

# Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters

Erin E. Dahlke and Donald G. Truhlar\*

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

## **Abstract:**

The use of background molecular charge to incorporate environmental effects on a molecule or active site is widely employed in quantum chemistry. In the present article we employ this practice in conjunction with many-body expansions. In particular, we present electrostatically embedded two-body and three-body expansions for calculating the energies of molecular clusters. The system is divided into fragments, and dimers or trimers of fragments are calculated in a field of point charges representing the electrostatic potential of the other fragments. We find that including environmental point charges can lower the errors in the electrostatically embedded pairwise additive (EE-PA) energies for a series of water clusters by as much as a factor of ten when compared to the traditional pairwise additive approximation, and that for the electrostatically embedded three-body (EE-3B) method the average mean unsigned error over nine different levels of theory for a set of six tetramers and one pentamer is only 0.05 kcal/mol, which is only 0.4% of the mean unsigned net interaction energy. We also test the accuracy of the EE-PA and EE-3B methods for a cluster of 21 water molecules and find that the errors relative to a full MP2/aug'-cc-pVTZ calculation to be only 2.97 and 0.38 kcal/mol, respectively, which are only 1.5% and 0.2% respectively of the net interaction energy. This method offers the advantage over some other fragment-based methods that it does not use an iterative method to determine the charges, and thus provides substantial savings for large clusters. The method is convenient to adapt to a variety of electronic structure methods and program packages, it has  $N^2$  or  $N^3$  computational scaling for large systems (where  $N$  is the number of fragments), and it is easily converted to an  $O(N)$  method, and its linearity allows for convenient analytic gradients.

## 1. Introduction

Many computational chemistry applications require one to calculate accurate energies for very large molecules such as proteins, for large molecular clusters or nanoparticles that have hundreds or even thousands of atoms, or for condensed-phase extended systems (such as liquid solvents or high-pressure solids in the earth's mantle). Such requirements have led to research on developing quantum mechanical electronic structure methods to efficiently and accurately describe large systems.<sup>1–45</sup> Correlated quantum mechanical methods, such as post-Hartree–Fock wave function theory<sup>46</sup> (WFT) or density functional theory<sup>47</sup> (DFT), while able to give highly accurate results, are limited — in their original computational formulation — to relatively small systems containing only tens, or in the case of DFT, hundreds or a few thousand atoms. If one is interested in studying a system with tens or hundreds of thousands of atoms, conventional quantum mechanical methods are useless, and for systems with only 50 – 100 atoms, they are often so computationally demanding as to be impractical with the time and resources available for a given project. Molecular mechanics,<sup>48,49</sup> on the other hand, while able to handle the large number of atoms present in extended systems does not provide the predictive quantitative accuracy needed to investigate many of the questions of interest, such as chemical reactions or phase equilibria under conditions (for example, the high pressure phases of the earth's mantle) where there is not enough data for a specific parameterization.

Perhaps the simplest of all approximations that is frequently invoked is to expand the system as a sum of many-body terms, and then to truncate the series after only a few terms. If the sum is truncated after inclusion of the two-body terms (interactions of two

monomers) one is said to have made the pairwise additive approximation, while if three-body terms are also included one is said to have made the three-body approximation. The benefit of making such approximations is that the total energy of the system can be written as a combination of the energy of all the fragments (called monomers), dimers, and, in the case of the three-body approximation, trimers within the structure. This reduces a single very large and expensive calculation to a large number of small and computationally efficient calculations, and it can allow quantum mechanics methods to be used for large systems where it would be otherwise impractical. Many-body expansions are applicable both to covalently connected monomers, such as the peptide residues in a protein, and to noncovalently connected monomers, such as water molecules in ice, liquid water, or water clusters. Additionally, these methods can be highly parallelized because each fragment can be calculated on a separate processor, making the calculation quite fast. Unfortunately, while the pairwise additive approximation may give qualitatively correct results it does not give good quantitative results, even for small clusters with highly accurate pair potentials.<sup>50</sup> While the inclusion of the three-body terms helps to improve quantitative accuracy,<sup>51</sup> the results may still be insufficiently accurate for systems in which many-body effects play an important role, such as water clusters.<sup>52</sup>

As a result of these shortcomings, some workers have tried to modify the pairwise and three-body approximations by adding additional terms to the monomer, dimer, or trimer Hamiltonian that account for the electrostatic field of the other atoms in the system.<sup>16,17</sup> These methods, which include the pair interaction molecular orbital method<sup>16</sup> (PIMO) and the fragment molecular orbital method<sup>17,24,29,30</sup> (FMO) of

Kitaura *et al.*, add terms to the fragment and dimer Hamiltonians that account for the electrostatic potential of the other fragments in the system. These methods involve an iterative procedure to determine the electrostatic potential that occurs in the fragment, dimer, and in some cases trimer, Hamiltonians. In this procedure one divides the system into the fragments and provides an initial guess for the electron distribution of each fragment. The Schrödinger equation is then solved for all fragments in the system to obtain both the fragment energies and also a new electron density distribution. The new guess is plugged back into the Hamiltonian, and the cycle is repeated until the electron density distributions are converged. FMO calculations in which a unit of two consecutive peptide residues are treated as a monomer have been shown to reproduce *ab initio* molecular orbital energies of polypeptides to within 2 kcal/mol; however, one might wonder if there is a way to attain similar accuracy without the use of an iterative procedure, particularly for noncovalently connected monomers where one can easily separate the system into fragments without having to cut through any covalent bonds. Some approximations to the electrostatic potential have been made<sup>24</sup> in which the electrostatic potential was treated using a system of Mulliken and fractional point charges, depending on the separation of the fragments, and the expression for the total energy was written in terms of a density difference matrix so that only the net contribution of the electrostatic energy was included. Since this method led to no significant loss of accuracy when compared to the original FMO method, it prompts one to wonder if there is a simpler way to approximate the electrostatic potential, without sacrificing accuracy.

Recent work by Jiang *et al.*<sup>43</sup> has modified the molecular fragmentation with conjugated caps (MFCC) method of Zhang *et al.*<sup>25,26</sup> so that the energy of each fragment is calculated in a field of point charges representing the other charge centers. Their method, called electrostatic field-adapted molecular fragmentation with conjugated caps (EFA-MFCC), shows errors within a few millihartrees of the *ab initio* energy for a set of model peptides and biological molecules. In the EFA-MFCC method the total energy is written as the sum of the energies of the capped fragments minus the energy of the conjugated caps (see references 25, 26, and 43). In light of the discussion of many-body expansions presented in the introduction, one could consider the energy expression of MFCC and EFA-MFCC to be an electrostatically embedded one-body expansion of the total energy. In light of the good results obtained with the by Jiang *et al.* one might ask whether using higher order many-body expansions (i.e., pairwise additive or three-body) in simple electrostatic fields would be even more accurate, particularly in light of the work of Kitaura *et al.*<sup>29, 30</sup> who have successfully used a many-body expansion with more complicated ways of representing the electrostatic potential (see previous paragraph and references 24, 28–30).

In this work, inspired by the successes of the many-body expansions within a self-consistent field representing the electrostatic potential of the other particles in the work of Federov and Kitaura<sup>29,30</sup> and similar in some respects to the work of Jiang *et al.*<sup>43</sup> in which a simpler electrostatic embedding is applied to fragments, we combine the use of a many-body expansion with the use of prespecified point charges to represent the electrostatic field of the other molecules. In order to assess the accuracy of the method we have used it to calculate the binding energies of a series of water clusters ranging in

size from trimer to pentamer. Because water clusters are known to have very large many-body effects,<sup>52</sup> this should provide a good test of the capabilities of the method.

## 2. Theory

Consider a system of  $N$  interacting units, called monomers. In the present paper we limit the treatment to noncovalently connected monomers, but, like the fragment molecular orbital method,<sup>29</sup> it can be extended to covalently connected monomers.

Without any approximation, the total energy of the system can be written as:

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

where

$$V_1 = \sum_i^N E_i, \quad (2)$$

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

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

with higher-order terms defined analogously, and where  $E_i$ ,  $E_{ij}$ ,  $E_{ijk}$ , ... are energies of embedded monomers, embedded dimers, embedded trimers, and so forth. By an embedded  $n$ -mer, we mean one whose energy is calculated in a field representing the  $N - n$  other particles. If one retains only the first two terms of equation 1, the total energy of the system can be approximated as:

$$E_{\text{EE-PA}} = \sum_{i<j}^N E_{ij} - (N - 2) \sum_i^N E_i \quad (5)$$

where EE-PA denotes the electrostatically embedded pairwise additive method. If one includes the first three terms in equation 1, the total energy becomes:

$$E_{\text{EE-3B}} = \sum_{i<j<k}^N E_{ijk} - (N-3) \sum_{i<j}^N E_{ij} + \frac{(N-2)(N-3)}{2} \sum_i^N E_i \quad (6)$$

where EE-3B denotes the electrostatically embedded three-body method. (EE-MB denotes an electrostatically embedded many-body approximation of unspecified order.) If one calculates the energies using equations 5 and 6 without the presence of the point charges, one gets the conventional pairwise additive (PA) and three-body (3B) energies, as discussed in the introduction. We emphasize that if the series of equation 1 is not truncated, the result is exact and is independent of whether or not embedding is employed; however, we will show that the rate of convergence of the series (i.e., the accuracy if one truncates after a given order of many-body terms) depends strongly on embedding.

In the present study, the electrostatic embedding is carried out by using point charges, as in ref. 43, to represent the other  $N - n$  particles for each  $n$ -mer. Two possible ways to obtain the set of point charges needed for these methods are considered in this work:

- A. Determine a charge representation for the entire cluster; then, for each monomer, dimer, or trimer, represent the other  $N - 1$ ,  $N - 2$ , or  $N - 3$  water molecules with the charges from this full-system charge calculation.
- B. Determine the charges for a gas-phase monomer, and, for each monomer, dimer, or trimer, represent the other  $N - 1$ ,  $N - 2$ , or  $N - 3$  monomers with the gas-phase monomer point charges.

In order to compare method A to method B we have used the AM1<sup>53</sup> Mulliken charges<sup>54</sup> to carry out both (denoted AM1 and AM1M for methods A and B, respectively). Since for a very large system the calculation of the charges for the entire cluster may become

very expensive we are especially interested in method B (monomer charges), and test the monomer-based charge scheme, using three other choices for the monomer charges: B3LYP<sup>55-57</sup>/6-31G\*<sup>58</sup> Mulliken charges (denoted B3LYPM), B3LYP/6-31G\* class IV charges<sup>59</sup> (denoted CM4M), and TIP3P<sup>61</sup> molecular mechanics charges (denoted TIP3P). (Note that Mulliken charges are class II charges and CM4 is a class IV charge model.<sup>60</sup>) All AM1 calculations were carried out using the *Gaussian03*<sup>62</sup> software package, while the B3LYP/6-31G\* CM4 and Mulliken charges were calculated using the MN-GSM software program, version 6.0<sup>63</sup> (Note that the B3LYP Mulliken charges can be calculated with any electronic structure package employing B3LYP that can calculate Mulliken charges.)

The procedure for the electrostatically embedded many-body method can be summarized as follows:

1. Determine the partial atomic charges for all  $N$  molecules in the cluster by one of the methods described in the previous paragraph.
2. Calculate the energy of each  $n$ -body cluster in the expansion (for EE-PA, one has  $n = 1$  and 2; for EE-3B, one has  $n = 1, 2,$  and 3) by representing the atoms of the other  $N - n$  molecules with nuclear-centered point charges corresponding to the partial atomic charges determined in step 1.
3. Calculate the total energy of the system using the appropriate equation (i.e., equation 5 for EE-PA, equation 6 for EE-3B).

We note that the internal geometries (O-H bond distances and H-O-H bond angles) of the monomers for those clusters taken from simulations vary within the cluster, and in some cases they are also quite different from those clusters obtained from



gas-phase optimizations (the bond lengths and angles for the optimized clusters tend to be similar). Thus in principle, many of the monomers in these clusters have slightly different charges, and indeed when one uses method A the charges are different. One could also laboriously calculate all the slightly different monomer charge sets, but when we use the monomer charges we use a simpler procedure – all charges correspond to the equilibrium structure of the monomer (taken to have the O-H distances of 0.9572 Å and a bond angle of 104.52 degrees).

### 3. Results and Discussion

In previous work<sup>50</sup> we have tested the ability of the pairwise additive approximation to reproduce full density functional calculations of binding energies of small water clusters ranging in size from trimer to pentamer. The database used in that work consists of a set of eight trimers, six tetramers, and a pentamer; the structures are taken from Monte Carlo and molecular mechanics simulations of bulk water and ice, as well as from gas-phase optimizations. We have chosen to use the same set of 15 clusters in this work, and we refer the reader to reference 50 for more information about the database. All tests of the many-body and electrostatically embedded many-body results are tests of how well these approximations can reproduce the total energy from full calculations at a given level of theory, not how well they agree with higher-level calculations. For example EE-PA and EE-3B calculations of the energy with PBE1W/6-311+G(2d,2p) monomers, dimers, and trimers are compared to full PBE1W/6-311+G(2d,2p) calculations on the whole cluster, which we refer to as the true energies.

While the main subject of the present work is the use of prespecified point charges in a many-body expansion, one might be interested in seeing the results for electrostatic embedding with only the one-body term. For the interested reader these results and a brief discussion have been placed in supporting information.

For each of the fifteen clusters we have calculated the true energies at a given level of electronic structure theory and basis set, the pairwise additive energies at the same level of theory and basis set, and the electrostatically embedded pairwise additive energies with that level and basis set. For the EE-PA energies we tested the five charge approximations described above: AM1, AM1M, B3LYPM, CM4M, and TIP3P. Each of these five energies were calculated at five different levels of theory with one or two basis sets. The levels of electronic structure theory are: BLYP,<sup>55,56</sup> B3LYP,<sup>55–57</sup> PBE,<sup>64</sup> PBE1W,<sup>49</sup> MP2.<sup>65</sup> Four of these are methods that were also used in reference 50; they include three generalized gradient approximations (BLYP, PBE, PBE1W), or GGAs, and one hybrid GGA (B3LYP), that also includes a percentage (20%) of Hartree–Fock exchange. The fifth level of theory is MP2 (second order Møller–Plesset perturbation theory<sup>65</sup>), both to illustrate the electrostatically embedded many-body expansion for a WFT method and because MP2 is known to be highly accurate for small water clusters.<sup>66,67</sup> For each of the density functionals we use the MG3S basis set (which for water is the same as the 6-311+G(2df,2p)<sup>58</sup> basis set), and also the basis set, different for each functional, that was determined previously<sup>50</sup> to be optimal for small water clusters. Three of the functionals (BLYP, B3LYP, and PBE1W) use a Pople basis set,<sup>58</sup> while one functional (PBE) uses an augmented Dunning basis set;<sup>68</sup> in particular we consider the BLYP/6-31+G(d,p), B3LYP/6-31+G(d,2p), PBE/aug-cc-pVTZ, and

PBE1W/6-311+G(2d,2p) levels of theory. For the MP2 calculations we use the cc-pVTZ<sup>69</sup> basis set on hydrogen and the aug-cc-pVTZ basis set on oxygen; this combination will be referred to as the aug'-cc-pVTZ basis set. All density functional and MP2 calculations were carried out with the *Gaussian03* program.<sup>62</sup>

Before discussing the results, it is useful to provide some characterization of the clusters. Seven of the eight trimers are bound with respect to three optimized monomers for all nine levels of theory tested. The net binding energies for these seven trimers range, depending on electronic structure level and the particular trimer, from 0.2 to 17.3 kcal/mol, with an average of 13.0 kcal/mol. The only exception, taken from a Monte Carlo simulation of liquid water, is predicted to be unbound by as much as 3.2 kcal/mol by MP2/aug'-cc-pVTZ and to be bound at most by 2.1 kcal/mol with BLYP/6-31+G(d,p). The tetramers are all bound with respect to four optimized monomers; depending on the electronic structure level and the particular tetramer, the net binding energy ranges from 1.9 to 30.8 kcal/mol, with an average binding energy of 13.6 kcal/mol. The pentamer, in contrast, taken from a simulation of liquid water, is unbound by 8.8 to 16.4 kcal/mol, depending on the electronic structure method.

Table 1 shows the mean errors of the pairwise additive and electrostatically embedded pairwise additive energies, relative to the true energies calculated at the same level of theory. We can see that even for the worst EE-PA result, the one with AM1M charges, the average mean unsigned error is reduced by a factor of two. Moreover, the best two methods, EE-PA-B3LYPM and EE-PA-TIP3P, show a tenfold reduction in error. For the EE-PA-AM1 and EE-PA-AM1M calculations we see improved performance of the full cluster calculation over the gas-phase monomer charges; however

the difference in the average mean unsigned error in the nine methods is only 0.15 kcal/mol indicating that the use of monomer charges is sufficient for this application. The best overall result comes from the use of the TIP3P charges, which were parameterized (along with two Lennard-Jones parameters) to improve the energetics and liquid density of simulations of liquid water.<sup>60</sup> We note that Jiang *et al.*<sup>43</sup> saw a similar result, in that while the energy obtained is not independent of the charges used, the choice of charge model does not greatly influence the quality of the result obtained.

As the two best results, EE-PA-B3LYPM and EE-PA-TIP3P, use only the gas-phase monomer charges, which are related by  $q_H = -q_O/2$ , one might be able to use  $q_O$  as a variational parameter in order to obtain even better results. Optimization of  $q_O$  to minimize the average mean unsigned error of the nine methods led to  $q_O = -0.8972$ . Table 2 compares the mean errors for the optimized charges (denoted QOPT) to the other EE-PA methods that use monomer charges. We find that optimizing  $q_O$  does not lead to significant improvement over the TIP3P charges. In fact, the average mean unsigned errors over all nine methods for the EE-PA approximation with the AM1M, B3LYPM, CM4M, TIP3P, and QOPT charges methods are 0.80, 0.21, 0.34, 0.19, and 0.19 kcal/mol, respectively. In light of the small difference in errors between the B3LYPM, TIP3P, and QOPT methods, and in the interest of developing a method that is generally transferable to other systems, we have decided to use the B3LYP/6-31G\* monomer Mulliken charges (B3LYPM) because (unlike the TIP3P charges) they are well defined for any system of interest, and thus they avoid the need to optimize a new set of charges for each system of interest.

Table 3 compares the three-body and electrostatically embedded three-body energies to the true energies for the tetramers and pentamer. Based on the results for the electrostatically embedded pairwise additive methods we test only the B3LYP/6-31G\* monomer Mulliken (B3LYPM) charges. In all cases we see that the EE-3B method gives a mean unsigned error of less than 0.1 kcal/mol; in the case of MP2/aug'-cc-pVTZ we see a mean unsigned error of only 0.01 kcal/mol. Again we see a minimum of a twofold decrease in the error upon inclusion of the point charges, and as much as a tenfold improvement for MP2. The average mean unsigned error for the nine methods is 0.05 kcal/mol, compared to an average mean unsigned error of 0.31 kcal/mol for the EE-PA scheme applied to the tetramers and pentamer. The average error of 0.05 kcal/mol in the EE-3B energies corresponds to a relative error of 0.4% when compared to the average unsigned net interaction energy of the tetramers and pentamer, which is 13.6 kcal/mol; the average mean unsigned error of 0.31 kcal/mol for the EE-PA scheme applied to this same set of clusters corresponds to 2.3%, which is also small enough to be useful for demanding applications.

In order to test the new methods for larger clusters, we have applied them to calculate the MP2/aug'-cc-pVTZ energy of a cluster of 21 water molecules (see Figure 1), taken from the Cambridge Cluster Database,<sup>70</sup> which corresponds to the global minimum-energy structure for  $N = 21$  with the TIP5P<sup>71</sup> empirical potential.<sup>72</sup> We find that the EE-PA method gives an error of 2.97 kcal/mol relative to the true energy, while EE-3B gives an error of only 0.38 kcal/mol. These values are larger than the mean unsigned errors for the smaller clusters because the 21-mer has a much larger interaction energy, due to the large number of hydrogen bonds (see Fig. 1). In particular the energy

of the 21-mer is 203.2 kcal/mol lower than 21 separated gas-phase water monomers. The errors in the EE-PA and EE-3B net interaction energies are thus 1.5% and 0.2%, respectively.

In order to put these results into perspective Table 4 shows FMO results for DFT calculations on water clusters containing 16 and 32 water molecules.<sup>30</sup> In general we see that our errors are lower than those for the FMO methods with  $N = 16$ . Federov *et al.* assume that the error scales linearly with the system size,<sup>29,30</sup> indicating that for the same system size the method presented here should be more accurate than FMO. The FMO2 and FMO3 methods perform better for small-basis-set Hartree–Fock calculations, but we have not tested the EE-MB schemes for those levels of calculation, which are less accurate for polarization and which do not include correlation.

It is also interesting to note the wide range in performance that one sees in Table 4 with respect to varying basis set. While we see some dispersion in Tables 1 and 2, the spread is not as severe as what is seen in Table 4, even after adjusting for system size.

An important feature of the EE-PA and EE-3B methods is their favorable scaling in the limit of large system size. For example, the calculation of the EE-3B energy for the 21-mer took only ~55 hours using a single processor on an SGI Altix, while the full MP2/aug'-cc-pVTZ calculation took ~86 hours using eight processors on the same machine (a factor of 12.5 times longer). In theory, in the limit of large system size and independent of what method is used for the trimer, the EE-3B method scales as  $N^3$  where  $N$  is the number of monomers. Thus the method shows great promise for allowing calculations of MP2 or CCSD(T) (coupled cluster theory with single and double excitations and quasiperturbative connected triples<sup>73</sup>) accuracy on large systems with  $N^3$

scaling, whereas conventional MP2 and CCSD(T) scale as  $N^5$  and  $N^7$  respectively. The EE-PA method has even more favorable  $N^2$  scaling. If one makes a further approximation of including only pairs or trimers within a certain cutoff distance of each other (for example, only neighboring monomers), the method scales linearly in  $N$ . The use of a cutoff has particular value if one is interested in using periodic boundary conditions, as it would allow for the study of a large periodic system with highly accurate quantum mechanical methods at a relatively low cost. This further approximation was not made in the present work, but is an interesting topic for further study.

Another important issue to emphasize is simplicity. Since essentially all modern electronic structure packages allow calculations in the field of point charges (or the one-electron part of the Hamiltonians can be easily modified to allow this), the present EE-PA and EE-3B methods can be carried out by writing a simple script to drive virtually any electronic structure package. Furthermore, if a new variant of correlated WFT becomes available, for example a new coupled cluster approximation, it can immediately be employed for the  $n$ -mers of the electrostatically embedded many-body series without method-specific programming. An additional benefit is that performing an EE-PA or EE-3B calculation is equivalent to carrying out a full quantum mechanical calculation on the system, unlike combined QM/MM models<sup>13,20,22</sup> that treat only part of the system quantum mechanically.

Finally we emphasize the linearity of the method. Equations 1 – 6 are all linear in the energies. Thus if analytic gradients or Hessians are available for the monomers, they are immediately available for the EE-MB energy. For example

$$\nabla E_{\text{EE-PA}} = \sum_{i<j}^N \nabla E_{ij} - (N-2) \sum_i^N \nabla E_i. \quad (7)$$

In contrast to this simplicity, most other fragment-based methods require extra programming or even assumptions to obtain analytic gradients. Fast and accurate gradients are essential for dynamics calculations.

#### 4. Summary and Conclusions

We present here a many-body series called the electrostatically embedded many-body expansion that merges the many-body expansion, as developed extensively by Federov and Kitaura,<sup>29,30</sup> with the use of prespecified point charges on fragments as proposed by Jiang *et al.*<sup>43</sup> We have implemented it using fixed monomer partial atomic charges for the electrostatic embedding, and we have demonstrated its accuracy for a set of water clusters ranging in size from trimer to 21-mer. For 8 trimers, 6 tetramers, and a single pentamer, the electrostatically embedded pairwise additive approximation yields a tenfold reduction in error as compared to the conventional pairwise additive method, with an average mean unsigned error of approximately 0.21 kcal/mol when the B3LYP/6-31G\* monomer Mulliken charges are used. When three-body terms are included with the same set of charges, the average mean unsigned error is only 0.05 kcal/mol, as compared to 0.17 kcal/mol if no electrostatic embedding is used, an improvement of more than a factor of three. In the case of the 21-mer, the EE-PA approximation leads to an error in the total energy of only 2.97 kcal/mol while the EE-3B approximation gives an error of only 0.38 kcal/mol relative to a full MP2/aug'-cc-pVTZ calculation.

While we have only tested the method up to the three-body terms, it is easily generalized to include as many of the  $n$ -body terms as desired. In the future we hope to



apply this method to the calculation of binding energies for much larger clusters and to evaluate possible choices of charge model for other types of molecular clusters.

**Supporting Information Available:** Results for truncating the many-body expansion after  $V_1$ . Sample scripts for carrying out EE-PA and EE-3B calculations on water clusters are available online at <http://comp.chem.umn.edu/truhlar/>.

### Acknowledgments

This work was supported in part by the National Science Foundation under grants nos. CHE03-49122 and ITR-0428774.

### References

- (1) Newton, M. D.; Boer, F. P.; Lipscomb, W. N. *J. Amer. Chem. Soc.* **1966**, *88*, 2353.
- (2) Degand, P.; Leroy, G.; Peeters, D. *Theor. Chim. Acta* **1973**, *30*, 243.
- (3) Niessen, W. V. *Theor. Chim. Acta* **1973**, *31*, 111.
- (4) Stechel, E. B. In *Domain-Based Parallelism and Problem Decomposition in Computational Science and Engineering*; Keyes, D. R., Saad, Y., Truhlar, D. G., Eds.; SIAM: Philadelphia, 1995; p. 217.
- (5) Maseras, F.; Morokuma, K. *J. Comp. Chem.* **1995**, *16*, 1170.
- (6) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Seiber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (7) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- (8) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744.
- (9) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654.

- (10) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268.
- (11) Challacombe, M.; Schwegler, E. *J. Chem. Phys.* **1997**, *106*, 5526.
- (12) Sierka, J.; Sauer, J. *Faraday Discuss. Chem. Soc.* **1997**, *106*, 41.
- (13) *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J.; Thompson, M. A.; Eds.; ACS Symp. Ser. 712; American Chemical Society: Washington, 1998.
- (14) Galli, G. *Comp. Mat. Sci.* **1998**, *12*, 242.
- (15) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
- (16) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319.
- (17) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
- (18) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336.
- (19) Papanicolaou, N. I.; Kallinteris, G. C.; Evangelakis, G. A.; Papaconstantopoulos, D. *Comp. Mat. Sci.* **2000**, *17*, 224.
- (20) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendurst, J.; Ed.; NIC Ser. 3; Neumann Institute for Computing: Jülich, 2000; p 285.
- (21) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
- (22) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

- (23) Titmuss, S. J.; Cummins, P. L.; Rendall, A. P.; Bliznyuk, A. A.; Gready, J. E. *J. Comp. Chem.* **2002**, *23*, 1314.
- (24) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
- (25) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
- (26) Zhang, D. W.; Chen, X. H.; Zhang, J. Z. H. *J. Comp. Chem.* **2003**, *24*, 1846.
- (27) Assfeld, X.; Ferré, N.; Rivail, J.-L. *Theor. Chem. Acc.* **2004**, *111*, 328.
- (28) Federov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comp. Chem.* **2004**, *25*, 872.
- (29) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
- (30) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.
- (31) Re, S.; Morokuma, K. *Theor. Chem. Acc.*, **2004**, *112*, 59.
- (32) Miyazaki, T.; Bowler, D. R.; Choudhury, R.; Gillan, M. J. *J. Chem. Phys.* **2004**, *121*, 6186.
- (33) Staszewska, G.; Staszewski, P.; Schultz, N. E.; Truhlar, D. G. *Phys. Rev. B* **2005**, *71*, 45423.
- (34) Wada, M.; Sakurai, M. *J. Comp. Chem.* **2005**, *26*, 160.
- (35) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 31103.
- (36) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- (37) Monard, G.; Bernal-Uruchurtu, van der Vaart, A.; Merz, K. M. Jr.; Ruiz-López, M. F. *J. Phys. Chem. A* **2005**, *109*, 3425.
- (38) Federov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
- (39) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.

- (40) Federov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. *J. Comp. Chem.* **2006**, *27*, 976.
- (41) Ishida, T.; Federov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457.
- (42) Canfield, P.; Dahlborn, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, *124*, 24301.
- (43) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
- (44) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *124*, 184703.
- (45) Vashista, P.; Kalia, R. K.; Nakano, A. *J. Phys. Chem. B* **2006**, *110*, 3727.
- (46) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
- (47) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253.
- (48) Bowen, J. P.; Allinger, N. L. *Rev. Comp. Chem.* **1991**, *2*, 81.
- (49) Dinur, U.; Hagler, A. T. *Rev. Comp. Chem.* **1991**, *2*, 99.
- (50) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *109*, 15677.
- (51) Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975.
- (52) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.
- (53) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Amer. Chem. Soc.* **1985**, *107*, 3902.
- (54) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- (55) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (56) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (57) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

- (58) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- (59) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Theor. Chem. Acc.* **2005**, *113*, 133.
- (60) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Design* **1995**, *9*, 87.
- (61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.

- (63) Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 6.0, University of Minnesota, Minneapolis, MN 55455-0431, 2006.
- (64) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (65) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (66) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. *J. Chem. Phys.* **2002**, *116*, 1493.
- (67) Xantheas, S. S.; Aprà, E. *J. Chem. Phys.* **2004**, *120*, 823.
- (68) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
- (69) Dunning, T. H. Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (70) Wales, D. J.; Doye, J. P. K.; Dullweber, A.; Hodges, M. P.; Naumkin, F. Y.; Calvo, F.; Hernández-Rojas, J.; Middleton, T. F. *Cambridge Cluster Database* <http://www-wales.ch.cam.ac.uk/CCD.html> (accessed 03/08/06).
- (71) Mohoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.
- (72) James, T.; Wales, D. J.; Hernández-Rojas, J. *Chem. Phys. Lett.* **2005**, *415*, 302.
- (73) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

**Table 1. Mean Errors<sup>a</sup> (kcal/mol) in Binding Energies for the Pairwise Additive Approximation and the Electrostatically Embedded Pairwise Additive Approximation with Five Point-Charge Models**

	PA			EE-PA														
				AM1			AM1M			B3LYPM			CM4M			TIP3P		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
BLYP/MG3S	1.99	2.11	2.66	0.74	0.76	1.01	0.91	0.94	1.23	0.21	0.21	0.32	0.42	0.42	0.58	0.14	0.15	0.24
PBE/MG3S	1.62	1.83	2.32	0.39	0.58	0.74	0.56	0.69	0.92	-0.13	0.21	0.32	0.07	0.28	0.38	-0.20	0.23	0.34
PBE1W/MG3S	1.58	1.80	2.28	0.35	0.56	0.72	0.52	0.67	0.89	-0.17	0.24	0.35	0.03	0.28	0.38	-0.24	0.26	0.38
B3LYP/MG3S	1.87	2.00	2.52	0.65	0.68	0.91	0.82	0.85	1.12	0.14	0.14	0.25	0.34	0.34	0.50	0.07	0.10	0.18
BLYP/6-31+G(d,p)	1.81	1.95	2.47	0.74	0.77	1.02	0.90	0.93	1.22	0.27	0.27	0.39	0.46	0.47	0.64	0.21	0.21	0.31
PBE/aug-cc-pVTZ	1.82	2.02	2.53	0.43	0.58	0.76	0.62	0.72	0.98	-0.19	0.25	0.36	0.05	0.27	0.37	-0.27	0.28	0.41
PBE1W/6-311+G(2d,2p)	1.59	1.80	2.29	0.36	0.56	0.72	0.53	0.67	0.90	-0.17	0.24	0.35	0.04	0.28	0.38	-0.23	0.26	0.38
B3LYP/6-31+G(d,2p)	1.84	1.97	2.49	0.68	0.71	0.95	0.84	0.87	1.15	0.17	0.17	0.29	0.37	0.38	0.54	0.11	0.13	0.21
MP2/aug'-cc-pVTZ <sup>b</sup>	1.93	2.04	2.55	0.65	0.68	0.88	0.82	0.85	1.10	0.13	0.13	0.20	0.33	0.33	0.45	0.07	0.09	0.13

<sup>a</sup> MSE, MUE, and RMSE denote mean signed, mean unsigned, and root mean squared errors respectively.

<sup>b</sup> aug'-cc-pVTZ denotes using the cc-pVTZ basis set on hydrogen and the aug-cc-pVTZ basis set on oxygen.

**Table 2. Comparison of the Mean Errors (kcal/mol) for Five Kinds of Point Charges in the EE-PA Method**

	AM1M <sup>a</sup>			B3LYPM <sup>b</sup>			CM4M <sup>c</sup>			TIP3P <sup>d</sup>			QOPT <sup>e</sup>		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
BLYP/MG3S	0.91	0.94	1.23	0.21	0.21	0.32	0.42	0.42	0.58	0.14	0.15	0.24	0.07	0.09	0.16
PBE/MG3S	0.56	0.69	0.92	-0.13	0.21	0.32	0.07	0.28	0.38	-0.20	0.23	0.34	-0.26	0.26	0.39
PBE1W/MG3S	0.52	0.67	0.89	-0.17	0.24	0.35	0.03	0.28	0.38	-0.24	0.26	0.38	-0.30	0.30	0.44
B3LYP/MG3S	0.82	0.85	1.12	0.14	0.14	0.25	0.34	0.34	0.50	0.07	0.10	0.18	0.01	0.08	0.13
BLYP/6-31+G(d,p)	0.90	0.93	1.22	0.27	0.27	0.39	0.46	0.47	0.64	0.21	0.21	0.31	0.15	0.15	0.23
PBE/aug-cc-pVTZ	0.62	0.72	0.98	-0.19	0.25	0.36	0.05	0.27	0.37	-0.27	0.28	0.41	-0.35	0.35	0.48
PBE1W/6-311+G(2d,2p)	0.53	0.67	0.90	-0.17	0.24	0.35	0.04	0.28	0.38	-0.23	0.26	0.38	0.01	0.06	0.09
B3LYP/6-31+G(d,2p)	0.84	0.87	1.15	0.17	0.17	0.29	0.37	0.38	0.54	0.11	0.13	0.21	0.04	0.09	0.15
MP2/aug'-cc-pVTZ <sup>f</sup>	0.82	0.85	1.10	0.13	0.13	0.20	0.33	0.33	0.45	0.07	0.09	0.13	-0.30	0.30	0.43

<sup>a</sup>  $q_O = -0.385115$ ,  $q_H = 0.1925575$

<sup>b</sup>  $q_O = -0.778311$ ,  $q_H = 0.3891555$

<sup>c</sup>  $q_O = -0.637$ ,  $q_H = 0.319$

<sup>d</sup>  $q_O = -0.834$ ,  $q_H = 0.417$

<sup>e</sup>  $q_O = -0.8972$ ,  $q_H = 0.4486$

<sup>f</sup> aug'-cc-pVTZ using the cc-pVTZ basis set on hydrogen and the aug-cc-pVTZ basis set on oxygen.



**Table 3. Mean Errors (kcal/mol) for Three-Body and Electrostatically Embedded Three-Body Binding Energies**

	3B			EE-3B <sup>a</sup>		
	MSE	MUE	RMSE	MSE	MUE	RMSE
BLYP/MG3S	-0.07	0.18	0.29	-0.02	0.03	0.05
PBE/MG3S	-0.13	0.18	0.32	-0.08	0.08	0.11
PBE1W/MG3S	-0.14	0.17	0.32	-0.09	0.09	0.12
B3LYP/MG3S	-0.07	0.16	0.27	-0.02	0.03	0.04
BLYP/6-31+G(d,p)	-0.06	0.17	0.27	-0.02	0.03	0.04
PBE/aug-cc-pVTZ	-0.10	0.16	0.29	-0.07	0.07	0.09
PBE1W/6-311+G(2d,2p)	-0.13	0.17	0.32	-0.08	0.08	0.12
B3LYP/6-31+G(d,2p)	-0.07	0.16	0.25	-0.02	0.02	0.04
MP2/aug'-cc-pVTZ <sup>b</sup>	-0.05	0.14	0.21	-0.01	0.01	0.01

<sup>a</sup> based on B3LYPM point charges

<sup>b</sup> aug'-cc-pVTZ denotes the use of the cc-pVTZ basis set on hydrogen and the aug-cc-pVTZ basis set on oxygen.

**Table 4 Errors (kcal/mol) in FMO2<sup>a</sup> and FMO3<sup>b</sup> Approximations to True Energies for Clusters of *N* Water Molecules<sup>c</sup>**

	<i>N</i> = 16	<i>N</i> = 32
FMO2		
B3LYP/6-31G*	-9.83	-26.27
B3LYP/6-1++G**	-6.78	-23.58
FMO3		
B3LYP/6-31G*	0.51	1.85
B3LYP/6-31++G**	3.12	14.71

<sup>a</sup> FMO2 corresponds to the two-body fragment molecular orbital method with one water molecule per fragment

<sup>b</sup> FMO3 corresponds to the three-body fragment molecular orbital method with one water molecule per fragment

<sup>c</sup> All errors are taken from reference 30.

**Figure 1: Geometry of 21-mer Used to Test EE-PA and EE-3B for Large Clusters.**

