

Evaluation of the Electrostatically Embedded Many-Body Expansion and the Electrostatically Embedded Many-Body Expansion of the Correlation Energy by Application to Low-Lying Water Hexamers

Erin E. Dahlke, Hannah R. Leverentz, and Donald G. Truhlar*

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

Received July 24, 2007

Abstract: We have applied a many-body (MB) expansion, the electrostatically embedded many-body (EE-MB) approximation, and the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE), each at the two-body (MB = PA, where PA denotes pairwise additive) and three-body (MB = 3B) levels, to calculate total energies for a series of low-lying water hexamers using eight correlated levels of theory including second-order and fourth-order Møller–Plesset perturbation theory (MP2 and MP4) and coupled cluster theory with single, double, and quasiperturbative triple excitations (CCSD(T)). Comparison of the expansion methods to energies obtained from full (i.e., unexpanded) calculations shows that the EE-3B-CE method is able to reproduce the full cluster energies to within 0.03 kcal/mol, on average. We have also found that the deviations of the results predicted by the expansion methods from those obtained with full calculations are nearly independent of the correlated level of theory used; this observation will allow validation of the many-body methods on large clusters at less expensive levels of theory (such as MP2) to be extrapolated to the CCSD(T) level of theory. Furthermore, we have been able to rationalize the accuracies of the MB, EE-MB, and EE-MB-CE methods for the six hexamers in terms of the specific many-body effects present in each cluster.

1. Introduction

The ability to calculate accurate energies for large systems has long been a goal of the quantum chemical community. Hartree–Fock theory, which neglects electron correlation, is inadequate but is used to generate orbitals for methods such as second-order Møller–Plesset perturbation theory, MP2,¹ coupled cluster theory with single and double excitations, CCSD,^{2,3} or CCSD with quasiperturbative connected triple excitations, CCSD(T),⁴ which do include electron correlation and are able to accurately predict energies, geometries, and frequencies for small to moderately sized chemical systems. However, these post-Hartree–Fock methods have thus far proven to be too expensive, in their original implementations, to be used for systems containing tens to hundreds of atoms. As a result, there has been considerable

research aimed at trying to develop highly efficient algorithms,^{5–9} including parallelization schemes,^{10–18} to make large systems tractable. However, because of the steep scaling of computational effort with respect to system size (CCSD(T), CCSD, and MP2 scale as N^7 , N^6 , and N^5 , respectively, where N is the number of atoms¹⁹), it is impractical to utilize even these more efficient implementations for systems containing hundreds to thousands of atoms.

A promising area of research has focused on developing variants of these methods that use localized molecular orbitals^{15,20–23} or fragmentation.^{13,24–40} One can also consider including a subset of interactions, for example Coulomb interactions, to high order or in full, with other interactions, e.g., those due to electron correlation energy considered only to a lower order, e.g., only pairwise.^{41–43} In past work^{44,45} we have developed the electrostatically embedded many-body method (EE-MB) and the electrostatically embedded

* Corresponding author e-mail: truhlar@umn.edu.

many-body expansion of the correlation energy (EE-MB-CE) methods and have applied both methods to the study of water clusters, ranging in size from 5 to 20 water molecules, at the MP2 level of theory. MP2 was chosen because it is the least expensive of the post-Hartree–Fock methods, allowing for direct comparison of the EE-MB and EE-MB-CE methods to the MP2 energies of the full cluster. In this way we were able to examine the performance of the expansion methods with respect to increasing system size, which would not have been possible with other more expensive post-Hartree–Fock methods like CCSD(T).

Despite the advantages of using MP2 for our initial studies, there is a concern that other levels of electronic structure theory could show different behavior. Because of the N^6 or N^7 scaling of the post-MP2 methods with respect to system size, if we wish to compare directly to full calculations (i.e., conventional calculations without a many-body expansion) at a post-MP2 level with a reasonably large basis set (polarized valence triple- ζ or higher) we are limited to relatively small systems (on the order of five heavy atoms). Recent work by Olson et al.¹⁸ has provided CCSD(T), CCSD, and MP2 energies for a series of five water hexamers using both the aug-cc-pVTZ^{46,47} and s-cc-pVTZ basis sets (s-cc-pVTZ denotes semidiffuse cc-pVTZ, and it uses the aug-cc-pVTZ basis set on oxygen and the cc-pVTZ⁴⁸ basis set on hydrogen). These results constitute a set of highly accurate energies against which to test our methods.

Water clusters in general exhibit large many-body effects,⁴⁹ and it is well-known that different structural motifs can lead to different many-body effects;^{50,51} therefore, these clusters are an excellent choice for examining the behavior of many-body methods. Along this line, Pedulla and Jordan⁵¹ have examined the many-body effects of three isomers of the water hexamer (cage, prism, and ring). Since there are a number of isomers of the water hexamer that all lie within a few kilocalories per mole of each other,^{18,52–58} this system serves as a good test of the predictive capabilities of electronic structure methods for water.

For any level of theory (e.g., MP2 or CCSD(T) with a given basis) we can either perform full (i.e., conventional) calculations of the potential energy, V , or many-body expansions, with the latter defined by a truncated version of

$$V = V_1 + V_2 + V_3 + \dots \quad (1)$$

where V_n is the n -body term. Truncating at V_2 is called the pairwise additive (PA) approximation, and truncating at V_3 is called the three-body (3B) approximation. For $(\text{H}_2\text{O})_N$, V_2 involves calculating $(N(N-1)/2)$ dimer calculations, and V_3 involves $(N(N-1)(N-2)/3!)$ trimer calculations. If the n -mer calculations are performed in vacuum, we have a conventional many-body method (PA or 3B), and if they are performed in a field of point charges at the nuclear positions of the $N-n$ missing monomers, we have the electrostatically embedded many-body method (EE-PA or EE-3B). If we perform a full (V_N) calculation at the Hartree–Fock (HF) level and expand the correlation energy $(V - V_{\text{HF}})$, we have the many-body expansion of the correlation energy method (PA-CE or 3B-CE without point charges and EE-PA-CE or EE-3B-CE with them). Further details of the

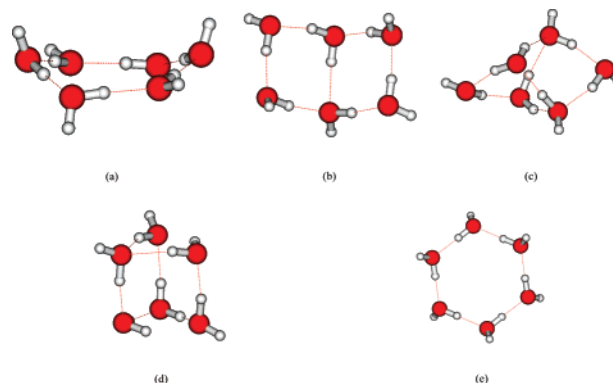


Figure 1. MP2/DH(d,p) optimized hexamers of Day et al. for the (a) boat, (b) book, (c) cage, (d) prism, and (e) ring structures.

many-body methods have been discussed in previous work^{44,45} and are not discussed here.

2. Computational Details

The hexamers used in this work are defined by the MP2/DH(d,p) geometries of Day et al.⁵⁵ for the boat, book, cage, ring (denoted as cyclic in that work), and prism isomers (see Figure 1). All single-point calculations in this work use the s-cc-pVTZ basis set (see section 1 for the definition of s-cc-pVTZ). For each hexamer a total of nine levels of electronic structure theory were considered: Hartree–Fock, MP2, MP3, MP4D, MP4DQ, MP4SDQ, MP4, CCSD, and CCSD(T), where MP3, MP4D, MP4DQ, MP4SDQ, and MP4 denote various high-order perturbation theory approximations.^{59–62} The CCSD(T), CCSD, and MP2 single-point energies for the hexamers were taken from the work of Olson et al.¹⁸ In addition, MP4 calculations were run with the *Gaussian 03*⁶³ software program to determine the MP3, MP4D, MP4DQ, MP4SDQ, and MP4 single-point energies for each hexamer. PA, 3B, EE-PA, EE-3B, PA-CE, 3B-CE, EE-PA-CE, and EE-3B-CE calculations were carried for each hexamer, at each of the nine levels of theory, using the *Gaussian 03* software package.

For the sake of clarity a combination of many-body method and electronic structure theory will be denoted by the name of the many-body method with the level of electronic structure level in parentheses. For example, EE-PA-CE(MP2) will denote an EE-PA-CE calculation carried out at the MP2 level of theory.

In the EE-MB and EE-MB-CE methods, charges of -0.778 and 0.389 were used for the oxygen and hydrogen atoms, respectively, as in refs 44 and 45.

3. Results and Discussion

3.1. Full Calculations. While the main goal of this paper is analysis of the ability of the many-body methods to accurately reproduce full energies obtained at the same level of electronic structure theory (e.g., comparing a EE-PA-CE(MP2) calculation to a full MP2 calculation on the same hexamer), it is useful to first examine the results of the full calculations. Table 1 shows the relative energy differences (relative to the prism structure) predicted by the full calculations for each level of theory. All eight correlated

Table 1. Relative^a Energies (kcal/mol) Predicted by Full Calculations Using the s-cc-pVTZ Basis Set^b

	boat	book	cage	prism	ring
HF	−1.79	−1.05	−0.27	0.00	−2.52
MP2	2.41	0.67	0.02	0.00	1.22
MP3	2.25	0.74	0.13	0.00	1.15
MP4D	2.33	0.76	0.14	0.00	1.21
MP4DQ	1.92	0.60	0.10	0.00	0.84
MP4SDQ	2.20	0.71	0.12	0.00	1.09
MP4	2.99	1.02	0.17	0.00	1.80
CCSD	2.23	0.74	0.14	0.00	1.12
CCSD(T)	2.99	1.06	0.20	0.00	1.81

^a All energies are relative to the prism isomer ^b The s-cc-pVTZ basis set denotes the aug-cc-pVTZ basis set on oxygen and the cc-pVTZ basis set on hydrogen.

levels of theory predict a relative energy ordering of prism < cage < book < ring < boat. The energy differences separating the isomers are all less than 3 kcal/mol, with the smallest energy gap (cage minus prism) in the range between 0.02 and 0.20 kcal/mol.

An interesting result in Table 1 is how the perturbation theory and CCSD results compare to the CCSD(T) results. As was noted by Olson et al.¹⁸ the MP2 results differ from the CCSD(T) results by 0.2–0.6 kcal/mol, with CCSD(T) predicting systematically larger energy gaps than MP2. Past work^{57,64,65} comparing MP2 and CCSD(T) for small water clusters showed that the differences obtained are small (on the order of 0.05 kcal/mol for the water dimer and 0.1 kcal/mol for the trimer), and, as a result, MP2 has become the method of choice for many workers when studying water clusters.^{51,56,65–68} The results in Table 1 indicate that the exceptionally good agreement of MP2 and CCSD(T) for small water clusters may begin to break down as larger clusters are considered, and thus caution should be used when applying MP2 to medium- to large-sized clusters. Another interesting result is how well MP4 is able to reproduce the CCSD(T) relative energies. Of the methods in Table 1, only MP4 and CCSD(T) include connected triple excitations. CCSD(T) includes not only the fourth-order connected triples of MP4 but also a fifth-order connected triple excitation operator involving singles amplitudes. The near equivalence of MP4SDQ and CCSD in Table 1 shows that disconnected triples (included in the latter but not the former) are unimportant for water clusters, and the near agreement of MP4 with CCSD(T) but not with MP4SDQ shows that for water clusters the connected triples are important, but it is adequate to include them at fourth order.

With five different structures there are a total of 10 energy differences that one can compute (for example, the energy difference between the boat and the book or between the cage and the ring). In order to better characterize how accurate the MP n ($n = 2–4$) and CCSD results are, as compared to CCSD(T), we have calculated the 10 energy differences at each of the 9 levels of theory and computed the mean unsigned and root mean squared errors relative to the CCSD(T) results. The results of this analysis are shown in Table 2. From this table it is clear that MP4 is very accurate for these systems, as compared to CCSD(T); it has a mean unsigned deviation of only 0.02 kcal/mol, and the

Table 2. Average Deviations between Various Full Calculated Energies and Full CCSD(T)^a

	MUE	RMSE
HF	2.68	3.08
MP2	0.31	0.36
MP3	0.41	0.47
MP4D	0.37	0.43
MP4DQ	0.60	0.69
MP4SDQ	0.44	0.51
MP4	0.02	0.03
CCSD	0.43	0.49

^a These results are averaged over the ten energy differences that can be obtained from the five structures in Table 1.

largest difference between the MP4 and CCSD(T) for any of the 10 possible energy differences is 0.04 kcal/mol. This conclusion is in good agreement with previous work by Xantheas et al.⁶⁹ on the water dimer. While MP2 has the second-lowest mean unsigned error (0.31 kcal/mol), its performance is considerably worse than that of MP4. The remaining correlated methods have mean unsigned errors ranging from 0.37 to 0.60 kcal/mol. Hartree–Fock does particularly poorly, as is expected from the results in Table 1. While both MP4 and CCSD(T) formally scale as N^7 , the use of CCSD(T) requires the completion of a CCSD calculation (which has an iterative N^6 step⁴) before the noniterative triples calculation. As a result, MP4 is less expensive than CCSD(T) which may allow MP4 to be used for benchmark calculations of water clusters that are too large for CCSD(T). It seems worthwhile to note that although CCSD(T) is well-known⁴ to be more accurate than MP4 in general, where singles amplitudes may be large, this need not be the case for particular interactions of noncovalent interactions of closed-shell species.⁷⁰ Even though the present comparison involves larger clusters than those for which MP n results ($n = 2, 3, 4$) have previously (prior to ref 18) been compared to CCSD(T) results, and hence involves comparisons that are assumed to be more relevant to the bulk water case, there is no guarantee that MP4 is better than MP2 in general (the series is often divergent, especially with basis sets containing diffuse functions); and the good agreement of MP4 with CCSD(T) for these small water clusters is not guaranteed to hold for larger clusters. Therefore caution must be exercised in choosing either MP2 or MP4 as an alternative to CCSD(T) for water clusters. Further, systematic validation studies would be useful.

3.2. Pairwise Additive Methods. In order to assess the accuracy of the many-body methods we begin by evaluating the average error in the electronic energy for each pairwise additive method, at each level of electronic structure theory, when compared to the full calculation (i.e., evaluating the deviation between the EE-PA(MP2) and full MP2 energy for each of the five hexamers). Table 3 shows the average deviations between the pairwise additive methods (PA, PA-CE, EE-PA, EE-PA-CE) and the full calculations at each of the nine levels of theory. The first interesting observation is that the errors for each pairwise additive method (PA or EE-PA) are similar for all the correlated levels of theory as indicated by the standard deviations being much smaller than the average MUE. A comparison of the average mean

Table 3. Average Deviations^a (kcal/mol) between Pairwise Additive Energies and Full Calculations at the Same Level of Theory

	PA			PA-CE			EE-PA			EE-PA-CE		
	MSD	MUD	RMSD	MSD	MUD	RMSD	MSD	MUD	RMSD	MSD	MUD	RMSD
HF	11.81	11.81	11.94	0.00	0.00	0.00	1.13	1.13	1.15	0.00	0.00	0.00
MP2	11.89	11.89	12.01	0.03	0.09	0.10	1.10	1.10	1.11	−0.04	0.04	0.04
MP3	11.75	11.75	11.89	0.01	0.11	0.13	0.96	0.96	1.01	−0.17	0.17	0.22
MP4D	11.75	11.75	11.89	0.01	0.10	0.13	0.96	0.96	1.00	−0.17	0.17	0.22
MP4DQ	11.70	11.70	11.83	−0.03	0.12	0.14	0.95	0.95	0.98	−0.19	0.19	0.21
MP4SDQ	11.72	11.72	11.86	−0.02	0.10	0.12	0.98	0.98	1.02	−0.15	0.15	0.18
MP4	11.80	11.80	11.93	0.03	0.08	0.09	1.02	1.02	1.06	−0.11	0.11	0.15
CCSD	11.74	11.74	11.88	−0.01	0.09	0.11	0.99	0.99	1.02	−0.14	0.14	0.18
CCSD(T)	11.82	11.82	11.96	0.05	0.09	0.10	1.03	1.03	1.07	−0.11	0.12	0.16
av MUE ^b		11.77			0.10			1.00			0.14	
SD ^c		0.06			0.01			0.05			0.05	

^a MSD, MUD, and RMSD denote mean signed, mean unsigned, and root-mean-squared deviations, respectively, in \AA as compared to the full calculations. Thus a positive MSE corresponds to underestimating the strength of binding, and a negative MSE corresponds to overestimating the strength of binding. ^b Average of the MUE for the correlated methods (the rows from MP2 to CCSD(T)). ^c Standard deviation of MUE for the correlated methods (the rows from MP2 to CCSD(T)).

Table 4. Average Deviations^a (kcal/mol) between Three-Body (3B) Methods and Full Calculations at the Same Level of Theory

	3B			3B-CE			EE-3B			EE-3B-CE		
	MSD	MUD	RMSD	MSD	MUD	RMSD	MSD	MUD	RMSD	MSD	MUD	RMSD
HF	1.08	1.08	1.22	0.00	0.00	0.00	0.08	0.09	0.11	0.00	0.00	0.00
MP2	1.25	1.25	1.42	0.17	0.17	0.20	0.09	0.12	0.15	0.01	0.03	0.04
MP3	1.21	1.21	1.37	0.12	0.12	0.15	0.08	0.11	0.14	0.00	0.03	0.03
MP4D	1.23	1.23	1.39	0.15	0.15	0.17	0.09	0.11	0.14	0.01	0.03	0.03
MP4DQ	1.21	1.21	1.37	0.13	0.13	0.15	0.08	0.11	0.13	0.00	0.03	0.03
MP4SDQ	1.23	1.23	1.39	0.15	0.15	0.17	0.08	0.11	0.14	0.00	0.03	0.04
MP4	1.28	1.28	1.45	0.20	0.20	0.23	0.09	0.13	0.16	0.01	0.04	0.05
CCSD	1.22	1.22	1.39	0.14	0.14	0.17	0.08	0.11	0.14	0.00	0.03	0.04
CCSD(T)	1.27	1.27	1.44	0.19	0.19	0.22	0.09	0.13	0.16	0.01	0.05	0.05
av MUE ^b		1.24			0.16			0.12			0.03	
SD ^c		0.03			0.03			0.01			0.01	

^a MSD, MUD, and RMSD denote mean signed, mean unsigned, and root-mean-squared deviations, respectively, as compared to full calculations. See footnote *a* of Table 3 for an explanation of the signs. ^b Average of the MUE for the correlated methods (the rows MP2 to CCSD(T)). ^c Standard deviation of the correlated methods (the rows MP2 to CCSD(T)).

unsigned errors shows that the PA method has a large error of 11.78 kcal/mol. The average binding energy of the six structures at the CCSD(T) level is 46.72 kcal/mol (taken from ref 18), so an average error of 11.78 kcal/mol corresponds to a percent error of approximately 25% (the next largest percent error is 2.2% for the EE-PA method). The EE-PA method shows an order-of-magnitude improvement over the PA method, and the PA-CE and EE-PA-CE methods have errors that are 2 orders of magnitude better.

Furthermore, one can see that for the PA and EE-PA methods the mean unsigned error for the Hartree–Fock level of theory is as large or larger than the errors for the correlated methods (by definition the Hartree–Fock errors for the PA-CE and EE-PA-CE methods are zero because they include a full Hartree–Fock calculation). The fact that the errors in the PA-CE and EE-PA-CE methods are much smaller than errors in the PA and EE-PA methods indicates that the largest breakdown in the many-body and electrostatically embedded many-body expansions are in the Hartree–Fock energy and that even a pairwise treatment of the correlation energy is sufficient to reproduce the correlation energy at a given level to within 0.2 kcal/mol. Given that even the CCSD(T) level of theory is believed to be accurate only to within ~ 1 kcal/mol the difference between the pairwise methods at a given

level of electronic structure theory (with the exception of the standard pairwise additive approximation) and the conventional calculation (i.e., a calculation on the full cluster at the same level of theory) is expected to be of the same magnitude or smaller than the difference between the conventional calculation and the result from a full configuration interaction calculation.

3.3. Three-Body Methods. We continue our analysis by comparing the electronic energies predicted by the three-body method, at each level of electronic structure theory, to the full calculation for each of the five hexamers. Table 4 shows the average errors between the three-body methods (3B, 3B-CE, EE-3B, EE-3B-CE) and the full calculations at each level of theory. As expected,⁴⁵ we see that the 3B calculations are approximately an order of magnitude better than the PA results (compare Table 4 to Table 3). Also, we see that including the full Hartree–Fock energy reduces the errors by approximately an order of magnitude as one goes from the 3B to the 3B-CE method and from the EE-3B to the EE-3B-CE method. We again see that all of the correlated methods have very similar errors for each of the many-body methods and that the standard deviations are even lower for

Table 5. Signed Errors (kcal/mol) for the Many-Body Methods for the Five Hexamer Structures Relative to the Full CCSD(T) Calculation

	boat	book	cage	prism	ring	MUE ^b
PA ^a	13.67	11.71	9.93	9.74	14.06	2.47
3B	1.99	1.14	0.52	0.59	2.11	0.92
PA-CE	-0.01	0.13	-0.10	0.13	0.09	0.14
3B-CE	0.30	0.19	0.06	0.06	0.32	0.15
EE-PA	1.25	1.06	0.78	0.65	1.40	0.39
EE-3B	0.21	0.08	-0.01	-0.09	0.25	0.18
EE-PA-CE	-0.01	-0.05	-0.23	-0.28	0.04	0.17
EE-3B-CE	0.05	0.02	-0.02	-0.06	0.07	0.07

^a In this table, PA denotes PA(CCSD(T)), EE-PA denotes EE-PA(CCSD(T)), etc. ^b MUE denotes mean unsigned error of the ten comparisons of the relative energies for two structures; see section 3 of the text for an explanation.

the three-body methods than at the pairwise additive level. The 3B-CE errors are smaller than the 3B ones, and the EE-3B-CE errors are smaller than the EE-3B errors, indicating again that the dominant errors associated with the 3B and EE-3B methods are due to the Hartree–Fock energy and not the correlation energy. As seen with the pairwise additive methods, the errors associated with the three-body methods (with the possible exception of the conventional three-body method, 3B) are smaller than the intrinsic errors of the levels of electronic structure theory that are tested.

3.4. Analysis of the Many-Body Methods. Based on the results in Tables 3 and 4, the many-body methods can be ranked in order of decreasing mean unsigned error as PA \gg 3B > EE-PA \gg 3B-CE > EE-PA-CE > EE-3B > PA-CE \gg EE-3B-CE; however, Tables 3 and 4 deal only with average over errors. In order to truly understand the sources of the errors in the many-body methods, it is useful to also look at the individual error for each hexamer. Because all the correlated electronic structure methods in Table 2 show very similar results for each many-body method, only one level of electronic structure theory need be discussed in this respect. We will discuss the analysis of the CCSD(T) case, for which the key results are given in Table 5. The first set of results shown in Table 5 is the error in absolute energy for each hexamer, calculated with each many-body method, compared to the full CCSD(T) calculation (for example, the error between the energy of the cage isomer calculated at the PA(CCSD(T)) level of theory and the energy of the cage isomer calculated using a full CCSD(T) calculation). As mentioned in section 3.1, with 5 hexamers there are 10 total energy differences that one can compute (for example, the energy difference between the boat and the book or between the cage and the ring). We have calculated these 10 energy differences for each many-body method listed in Table 5 and compared them to the results from the full CCSD(T) calculations; the resulting mean unsigned error can be found in the last column of Table 5. The purpose of this analysis is not to assess the accuracy of the many-body methods (which has been done in the two previous sections) but to see if we can gain any insight into the performance of the many-body methods with respect to the many-body terms present in the structures.

Table 5 shows that the mean unsigned errors for the PA-CE, 3B-CE, EE-3B, and EE-PA-CE methods all lie within 0.04 kcal/mol of each other—as was stated in our previous discussions of Tables 2 and 3—which make them all appropriate for use on systems that require high accuracy; however, if the errors for individual structures are compared, the four methods behave quite differently. For example, the EE-PA-CE method has smaller errors for the boat, book, and ring structure than for the cage and prism, whereas the EE-3B method has smaller errors for the book, cage, and prism than for the boat and ring structures.

Because each method in Table 5 uses a different approximation to calculate the many-body effects in these clusters (i.e., neglecting some terms or including them in an average way via the point charges) it is reasonable to assume that their performance for the hexamers is directly related to the many-body effects present in the hexamers. Therefore, in order to better understand any systematic shortcomings of each many-body method we must first have a good understanding of the many-body effects in each structure.

3.4.1. General Discussion of Many-Body Effects. Before we begin this analysis we will take a moment to clarify a few terms necessary in our discussion. First of all, when we refer to smaller errors, we mean smaller absolute values of errors. Second, within a many-body expansion the total energy of the system is written as a sum of n -body terms denoted by V_n (see eq 1 in the Introduction) in which the one-body (V_1), two-body (V_2), and three-body (V_3) terms can be written as

$$V_1 = \sum_i E_i \quad (2)$$

$$V_2 = \sum_{i < j} (E_{ij} - E_i - E_j) \quad (3)$$

$$V_3 = \sum_{i < j < k} [(E_{ijk} - E_i - E_j - E_k) - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k)] \quad (4)$$

respectively, and so on for higher-order terms, and where E_i , E_{ij} , and E_{ijk} , ..., are the energies of the monomers, dimers, trimers, and so forth, in the system.

Because the electronic energies of all the monomers in the system are negative, each term of the series in eq 2 is negative, and, therefore, V_1 must be negative. For the series in eqs 3 and 4, however, each term may be positive or negative. For example, if the energy of dimer E_{ij} is higher in energy than the sum of its constituent monomer energies (i.e., an unfavorable interaction) the corresponding term in the series will be positive.

Throughout this analysis if we are talking about the series in eqs 2, 3, or 4 we will refer to them as the one-body, two-body, or three-body terms. If we are talking about the individual terms making up these series we will refer to them as an individual one-body, individual two-body, or individual three-body terms. We will also use the phrase “beyond-three-body terms” to denote the sum of the four-, five-, and six-body terms.

Table 6. Two-, Three-, and Beyond-Three-Body Terms (kcal/mol) at the CCSD(T) Level of Theory

	V_2	V_3	beyond V_3^a
boat	-31.99	-11.68	-1.99
book	-36.07	-10.57	-1.14
cage	-38.62	-9.41	-0.52
prism	-39.06	-9.16	-0.59
ring	-32.82	-11.95	-2.11

^a Beyond V_3 denotes the sum of four-, five-, and six-body terms.**Table 7.** Contribution of the Correlation Energy to the Two-, Three-, and Beyond-Three-Body Terms (kcal/mol) at the CCSD(T) Level of Theory

	V_2	V_3	beyond V_3^a
boat	-11.87	0.25	-0.30
book	-14.24	0.06	-0.19
cage	-16.15	0.16	-0.06
prism	-16.43	0.19	-0.06
ring	-12.08	0.24	-0.32

^a Beyond V_3 denotes the sum of four-, five-, and six-body terms.

3.4.2. Many-Body Effects in Water Hexamers. To analyze the many-body effects of each hexamer we will build on the insights of Pedulla and Jordan⁵¹ who have carried out a many-body analysis using the MP2 level of theory with the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets for the cage, ring, and prism isomers (optimized at the MP2/6-31+G(2d,p) level of theory). The relevant conclusions of ref 51 are as follows: (i) the cage and prism have larger two-body terms than the ring due to the presence of more hydrogen bonds, (ii) the ring isomer has larger three- and four-body terms because all of the individual two- and three-body terms have the same sign, (iii) five- and six-body terms are ≤ 0.05 kcal/mol for the cage and prism structure but are as large as 0.20 kcal/mol for the ring, and (iv) the effects of electron correlation are relatively unimportant for many-body terms beyond third order.

Table 6 shows the two-, three-, and beyond-three-body terms, at the CCSD(T)/s-cc-pVTZ level of theory, for each hexamer considered in this work. Following the work of Pedulla and Jordan we have also computed the contribution of the correlation energy to the two-, three-, and beyond-three-body terms; these results are shown in Table 7. Tables 6 and 7 show that the results obtained for our clusters at the CCSD(T)/s-cc-pVTZ level of theory are consistent with the work of Pedulla and Jordan at the MP2/aug-cc-pVTZ level of theory. We see that the three-body and beyond-three-body terms are approximately four times larger for the ring and the boat structure than for the cage and the prism and that the book structure is intermediate between these two groups. We also see that the contribution of the correlation energy to the three-body and beyond-three-body terms is approximately 2 orders of magnitude smaller than its contribution to the two-body terms, and finally we see that the magnitudes of the three-body and beyond-three-body terms are very similar.

3.4.3. Analysis of the Nonelectrostatically Embedded Methods. Table 6 shows that the terms beyond the two-body terms are smallest for the prism and cage structure and largest for the boat and ring structure. Therefore, the PA method

(which neglects all three-body and higher terms) will perform the best for the prism and the worst for the ring, which is confirmed by Table 5. At the 3B level of theory the errors are significantly reduced compared to the PA method (because only four-body and higher terms are neglected), but the results are still best for structures with small four-, five-, and six-body effects (i.e., prism and cage) and worst for structures with larger many-body effects (i.e., ring); this agrees with the results in Table 5.

At the PA-CE level, $V_1 - V_6$ are accounted for at the Hartree-Fock level, but correlation effects are considered only for the one- and two-body terms. The work of Pedulla and Jordan and the results of Table 7 show that inclusion of correlation energy has only a relatively small effect on beyond-three-body terms. As a result Table 5 shows that the PA-CE errors are much lower (nearly 2 orders of magnitude lower) than the PA errors (due to inclusion of $V_3 - V_6$ at the Hartree-Fock level). Table 5 also shows that the PA-CE method performs better than the 3B method (due to inclusion of $V_4 - V_6$ at the Hartree-Fock level). Finally, for the 3B-CE method one would expect improved performance over the three previously discussed methods, because only the four-, five-, and six-body correlation terms are neglected; however, the errors for the boat and ring are substantially larger at the 3B-CE level than at the PA-CE level; however, this can be explained by examining the contributions of correlation energy to the three-body and beyond-three-body terms. Table 7 shows that if only correlation effects are considered, the magnitude of the three-body and beyond-three-body terms are similar, but that they have different signs. Because these terms are nearly equal and opposite when both are neglected (i.e., in the PA-CE method) the errors cancel each other, and the overall error is lower than may have been expected; however, when only the latter is neglected (i.e., in the 3B-CE method) there is no such cancellation and the errors increase, particularly for structures like the ring and boat.

3.4.4. Analysis of the Electrostatically Embedded Methods. In the EE-PA approximation all two-body terms are taken into account explicitly, and the beyond-pairwise terms are accounted for in an average way by the presence of the point charges. As a result, Table 5 shows that the overall errors are substantially (approximately 1 order of magnitude) smaller than for the PA method. As a further consequence, the EE-PA method performs best for the prism and cage (which have smaller three-body and beyond-three-body terms). Table 5 also shows that the EE-PA method has lower errors than the 3B method for the ring, boat, and book structures and slightly higher errors for the cage and prism structure. These results are most likely due to not explicitly accounting for the three-body correlation terms discussed at the end of the previous section. The EE-3B method explicitly includes the three-body correlation terms, and, as a result, the errors are reduced by nearly an order of magnitude compared to the EE-PA method.

Table 5 also shows that EE-PA-CE method, by including the full Hartree-Fock energy, has smaller errors than the EE-PA method; this result is expected based on the non-EE results. The non-EE results also suggest that the largest errors

Table 8. Timings^a for MP4 and Many-Body Methods at the CCSD(T) Level of Theory, Relative to MP2 with the Same s-cc-pVTZ Basis Set^b

	timing
MP4	178
PA or EE-PA	5.0
3B or EE-3B	110
PA-CE or EE-PA-CE	5.4
3B-CE or EE-3B-CE	111

^a All calculations use the s-cc-pVTZ basis set. ^b In this table, PA denotes PA(CCSD(T)), EE-PA denotes EE-PA(CCSD(T)), etc.

should be for the cage and prism (due to not explicitly including the three-body correlation energy), which is consistent with the results in Table 5. Table 5 also shows that the EE-PA-CE errors for the ring, boat, and book structures are lower than the EE-3B errors, while the opposite is true for the cage and prism. This result can be rationalized by considering the largest error in each method, in particular failing to include the full Hartree–Fock energy in the EE-3B method and not explicitly accounting for the three-body correlation terms in the EE-PA-CE method. The EE-MB-(HF) errors for the boat, book, cage, prism, and ring are 0.16, 0.06, 0.01, -0.02 , and 0.18 kcal/mol, whereas the errors associated with not explicitly accounting for the electrostatically embedded three-body correlation energy are -0.06 , -0.07 , -0.21 , -0.21 , and -0.03 kcal/mol. The dominant errors for the two methods have quite different effects on the different structures. Based on this observation one might have predicted that if an EE-3B expansion of the Hartree–Fock energy were used with an EE-PA expansion of the correlation energy that the errors for the book, boat, cage, prism, and ring would be 0.10 , -0.01 , -0.19 , -0.24 , and 0.15 kcal/mol, respectively (assuming the errors are purely additive); the actual errors obtained are 0.15 , 0.01 , -0.21 , -0.30 , and 0.22 kcal/mol.

In the EE-3B-CE method, $V_1 - V_6$ are accounted for explicitly in the Hartree–Fock energy, the contribution to the two- and three-body terms from the correlation energy is explicitly taken into account, and the contribution of correlation energy to the higher-order terms is included in an average way. As a result, the EE-3B-CE method has the lowest errors; in particular, Table 5 shows that the error at the EE-3B-CE level has a magnitude of 0.07 kcal/mol or less for all five of the hexamers.

3.5. Timings. In order to evaluate the usefulness of the EE-MB and EE-MB-CE expansions, we must consider not only their accuracies but also their costs relative to competitive, or potentially competitive, methods. Therefore, we have computed the average times needed, on a single processor, to calculate the hexamer energies at the MP4/s-cc-pVTZ levels of theory and also for the eight many-body methods at the CCSD(T)/s-cc-pVTZ level of theory and have expressed these timings relative to the time needed to calculate the same energies at the MP2/s-cc-pVTZ level of theory with the same computer program and on the same computer (note that even ratios of timings depend on the computer program and computer). These timings are given in Table 8. First, the table shows that all eight many-body methods at the CCSD(T) level of theory are less expensive than an MP4

calculation on the full system. As the system size increases the many-body methods will become increasingly cost-effective relative to full MP4 calculations. Second, inclusion of the point charges does not change the amount of time needed to carry out the many-body expansion. Third, inclusion of the full Hartree–Fock energy causes a negligible increase in cost for these small systems.

Perhaps most importantly is that the pairwise methods are only five times more expensive than an MP2 calculation. This is important because both the PA-CE(CCSD(T)) and EE-PA-CE(CCSD(T)) methods perform better than MP2 when compared to the full CCSD(T) calculations. While the three-body methods are approximately 100 times more expensive than full MP2 calculations, all of the monomer, dimer, and trimer calculations are independent of each other which allows them each to be run on a different processor. As a result, the many-body methods are all highly parallelizable and so for moderately sized systems (on the order of 10 monomers) can be run in under a day, even at the CCSD(T) level of theory.

4. Conclusions

Many-body (MB), electrostatically embedded many-body (EE-MB), and electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE) calculations were carried out on five low-lying water hexamers and compared to full calculations at eight correlated levels of electronic structure theory ranging from MP2 theory to CCSD(T). We found that the average absolute errors associated with the many-body methods are consistent over the correlated levels of theory tested. Furthermore, when the errors obtained with each many-body method for each structure are compared they are also consistent across all levels of theory.

The mean unsigned errors in the relative energies of the structures are 0.17 and 0.07 kcal/mol, respectively, for EE-PA-CE and EE-3B-CE calculations, as compared to mean unsigned errors of 2.47 and 0.92 kcal/mol for conventional PA and 3B calculations, although the EE improvement adds negligibly to the computational cost, and the CE improvement requires only adding a Hartree–Fock calculation of the full system (which, for small- or moderate-sized systems is negligible or small, respectively). Finally, if one compares the accuracy of the many-body methods for reproducing the CCSD(T) relative (between structures) energy differences to the accuracy of full MP2 calculations (where “full” denotes without a many-body expansion, and where we note that full MP2 is commonly used in the literature for water clusters), we find that carrying out EE-MB, MB-CE, and EE-MB-CE calculations at the CCSD(T) level gives far better results, despite the method being competitive in speed if the many-body methods are run in parallel. We have also found that MP2 appears to be somewhat anomalous in that it is the only method that has a lower mean unsigned error for the EE-PA-CE method than for the EE-3B-CE method, which is probably just an accident.

In addition, we have been able to rationalize the performance of the EE-MB and EE-MB-CE methods on the various isomers in terms of the many-body effects of the clusters themselves. This will allow us to use the most cost-

effective method possible for future studies and can provide insight into the performance of these methods on other systems.

Acknowledgment. The authors are grateful to Ryan Olson for an advance copy of ref 18. This work was supported in part by the National Science Foundation under grant nos. ITR-0428774 and CHE-0704974.

References

- (1) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, 46, 618.
- (2) Cizek, J. *Adv. Chem. Phys.* **1969**, 14, 35.
- (3) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, 76, 1910.
- (4) Raghavachari, K.; Anderson, J. B. *Chem. Phys. Lett.* **1989**, 157, 479.
- (5) Scuseria, G. E.; Lee, T. J.; Schaefer, H. F., III *Chem. Phys. Lett.* **1986**, 130, 236.
- (6) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, 153, 503–506.
- (7) Scuseria, G. E.; Scheiner, A. C.; Lee, T. J.; Rice, J. E.; Schaefer, H. F., III *J. Chem. Phys.* **1987**, 86, 2881.
- (8) Lee, T. J.; Rice, J. E. *Chem. Phys. Lett.* **1988**, 150, 406.
- (9) Stanton, J. F.; Gauss, J.; Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1991**, 94, 4334.
- (10) Rendell, A. P.; Lee, T. J.; Komornicki, A. *Chem. Phys. Lett.* **1991**, 178, 462.
- (11) Rendell, A. P.; Lee, T. J.; Lindh, R. *Chem. Phys. Lett.* **1992**, 194, 845.
- (12) Rendell, A. P.; Guest, M. F.; Kendall, R. A. *J. Comput. Chem.* **1993**, 14, 1429.
- (13) Stechel, E. B., Ed. In *Domain Based Parallelism and Problem Decomposition in Computational Science and Engineering*; Keyes, D. R., Saad, Y., Truhlar, D. G., Eds.; SIAM: Philadelphia, PA, 1995.
- (14) Baker, J.; Pulay, P. *J. Comput. Chem.* **2002**, 23, 1150–1156.
- (15) Nakao, Y.; Hirao, K. *J. Chem. Phys.* **2004**, 120, 6375.
- (16) Haettig, C.; Hellweg, A.; Koehn, A. *Phys. Chem. Chem. Phys.* **2006**, 8, 1159–1169.
- (17) Ishimura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2006**, 27, 407–413.
- (18) Olson, R. M.; Bentz, J. L.; Kendall, R. A.; Schmidt, M. W.; Gordon, M. S. *J. Chem. Theory Comput.* **2007**, 3, 1312.
- (19) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, 100, 12960.
- (20) Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, 86, 914.
- (21) Galli, G.; Parrinello, M. *Phys. Rev. Lett.* **1992**, 69, 3547.
- (22) Murphy, R. B.; Beachy, M.; Ringnalda, M.; Friesner, R. J. *Chem. Phys.* **1995**, 103, 1481.
- (23) Nielsen, I. M. B.; Janssen, C. L. *J. Chem. Theory Comput.* **2007**, 3, 71.
- (24) Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, 96, 2408.
- (25) Baroni, S.; Giannozzi, P. *Europhys. Lett.* **1992**, 17, 547.
- (26) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maignet, B.; Ferenczy, G. C. *J. Comput. Chem.* **1994**, 15, 269.
- (27) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, 263, 100.
- (28) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, 313, 701.
- (29) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2002**, 119, 3599.
- (30) Christie, R. A.; Jordan, K. D. *Struct. Bonding (Berlin)* **2005**, 116, 27.
- (31) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, 122, 154102.
- (32) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, 123, 134103.
- (33) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, 110, 8777.
- (34) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, 125, 104104.
- (35) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, 433, 182.
- (36) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**.
- (37) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, 124, 114112.
- (38) Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, 28, 222.
- (39) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagese, S. *J. Comput. Chem.* **2007**, 28, 1476.
- (40) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, 111, 2722.
- (41) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, 312, 319.
- (42) Sugiki, S.; Kurita, N.; Sengoku, Y.; Sekino, H. *Chem. Phys. Lett.* **2003**, 382, 611.
- (43) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, 103, 2255.
- (44) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, 3, 46.
- (45) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, 3, 1342.
- (46) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, 98, 1358.
- (47) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1995**, 96, 6796.
- (48) Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.
- (49) Xantheas, S. S. *J. Chem. Phys.* **1994**, 100, 7523.
- (50) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. *J. Chem. Phys.* **1970**, 53, 4544.
- (51) Pedulla, J. M.; Jordan, K. D. *Chem. Phys. Lett.* **1998**, 291, 78.
- (52) Mhin, B. J.; Kim, H. S.; Kim, H. S.; Yoon, C. W.; Kim, K. S. *Chem. Phys. Lett.* **1991**, 176, 41.
- (53) Lee, C.; Chen, H.; Fitzgerald, G. *J. Chem. Phys.* **1994**, 101, 4472.
- (54) Estrin, D. A.; Paglieri, L.; Corongiu, G.; Clementi, E. *J. Phys. Chem.* **1996**, 100, 8701.
- (55) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, 112, 2063.
- (56) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2000**, 112, 9759.
- (57) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. *J. Chem. Phys.* **2002**, 116, 1493.
- (58) Losada, M.; Leutwyler, S. *J. Chem. Phys.* **2002**, 117, 2003.

- (59) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91.
- (60) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 4244.
- (61) Frisch, M. J.; Krishnan, R.; Pople, J. A. *Chem. Phys. Lett.* **1980**, *75*, 66.
- (62) Adams, G. F.; Bent, G. D.; Bartlett, R. J. In *Potential Energy Surfaces and Dynamics Calculations*; Truhlar, D. G., Ed.; Plenum: New York, 1981; p 133.
- (63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Robb, G. E. S. 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.; 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.; S. Clifford; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; P. Piskorz; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; M. A. Al-Laham; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; P. M. W. Gill; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian03—version c01* eds.; Gaussian Inc.: Wallingford, CT, 2004.
- (64) Halkier, A.; Koch, H.; Jorgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150.
- (65) Nielsen, I. M. B.; Seidl, E. T.; Janssen, C. L. *J. Chem. Phys.* **1999**, *110*, 9435.
- (66) Xantheas, S. S.; Aprà, E. *J. Chem. Phys.* **2004**, *120*, 823.
- (67) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. *J. Chem. Phys.* **2004**, *121*, 2655.
- (68) Su, J. T.; Xu, X.; Goddard, W. A., III *J. Phys. Chem. A* **2004**, *108*, 10518.
- (69) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. *J. Chem. Phys.* **2002**, *116*, 1493.
- (70) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2002**, *2*, 1009.

CT700183Y