Assessment of the Pairwise Additive Approximation and Evaluation of Many-Body Terms for Water Clusters

Erin E. Dahlke and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received: February 17, 2006; In Final Form: April 25, 2006

We have tested the ability of four commonly used density functionals (three of which are semilocal and one of which is nonlocal) to outperform accurate pairwise additive approximations in the prediction of binding energies for a series of water clusters ranging in size from dimer to pentamer. Comparison to results obtained with the Weizmann-1 (W1) level of wave function theory shows that while all density functionals are capable of outperforming the accurate pairwise data, the choice of basis set used is crucial to the performance of the method, and if a poor choice of basis set is made the errors obtained with density functional theory (DFT) can be larger than those obtained with the simple pairwise approximation. We have also compared the binding energies and many-body terms determined with DFT to those obtained with W1, and have found that all semilocal functionals have significant errors in the many-body components of the full interactions energy. Despite this limitation, however, we have found that, of the four functionals tested, PBE1W/MG3S is the most accurate for predicting the binding energies of the clusters.

1. Introduction

For a system of N interacting particles, one can write the total potential energy of the system as a series of n-body potentials:[1]

$$V(r_1,r_2,...,r_N) = \sum_{i<j}^{N} V_2(r_i,r_j) + \sum_{i<j<k}^{N} V_3(r_i,r_j,r_k) + ... + \sum_{i<j<k,...,n}^{N} V_n(r_i,r_j,r_k,...,r_n)$$

(1)

where V_n is the n-body interaction term. Truncation of this series after the first term leads to the pairwise additive approximation; that is, the total potential is a sum of all two-body interactions in the system. While such an approximation can often provide qualitatively correct results, the many-body terms ($n > 2$) are often necessary to achieve quantitative accuracy.[1] Despite this limitation, though, pairwise potentials are widely invoked both because of an incomplete understanding of the role of many-body effects and also because of the ease with which pairwise methods can be used.

The interest in simulating bulk water and ice has led to the development of a number of analytic potential functions designed specifically for water. The complexity of these potentials ranges from simple multisite fixed-charge models, such as the TIPnP ($n = 3,4,5$) family of potentials[2] and the SPC/E model,[3] to more complicated potentials that introduce fluctuating charges[4−7] or polarization terms.[6,8−13] These potentials are generally parameterized in order to account as fully as possible for many-body effects in an average way. Recently, however, direct dynamics[14] and direct Monte Carlo[15] techniques utilizing density functional theory (DFT) have also been used to investigate the properties of condensed-phase water. The use of DFT instead of an analytic potential to calculate the energy of the system greatly increases the cost of the calculation, but the added cost is generally rationalized by the belief that the energies obtained are far more accurate than those obtained with simple pair potentials. While in principle the use of DFT to evaluate the energy of the system should give far more accurate results, the increased cost often results in the use of smaller simulation boxes, and compromises are made in both the size of basis set used and the length of the simulation. If the results are not well converged with respect to all of these issues, serious errors can be introduced that may render the simulations less accurate than if a simple pair potential had been used. Such concerns may lead one to question whether the use of ab initio simulation methods is worth the substantial increase in cost, at least for pure liquids without autoionization or reactive solutes. Furthermore, even some recent simulations of bulk water[15] that are well converged with respect to such numerical parameters have yielded results far worse than those obtained from the simulations with analytic potentials.

Past work by Hodges et al.[16] evaluated the accuracy, as compared to the MP2 level of theory, of two analytic potentials for predicting the pairwise and many-body terms for a set of water clusters, and they used this work to rationalize the use of these potentials for simulation. Here we carry out a similar analysis of density functional theory in comparison to highly accurate data from wave function theory. We will address (i) whether DFT is able to outperform accurate, pairwise data for a set of water clusters, ranging in size from trimer to pentamer and (ii) whether DFT is able to predict separately the pairwise and many-body terms obtained from high-level wave function calculations.
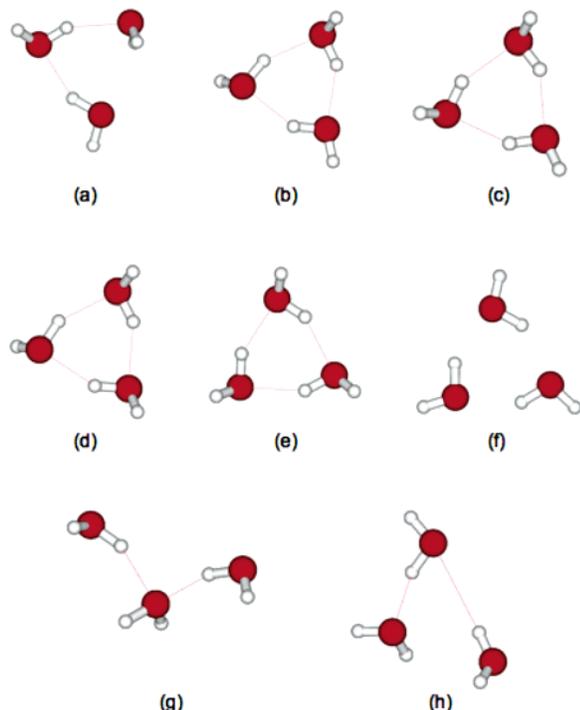
10596 *J. Phys. Chem. B, Vol. 110, No. 22, 2006*

Letters

Figure 1. Trimer water clusters: (a−f) optimized gas-phase clusters; (g,h) taken from Monte Carlo simulations of bulk water.

2. Pairwise Additive Database

The pairwise additive database used in the present study contains eight trimers, six tetramers, and a single pentamer. The trimers used in this study are the same eight as described in previous work.[17] The tetramers can be broken down into three classes of structures: branched, cyclic, and linear. One tetramer from each class was obtained from a Monte Carlo simulation of bulk water in the NPT ensemble.[15] The second geometries for both the branched and cyclic tetramers were optimized at the MP2/MG3S (where MP2 denotes Møller−Plesset second-order perturbation theory[18] and the MG3S[19] basis set is equivalent to 6-311+G(2df,2p)[20] for water) level of theory, and they were verified to be local minima by harmonic frequency analysis. The second linear tetramer was taken from a molecular dynamics simulation of ice VIII.[21] The pentamer structure was taken from the same Monte Carlo simulation as the three tetramers. The structures for the clusters are shown in Figures 1 and 2, and the Cartesian coordinates can be found in Supporting Information.

After selection of the structures to be used in this study, each cluster was broken down into its smaller, constituent clusters. For example, the pentamer was broken down into five tetramers, ten trimers, ten dimers, and five monomers, with the tetramers and trimers broken down in a similar manner. Accurate energies for all structures were obtained using the Weizmann-1[22,23] (W1) level of wave function theory with fixed geometries (i.e., omitting the B3LYP geometry optimization in the W1 protocol, see refs 22 and 23 for more information), performed with the MOLPRO[24] quantum chemistry package.

We define the binding energy of a cluster containing N water molecules as

$$E_{bind} = E_A + E_B + \cdots + E_N - E_{AB\cdots N} \quad (2)$$

For this work, all binding energies are taken relative to the unrelaxed monomers. For those readers who would like to compare this work to that found in ref 17, we note that two
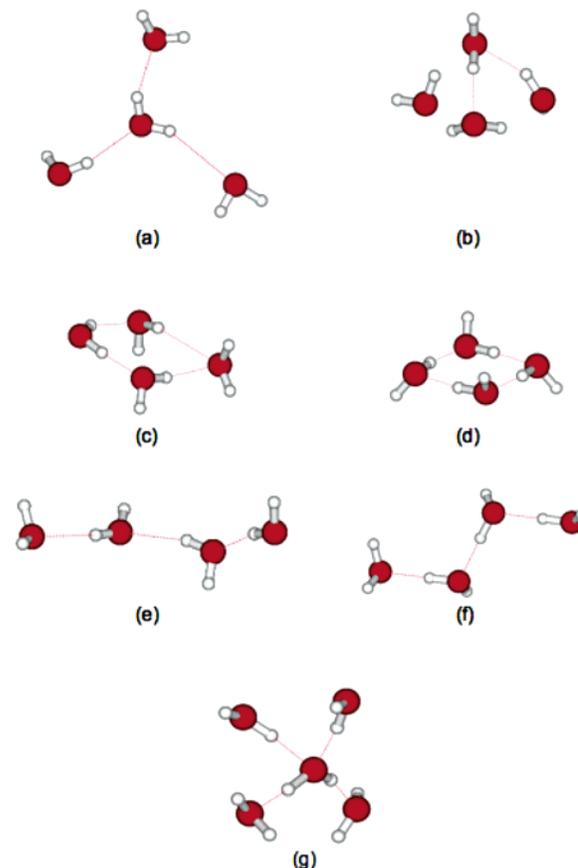

Figure 2. Tetramer and pentamer water clusters: (a,c,e,g) taken from Monte Carlo simulations of bulk water; (b,d) gas-phase optimized structures; (f) taken from a molecular dynamics simulation of ice VIII.

changes have been made with regard to the trimer data. First, in this work the W1 level of theory is used to compute accurate energies for all of the trimers and their constituent structures; previously the accurate energies for a subset of the trimers were obtained using focal point analysis, as described in the paper[25] from which the structures were taken. Second, the binding energies for all trimers are calculated relative to the unrelaxed monomers, whereas previously some were calculated relative to the relaxed gas-phase monomer. The reader is referred to ref 17 for more information on both of these points.

For each cluster, the accurate pairwise additive energy is given by taking a sum of the Weizmann-1 binding energies for the constituent dimers. The accurate three-body potential term from eq 1 is given by

$$V_3(r_A,r_B,r_C) = V(r_A,r_B,r_C) - \sum_{i<j}^{3} V_2(r_i,r_j) \quad (3)$$

$$= V(r_A,r_B,r_C) - V_2(r_A,r_B) - V_2(r_A,r_C) - V_2(r_B,r_C)$$

where $V(r_A,r_B,r_C)$ is the binding energy of the trimer, evaluated using eq 2. In the same way the four-body term is defined as

$$V_4(r_A,r_B,r_C,r_D) =$$
$$V(r_A,r_B,r_C,r_D) - \sum_{i<j<k}^{4} V_3(r_i,r_j,r_k) - \sum_{i<j}^{4} V_2(r_i,r_j) \quad (4)$$

where each V_3 is evaluated using eq 3. The five-body terms are obtained in a similar manner.

Letters

J. Phys. Chem. B, Vol. 110, No. 22, 2006 **10597**

3. Computational Methods

We classify the functionals as strictly local, semilocal, and nonlocal. Strictly local and semilocal functionals are both local. Strictly local functionals depend only on the local spin density; such functionals are almost always parameterized to the homogeneous electron gas, and as such they are not usually accurate enough for quantitative applications on light elements (like oxygen) and are not considered here. Semilocal functionals depend in addition on the gradient and/or Laplacian of the density and/or on the kinetic energy density. The simplest semilocal functionals are generalized gradient approximations (GGAs), which depend only on the density and its gradient. Functionals that include Hartree−Fock exchange, which are usually called hybrid functionals, are nonlocal and are much less computationally convenient for bulk simulations. We have chosen four density functionals to examine in this study: PBE,[26] BLYP,[27,28] PBE1W,[17] and B3LYP.[27−29] PBE and BLYP were chosen because they are the two most commonly used GGA functionals in water and ice simulations.[30−36] The present study also includes the semilocal PBE1W functional, which is also a GGA, because it is a variation of PBE that has been optimized specifically for water and ice. Finally, we include the hybrid B3LYP functional both because of its widespread use in chemistry and also because of its recent use in the simulation of liquid water.[64] All density functional calculations were carried out using the Gaussian03[37] electronic structure package.

In previous work on water clusters, we have utilized the MG3S basis set for the assessment of all density functional methods. However, recent work by Csonka et al.[38] has shown that for the calculation of the binding energies of small water clusters, the results are highly basis-set dependent, and the empirically optimal basis set is specific to each functional. For this reason we give the results for the MG3S basis set and also, for each functional, for the optimal basis set as determined using the method in ref 38. The density functionals with their optimal basis sets are: PBE1W/6-311+G(2d,2p), PBE/aug-cc-pVTZ,[39] BLYP/6-31+G(d,p),[40] and B3LYP/6-31+G(d,2p).[40] The combination of a density functional and a basis set forms the basis for a theoretical model chemistry.[41] Although there is fundamental interest in examining the performance of functionals in the limit of a complete one-electron basis, selecting and validating well-balanced basis sets appropriate to each method and application is also an important objective of modern theoretical chemistry since practical predictions with well defined theoretical models "is of most interest to the larger chemical community."[41] Our original goal in developing the PBE1W functional was to optimize a semilocal functional for water because of the computational efficiencies achievable with local functionals. A special focus of the present study is therefore to analyze whether theoretical models built on semilocal functionals can achieve adequate performance for larger clusters and whether they correctly predict the separate pairwise and many-body components of the total interaction energies in these clusters.

Another issue worth addressing is that of basis set superposition error and whether counterpoise corrections[42] should be applied to the results in this paper. Xantheas has previously applied MP2 theory with a counterpoise correction scheme to evaluate many-body terms for small water clusters ($n = 2−6$).[43] For the current work, however, the use of W1 theory, which involves an extrapolation to the complete basis set limit, should be sufficiently accurate that the inclusion of counterpoise corrections, despite their enormous cost, would give only a very small change in the calculated energies. For the DFT calcula-

TABLE 1: Comparison of Accurate Many-Body Terms (kcal/mol) for the Global Minima Structures of the Water Trimer and Tetramer as Compared to Previous Work[25]

	$E_{bind}{}^c$	PA	V_3	V_4
C1GM[a]	16.39	13.92 (85)[d]	2.46 (15)	
MP2/aug-cc-pVDZ[b]	16.94	14.68 (87)	2.27 (13)	
MP4/aug-cc-pVDZ	16.92	14.71 (87)	2.20 (13)	
MP2/aug-cc-pVTZ	17.99	15.65 (87)	2.34 (13)	
MP2/aug-cc-pVDZ cp[e]	14.25	11.80 (83)	2.45 (17)	
MP4/aug-cc-pVDZ cp	13.97	11.57 (83)	2.40 (17)	
MP2/aug-cc-pVTZ cp	15.59	13.14 (84)	2.45 (16)	
CT2[a]	28.62	22.23 (78)	5.86 (20)	0.53 (2)
MP2/aug-cc-pVDZ	30.12	23.60 (78)	5.88 (20)	0.64 (2)
MP4/aug-cc-pVDZ	24.37	23.48 (79)	5.76 (19)	0.67 (2)
MP2/aug-cc-pVDZ cp	25.33	18.55 (73)	6.24 (25)	0.54 (2)
MP4/aug-cc-pVDZ cp	24.56	17.97 (73)	6.16 (25)	0.56 (2)

[a] Structures taken from current study. [b] All MP2 energies taken from ref 25. [c] The MP2 binding energies are given by $Ebind = BEn − ER$ where BEn and ER are the binding and relaxation energies, respectively, given in ref 25 for the trimer ($n = 3$) and tetramer ($n = 4$). [d] The percentage of the total energy is given in parentheses. [e] cp denotes counterpoise corrected values.

tions, we have chosen to omit the counterpoise correction for several reasons: (i) one often finds that for small basis sets the counterpoise corrections are significant, and in the right direction, whereas for larger basis sets the corrections are often small and may not improve the accuracy of the calculation,[44−47] (ii) for clusters larger than the dimer, the definition of the counterpoise correction becomes ambiguous,[48] and (iii) the magnitude of the counterpoise corrections is often smaller in DFT than it is in wave function theory.[49]

4. Results and Discussion

4.1. Comparison of Accurate Many-Body Terms to Previous Work. Previous work by Xantheas has examined the nature of many-body effects in cyclic water clusters from the trimer through the hexamer.[43] As the global minima for the water trimer and tetramer are examined both in Xantheas' paper and the current study (C1GM and CT2 respectively), it may be worthwhile to compare the accurate many-body terms between the two works. Before such a comparison can be made, however, a few differences in procedure must be noted. In Xantheas' work, both the trimer and the tetramer were optimized at the MP2/aug-cc-pVDZ level of theory, and the many-body terms were calculated at the MP2 and MP4 levels with the same basis set. For the trimer, the many-body terms were also calculated at the MP2 level of theory using the aug-cc-pVTZ basis set. In the current work the trimer and tetramer were optimized at the MP2/aug-cc-pVQZ and MP2/MG3S level of theories respectively (the trimer was taken from a previous study of the trimer potential energy surface[25]), and the many-body effects were calculated using the Weizmann-1 level of theory. Thus, the present article employs higher levels of theory for both geometries and energies than the levels employed by Xantheas.

Table 1 shows the comparison of the binding energies and many-body effects between the two papers. Looking at the magnitudes of the binding energies and many-body terms shows that differences may be as large as a couple of kilocalories per mole for the binding and pairwise energies, with smaller differences seen for the three- and four-body terms. Owing to the differences in geometries and the levels of theory used in the two works, a more valuable comparison may be to look at the percentage of the total energy accounted for by each term. A comparison of these values (given in parentheses in Table 1) shows that for the trimer structure all of the results of Xantheas

10598 *J. Phys. Chem. B, Vol. 110, No. 22, 2006*

Letters

TABLE 2: Comparison of DFT Binding Energies (kcal/mol) to Weizmann-1 (W1) Binding Energies and W1 Pairwise Additive (W1 PA) Energies

| structure | W1 PA | W1 | MG3S | | | | 6-311+G(2d,2p) | aug-cc-pVTZ | DIDZ | 6-31+G(d,2p) |
			PBE1W	PBE	BLYP	B3LYP	PBE1W	PBE	BLYP	B3LYP
trimers[a]										
BTS	12.45	14.15	13.56	14.51	11.67	13.17	13.63	13.39	14.05	13.96
C1C1TS	13.81	16.13	15.38	16.45	13.55	15.11	15.41	15.44	16.00	16.02
C1C3TS	13.29	15.56	14.80	15.84	12.98	14.51	14.81	14.86	15.29	15.37
C1GM	13.92	16.39	15.89	17.02	14.00	15.49	15.97	15.97	16.37	16.35
C3LM	13.25	15.59	14.95	16.01	11.90	14.59	14.98	15.01	15.34	15.41
C3h3rdOSP	13.03	15.00	13.57	14.47	13.10	13.61	13.46	13.62	14.44	14.65
G323	7.50	7.04	7.02	7.64	5.54	6.51	7.05	6.89	6.88	6.97
NVT	7.15	7.05	6.81	7.46	5.12	6.08	6.84	6.79	6.18	6.43
tetramers[b] and pentamer										
BT1	10.17	10.72	10.67	11.11	9.04	10.13	10.65	10.15	11.12	10.93
CT1	11.88	11.55	11.62	12.49	9.18	10.51	11.71	11.45	10.73	10.98
LT1	11.87	13.02	12.98	13.62	11.16	12.44	12.94	12.65	13.39	13.31
BT2	20.69	23.57	22.44	24.13	19.30	21.67	22.59	22.39	22.42	22.63
CT2	22.23	28.62	28.51	30.28	25.55	27.86	28.65	28.75	29.84	29.58
LT2	14.87	18.31	18.58	20.21	15.87	17.84	18.73	19.00	19.01	19.18
pentamer	15.81	17.84	17.16	19.86	12.55	15.62	17.42	18.52	15.48	16.87

[a] Names of structures correspond to those used to label geometries in the Supporting Information of ref 17. [b] Abbreviations stand for branched (BT), cyclic (CT) and linear (LT) tetramer. Abbreviations with a 1 indicate that the structure was taken from a NPT Monte Carlo simulation; those with a 2 were obtained as described in the text.

agree quite well with our results, with his counterpoise-corrected and noncorrected results straddling our results, with particularly good agreement for his counterpoise-corrected MP2/aug-cc-pVTZ calculation. For the tetramer we see slightly worse agreement than for the trimer, although the percentages of the total energy that each term comprises are quite similar, and we see the same trends in the percentages as one goes from the pairwise to the three-body to the four-body terms.

4.2. Comparison of Density Functionals and the Accurate Pairwise Additive Approximation. Table 2 compares the DFT binding energies both to the W1 binding energy and to the accurate pairwise value. The reader should keep in mind that a higher binding energy corresponds to a lower cluster energy or a higher monomer energy, or both. A comparison of the W1 binding energies to the W1 pairwise additive ones shows that the pairwise approximation underestimates the binding energy of the cluster for all but three of the structures: the two trimers and cyclic tetramer taken from Monte Carlo simulations. The structures in Figure 1 show that all of the trimers taken from the literature (see Figure 1a–1f) are arranged in space so that the formation of any of the hydrogen bonds in the cluster polarizes the water molecules in a way that makes the formation of the remaining hydrogen bonds more favorable. In the case of the two trimers taken from the simulation we see more complicated hydrogen bonding patterns. In the G323 structure (see Figure 1g) the central molecule is acting as a double hydrogen bond acceptor, leading to a destabilization of the trimer with respect to the constituent dimers. For the NVT structure (Figure 1h) the arrangement of the hydrogen bonds should be favorable; however, comparing this structure to the literature structures, which are all well defined minima on the trimer potential energy surface, shows that the angle between the three oxygen atoms is only 51.1°, which is ~9 degrees smaller than the average value (60.0°) from the literature structures. In addition, the O–H–O bond angle of the hydrogen bond in which the central water molecule is acting as a hydrogen bond acceptor is substantially closer to 180° (175°) than the O–H–O bond angles in the optimized structures (average value 149°). These two factors together may account for the overestimation of the binding energy for this structure. Figure 2c shows the structure of the cyclic tetramer taken from simulation. This

cluster has both a double hydrogen bond acceptor and a double hydrogen bond donor, leading to a similar destabilization as in the G323 dimer.

One measure of how the density functional methods perform is whether they predict binding energies that are higher than those underestimated by the pairwise approximation and lower that those that are overestimated. Looking at the results in Table 2 shows that when the optimal basis set is used for each functional, both PBE1W and B3LYP always meet this criterion. PBE has one exception, namely structure BT1; however, the deviation from the pairwise value is only 0.02 kcal/mol. BLYP also has only one exception, the pentamer structure, for which it underestimates the binding energy by 0.33 kcal/mol. When the MG3S basis set is considered, however, we see that PBE predicts higher binding energies than the pairwise approximation for all structures. On the other hand, BLYP predicts binding energies that are almost always lower than the accurate pairwise ones. B3LYP performs reasonably well with the MG3S basis set, as it only gives binding energies lower than the pairwise for two structures, BT1 and the pentamer, and both of these underestimates are small (0.04 and 0.19 kcal/mol respectively).

Table 3 summarizes the mean deviations of both the converged pairwise binding energies and the DFT binding energies from the W1 binding energies. We see that on average, if an optimal basis set is used for each functional, all of the density functional methods do perform better than the accurate pairwise potential with respect to both the mean unsigned and root-mean-squared errors. However, if a nonoptimal basis set is used, the density functional results can be considerably worse than the accurate pairwise potential, as in the case of BLYP/MG3S.

4.3. Binding Energies. Of special interest in Table 3 are the mean errors in the tetramer and pentamer binding energies. In past work[17] we have assessed the accuracy of a variety of density functional methods for predicting the binding energy of water dimers and trimers, and we used these data to develop the PBE1W functional. It has recently been proposed in the literature[38] that the optimization procedure used in the development of PBE1W led to a functional that was optimized specifically for the trimers present in the test set, and it has been implied that its performance will suffer for larger clusters

TABLE 3: Mean Errors[a] (kcal/mol) in DFT Binding and W1 Pairwise Additive Energies Compared to W1 Binding Energies

	trimers			tetramers and pentamer			all		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
W1 PA	−1.56	1.70	1.91	−2.30	2.40	3.09	−1.91	2.03	2.53
PBE1W/MG3S	−0.62	0.62	0.73	−0.24	0.33	0.51	−0.44	0.48	0.64
PBE1W/6−311+G(2d,2p)	−0.60	0.60	0.73	−0.14	0.27	0.41	−0.38	0.46	0.61
PBE/MG3S	0.31	0.44	0.46	1.15	1.15	1.32	0.70	0.77	0.96
PBE/aug−cc−pVTZ	−0.62	0.62	0.71	−0.11	0.53	0.63	−0.38	0.58	0.68
BLYP/MG3S	−2.38	2.38	2.46	−3.00	3.00	3.24	−2.67	2.67	2.85
BLYP/6−31+G(d,p)	−0.29	0.29	0.40	−0.23	1.00	1.19	−0.27	0.63	0.86
B3LYP/MG3S	−0.98	0.98	1.00	−1.08	1.08	1.26	−1.03	1.03	1.13
B3LYP/6−31+G(d,2p)	−0.22	0.22	0.28	−0.02	0.69	0.75	−0.13	0.44	0.55

[a] MSE, MUE, and RMSE denote mean signed error, mean unsigned error, and root-mean-square error, respectively.

TABLE 4: Mean Errors (kcal/mol) in DFT Binding Energies for Large Data Set,[a] Compared to W1 Binding Energies

	dimers			trimers			tetramers and pentamer			all		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
PBE1W/MG3S	−0.01	0.10	0.15	−0.13	0.28	0.40	−0.21	0.31	0.45	−0.07	0.19	0.30
PBE1W/6-311+G(2d,2p)	0.00	0.11	0.16	−0.09	0.29	0.41	−0.09	0.28	0.39	−0.04	0.19	0.30
PBE/MG3S	0.21	0.22	0.31	0.55	0.58	0.73	1.27	1.27	1.44	0.44	0.46	0.68
PBE/aug-cc-pVTZ	−0.05	0.15	0.22	−0.11	0.39	0.51	0.17	0.57	0.69	−0.05	0.28	0.41
BLYP/MG3S	−0.56	0.56	0.66	−1.59	1.59	1.76	−3.06	3.06	3.26	−1.18	1.18	1.57
BLYP/6-31+G(d,p)	−0.06	0.17	0.23	−0.23	0.42	0.53	−0.71	1.16	1.30	−0.19	0.36	0.56
B3LYP/MG3S	−0.21	0.21	0.25	−0.61	0.61	0.69	−1.17	1.17	1.28	−0.45	0.45	0.61
B3LYP/6-31+G(d,2p)	−0.03	0.12	0.17	−0.10	0.30	0.36	−0.24	0.63	0.71	−0.08	0.24	0.34

[a] The large data set consists of 61 dimers, 39 trimers, 11 tetramers, and one pentamer for a total of 112 structures.

that are outside of the database used for the parameterization. Table 3 provides a test of that speculation.

Comparing the PBE1W errors for the trimers to those for the tetramers and pentamers, for both the MG3S and 6-311+G-(2d,2p) basis sets, shows that both methods have mean unsigned errors that are approximately a factor of 2 smaller for the larger clusters than the smaller ones, and have root-mean-square errors that also decrease with increasing cluster size. Thus the speculations were not correct. The only other density functional that improves its mean errors upon moving to larger clusters is PBE/aug-cc-pVTZ, for which the mean unsigned error decreases by approximately 15%. A comparison of the PBE1W results, with either basis set, to the PBE/aug-cc-pVTZ results shows that they give nearly identical results for the trimer data, with PBE1W giving better results for the larger clusters. The remaining density functionals show decreased performance with increasing cluster size.

To get better statistics for the accuracy of the density functionals, we have calculated the mean errors in the binding energies for not only the 15 structures listed in Table 2, but also over all the constituent dimers, trimers, and tetramers used to determine the pairwise and many-body terms. After taking into account structures that were equivalent by symmetry, this gave a total of 112 pieces of data, hereafter referred to as the large data set. Full W1 calculations were carried out for all 112 structures to perform this test. The results for the large data set are given in Table 4. For the large data set the mean errors increase with increasing cluster size for all four density functionals tested. PBE1W/MG3S and PBE1W/6-311+G(2d,-2p) give the smallest mean unsigned error (0.19 kcal/mol) for all cluster sizes. Additionally, we see that for both of these levels of theory their performance with respect to the other functionals get better with increasing cluster size. The error seen in PBE1W for the tetramers and pentamer is almost a factor of 2 smaller than the error of the next best method, PBE/aug-cc-pVTZ.

4.4. Many-Body Terms. The investigation of many-body effects in small water clusters is not a new area of

TABLE 5: Weizmann-1 Many-Body Terms (kcal/mol)

	V_3	V_4	V_5
BTS	1.70		
C1C1TS	2.32		
C1C3TS	2.28		
C1GM	2.46		
C3LM	2.34		
C3h3rdOSP	1.97		
G323	−0.46		
NVT	−0.10		
BT1	0.55	0.00	
CT1	−0.44	0.05	
LT1	1.11	0.04	
BT2	2.98	−0.09	
CT2	5.86	0.53	
LT2	3.36	0.04	
P	2.26	−0.24	0.01

research,[16,43,50−60] however, much of the work in the literature has only looked at the many-effects in structures obtained from ab initio geometry optimizations of small water clusters. In contrast, the data set used in the present work includes not only well-defined minima but also structures that have come directly out of simulations of bulk water and ice. Due to environmental effects present in the condensed-phase, one might expect the many-body effects in these structures to be different than those in gas-phase clusters. If one finds that the many-body effects are quite different for these two different types of clusters, and if one is interested in assessing the applicability of different functionals to the simulation of condensed-phase water, then it may of be of use to see how the functionals do at predicting the many-body effects of clusters taken out of condensed-phase simulations.

Another concern is whether the assessment of the results for these methods on small water clusters will be able to help predict their performance in bulk simulations. As mentioned in the Introduction, the evaluation of the ability of analytic potentials to accurately reproduce many-body effects in small clusters has been used in the past as a means to validate their use in

TABLE 6: Mean Errors (kcal/mol) for DFT Pairwise Additive and Many Body Terms Compared to W1 Results

	PA			V_3			V_4 and V_5			all		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
PBE1W/MG3S	−0.11	0.35	0.47	−0.37	0.37	0.54	0.08	0.08	0.12	−0.17	0.30	0.45
PBE1W/6-311+G(2d,2p)	−0.06	0.38	0.52	−0.36	0.36	0.53	0.08	0.08	0.12	−0.15	0.31	0.47
PBE/MG3S	0.99	1.04	1.30	−0.33	0.33	0.49	0.08	0.08	0.12	0.28	0.56	0.87
PBE/aug-cc-pVTZ	−0.29	0.60	0.73	−0.12	0.17	0.30	0.05	0.05	0.09	−0.15	0.32	0.50
BLYP/MG3S	−2.75	2.75	2.91	0.06	0.13	0.18	0.02	0.05	0.07	−1.05	1.14	1.83
BLYP/6-31G+(d,p)	−0.17	0.57	0.79	−0.10	0.11	0.16	0.02	0.04	0.06	−0.11	0.28	0.51
B3LYP/MG3S	−0.99	0.99	1.07	−0.05	0.07	0.11	0.03	0.03	0.05	−0.41	0.43	0.68
B3LYP/6-31+G(d,2p)	−0.06	0.37	0.47	−0.08	0.10	0.13	0.02	0.03	0.04	−0.05	0.19	0.31

condensed phase simulation.[16] Additionally, work by Goldman et. al has shown that the VRT(AP−W)III potential,[61] which is based on ab initio and spectroscopic data available for the water dimer, is able to accurately reproduce the vibrational ground-state structures of small water clusters up to the hexamer,[62,63] as well as reproduce many of the condensed-phase properties of liquid water.[63] In light of this kind of success, it is our hope that the study of the performance of density functionals for small water clusters will be able to lend some insight to their performance in condensed-phase simulations.

Table 5 shows the Weizmann-1 many-body terms for the fifteen structures considered in Tables 1 and 2. Looking at their magnitude shows that the structures taken from simulation tend to have smaller many-body effects than the optimized clusters. We also see that with the exception of the four-body term of the BT2 tetramer, all of the negative many-body terms correspond to structures that are taken from simulations. Additionally, Table 4 shows the rapid convergence of the series in eq 1. One can see that the average signed value of the interaction terms decreases from 13.46 to 1.89 to 0.04 to 0.01 kcal/mol as one goes from V_2 to V_3 to V_4 to V_5, respectively. (The average absolute magnitudes are 13.46, 2.00, 0.15, and 0.01 kcal/mol, respectively.) Of the seven structures larger than trimer, five of them (the exceptions being the CT2 tetramer and the pentamer structure) have converged within 0.09 kcal/mol after inclusion of the three-body term. In the case of the CT2 tetramer, the high symmetry of the cluster and the cooperative nature of the hydrogen bonding around the ring leads to trimers that are highly cooperative in nature as well. For the pentamer structure, the nature of the three- and four-body effects has been investigated by Hermansson[54] and has been explained by the electron density distribution of the constituent water molecules.

Table 6 shows the mean errors in the pairwise-additive and many-body terms as predicted by the density functionals, and illustrates how a theoretical model may be analyzed for predicting the constituents of an interaction energy as well as its total magnitude. For three of the four functionals tested, we see that the largest error is in the pairwise term with substantial improvement upon moving to the three-body term, which is consistent with its smaller magnitude. The two exceptions, PBE1W/MG3S and PBE1W/6-311+G(2d,2p), have similar errors for both the pairwise and three-body terms, with a difference of only 0.02 kcal/mol in the mean unsigned errors. While at first glance this may be somewhat alarming, the good performance of these methods for the pairwise term may make this result appear worse than it really is, as the mean unsigned errors for the pairwise term are substantially lower (39% and 33% respectively) than the next best result for a local functional (BLYP/6-31+G(d,p)). Averaging over all the many-body terms shows that the nonlocal B3LYP/6-31+G(d,2p) functional has the smallest error of the eight model chemistries studied for the many-body terms. Table 6 also shows that BLYP/6-31+G-(d,p) is the best of the three semilocal methods for predicting

many-body effects; however, it should be noted that if an optimal basis set is used the three semilocal functionals differ from each other in the mean unsigned error by at most 0.04 kcal/mol. Comparing the results of the functionals with and without use of the optimal basis set again illustrates that the use of the optimal basis set is crucial to maximizing the performance of the functionals, as the mean unsigned error can vary by as much as a factor of 5. One expects that there could be similar sensitivities to plane wave cutoff values and effective core potentials for basis sets consisting of plane waves.

Recent work by Todorova et al.[64] has involved a series of water simulations with the following functionals: LDA, BLYP, XLYP, PBE, rPBE, TPSS, B3LYP, X3LYP, PBE1PBE, and Hartree−Fock theory. In general they find that the hybrid functionals (B3LYP, X3LYP, and PBE1PBE) outperform the semilocal GGA and meta-GGA functionals (BLYP, XLYP, PBE, rPBE, TPSS), which, in turn, outperform LDA. These results are consistent with our previous study[17] in which we evaluated the accuracy of 25 density functionals to predict the binding energies for a collection of water dimers and trimers and found, in general, that hybrid functionals outperform nonhybrid functionals. We would like to note that there was a wide range of performance between the different hybrid methods (as with all classes of density functionals tested), and that B3LYP was not the best hybrid functional tested in ref 17. The better performance of B3LYP as compared to the nonhybrid methods is also consistent with the results of this study, in which we see that B3LYP/6-31+G(d,2p) gives the lowest errors in predicting the many-body terms for our set of clusters; however, it is likely that other hybrid methods would have given results as good, if not better than B3LYP. Additionally, B3LYP/6-31+G(d,2p) has the second lowest error in predicting the binding energies for our large data set, having a mean unsigned error only 0.05 kcal/mol higher than either PBE1W result. Since PBE1W does better for predicting binding energies than for predicting many-body terms, and B3LYP does the opposite, using PBE1W in the simulation of liquid water may be useful for determining which of these two criteria is more important for correctly predicting the properties of bulk water.

5. Concluding Remarks

In this study we have evaluated the accuracy of a set of density functional methods for predicting the binding energies and many-body terms in a set of water clusters. We have compared the binding energies obtained from DFT to accurate pairwise additive energies obtained using the very accurate-Weizmann-1 level of theory and have found that while all density functionals tested are capable of outperforming the accurate pairwise data, the correct choice of basis set used is crucial to ensure good results. This illustrates the importance of the choice of basis set used in simulation of bulk water and ice.

Letters

J. Phys. Chem. B, Vol. 110, No. 22, 2006 **10601**

We have also shown that of the four density functionals tested, PBE1W/MG3S and PBE1W/6-311+G(2d,2p) provide the most accurate binding energies for a large database containing 112 water clusters ranging in size from dimer to pentamer. The success of PBE1W in predicting the binding energy for a wide range of trimers and tetramers that were not included in its parameterization set is encouraging for its performance when used in simulation of bulk systems or for larger water clusters.

Finally, we have shown that the ability of DFT to correctly predict many-body terms for small water clusters is also dependent both on the density functional used and on the basis set employed. We have found that all semilocal functionals have significant errors in predicting the many-body terms, but that this does not necessarily mean that they are unable to accurately predict useful binding energies. Which of these two factors will prove to be more crucial for good performance in simulations of bulk water remains to be seen.

Supporting Information Available: Tables giving the accurate binding energies for all structures in the large data set, as well as the Cartesian coordinates for all structures listed in Table 2. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975.
(2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(3) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *97*, 6269.
(4) Rick, S. W.; Stewart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.
(5) Rick, S. W. *J. Chem. Phys.* **2001**, *114*, 2276.
(6) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237.
(7) Yang, Z.-Z.; Wu, Y.; Zhao, D.-X. *J. Chem. Phys.* **2004**, *120*, 2541.
(8) Stillinger, F. H.; David, C. W. *J. Chem. Phys.* **1978**, *69*, 1473.
(9) Ahlström, P.; Wallquist, A.; Engström, S.; Jönsson, B. *Mol. Phys.* **1989**, *68*, 563.
(10) Dykstra, C. E. *J. Chem. Phys.* **1989**, *91*, 6472.
(11) Svishchev, I. M.; Kusalik, P. G.; Wang, J.; Boyd, R. J. *J. Chem. Phys.* **1996**, *108*, 4742.
(12) Chen, B.; Xing, H.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391.
(13) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 1479.
(14) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080.
(15) McGrath, M. J.; Siepmann, I. J.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *ChemPhysChem* **2005**, *6*, 1894.
(16) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163.
(17) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 15677.
(18) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(19) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
(20) Krishnan, R.; Binkly, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
(21) Umemoto, K.; Wentzcovich, R., unpublished.
(22) Oliveira, G. d.; Martin, J. M. L. *J. Chem. Phys.* **1999**, *111*, 1843.
(23) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.
(24) Werner, H.-L.; Knowles, P. J.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schÿtz, M.; Schumann, U.; Stoll, H.;

Stone, A. J.; Tarroni, R.; Thorsteinsson, T. MOLPRO, 2002.6 ed.; University of Birmingham: Birmingham, 2002.
(25) Anderson, J. A.; Crager, J.; Federoff, L.; Tschumper, G. S. *J. Chem. Phys.* **2004**, *121*, 11023.
(26) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(27) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(28) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(29) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
(30) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572.
(31) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2002**, *116*, 10372.
(32) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *120*, 300.
(33) Schwegler, E.; Grossman, J. C.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *121*, 5400.
(34) Kuo, I.-F. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 12990.
(35) Umemoto, K.; Wentzcovich, R.; Baroni, S.; Gironcoli, S. d. *Phys. Rev. Lett.* **2004**, *92*, 105502-1.
(36) Umemoto, K.; Wentcovich, R. *Phys. Rev. B* **2004**, *69*, 180103.
(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian 03, Revision C.01 ed.; Gaussian, Inc.: Wallingford CT, 2004.
(38) Csonka, G. I.; Ruzsinszky, A.; Perdew, J. P. *J. Phys. Chem. B* **2005**, *109*, 21471.
(39) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(40) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(41) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
(42) Boys, S. F.; Bernardi, D. *Mol. Phys.* **1970**, *19*, 553.
(43) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.
(44) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418. Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1987**, *86*, 3760(E).
(45) Collins, J. R.; Gallup, G. A. *Chem. Phys. Lett.* **1986**, *123*, 56.
(46) Masamura, M. *Theor. Chem. Acc.* **2001**, *100*, 301.
(47) Weck, G.; Millet, A.; Kochanski, E. *J. Phys. Chem. A* **2002**, *106*, 12084.
(48) Mierzwicki, K.; Latajka, Z. *Chem. Phys. Lett.* **2003**, *380*, 654.
(49) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
(50) Hankins, D.; Miskowitz, J. W.; Stillinger, F. H. *J. Chem. Phys.* **1970**, *53*, 4544.
(51) Clementi, E.; Kolos, W.; Lie, G. C.; Ranghino, G. *Int. J. Quantum Chem.* **1980**, *17*, 377.
(52) Yoon, B. J.; Morokuma, K.; Davidson, E. R. *J. Chem. Phys.* **1985**, *83*, 1223.
(53) Kim, K. S.; Dupuis, M.; Lie, G. C.; Clementi, E. *Chem. Phys. Lett.* **1986**, *131*, 451.
(54) Hermansson, K. *J. Chem. Phys.* **1988**, *89*, 2149.
(55) Chen, W.; Gordon, M. S. *J. Phys. Chem.* **1996**, *100*, 14316.
(56) Gregory, J. K.; Clary, D. C. *J. Phys. Chem.* **1996**, *100*, 18014.
(57) Masella, M.; Flament, J. P. *J. Chem. Phys.* **1997**, *107*, 9105.
(58) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. *J. Phys. Chem. A* **1999**, *103*, 6811.
(59) Ojamäe, L.; Hermansson, K. *J. Phys. Chem.* **1994**, *98*, 4271.
(60) Pedulla, J. M.; Kim, K.; Jordan, K. D. *Chem. Phys. Lett.* **1998**, *291*, 78.
(61) Goldman, N.; Fellers, R. S.; Brown, M. G.; Braly, L. B.; Keoshian, C. J.; Leforestier, C.; Saykally, R. J. *J. Chem. Phys.* **2002**, *116*, 10148.
(62) Goldman, N.; Leforestier, C.; Saykally, R. J. *Philos. Trans. R. Soc. A* **2005**, *363*, 493.
(63) Goldman, N.; Saykally, R. J. *J. Chem. Phys.* **2004**, *120*, 4777.
(64) Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. *J. Phys. Chem. B* **2006**, *110*, 3685.