 4. The analysis of the four-body interactions shown in Figure 5

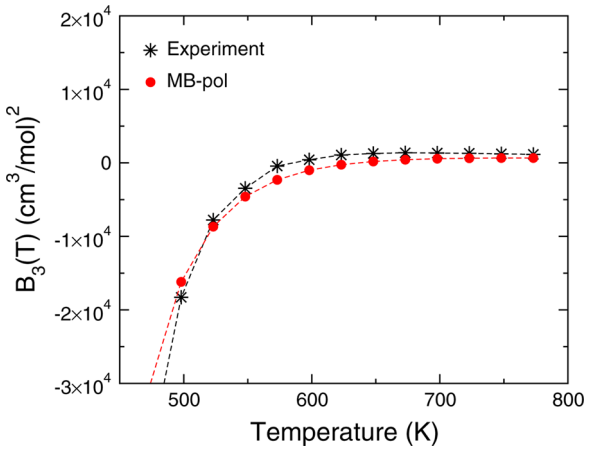


Figure 4. MB-pol third virial coefficient compared with the corresponding experimental data from ref 54.

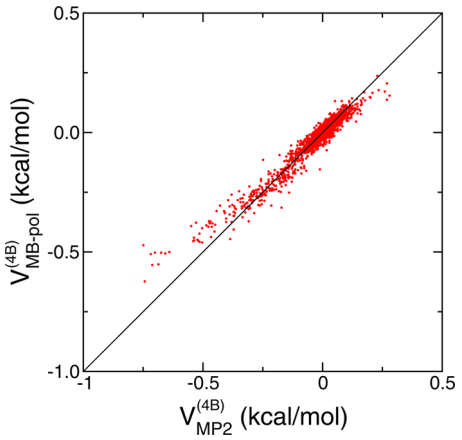


Figure 5. Correlation plot for the four-body interaction energies obtained from the analysis of 2674 water tetramers. Plotted on the *x*-axis are the BSSE-corrected MP2/aug-cc-pVTZ+mb values while on the *y*-axis are the corresponding MB-pol values.

thus enables a quantitative assessment of the accuracy of MB-pol in describing higher-order many-body contributions. For this analysis, 2674 tetramer configurations were extracted (1) from PIMD simulations of both cage (881 tetramers) and prism (925 tetramers) isomers of the water hexamer carried out with the HBB2-pol potential[44] and (2) from PIMD simulations of liquid water (868 tetramers) carried out with the HBB2-pol potential.[27] In both cases, a significantly larger number of tetramer configurations was initially selected and only geometries that were separated by a uniformly weighted RMSD ≥ 0.2 Å were included in the final set. The four-body reference interaction energies were computed at the MP2 level of theory using the aug-cc-pVTZ basis set[38,39] supplemented with midbond functions[40] placed at the center of mass of each tetramer. The interaction energies were evaluated via the counterpoise method using the energies of the constituent fragments computed in the full tetramer basis.[41,42] As shown in Figure 5, the mean absolute value of the four-body interaction is ∼0.06 kcal/mol. The RMSD associated with the MB-pol four-body energies is 0.03 kcal/mol, with the largest deviation being 0.392 kcal/mol (the mean total energy of these tetramer configurations relative to the tetramer global minimum is ∼42.89 kcal/mol).

dx.doi.org/10.1021/ct500079y | J. Chem. Theory Comput. 2014, 10, 1599−1607

5.4. Energetics of Small Water Clusters. The relative energies (with respect to the corresponding global minima) of small water clusters with more than three molecules allows for a further assessment of the ability of MB-pol to correctly describe n-body interactions with $n > 3$. For this purpose, the relative energies of the low-lying isomers of the water tetramer, pentamer, and hexamer calculated with the iAMOEBA-rev5, WHBB, HBB2-pol, and MB-pol potentials are compared in Figure 6 with the corresponding *ab initio* values reported in the literature.[55,56] Water clusters in this size range are among the largest ones for which highly correlated electronic structure calculations are still feasible. All water potentials reproduce the *ab initio* data semiquantitatively, with iAMOEBA-rev5, which was specifically parametrized to reproduce the relative energies of small water clusters, providing the closest agreement. Except for the CA-B isomer of the water pentamer, MB-pol reproduces (within chemical accuracy) the energy ordering of all cluster isomers as predicted by the *ab initio* calculations. In this regard, it is important to note that all *ab initio* data were calculated for geometries optimized at the MP2 level, which may also be source of some discrepancies with predictions made by potentials that were parametrized at the CCSD(T) level as WHBB, HBB2-pol, and MB-pol. Further investigation of these differences will be the subject of future work.

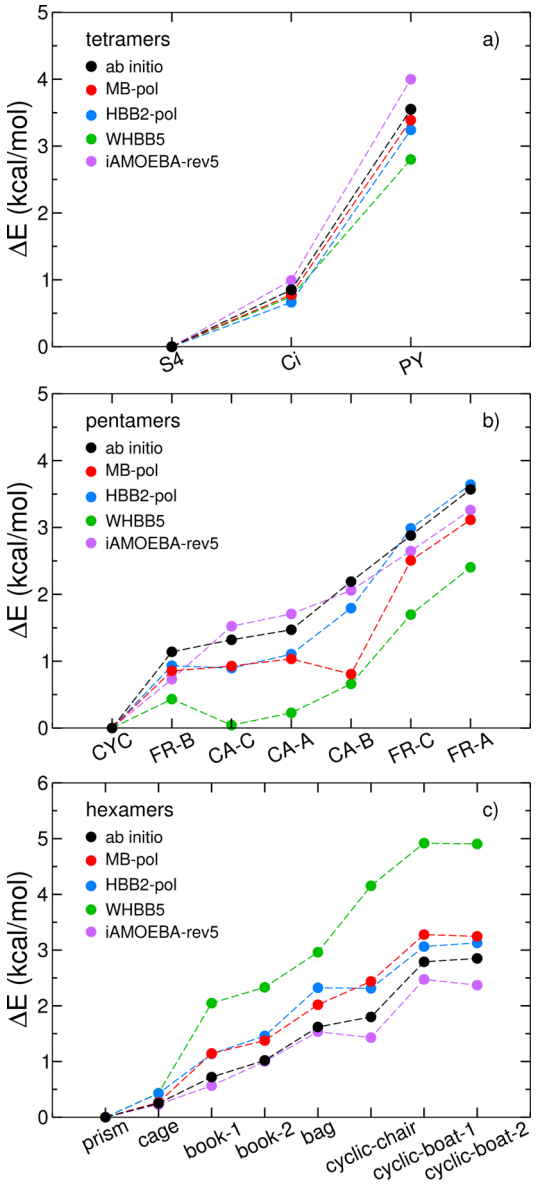


Figure 6. Relative energies (in kcal/mol) of the water tetramer, pentamer and hexamer isomers calculated with the MB-pol (red), HBB2-pol (blue), WHBB5 (green), and iAMOEBA-rev5 (violet) potentials. The reference *ab initio* values from refs 55 and 56 are shown in black.

6. SUMMARY

The full-dimensional many-body MB-pol water potential derived entirely from "first principles" is presented, with a specific emphasis on the three-body contribution. The latter is constructed through the combination of the three-body induction energy, which dominates at long-range, with an explicit short-range correction that accounts for the complexity of the underlying multidimensional potential energy surface in regions where the electron densities of the individual monomers overlap significantly. The parameters entering the potential were determined by fitting more than 12000 three-body energies calculated at the CCSD(T)/aug-cc-pVTZ+mb level of theory. The accuracy of higher-body contributions to the total interaction energy is investigated through the comparison of MB-pol and MP2/aug-cc-pVTZ+mb four-body energies computed for 2674 water tetramers. The accuracy of the MB-pol potential is further validated through the analysis of both energies and geometries corresponding to the stationary points on the trimer potential energy surface, third virial coefficient, and energetics of small water clusters. In forthcoming studies, we will report on finite-temperature molecular simulations of both water clusters and condensed-phase systems.

APPENDIX A: TWO-BODY FIT

The original representation of the two-body term presented in ref 29 has been slightly modified here to further improve both accuracy and performance of the full many-body MB-pol potential in applications to water systems larger than the dimer. Specifically, three minor changes were adopted: (a) all $1/r^8$ terms have been removed from the expression of the dispersion energy; (b) for consistency with the three-body term, the switching function [eq 8 in ref 29] has been replaced by the following expression

$$s_2(x) = \begin{cases} 1 & \text{if} \quad x < 0 \\ \cos^2 \dfrac{\pi}{2} x & \text{if} \quad 0 \le x < 1 \\ 0 & \text{if} \quad 1 \le x \end{cases} \tag{9}$$

and (c) the corresponding cutoff distances have been reduced to $R_{i,o}$ = 4.5/6.5 Å. Both long- and short-range parameters have been determined as described in ref 29 using the same reference energies. The new short-range fit is marginally more accurate than the original one despite the reduced cutoffs. The differences between the new and old two-body interaction energies are shown in Figure 7 as a function of the dimer total energy for configurations taken from the short-range training set.

dx.doi.org/10.1021/ct500079y | J. Chem. Theory Comput. 2014, 10, 1599−1607

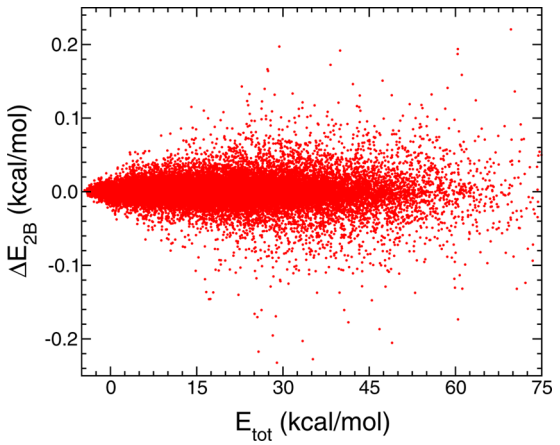


Figure 7. Differences between two-body interaction energies calculated with the current and original[29] MB-pol two-body terms as a function of the dimer total energy. The RMSD over the lower 100 kcal/mol is 0.017 kcal/mol.

■ ASSOCIATED CONTENT

Supporting Information

(1) Training set of 12347 water trimers, including both geometries and CCSD(T)/aug-cc-pVTZ+mb three-body energies. (2) List of the symmetrized monomials that enter $V_{poly}^{(3B)}$ in eq 5. (3) Reference implementation of the MB-pol potential. (4) Trimer geometries along the scan shown in Figure 3. (5) Trimer stationary points optimized on the MB-pol PES. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: fpaesani@ucsd.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by National Science Foundation through grant CHE-1111364. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number OCI-1053575 (allocation TG-CHE110009).

■ REFERENCES

(1) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(2) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.
(3) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177−193.
(4) Chang, D. T.; Schenter, G. K.; Garrett, B. C. *J. Chem. Phys.* **2008**, *128*, 164111.
(5) Murdachaew, G.; Mundy, C. J.; Schenter, G. K.; Laino, T.; Hutter, J. *J. Phys. Chem. A* **2011**, *115*, 6046−53.
(6) Del Ben, M.; Schönherr, M.; Hutter, J.; VandeVondele, J. *J. Phys. Chem. Lett.* **2013**, *4*, 3753−3759.
(7) Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. *Phys. Rev. B* **2013**, *88*, 054104.
(8) Wen, S.; Beran, G. J. O. *J. Chem. Theory Comput.* **2011**, *7*, 3733−3742.
(9) Zhang, C.; Donadio, D.; Gygi, F.; Galli, G. *J. Chem. Theory Comput.* **2011**, *7*, 1443−1449.
(10) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
(11) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.
(12) Zhang, C.; Wu, J.; Galli, G.; Gygi, F. *J. Chem. Theory Comput.* **2011**, *7*, 3054−3061.
(13) Møgelhøj, A.; Kelkkanen, A. K.; Wikfeldt, K. T.; Schiøtz, J.; Mortensen, J. J.; Pettersson, L. G. M.; Lundqvist, B. I.; Jacobsen, K. W.; Nilsson, A.; Nørskov, J. K. *J. Phys. Chem. B* **2011**, *115*, 14149−60.
(14) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Theory Comput.* **2012**, *8*, 1929−1934.
(15) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.
(16) Stone, A. J. *Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1997.
(17) ref 16, Chapter 9.
(18) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523−7534.
(19) Pedulla, J. M.; Vila, F.; Jordan, K. D. *J. Chem. Phys.* **1996**, *105*, 11091.
(20) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163−9168.
(21) Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225−231.
(22) Cui, J.; Liu, H.; Jordan, K. D. *J. Phys. Chem. B* **2006**, *110*, 18872−18878.
(23) Hermann, A.; Krawczyk, R.; Lein, M.; Schwerdtfeger, P.; Hamilton, I.; Stewart, J. J. P. *Phys. Rev. A* **2007**, *76*, 013202.
(24) Khaliullin, R.; Cobar, E.; Lochan, R.; Bell, A.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15328−15339.
(25) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van Der Avoird, A. *J. Chem. Phys.* **2008**, *128*, 094313.
(26) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.
(27) Babin, V.; Medders, G. R.; Paesani, F. *J. Phys. Chem. Lett.* **2012**, *3*, 3765−3769.
(28) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 1103−1114.
(29) Babin, V.; Leforestier, C.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 5395−5403.
(30) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341−350.
(31) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. *J. Chem. Phys.* **2008**, *128*, 154519.
(32) Braams, B. J.; Bowman, J. M. *Int. Rev. Phys. Chem.* **2009**, *28*, 577−606.
(33) Xie, Z.; Bowman, J. M. *J. Chem. Theory Comput.* **2010**, *6*, 26−34.
(34) Partridge, H.; Schwenke, D. W. *J. Chem. Phys.* **1997**, *106*, 4618.
(35) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956−72.
(36) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249−52.
(37) Raghavachari, K.; Truck, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479−483.
(38) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(39) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(40) Tao, F.-M.; Pan, Y.-K. *J. Chem. Phys.* **1992**, *97*, 4989−4995.
(41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(42) Wells, B. H.; Wilson, S. *Chem. Phys. Lett.* **1983**, *101*, 429−434.
(43) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, version 2012.1, a package of ab initio programs. 2012; http://www.molpro.net (accessed March 7, 2014).
(44) Babin, V.; Paesani, F. *Chem. Phys. Lett.* **2013**, *580*, 1−8.
(45) ref 16., Chapter 5.

(46) Leforestier, C.; Szalewicz, K.; van der Avoird, A. *J. Chem. Phys.* **2012**, *137*, 014305.
(47) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. *J. Chem. Phys.* **2009**, *131*, 024501.
(48) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A. *J. Chem. Phys.* **2006**, *125*, 184507.
(49) Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.
(50) Kumar, R.; Skinner, J. L. *J. Phys. Chem. B* **2008**, *112*, 8311−8318.
(51) Anderson, J. A.; Crager, K.; Fedoroff, L.; Tschumper, G. S. *J. Chem. Phys.* **2004**, *121*, 11023−11029.
(52) Hill, T. *An Introduction to Statistical Thermodynamics*; Dover: Mineola, NY, 1986.
(53) ref 52., Chapter 15.
(54) Kell, G.; McLaurin, G.; Whalley, E. *Proc. R. Soc. Lond. A* **1989**, *425*, 49−71.
(55) Temelso, B.; Archer, K. A.; Shields, G. C. *J. Phys. Chem. A* **2011**, *115*, 12034−12046.
(56) Bates, D. M.; Tschumper, G. S. *J. Phys. Chem. A* **2009**, *113*, 3555−3559.