The Electrostatically Embedded Many-Body Method for the Efficient Computation of Properties of Atmospherically Relevant Nanoparticles

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Hannah R. Leverentz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Donald G. Truhlar, Adviser

September 2012

Acknowledgements

I would like to thank Professor Don Truhlar for his excellence as an adviser: his ability to devote quality time, thought, talent, and creativity to each of the many projects and people that he oversees is astounding. I would also like to thank his group members, both current and former, who have been kind and helpful in more ways than it would be reasonable for me to enumerate here. Erin Dahlke-Speetzen, Ewa Papajak, Steven Mielke, Alek Marenich, Boris Averkiev, Rubén Meana Paneda, and Bo Wang deserve special thanks because they have devoted many hours over several years to answering questions, proofreading manuscripts, and helping me solve and circumvent problems. I also am indebted to Professor J. Ilja Siepmann and two of his former group members, Katie Maerzke and Neeraj Rai, for their patience in explaining fundamental concepts of Monte Carlo and molecular dynamics simulations to me and for generating data that served as the foundation of many of the results presented in this thesis.

Outside of research, I am extremely grateful to Doreen Leopold and George Barany for providing me with valuable teaching opportunities and for mentoring me through those experiences. Chuck Tomlinson, Lynn Johnsrud, and Nancy Thao have been invaluable resources to guide me through the requirements for obtaining a graduate degree: I am especially thankful for their patience with me and my many questions! I am grateful to Professor Sanford Lipsky for being a teacher who went the extra mile to help his students understand the material presented in his classes.

Finally, I would like to thank my family members and especially my parents, Tom and Janice Leverentz, for the encouragement, help, and love with which they have surrounded and supported me every day of my life.

Dedication

This dissertation is dedicated to my talented and loving parents, Thomas and Janice Leverentz. I love and admire you both so much! Your lives serve as an example that with God's help I hope to immitate: you manage to achieve excellence in all that you do and yet you never allow your personal goals to take priority over other people; rather, you use your abilities to serve others and bring them joy.

Abstract

Condensed-phased particles suspended in the earth's atmosphere are called aerosols. These particles have important effects on climate and human health, but the mechanisms by which many of these particles form is not well understood. Experimental techniques for studying the early stages of the formation of these particles currently are not available, so computational methods that are capable of accurately handling property calculations on tens of thousands of configurations of the molecular clusters that serve as precursors to aerosols are needed in order to make predictions about the mechanisms of aerosol formation in the atmosphere. Fragment-based computational methods are a promising avenue for the affordable and accurate calculation of properties of large molecular clusters. Fragment-based methods use linear combinations of property calculations done on fragments of the system to obtain an approximation of those properties for the entire system, rather than attempting a highly demanding computation of those properties based on considering all of the molecules in the system at once. Many successful fragment-based methods have been presented in the literature, but this thesis focuses on one fragment-based method that is particularly straightforward and easy to implement: the electrostatically embedded many-body (EE-MB) method. The thesis demonstrates the ability of the EE-MB method to make accurate predictions of properties of atmospherically relevant clusters and explores a few variations of the EE-MB approximation that were developed specifically for Monte Carlo simulations of atmospheric nucleation.

Table of Contents

List of Tables

List of Figures

Preface

The following are citations for previously published works reproduced in this thesis:

Chapter 2.

Adapted with permission from Leverentz, H. R.; Truhlar, D. G. "Assessment of New Meta and Hybrid Meta Density Functionals for Predicting the Geometry and Binding Energy of a Challenging System: The Dimer of H_2S and Benzene" *J. Phys. Chem. A* **2008**, *112*, 6009.

Chapter 3.

Adapted with permission from Leverentz, H. R.; Truhlar, D. G. "The Electrostatically Embedded Many-Body Approximation for Systems of Water, Ammonia, and Sulfuric Acid and the Dependence of its Performance on Embedding Charges" *J. Chem. Theory Comput*. **2009**, *5*, 1573.

Chapter 5.

Reproduced by permission of the PCCP Owner Societies: Leverentz, H. R.; Maerzke, K. A.; Keasler, S. J.; Siepmann, J. I.; Truhlar, D. G. "Electrostatically Embedded Many-Body Method for Dipole Moments, Partial Atomic Charges, and Charge Transfer" *Phys. Chem. Chem. Phys*. **2012**, *14*, 7669.

http://pubs.rsc.org/en/content/articlelanding/2012/cp/c2cp24113g

Chapter 1. Introduction

1.1 General overview

The spontaneous formation of liquid or solid particles from gases in the earth's atmosphere is a phenomenon that has a significant impact on the earth's climate[1–3] but whose mechanism still is not well understood.[4] This process is called "secondary aerosol formation" or "atmospheric nucleation". One of the primary reasons for which the mechanisms of secondary aerosol formation are difficult to ascertain is that current experimental equipment cannot detect neutral aerosols that are smaller than about three nanometers in diameter, but the process of nucleation begins with the formation of neutral molecular clusters that are much smaller than this.[4] In order to investigate the initial stages of the atmospheric nucleation process, theoretical calculations that account for the distribution of electrons around the individual nuclei that are involved in the preliminary molecular clusters are needed. Additionally, these theoretical methods should be as inexpensive as possible in terms of computer time and memory requirements because (1) we need to be able to study the growth of the clusters from their formation from separated gas-phase molecules up to as large a size as possible and (2) thousands of possible nuclear configurations of many molecules, each with a unique distribution of electrons around the nuclei, must be sampled in order to obtain a correct picture of the cluster formation process because, of course, the nuclei are constantly in motion.

Theoretical methods that can make accurate predictions about the electronic structure and other related properties of small molecular systems do exist, but their computational costs (in terms of both time and computer memory requirements)

generally do not scale favorably with system size. The most accurate practical electronic structure methods, such as CASPT2[6] and CCSD(T),[7] scale as N^5 and N^7, respectively, where N is the number of atoms in a given system. These methods can only be used for small systems. Less computationally expensive methods based on density functional theory[8] that have been shown to achieve accuracy close to that of the most accurate methods for small molecular clusters typically scale as N^3 and often as N^4.[9] These density functionals can handle systems containing hundreds or even thousands of atoms on state-of-the-art computer systems.[8] A particle of three nanometers in diameter would contain more than one thousand atoms, which might be achievable using density functional theory for the computation of the electronic structure and related properties of a single configuration of nuclei of a small aerosol. However, thousands of these calculations still would not be reasonable at these levels of theory for such systems. Therefore it is clear that computational methods whose computational costs scale more favorably with system size than N^3 and that can complete cluster property calculations of thousands of configurations of hundreds to thousands of atoms in a reasonable amount of time are needed.

Fortunately, a number of approximations whose costs scale more favorably with system size than N^3 have been developed. These methods typically fall into one of two (not mutually exclusive) categories: linear scaling methods and fragment-based methods.

Linear scaling methods[11–26] are modifications of existing electronic structure methods that make the computational cost of a given method scale linearly with system size (i.e., to make the cost of that method scale as $O(N)$ in the limit of large N.

Examples of this type of method are the divide and conquer method of Yang,[11] the linear scaling density functional and post-Hartree–Fock calculations of Scuseria,[14,22] and the linear scaling coupled cluster algorithms of Head-Gordon and coworkers.[20,26] The advantage of these methods is that in the large-system limit these methods will attain the most favorable scaling possible. However, the drawbacks of linear scaling methods are (1) they require specialized knowledge of the details of a particular electronic structure method and the modification of a particular electronic structure program in order to be implemented and (2) the linear scaling of these methods often is not seen until the system being studied is quite large: Hua *et al.*[29] point out that "as the crossover between these linear scaling approaches and the corresponding conventional approaches occurs at quite large molecules (usually with more than two hundred atoms), they have not been established as a practical tool for electronic structure and property calculations of large systems."

Fragment-based methods [29–59] require dividing a chemical system into smaller parts, which usually are called fragments, monomers, or subgroups, and using the properties calculated from the individual fragments to obtain approximations of the properties of the entire system. The scaling of the cost of these methods with system size ranges from $O(N)$ to $O(N^3)$, but any of these methods ultimately can be made to scale linearly with system size by neglecting interactions between distant fragments. However, although these methods save computational time and memory, many of them have a high cost in "human time" because they are difficult to implement if one does not have programming expertise and/or a good grasp of the essentials of theoretical and computational chemistry. The approximation described in this work, the

electrostatically embedded many-body (EE-MB) approximation,[55] is an exception to this generalization. The EE-MB approximation is elegant in that it is straightforward and requires minimal programming or theoretical chemistry knowledge to implement, and yet it also has been shown to be a cost-effective way to obtain accurate results for molecular clusters of atmospheric relevance.

1.2 Overview of Existing Fragment-Based Methods

Fragment-based methods can be divided into the two categories used by Hua *et al*.[29]: density-matrix based methods and energy-based methods (however, the reader should bear in mind that these are loose classifications that are intended to facilitate the discussion of the many fragment-based methods that are available rather than technical terms with strict definitions; one could find reasons to place any given fragment-based method in either category).

Density-matrix based methods[30–39] emphasize the construction of the density matrix of a large system from the density matrices of the fragments into which the system has been divided. For all of the electronic structure methods that assume that a system's wave function or electron density is the antisymmetrized product of molecular orbitals (i.e., one-electron wave functions) that are in turn linear combinations of predefined one-electron basis functions, the density matrix is the entity that defines a system's wave function or electron density. Therefore, density-matrix based methods essentially build an approximation of the electron density of the full system from the electron densities of the individual fragments. Other properties, such as the system's energy and dipole moment, can be calculated from the resulting approximation of the full system's density matrix. Examples of density-matrix based approaches are the

molecular tailoring approach (MTA),[30] the adjustable density matrix assembler approach (ADMA),[31] and the molecular fractionation with conjugated caps density matrix (MFCC-DM) approach.[34] In several of these approaches, the accuracy of the approximation is often improved by embedding each fragment in an electrostatic field that represents the effect of the other fragments in the system. Density-matrix based approaches typically display greater cost-effectiveness than their conventional counterparts for systems much smaller than those required to see the time advantages in linear scaling methods. However, density-matrix based approaches do require some technical expertise to implement, especially if they are applied to electronic structure methods that assume that the form of a system's wave function is more complicated than an antisymmetrized product of molecular orbitals.

Energy-based methods [29,40–59] focus on approximating a system's energy as a linear combination of the fragment energies (however, these methods can also be used to obtain approximations of other properties such as dipole moments and partial atomic charges, as will be demonstrated in Chapter 5). The EE-MB method falls into this category, along with the majority of other fragment-based approaches. Energy-based methods are advantageous in that they are relatively easy to implement for use with any available electronic structure method. As is the case with the density-matrix based fragment approaches, the accuracy of the predictions made by energy-based approaches can be improved by embedding each fragment in an electrostatic field that represents the effects of the other fragments on the given fragment. The energy-based methods can be further divided into two subcategories: overlapping fragment methods and non-overlapping fragment methods.

Overlapping fragment methods[29,40–48] involve partitioning a chemical system into fragments that, as the name suggests, overlap with one another; that is, an atom belonging to fragment *A* might also belong to fragment *B*. These methods then use the inclusion-exclusion principle to derive the final expression for the linear combination of the fragment energies that approximates the full-system energy. Examples of overlapping fragment methods include systematic molecular fragmentation (SMF),[40–42] the 3B:MB QM:QM method,[43–45] isodesmic fragmentation (IF),[46–47] the generalized energy-based fragmentation approach (GEBF),[29] and the extended oniom (XO) method.[48] These methods have been applied successfully to a wide variety of chemical systems, from noncovalently-interacting molecular clusters to crystalline solids. One drawback of these methods, however, is that the algorithms used to determine the overall fragmentation scheme (and which are often important part of the definition of a particular method) can be quite complicated, so that even though these methods require less technical expertise to interface with currently existing electronic structure programs, some programming expertise still is required to implement the fragmentation algorithms that must be used prior to the computation of fragment energies.

Non-overlapping fragment methods[49–59] involve partitioning a chemical system into fragments that do not overlap; that is, an atom belonging to fragment *A* cannot belong to fragment *B*. For the remainder of this thesis, a non-overlapping fragment will be called a monomer, a pair of monomers will be called a dimer, and a set of three monomers will be called a trimer, and the word "oligomer" will be a generic term used to refer to a monomer, dimer, or trimer. Non-overlapping fragment methods typically are based on a truncation of a many-body expansion of the total energy of a system[27–28]

as an approximation of the system's total energy. The many-body expansion of the total energy (E) of a system composed of N monomers is as follows:

$$E = V_1 + V_2 + V_3 + ... + V_N \tag{1.1}$$

In eq 1.1, V_1 is simply the sum of the energies of all of the monomers in the system,

$$V_1 = \sum_{i=1}^{N} E_i \tag{1.2}$$

where E_i is the electronic plus internuclear repulsion energy of monomer i. V_2 is the sum of the "two-body" (or "pairwise additive") energy contributions of all possible dimer energies in the system. The two-body energy contribution of dimer ij (which is the dimer composed of monomers i and j), v_{ij}, is defined as the difference between the energy of the dimer (E_{ij}) and the sum of the energies of its constituent monomers:

$$v_{ij} = E_{ij} - E_i - E_j \tag{1.3}$$

Therefore, V_2 is defined as follows:

$$V_2 = \sum_{i<j}^{N} v_{ij} \tag{1.4}$$

One can also define the three-body energy contribution v_{ijk} of trimer ijk (the trimer composed of monomers i, j, and k) as the difference between the energy of the trimer (E_{ijk}) and the sum of the one- and two-body energy contributions of the dimers and monomers within it:

$$v_{ijk} = E_{ijk} - v_{ij} - v_{ik} - v_{jk} - E_i - E_j - E_k \tag{1.5}$$

Then V_3 is the sum of the three-body energy contributions of all possible trimers in the system:

$$V_3 = \sum_{i<j<k}^{N} v_{ijk} \tag{1.6}$$

One can go on to define V_4, V_5, etc. in a similar manner. If the many-body expansion is carried out to the N^{th} term, one obtains an exact expression for the system's total energy. However, none of the non-overlapping fragment methods based on this expansion carry this expression past the V_3 term, because to do so would incur a scaling in cost of N^4 or higher (where here N refers to the number of monomers, not to the number of atoms, but where N could be replaced by the number of atoms if each monomer contained the same number of atoms). Nonetheless, in many cases these methods are able to reproduce conventionally calculated full-system energies at a given level of electronic structure theory to within 1.0%, especially if some form of electrostatic embedding is used. One can include electrostatic embedding by placing each monomer, dimer, and trimer in an electrostatic potential that replicates the electrostatic effects of the rest of the system on the given oligomer.

Examples of non-overlapping fragment methods are the kernel energy method (KEM),[49] the fragment molecular orbital method (FMO),[50] the fast electron correlation method (FEC),[51] the explicit polarization (X-Pol) method,[52–53] electrostatic field-adapted molecular fractionation with conjugated caps (EFA-MFCC),[54] the electrostatically embedded many-body (EE-MB)[55] and electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE)[56] methods, the multilevel fragment-based approach (MFBA),[57] and the hybrid many-body interaction (HMBI) method.[58] These methods are all similar in concept, but they differ from one another in

details such as the ways in which the electrostatic embedding is obtained and which terms of the expansion are retained. The KEM method is the simplest of the non-overlapping fragment-based methods: it does not include electrostatic embedding and it estimates a given system's energy as the sum of the V_1 and V_2 terms. The FMO, FEC, and X-Pol methods use an iterative approach to determine the electrostatic embedding that depends on the specific configuration of nuclei being studied, whereas the other methods use a previously determined set of external potentials that potentially can be used with a variety of nuclear configurations. The EFA-MFCC method was primarily designed to calculate interactions between proteins and substrates and does not include the V_2 and V_3 terms. The X-Pol method also was designed primarily for use with proteins and does not include the V_2 and V_3 terms but instead uses empirically derived corrections to improve the accuracy of the V_1 term. The EE-MB method, which is the subject of this work and is described more fully in Chapter 3, uses point charges (which in some cases can be "screened"[59] to improve the accuracy of the calculation) to embed each oligomer and uses the sum of the V_1, V_2, and often the V_3 terms to approximate a system's total energy. The MFBA uses a higher level of electronic structure theory to calculate all monomer energies needed for V_1 and also uses that higher level of theory to compute the contribution to V_2 of any dimer that falls within a certain distance cutoff, but it computes the contribution to V_2 of any dimer outside of the distance cutoff at a lower level of theory in order to lower the computational cost of the calculation. The EE-MB-CE and HMBI methods both perform a calculation of the energy of the entire system at some low level of theory and then correct that energy by including only the

first two or three terms of the many-body expansion of the *difference* between the system's energy at the low level of theory and at a higher level of theory. These methods are essentially the same, but in the EE-MB-CE method the low level of theory is a quantum mechanical electronic structure method, whereas in the HMBI method the low level of theory is a classical force field.

1.3 Organization of the Thesis

The purpose of this thesis is two-fold: (1) to show that the EE-MB approximation is an accurate and cost-effective way to obtain relative energies and other properties of noncovalently interacting clusters and (2) to explore some potentially cost-saving variations of the EE-MB method that were designed specifically for simulations of the beginning stages of atmospheric nucleation.

Chapter 2 of this thesis does not relate directly to the EE-MB method or to any other fragment-based method, but it lays important groundwork by providing validation for the density functional methods used in conjunction with the EE-MB approximation later in the thesis. This chapter shows that these density functional methods are able to accurately reproduce the binding energy and equilibrium geometry of a weakly hydrogen-bonded system, an H_2S-benzene dimer, which is a difficult case for most density functionals. This chapter therefore builds confidence that the density functional methods chosen for use with the EE-MB method will be able to capture accurately the two- and three-body interactions occurring in dimers and trimers. Donald G. Truhlar is acknowledged as a co-author of chapter 2.

Chapter 3 of this thesis provides validation for the use of the EE-MB method on atmospherically relevant ion-containing clusters and also shows one does not need a

complicated or geometry-dependant scheme in order to obtain a set of embedding charges that yields accurate results in the binding energies of these clusters. The clusters chosen for this study were composed of water, sulfuric acid, and ammonium and bisulfate ions, which are thought to be important components in the preliminary stages of secondary aerosol formation.[5] Donald G. Truhlar is acknowledged as a co-author of chapter 3.

Chapter 4 of this thesis explores two variations of the EE-MB method that were designed specifically for use in simulations of the beginning stages of the atmospheric nucleation process. The two variations of the method are called the electrostatically embedded three-body method with consistently screened perturbations (EE-3B-CSP) and the electrostatically embedded pairwise additive specific reaction parameter (EE-PA-SRP) approximation. The goal of these variations was to find an approximation that would significantly reduce the number of trimer calculations required and yet closely maintain the excellent accuracy seen in the EE-3B method. Neither variation fully achieved this goal, but the results are presented in this thesis because these ideas were worth trying and because they might serve as a springboard for ideas of alternative cost-saving variations to the EE-MB or other fragment-based methods. Donald G. Truhlar is acknowledged as a co-author of chapter 4.

Chapter 5 demonstrates that the EE-MB method can be used to obtain not only energies of large system but also properties of the electron density, specifically dipole moments, partial atomic charges, and amounts of charge transfer. The systems chosen for this study were aqueous ammonia droplets, microhydrated chloride clusters, hydrogen fluoride clusters, and microhydrated hydrogen fluoride clusters. The aqueous

ammonia droplets were chosen because of their atmospheric relevance, and the other systems were chosen because they were expected to have significant amounts of charge transfer and therefore would be a good test of how well the EE-MB approximation could estimate that property. Donald G. Truhlar, J. Ilja Siepmann, Katie A. Maerzke, and Samuel J. Keasler are acknowledged as co-authors of chapter 5.

1.4 References

(1) Kurtén, T.; Vehkamäki, H. "Investigating Atmospheric Sulfuric Acid-Water-Ammonia Particle Formation Using Quantum Chemistry" *Advances in Quantum Chemistry* **2008**, *55*, 407.

(2) Rosenfeld, D.; Lohmann, U.; Raga, G. B.; O'Dowd, C. D.; Kulmala, M.; Fuzzi, S.; Reissell, A.; Andreae, M. O. *Science* **2008**, *321*, 1309.

(3) Tao, W.-K.; Chen, J.-P.; Li, Z.; Wang, C.; Zhang, C. "Impact of Aerosols on Convective Clouds and Precipitation" *Rev. Geophys*. **2012**, *50*, RG2001.

(4) Kulmala, M.; Riipinen, I.; Sipilä, M.; Manninen, H. E.; Petäjä, T.; Junninen, H.; Dal Maso, M.; Mordas, G.; Mirme, A.; Vana, M.; Hirsikko, A.; Laakso, L.; Harrison, R. M.; Hanson, I.; Leung, C.; Lehtinen, K. E. J.; Kerminen, V.-M. "Toward Direct Measurement of Atmospheric Nucleation" *Science* **2007**, *318*, 89.

(5) Kulmala, M. "How particles nucleate and grow" *Science* **2003**, *302*, 1000.

(6) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. "Second-order perturbation theory with a complete active space self-consistent field reference function" *J. Chem. Phys*. **1992**, *96*, 1218.

(7) Raghavachari, K.; Anderson, J. B. "A fifth-order perturbation comparison of electron correlation theories" *Chem. Phys. Lett*. **1989**, *157*, 479.

(8) Kohn, W. “Electronic Structure of Matter – Wave Functions and Density Functionals” Nobel Lecture, January 28, 1999. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/kohn-lecture.html. Accessed May 17, 2012.

(9) Zheng, J.; Zhao, Y.; Truhlar, D. G. "The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights" *J. Chem. Theory Comput*. **2009**, *5*, 808.

(10) Zhao, Y.; Truhlar, D. G. "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals" *Theor. Chem. Acc*. **2008**, *120*, 215.

(11) Yang, W. "Direct calculation of electron density in density-functional theory" *Phys. Rev. Lett*. **1991**, *66*, 1438.

(12) Millam, J. M.; Scuseria, G. E. "Linear scaling conjugate gradient density matrix

search as an alternative to diagonalization for first principles electronic structure calculations" *J. Chem. Phys*. **1997**, *106*, 5569.

(13) Li, X. P.; Nunes, R. W.; Vanderbilt, D. "Density-matrix electronic-structure method with linear system-size scaling" *Phys. Rev. B* **1993**, *47*, 10891.

(14) Scuseria, G. E. "Linear Scaling Density Functional Calculations with Gaussian Orbitals" *J. Phys. Chem. A* **1999**, *103*, 4782.

(15) Pulay, P. "Localizability of dynamic electron correlation" *Chem. Phys. Lett*. **1983**, *100*, 151.

(16) Förner, W.; Ladik, J.; Otto, P.; Cízek, J. "Coupled-cluster studies. II. The role of localization in correlation calculations on extended systems" *Chem. Phys*. **1985**, *97*, 251.

(17) Almlöf, J. "Elimination of energy denominators in Møller—Plesset perturbation theory by a Laplace transform approach" *Chem. Phys. Lett*. **1991**, *181*, 319.

(18) Saebø, S.; Pulay, P. "Local Treatment of Electron Correlation" *Annu. Rev. Phys. Chem*. **1993**, *44*, 213.

(19) Hampel, C.; Werner, H.-J. "Local treatment of electron correlation in coupled cluster theory" *J. Chem. Phys*. **1996**, *104*, 6286.

(20) Head-Gordon, M.; Maslen, P. E.; White, C. A. "A tensor formulation of many-electron theory in a nonorthogonal single-particle basis" J. Chem. Phys. **1998**, 108, 616.

(21) Ayala, P. Y.; Scuseria, G. E. "Linear scaling second-order Moller–Plesset theory in the atomic orbital basis for large molecular systems" *J. Chem. Phys*. **1999**, *110*, 3660.

(22) Scuseria, G. E.; Ayala, P. Y. "Linear scaling coupled cluster and perturbation theories in the atomic orbital basis" *J. Chem. Phys*. **1999**, *111*, 8330.

(23) Li, S.; Ma, J.; Jiang, Y. "Linear scaling local correlation approach for solving the coupled cluster equations of large systems" *J. Comput. Chem*. **2002**, *23*, 237.

(24) Nakao, Y.; Hirao, K. "A local second-order Møller–Plesset method with localized orbitals: A parallelized efficient electron correlation method" *J. Chem. Phys*. **2004**, *120*, 6375.

(25) Li, S.; Shen, J.; Li, W.; Jiang, Y. "An efficient implementation of the "cluster-in-molecule" approach for local electron correlation calculations" *J. Chem. Phys*. **2006**, *125*, 074109.

(26) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. "A near linear-scaling smooth local coupled cluster algorithm for electronic structure" *J. Chem. Phys*. **2006**, *125*, 074116.

(27) Elrod, M. J.; Saykally, R. J. "Many-Body Effects in Intermolecular Forces" *Chem. Rev*. **1994**, *94*, 1975.

(28) Xantheas, S. S. "Ab initio studies of cyclic water clusters (H2O)n, n=1-6. II. Analysis of many-body interactions" *J. Chem. Phys*. **1994**, *100*, 7523.

(29) Hua, S.; Hua, W.; Li, S. "An Efficient Implementation of the Generalized Energy-Based Fragmentation Approach for General Large Molecules" *J. Phys. Chem. A* **2010**, *114*, 8126.

(30) Gadre, S. G.; Shirsat, R. N.; Limaye, A. C. "Molecular Tailoring Approach for Simulation of Electrostatic Properties" *J. Phys. Chem*. **1994**, *98*, 9165.

(31) Exner, T. E.; Mezey, P. G. "The Field-Adapted ADMA Approach: Introducing Point Charges" *J. Phys. Chem. A* **2004**, *108*, 4301.

(32) He, X.; Zhang, J. Z. H. "A new method for direct calculation of total energy of protein" *J. Chem. Phys*. **2005**, *122*, 031103.

(33) Chen, X.; Zhang, Y.; Zhang, J. Z. H. "An efficient approach for ab initio energy calculation of biopolymers " *J. Chem. Phys*. **2005**, *122*, 184105.

(34) Chen, X.; Zhang, J. Z. H. "Molecular fractionation with conjugated caps density matrix with pairwise interaction correction for protein energy calculation" *J. Chem. Phys*. **2006**, *125*, 044903.

(35) Li, W.; Li, S. "A localized molecular-orbital assembler approach for Hartree–Fock calculations of large molecules" *J. Chem. Phys*. **2005**, *122*, 194109.

(36) Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. "A new localization scheme for the elongation method" *J. Chem. Phys*. **2004**, *121*, 10385.

(37) Akama, T.; Kobayashi, M.; Nakai, H. "Implementation of divide-and-conquer method including Hartree-Fock exchange interaction" *J. Comput. Chem*. **2007**, *28*, 2003.

(38) Kobayashi, M.; Nakai, H. "Dual-level hierarchical scheme for linear-scaling divide-and-conquer correlation theory" *Int. J. Quantum Chem*. **2009**, *109*, 2227.

(39) Kobayashi, M.; Imamura, Y.; Nakai, H. "Alternative linear-scaling methodology for the second-order Møller-Plesset perturbation calculation based on the divide-and-conquer method" *J. Chem. Phys*. **2007**, *127*, 074103.

(40) Collins, M. S.; Deev, V. A. *J. Chem. Phys*. **2006**, *125*, 104104.

(41) Addicoat, M. A.; Collins, M. A. *J. Chem. Phys*. **2009**, *131*, 104103.

(42) Mullin, J. M.; Roskop, L. B.; Pruitt, S. P.; Collins, M. A.; Gordon, M. S. *J. Phys. Chem. A* **2009**, *113*, 10040.

(43) Hopkins, B. W.; Tschumper, G. S. *J. Comput. Chem*. **2003**, *24*, 1563.

(44) Hopkins, B. W.; Tschumper, G. S. *Mol. Phys*. **2005**, *103*, 309.

(45) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. *J. Chem. Phys*. **2011**, *135*, 044123.

(46) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A*, **2006**, *110*, 8777.

(47) Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A*, **2009**, *113*, 10527.

(48) Guo, W.; Wu, A.; Xu, X. *Chem. Phys. Lett*. **2010**, *498*, 203.

(49) Huang, L.; Massa, L.; Karle, *J. Proc. Natl. Acad. Sci*. **2006**, *103*, 1233.

(50) Fedorov, D. G.; Kitaura, K. "The three-body fragment molecular orbital method for accurate calculations of large systems" *Chem. Phys. Lett.* **2006**, *433*, 182.

(51) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. "Fast electron correlation methods for molecular clusters in the ground and excited states" *Mol. Phys*. **2005**, *103*, 2255.

(52) Xie, W.; Gao, J. *J. Chem. Theory Comput*. **2007**, *3*, 1890.

(53) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys*. **2008**, *128*, 234108-1.

(54) Jian, N.; Ma, J.; Jiang, Y. "Electrostatic field-adapted molecular fractionation with conjugated caps for the energy calculations of charged biomolecules" *J. Chem. Phys*. **2006**, *124*, 114112.

(55) Dahlke, E. E.; Truhlar, D. G. "Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters" *J. Chem. Theory Comput*. **2007**, *3*, 46.

(56) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. "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" *J. Chem. Theory Comput*. **2008**, *4*, 33.

(57) Rezac, J.; Salahub, D. R. *J. Chem. Theory Comput*. **2010**, *6*, 91.

(58) Beran, G. J. O. *J. Chem. Phys*. **2009**, *130*, 164115.

(59) Tempkin, J. O. B.; Leverentz, H. R.; Wang, B.; Truhlar, D. G. "Screened Electrostatically Embedded Many-Body Method" *J. Phys. Chem. Lett*. **2011**, *2*, 2141.

Chapter 2. Assessment of New Meta and Hybrid Meta Density Functionals for Predicting the Geometry and Binding Energy of a Challenging System: The Dimer of H_2S and Benzene[1]

2.1. Introduction

To accurately and inexpensively model systems of noncovalently bound molecules is an ongoing challenge for computational chemists.[1, 2] Coupled cluster theory with single and double excitations and quasiperturbative connected triple excitations (CCSD(T))[3] is an *ab initio* computational method that is generally thought to be capable of accurately predicting the energies of a wide variety of systems including the energies of binding for noncovalently bound molecules, but it is usually too expensive to apply to systems containing more than ten or fifteen atoms. Other wave-function-based correlated methods, such as second- and higher-order Møller-Plesset perturbation theory (MP2[4], MP4[5], etc.) become impractical for very large systems and are too expensive for full simulations even for many smaller systems. On the other hand, methods that are based on density functional theory (DFT) are much less computationally expensive and can be used on systems containing thousands of atoms, but, until recently, available density functionals were incapable of describing dispersion-like interactions without empirical parameterization or a fortuitous cancellation of errors.[6, 7]

However, the situation is not so bleak as it once appeared, and "it would be an oversimplification to dismiss DFT methods for noncovalent interaction in general." [7] Recent work has shown that one can design functionals that capture the medium-range

---

[1] This work was supported in part by the U. S. Department of Energy, Office of Basic Energy Sciences.

dispersion-like interactions and the electrostatic, induction, and charge transfer interactions that determine the binding energies of van der Waals molecules and the strengths of hydrogen bonds.[2, 7-9] One must be able to treat all four of these kinds of attractive interactions accurately for problems such as the competition[2, 10] between π-π stacking (dominated by medium-range dispersion-like interactions) and conventional hydrogen bonding (dominated by the other three). Density functional methods have been used to calculate interaction energies for a variety of hydrogen-bonded systems, ranging from "true" hydrogen bonds (H–O···H–O, H–N···H–N, etc.) to "pseudo" hydrogen bonds, such as O–H···π and S–H···π, where "π" refers to the π system, especially one in an aromatic ring.[7, 11-16] The accuracy of the result of a binding energy calculation on a hydrogen-bonded system of any kind may vary drastically with the density functional chosen to perform the calculation.[11, 17] Ideally one would hope to find a single density functional that works well for the entire range of hydrogen bond types, from the strong short-range "true" hydrogen bonds such as those found in water to weaker, longer-range "pseudo" hydrogen bonds such as those found between second-row hydrides and the π system of an aromatic ring. Such a functional would prove invaluable for the computational modeling of protein folding and protein crystal packing because a variety of hydrogen bond types, including both stronger and weaker hydrogen bonds, play essential roles in these processes;[18-20]such noncovalent interactions may also be important in protein-carbohydrate interactions, carbon nanotubes, materials built with fullerenes and graphene sheets, and conducting polymers.

Previous studies[7, 15, 17] have demonstrated the ability of several newly developed meta and hybrid meta density functionals to accurately predict the interaction energies of shorter-range hydrogen bonds and even of longer-range hydrogen bonds between first-row hydrides and an aromatic ring. Wang and Paulus[16] recently tested eleven density functionals for their abilities to predict the binding energy of an even weaker and longer-range hydrogen bond: the interaction between a second-row hydride and an aromatic ring that occurs in the H_2S-benzene dimer. This particular interaction (i.e. the relatively weak S-H···π interaction) is of interest because it makes a noteworthy contribution to the stability of the conformations of certain folded proteins[20, 21] and because it could potentially be the basis for the creation of an H_2S sensor composed of single-walled carbon nanotubes (SWCNTs).[16] The present study puts the newly developed density functionals that are recommended in previous work[7, 15, 17] as being the best for hydrogen bonds (and for noncovalently bound systems in general) to a test similar to that posed by Wang and Paulus in order determine whether or not these newer functionals capably reproduce the dissociation energies and the geometries of three bound conformations of the H_2S-benzene dimer and thus (because these density functionals have already performed well on shorter and middle-range hydrogen bonds) demonstrate their ability to qualitatively describe an entire array of possible hydrogen bond types.

2.2. Computational Details

Wang and Paulus[16] determined three bound conformations of the H_2S-benzene dimer by fully optimizing several possible geometries of the dimer at the MP2 level of theory with the aug-cc-pVXZ (where X = D, T, and Q)[22, 23] and 6-311++G(d,p)[24] basis

sets. The structures of these three conformations are shown in Figure 2.1 and have been used as the input structures for full optimization and frequency calculations in the present work by using the recently developed density functionals recommended[7, 15, 17] for hydrogen-bonded systems. The structures of the three bound conformations are named in accordance with those assigned by Wang and Paulus: **1a** denotes a structure with C_{2v} symmetry wherein the sulfur atom is located directly above benzene's center of mass and the two hydrogen atoms of H_2S are in a plane determined by the C_6 axis of benzene and a line directly through two opposite carbon atoms of the benzene ring. The **1b** and **1c** structures both have C_s symmetry; in the **1b** structure the angle formed by the sulfur atom, one of the hydrogen atoms of H_2S, and one of the carbon atoms of benzene is nearly 180˚, whereas in the **1c** structure that hydrogen atom points almost directly to benzene's center of mass.

The dissociation energies of the H_2S-benzene dimer given in this paper and in the paper by Wang and Paulus were calculated as:

$$D_e = -E(\mathrm{D}) + E(C_6H_6) + E(H_2S) \qquad (1)$$

where E is the electronic energy (including nuclear repulsion), D is the gas-phase H_2S-benzene dimer, C_6H_6 is the gas-phase benzene monomer, and H_2S is the gas-phase H_2S monomer. In most of the calculations done for this paper, both of the monomers and the dimer were fully optimized at a given level of theory and with a given basis set in order to determine E(D), $E(C_6H_6)$, and $E(H_2S)$. We use the notation L/B to indicate a calculation that was fully optimized at level of theory L and basis set B. In some cases, however, each term in Equation (1) was not fully optimized. For those cases we use the notations L1/B1//L2/B2 and L/B/*d*. The notation L1/B1//L2/B2 means that a full

optimization was done at the level of theory L2 with basis set B2 to obtain a reasonable geometry for the dimer and that this geometry was then used to perform a single-point energy calculation at level of theory L1 with basis set B1. L/B/*d* indicates that the energy of the dimer was optimized with respect to only a single coordinate, *d*. For both of these cases (i.e., those cases in which either the notation L1/B1//L2/B2 or the notation L/B/*d* is used) the monomer geometries were not optimized at any level of theory but were taken to have the structural parameters listed in Table 2.1, which are the best literature values currently available for the true equilibrium geometries of the H_2S and benzene monomers.

The eight density functionals tested in this study are the hybrid meta GGA functionals PWB6K,[25] M05-2X,[9] M06,[15] M06-2X,[15] M06-HF,[26] and MPWB1K,[17] the meta GGA functional M06-L,[27] and the hybrid GGA functional B3LYP.[28] The MPWB1K and PWB6K functionals were selected because they have already shown good performance for systems wherein the hydrogen of a first-row hydride (such as water or NH_3) or even the hydrogen of HCl are noncovalently bound to the π network of an aromatic ring.[7, 29] These two functionals have also performed well in studies testing their ability to find the dissociation energies of a variety of other noncovalently bound systems.[2, 29-31] Previous research[2, 15, 27, 32, 33] has shown that the relatively new M05-2X, M06, M06-2X, M06-L, and M06-HF functionals give good results (compared to other density functionals of their respective kinds; i.e., the meta GGA functional listed does well compared to older meta GGA and GGA functionals but not necessarily as well as older hybrid or hybrid meta functionals) for various types of noncovalent interactions, including hydrogen bonding and even, in some cases, π-π stacking and have therefore

been chosen for further testing in this study. The B3LYP functional was included in this study because, even though it is an older functional, it is still by far the most widely used functional in a variety of applications. The basis set used for all but three of the optimization and single-point energy calculations done for the present study was the MG3S basis, which for hydrogen is the same as 311G(2p), for carbon is 6-311+G(2df) and for sulfur is an improved version[34] of 6-311+G(3d2f). The aug-cc-pVTZ basis set was used in conjunction with the MPWB1K functional to perform geometry optimization calculations on the three conformations of the H_2S-benzene dimer for comparison with the Wang and Paulus study.

The *Gaussian 03* software package[35] in conjunction with the Minnesota Gaussian Functional Module (MN-GFM)[36] was employed to carry out all optimization and single-point energy calculations done in the present study. The counterpoise correction[37] (CP) for basis set superposition error (BSSE) was not added to any of the DFT energies calculated for the present study, although such a correction was added for some of the energies of the H_2S-benzene dimer in the other references cited in this work. Whether or not the counterpoise correction for BSSE was used to compute an energy value listed in this work will be clearly indicated near the energy value wherever it is presented, be it in the body of the text or in a table.

2.3. Results and Discussion

2.3.1 The dissociation energies of the three bound conformations of the H_2S-benzene dimer predicted by DFT calculations compared to those predicted by MP2 calculations

The dissociation energies (D_e) of the noncovalent interaction energies between benzene and H_2S in each of the three bound dimer conformations were determined by Wang and Paulus to be 3.63, 3.70, and 3.75 kcal/mol for structures **1a**, **1b**, and **1c** respectively after full optimizations at the MP2 level of theory with the aug-cc-pVQZ basis set and with counterpoise correction for basis set superposition error (BSSE).[16] For the present study, the dissociation energies of these conformations were calculated by several more recently developed density functionals. These dissociation energies, along with some of those calculated by Wang and Paulus and others calculated by Tauer, Derrick, and Sherrill in a separate study[20] are presented in Table 2.2.

The "CP?" column of Table 2.2 indicates whether or not a counterpoise correction for BSSE was included in the calculation of the dissociation energy; a "Y" in this column denotes that yes, counterpoise correction was used, whereas an "N" denotes that counterpoise correction was not used.

The "MSD (MP2)" column of Table 2.2 denotes the mean signed deviation across all three conformers between the dissociation energies calculated at the given level of theory and basis set and the dissociation energies calculated with MP2/aug-cc-pVQZ. A negative value in this column indicates that on average the given level of theory and basis set predicted lower dissociation energies for the conformers than those that were predicted at MP2/aug-cc-pVQZ. The reader must note that this column should

not necessarily be taken as the best measure of the accuracy of the binding energies calculated at the various levels of DFT; a better measure of this kind of accuracy will be discussed in Section 2.3.3. The reasons the dissociation energies of the conformers calculated by MP2 may not be the best benchmark will be more clear after Sections 2.3.2 and 2.3.3, but we summarize three of those reasons here: 1) Although MP2 does a good job of predicting geometries of various conformers (see Section 2.3.2) the lowest-energy conformer predicted by MP2 is **1c**, whereas experimentally observed conformations of the H_2S-benzene dimer[38] and the similar water-benzene dimer[39] are best represented by **1b**, so it would appear from experiment that the most stable conformer of the H_2S-benzene dimer is really **1b**. 2) Compared to CCSD(T) calculations done on the three bound dimer conformations, MP2 overestimates the dissociation energy by nearly 1 kcal/mol. 3) MP2 is well known to suffer from large basis-set superposition error. For those three reasons, the MSD (MP2) column should be seen as a book-keeping column that displays a measure of how much each level of theory tends to "bind" the three dimers; for example, with a negative MSD (MP2) of –1.98 kcal/mol, one can see that MPWB1K/MG3S predicts that none of the three conformers will be strongly bound, whereas at the positive MSD of 1.50 kcal/mol for MP2/aug-cc-pVDZ shows that this method results because MP2 predicts too much binding.

The newer density functionals tested in this study all predict that the three conformers of the H_2S-benzene dimer are more weakly bound than MP2 calculations suggest, but that they are more strongly bound than a B3LYP calculation predicts. Also, MP2 calculations done with or without counterpoise correction for BSSE and

with any of the Dunning (aug-cc-pVXZ) basis sets predict that the **1c** conformer is the most stable, followed by **1b** and lastly by **1a**, whereas every density functional tested in this study (with the exception of B3LYP, which will be discussed later in this section) predicts that **1b** is the most stable conformer. Frequency analysis of the optimal **1b** conformation by each density functional show that four functionals, M06-2X, M06, M06-HF, and M05-2X yield the **1b** conformation as a local minimum on the potential energy surface (PES) for the H_2S-benzene dimer. On the other hand, PWB6K, M06-L, and MPWB1K/MG3S all classify the **1b** conformer as a hilltop or a saddle point, with 2, 1, and 2 imaginary frequencies respectively. B3LYP does not yield the **1b** conformation as a stationary point of any kind; none of the B3LYP geometry optimizations converged to a **1b** conformation of the dimer.

Of the eight density functionals assessed with the MG3S basis set in this work, four predict that **1c** is the least stable conformer and four predict that **1a** is the least stable conformer. All four functionals that predict **1b** to be the most stable conformer and **1c** to be the least stable (PWB6K, M06-2X, M06, and M06-L) agree that the **1b** conformer is much more stable than either **1a** or **1c**; that is, those four functionals all agree that the dissociation energy of the **1a** conformer is much closer that of the **1c** conformer than it is to that of the **1b** conformer. No such clear trend exists among the four density functionals (B3LYP, M06-HF, M05-2X, and MPWB1K) that predict **1b** to be the most stable conformation of the dimer and **1a** to be the least stable. M05-2X and MPWB1K predict that **1c** is closer in energy to **1a** than to **1b**, but B3LYP and M06-HF predict that **1c** is closer in energy to **1b** than to **1a**. The reason B3LYP places the D_e of **1c** so close to that of **1b** becomes obvious when one visualizes the structures; even

when the **1b** structure is used as the starting point for a geometry optimization by B3LYP, the result of the optimization is a structure that looks just like **1c**, with a hydrogen-sulfur bond pointing almost directly to the center of the benzene ring. M06-HF, however, does locate a minimum on the PES for the H_2S-benzene dimer that has the **1b** structure and still predicts that the **1c** conformer will be closer in energy to that **1b** conformer than to the **1a** conformer.

The dissociation energies calculated at the MPWB1K/aug-cc-pVDZ level (without correction for BSSE) for the **1b** and **1c** conformers in the present study disagree with those calculated (also without correction for BSSE) with the same functional and basis set by Wang and Paulus. For the **1b** conformer, this discrepancy is nearly negligible: Wang and Paulus found the D_e of this conformer to be 2.37 kcal/mol, and the present study found it to be 2.36 kcal/mol. However, Section 2.3.2 will show that the geometries predicted at this level of theory by the two studies for the **1b** conformer may be quite different. For the **1c** conformer, the discrepancy between the dissociation energy values is significant: Wang and Paulus get a value of 2.38 kcal/mol whereas we obtained a value of 1.73 kcal/mol. Clearly this affects the relative stabilities predicted for the three conformations of the dimer: the Wang and Paulus study showed that MPWB1K/aug-cc-pVDZ without correction for BSSE predicts that **1c** is the most stable conformer, followed by **1b** and finally by **1a** as the least stable conformer. The present study finds that what should be the same calculation predicts **1b** to be the most stable conformer, followed by **1a** and finally **1c**. One possible explanation for these differences is that different input structures were used in each study for the geometry optimizations. Both studies used structures that fit the symmetry

and other general structural requirements given in Figure 2.1, but parameters such as bond lengths and tilt angles were not exactly the same. Perhaps the region of the PES that incorporates all H_2S-benzene geometries that could be classified as **1c** conformers is relatively flat and contains several shallow minima or stationary points. We note that our own studies used tight SCF convergence parameters, and all structures are well converged.

*2.3.2 The geometry of the **1b** conformation of the H_2S-benzene dimer predicted by DFT compared to that obtained by experiment*

Prior to the publication of the theoretical work done by Wang and Paulus on the gas-phase H_2S-benzene dimer, an experimental study on this dimer had been performed by Arunan *et al.*[38] The equilibrium geometry of the dimer was found to have a structure similar to that of **1b** in Figure 2.1, with a distance of 3.818 Å between the sulfur atom and benzene's center of mass and with an angle of 28.5° between the C_{2v} axis of H_2S and approximately the C_6 axis of benzene. (The "approximately" modifier is added because the angle measured in the experiment was actually the angle between the C_{2v} axis of H_2S and the "*a*" axis of the dimer; fortunately, the *a* axis of the dimer happened to almost exactly coincide with the C_6 axis of benzene.) This geometry was used as the standard against which the geometry predicted by each density functional for the **1b** conformation was compared. Because B3LYP did not predict any geometry for a **1b** conformation of this dimer as explained in Section 2.3.1, no B3LYP data has been included in Tables 2.3 and 2.4.

Table 2.3 presents a few of the key structural parameters of the **1b** conformation of the H_2S-benzene dimer. The "**1b** Angle" refers to the angle (in degrees) formed

between the C_6 axis of benzene and the C_{2v} axis of H_2S in the **1b** conformer. The distance "*d*" is the distance in Angstroms between the sulfur atom and benzene's center of mass. Not surprisingly, the density functionals that predicted higher dissociation energies for the dimer also predict shorter "*d*" values. The "Δ S–H" and the "Δ HSH" columns give the predicted distortions in the length of the S–H bond (in Angstroms) and the HSH bond angle (in degrees) as the H_2S molecule goes from being the gas-phase monomer to forming a hydrogen bond with benzene's π system in the **1b** conformation. That all the values in the Δ S–H column are positive and that all the values in the Δ HSH column are negative indicate that every density functional tested predicts that the S–H bond is stretched and that the H-S-H bond angle is compressed in the complex. The "SHC Angle" column gives the angle (in degrees) formed between the sulfur atom, the hydrogen atom closest to the benzene molecule, and the nearest carbon to that hydrogen on the benzene molecule.

Table 2.4 enables one to see how accurately each method is able to reproduce the experimentally determined geometry of the **1b** conformation of the H_2S-benzene gas-phase dimer. A negative value in either column indicates that the theoretically predicted value is less than the experimental value, whereas a positive value implies that the theoretically predicted value is greater than the experimental value. The "MUPE" column lists the mean unsigned percentage errors across the two key experimentally verifiable structural parameters for the **1b** conformation of the dimer; that is, the values in this column were calculated by averaging the unsigned percent error in the **1b** angle and the unsigned percent error in *d*.

Table 2.4 shows that, even with a polarized double-zeta (i.e., smaller) basis set, MP2 is the method that most accurately predicts the key structural parameters of the **1b** conformer with an MUPE of 5%. With a polarized triple-zeta (larger) basis set, three of the density functionals are able to do nearly as well as MP2 did with the polarized double-zeta basis: MPWB1K, PWB6K, and, encouragingly because it is a meta-GGA functional, M06-L, each having an MUPE of only 8%. M06 also does well with an MUPE of 9%. Even though a larger basis set was needed in order to achieve these levels of accuracy for the density functionals, the computational cost was still quite low: most of the full geometry optimizations took only a few hours to complete on a single processor with 600 Megabytes of memory. The exception to this was B3LYP, which never converged to a **1b** conformation of the dimer.

A discrepancy exists between the structural parameters predicted by MPWB1K with the polarized double-zeta basis set (aug-cc-pVDZ) calculated by Wang and Paulus and those calculated for the present study with the same density functional and basis set: the result of Wang and Paulus' calculation gave a **1b** Angle of only 16˚ and an overall MUPE of 22%, whereas this study found that angle to be 23.7˚, quite a bit closer to the experimental value and therefore with a much better MUPE of 8%. We already mentioned, in discussing Table 2.2, that the D_e value for the **1b** conformer calculated by Wang and Paulus does not match the D_e value calculated for the **1b** conformer in the present study. This may indicate that two different structures, both of which are quite close in energy and can be classified as "**1b** conformers" with C_s symmetry and an SHC Angle of nearly 180˚, are stationary points on a PES generated by a series of MPWB1K/aug-cc-pVDZ calculations.

*2.3.3 The binding energy of a **1a** conformation of the H_2S-benzene dimer compared to that calculated by CCSD(T)*

Tauer, Derrick, and Sherrill have also previously explored the energy of binding of the H_2S-benzene dimer in a theoretical study.[20] Tauer *et al*. desired to find the lowest-energy conformation of this dimer using the highest level of electronic structure theory available both then and now: CCSD(T). The great computational expense of this method unfortunately precluded (and still precludes) a full optimization of this system with an adequate basis set, so Tauer *et al*. instead did several series of single-point energy calculations using CCSD(T) with aug-cc-pVDZ and aug-cc-pVTZ basis sets on various geometries of the dimer. These calculations included the counterpoise correction for BSSE. In each series of single-point energy calculations, Tauer *et al.* held all general coordinates of the molecule fixed but one; this one coordinate was varied over its entire range of reasonably possible values, and the minimum energy with respect to that coordinate was determined. Ultimately three general coordinates of the dimer were varied one at a time: the "swing" angle θ, the "twist" angle ϕ, and the intermonomer distance *d* between the sulfur atom and benzene's center of mass (see Figure 2.2 of this paper or Figure 1 of Reference 20; the distance *d* in this work is equivalent to the distance *R* in Reference 20, θ is equivalent to A1, and ϕ is equivalent to A2). It was determined that varying ϕ made a less than 0.01 kcal/mol difference in the various energies of binding the dimer, so for simplicity this angle was left at 0˚ for all subsequent calculations. Varying the swing angle θ showed that the lowest binding energy was reached at an angle of about 30˚, but because this minimum was only 0.06 kcal/mol lower than the saddle point at 0˚, θ was also set at 0˚ for the remaining

calculations. Finally, with the monomers frozen at their literature values given in Table 2.1 and in References 40 and 41 and with θ and ϕ set at 0˚, the intermolecular distance *d* was varied from 3.00 to 7.00 Å. The value of *d* that yielded the strongest binding at the CCSD(T)/aug-cc-pVTZ level was 3.80 Å. The energy of binding of this dimer, which looks like the **1a** dimer shown in Figure 2.1, was calculated as

$$BE(a) = E(D,a) - E(C_6H_6) - E(H_2S) \qquad (2)$$

where in this case E(D,a) is the electronic energy of the **1a** dimer with the structural parameters given above, and $E(C_6H_6)$ and $E(H_2S)$ are the electronic energies of the monomers with the structural parameters given in Table 2.1. The energy of binding for this particular geometry of the dimer was extrapolated to the complete basis set (CBS) limit and this best estimate for the energy of binding of this system was found to be –2.81 kcal/mol. BE(a) was calculated according to Equation (2) by each of the density functionals included in this study and each value was compared to the best estimate of BE(a) = –2.81 kcal/mol from Tauer *et al.* These results are displayed in Table 2.5.

In Table 2.5, the "BE(a)" column lists the binding energy for the dimer calculated according to Eq. (2). The "Error" column is the BE(a) value (in kcal/mol) calculated at the given level of theory minus our best estimate of –2.81 kcal/mol, which is the BE(a) value for the same system calculated by Tauer *et al.* as described above and in Reference 20. The "% Error" column is the error column value at the same level of theory divided by the absolute value of the best estimate of BE(a). In both of those columns a negative value implies that the given level of theory "overbinds" the dimer (i.e. the level of theory yields a BE(a) that is too negative relative to the best estimate) and a positive value implies that the level of theory "underbinds" the dimer.

With an error of 0.02 kcal/mol and a relative error of merely 1%, the M05-2X density functional is clearly able to closely reproduce the best estimate of the binding energy of this specific geometry of the **1a** conformation of the H_2S-benzene dimer. Three other functionals also did quite well: M06, M06-HF, and M06-2X, with errors of only 0.19, –0.20, and –0.25 kcal/mol and relative errors of 7%, –7%, and –9%, respectively. Every density functional tested, with the exceptions of B3LYP and MPWB1K, was able to calculate a significantly more accurate binding energy for this system than MP2 with the same basis set. MP2 overbound the system by 0.98 kcal/mol for a large relative error of –35%, whereas the remaining functionals (excluding the two exceptions noted earlier) all achieved absolute relative errors of less than 20% (which would correspond to an absolute error of less than 0.56 kcal/mol). MPWB1K performed only slightly worse than MP2 with an error of 1.05 kcal/mol and a relative error of 37%, but the popular B3LYP significantly underbound the system with an error of 2.80 kcal/mol and therefore a relative error of 100%.

2.4. Conclusions

To obtain accurate dissociation energies and geometries for weakly hydrogen-bonded systems is a difficult task for any electronic structure method and especially for density functional theory. The goal of this work was to determine whether or not several relatively new meta and hybrid meta density functionals would be able to capture most of the noncovalent interaction energy of such a system and also to see whether or not they would be able to accurately reproduce the geometry of such a system. The prototypical weakly hydrogen-bonded system chosen for this study was the H_2S-benzene dimer, which involves a weak hydrogen-bond interaction between a

second-row hydride and the conjugated π system of the benzene ring. The density functionals were used to optimize the geometries of three conformations of this dimer (**1a**, **1b**, and **1c** shown in Figure 2.1) and to find the dissociation energy of each of the three conformations. MP2 optimizations done on the three conformations with various basis sets all predict that the **1c** conformation of this dimer is the most stable; however, every density functional tested in this study (with the exception of B3LYP) predicts that the **1b** conformation of the dimer is the most stable. The results of the dissociation energy calculations by the density functionals appear to be more qualitatively accurate than those obtained from MP2, because an experimental study on this dimer with microwave spectroscopy implies that the equilibrium geometry of the dimer is best classified as the **1b** conformation.[38] MP2, even with only a polarized double-zeta basis set, does do an excellent job of reproducing the geometry of this equilibrium structure, but several of the density functionals when used with a polarized triple-zeta basis do nearly as well as MP2 with the polarized double-zeta, and still at a low computational cost. The density functionals that were able to most closely reproduce the equilibrium geometry (i.e. the **1b** conformation) of the dimer when used with the MG3S basis set were the meta-GGA functional M06-L and the hybrid meta functionals MPWB1K and PWB6K. The hybrid meta functional M06 was also able to reproduce key structural parameters of the **1b** conformer reasonably well with a mean unsigned percent error of less than 10%. The quantitative accuracy of the density functionals' calculations of the dissociation energy of a particular **1a** conformation of the dimer was tested by comparing each functional's result for the dissociation energy of the dimer with the dissociation energy calculated at the CCSD(T) level and extrapolated to the limit of an

infinite basis set. The density functional that yielded the greatest accuracy for the dissociation energy is M05-2X with a relative error of only 1%, but the M06, M06-HF, and M06-2X functionals also did quite well with absolute errors of less than 10%. Every density functional tested except for MPWB1K and B3LYP yielded a more accurate dissociation energy than MP2 with the same basis set.

The fact that all of the new density functionals were able to qualitatively and quantitatively predict dissociation energies more accurately than the older functional B3LYP and even the wave-function-based MP2 method, combined with the facts that most of the new functionals were able to produce good geometries for one of the conformers, that they did so at a relatively low computational cost, and that all of these functionals have already done relatively well for systems containing stronger hydrogen bonds and other types of noncovalent interactions, indicates that any one of these new density functionals would be an excellent candidate for modeling larger systems that contain hydrogen bonds of various strengths. The functional that showed the most versatility in this work and could therefore be awarded this study's "Best in Show" was the hybrid meta M06 functional, performing well both in its prediction of key geometric parameters and in the accuracy of its calculations of dissociation energies.

An area for further exploration of the versatility of these functionals would be to model the interaction of the sulfur lone pairs with various parts of the benzene molecule as Ringer *et. al*. have done with wave function theory in a recent paper;[21] however, this is beyond the scope of the present study.

2.5. Summary

Four new density functionals, three slightly older functionals, and the popular B3LYP functional were tested for their abilities to predict the dissociation energies of three conformers of the H_2S-benzene dimer and to reproduce the key geometric parameters of the equilibrium conformation of this dimer. All of the functionals tested except B3LYP were able to correctly determine which of the three conformations of the dimer is the most stable. The functionals that are best able to reproduce the geometry of the equilibrium conformation of the dimer with a polarized triple-zeta basis set are M06-L, PWB6K, and MPWB1K, each having a mean unsigned relative error across the two experimentally verifiable geometric parameters of only 8%. The M05-2X functional yields the most accurate binding energy of a conformation of the dimer for which a binding energy calculated at the CCSD(T) level of theory is available; M05-2X predicts a binding energy only 0.02 kcal/mol different from that obtained by the CCSD(T) calculation. The M06 functional did well in both categories by yielding a good representation of the geometry of the equilibrium structure and by calculating a binding energy for a given conformation that is only 0.19 kcal/mol different from that calculated by CCSD(T).

2.6. References for Chapter 2

(1) Rappé, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000***, 104*, 6117.

(2) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comp.* **2007***, 3*, 289.

(3) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989***, 157*, 479.

(4) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934***, 46*, 618.

(5) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980***, 72*, 4244.

(6) Cramer, C. J. In *Essentials of Computational Chemistry. Theories and Models;* John Wiley and Sons Ltd: Chichester, 2004; .

(7) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J Phys Chem B* **2005***, 109*, 19046.

(8) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006***, 100*, 5121.

(9) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comp.* **2006***, 213*, 364.

(10) Mignon, P.; Loverix, S.; Geerlings, P. *Chem. Phys. Lett.* **2005***, 401*, 40.

(11) Turki, N.; Milet, A.; Ouameralli, O.; Moszynski, R.; Kockanski, E. *THEOCHEM* **2002***, 577*, 239.

(12) Korth, H. G.; de Heer, M. I.; Mulder, P. *J. Phys. Chem. A* **2002***, 106*, 8779.

(13) Klein, R. A. *J. Comp. Chem.* **2003***, 24*, 1120.

(14) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem.* **2004***, 108*, 5692.

(15) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215.

(16) Wang, Y.; Paulus, B. *Chem. Phys. Lett.* **2007***, 441*, 187.

(17) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004***, 108*, 6908.

(18) Perutz, M. F. *Phil. Trans. Phys. Sci. Eng.* **1993***, 345*, 105.

(19) Myers, J. K.; Pace, C. N. *Biophys. J.* **1996***, 71*, 2033.

(20) Tauer, T. P.; Derrick, M. E.; Sherrill, C. D. *J. Phys. Chem.* **2005***, 109*, 191.

(21) Ringer, A. L.; Senenko, A.; Sherrill, C. D. *Prot. Sci.* **2007***, 16*, 2216.

(22) Woon, D. E.; Dunning, T. H.,Jr. *J. Chem. Phys.* **1993***, 98*, 1358.

(23) Kendall, R. A.; Dunning, T. H.,Jr.; Harrison, R. J. *J. Chem. Phys.* **1995***, 96*, 6796.

(24) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory;* Wiley: New York, 1986; .

(25) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005***, 109*, 5656.

(26) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006***, 110*, 13126.

(27) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006***, 125*, 194101.

(28) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994***, 98*, 11623.

(29) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005***, 109*, 6624.

(30) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005***, 7*, 2701.

(31) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005***, 123*, 161103.

(32) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comp.* **2006***, 2*, 1009.

(33) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008** 112, 6860.

(34) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999***, 110*, 4703.

(35) 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*; Gaussian, Inc.: Wallingford, CT, 2003.

(36) Zhao, Y.; Truhlar, D. G. **2006***, MN-GFM: Minnesota Gaussian Functional Module; version 3.0*, University of Minnesota, Minneapolis, MN, 2007.

(37) Boys, S. F.; Bernardi, D. *Mol. Phys.* **1970***, 19*, 553.

(38) Arunan, E.; Emilsson, T.; Gutowsky, H. S.; Fraser, G. T.; Oliveira, G. D.; Dykstra, C. E. *J. Chem. Phys.* **2002***, 117*, 9766-9776.

(39) Dobrowolski, J.; Jamróz, M. *J. Mol. Struc.* **1993***, 293*, 147.

(40) Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000***, 104*, 2865.

(41) Edwards, T. H.; Moncur, N. K.; Snyder, L. E. *J. Chem. Phys.* **1967***, 46*, 2139.

TABLE 2.1. The best literature values for the structural parameters of the H_2S and benzene monomers

H2S Monomer[a]	
S–H bond length	1.3356 Å
H-S-H bond angle	92.12˚
Benzene Monomer[b]	
C–C bond length	1.3915 Å
C–H bond length	1.0800 Å

[a] From Reference 41

[b] From Reference 40

TABLE 2.2. Dissociation energies (in kcal/mol) of the three bound H_2S-benzene dimer conformations

Method/Basis Set	Ref[d]	CP?	D_e **1a**	D_e **1b**	D_e **1c**	MSD (MP2)
B3LYP/MG3S	P	N	0.47	N/A[c]	0.71	N/A
MPWB1K/MG3S	P	N	1.57	2.00	1.58	-1.98
MPWB1K/aug-cc-pVDZ	P	N	2.14	2.36	1.73	-1.62
MPWB1K/aug-cc-pVDZ	16	N	2.14	2.37	2.38	-1.40
MPWB1K/aug-cc-pVDZ	16	Y	1.76	1.91	N/A	N/A
PWB6K/MG3S	P	N	2.01	2.49	1.93	-1.55
M05-2X/MG3S	P	N	2.81	2.97	2.87	-0.81
M06-L/MG3S	P	N	1.80	2.56	1.69	-1.68
M06-HF/MG3S	P	N	3.12	3.51	3.40	-0.35
M06/MG3S	P	N	1.71	2.70	1.49	-1.73
M06-2X/MG3S	P	N	3.17	3.34	3.12	-0.48
MP2/aug-cc-pVDZ	16	N	5.05	5.16	5.36	1.50
MP2/aug-cc-pVDZ	16	Y	2.97	3.03	3.05	-0.68
MP2/aug-cc-pVDZ/*d*	20	Y	3.06	N/A	N/A	N/A
MP2/aug-cc-pVTZ	16	N	4.38	4.52	4.61	0.81
MP2/aug-cc-pVTZ	16	Y	3.45	3.53	3.60	-0.17
MP2/aug-cc-pVTZ/*d*	20	Y	3.47	N/A	N/A	N/A
MP2/aug-cc-pVQZ	16	N	3.98	4.05	4.15	0.37
MP2/aug-cc-pVQZ	16	Y	3.63	3.70	3.75	0.00
MP2/aug-cc-pVQZ/*d*	20	Y	3.60	N/A	N/A	N/A
CCSD(T)/aug-cc-pVTZ/(inc.)[a]	16	Y	2.73	2.69	N/A	N/A
CCSD(T)/aug-cc-pVTZ/*d*	20	Y	2.64	N/A	N/A	N/A
CCSD(T)/aug-cc-pVQZ// CCSD(T)/aug-cc-pVTZ/*d* (est.)[b]	20	Y	2.74	N/A	N/A	N/A
CCSD(T)/CBS limit// CCSD(T)/aug-cc-pVTZ/*d*[c]	20	Y	2.81	N/A	N/A	N/A

[a] This calculation was not a full CCSD(T) optimization but was obtained by the method of increments; see Reference 16 for details.

[b] This calculation was an estimate of the CCSD(T)/aug-cc-pVQZ based on a full optimization done at MP2/aug-cc-pVQZ and the CCSD(T)/aug-cc-pVTZ/*d* calculation. See Reference 20 for details.

[c] See Section 2.3.1 for an explanation.

[d] The "Ref" column gives the number of the reference from which the dissociation energy values were taken; a "P" in this column indicates that the calculation was done in the present study.

TABLE 2.3. Key structural parameters of the 1b conformation of the H_2S-benzene dimer

Method/Basis Set	Ref[a]	**1b** Angle (˚)	*d* (Å)	Δ S-H (Å)	Δ HSH (˚)	SHC Angle (˚)
PWB6K/MG3S	P	23.2	3.87	0.002	-0.4	180
MPWB1K/MG3S	P	23.6	3.89	0.002	-0.5	180
MPWB1K/aug-cc-pVDZ	P	23.7	3.83	0.002	-0.3	179
MPWB1K/aug-cc-pVDZ	16	16	3.83	0.002	-0.5	170
M05-2X/MG3S	P	20.0	3.78	0.003	-0.3	177
M06-L/MG3S	P	24.4	3.79	0.003	-0.2	179
M06-HF/MG3S	P	19.6	3.65	0.004	-0.4	178
M06/MG3S	P	23.2	3.84	0.004	-0.3	176
M06-2X/MG3S	P	19.0	3.70	0.003	-0.4	178
MP2/aug-cc-pVDZ	16	27	3.63	0.004	-1.1	179
Experimental	38	28.5	3.82	N/A	N/A	N/A

[a] The "Ref" column gives the number of the reference from which the values were taken; a "P" in this column indicates that the values were obtained in the present study.

TABLE 2.4. Errors between calculations and experiment for key structural parameters of the 1b conformation of the H_2S-benzene dimer

Method/Basis Set	Ref[a]	Error in **1b** Angle (˚)	Error in *d* (Å)	MUPE[b]
PWB6K/MG3S	P	-5.3	0.06	8
MPWB1K/MG3S	P	-4.9	0.07	8
MPWB1K/aug-cc-pVDZ	P	-4.8	0.02	8
MPWB1K/aug-cc-pVDZ	16	-12.5	0.01	22
M05-2X/MG3S	P	-8.5	-0.04	15
M06-L/MG3S	P	-4.1	-0.03	8
M06-HF/MG3S	P	-8.9	-0.17	18
M06/MG3S	P	-5.3	0.02	9
M06-2X/MG3S	P	-9.5	-0.11	18
MP2/aug-cc-pVDZ	16	-1.5	-0.19	5
Experimental[c]	38	0.0	0.00	0

[a] The "Ref" column gives the number of the reference from which the values were taken; a "P" in this column indicates that the values were obtained in the present study.

[b] Mean unsigned percent error. See Section 2.3.2 for an explanation of how the values in this column were obtained.

[c] The **1b** angle is 28.5˚; *d* is 3.818 Å.

TABLE 2.5. Binding energies for a specific geometry[a] of the 1a conformation of the H_2S-benzene dimer

Method/Basis Set	Ref[b]	CP?	BE(a) (kcal/mol)	Error (kcal/mol)	% Error
B3LYP/MG3S//b	P	N	-0.01	2.80	100
MPWB1K/MG3S//b	P	N	-1.76	1.05	37
PWB6K/MG3S//b	P	N	-2.28	0.53	19
M05-2X/MG3S//b[c]	P	N	-2.79	0.02	1
M06-L/MG3S//b	P	N	-2.44	0.37	13
M06-HF/MG3S//b	P	N	-3.01	-0.20	-7
M06/MG3S//b	P	N	-2.62	0.19	7
M06-2X/MG3S//b	P	N	-3.06	-0.25	-9
MP2/MG3S//b	P	N	-3.79	-0.98	-35
Best Estimate[d]	20	Y	-2.81	0.00	0

[a] See Section 2.3.3 for a description of this geometry.

[b] The "Ref" column gives the number of the reference from which the values were taken; a "P" in this column indicates that the values were obtained in the present study.

[c] b = CCSD(T)/aug-cc-pVTZ/*d*. See Sections 2.2 and 2.3.3 for an explanation of this notation.

[d] CCSD(T)/CBS limit//CCSD(T)/aug-cc-pVTZ/*d*. See Section 2.3.3 for a summary of the origin of these values and see Reference 20 for details.

Figure 2.1. Input structures used for DFT optimizations of the H_2S-benzene dimer (based on Figure 1 of Reference 16). The top row gives views from above, and the bottom row gives views from the side. Structure **1a** has C_{2v} symmetry. Structure **1b** has C_s symmetry, with an S-H-C bond angle of ~180˚. Structure **1c** has C_s symmetry with one of the H–S bonds of H_2S pointing to the center of the benzene ring.

Figure 2.2. Coordinates varied in a partial optimization of the **1a** conformation of the H_2S-benzene dimer at the CCSD(T) level of theory. See Section 2.3.3 and Reference 20 for definitions of the coordinates.

Figure 2.1

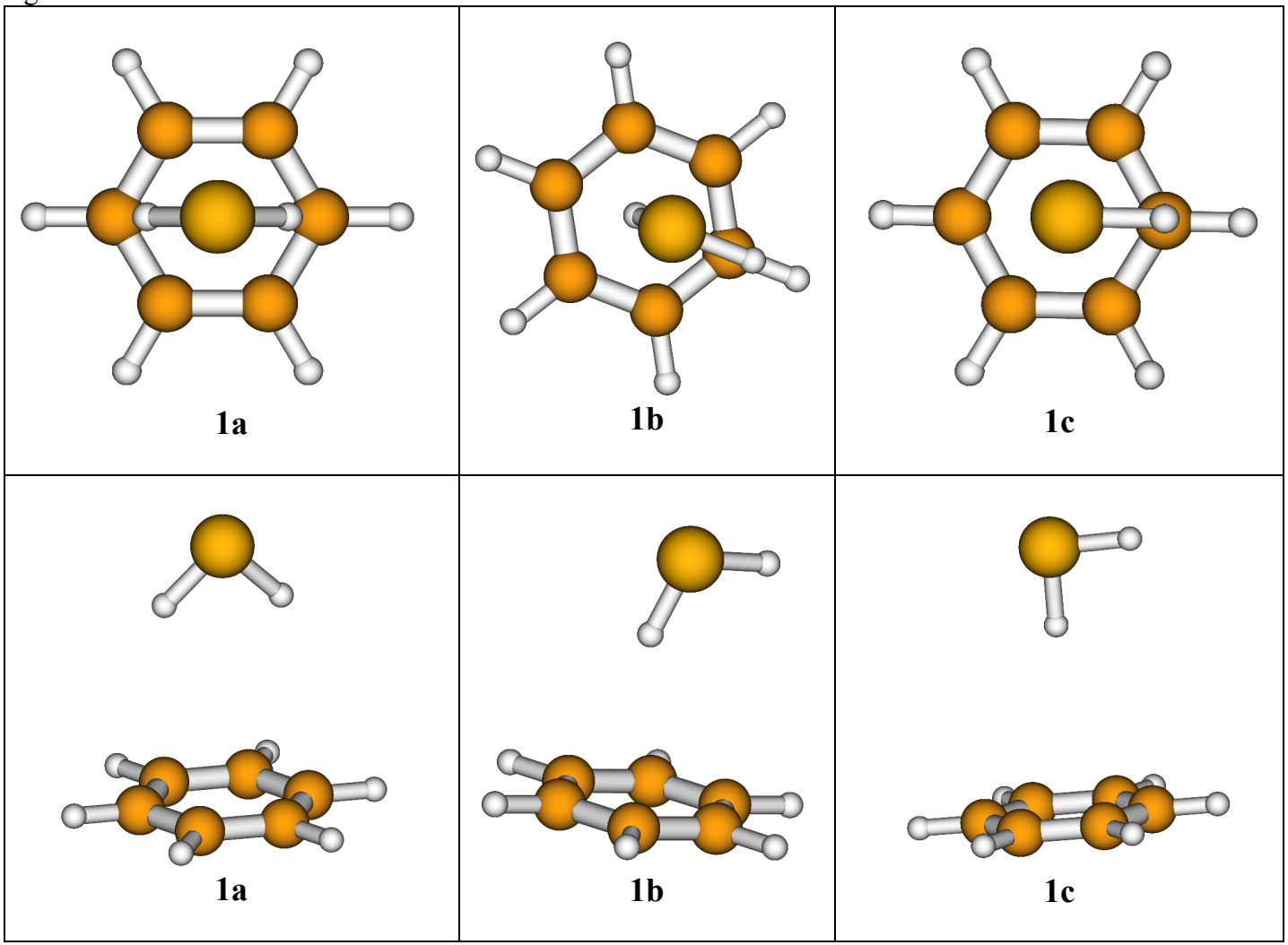
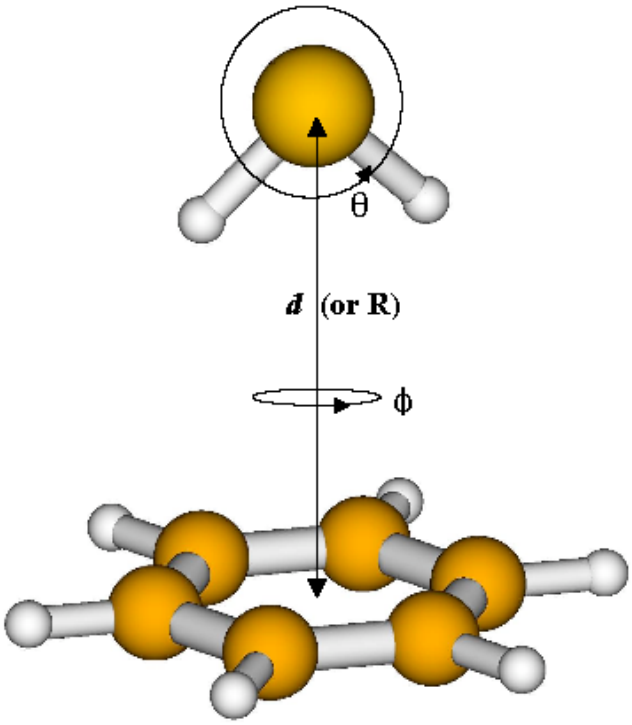


Figure 2.2



Chapter 3. The Electrostatically Embedded Many-Body Approximation for Systems of Water, Ammonia, and Sulfuric Acid and the Dependence of its Performance on Embedding Charges[2]

3.1 Introduction

To compute properties of a chemical system often requires one to find a balance between computational cost and accuracy. A variety of relatively low-cost classical mechanical and semi-empirical quantum mechanical methods allow one to calculate the properties of large (hundreds to thousands of atoms) systems quickly (sometimes within a fraction of a second), but, without problematic parameterization against experimental data, these methods are often incapable of providing more than qualitative accuracy for properties derived from a potential energy surface (PES). At the other extreme, calculations based on the first principles of quantum mechanics [such as coupled cluster[1] (CC) or configuration interaction[2] (CI) theory] have been developed that in principle could be carried to nearly arbitrary levels of quantitative accuracy[3] but that in practice may be used to calculate the energies only of systems containing a few atoms due to the methods' high computational cost. Thus, much effort has been expended in order to find a broadly applicable method that can accurately calculate the energy of a large system at a cost that would be reasonable for use in either molecular dynamics (MD) or Monte Carlo (MC) simulations. Fragment-based approaches[4-13] are one class of methods that attempt to accomplish this goal. These methods involve breaking the large system into subsystems (which will be called fragments) that are small enough to be treated at some desired level of electronic structure theory. Often, an attempt is

---

[2] The authors are grateful to the Minnesota Supercomputing Institute (www.msi.umn.edu) for computer time. This work was supported in part by the National Science Foundation grant no. CHE07-04974.

made to polarize each fragment by representing the "missing" fragments as point charges or continuous charge density distributions, and the large system's total energy is then calculated as some linear combination of the fragments' energies and sometimes of the energies of pairs and trimers of the fragments as well.

The electrostatically-embedded many-body (EE-MB) method,[13-17] which will be described in greater detail in Section 3.2, is a relatively simple fragment-based method that is computationally inexpensive because it does not involve the self-consistent determination of embedding point charges or charge distributions. In the formal EE-MB approximation, each fragment (or monomer), pair of fragments (dimer), and sometimes group of three or more fragments (trimer or higher oligomer) is embedded in a predetermined set of point charges (called embedding charges or background charges) that represents the fragments that are not explicitly included in the electronic structure calculation of a given monomer, dimer, or trimer. When tested on water clusters and on mixed clusters of water and ammonia, the EE-MB approximation showed itself to be a cost-effective way to accurately calculate the total energy of a system of noncovalently interacting molecules at virtually any desired level of electronic structure theory.[13-17] The present work continues to explore the EE-MB approximation by looking at two additional aspects of the EE-MB calculations, as described in the next two paragraphs.

First, the present study applies the EE-MB approximation to more complicated systems than any on which it has yet been tested; the largest clusters considered in this article are formed from six water molecules, one ammonia molecule, and two sulfuric acid molecules. Clusters of this type were selected because these molecules are thought to be the fundamental components of clusters formed during the early stages of

atmospheric nucleation processes.[18] In addition, these clusters test the EE-MB approximation's ability to predict accurate energies (compared to the "full" quantum mechanical calculation by the same electronic structure method) for systems involving both large and small fragments (the large fragment being sulfuric acid with five heavy atoms and the small fragments being water and ammonia with only one heavy atom each) as well as ions or charge transfer complexes because several of the configurations considered in this article correspond to clusters of ammonia, sulfuric acid, bisulfate ion, ammonium ion, and water rather than clusters of only ammonia, sulfuric acid, and water.

Second, the present study compares various ways to obtain the embedding charges and tests how sensitively the accuracy of the EE-MB approximation depends on the resulting sets of embedding charges. Typically the sets of background charges that represent the "missing" monomers are determined by performing some kind of population analysis or charge analysis on the electron density matrices of the isolated and optimized gas-phase monomers. Using these predetermined sets of background charges has several advantages relative to using charges that depend on the configuration under consideration: (1) it lowers the cost of the EE-MB calculation by precluding the need to perform additional self-consistent field calculations to determine the "best" background charges for each configuration, and (2) it maintains the straightforward availability of analytic gradients and Hessians (if they are already available for a given method of electronic structure theory) by removing the embedding charges' dependence on the specific geometry of the system. However, one might argue that using such an inflexible set of embedding charges may not adequately

polarize each fragment and could potentially compromise the accuracy of the EE-MB approximation. Therefore, in the present study we also test some inexpensive ways to obtain embedding charges that *do* depend on the specific geometry of each system being studied, and we compare the EE-MB results from those geometry-dependent (GD) charges with those from the geometry-independent (GI) charges that would be used in the formal EE-MB approximation. One should note that the formal EE-MB approximation would be more easily applied to dynamical simulations[17] that require fast calculations of PES gradients, but that either the formal EE-MB approximation or one that uses geometry-dependent background charges would be convenient for Monte Carlo simulations, where the calculation of PES gradients is not required.

The outline of the rest of this paper is as follows: Section 3.2 briefly reviews the theoretical underpinnings of the EE-MB approximation, Section 3.3 describes the computational methods used to perform the tests in this study and also gives the details of how the various sets of background charges were obtained, Section 3.4 presents the results and discusses their significance, and Section 3.5 summarizes our conclusions.

3.2 Theory

The EE-MB approximation, like several other fragment-based methods, is based on the many-body expansion of a system's total energy. Once a system has been fragmented into N monomers, the many-body expansion expresses the system's total energy as a sum of the energetic contributions of the one-body (i.e., individual monomer) interactions (V_1), the two-body interactions (V_2), the three-body interactions (V_3), and so on up to the N-body term, as shown in eq 1.

$$E = V_1 + V_2 + V_3 + \ldots + V_N \tag{1}$$

If one denotes the energy of one of the monomers as though it had the geometry it has in the cluster but were alone in a vacuum as E_i (where i runs over the arbitrary labels given to the monomers), the energy of dimer as E_{ij}, and the energy of a trimer as E_{ijk}, then the first three terms on the right hand side of eq 1 are defined in eqs 2 through 4; the definitions of the remaining terms can be inferred from these equations.

$$V_1 = \sum_{i=1}^{N} E_i \tag{2}$$

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

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

One could approximate the total energy of the system by truncating eq 1 at some term V_M with M less than N; this is the many-body (MB) approximation of the system's energy given by

$$E \approx V_1 + V_2 + V_3 + \ldots + V_M \tag{5}$$

If one truncates eq 1 after $M = 2$, one has made the two-body (2B) or pairwise additive (PA) approximation. If one truncates eq 1 after $M = 3$, one has made the three-body (3B) approximation.

The same equations as above underlie the EE-MB approximation, but the EE-MB approximation accounts for some of the higher-body interactions in the lower-order terms by calculating the monomer, dimer, trimer, etc. energies (E_i, E_{ij}, and E_{ijk}) as though each monomer, dimer, trimer, etc. were embedded in a field of point charges located at the coordinates of the missing nuclei. The surrounding point charges polarize

or distort the electronic orbitals of each monomer (or group of monomers) so that they take on shapes and amplitudes that more closely resemble those that they might have in the overall system's wave function or electron density. Some of the specific methods by which such sets of point charges could be obtained are described in Section 3.3.

3.3. Methods

3.3.1 Choices of embedding point charges

The paper that introduced the EE-MB approximation[13] pointed out that there are two major categories by which background charges may be determined for use in an EE-MB calculation: The first category, which yields what in this work we call the geometry-dependent or GD charges, calculates the density matrix corresponding to the wave function or the electron density function of the entire system at a computationally inexpensive level of electronic structure theory, such as the semi-empirical method AM1,[19] and performs a charge analysis (such as a Mulliken,[20] Löwdin,[21] or redistributed Löwdin[22] analysis) on that density matrix to calculate partial charges located at the system's atomic centers. The second category, which yields what in this work we call the geometry-independent or GI charges, calculates the optimized density matrix of each type of monomer involved in the system and performs a charge analysis on each of those density matrices. The individual monomers can have their density matrices optimized in either the gas phase or a liquid solution phase. For example, if the system being studied were a cluster of water molecules, one could optimize a water molecule as though it were isolated in the gas phase or one could use an implicit solvation model to mimic an aqueous solution around the water molecule. The atom-centered partial charges calculated from each individually optimized density matrix are then used as the

point charges representing that type of monomer in the EE-MB calculation, regardless of that monomer's position or shape in the overall system. The original (i.e., formal) EE-MB approximation calculates GI charges from monomers optimized in the gas phase, but we test the following three general types of point charges in the present work: GD charges, GI charges from monomers optimized in the gas phase, and GI charges from monomers optimized in a solution phase. (One could imagine another type of GD charge where charges are calculated for monomers but at the geometry they have in the particular configuration of the whole system that is under consideration, but we will not consider this method.)

3.3.2 Computational Methods

All EE-MB calculations carried out in the present work were conducted using the M06-2X[23] density functional, which was chosen because it performs better than other density functionals for noncovalent interactions between molecules composed of main-group elements.[23] Three different basis sets were used to test the overall accuracy of the EE-PA and the EE-3B approximation: MG3S,[24] cc-pV(T+d)Z+,[25] and aug-cc-pV(T+d)Z.[26]

In order to calculate the geometry-dependent (GD) sets of background charges, the AM1 wave function of each configuration studied was calculated, and the following methods of charge analysis were used on those wave functions: Mulliken population analysis,[20] Charge Model 1 (CM1A, where the A indicates that the version of CM1 used was specifically parameterized to be used with AM1 wave functions),[27] Charge Model 2 (CM2),[28] and Charge Model 3 (CM3).[29, 30] One should note that Mulliken charges are Class II charges[27] and at best give electrostatic properties corresponding to an

approximate level of theory with a finite set of basis functions, whereas CM*x* (*x* = 1A, 1P, 2, 3, 4, or 4M) charges are Class IV charges because they include empirical parameters that map Class II charges (such as Mulliken or Löwdin charges) to charges that more realistically reproduce experimental dipole moments. Following the recommendation of Udier-Blagović et al.,[31] a final set of GD charges has also been tested: these charges are simply CM1A charges scaled by 1.14, and they are labeled "CM1A*1.14" or "scaled CM1A" charges. The scaling is designed to make the charges (although computed in the gas-phase) more appropriate for liquid simulations.

Geometry-independent (GI) charges were obtained from the density matrices of both gas-phase and liquid-phase monomers. The optimized density matrices of gas-phase monomers were used to calculate point charges according to eight different methods of charge analysis: ChElPG,[32] Merz-Singh-Kollman (MK),[33, 34] the MK method with the additional constraint to reproduce dipole moments as well as electrostatic potentials (ESP-Dipole; see the *Gaussian 03* online manual[35] for details), Natural Bond Orbital (NBO),[36] CM1A, CM2, CM3, and CM4M.[37, 38] NBO can be considered to be a Class II charge model, but rather than calculating charges from a density matrix expressed in terms of the original basis set functions, NBO charges are calculated from a density matrix expressed in terms of a set of functions that adopt the "natural" shapes that a chemist would expect to describe various types of chemical bonds. The ChElPG, ESP-Dipole, and MK methods are quite similar to one another and yield what are classified as Class III charges; these charges are those that best reproduce the electrostatic potential due to a system's electron density distribution function at various points in space around the system (which is a gas-phase monomer

for GI charge analysis). The ChElPG, ESP-Dipole, MK, and NBO charges were determined from electron density functions computed by M06-2X/cc-pV(T+d)Z+//M06-2X/cc-pV(T+d)Z+ (we adopt the common notation *W/X//Y/Z*, where *Y* is the level of electronic structure theory or density functional and *Z* is the basis set with which the geometry of the system was optimized, and where *W* is the level of electronic structure theory or density functional and *X* is the basis set with which the electron density and/or energy to be used in subsequent calculations was optimized). Because CM4M contains parameters that depend on the density functional and basis set chosen and is currently parameterized for a variety of double-zeta but not triple-zeta quality basis sets, the CM4M charges were calculated from the M06-2X/MIDI!//M06-2X/cc-pV(T+d)Z+ density matrix (the MIDI![39] basis set is of double-zeta quality and was designed specifically for the efficient calculation of accurate geometries and partial charges). Charge Models 1, 2, and 3 were originally parameterized for semi-empirical methods, and to obtain the geometry-independent CM1A, CM2, and CM3 charges we used the AM1//AM1 wave functions of the isolated gas-phase monomers. The scaled CM1A charges (CM1A*1.14) of the gas-phase monomers were also used as GI background charges for EE-MB calculations.

Three sets of background charges were based on liquid-phase monomers: these charges are denoted SM5.42/CM2, SM8/CM4M, and SMD/CM4M. To describe the density matrices used to obtain the charges, we adopt the following notation: *slvnt-SMx/W/X//slvnt-SMy/Y/Z*, where *W*, *X*, *Y*, and *Z* are as defined above and where *SMy* is the solvation model applied to perform a liquid-phase geometry optimization of the monomer, where *SMx* is the solvation model used to obtain a liquid-phase optimized

wave (or density) function for subsequent charge analysis (for this study, *x* and *y* can be 5.42, 8, or D), and where *slvnt* indicates the solvent in which the monomer was theoretically immersed (for this study, *slvnt* = aq to signify that the calculation was performed in an aqueous solution). For SM5.42/CM2 charges, CM2 charges were calculated from the aq-SM5.42/AM1//AM1 monomer density matrices; that is, the monomer geometries were optimized by AM1 in the gas phase, and the wave functions were then optimized by AM1 in the aqueous phase using Solvation Model 5.42. "SM8" and "SMD" indicate Solvation Model 8[40] and Solvation Model D,[41] respectively. The SM8/CM4M charges are CM4M charges calculated from the aq-SM8/M06-L/6-31G(d)//M06-L/6-31G(d) monomer density matrices based on the M06-L density functional[42] with the 6-31G(d) basis,[43, 44] and the SMD/CM4M charges are CM4M charges calculated from the aq-SMD/M06-L/6-31G(d)//M06-L/6-31G(d) monomer density matrices.

All geometries were optimized using the *Minnesota Gaussian Functional Module, version 3.0* (*MN-GFM-v3.0*),[45] a locally modified version of the GAUSSIAN 03[46] electronic structure package, revision D.01. *MN-GFM-v3.0* was also used to perform the charge analyses for the ChElPG, ESP-Dipole, MK, and NBO charges and to carry out single-point energy calculations on the clusters and molecules involved in this study. The *Minnesota Gaussian Solvation Module, version 2008* (*MN-GSM-v.2008*),[47] a module for performing solvation calculations in GAUSSIAN 03, revision D.01, was used to compute the CM4M and SM8/CM4M charges. The SMD/M06-L/6-31G(d)//M06-L/6-31G(d) wave function was computed using the *GESOL*[48] program (an external module for GAUSSIAN 03), but the SMD/CM4M

charges based on this wave function were computed using *MN-GSM-v.2008*. Calculations done on AM1 wave functions to find the geometry-dependent and geometry-independent CM1A, CM2, and CM3 charges as well as the geometry-independent SM5.42/CM2 charges were performed using *AMSOL-version 7.1*.[49] All EE-MB calculations were executed using MBPAC 2007-2,[50] a program that calls GAUSSIAN 03 or *MN-GFM* to perform electrostatically-embedded many-body calculations.

3.4. Results and Discussion

The first goal of this study was simply to test how well the EE-PA and EE-3B approximations are able to reproduce the energies of systems containing water, ammonia, and sulfuric acid and/or their conjugate acids or bases calculated in the conventional manner for a given model chemistry (for this discussion, "model chemistry" or "method" implies a specific combination of electronic structure theory level or density functional with a specific basis set). To do this, the binding energies of eight clusters, each composed of between three and nine molecules, were calculated by M06-2X with three different basis sets: MG3S, cc-pV(T+d)Z+, and aug-cc-pV(T+d)Z.

The eight clusters considered in the first part of this work and their names are shown in Figure 3.1. The smallest cluster, called 1_H2O, contains one bisulfate ion, one ammonium ion, and only one water molecule. Because the 1_H2O cluster contains only three molecules and because each molecule is defined as a monomer for the EE-MB calculations done in this study, only the EE-PA approximation is applied to this cluster (the EE-3B approximation necessarily yields the same result as the conventional calculation for the same method). The next smallest cluster contains one bisulfate ion,

one ammonium ion, and two water molecules and is called the 2_H2O cluster. Clusters 1_H2O and 2_H2O are based on structures shown in Figure 1 of Reference 51; however, the precise coordinates used for the single-point energy calculations carried out in this study (which have been included in Appendix to this chapter) were the result of M06-2X/MG3S geometry optimizations done on these two clusters. Six clusters, three of which comprise six water molecules, one ammonia molecule, and one sulfuric acid molecule and three of which comprise six water water molecules, one ammonium ion, one bisulfate ion, and one sulfuric acid molecule, were also studied. These configurations are called 1A, 1B, 1C, 2A, 2B, and 2C. They were generated during an MD simulation;[52] a "1" in the name indicates that one sulfuric acid molecule was used in the starting configuration of the simulation and a "2" indicates that two sulfuric acid molecules (one of which is in bisulfate form) were used in the starting configuration of the simulation.

The binding energy (E_{bind}) of each of the clusters described in the preceding paragraph was first calculated in the conventional way with respect to the neutral gas-phase monomers with geometries optimized by the M06-2X/MG3S method. That is,

$$E_{\text{bind}} = E_{\text{cluster}} - \Sigma\, E_{\text{molecule}} \qquad (6)$$

where E_{cluster} is the M06-2X/*basis*//M06-2X/MG3S absolute electronic energy of the cluster and E_{molecule} is the M06-2X/*basis*//M06-2X/MG3S absolute energy of the neutral version of each molecule from which the cluster is formed [*basis* = MG3S, cc-pV(T+d)Z+, or aug-cc-pV(T+d)Z]. Table 3.1 lists the binding energy of each cluster when calculated in the conventional (or "full") manner by each method and the difference (or error) between the EE-PA calculation with different sets of geometry-

independent background charges calculated according to the formal EE-MB prescription; that is, the background charges were calculated from the optimized gas-phase monomers. These full binding energies were then used to test the truncated EE-MB expansions. Note that although the binding energies are defined with respect to the neutral versions of the monomers, in applying the EE versions of eqs 1–4 the clusters were fragmented into both neutral and ionic monomers and the background charges that were used to represent the missing monomers were calculated from the corresponding optimized gas-phase monomers, which are both neutral and ionic.

Table 3.1 shows that the EE-MB approximation, and in particular the EE-3B approximation, continues to yield accurate results when compared to the conventional calculations when it is used to calculate the binding energies of these more complicated clusters than others on which it has so far been tested. The maximum absolute error for the EE-PA calculations is 5.28 kcal/mol; this error occurs for structure 2C, which has a large binding energy, and so it corresponds to an error of only 8.4%. Furthermore, the average relative absolute error over all eight configurations, three basis sets, and five background charge sets for the EE-PA approximation is only 3.0%. The maximum absolute error for the EE-3B calculations (seven configurations) is a mere 0.41 kcal/mol, again for structure 2C, and it corresponds to a relative absolute error of only 0.7%. The average relative absolute error over all seven configurations, three basis sets, and five background charge sets for the EE-3B approximation is 0.3%. These results imply that the EE-3B approximation is capable of handling complicated systems involving ions and/or charge transfer complexes and a wide range of monomer sizes and monomer complexity.

One particularly striking aspect of Table 3.1 is the comparative size of the EE-3B error in reproducing the full calculations and the deviations of the various full calculations from one another. Although all three basis sets are multiply polarized valence triple zeta sets with diffuse functions, the results for a given cluster with a pair of basis sets differ from one another on average by 0.9 kcal/mol whereas a typical error due to the EE-3B approximation is ~0.1 kcal/mol. Thus the error incurred by truncating the EE-MB expansion with geometry-independent background charges is much less than the uncertainty due to choice of basis set.

The second goal of this study was to compare two major categories of methods by which background charges for EE-MB calculations can be obtained: geometry-dependent and geometry-independent charges. From the geometry-independent charges, two subcategories of partial charge calculation methods are also compared: those that use the gas-phase monomer density matrices and those that use liquid-phase monomer density matrices. (See Section 3.3.1 for a more detailed explanation of these categories.) This portion of the study focuses on the binding energies of the three largest (and most complex) clusters (2A, 2B, and 2C) calculated by the M06-2X/cc-pV(T+d)Z+ method both conventionally and with the EE-MB approximation (MB = PA or 3B for this study) using background charge sets from each of the above categories and subcategories. [The cc-pV(T+d)Z+ basis set was selected because, of the three basis sets shown in Table 3.1, it is generally the most efficient at reducing basis set superposition error (BSSE); that is, on average the cc-pV(T+d)Z+ basis set yields about the same amount of BSSE as aug-cc-pV(T+d)Z but at lower cost.[25] The MG3S basis set tends to yield larger amounts of BSSE than the other two

basis sets. Using a basis set that in general yields low BSSE diminishes the need to attempt to correct for BSSE by methods such as counterpoise correction,[53] which would significantly increase the overall cost of a calculation of the binding energy of a nine-molecule system.[54]]

Table 3.2 lists the errors from the EE-PA and EE-3B calculations relative to the conventionally calculated binding energy for each of the three clusters. The charge sets labeled with the prefix "GI_" are geometry-independent charge sets; that is, these charges were calculated from the density matrices of individual monomers and therefore do not depend on a given cluster's geometry (in still other words, these charge sets would remain the same for any cluster containing the same types of molecules and/or ions). Of the geometry-independent charge sets, those that were computed from aqueous-phase monomer density matrices contain the letters "SM" (for "solvation model") in their labels immediately following the "GI_" prefix; those that were computed from density matrices of gas-phase monomers do not. The "GD_" prefix indicates that the given charge set is geometry dependent; that is, the charge set is calculated from a semi-empirical wave function for the entire cluster and therefore depends on the geometry of the cluster. Table 3.3 summarizes the results shown in Table 3.2 by listing the mean unsigned errors (MUE) and root mean squared errors (RMSE) of the EE-PA and EE-3B approximations over all three clusters. Section 3.3.2 contains the specific description of the meaning of the name of each charge set and the method by which each charge set was calculated.

First, Tables 3.2 and 3.3 show that electrostatic embedding significantly enhances the accuracy of the PA and 3B approximations. Without electrostatic

embedding, the pairwise additive MUE is 9.82 kcal/mol, whereas the maximum electrostatically-embedded pairwise additive MUE is 4.40 kcal/mol and the average EE-PA MUE is 2.96 kcal/mol. Similarly, the three-body MUE without electrostatic embedding is 0.70 kcal/mol, whereas the maximum EE-3B MUE is 0.30 kcal/mol and the average EE-3B MUE is 0.20 kcal/mol.

A second point illustrated by Tables 3.2 and 3.3 is that GD charge sets yield EE-MB results that are only slightly better than those from GI sets. For the EE-PA approximation, the average GI MUE is 3.00 kcal/mol and the average GD MUE is 2.84 kcal/mol. For the EE-3B approximation, the average GI MUE is 0.21 kcal/mol and the average GD MUE is 0.18 kcal/mol. The small (nearly insignificant for the EE-3B approximation) improvement in accuracy afforded by the GD charge sets is not worth the loss of convenient analytic gradients when performing MD simulations, nor does it even seem to be worth the tiny relative increase in cost that would be incurred during MC simulations. Therefore, the original EE-MB approximation where the electrostatic embedding is based on GI charges continues to be the recommended approach for EE-MB calculations.

A third conclusion that may be drawn from Tables 3.2 and 3.3 is that using gas-phase monomer wave or density functions as a starting point for the determination of GI charges is just about as good as using charges derived from liquid-phase monomers. The average MUE of the charge sets obtained from gas-phase monomers is 3.01 kcal/mol for the EE-PA approximation and 0.21 kcal/mol for the EE-3B approximation. The average MUEs of the charge sets obtained from aqueous-phase monomers are 2.99 kcal/mol and 0.20 kcal/mol respectively. This may come as a

surprise because one might expect that the electron density around a monomer in a cluster would more closely resemble the electron density distribution of a solvated monomer than it would the electron density distribution of a gas-phase monomer. This is because a monomer in a cluster or a monomer in solution is polarized by the surrounding monomers and might experience more charge separation than would a monomer in the gas phase; i.e., charges derived from either a monomer in a cluster or a monomer in solution might take on more extreme magnitudes than charges derived from a monomer in the gas phase. However, this did not turn out to be the case. The charges derived from liquid-phase monomers were on the whole quite similar to those derived from gas-phase monomers, as shown in Tables 3.4 to 3.6. This explains why these charge sets produce similar results when used as the background charges in EE-MB calculations. Once again, the original EE-MB approximation (taking GI charges from monomers in the gas phase) remains the recommended approach because gas-phase monomer calculations are less costly (even if only by a little) than liquid-phase calculations and because potential ambiguity regarding which solvent to choose for monomers involved in mixed clusters is avoided when the gas-phase monomers are used to generate embedding charges.

To summarize the discussion of the results presented so far, one could simply state that the accuracy of the EE-MB approximation does not appear to be heavily dependent on the set of background charges chosen. Compared to the average binding energy of the three clusters, -65 kcal/mol, a 5% error would be 3.3 kcal/mol and a 1% error would be 0.65 kcal/mol. Thus, most of the EE-PA calculations yield MUEs of

less than 5% and all of the EE-3B calculations yield MUEs of less than 1%, regardless of the charge model used.

The reason for which the two subcategories of GI charges (gas-phase vs. liquid-phase monomers) do not produce significantly different EE-MB results was addressed in an earlier paragraph, but one is still left to wonder why GI charges manage to do about as well as GD charges. In an attempt to understand this, one can investigate (1) how the dipoles of individual water molecules vary within a cluster and (2) how the dipole of an individual water molecule varies when the water molecule is embedded in different sets of point charges. Because water is a planar molecule and generally possesses close to C_{2v} symmetry, its dipole is a good indicator of the extent to which the water molecule is polarized. (Note that the word "dipole" is used to mean "the magnitude of the dipole moment".) If the dipoles of the water molecules in the entire cluster do not vary much (Test 1), then one can see how the "inflexible" charges from a gas-phase monomer could adequately mimic the effects of other water molecules in the cluster. Additionally, if the dipole of a single water molecule embedded in point charges does not vary much with different background charge sets (Test 2), then one can infer that the choice of background charges will not strongly impact an EE-MB calculation, because the purpose of the background charges is to polarize the monomers, dimers, trimers, etc. If the embedding charges do not have a strong effect on the polarization of a monomer, then it is unlikely that they would have a strong effect on the result of an EE-MB calculation.

As stated above, Test 1 investigates the variation in the dipoles of the water molecules within a given cluster. The dipole moment of each water molecule in

clusters 2A, 2B, and 2C was calculated using the Mulliken, CM1A, scaled CM1A, CM2, and CM3 point charges that were obtained from the AM1 wave functions of those clusters; these dipoles are listed in Table 3.7. The point charges were used to calculate the dipoles because one cannot determine the expectation value of the dipole moment of an individual water molecule from the wave function of the entire cluster. The dipole moment of each water molecule was calculated with respect to that particular molecule's center of nuclear charge. To give an idea of the locations of the water molecules within a cluster, Figure 3.2 shows configuration 2A with each water molecule labeled by the arbitrary fragment number that it was assigned for the EE-MB calculations; these numbers correspond to the monomer labels listed in Table 3.7. The water molecules of configurations 2B and 2C were labeled in essentially the same way, although of course their locations within the cluster are slightly different in each case. Table 3.7 shows that for every type of point charge representation other than Mulliken charges, the relative standard deviation is only 5%. In the case of dipoles derived from Mulliken charges the relative standard deviation is 8%. These results indicate that the variation in the extent of polarization of water molecules in a semi-empirical calculation of an entire cluster's wave function is not large.

Test 2 investigates the dependence of the extent of polarization of a single embedded water molecule on the choice of background charges. Monomer number 5 of configuration 2A was selected as the embedded water molecule for this test. The remaining water molecules were represented by the same sets of point charges described in Section 3.3.2, and the M06-2X/MIDI! density matrix of the water molecule in each charge set was calculated. From this density matrix, one can calculate the

dipole of the water molecule in two different ways: (a) quantum mechanically as the expectation value of the dipole moment operator or (b) classically from a set of point charges that has been determined through charge analysis of the water molecule's density matrix. Dipoles computed by methods (a) and (b) are called density dipoles and point-charge dipoles, respectively.

Method (b) begs the question of which specific charge model should be used to assign point charges to the embedded water molecule. To decide which charge model to use for this specific task, the dipoles of several non-embedded gas-phase molecules and ions according to nine charge models listed in Table 3.8. The ChElPG, ESP-Dipole, MK, and NBO charge analyses were carried out on the molecules' M06-2X/cc-pV(T+d)Z+ density matrices, the CM4M charges were determined from the M06-2X/MIDI! density matrices, and the CM1A, scaled CM1A, CM2, and CM3 charges were extracted from the AM1 wave functions. The geometries of these compounds had previously been optimized at the same level of theory at which the charge analyses were performed, except for the CM4M charges which had been optimized at M06-2X/cc-pV(T+d)Z+. The classical dipoles of these molecules calculated from these charge representations are shown in Table 3.8 and are compared to our best estimates of these dipoles. In the case of the neutral compounds, our best estimates are based on experimental values, but in the case of ions these have been determined with respect to each given ion's center of nuclear charge by finite-field calculations done with the CCSD(T)/aug-cc-pV(T+d)Z//M06-2X/cc-pV(T+d)Z+ method. The MUEs and RMSEs over all five compounds with respect to the best-estimate dipoles are also given in Table 3.8. Based on the MUEs, the CM4M charges

appear to reproduce the best estimate dipoles better than the other methods, so the CM4M charges of the embedded water monomer were used to calculate the classical dipoles of method (b).

The dipoles arising from methods (a) and (b) are given in Table 3.9, along with the individual point charges used to calculate the dipoles for method (b). One should first notice, by comparing the point charge dipole of the un-embedded water molecule (in the row labeled "None") to the point-charge dipoles of the embedded water molecule, that the electrostatic embedding does increase the dipole of the water molecule by about 0.3 D. Once embedded, however, the specific background charge set chosen does not affect the point-charge dipole by more than 0.05 D (or the density dipole by more than 0.09 D). The relative standard deviations over all sets of embedding charges for the water molecule's CM4M charges, density dipoles [method (a)], and point-charge dipoles [method (b)] are all under 1.3%. Thus, the extent to which a water molecule is polarized is affected by whether or not the water molecule is embedded in point charges, but the extent of polarization does not depend heavily on the specific set of embedding charges used.

3.5. Conclusions

The primary goals of this paper were (i) to test the overall accuracy of the EE-MB approximation for clusters involving both large and small monomers as well as a mix of ions and neutral molecules and (ii) to observe the dependence of the EE-MB approximation's accuracy on the background charges used as the electrostatic embedding.

Regarding the first goal, this study shows that the EE-MB approximation is capable of providing accurate binding energies for relatively complicated systems. For five sets of embedding charges used with three different basis sets on a test set of mixed clusters ranging in size from two to nine molecules, the errors in the binding energies from the EE-PA approximation relative to the binding energies of the full calculations at the same level of theory do not exceed 10%, and in many cases are closer to 5%. The EE-3B approximation does even better, with relative errors that do not exceed 0.7%.

Regarding the second goal, this study shows (in accord with results from previous studies on less complicated systems[13, 16]) that electrostatic embedding does significantly improve the performance of the PA and 3B approximations, but that the specific set of point charges used for the electrostatic embedding does not strongly influence the accuracy of the EE-PA or EE-3B approximations. Two general categories of background charge sets were tested: geometry-dependent (GD) and geometry-independent (GI) charge sets. On the whole, GD and GI charges yield EE-MB results of almost equal accuracy; over three configurations of a mixed nine-molecule system, the EE-3B MUE over the GI charge sets is 0.21 kcal/mol and over GD charge sets is 0.18 kcal/mol. Although the GD charge sets perform slightly better, they are also slightly more complicated to implement for energies and much more complicated to implement for gradients. Of the GI charge sets, those that were obtained from gas-phase monomer density matrices do not perform significantly differently than those that were obtained from liquid-phase monomer density matrices: over the EE-3B binding energies of three configurations of the nine-molecule system, the gas-phase monomer-

derived charge sets yielded an MUE of 0.21 kcal/mol, and the liquid-phase monomer-derived charge sets yielded an MUE of 0.20 kcal/mol.

A third objective that arose during the course of this study was to investigate why the GD charge sets do not perform as much better than the GI charge sets as one might have expected. The conclusions reached from that portion of the study are these: (1) The polarization of the water molecules within a given nine-molecule cluster does not vary much from molecule to molecule, implying that the "rigid" point charges from a gas-phase monomer are adequate to represent that type of monomer regardless of where it is located within a cluster. (2) The polarization of a water molecule is affected by the presence of embedding charges, but the specific set of embedding charges used does not strongly affect the extent of the water molecule's polarization. The purpose of the background charges in an EE-MB calculation is to include higher-order effects in lower orders of the many-body expansion through the polarization of individual monomers and groups of monomers. Because the extent to which a water molecule is polarized is not greatly influenced by the choice of embedding charges, one can understand why the overall accuracy of the EE-MB approximation is not greatly influenced by the choice of embedding charges.

A concise summary of the major conclusions reached by this study is as follows: the EE-3B approximation as it was originally formulated (i.e., using geometry-independent background charges derived from equilibrium gas-phase monomer wave or density functions) can be trusted to provide accurate results for relatively complicated systems involving ions and noncovalently interacting monomers of widely varying sizes.

3.6. Supporting Information Available

Tables listing the Cartesian coordinates of the eight structures shown in Figure 3.1 are available free of charge via the Internet at http://pubs.acs.org and in the appendix to this chapter.

3.7. References for Chapter 3

(1) Cizek, J. *J. Chem. Phys.* **1966***, 45*, 4256.

(2) Shavitt, I. In Schaefer, H. F. I., Ed.; Methods of Electronic Structure Theory; Plenum: New York, 1977; pp 189-275.

(3) Bytautas, L.; Matsunaga, N.; Nagata, T.; Gordon, M. S.; Reudenberg, K. *J. Chem. Phys.* **2007***, 127*, 204301.

(4) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999***, 313*, 701.

(5) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005***, 122*, 154102.

(6) Hirata, S.; Valiev, M.; Dupuis, M. X.,S.S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005***, 103*, 2255.

(7) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2006***, 125*, 44903.

(8) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006***, 124*, 114112.

(9) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007***, 111*, 2193.

(10) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007***, 111*, 6904.

(11) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008***, 128*, 234108.

(12) Hirata, S. *J. Chem. Phys.* **2008***, 129*, 204104.

(13) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2007***, 3*, 46.

(14) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comp.* **2007***, 3*, 1342.

(15) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J Chem Theory Comput* **2008***, 4*, 33.

(16) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008***, 4*, 683.

(17) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008***, 4*, 1.

(18) Kulmala, M. *Science* **2003***, 302*, 1000.

(19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985***, 107*, 3902.

(20) Mulliken, R. S. *J. Chem. Phys.* **1955***, 23*, 1833.

(21) Baker, J. *Theor. Chim. Acta* **1985***, 68*, 221.

(22) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002***, 5*, 117.

(23) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008***, 120*, 215.

(24) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003***, 107*, 1384.

(25) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1197.

(26) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001***, 114*, 9244.

(27) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput. -Aided Mol. Design* **1995***, 9*, 872.

(28) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998***, 102*, 1820.

(29) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002***, 106*, 10707.

(30) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003***, 24*, 1291.

(31) Udier-Blagović , M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004***, 25*, 1322.

(32) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990***, 11*, 361.

(33) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984***, 5*, 129.

(34) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990***, 11*, 431.

(35) 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.; Pople, J. A. Gaussian 03 Online Manual. http://www.gaussian.com/g_ur/g03mantop.htm (accessed 2/13, 2009).

(36) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980***, 102*, 7211.

(37) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005***, 113*, 133.

(38) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007***, 3*, 2046.

(39) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996***, 93*, 281.

(40) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007***, 3*, 2011.

(41) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378.

(42) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006***, 125*, 194101.

(43) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972***, 16*, 217.

(44) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998***, 109*, 1223.

(45) Zhao, Y.; Truhlar, D. G. *Minnesota Gaussian Functional Module*, version 3.0; University of Minnesota: Minneapolis, MN, 2007.

(46) *Gaussian 03*, revision D.01; 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.; Pople, J. A.; Gaussian, Inc.: Wallingford, CT, 2004.

(47) Olson, R. M.; Marenich, A. V.; 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. *Minnesota Gaussian Solvation Module,* version 2008; University of Minnesota: Minneapolis, MN, 2008.

(48) Marenich, A. V.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *GESOL*, version 2008; University of Minnesota: Minneapolis, MN, 2008.

(49) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Zhu, T.; Thompson, J. D.; Winget, P.; Lynch, B. J.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 7.1; University of Minnesota: Minneapolis, MN, 2004.

(50) Dahlke, E. E.; Truhlar, D. G. *MBPAC*, version 2007-2; University of Minnesota: Minneapolis, MN, 2007.

(51) Larson, L. J.; Largent, A.; Tao, F. *J. Phys. Chem. A* **1999***, 103*, 6786.

(52) Anderson, K. E.; Siepmann, J. I.; McMurry, P. H.; VandeVondele, J. *J. Am. Chem. Soc.* **2008***, 130*, 14144.

(53) Boys, S. F.; Bernardi, D. *Mol. Phys.* **1970***, 19*, 553.

(54) Valiron, P.; Mayer, I. *Chem. Phys. Lett.* **1997***, 275*, 46.

(55) Dyke, T. R.; Muenter, J. S. *J. Chem. Phys.* **1973***, 59*, 3125.

(56) Clough, S. A.; Beers, Y.; Klein, G. P.; Rotham, L. S. *J. Chem. Phys.* **1973***, 59*, 2254.

(57) Shostak, S. L.; Ebenstein, W. L.; Muenter, J. S. *J. Chem. Phys.* **1991***, 94*, 5875.

(58) Sedo, G.; Schultz, J.; Leopold, K. R. *J. Mol. Spectrosc.* **2007***, 251*, 4.

(59) Iwahori, J.; Ueda, Y.; Nakagawa, K. *J. Mol. Spectrosc.* **1986***, 117*, 1.

Table 3.1. Binding Energies (E_{bind}, kcal/mol) of Eight Clusters with M06-2X Density Functional and Three Basis sets and the Corresponding Errors[a] (kcal/mol) from the EE-PA and EE-3B Calculations using Five Different Sets of Geometry-Independent Background Charges

		E_{bind}	*EE-PA Errors*					*EE-3B Errors*				
Basis	System	Full	ChElPG	ESP-Dipole	MK	NBO	CM4M	ChElPG	ESP-Dipole	MK	NBO	CM4M
MG3S	1_H2O	-30.50	-0.57	-0.62	-0.59	-0.12	-0.98					
cc-pV(T+d)Z+	1_H2O	-29.69	-0.61	-0.65	-0.63	-0.17	-1.04					
aug-cc-pV(T+d)Z	1_H2O	-29.57	-0.34	-0.38	-0.36	-0.06	-0.66					
MG3S	2_H2O	-46.14	-1.36	-1.42	-1.36	-0.74	-1.87	0.06	0.06	0.06	0.02	0.06
cc-pV(T+d)Z+	2_H2O	-45.31	-1.43	-1.48	-1.43	-0.82	-1.96	0.03	0.04	0.03	-0.01	0.04
aug-cc-pV(T+d)Z	2_H2O	-45.08	-1.16	-1.20	-1.16	-0.65	-1.56	0.03	0.04	0.03	0.00	0.01
MG3S	1A	-52.23	1.11	1.17	1.06	0.04	2.19	0.17	0.17	0.17	0.11	0.22
cc-pV(T+d)Z+	1A	-51.80	1.14	1.18	1.08	0.01	2.23	0.15	0.15	0.14	0.10	0.19
aug-cc-pV(T+d)Z	1A	-50.87	1.23	1.31	1.17	-0.06	2.18	0.16	0.16	0.15	0.08	0.19
MG3S	1B	-31.39	0.48	0.50	0.44	-0.26	1.31	-0.06	-0.07	-0.06	-0.06	-0.05
cc-pV(T+d)Z+	1B	-31.80	0.50	0.51	0.45	-0.27	1.32	-0.05	-0.05	-0.05	-0.05	-0.04
aug-cc-pV(T+d)Z	1B	-30.89	0.53	0.58	0.48	-0.39	1.27	-0.04	-0.04	-0.04	-0.03	-0.03
MG3S	1C	-43.89	0.61	0.63	0.55	0.07	1.29	0.14	0.14	0.13	0.10	0.17
cc-pV(T+d)Z+	1C	-43.36	0.68	0.70	0.62	0.09	1.37	0.15	0.15	0.15	0.12	0.18
aug-cc-pV(T+d)Z	1C	-42.31	0.79	0.84	0.73	0.06	1.38	0.13	0.13	0.13	0.11	0.15
MG3S	2A	-66.84	-2.18	-2.23	-2.14	-1.58	-2.83	0.11	0.11	0.10	0.10	0.12
cc-pV(T+d)Z+	2A	-66.30	-2.21	-2.25	-2.16	-1.67	-2.82	0.14	0.13	0.13	0.12	0.15
aug-cc-pV(T+d)Z	2A	-65.24	-2.24	-2.28	-2.20	-1.76	-2.73	0.22	0.21	0.21	0.06	0.26
MG3S	2B	-68.53	-2.54	-2.61	-2.50	-1.79	-3.18	0.12	0.13	0.12	0.06	0.17
cc-pV(T+d)Z+	2B	-67.65	-2.63	-2.68	-2.58	-1.91	-3.23	0.14	0.14	0.13	0.07	0.20
aug-cc-pV(T+d)Z	2B	-66.71	-2.56	-2.60	-2.51	-1.83	-3.01	0.22	0.23	0.22	0.06	0.30
MG3S	2C	-62.70	-4.00	-4.15	-3.97	-2.55	-5.28	0.34	0.34	0.33	0.31	0.37
cc-pV(T+d)Z+	2C	-62.13	-4.03	-4.17	-3.99	-2.65	-5.27	0.29	0.29	0.27	0.26	0.30
aug-cc-pV(T+d)Z	2C	-61.29	-4.07	-4.22	-4.04	-2.58	-5.12	0.38	0.38	0.37	0.27	0.41

[a]The errors are calculated as E_{bind}(EE-MB) – E_{bind}(full).

Table 3.2. Binding Energies [E_{bind}(full), kcal/mol] from Conventional Calculations at M06-2X/cc-pV(T+d)Z+ and the Corresponding Errors[a] (kcal/mol) from the EE-MB Calculations when Different sets of Background Charges are Used

	2A		*2B*		*2C*	
	EE-PA	EE-3B	EE-PA	EE-3B	EE-PA	EE-3B
E_{bind}(full)	-66.30		-67.65		-62.13	
Charge Model			*Errors*			
None[b]	-4.94	-0.85	-8.51	-0.14	-15.99	-1.11
GI_ChElPG[c]	-2.21	0.14	-2.63	0.14	-4.03	0.29
GI_ESP-Dipole	-2.25	0.13	-2.68	0.14	-4.17	0.29
GI_MK	-2.16	0.13	-2.58	0.13	-3.99	0.27
GI_NBO	-1.67	0.12	-1.91	0.07	-2.65	0.26
GI_CM4M	-2.82	0.15	-3.23	0.20	-5.27	0.30
GI_CM1A	-2.29	0.14	-2.71	0.15	-4.21	0.31
GI_(CM1A*1.14)	-2.37	0.21	-2.76	0.23	-3.75	0.37
GI_CM2	-2.33	0.19	-2.83	0.16	-4.06	0.44
GI_CM3	-2.42	0.18	-2.93	0.17	-4.31	0.43
GI_SM5.42/CM2	-2.10	0.17	-2.50	0.12	-3.53	0.39
GI_SM8/CM4M	-2.40	0.14	-2.76	0.16	-4.36	0.26
GI_SMD/CM4M	-2.35	0.14	-2.70	0.15	-4.24	0.26
GD_AM1-Mulliken[d]	-2.96	0.18	-3.82	0.20	-6.42	0.52
GD_CM1A	-1.81	0.10	-2.04	0.08	-3.36	0.18
GD_CM1A*1.14	-1.98	0.15	-2.16	0.12	-2.85	0.21
GD_AM1-CM2	-1.79	0.11	-2.20	0.06	-3.67	0.29
GD_AM1-CM3	-1.80	0.11	-2.18	0.08	-3.50	0.26

[a]The errors are calculated as E_{bind}(EE-MB) – E_{bind}(full).

[b]"None" implies that no electrostatic embedding was used for these calculations; i.e., this row gives PA and 3B errors, not EE-PA and EE-3B errors.

[c]The "GI_" prefix indicates that these background charges are geometry independent. An "SM" following this prefix indicates that the charges were obtained from aqueous-phase monomers; all others were obtained from gas-phase monomers.

[d]The "GD_" prefix indicates that these background charges are geometry dependent.

Table 3.3. Mean Unsigned Errors (MUE) and Root Mean Squared Errors (RMSE) in kcal/mol Over Three Configurations (2A, 2B, and 2C) of an $(H_2SO_4)(HSO_4^-)(NH_4^+)(H_2O)_6$ System[a]

	EE-PA		*EE-3B*	
Charge Model	MUE	RMSE	MUE	RMSE
Full	0.00	0.00	0.00	0.00
None[b]	9.82	10.84	0.70	0.81
GI_ChElPG[c]	2.96	3.06	0.19	0.20
GI_ESP-Dipole	3.03	3.14	0.19	0.20
GI_MK	2.91	3.01	0.18	0.19
GI_NBO	2.08	2.12	0.15	0.17
GI_CM4M	3.77	3.92	0.22	0.23
GI_CM1A	3.07	3.18	0.20	0.22
GI_(CM1A*1.14)	2.96	3.02	0.27	0.28
GI_CM2	3.07	3.16	0.26	0.29
GI_CM3	3.22	3.32	0.26	0.29
GI_SM5.42/CM2	2.71	2.78	0.23	0.26
GI_SM8/CM4M	3.17	3.28	0.19	0.19
GI_SMD/CM4M	3.10	3.21	0.18	0.19
GD_AM1-Mulliken[d]	4.40	4.64	0.30	0.34
GD_CM1A	2.41	2.50	0.12	0.13
GD_CM1A*1.14	2.33	2.36	0.16	0.17
GD_AM1-CM2	2.56	2.68	0.15	0.18
GD_AM1-CM3	2.50	2.60	0.15	0.17

[a]The full (or conventional) and EE-MB calculations were performed with the M06-2X/cc-pV(T+d)Z+ method.

[b]No background charges were used for these calculations; i.e., these are the PA and 3B approximations to the total energy without electrostatic embedding.

[c]The "GI_" prefix indicates that these background charges are geometry independent. An "SM" following this prefix indicates that the charges were obtained from aqueous-phase monomers; all others were obtained from gas-phase monomers.

[d]The "GD_" prefix indicates that these background charges are geometry dependent.

Table 3.4. Geometry-Independent Background Charges (in *e*) Based on the Geometries of the Gas-Phase Monomers Optimized with the M06-2X/cc-pV(T+d)Z+ Method

Molecule	Atom Type	ChElPG[a]	Dipole[a]	ESP-MK[a]	NBO[a]	CM4M[b]
H_2O	O	-0.726	-0.709	-0.731	-0.930	-0.601
H_2O	H	0.363	0.355	0.365	0.465	0.300
HSO_4^-	S	1.428	1.329	1.328	2.591	0.403
HSO_4^-	O	-0.700	-0.677	-0.676	-1.016	-0.423
HSO_4^-	H	0.372	0.378	0.375	0.472	0.289
H_2SO_4	S	1.164	1.042	1.049	2.602	0.499
H_2SO_4	O	-0.509	-0.482	-0.483	-0.910	-0.298
H_2SO_4	H	0.435	0.443	0.442	0.519	0.347
NH_4^+	N	-0.784	-0.834	-0.834	-0.859	-0.604
NH_4^+	H	0.446	0.458	0.458	0.465	0.401

[a]Charge analyses were done on the M06-2X/cc-pV(T+d)Z+ gas-phase monomer density matrices.

[b]Charge analyses were done on the M06-2X/MIDI! gas-phase monomer density matrices.

Table 3.5. Geometry-Independent Background Charges (in *e*) Based on the Geometries of the Gas-Phase Monomers Optimized with the AM1 Method

Molecule	Atom Type	CM1A[a]	CM1A*1.14[a]	CM2[a]	CM3[a]	SM5.42/ CM2[b]
H_2O	O	-0.706	-0.805	-0.711	-0.679	-0.783
H_2O	H	0.353	0.402	0.356	0.340	0.392
HSO_4^-	S	1.433	1.634	2.878	2.491	2.934
HSO_4^-	O	-0.709	-0.808	-1.061	-0.963	-1.086
HSO_4^-	H	0.401	0.457	0.368	0.360	0.409
H_2SO_4	S	1.440	1.642	2.874	2.487	2.961
H_2SO_4	O	-0.601	-0.685	-0.937	-0.842	-0.974
H_2SO_4	H	0.481	0.548	0.438	0.439	0.468
NH_4^+	N	-0.514	-0.586	-0.793	-0.829	-0.792
NH_4^+	H	0.378	0.431	0.448	0.457	0.448

[a]Charge analyses were done on the AM1 gas-phase monomer density matrices.
[b]Charge analyses were done on the SM5.42/AM1 aqueous-phase monomer density matrices.

Table 3.6. Geometry-Independent Background Charges[a] (in *e*) Based on the Geometries of the Gas-Phase Monomers Optimized by the M06-L/6-31G(d) Method

Molecule	Atom + Label	SM8/CM4M	SMD/CM4M
H_2O	O	-0.695	-0.708
H_2O	H	0.347	0.354
HSO_4^-	S	0.751	0.775
HSO_4^-	O	-0.525	-0.534
HSO_4^-	H	0.350	0.361
H_2SO_4	S	0.875	0.867
H_2SO_4	O	-0.416	-0.418
H_2SO_4	H	0.393	0.401
NH_4^+	N	-0.584	-0.592
NH_4^+	H	0.396	0.398

[a]Charge analyses were done on the SM*x*/M06 L/6-31G(d) (*x* = 8, D) aqueous-phase monomer density matrices.

Table 3.7. Dipoles[a] (in debye) of Individual Water Molecules Within Different Configurations of an $(H_2SO_4)(HSO_4^-)(NH_4^+)(H_2O)_6$ System from Various Point Charge Representations of those Configurations

Cluster label	Monomer label	Mulliken	CM1A	CM1A*1.14	CM2	CM3
2A	2	1.39	2.26	2.58	2.24	2.21
2A	4	1.37	2.20	2.50	2.18	2.15
2A	5	1.58	2.47	2.81	2.43	2.42
2A	6	1.35	2.30	2.62	2.31	2.24
2A	7	1.23	2.13	2.42	2.12	2.06
2A	9	1.30	2.22	2.53	2.21	2.15
2B	2	1.48	2.40	2.74	2.37	2.34
2B	4	1.39	2.24	2.55	2.22	2.19
2B	5	1.39	2.20	2.51	2.17	2.16
2B	6	1.30	2.19	2.49	2.20	2.13
2B	7	1.25	2.24	2.55	2.24	2.16
2B	9	1.31	2.19	2.50	2.18	2.13
2C	2	1.54	2.46	2.81	2.39	2.40
2C	4	1.42	2.25	2.56	2.21	2.20
2C	5	1.34	2.19	2.50	2.18	2.14
2C	6	1.44	2.44	2.79	2.45	2.38
2C	7	1.21	2.12	2.42	2.12	2.05
2C	9	1.18	2.08	2.38	2.10	2.01
Average (debye)		1.36	2.25	2.57	2.24	2.20
Standard Deviation (debye)		0.11	0.12	0.13	0.11	0.12
% Standard Deviation		8.0	5.2	5.2	4.7	5.4

[a]Dipoles are calculated with respect to each water molecule's center of nuclear charge.

Table 3.8. Dipoles[a] (in debye) of Gas-Phase Monomers Calculated from Point Charges and Compared to Best Estimates

Charge Model	H_2O	H_2SO_4	NH_3	HSO_4^-	OH^-	MUE[b]	RMSE[b]
ChElPG	2.03	3.25	1.64	2.65	1.59	0.19	0.21
ESP-Dipole	1.98	3.31	1.59	2.72	1.59	0.20	0.22
MK	2.04	3.27	1.67	2.68	1.59	0.21	0.22
NBO	2.60	3.70	1.89	3.24	2.20	0.68	0.70
CM4M	1.68	3.02	1.55	2.40	0.95	0.18	0.21
CM1A	2.02	2.25	1.76	1.83	1.38	0.40	0.49
CM1A*1.14	2.30	2.56	2.01	2.08	1.57	0.43	0.44
CM2	2.03	1.89	1.61	1.59	1.40	0.49	0.67
CM3	1.94	1.87	1.60	1.55	1.29	0.48	0.68
Best Estimate	1.85[c]	2.96[d]	1.47[e]	2.60[f]	1.33[f]		

[a]The dipole of each compound was calculated with respect to the compound's center of nuclear charge.

[b]The MUEs/RMSEs are averages of the differences between the point-charge derived and the best estimate dipoles over all five compounds. The MUEs/RMSEs are given in debye.

[c]References 55-57.

[d]Reference 58.

[e]Reference 59.

[f]Determined in the present study by a finite-field calculation at CCSD(T)/aug-cc-pV(T+d)Z//M06-2X/cc-pV(T+d)Z+.

Table 3.9. CM4M Charges (in *e*) and Dipoles (in debye) of Monomer Number 5 from Configuration 2A Embedded in the Given Sets of Background Charges

Background Charges	H1[a] (*e*)	O (*e*)	H2[b] (*e*)	Density Dipole[c] (D)	Point-Charge Dipole[d] (D)
Full[e]	0.321	-0.593	0.321	N/A[e]	1.87[e]
None[f]	0.297	-0.596	0.298	1.95	1.76
GI_ChElPG	0.371	-0.687	0.317	2.61	2.05
GI_ESP-Dipole	0.371	-0.687	0.317	2.61	2.05
GI_MK	0.370	-0.687	0.317	2.61	2.05
GI_NBO	0.376	-0.690	0.314	2.63	2.06
GI_CM4M	0.366	-0.684	0.318	2.59	2.04
GI_CM1A	0.370	-0.686	0.315	2.60	2.05
GI_CM1A*1.14	0.380	-0.697	0.318	2.68	2.09
GI_CM2	0.377	-0.693	0.316	2.65	2.07
GI_CM3	0.375	-0.692	0.316	2.64	2.07
GI_SM5.42/CM2	0.377	-0.692	0.315	2.64	2.07
GI_SM8/CM4M	0.367	-0.684	0.317	2.59	2.04
GI_SMD/CM4M	0.368	-0.684	0.317	2.59	2.04
GD_Mulliken	0.378	-0.690	0.312	2.63	2.07
GD_CM1A	0.371	-0.684	0.313	2.59	2.05
GD_CM1A*1.14	0.380	-0.696	0.316	2.68	2.08
GD_CM2	0.377	-0.689	0.311	2.62	2.06
GD_CM3	0.377	-0.690	0.314	2.63	2.07
Average[g]	0.374	-0.689	0.315	2.62	2.06
Standard Deviation[g]	0.005	0.004	0.002	0.03	0.01
% Standard Deviation[g]	1.2	0.6	0.7	1.1	0.7

[a]The hydrogen atom that forms an H-bond with an oxygen atom of the nearby HSO_4^- ion.

[b]The hydrogen atom that is not involved in an H-bond.

[c]Dipole calculated from the M06 2X/MIDI! wave function of the embedded water molecule. 1 D ≡ 1 debye.

[d]Dipole calculated from the point charges given in the columns labeled H1, O, and H2. 1 D ≡ 1 debye.

[e]The CM4M charges assigned to the water molecule when the M06-2X/MIDI! calculation is performed on the entire 2A configuration. Because in this case the sum of the point charges on this fragment is not zero, the point charge dipole moment was calculated with respect to this fragment's center of nuclear charge.

[f]"None" indicates that water monomer 5 was left in the geometry that it has in configuration 2A but that it was not embedded in point charges.

[g]The values found in the rows labeled "Full" and "None" were not included in the calculations of the averages, standard deviations, or % standard deviations.

Figures

Figure 3.1: Eight clusters formed from water, ammonia, and sulfuric acid (note that many of the clusters contain ammonium and bisulfate ions rather than neutral ammonia and sulfuric acid molecules). The 1_H2O and 2_H2O structures were optimized at M06-2X/MG3S. The remaining structures were taken from molecular dynamics simulations. The composition of each cluster is as follows:

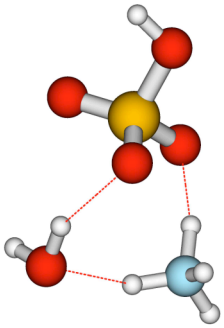
(a) 1_H2O = $(HSO_4^-)(NH_4^+)(H_2O)$,

(b) 2_H2O = $(HSO_4^-)(NH_4^+)(H_2O)_2$,

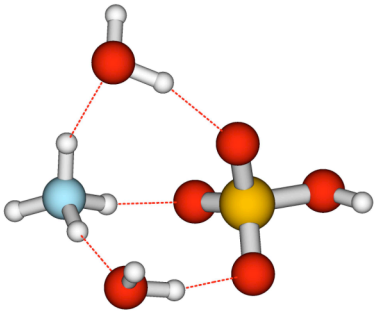
(c) 1A = 1B = 1C = $(H_2SO_4)(NH_3)(H_2O)_6$, and

(d) 2A = 2B = 2C = (H_2SO_4) $(HSO_4^-)(NH_4^+)(H_2O)_6$.

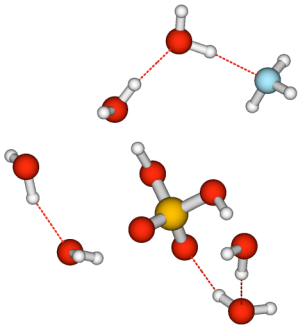
Figure 3.2: The 2A configuration with the water molecules labeled with the arbitrary fragment numbers that they were assigned for the EE-MB calculations. Note that two hydrogen atoms are obstructed from view in this figure: one is obstructed by the nitrogen atom of the ammonium ion, and the other is obstructed by the oxygen atom of water molecule 7.



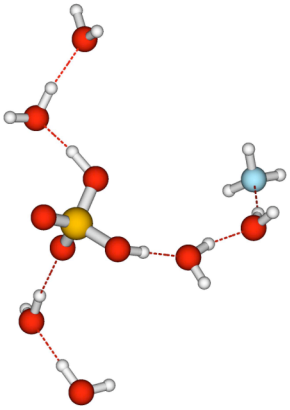
1_H2O



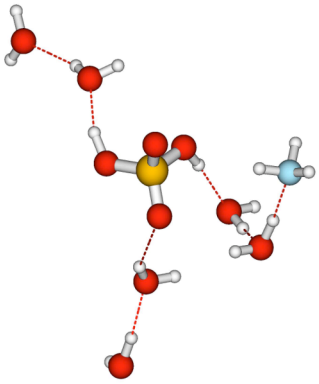
2_H2O



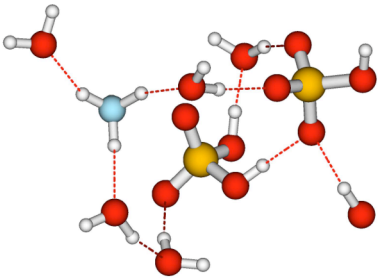
1A



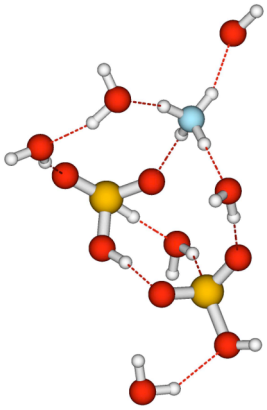
1B



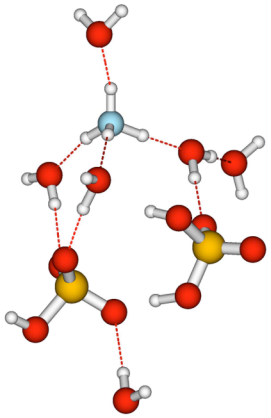
1C



2A



2B



2C

Figure 3.1

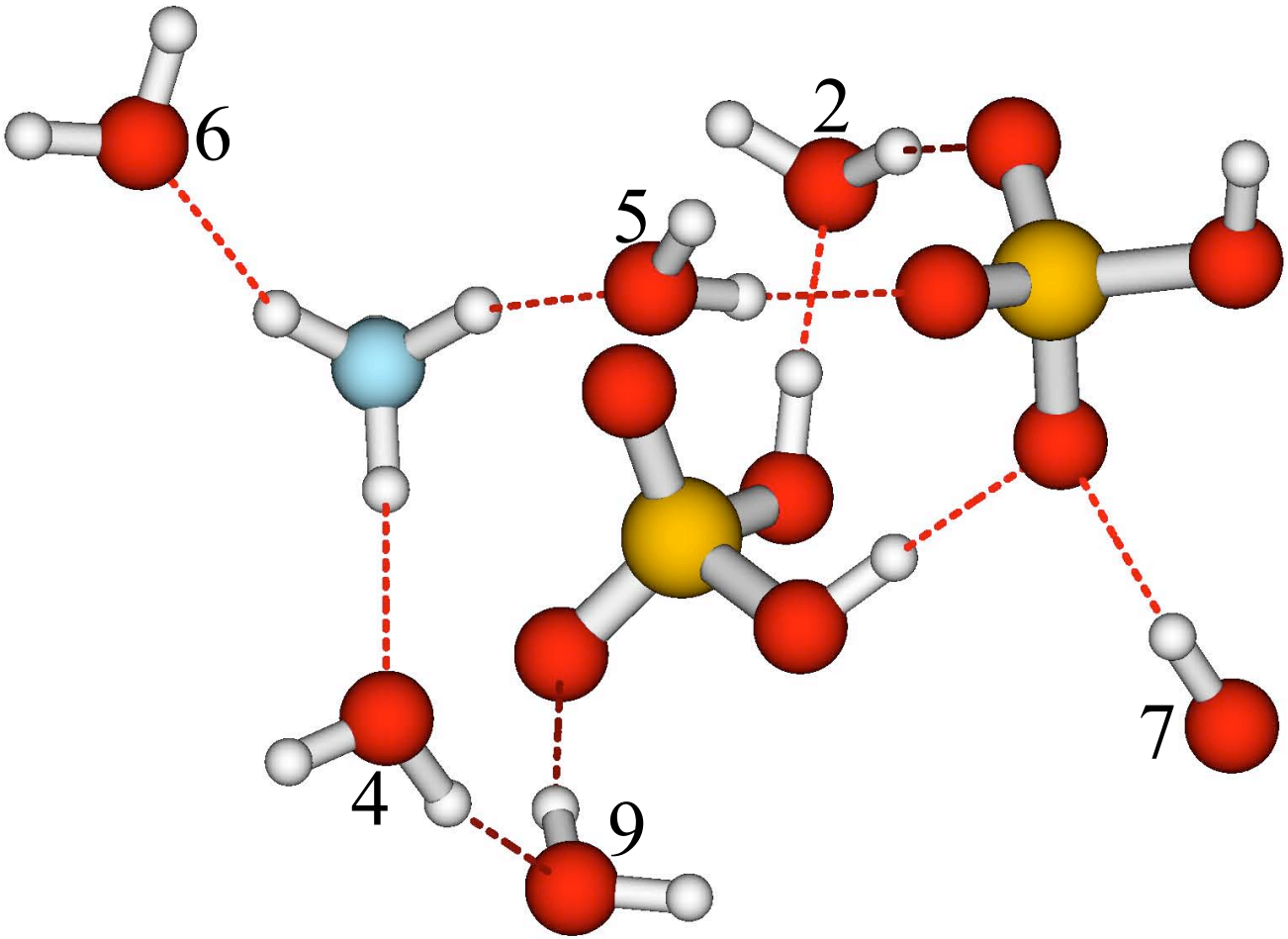


Figure 3.2

Chapter 4. Exploring Variations of the Electrostatically Embedded Many-Body Approximation for Monte Carlo Simulations

4.1. The Electrostatically-Embedded Three-Body Approximation with Consistently Screened Perturbations for Monte Carlo Simulations[3]

4.1.1 Introduction

The electrostatically-embedded three-body (EE-3B) approximation has shown excellent agreement with conventional (or full) calculations done at the same level of electronic structure theory,[1-4] and can save significant amounts of computer time even for clusters containing 6–20 heavy atoms, depending on the level of electronic structure theory used (the savings afforded by the EE-3B method when it is used with correlated levels of wave function theory are more dramatic than those seen when it is used with density functional theory).[1, 2] However, using the EE-3B method to compute the 100,000+ single-point energies of cluster configurations generated during a Monte Carlo (MC) simulation still would be far too expensive to be practical. Thus, we continue to seek ways to cut the cost of *ab initio* electronic structure calculations so that they may be used to study dynamic systems through MC or MD (molecular dynamics) simulations. In this study, we investigate the successes and failures of one approximation method that has been proposed to pare the cost of the EE-3B method: the electrostatically-embedded three-body approximation with consistently screened perturbations (EE-3B-CSP).

---

[3] The authors would like to thank Steven Mielke for many helpful insights, Katie Maerzke for providing the configurations of the water–ammonia systems that were used in this study, and the Minnesota Supercomputing Institute for computer time.

4.1.2 Theory

4.1.2.A The EE-3B Method

The EE-3B method approximates the total energy of a system (E) as the sum of the electrostatically-embedded one-, two-, and three-body interaction energies (V_1', V_2', and V_3', respectively; the primes denote that energies are calculated with electrostatic embedding). That is,

$$E \approx E_{\mathrm{EE-3B}} = V_1' + V_2' + V_3' \tag{1}$$

where

$$V_1' = \sum_i E_i' \tag{2}$$

$$V_2' = \sum_{i<j} E_{ij}' - E_i' - E_j' = \sum_{i<j} v_{ij}' \tag{3}$$

$$V_3' = \sum_{i<j<k} E_{ijk}' - v_{ij}' - v_{ik}' - v_{jk}' - E_i' - E_j' - E_k' = \sum_{i<j<k} v_{ijk}' \tag{4}$$

In eqs 2–4, E_i' denotes the energy of monomer *i* embedded in a constellation of point charges located at the nuclei of all other atoms in the entire system. Similarly, E_{ij}' and E_{ijk}' represent the electrostatically-embedded energies of dimer *ij* and trimer *ijk*. The sums in eqs 2–4 run over all possible monomers, dimers, and trimers in the system. The definitions of the terms v_{ij}' and v_{ijk}' are implied in eqs 3 and 4, respectively. Physically, the terms v_{ij}' and v_{ijk}' are dimer *ij*'s contribution to the two-body interaction energy (V_2') and trimer *ijk*'s contribution to the three-body interaction energy (V_3'), respectively.

4.1.2.B The EE-3B-CSP Method

The EE-3B-CSP method starts out the same as the EE-3B method: for a given initial configuration (step 1), the EE-3B energy is determined as usual, but if the site-site version of the EE-3B-CSP method is being used (see below for clarification), an inexpensive partial charge analysis on the electron density of every possible monomer, dimer, and trimer in the initial configuration is done as well. After moving an atom or molecule to form a slightly different configuration (as one might do in a sequence of Monte Carlo moves), as long as the fragmentation scheme did not change, the EE-3B-CSP energy may be calculated for the new configuration, using the original configuration as a reference. Eqs 1-4 are still used to govern the EE-3B-CSP approximation, but the difference between the EE-3B and the EE-3B-CSP approximations lies in the way in which the individual trimers' electrostatically-embedded three-body contributions (v_{ijk}') are defined. If a trimer contains an atom that moved with respect to a reference configuration, then the trimer energy v_{ijk}' is defined in the same way as above and its energy must be calculated using the chosen level of electronic structure theory. If a trimer does not contain an atom that moved, however, then it might be reasonable to assume that the changes in the trimer's energy and wave function due to the change in the position of one (or a few) of its embedding charges are small, and we might be able to use first-order perturbation theory to approximate its contribution to the three-body energy in the system's new configuration based on the monomer, dimer, and trimer energies computed in some earlier configuration of the system. In this way, one might not need to use a relatively expensive electronic

structure calculation to obtain the energy of each trimer at every step of the MC simulation but instead might be able to quickly estimate the trimer's energy in a given configuration using perturbation theory. However, in some cases the assumption that the change in a trimer's energy due to the change in the locations of a few external charges is small is not valid. Therefore, one must "screen" the estimated change in a given trimer's energy to determine whether the estimated energy value will be accepted and used to approximate that trimer's v_{ijk}' value in the current configuration or the estimated energy value will not be used and the trimer's energy will need to be recalculated in the current constellation of point charges using the chosen level of electronic structure theory.

In the EE-3B-CSP method, the user must define a parameter, ε, that is used to screen the perturbation energies to see if they may be used to approximate a trimer's energy in a new constellation of point charges. If the change in the trimer's energy estimated by first-order perturbation theory is larger than ε, then the trimer's energy in the new constellation must be recalculated accurately using electronic structure theory; otherwise, the perturbation theory estimate is used to approximate the trimer's energy. If for any reason a trimer's energy must be calculated accurately at a certain MC step, then that step is considered a reference step for that particular trimer, and future perturbation theory estimates of that trimer's energy will be based on that reference configuration. The user must also define a positive integer parameter, n, so that all trimer energies will be "refreshed" – recalculated accurately using electronic structure theory in the current configuration – every n^{th} accepted MC step.

In order to use first-order perturbation theory to estimate a trimer's contribution to the system's three-body energy, we approximate v_{ijk}' according to eq 5a:

$$v_{ijk}' \approx \tilde{v}_{ijk}' \equiv \tilde{E}_{ijk}^{(ijk)} - \left(\tilde{E}_{ij}^{(ijk)} - \tilde{E}_{i}^{(ijk)} - \tilde{E}_{j}^{(ijk)}\right) - \left(\tilde{E}_{ik}^{(ijk)} - \tilde{E}_{i}^{(ijk)} - \tilde{E}_{k}^{(ijk)}\right) - \left(\tilde{E}_{jk}^{(ijk)} - \tilde{E}_{j}^{(ijk)} - \tilde{E}_{k}^{(ijk)}\right) - \tilde{E}_{i}^{(ijk)} - \tilde{E}_{j}^{(ijk)} - \tilde{E}_{k}^{(ijk)} \quad (5a)$$

In eq 5a, $\tilde{E}_{L}^{(ijk)}$ is the first-order perturbation theory estimate of the electrostatically embedded energy of oligomer *L*, where *L* can be replaced by the subscripts that label a monomer, dimer, or trimer (for example, *L* could be replaced by *i*, *ij*, or *ijk*). The superscript (*ijk*) is used because each estimated energy depends on the configuration of the system at the Monte Carlo step when the most recent reference energy of trimer *ijk* was calculated, which might be different for each trimer.

$\tilde{E}_{L}^{(ijk)}$ is given by:

$$\tilde{E}_{L}^{(ijk)} \approx E_{L}^{(0,ijk)} + E_{L}^{(1,ijk)} \quad (5)$$

In the eq. above, $E_{L}^{(0,ijk)}$ is the energy of oligomer *L* that was calculated at trimer *ijk*'s most recent reference step, when the oligomer was embedded in a slightly different constellation of point charges. $E_{L}^{(1,ijk)}$ can be either the true first-order perturbation theory estimate of the change in oligomer *L*'s energy due to the movement of the embedding charges (given in eqs 6 and 7), or it can be an approximation of the first-order perturbation energy (see eq 8). Each of these options is described in the following paragraphs. Before proceeding with the details, however, it is important to notice that, although the monomer and dimer energies are calculated accurately using electronic

structure theory at each MC step, one must use first-order perturbation theory to estimate the energies of trimer *ijk*'s constituent monomers and dimers when computing $\tilde{v}_{ijk}'$. This improves the accuracy of our estimate of v_{ijk}' because some of the errors inherent in the perturbation theory estimates of oligomer energies are likely to cancel out when the differences between the estimated energies are taken. This is why the method is called EE-MB with *consistently* screened perturbations. If accurate monomer and dimer energies were used with the estimated trimer energy in the expression for $\tilde{v}_{ijk}'$, then this error cancellation would not occur and our estimates of v_{ijk}' would actually be *worse*.

Using true first-order perturbation theory, one writes the following expression for the first-order perturbation energy of oligomer *L*:

$$E_L^{(1,ijk)} = \left\langle \Psi_L^{(0,ijk)} \middle| V_L^{(ijk)} \middle| \Psi_L^{(0,ijk)} \right\rangle \tag{6}$$

where $\left| \Psi_L^{(0,ijk)} \right\rangle$ is the wave function of oligomer *L* when it is in the configuration corresponding to the most recent reference state of trimer *ijk* and where $V_L^{(ijk)}$ is a perturbation energy operator given by:

$$V_L^{(ijk)} = \sum_{\lambda=1}^{n_p^{(ijk)}} \sum_{A=1}^{n_2^{(L)}} \left[\sum_{\mu=1}^{n_e^{(L)}} q_A \left(\frac{1}{\left| \mathbf{r}_\mu - \mathbf{R}_A^{(\lambda)} \right|} - \frac{1}{\left| \mathbf{r}_\mu - \mathbf{R}_A^{(\lambda-1)} \right|} \right) + \sum_{a=1}^{n_1^{(L)}} Z_a q_A \left(\frac{1}{\left| \mathbf{R}_a - \mathbf{R}_A^{(\lambda)} \right|} - \frac{1}{\left| \mathbf{R}_a - \mathbf{R}_A^{(\lambda-1)} \right|} \right) \right] \tag{7}$$

Individual terms of eq 7 are defined as follows: $n_e^{(L)}$ is the number of electrons in oligomer *L*, $n_1^{(L)}$ is the number of atoms in oligomer *L*, $n_2^{(L)}$ is the number of

embedding charges that surround oligomer L, Z_a is the nuclear charge on atom a of oligomer L, $\mathbf{R}_a$ is the three-dimensional vector indicating the position of atom a, and $\mathbf{r}_\mu$ is the three-dimensional vector indicating the position of electron μ of oligomer L. If more than one Monte Carlo move is made between reference calculations, then the perturbation energy for a given oligomer should actually be the *sum* of the perturbation energies from all of the preceding non-references steps, so $n_{\mathrm{p}}^{(ijk)}$ is the number of Monte Carlo steps counted from the most recent reference calculation of trimer *ijk*'s energy (i.e., $n_{\mathrm{p}}^{(ijk)}$is the number of perturbation steps), and λ labels a specific perturbation step. One should be aware that $n_{\mathrm{p}}^{(ijk)}$ depends on the specific trimer: every time a trimer needs its energy recalculated, either because it contained an atom that moved or because it failed to pass an energy screening test (which will be explained below), that recalculated energy will then be used as the reference energy for that trimer in future steps. q_A and $\mathbf{R}_A^{(\lambda)}$ are the charge value and location, respectively, of embedding charge A at perturbation step λ, and $\mathbf{R}_A^{(\lambda-1)}$ is the location of embedding charge A at the preceding accepted perturbation step, with $\mathbf{R}_A^{(0)}$ corresponding to the location of embedding charge A in the most recent reference configuration for trimer *ijk*. In order to ensure that the geometry of a system does not change so much over a series of steps that it no longer can be considered a perturbation of the reference geometry, a full EE-3B calculation may be done every n steps to reset the reference energies for *all* oligomers simultaneously (the specific value used for n is defined by the user of the EE-3B-CSP method).

In order to cut computational costs further, one could eliminate the need to calculate the one-electron integrals implied in eq 6 by using an additional set of point charges (*not* the same as the embedding charges) that roughly represent the charge density distribution of oligomer *L* instead of using the squared modulus of the wave function and the set of nuclear charges $\{Z_a\}$. That is, one could estimate the perturbation energy as follows:

$$E_L^{(1,ijk)} = \sum_{\lambda=1}^{n_{\mathrm{p}}^{(ijk)}} \sum_{A=1}^{n_2^{(L)}} \sum_{a=1}^{n_1^{(L)}} Q_a q_A \left(\frac{1}{\left| \mathbf{R}_a - \mathbf{R}_A^{(\lambda)} \right|} - \frac{1}{\left| \mathbf{R}_a - \mathbf{R}_A^{(\lambda-1)} \right|} \right) \qquad (8)$$

The only term in eq 8 that does not also appear in eq 7 is Q_a, which is the partial charge assigned to atom *a* of oligomer *L* when a charge analysis is performed on oligomer *L* while it is surrounded by the constellation of embedding charges corresponding to the most recent reference state of trimer *ijk*. All other terms are defined in the same way that they are in eq 7.

When one uses the true first-order perturbation theory expressed in eqs 6 and 7 to estimate an oligomer's energy in a new constellation of point charges, we say that he or she has used first-order site-orbital perturbation theory. When one uses eq 8 to estimate the change in the oligomer's energy due to the moved point charges, we say that he or she has used first-order site-site perturbation theory.

Regardless of whether one uses first-order site-orbital or first-order site-site perturbation theory to obtain an oligomer's energy, one is assuming that the wave function/charge distribution of the oligomer hardly changes at all due to small changes in the locations of a few embedding charges. In some cases, however, particularly if the

embedding charges that moved are quite close to the oligomer in question, this assumption may not be valid, and it might be necessary to recalculate E_L' more accurately (i.e., do a full quantum mechanical calculation of the energy of embedded oligomer *L* or use second-order perturbation theory) instead of relying on first-order perturbation theory. We therefore attempt to "screen out" such dangerous instances by requiring that the absolute value of the perturbation energy, $\left|E_L^{(1,ijk)}\right|$, be less than or equal to a user-defined energy parameter, ε. If $\left|E_L^{(1,ijk)}\right| > \varepsilon$, then we assume that this implies that the energy change due to the moved point charges in the true E_L' is also relatively large, which could be an indication that the wave function changed significantly and that first-order perturbation theory is no longer valid; conversely, we also assume that if the change in E_L' is large, then it is likely that $\left|E_L^{(1,ijk)}\right| > \varepsilon$. In other words, we assume that $E_L^{(1,ijk)}$ correlates with the true change in E_L'. When $\left|E_L^{(1,ijk)}\right| > \varepsilon$, trimer *ijk*'s energy (E_{ijk}') is calculated quantum mechanically in the new constellation of embedding charges, the true value of v_{ijk}' is used in eq 4 rather than the approximate value $\tilde{v}_{ijk}'$, and the current configuration becomes the new reference configuration for that trimer for subsequent Monte Carlo moves until (1) a Monte Carlo move changes the position of one of the atoms belonging to trimer *ijk*, (2) a new value of $\left|E_L^{(1,ijk)}\right|$ exceeds ε, or (3) an n^{th} accepted MC step is reached and all embedded oligomer energies are calculated without approximations in the current configuration.

4.1.3 Methods

In order to test the accuracy of the EE-3B-SP method, forty-four configurations of $NH_3(H_2O)_{11}$, $(NH_3)_2(H_2O)_{14}$, and $(NH_3)_2(H_2O)_{18}$ were generated by a Metropolis[5] Monte Carlo simulation that was done using the MCCCS (Monte Carlo for Complex Chemical Systems) code developed by the Siepmann group at the University of Minnesota. An example of one of the configurations of each of these systems is shown in Figure 4.1.1. During the simulation, each molecule was kept rigid and the system's potential energy at each step was calculated according to the OPLS[6] force field parameters for ammonia and the SPC/E[7] force field parameters for water (standard Lorentz-Berthelot[8] combining rules were used to obtain parameters for water-ammonia interactions). The only moves allowed during the simulation were translations of less than 0.39 Å for water or of less than 0.88 Å for ammonia and rotations of less than 0.49 radians (23˚) for water or of less than 1.14 radians (65˚) for ammonia. These forty-four configurations of each system include moves that were ultimately rejected by the algorithm and that would not have been included in the final averages used to calculate thermodynamic properties of the system; however, we include these rejected configurations as part of our test suite for the EE-3B-CSP approximation because any method used to calculate single-point energies during an MC simulation must be able to accurately predict which are high-energy, unlikely configurations as well as be able to obtain accurate energies for lower-energy configurations.

The energy of each of the forty-four configurations of each system was calculated using the M06-2X[9] density functional with the cc-pVTZ+[10] basis set in five ways: (a) in the conventional manner (that is, at M06-2X/cc-pVTZ+ without additional

approximations), (b) using the EE-3B approximation with CM4[11] embedding charges (see below for a more detailed explanation of how these embedding charges were obtained), (c) using the EE-3B approximation with CHelpG[12] embedding charges, (d) using the EE-3B-CSP approximation with CM4 embedding charges and the site-site protocol, and (e) using the EE-3B-CSP approximation with CHelpG embedding charges and the site-site protocol. The embedding charges used to represent each water or ammonia molecule in the EE-3B and EE-3B-SP calculations were the CM4 or CHelpG charges obtained from the M06-2X/cc-pVTZ+ density of the gas-phase water or ammonia molecule in its M06-2X/cc-pVTZ+ optimized geometry. The CHelpG method was also used to obtain the partial charges assigned to the atoms of each embedded monomer, dimer, or trimer when the site-site version of the EE-3B-CSP method was tested. All electronic structure calculations were done using *Minnesota Gaussian Functional Module, version 3.0* (*MN-GFM-v3.0*),[13] a locally modified version of the GAUSSIAN 03[14] electronic structure package, revision D.01. The EE-3B calculations were done using a slightly modified version of MBPAC 2007-2,[15] and the EE-3B-CSP calculations were done using a significantly modified version of MBPAC 2007-2.

For all EE-3B-CSP calculations done in this study, a reference calculation (i.e., a complete EE-3B calculation) was performed every fifth accepted step; in other words, $n = 5$ for all EE-3B-CSP calculations. Seven different values for the energy threshold, ε, were tested within the site-site EE-3B-CPS approximation for each configuration of each system used in this study.

4.1.4 Results and Discussion

4.1.4.A The Relative Costs of the EE-PA, EE-3B, and Site-Site EE-3B-CSP Approximations

The goal of the EE-3B and the EE-3B-CSP approximations is to lower the cost of an energy calculation relative to the cost of a conventional (or "full") calculation performed at the same level of theory. In the large-system limit, the EE-3B and EE-3B-CSP approximations will necessarily be less computationally expensive than a conventional M06-2X calculation, because the cost of an M06-2X calculation scales as $O(N^4)$ with system size (where N is the number of monomers in the system) whereas the cost of an EE-3B calculation scales as $O(N^3)$ and the cost of an EE-3B-CSP calculation will have an even more favorable scaling. However, one might need to go to a very large system before seeing the advantage in the EE-3B or EE-3B-CSP calculations. Table 4.1.1 investigates whether or not the EE-3B and EE-3B-CSP calculations are able to save computer time relative to that required for conventional calculations on the systems chosen for this study. Table 4.1.1 shows estimates of the relative central processor unit (CPU) times required to perform the 44-step simulation. The time given for the conventional calculation is the sum of the times taken directly from the *Gaussian* 09 output files and then multiplied by four (because each conventional calculation was performed on four processors), but the EE-MB and EE-3B-CSP calculations are based on the average time required for water monomer, dimer, and trimer calculations at M06-2X/cc-pVTZ+ on a single processor and on the average number of accurate trimer calculations given in Table 4.1.2. Table 4.1.2 shows the average number of trimer energies that had to be recalculated accurately (i.e., without

using site-site first-order perturbation theory) over the last 43 steps. All averages (including those shown in Table 1) were calculated over configurations 2 through 44; i.e., none of the averages include configuration 1.

Table 4.1.1 shows that an ε values of at least 0.15 kcal/mol must be used before the EE-3B-CSP calculations become advantageous over the conventional calculations of $NH_3(H_2O)_{11}$, that an epsilon value of at least 0.20 kcal/mol must be used before the EE-3B-CSP calculations become advantageous over the conventional calculations of $(NH_3)_2(H_2O)_{14}$, and that an epsilon value greater than 1.00 kcal/mol must be used before the EE-3B-CSP calculations become advantageous over the conventional calculations of $(NH_3)_2(H_2O)_{18}$. Table 4.1.1 also shows that the EE-PA method is at most a mere 14% of the cost of the least expensive EE-3B-CSP calculation possible (with $\varepsilon \rightarrow \infty$).

Although the timings of site-orbital calculations have not been tested in this study, one can assume that they will add to the overall cost of an EE-3B-CSP calculation because they require the calculation of one-electron Coulomb interaction integrals rather than a simple finite sum of Coulomb interactions between point charges. This additional cost may make the site-orbital version of EE-3B-CSP approximation impractical for Monte Carlo simulations, because even the site-site version does not save as much computer time as had been hoped originally.

Section 4.1.4.B investigates whether or not the savings in cost afforded by the EE-PA, EE-3B, and site-site EE-3B-CSP approximations is worth the the loss of accuracy that they incur.

4.1.4.B The Accuracy and Efficiency of the EE-PA, EE-3B, and Site-Site EE-3B-CSP Approximations

For a Monte Carlo simulation, the most important quantity to calculate accurately is the energy difference between two configurations of the same system. These energy differences are what determine the course of the MC simulation and dictate which geometries are ultimately included in the final averages that are used to calculate thermodynamic properties of the system. Thus, our analysis of the accuracy of a given approximation method is based on its ability to calculate accurate energies relative to the energy of the starting configuration (step 1) of each of the water–ammonia systems used in this study. The best estimate of the relative energy of each configuration comes from the conventional calculations performed at M06-2X/cc-pVTZ+. The conventionally calculated relative energy values range from -0.52 kcal/mol to 7.24 kcal/mol. The average of the absolute values of the conventionally calculated relative energies is 1.89 kcal/mol, with a standard deviation of 1.79 kcal/mol.

Errors in the EE-PA, EE-3B, and site-site EE-3B-CSP relative energies are calculated with respect to the conventionally calculated M06-2X/cc-pVTZ+ values. Tables 4.1.3, 4.1.4, and 4.1.5 show the mean signed errors (MSE), mean unsigned errors (MUE), and root mean squared errors (RMSE) of the EE-PA, EE-3B, and EE-3B-CSP relative energies with respect to the conventionally calculated relative energies in the three systems selected for this study.

Tables 4.1.3–4.1.5 show that the EE-PA approximation does reasonably well in the prediction of the relative energies of these water–ammonia clusters and that adding

all of the trimer calculations for an EE-3B approximation generally cuts the MUEs down to half of their original value. As expected, these tables show that increasing the value of ε increases the error in the EE-3B-CSP approximation relative to that of the EE-3B approximation. Unexpectedly, however, these tables also show that increasing the value of ε beyond a certain point actually yields MUEs higher than those of the much less expensive EE-PA approximation. It seems that including too many approximate trimer energies in the three-body term in the EE-MB expansion of the total energy is actually worse than neglecting the three-body term entirely. This shows that one must be careful not to use large values of ε. The definition of a "large" value of ε depends on the type of system being studied, but for the water–ammonia systems shown here a large value of ε is defined most conservatively as one that is greater than 0.1 kcal/mol.

Combining information from Tables 4.1.1–4.1.5 shows that for the level of theory used in this work, the EE-3B-CSP approximation is not an efficient way to obtain relative energies of various conformations of moderately-sized molecular clusters. Table 4.1.1 shows that in order reduce the cost of an EE-3B-CSP calculation to less than that of a conventional M06-2X calculation one must use values of ε that are greater than 0.1 kcal/mol in each cluster studied. However, as discussed in the previous paragraph, using values of ε greater than 0.1 kcal/mol can lead to errors worse than those of the much simpler EE-PA approximation. Therefore, for these systems it appears that one cannot find a value of ε that yields an efficient compromise between computational cost and accuracy in the EE-3B-CSP approximation.

4.1.5 Conclusions

The purpose of this study is to assess the cost and accuracy of an approximation method that has been proposed to streamline the electrostatically-embedded many-body method for use in Monte Carlo simulations. The proposed approximation has been dubbed the electrostatically-embedded many-body approximation with consistently screened perturbations, or EE-MB-CSP. One can perform an EE-MB-CSP approximation in two ways: one can apply the site-orbital version of the EE-MB-CSP approximation and use changes in Coulomb interactions between embedding charges and electron orbitals to estimate the change in an oligomer's energy, or one can apply the site-site version of the EE-MB-CSP approximation and use changes in Coulomb interactions between embedding charges and point charges that represent an oligomer's charge density to estimate the change in the oligomer's energy. Because the site-site version of the EE-MB-CSP approximation is less costly and easier to implement, it has been chosen as the subject of this preliminary investigation.

Our findings regarding the site-site version of the EE-MB-CSP approximation are summarized as follows:

- The accuracy of the EE-3B-CSP approximation can be controlled by adjusting the value of the threshold energy screening parameter ε. A smaller value of ε yields more accurate relative energies and a larger value of ε yields less accurate relative energies.
- Using values of ε greater than 0.1 kcal/mol can be dangerous because they can lead to mean unsigned errors that are higher than those of the much less computationally costly EE-PA approximation.

- Using values of epsilon less than or equal to 0.1 kcal/mol does not make the EE-3B-CSP approximations cost less (in terms of computer time) than the conventional M06-2X calculations for the water–ammonia systems selected for this study.

These findings indicate that the site-site version of the EE-3B-CSP approximation is not a cost-effective way to obtain reliable relative energies of various configurations of molecular clusters. Future research might investigate other possibilities for trimming the cost of an EE-3B calculation without degrading its accuracy by a significant amount.

4.1.6 References for Section 4.1

(1) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2007***, 3*, 46.

(2) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J Chem Theory Comput* **2008***, 4*, 33.

(3) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008***, 4*, 683.

(4) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009***, 5*, 1573.

(5) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953***, 21*, 1087.

(6) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999***, 121*, 4827.

(7) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987***, 91*, 6269.

(8) Maitland, G. C.; Rigby, M.; Smith, E. B.; Wakeham, A. In *Intermolecular Forces: Their Origin and Determination;* Pergamon Press: Oxford, 1987.

(9) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008***, 120*, 215.

(10) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009***, 5*, 1197.

(11) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 2046.

(12) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990***, 11*, 361.

(13) Zhao, Y.; Truhlar, D. G. *Minnesota Gaussian Functional Module*, version 3.0; University of Minnesota: Minneapolis, MN, 2007.

(14) *Gaussian 03*, revision D.01; 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.; Pople, J. A.; Gaussian, Inc.: Wallingford, CT, 2004.

(15) Dahlke, E. E.; Truhlar, D. G. *MBPAC*, version 2007-2; University of Minnesota: Minneapolis, MN, 2007.

Table 4.1.1. Relative Times[a] Required for Forty-Four Single-Point Energy Calculations of Each Given System at M06-2X/cc-pVTZ+

	ε	$(NH_3)(H_2O)_{11}$		$(NH_3)_2(H_2O)_{14}$		$(NH_3)_2(H_2O)_{18}$	
Method[b]	(kcal/mol)	ChelpG	CM4	ChelpG	CM4	ChelpG	CM4
Conventional	N/A	1.00	1.00	1.00	1.00	1.00	1.00
EE-PA	0.000	0.08	0.08	0.07	0.07	0.11	0.11
EE-3B	0.000	1.51	1.51	1.82	1.82	3.57	3.57
EE-3B-CSP	0.010	1.39	1.38	1.64	1.61	3.07	3.02
EE-3B-CSP	0.100	1.06	1.01	1.17	1.12	2.14	2.03
EE-3B-CSP	0.150	0.97	0.94	1.07	1.01	1.92	1.81
EE-3B-CSP	0.200	0.92	0.88	0.99	0.95	1.77	1.68
EE-3B-CSP	0.250	0.88	0.85	0.95	0.90	1.67	1.58
EE-3B-CSP	0.500	0.76	0.74	0.80	0.76	1.38	1.32
EE-3B-CSP	1.000	0.69	0.67	0.70	0.67	1.18	1.12
EE-3B-CSP	10^6	0.59	0.59	0.60	0.60	0.96	0.96

[a]Relative time = [time required for the given method] / [time required for the conventional calculation]; relative times are unitless.

[b]For all EE-3B-CSP calculations, $n = 5$.

Table 4.1.2. Average Numbers of Accurately Calculated Trimer Energies Using Different Values of Epsilon Within the EE-3B-CSP Approximation[a] Over Forty-Three Configurations of Each Given System at M06-2X/cc-pVTZ+

	$(NH_3)(H_2O)_{11}$		$(NH_3)_2(H_2O)_{14}$		$(NH_3)_2(H_2O)_{18}$	
ε (kcal/mol)	ChelpG	CM4	ChelpG	CM4	ChelpG	CM4
0.000	220	220	560	560	1140	1140
0.010	199	196	491	482	954	935
0.100	139	131	318	300	608	566
0.150	124	117	280	259	526	486
0.200	114	108	252	237	472	436
0.250	107	101	235	219	434	401
0.500	86	83	181	167	327	304
1.000	73	70	142	133	252	228
10^6	55	55	105	105	171	171

[a]For all EE-3B-CSP calculations, $n = 5$.

Table 4.1.3. MSE, MUE, and RMSE Relative to Conventional M06-2X/cc-pVTZ+ Relative Energies Over Forty-Three Configurations of $(NH_3)(H_2O)_{11}$

	ε	MSE (kcal/mol)		MUE (kcal/mol)		RMSE (kcal/mol)	
Method[a]	(kcal/mol)	ChelpG	CM4	ChelpG	CM4	ChelpG	CM4
EE-PA	0.000	0.021	-0.062	0.067	0.068	0.079	0.088
EE-3B	0.000	-0.042	-0.023	0.043	0.023	0.047	0.029
EE-3B-CSP	0.010	-0.044	-0.024	0.044	0.024	0.048	0.030
EE-3B-CSP	0.100	-0.052	-0.032	0.061	0.045	0.074	0.058
EE-3B-CSP	0.150	-0.061	-0.035	0.071	0.052	0.086	0.066
EE-3B-CSP	0.200	-0.058	-0.032	0.077	0.056	0.095	0.071
EE-3B-CSP	0.250	-0.057	-0.039	0.079	0.064	0.098	0.081
EE-3B-CSP	0.500	-0.058	-0.047	0.103	0.092	0.136	0.125
EE-3B-CSP	1.000	-0.066	-0.046	0.123	0.115	0.162	0.144
EE-3B-CSP	10^6	-0.022	-0.011	0.199	0.164	0.260	0.215

[a]For all EE-3B-CSP calculations, $n = 5$.

Table 4.1.4. MSE, MUE, and RMSE Relative to Conventional M06-2X/cc-pVTZ+ Relative Energies Over Forty-Three Configurations of $(NH_3)_2(H_2O)_{14}$

	ε	MSE (kcal/mol)		MUE (kcal/mol)		RMSE (kcal/mol)	
Method[a]	(kcal/mol)	ChelpG	CM4	ChelpG	CM4	ChelpG	CM4
EE-PA	0.000	0.000	-0.063	0.080	0.101	0.102	0.129
EE-3B	0.000	-0.044	-0.049	0.052	0.054	0.058	0.061
EE-3B-CSP	0.010	-0.047	-0.049	0.056	0.056	0.063	0.063
EE-3B-CSP	0.100	-0.038	-0.045	0.064	0.064	0.073	0.070
EE-3B-CSP	0.150	-0.042	-0.052	0.078	0.081	0.090	0.091
EE-3B-CSP	0.200	-0.055	-0.054	0.101	0.095	0.121	0.111
EE-3B-CSP	0.250	-0.049	-0.058	0.110	0.103	0.136	0.118
EE-3B-CSP	0.500	-0.072	-0.081	0.161	0.148	0.222	0.208
EE-3B-CSP	1.000	-0.063	-0.065	0.205	0.184	0.297	0.253
EE-3B-CSP	10^6	-0.121	-0.103	0.386	0.313	0.552	0.440

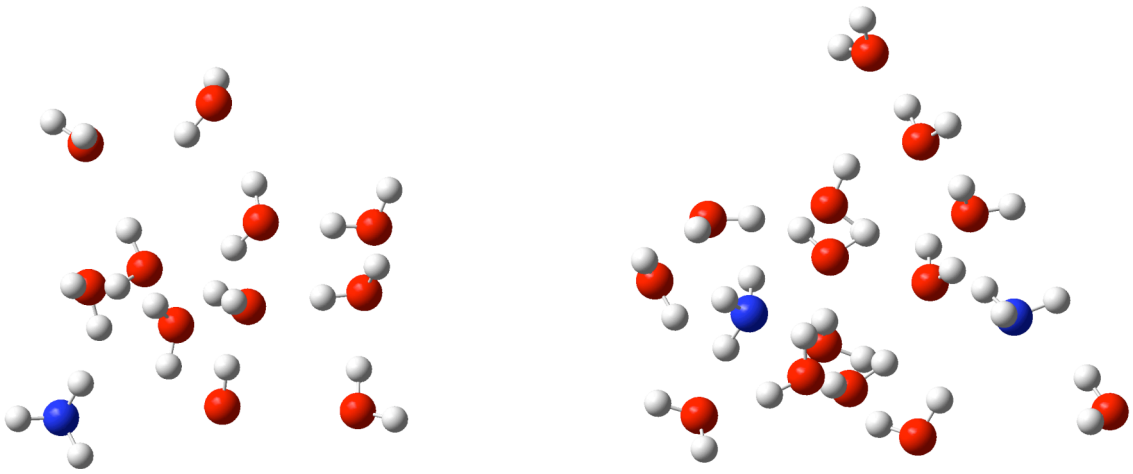
[a]For all EE-3B-CSP calculations, $n = 5$.

Table 4.1.5. MSE, MUE, and RMSE Relative to Conventional M06-2X/cc-pVTZ+ Relative Energies Over Forty-Three Configurations of $(NH_3)_2(H_2O)_{18}$

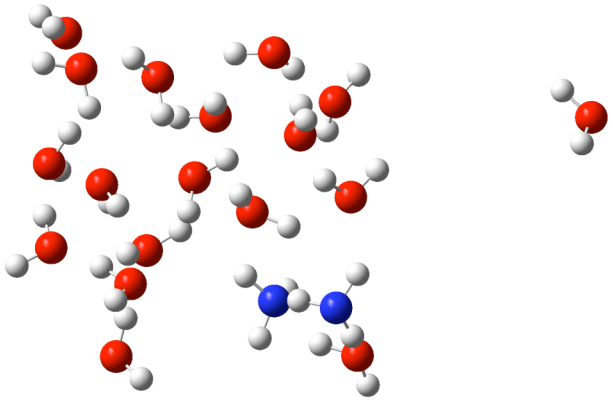
	ε	MSE (kcal/mol)		MUE (kcal/mol)		RMSE (kcal/mol)	
Method[a]	(kcal/mol)	ChelpG	CM4	ChelpG	CM4	ChelpG	CM4
EE-PA	0.000	0.124	0.190	0.136	0.204	0.164	0.245
EE-3B	0.000	-0.101	-0.091	0.101	0.091	0.111	0.101
EE-3B-CSP	0.010	-0.101	-0.091	0.102	0.091	0.113	0.101
EE-3B-CSP	0.100	-0.101	-0.088	0.127	0.111	0.143	0.129
EE-3B-CSP	0.150	-0.101	-0.091	0.140	0.127	0.157	0.145
EE-3B-CSP	0.200	-0.098	-0.087	0.147	0.133	0.163	0.153
EE-3B-CSP	0.250	-0.097	-0.106	0.159	0.146	0.178	0.177
EE-3B-CSP	0.500	-0.119	-0.104	0.211	0.184	0.281	0.251
EE-3B-CSP	1.000	-0.090	-0.078	0.225	0.228	0.308	0.321
EE-3B-CSP	10^6	0.068	0.026	0.446	0.367	0.676	0.550

[a]For all EE-3B-CSP calculations, $n = 5$.

Figure 4.1.1. (a) $NH_3(H_2O)_{11}$, (b) $(NH_3)_2(H_2O)_{14}$, and (c) $(NH_3)_2(H_2O)_{18}$ at step 1 of the Monte Carlo simulations used for this study. In cluster (c), the water molecule that appears slightly outside the cluster is actually ~90 Å away from the center of the cluster.



(c) (b)



(c)

Figure 4.1.1

4.2 The Electrostatically-Embedded Pairwise-Additive Specific Reaction Parameter Approximation for Monte Carlo Simulations[4]

4.2.1 Introduction

In order to predict the properties of chemical mixtures and compounds, one would ideally (a) use a level of theory wherein the quantum mechanical (QM) nature of the electrons is considered, and (b) take averages over properties of many configurations of the system of interest in order to predict experimentally observable quantities. Although a variety of levels of theory exist that satisfy (a) and molecular dynamics (MD) or Monte Carlo (MC) techniques can be used to adequately sample configuration space and therefore satisfy (b), the difficulty arises when one attempts to satisfy both criteria simultaneously. To use one of the existing levels of electronic structure theory [such as Hartree Fock (HF), coupled-cluster theory, or even density functional theory (DFT)] to perform a QM potential energy calculation on a single configuration of a system containing more than 100 heavy atoms requires more computational effort than most scientists have available. Clearly, using any of these levels of theory without additional approximations to perform the 10^5 to 10^7 potential energy calculations required to obtain a free energy profile of a large system is presently out of the question.

The electrostatically-embedded many-body (EE-MB) approximation[1-5] (described in more detail in Section 4.2.4 and the cited sources) attempts to reduce the cost of a QM potential energy calculation by using any desired level of electronic

---

[4] The authors would like to thank the Minnesota Supercomputing Institute for computer time.

structure theory to perform QM potential energy calculations on individual fragments (or monomers) and combinations of fragments, but never on the entire system at once. A linear combination of the potential energy of each fragment and of each combination of fragments is then used to approximate the potential energy of the entire system. In terms of total computational cost (which considers both the time and the number of processors required to complete a calculation), only the pairwise-additive version of the EE-MB approximation (i.e., the EE-PA approximation) is truly advantageous over the least expensive density functionals: if N is the number of monomers in the system, the EE-PA method ultimately scales as N^2 whereas the three-body version of EE-MB (i.e., EE-3B) and the least costly density functionals scale as N^3. However, previous studies[1-5] have shown that only the EE-3B approximation is consistently capable of yielding energies that come within 1% of a full-system calculation done at the same level of electronic structure theory, whereas the EE-PA method does not consistently perform that well. Current research, however, has indicated that the EE-PA energies may be more sensitive to the values assigned to the embedding charges (i.e., the partial charges that are used to represent atoms that are not treated quantum mechanically in a given fragment energy calculation) than are the EE-3B energies. One could potentially take advantage of this trait by finding a set of embedding charges that minimizes the mean unsigned error (MUE) of a representative collection of small clusters and then using those embedding charges to perform EE-PA calculations on larger systems that comprise the same kinds of monomers that appear in the smaller clusters. In this way, the embedding charges are treated as a sort of specific reaction parameter (SRP), and so we call this method the "EE-PA-SRP" approximation. The fundamental assumption of

the EE-PA-SRP approximation is that the embedding charges that minimize the absolute errors in one system (or a few systems) are likely to minimize the errors in other chemically similar systems. The present study investigates whether or not the fundamental assumption of the EE-PA-SRP approximation holds for systems containing water and ammonia molecules, which are of interest to researchers studying atmospheric nucleation.[6]

4.2.2 Theory

The electrostatically-embedded many-body (EE-MB) approximation is based on the many-body expansion of a system's total potential energy (E):

$$E = V_1 + V_2 + V_3 + ... + V_N \tag{1}$$

where N is the total number of monomers into which the system has been divided. In the present article, the monomers are not bonded to one another, but they interact noncovalently. If E_i denotes the energy of monomer i in the geometry that it has in the given configuration of the system, E_{ij} denotes the energy of dimer ij (that is, the dimer composed of monomers i and j), and E_{ijk} denotes the energy of trimer ijk, then the first three terms of eq 1 may be written:

$$V_1 = \sum_{i=1}^{N} E_i \tag{2}$$

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

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

If one uses the sum of only the first two terms of the many-body expansion to approximate *E*, then one has made the two-body (2B) or pairwise-additive (PA) approximation; if one uses the sum of the first three terms to approximate *E*, then one has made the three-body (3B) approximation. In order to make the electrostatically-embedded many-body (EE-MB) approximation, one defines E_i, E_{ij}, or E_{ijk} as the energy of a monomer, dimer, or trimer in the geometry that it has in the given configuration and embedded in a field of partial point charges located at the coordinates of the nuclei of the atoms that do not belong to the specific *m*-mer in question. Including the embedding charges in the calculation of a given oligomer's energy adds almost nothing to its computational cost, but can significantly speed the convergence of the many-body series to the correct answer, as higher-order effects are partially included in lower-order terms. The EE-PA energy ($E_{\mathrm{EE-PA}}$) is in general much more accurate than the PA energy (E_{PA}) but worse than the straight 3B energy (E_{3B}) when each is compared to the conventional calculation of a system's potential energy at some given level of electronic structure theory;[1-5] the EE-3B approximation, however, consistently produces excellent accuracy,[1-5] often yielding errors that are less than 1% of the total intermolecular interaction energy of a given system.

Although the accuracy of the EE-3B approximation is generally excellent, its cost may still be too great for the thousands of energy calculations required to create a free energy profile of the system (for example, a plot that shows how the free energy of a cluster is related to the number of molecules in the cluster). The EE-PA approximation, however, is significantly less costly than the EE-3B method, and could be plausible for use in MC simulations of small-molecule systems. The EE-PA

approximation requires that QM energy calculations be performed only on the monomers and dimers of the system, whereas the EE-3B approximation requires the trimer calculations as well. The number of dimers present in a system containing N monomers is only $\frac{N(N-1)}{2}$, or roughly $\frac{N^2}{2}$, whereas the number of trimer calculations is $\frac{N(N-1)(N-2)}{6}$, or roughly $\frac{N^3}{6}$. Clearly, the bulk of the expense of an EE-3B calculation lies in the trimer calculations, not only because the trimers are larger and contain more electronic degrees of freedom, but also because the number of trimers ultimately grows much more rapidly with system size than the number of monomers and dimers combined.

Past studies[1, 4, 5] have indicated that the EE-3B energy is not heavily dependent on the chosen values of the embedding charges. Typically, embedding charges are assigned based on results from charge analyses performed on the gas-phase monomers that compose the system. Although different methods of charge analysis often do assign different partial charges to the same atom in a given system, they often do not differ from one another by more than 0.3 *e*. Thus, although the past studies demonstrated that the EE-PA energies were a little more sensitive to the set of embedding charges chosen, only a relatively small range of embedding charges was sampled. The present study investigates (a) whether or not the EE-PA calculations can yield results that are as accurate as those of the EE-3B approximation by allowing the embedding charges to take on values outside of the range typically considered physically realistic for a given type of monomer, and (b) whether or not the set of embedding charges that minimizes the EE-PA error for one system will also minimize

(or come close to minimizing) the error for other chemically similar systems. We define the error as

$$Error(X) = E(X) - E(\text{Conv}) \qquad (5)$$

where X = PA, 3B, EE-PA, or EE-3B, $E(X)$ = the electronic energy of the given configuration as calculated within approximation method X, and E(Conv) = the electronic energy of the given configuration as calculated in the conventional manner.

4.2.3 Methods

In order to observe the dependence of the EE-PA and EE-3B energies on a wide range of values assigned to embedding charges in a two-component system, a series of single-point energy calculations were performed on four different configurations of $(NH_3)_2(H_2O)_{18}$ with the M06-2X[7] density functional and the cc-pVTZ+[8] basis set. In addition to the conventional M06-2X/cc-pVTZ+ calculation [throughout the paper a "level of theory" will be labeled *X/Y*, where *X* denotes a level of electronic structure theory and *Y* denotes a basis set] the EE-PA and EE-3B energy of each configuration was calculated several times, each time with a systematically varied set of embedding charges. The first EE-MB calculation for each structure was performed using the kind of embedding charges that would typically be recommended; that is, the embedding charges are obtained from a charge analysis of the gas-phase monomer that they are intended to represent. In this particular study, the ChelpG[9] charges of the M06-2X/cc-pVTZ+ optimized gas-phase water and ammonia molecules were used, and their specific values are given in Table 4.2.1. The embedding charge used to represent the oxygen atom of a neutral water molecule (q_O) was then varied in increments of ±0.05 *e* from its starting ChelpG value. In order to maintain the neutrality of the

embedding charge representation of each water molecule, the charge on the hydrogen atoms of the water molecule (q_H) was also varied according to the following formula:

$$q_H = -\frac{q_O}{2} \qquad (6)$$

The embedding charges used to represent the ammonia molecules were left at their original ChelpG values for each EE-MB calculation.

The four configurations of $(NH_3)_2(H_2O)_{18}$ are labeled A, B, C, and D, and are shown in Figure 4.2.1. None of these configurations is a stationary point on the $(NH_3)_2(H_2O)_{18}$ potential energy surface (PES). Configuration A was taken from a brief Monte Carlo simulation that was done using a molecular mechanics force field wherein all molecules were held rigid, and so all of the water molecules in configuration A share the same geometry, as do the ammonia molecules. Configurations B through D are based three $(H_2O)_{20}$ clusters that are random segments of an $(H_2O)_{64}$ cluster. Two of the water molecules in each of the $(H_2O)_{20}$ clusters were replaced with ammonia molecules, and several steps of a geometry optimization were performed in order to relax the ammonia molecules, but none of the geometry optimizations reached completion.

In order to determine whether or not the set of embedding charges that minimizes the EE-PA absolute errors in a smaller system will also minimize (or nearly minimize) the EE-PA absolute errors in a larger but chemically similar system, five water hexamers were subject to the same calculations as the $(NH_3)_2(H_2O)_{18}$ clusters. These five hexamers are labeled “boat”, “book”, “cage”, “prism”, and “ring”; they are

shown in Figure 4.2.2. The five water hexamers *are* stationary points on an MP2[10]/DH(*d,p*)[11] PES, as calculated in Reference 12.

All electronic structure calculations were performed using *Minnesota Gaussian Functional Module, version 3.0* (*MN-GFM-v3.0*),[13] a locally modified version of the GAUSSIAN 03[14] electronic structure package, revision D.01. The EE-MB calculations were done using MBPAC 2009.[15]

4.2.4 Results and Discussion

Tables 4.2.2 and 4.2.3 show the errors (calculated according to eq 5) in the EE-PA and EE-3B approximations of the energies of the four configurations of $(NH_3)_2(H_2O)_{18}$ when q_O is varied in 0.05 *e* increments from –0.476 to –0.976 *e*. Even over this wider range of q_O values, the error in the EE-3B energy (Table 4.2.3) remains almost constant for each configuration, although the absolute errors associated with more negative q_O values are generally lower than the absolute errors associated with more positive q_O values. The same trend is not evident for the EE-PA errors. First, the EE-PA errors are clearly more sensitive to the q_O values than are the EE-3B errors. Second, for these four configurations, the absolute values of the EE-PA errors reach clear minima within the range of q_O values tested, whereas it appears that the EE-3B absolute errors would reach a minimum value somewhere outside of the range tested. In fact, for all four configurations, the value of q_O that minimizes the EE-PA error lies between –0.576 and –0.726 *e*. Although the minimum falls at a slightly different q_O for each configuration, if we were to pick just one of the q_O values in that range and use it for all four configurations, the worst absolute error that we would incur in the EE-PA

energies would be 1.53 kcal/mol, which is still less than the worst absolute error incurred in the EE-3B energy when the original ChelpG value of q_O is used (1.86 kcal/mol). Finally, the minimum value of the absolute error in the EE-PA method is 0 kcal/mol for each configuration, whereas it is not clear from the tables whether or not the minimum absolute error in the EE-3B method could ever be made to reach 0 kcal/mol simply by varying q_O.

The results shown in Table 4.2.2 have both positive and negative implications. On the positive side, the EE-PA absolute errors can essentially be reduced to 0 kcal/mol if the correct set of embedding charges is chosen. In such a case, the inexpensive EE-PA calculations could certainly outperform the EE-3B calculations in terms of both cost and accuracy! On the negative side, the EE-PA errors are more sensitive to q_O than are the EE-3B errors, so if one does not use a set of embedding charges that happens to fall in the relatively small optimal region (which appears to be between –0.576 and –0.726 *e* for the four configurations tested), the EE-PA errors will be too large for the method to be useful. Thus, one hopes that the optimal range of q_O for these larger systems will match the optimal range of q_O for smaller systems, on which it would be more feasible to parameterize the embedding charges. For example, if the system that we wanted to study were $(NH_3)_2(H_2O)_{18}$, a simple system that we might choose to parameterize the embedding charges (in this case, just q_O and, concomitantly, q_H) could be a water hexamer. We would then hope that the value of q_O that minimizes the EE-PA absolute error for the hexamer would fall somewhere between -0.576 and -0.726 *e*. Tables 4.2.4 and 4.2.5 show how the errors in the EE-PA and

EE-3B energies of five water hexamers correspond to the values chosen for q_O. Again, we see that the EE-3B errors remain fairly insensitive to the q_O, although, as with $(NH_3)_2(H_2O)_{18}$, more negative values of q_O tend to yield slightly lower absolute errors in the EE-3B energies. Unfortunately, the similarities between the trends in the $(NH_3)_2(H_2O)_{18}$ errors and the trends in the $(H_2O)_6$ errors end here. For the $(NH_3)_2(H_2O)_{18}$ clusters, the q_O value that minimized the error in each configuration was always less negative (i.e., smaller in magnitude) than the ChelpG charge that would normally be used (–0.726 *e*). For the $(H_2O)_6$ clusters, the error-minimizing q_O is always more negative (i.e., greater in magnitude) than the ChelpG charge. For the EE-PA-SRP approximation, we would have concluded that the best q_O would be –0.926 *e*, because this is the value that minimizes the mean unsigned error in the EE-PA approximation over all five water hexamers. We then would have used this q_O to obtain EE-PA energies for the $(NH_3)_2(H_2O)_{18}$ clusters. In so doing, we would have unwittingly incurred absolute errors between 1.19 and 3.96 kcal/mol and would be worse off than if we had simply stuck with the original ChelpG charges, in which case we would have incurred absolute errors of at most 1.53 kcal/mol.

4.2.5 Conclusions

The EE-3B approximation is generally capable of yielding energies that are within 1% of a conventional QM energy calculation on the entire system, but, unfortunately, the cumulative cost of the large number of trimer calculations that are usually required could prevent the EE-3B method from being affordable for MC simulations on most interesting systems. The EE-PA approximation, on the other hand,

could potentially be affordable for MC simulations on a larger variety of systems, but it is not consistently capable of yielding a reliable level of accuracy. The present study has observed that EE-PA energies are significantly more sensitive to the choice of embedding charges than are the EE-3B energies, and that the EE-PA errors will essentially vanish with a judicious (or perhaps fortuitous) choice of embedding charges. This leads to the concept of the EE-PA-SRP method, where the embedding charges are treated as parameters that may be adjusted in order to nearly eliminate the EE-PA error. However, this study has indicated that attempting to parameterize the embedding charges on small systems for use in studies of larger (but chemically similar) systems is highly risky and could even end up being detrimental. Therefore, the EE-PA-SRP method is not recommended as a way to reduce the cost of single-point energy calculations.

4.2.6 References for Section 4.2

(1) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2007***, 3*, 46.

(2) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comp.* **2007***, 3*, 1342.

(3) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J Chem Theory Comput* **2008***, 4*, 33.

(4) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008***, 4*, 683.

(5) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009***, 5*, 1573.

(6) Kulmala, M. *Science* **2003***, 302*, 1000.

(7) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008***, 120*, 215.

(8) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009***, 5*, 1197.

(9) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990***, 11*, 361.

(10) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934***, 46*, 618.

(11) Dunning, T. H. J.; Hay, P. J. In *Methods of Electronic Structure Theory;* Plenum: New York, 1977.

(12) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000***, 112*, 2063.

(13) Zhao, Y.; Truhlar, D. G. *Minnesota Gaussian Functional Module*, version 3.0; University of Minnesota: Minneapolis, MN, 2007.

(14) *Gaussian 03*, revision D.01; 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.; Pople, J. A.; Gaussian, Inc.: Wallingford, CT, 2004.

(15) Dahlke, E. E.; Lin, H.; Leverentz, H. R.; Truhlar, D. G. *MBPAC*, version 2009; University of Minnesota: Minneapolis, MN, 2009.

Table 4.2.1. ChelpG Charges Assigned to the M06-2X/cc-pVTZ+ Optimized Gas-Phase Water and Ammonia Molecules

Molecule	Atom	ChelpG Charge
water	O	-0.726
water	H	0.363
ammonia	N	-0.925
ammonia	H	0.308

Table 4.2.2. Errors (kcal/mol) in the M06-2X/cc-pVTZ+ EE-PA Energies of Configurations A to D as a Function of the Embedding Charges Assigned to the Oxygen Atoms of Water Monomers[a]

q_O (in *e*)	A	B	C	D
-0.976	-2.50	-4.43	-1.98	-1.40
-0.926	-2.30	-3.96	-1.72	-1.19
-0.876	-2.07	-3.44	-1.43	-0.94
-0.826	-1.80	-2.86	-1.12	-0.66
-0.776	-1.50	-2.22	-0.79	-0.35
-0.726	-1.16	-1.53	-0.43	0.00
-0.676	-0.79	-0.77	-0.04	0.37
-0.626	-0.39	0.04	0.37	0.78
-0.576	0.05	0.92	0.81	1.23
-0.526	0.53	1.85	1.28	1.70
-0.476	1.04	2.85	1.77	2.21

[a]Light gray highlighting indicates the errors when q_O is the ChelpG charge assigned to oxygen in the gas-phase water molecule at M06-2X/cc-pVTZ+; dark gray highlighting indicates the minimum absolute error value for each configuration.

Table 4.2.3. Errors in the M06-2X/cc-pVTZ+ EE-3B Energies of Configurations A to D as a Function of the Embedding Charges Assigned to the Oxygen Atoms of Water Monomers[a]

q_O (in e)	A	B	C	D
-0.976	1.70	0.72	0.88	0.52
-0.926	1.73	0.74	0.90	0.53
-0.876	1.76	0.76	0.92	0.55
-0.826	1.79	0.79	0.94	0.57
-0.776	1.82	0.82	0.95	0.60
-0.726	1.86	0.85	0.97	0.62
-0.676	1.89	0.88	1.00	0.65
-0.626	1.93	0.91	1.02	0.69
-0.576	1.96	0.94	1.04	0.72
-0.526	2.00	0.98	1.06	0.76
-0.476	2.04	1.01	1.08	0.80

[a]Light gray highlighting indicates the errors when q_O is the ChelpG charge assigned to oxygen in the gas-phase water molecule at M06-2X/cc-pVTZ+; dark gray highlighting indicates the minimum absolute error value for each configuration.

Table 4.2.4. Errors (kcal/mol) in the M06-2X/cc-pVTZ+ EE-PA Energies of Water Hexamers as a Function of the Embedding Charges Assigned to the Oxygen Atoms of Water Monomers[a]

q_O (in e)	boat	book	cage	prism	ring
-0.976	-0.22	-0.45	-0.02	-0.20	-0.09
-0.926	0.04	-0.19	0.22	0.04	0.17
-0.876	0.34	0.11	0.49	0.30	0.48
-0.826	0.69	0.44	0.78	0.60	0.84
-0.776	1.08	0.81	1.10	0.91	1.23
-0.726					

[a]Light gray highlighting indicates the errors when q_O is the ChelpG charge assigned to oxygen in the gas-phase water molecule at M06-2X/cc-pVTZ+; dark gray highlighting indicates the minimum absolute error value for each configuration.

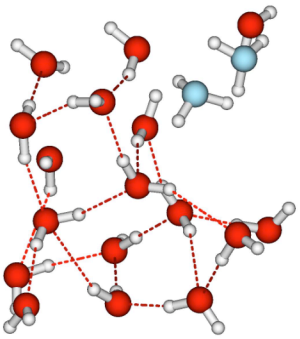
Table 4.2.5. Errors (kcal/mol) in the M06-2X/cc-pVTZ+ EE-3B Energies of Water Hexamers as a Function of the Embedding Charges Assigned to the Oxygen Atoms of Water Monomers[a]

q_O (in e)	boat	book	cage	prism	ring
-0.976	0.15	0.26	0.17	0.44	0.17
-0.926	0.19	0.29	0.20	0.47	0.22
-0.876	0.24	0.33	0.23	0.51	0.27
-0.826	0.29	0.37	0.26	0.55	0.32
-0.776	0.35	0.42	0.29	0.59	0.39
-0.726					

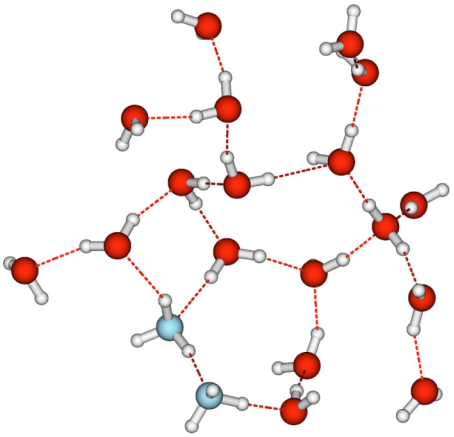
[a]Light gray highlighting indicates the errors when q_O is the ChelpG charge assigned to oxygen in the gas-phase water molecule at M06-2X/cc-pVTZ+; dark gray highlighting indicates the minimum absolute error value for each configuration.

Figure 4.2.1. Four configurations of $(NH_3)_2(H_2O)_{18}$.

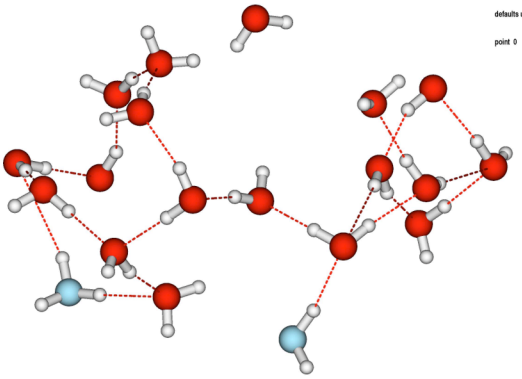
Figure 4.2.2. Five configurations of $(H_2O)_6$.



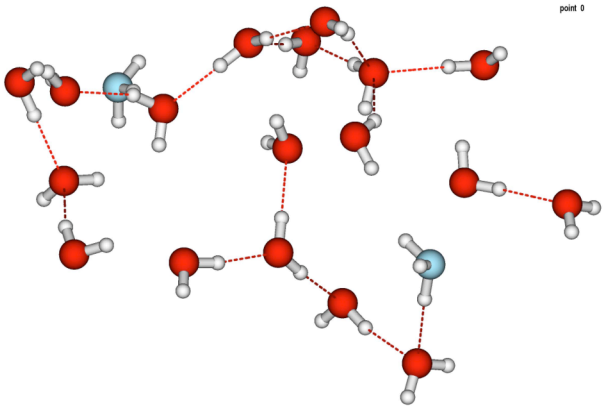
A



B

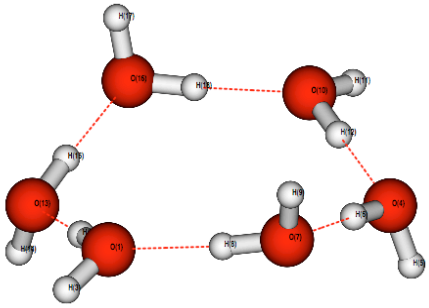


C

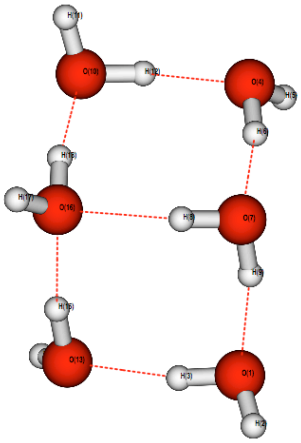


D

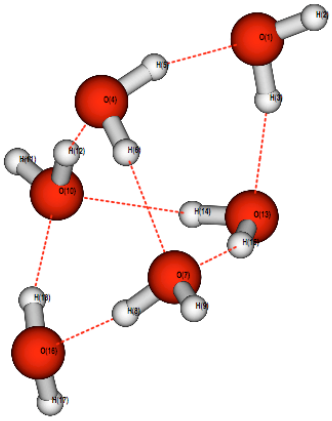
Figure 4.2.1



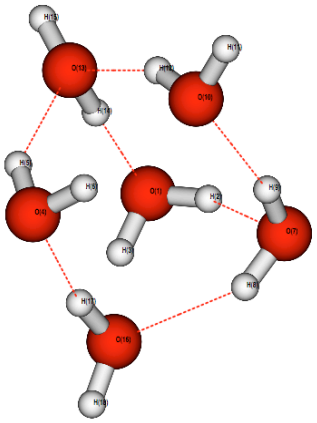
boat



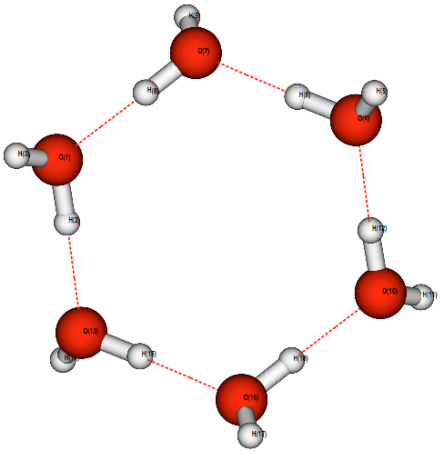
book



cage



prism



ring

Figure 4.2.2

Chapter 5. Electrostatically Embedded Many-Body Method for Dipole Moments, Partial Atomic Charges, and Charge Transfer[5]

5.1. Introduction

Fragment methods constitute one of the most powerful methods for modeling large and complex chemical systems.[1–33] To date, most fragment models have been developed for the purpose of calculating energies. However there is also a need to calculate other properties, such as the charge distribution. In this article we consider how well the electrostatically embedded many-body (EE-MB) approximation[7,11] can predict charge distributions in extended clusters, in particular cluster dipole moments, dipole moments of individual molecules in clusters, partial atomic charges, and charge transfer. The inclusion of charge transfer in our study is of particular interest because it is usually considered a shortcoming of fragment methods that they do not include charge transfer between fragments (for a recent study showing one way to overcome this restriction, see the work of Isegawa et al.[12]).

There has been considerable interest in devising quantum mechanical methods for calculating molecular dipole moments in condensed-phase materials. For example, much work has been done using maximally localized Wannier functions (MLFW) to estimate dipole moments of individual molecules, charge distributions, and bond orders in aqueous solutions and crystalline solids with defects.[34–46] These methods require the user to perform an electronic structure calculation using a plane wave basis set; the final molecular orbitals obtained in the plane wave basis are then transformed into localized

---

[5] This work was supported in part by the National Science Foundation under Grant No. CHE1051396.

functions that can be assigned to a specific molecule or fragment and used to calculate molecular dipole moments and charge distributions. The computational cost of MLWF-based calculations therefore scales asymptotically with system size in the same way that the chosen electronic structure method does. The difficulties of calculating partial atomic charges in condensed phases have also been recently discussed.[47–48]

In this work, we investigate the ability of the EE-MB method to predict charge distributions in molecular clusters. We examine several aspects:: (i) modeling the full-system system dipole moments of clusters directly (that is, without using the concept of partial atomic charges charges), (ii) modeling partial atomic charges, (iii) using the partial atomic charges to calculate dipole moments of individual molecules in a cluster, and (iv) using the partial atomic charges to calculate net charges on individual molecules in a cluster. We note that we are working with concepts (partial atomic charges in ii, subsystem dipole moments in iii, and subsystem net charges in iv) that are not uniquely defined by quantum mechanics, but they are broadly used modeling constructs whose meaning is rather intuitively obvious and is widely understood.

The EE-MB calcuations have the advantage that the asymptotic scaling of their computational cost with system size ultimately is not dependent on the chosen level of electronic structure theory but rather on the level to which one chooses to take the EE-MB expansion.[7] Also, the EE-MB approximation is relatively easy to implement because it does not require minimization procedures and matrix transformations of molecular orbitals; rather, the EE-MB approximation requires only linear combinations of fragment properties. Therefore, it is worthwhile to investigate the ability of various

levels of the EE-MB approximation to predict charge distributions in noncovalently interacting molecular clusters.

5.2. Theory

The EE-MB method has been formulated for predicting energies of an entire system from the energies of its constituent monomers, dimers, and trimers, each embedded in the electrostatic field of the other monomers.[7] The monomers of which the dimers and trimers are composed are nonbonded fragments, e.g., water molecules in a water cluster or water liquid. Here we develop the EE-MB approximations to molecular dipole moments and partial atomic charges. From the latter we can also calculate interfragment charge transfer.

5.2.1. Dipole moments

The EE-MB approximation to the magnitude of the full-system dipole moment (μ) is calculated as follows: (1) the x, y, and z components of the dipole moment are calculated from the electron densities of the various electrostatically embedded oligomers (monomers, dimers, or trimers) in the system, (2) each of the three components of the full-system dipole moment (μ_p, where $p = x$, y, or z) is calculated as a linear combination of the corresponding components of the oligomer dipoles (the linear combination used is the same as that used for the many-body approximation of the system's potential energy), and (3) the magnitude of the vector resulting from step (2) is taken to be μ. Let N be the total number of monomers into which the system has been divided. Let p be set equal to x, y, or z, let μ_p^i indicate the p-component of the dipole moment vector of electrostatically embedded monomer i, let μ_p^{ij} indicate the

p-component of the dipole moment vector of electrostatically embedded dimer ij, and let μ_p^{ijk} indicate the p-component of the dipole moment vector of electrostatically embedded trimer ijk. Let us define $\mu_p^{(1)}$, $\mu_p^{(2)}$, and $\mu_p^{(3)}$ according to eqs 1a, 1b, and 1c, respectively:

$$\mu_p^{(1)} \equiv \sum_{i=1}^{N} \mu_p^i \tag{1a}$$

$$\mu_p^{(2)} \equiv \sum_{i<j} \left(\mu_p^{ij} - \mu_p^i - \mu_p^j \right) \tag{1b}$$

$$\begin{aligned} \mu_p^{(3)} \equiv \sum_{i<j<k} \Big[& \mu_p^{ijk} - \left(\mu_p^{ij} - \mu_p^i - \mu_p^j \right) - \left(\mu_p^{ik} - \mu_p^i - \mu_p^k \right) \\ & - \left(\mu_p^{jk} - \mu_p^j - \mu_p^k \right) - \mu_p^i - \mu_p^j - \mu_p^k \Big] \end{aligned} \tag{1c}$$

Then we may write the electrostatically embedded one-body (EE-1B), two-body (or pairwise additive, EE-PA), and three body (EE-3B) approximations to μ_p (the p-component of the full-system dipole moment vector) according to eqs 2a, 2b, and 2c, respectively:

$$\mu_p \approx \mu_p^{\text{EE-1B}} = \mu_p^{(1)} \tag{2a}$$

$$\mu_p \approx \mu_p^{\text{EE-PA}} = \mu_p^{(1)} + \mu_p^{(2)} \tag{2b}$$

$$\mu_p \approx \mu_p^{\text{EE-3B}} = \mu_p^{(1)} + \mu_p^{(2)} + \mu_p^{(3)} \tag{2c}$$

Finally, we can approximate the magnitude of the full-system dipole moment vector, μ, using the EE-1B, EE-PA, or EE-3B approximation according to eq 3a, 3b, or 3c, respectively:

$$\mu \approx \mu^{\text{EE-1B}} = \sqrt{\left[\left(\mu_x^{\text{EE-1B}}\right)^2 + \left(\mu_y^{\text{EE-1B}}\right)^2 + \left(\mu_z^{\text{EE-1B}}\right)^2\right]} \tag{3a}$$

$$\mu \approx \mu^{\text{EE-PA}} = \sqrt{\left[\left(\mu_x^{\text{EE-PA}}\right)^2 + \left(\mu_y^{\text{EE-PA}}\right)^2 + \left(\mu_z^{\text{EE-PA}}\right)^2\right]} \tag{3b}$$

$$\mu \approx \mu^{\text{EE-3B}} = \sqrt{\left[\left(\mu_x^{\text{EE-3B}}\right)^2 + \left(\mu_y^{\text{EE-3B}}\right)^2 + \left(\mu_z^{\text{EE-3B}}\right)^2\right]} \tag{3c}$$

The origin and orientation of each oligomer are kept the same as its origin and orientation in the full cluster to ensure that the EE-MB approximation of the total dipole moment vector retains its intended physical meaning even for charged clusters.

One can also find the "pure" many-body (MB) approximations to the full-system dipole moment using eqs 1–3 without electrostatic embedding; in this case one uses the components of the dipole moments of monomers, dimers, and trimers that are not surrounded by some representation of the electrostatic potential of the rest of the system.

5.2.2. Point Charges

Assume that we have a system that has been divided into N monomers. If we had the wave function or electron density of the entire system, we could use one of several available charge analysis methods (such as Mulliken population analysis,[49] CHelpG electrostatic potential fitting,[50] or Natural Population Analysis[51]) to assign a

partial charge to each atom in the system. This set of partial charges could in turn be used as a rough approximation of the entire system's electron density. Let us now assume that we do not wish to expend the computational resources necessary to calculate the electron density of the entire system directly but that we do have enough computational resources to calculate the wave functions of the individual monomers, dimers, and trimers in the system. Our objective is to show here that from the wave functions of the electrostatically embedded monomers, dimers, and trimers we may obtain a set of partial charges that represents the electron density of the entire system. As an example, let us focus on the partial charge that we would like to assign to atom A; call this partial charge q_A. Let us say that atom A belongs to monomer i; that is, $A \in i$. We may then approximate q_A (the charge on atom A from the electron density of the entire system) according to the EE-1B, EE-PA, or EE-3B approximations according to equations 4, 5, and 6, respectively. In the EE-1B approximation we have

$$q_A \approx q_A^{\text{EE-1}B} = q_A^i \tag{4}$$

where q_A^i is the partial charge assigned to atom A from the wave function or electron density of electrostatically embedded monomer *i.* In the EE-PA approximation we have

$$q_A \approx q_A^{\text{EE}-\text{PA}} = q_A^i + \sum_{j \neq i} \left(q_A^{ij} - q_A^i \right) \tag{5a}$$

$$= \sum_{j \neq i} q_A^{ij} - (N-2) q_A^i \tag{5b}$$

where q_A^{ij} is the partial charge assigned to atom A from the wave function or electron density of electrostatically embedded dimer ij (which is composed of monomers i and j). The EE-3B approximation yields

$$q_A \approx q_A^{\mathrm{EE-3B}} = q_A^{\mathrm{EE-PA}} + \sum_{j\neq i}^{N} \sum_{\substack{k\neq i \\ k>j}}^{N} \left[q_A^{ijk} - \left(q_A^{ij} - q_A^{i} \right) - \left(q_A^{ik} - q_A^{i} \right) - q_A^{i} \right] \tag{6a}$$

$$= \sum_{j\neq i}^{N} \sum_{\substack{k\neq i \\ k>j}}^{N} q_A^{ijk} - (N-3) \sum_{j\neq i}^{N} q_A^{ij} + \frac{(N-2)(N-3)}{2} q_A^{i} \tag{6b}$$

where q_A^{ijk} is the partial charge assigned to atom A from the wave function or electron density of electrostatically embedded trimer ijk (which is composed of monomers i, j, and k).

As with the dipole moments, one can calculate the pure MB approximations to the atomic partial charge distribution of the entire system by using the partial charge distributions of the isolated monomers, dimers, and trimers.

In addition to calculating the partial atomic charges for their own sake, one can also use them to calculate dipole moments of individual molecules in the cluster.

5.3. Systems

Several configurations of seven different systems were chosen for this study: forty-four configurations of $(NH_3)(H_2O)_{11}$ (Figure 5.1), forty-four configurations $(NH_3)_2(H_2O)_{14}$ (Figure 5.2), six configurations of $[Cl(H_2O)_6]^-$ (Figures 5.3 and 5.4), six configurations of $(HF)_4$ (Figure 5.5), ten configurations of $(HF)_5$ (Figure 5.5), six

configurations of $(HF)_3(H_2O)$ (Figure 5.6), and two configurations of $(HF)_3(H_2O)_2$ (Figure 5.7). The EE-MB (and in some cases the MB) approximations of the dipole moment of each configuration are compared below to the conventionally calculated dipole moment of that configuration using the M06-2X[52] density functional and the cc-pVTZ+[53] basis set. MB and EE-MB approximations of the full-system M06-2X/cc-pVTZ+ CHelpG[50] point charge distributions are compared below to the conventional M06-2X/cc-pVTZ+ CHelpG point charge distributions of the microhydrated chloride, hydrogen fluoride, and microhydrated hydrogen fluoride systems described in Sections 5.3.2 and 5.3.3.

5.3.1. Aqueous Ammonia Droplets

The configurations for both of the water–ammonia systems were obtained as follows: a short Monte Carlo (MC) simulation with the SPC/E water model[54] and the OPLS ammonia model[55] was run using the configurations shown in Figures 5.1 and 5.2 as the starting points. The configurations generated during the first 44 steps of these simulations (including the starting configuration and rejected moves as well as accepted ones) were chosen as the test cases of the EE-MB method for dipoles. A conventional (a.k.a. "full-system") M06-2X/cc-pVTZ+ single-point energy and dipole calculation was run on each configuration of each system using *Gaussian 09*.[56] EE-PA and EE-3B dipole moment calculations were performed on each configuration by the M06-2X/cc-pVTZ+ method using a modified version of MBPAC 2007-2.[57] The M06-2X/cc-pVTZ+ method was used to calculate CM4M[58] charges of the equilibrium gas-phase water and ammonia monomers, and these CM4M charges were used as the embedding charges in the EE-MB calculations of the water–ammonia systems.

5.3.2. Microhydrated Chloride Ions

The water–chloride systems were also taken from an MC simulation at 250 K with the SPC/E water model and the CHARMM parameters for Cl^-.[59–60] Three of the water–chloride structures (int1, int2, and int3 in Figure 5.3) were selected because they represent an "interior" chloride ion in liquid water. Each of the "interior" structures contains at least four hydrogen bonds to the chloride ion where the O–Cl distance is less than 3.7 Å and where the O–H–Cl angle is greater than 150˚. The other three water–chloride structures (surf1, surf2, and surf3 in Figure 5.4) represent a "surface" chloride ion in liquid water. Each "surface" structure contains fewer than two hydrogen bonds to the chloride ion, where in this case a hydrogen bond is defined as having an O–Cl distance of less than 4.7 Å and an O–H–Cl angle greater than 145˚.

Conventional full-system M06-2X/cc-pVTZ+ dipole moment and CHelpG partial atomic charge calculations were run on each configuration of each water–chloride system using *Gaussian 09*. EE-1B, EE-PA, EE-3B, 1B, PA, and 3B dipole moment and CHelpG point charge calculations were also performed for each configuration by the M06-2X/cc-pVTZ+ method using a modified version of MBPAC 2007-2.

For the MB and EE-MB calculations of the water–chloride structures, two different fragmentation schemes were used: in fragmentation scheme 1 (S1), the chloride ion is considered a separate monomer and a point charge of –1 *e* is used as its embedding charge. In fragmentation scheme 2 (S2), the chloride ion is paired with the water molecule containing the atom nearest to the chloride ion and the resulting $[Cl{\bullet}H_2O]^-$ group is considered a single monomer. In Figures 5.3 and 5.4, the water

molecule that contains the atom closest to the chloride ion is always labeled "A", so that in S2 the fragmentation scheme is always ACl^-, B, C, D, E, F (yielding six monomers) whereas in S1 the fragmentation scheme is always Cl^-, A, B, C, D, E, F (yielding seven monomers). The CHelpG charges of the equilibrium gas-phase water molecule and the equilibrium gas-phase Cl^- (S1) or $[Cl(H_2O)]^-$ (S2) system were used as the embedding charges in the EE-MB calculations of the water–chloride clusters.

5.3.3. Hydrogen Fluoride Clusters and Microhydrated Hydrogen Fluoride Clusters

The $(HF)_m$ clusters (m = 4–5) shown in Figure 5.5 were constructed in the following way: (1) A partial geometry optimization at M06-2X/cc-pV(T+d)Z+ was attempted on each cluster with the constraints that the molecules must lie in a plane and that the molecules retain their initial orientations relative to one another. (2) After performing step 1, it was noted that the geometries of structures that had two fluorine atoms next to each other never converged to a minimum; rather, the molecules with the adjacent fluorine atoms kept drifting farther away from one another. (3) Therefore, a second partial geometry optimization of each $(HF)_m$ cluster was performed with both of the constraints described in step 1 but also with the constraint that the distance between adjacent fluorine atoms be fixed at 3.4 Å. This distance was chosen after performing a quick scan of a slice of the $(HF)_2$ potential energy surface (PES) where the two fluorine atoms are forced to be adjacent to one another. Of the distances used in this scan, 3.4 Å is the distance where the dimer binding energy (the energy of the infinitely separated monomers minus the energy of the dimer) is closest to –1.0 kcal/mol.

The $(HF)_3(H_2O)$ clusters in Figure 5.6 began with the M06-2X/cc-pV(T+d)Z+-optimized isolated gas-phase HF and H_2O molecules; the geometries of these molecules are given in Table 5.1. In each $(HF)_m(H_2O)$ cluster, the H_2O molecule (frozen in its optimized geometry) is placed in the center and *m* HF molecules (also frozen in their optimized geometries) are placed around it. HF molecule A is always placed so that the H, the F, and the O (of water) form a straight line perpendicular to the line through the two H atoms of water. If the F atom of A is closest to the O atom of water, then the F atom is placed R_{vdW}(F–O) Å away from the O of water; otherwise, the H atom of A is placed R_{vdW}(H–O) Å away from the O of water, where R_{vdW}(F–O) and R_{vdW}(H–O) are given in Table 5.2. HF molecules B and C are always placed so that the H, F, the nearest H of water and the O of water form a straight line. If the F atom of B or C is closet to the nearest H atom of water, then F is placed R_{vdW}(F–H) Å away from the nearest H atom of water. Otherwise, the H atom of B or C is placed R_{vdW}(H–H) Å away from the nearest H atom of water.

The $(HF)_3(H_2O)_2$ clusters are shown in Figure 5.7. The structure called w2hf3_a was constructed in exactly the same way as structure whf3_AaBaCa of Figure 5.6: the HF and H_2O molecules were kept in their M06-2X/cc-pV(T+d)Z+-optimized isolated gas-phase geometries and were placed at the appropriate R_{vdW} distance away from one another. In w2hf3_a, all molecules lie in the same plane. The w2hf3_a structure was then used as a starting point for an M06-2X/cc-pV(T+d)Z+ geometry optimization of the entire cluster. The result of this geometry optimization is the structure named

w2hf3_opt. This structure (w2hf3_opt) is a minimum-energy structure on the M06-2X/cc-pV(T+d)Z+ $(HF)_3(H_2O)_2$ PES; it contains no imaginary frequencies.

The dipole moment and CHelpG partial charge distribution of each hydrogen fluoride and microhydrated hydrogen fluoride cluster were calculated conventionally and within the EE-1B, EE-PA, EE-3B, 1B, PA, and 3B approximations using the M06-2X density functional and the cc-pV(T+d)Z+ basis set. For the EE-MB calculations, the M06-2X/cc-pV(T+d)Z+ CHelpG charges of the equilibrium gas-phase hydrogen fluoride or water molecule were used as the embedding charges.

5.4. Results

In order to assess the accuracy of the EE-MB and MB approximations for the calculation of full-system dipole moments, partial atomic charges, and partial charge transfer between fragments, the EE-MB and MB approximations of these values are compared to the conventionally calculated values at the same level of electronic structure theory (which, for all results shown, is the M06-2X density functional with the cc-pVTZ+ basis set). The errors summarized in the tables described in this section are calculated as the EE-MB (or MB) approximate values minus the conventionally calculated values.

Next we describe the results in Tables 5.2–5.6; the results will be discussed in Section 5.6. First we note that Tables 5.3–5.5 each include the number, n, of configurations over which each average is taken and the number, N, of monomers into which each configuration of each system is divided.

5.4.1. Full-System Dipole Moments

For simplicity, throughout the remainder of this paper we will refer to the magnitude of the dipole moment as the dipole moment. Table 5.3 shows the mean unsigned errors (MUEs) of the EE-MB calculations of cluster dipole moments (and, in some cases, MB calculations of these dipole moments) over all of the configurations of each type of system studied in this work. Table 5.3 also shows, to help put these values in perspective, the average, $\langle \mu \rangle$, of the conventionally calculated cluster dipole moments over all configurations of each type of system; not that this is the same as $\langle |\mu| \rangle$

5.4.2. Partial Atomic Charges

Table 5.4 shows the MUEs in the EE-MB and MB approximations of the CHelpG partial atomic charge distributions of the microhydrated chloride, hydrogen fluoride, and microhydrated hydrogen fluoride clusters used in this work. Table 5.4 also shows the average, $\langle |q| \rangle$, of the absolute values of the conventionally calculated atomic charges over all atoms in all configurations of each type of system.

5.4.3. Partial Charge Transfer

Table 5.5 shows the MUEs in the EE-MB and MB approximations of the net fragment charges, Q, of the microhydrated chloride, hydrogen fluoride, and microhydrated hydrogen fluoride clusters considered in this work. The net charge on a fragment (in this case, on a monomer of the cluster) is calculated as the sum of the CHelpG charges of the atoms that constitute the given fragment. Additionally, Table 5.5 shows the average, $\langle |Q| \rangle$, of the absolute values of the conventionally calculated net fragment charges over all configurations of each type of system.

5.4.4. Molecular Dipole Moments

The full-system dipole moment is well defined, but fragment dipole moments are model quantities. Despite their model character, molecular dipole moments of molecules in clusters and liquids are widely used for interpretative purposes, and here we show that we can compute them by the EE-MB method. Table 5.6 shows the average water fragment dipole moments for the $Cl(H_2O)^-$ clusters and also the average deviation from those calculated by CHelpG analysis of full-system density functional calculations.

5.5. Discussion

5.5.1. Dipole Moments

Table 5.3 shows that, as expected, the EE-MB full-system dipole moment becomes more accurate as the order of the EE-MB approximation increases. However, even for larger clusters, the EE-PA approximation is able to get the full-system dipole moment quantitatively correct. This is good news because of the more favorable scaling of the cost of the EE-PA approximation with system size than that of the EE-3B approximation. The mean unsigned percentage error (MUPE) over all configurations of all aqueous ammonia droplets is 5.0% in the EE-1B approximation, 0.9% in the EE-PA approximation, and 0.5% in the EE-3B approximation. Table 5.3 also shows that electrostatic embedding significantly improves the accuracy of the one-body approximation and slightly improves the accuracies of the two- and three-body approximations.

5.5.2. Partial Atomic Charges

The $[Cl(H_2O)_6]^-$ system provides an interesting challenge to many-body methods. The charge on Cl^- is of course –1.0 atomic units (a.u.), and the partial atomic and standard deviation on Cl in $Cl^{\bullet}(H_2O)^-$ in the six configurations (where we always consider the H_2O that contains the atom closest to Cl^- in the overall cluster) is –0.91±0.01 a.u., and when the $Cl^{\bullet}(H_2O)^-$ fragment is embedded in fifteen embedding charges representing the rest of the cluster, the average charge and its standard deviation on Cl become –0.94±0.04 a.u. However the average charge (and standard deviation) of Cl^- in the entire cluster is –0.70±0.07 a.u. We find that the EE-PA and EE-3B methods account for the reduction of this partial atomic charge from –1.0, –0.91, or –0.94 to –0.70 a.u. quite well. By definition, the EE-1B and 1B approximations cannot predict this reduction in charge because these methods do not allow charge transfer between monomers. However, the EE-PA and EE-3B approximations are able to account for this reduction in the charge assigned to Cl: using the simpler fragmentation scheme where the Cl^- ion is considered a monomer (S1; see Section 5.3.2), the average charge (and standard deviation) on Cl is –0.78±0.05 a.u. using the EE-PA approximation and is –0.75±0.05 a.u. using the EE-3B approximation. Using the fragmentation scheme where the Cl^- is paired with the closest water molecule to form a single monomer (S2), both the EE-PA and EE-3B approximations do a little better: the average charge (and standard deviation) on Cl is –0.75±0.06 a.u. using the EE-PA approximation and is –0.74±0.07 a.u. using the EE-3B approximation.

Table 5.4 shows trends in partial atomic charges similar to those noticed with the dipole moments: again, the accuracy of the EE-MB or MB approximation generally increases with the order to which the MB expansion is taken and, again, in most cases (except for the microhydrated hydrogen fluoride clusters) the PA-level approximation is already capable of yielding quantitative accuracy in the partial charge distribution relative to the conventionally calculated CHelpG charge distribution of the entire cluster. However, unlike the trend seen in the dipole moment calculations, the accuracy of the many-body approximations for the prediction of partial charge distributions depends strongly on whether or not electrostatic embedding is included only in the one-body case; in the pairwise additive and three-body cases, the accuracy of the MB approximation is similar to that of the EE-MB approximation of the same order.

An interesting point to consider is that as the size of a cluster increases, more and more atoms become "buried" within the cluster. Charge analysis methods that are based on finding a set of partial atomic charges that most closely reproduces the electrostatic potential due to the charge density of the entire system (like CHelpG) often have difficulty assigning stable physical charges to buried atoms. Using the EE-PA or the EE-3B method to calculate partial atomic charge distributions of large clusters could be a way to obtain stable physical charges on buried atoms without requiring the imposition of empirical constraints on the magnitudes of the charges assigned to various atoms.

5.5.3. Partial Charge Transfer

Partial charge transfer is ubiquitous in liquid-phase systems, nanoparticles, and clusters, but it is notoriously hard to treat by fragment methods.

The net charge on a fragment (calculated as the sum of the partial charges assigned to the atoms that compose the fragment) is an indication of how much charge transfer has occurred to or from that fragment when it moves from being an isolated state to the state of being a member of a given cluster: for example, an isolated gas-phase water molecule has a net charge of zero a.u., but if it is placed in a cluster containing a negatively-charged ion, its net charge might be –0.20 a.u., which shows that some of the electron density from the anion has been transferred to that water molecule, and molecules with net charges in the range of ±0.2 a.u. were also observed in simulations for liquid water which allowed for inter- and intramolecular charge equilibration.[61] The MUEs in net fragment charges shown in Table 5.5 are therefore a measure of how well the EE-MB approximations are able to capture the amount charge transfer occurring between the molecules in the clusters studied.

The one-body approximations are incapable of predicting charge transfer between molecules, so for systems containing all neutral fragments the EE-1B and 1B approximations necessarily have errors of 100% in the calculation of net monomer charges. Moving to the PA or EE-PA approximation reduces the error in net monomer charges to between 20 and 30% in most cases. Using the 3B or EE-3B approximation generally reduces the error in net monomer charges to between 13 and 20%. As with the partial charge calculations, a strong dependence on whether or not electrostatic embedding is used is not seen in the accuracy of the MB net monomer charge calculations.

Although neither the EE-MB nor the MB approximations were capable of quantitatively accurate predictions of net monomer charges for the cases chosen in this

work, the PA- and 3B-level approximations were able to qualitatively describe the amount of charge transfer between fragments most of the time.

5.5.4. Molecular Dipole Moments

The mean, median, minimum, and maximum molecular dipole moments for water are 2.19, 2.18, 2.03, and 2.34 D respectively, in the $(NH_3)(H_2O)_{11}$ clusters, 2.51, 2.53, 2.120, and 2.76 D, respectively, in the $[Cl(H_2O)_6]^-$ clusters, and 2.470, 2.472, 2.20, and 2.72 D, respectively, in the $[Cl•(H_2O)(H_2O)_5]^-$ clusters. A histogram showing the distribution of the EE-1B water dipole moments over all forty-four $(NH_3)(H_2O)_{11}$clusters is shown in Figure 5.8.

As a further illustration of the application of the EE-1B method for the computation of molecular dipole moments, we note that the mean molecular dipole moments for individual water molecules (1B approximation) are 2.07 D in $[Cl(H_2O)_6]^-$ but these values change when one considers the clusters. In fact, if the molecular dipole moments are calculated from the CHelpG charges of the full-system calcuations, the mean water monomer dipole moment changes by 46% (the new mean is 2.93 D). Table 5.6 shows that the EE-PA and EE-3B approximations give similar accuracyfor the average molecular dipole moments in the chloride–water clusters, and they both agree better then the 3B approximation with the average molecular dipole moments calculated from full-system density functional calcuations. The performance of the EE-PA method is very encouraging.

5.6. Summary

Our first set of tests using the EE-MB method for predicting charge distributions was concerned with full-system dipole moments. We compared the EE-MB results to

the one-body (1B) approximation, in which one simply carries out vectorial addition of the dipole moments of the noninteracting constituents, and to the EE-1B approximation, in which one vectorially adds the dipole moments of the electrostatically embedded monomers. Both of these approximations led to large errors, but the MUEs of EE-1B were reduced by factors of 3 to 18 by the EE-PA approximation and by factors of 6 to 55 by the EE-3B approximation, and the MUEs of the 1B approximation were reduced by factors of 7 to 63 by the PA approximation and by factors of 19 to 123 by the 3B approximation.

The second set of tests using the EE-MB method for predicting charge distributions was concerned with full-system partial atomic charge distributions. The 1B approximation of a particular atom's partial charge is simply that atom's partial charge when it is in an isolated monomer; the EE-1B approximation of an atom's partial charge is that atom's partial charge when its monomer is surrounded by embedding charges. On average, electrostatic embedding reduces the MUE of the 1B approximation by a factor of two. Including the pairwise additive terms usually reduces the MUEs of the 1B approximation by factors of 3 to 7, and using the 3B approximation also reduces the MUEs by factors of 3 to 7. The EE-PA approximation does not improve upon the EE-1B approximation by as large an amount for this quantity: the MUEs are reduced by factors of 1.0 to 1.5 in going from the EE-1B to the EE-PA approximation. A more significant improvement is seen in going from the EE-1B approximation to the EE-3B approximation: in this case, the MUEs are reduced by factors of 2 to 7.

A more difficult set of tests using the EE-MB method for predicting charge distributions was concerned with the amount of intermolecular charge transfer. The

EE-1B and 1B approximations (and many other linear-scaling fragment based methods) do not allow any charge transfer to occur between monomers at all, so even being able to get a qualitative picture of the amount of intermolecular charge transfer occurring in a system using the EE-PA or EE-3B approximations is an improvement. We find that in going from the EE-1B approximation to the EE-PA approximation the MUEs are reduced by factors of 2 to 3 and in going from the EE-1B approximation to the EE-3B approximation the MUEs are reduced by factors of 3 to 6. Similarly, in going from 1B to PA the errors are reduced by factors of 2 to 4 and in going from 1B to 3B the errors are reduced by factors of 3 to 5.

Finally we showed that the EE-MB approximation can also be used to calculate dipole moments for fragments of the cluster.

We conclude that the EE-PA and EE-3B approximations are capable of yielding accurate descriptions of full-system charge distributions through dipole moment and atomic partial charge distribution calculations. The EE-PA and EE-3B approximations are also capable of qualitatively describing the amount of intermolecular charge transfer occurring in a system, which is a difficult task for fragment-based methods.

5.7. References for Chapter 5

1 N. Ferre and X. Assfeld, *J. Mol. Struct. Theochem*., 2003, **632**, 83.
2 T. E. Exner and P. G. Mezey, *Phys. Chem. Chem. Phys*., 2005, **7**, 4061.
3 D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 6904.
4 W. Hua, T. Fang, W. Li, J. Yu, and S. Li, *J. Phys. Chem. A*, 2008, **112**, 10864.
5 W. Xie, L. Song, D. G. Truhlar, and J. Gao, *J. Chem. Phys*. 2008, **128**, 234108.
6 M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko, and J. A. Boatz, *J. Phys. Chem. B,* 2009, **113**, 9646.
7 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput*., 2007, **3**, 46.
8 Y. Tong, Y. Mei, Y. L. Li, C. G. Ji, and J. Z. H. Zhang. *J. Am. Chem. Soc*., 2010, **132**, 5137.
9 R. Mata, *Mol. Phys*., 2010, **108**, 381.
10 S. Hirata, *Mol. Phys*., 2010, **108**, 3113.
11 H. Stoll, H. Preuss, *Theor. Chem. Acc.* 1977, **46**, 12.
12 W. Yang, *Phys. Rev. Lett.* 1991, **66**, 1438.
13 V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. C. Ferenczy, *J. Comput. Chem*., 1994, **15**, 269.
14 J. Gao, *J. Phys. Chem. B* 1997, **101**, 657.
15 J. M. Pedulla, K. Kim, and K. D. Jordan, *Chem. Phys. Lett*., 1998, **291**, 78.
16 K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, *Chem. Phys. Lett.* 1999, **313**, 701.
17 C. Amovilli, I. Cacelli, S. Campanile, and G. Prampolini, *J. Chem Phys*., 2002, **117**, 3003.
18 D. W. Zhang, Y. Xiang, and J. Z. H. Zhang, *J. Phys. Chem. B* 2003, **107**, 12039.
19 S. -I. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, *Chem. Phys. Lett*., 2003, **382**, 611.
20 S. Li, W. Li, and T. Fang, *J. Am. Chem. Soc.*, 2005, **127**, 7215.
21 S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, *Mol. Phys*., 2005, **103**, 2255.
22 N. Jiang, J. Ma, and Y. Jiang, *J. Chem Phys*., 2006, **124**, 114112.
23 M. A. Collins, V. A. Deev, *J. Chem Phys*., 2006, **125**, 104104.
24 R. P. A. Bettens and A. M. Lee, *J. Phys. Chem. A*, 2006, **110**, 8777.
25 E. E. Dahlke and D. G. Truhlar, *J. Chem.Theory Comput*., 2007, **3**, 1342.
26 E. Suárez, N. Díaz, and D. Suarez, *J. Chem.Theory Comput*., 2009, **5**, 1667.
27 P. Söderhjelm, F. Aquilante, and U. Ryde, *J. Phys. Chem. B*, 2009, **113**,11085.
28 J. O. B. Tempkin, H. R. Leverentz, B. Wang, and D. G. Truhlar, *J. Phys. Chem. Lett*., 2011, **2**, 2141.
29 L. D. Jacobson and J. M. Herbert, *J. Chem Phys*., 2011, **134**, 094118.
30 N. J. Mayhall and K. Raghavachari, *J. Chem.Theory Comput*., 2011, **7**, 1336.
31 D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, *J. Chem Phys*., 2011, **135** , 044123.
32 W. Wen and G. J. O. Beran, *J. Chem.Theory Comput*., 2011, **7**, 3733.
33 M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem. Rev.* 2011, ASAP, doi: 10.1021/cr200093j.

34 E. D. Speetzen, H. R. Leverentz, H. Lin, and D. G. Truhlar, in *Accurate Condensed-Phase Electronic Structure Theory*, ed. F. R. Manby, CRC Press, Boca Raton, FL, 2011, p. 105.
35 M. Isegawa, J. Gao, and D. G. Truhlar, *J. Chem. Phys*., 2011, **135**, 084107.
36 N. Marzari and D. Vanderbilt, *Phys. Rev. B*, 1997, **56**, 12847.
37 P. L. Silvestrelli and M. Parrinello, *J. Chem. Phys*., 1999, **111**, 3572.
38 G. Berghoid, C. J. Mundy, A. H. Romero, J. Hutter, and M. Parrinello, *Phys. Rev. B*, 2000, **61**, 10040.
39 M. P. Gaigeot and M. Sprik, *J. Phys. Chem. B*, 2003, **107**, 10344.
40 M. Sharma, R. Resta, and R. Car, *Phys. Rev. Lett*., 2005, **95**, 187401.
41 C. Dellago and M. M. Naor, *Computer Phys. Commun*., 2005, **169**, 36.
42 M. Sharma, R. Resta, and R. Car, *Phys. Rev. Lett*., 2007, **98**, 247401.
43 M. J. Mcgrath, J. I. Siepmann, I. F. W. Kuo, C. J. Mundy, *Mol. Phys*., 2007, **105**, 1411.
44 A. Ferratti, A. Calzolari, B. Bonferroni, and R. Di Felice, *J. Phys. Condens. Matter*, 2007, **19**, 036215.
45 D. Kang, J. Dai, and J. Yuan, *J. Chem. Phys*., 2011, **135**, 024505.
46 B. Kirchner, P. G. Di Dio, and J. Hutter, *Top. Curr. Chem*., 2012, **307**, 109.
47 D. -L. Chen, A. C. Stern, B. Space, and J. K. Johnson, *J. Phys. Chem. A*, 2010 **114**, 10225.
48 T. Watanabe, T. A. Manz, and D. S. Sholl, *J. Phys. Chem. C* 2011, **115**, 4824.
49 R. S. Mulliken, *J. Chem. Phys*., 1955, **23**, 1833.
50 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem*., 1990, **11**, 361.
51 J. P. Foster and F. Weinhold, *J. Am. Chem. Soc*., 1980, **102**, 7211.
52 Y. Zhao and D. G. Truhlar, *Theo. Chem. Acc*., 2008, **120**, 215.
53 E. Papajak, H. R. Leverentz, J. Zheng, and D. G. Truhlar, *J. Chem. Theory Comput*., 2009, **5**, 1197.
54 H. J. C. Berendsen, J. R. Grigera, and T. P Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269.
55 R. C. Rizzo and W. L. Jorgensen, *J. Am. Chem. Soc*., 1999, **121**, 4827.
56 M. J. Frisch, et al., *Gaussian 09*, Revision A.2, Gaussian, Inc., Wallingford, CT, 2009.
57 E. E. Dahlke and D. G. Truhlar, *MBPAC 2007-2*, University of Minnesota, Minneapolis, MN, 2007.
58 R. M. Olson, A. V. Marenich, C. J. Cramer, and D. G. Truhlar. *J. Chem. Theory Comput*., 2007, **3**, 2046.
59 D. Beglov and B. Roux, *J. Chem. Phys.*, 1994, **100**, 9050.
60 B. Roux, *Biophys. J.*, 1996, **71**, 3177.
61 B. Chen, J. J. Potoff, and J. I. Siepmann, *J. Chem. Phys.*, 2000, **104**, 2378.
62 M. Mantina, R. Valero, and D. G. Truhlar, *J. Phys. Chem. A*, 2009, **113**, 5806.

Table 5.1. Geometric Parameters of the M06-2X/cc-pV(T+d)Z+ Optimized Gas-Phase Water and Hydrogen Fluoride Molecules

Molecule	Parameter	Value
HF	R_{HF}	0.9194 Å
H_2O	R_{OH}	0.9592 Å
H_2O	θ_{HOH}	105.2º

Table 5.2. Van der Waals Radii of Atoms and Combined van der Waals Radii of Atom Pairs

Radius	Value (Å)	Reference
R_{vdW}(F)	1.47	62
R_{vdW}(H)	1.10	62
R_{vdW}(O)	1.52	62
R_{vdW}(F–F)	2.94	R_{vdW}(F) + R_{vdW}(F)
R_{vdW}(F–H)	2.57	R_{vdW}(F) + R_{vdW}(H)
R_{vdW}(F–O)	2.99	R_{vdW}(F) + R_{vdW}(O)
R_{vdW}(H–H)	2.20	R_{vdW}(H) + R_{vdW}(H)
R_{vdW}(H–O)	2.62	R_{vdW}(H) + R_{vdW}(O)
R_{vdW}(O–O)	3.04	R_{vdW}(O) + R_{vdW}(O)

Table 5.3. Average cluster dipole moments and mean unsigned errors in cluster dipole moments (in debye)[a]

Ensemble	*n*	*N*	$\langle \mu \rangle$	$\langle \lvert \mu^{MB} - \mu \rvert \rangle$ 1B	PA	3B
			EE-MB			
$NH_3(H_2O)_{11}$	44	12	13.89	1.02	0.10	0.08
$(NH_3)_2(H_2O)_{14}$	44	16	17.97	0.50	0.18	0.08
$[Cl(H_2O)_6]^-$	6	7	6.36	0.42	0.10	0.04
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	6.36	0.33	0.07	0.04
$(HF)_{4–5}$	16	4–5	3.198	0.003	0.001	0.000
$(HF)_3(H_2O)_{1–2}$	8	4–5	5.262	0.164	0.009	0.003
			MB			
$[Cl(H_2O)_6]^-$	6	7	6.36	1.60	0.17	0.08
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	6.36	1.13	0.17	0.06
$(HF)_{4–5}$	16	4–5	3.198	0.125	0.002	0.000
$(HF)_3(H_2O)_{1–2}$	8	4–5	5.262	0.613	0.023	0.005

[a] *n* is the number of configurations in the ensemble, *N* is the number of fragments, μ is the dipole moment from a conventional calculation on the entire system, and μ^{MB} is the dipole moment calculated by the EE-MB or MB method.

Table 5.4. Mean unsigned errors (MUEs) in partial atomic charges (in atomic units)[a]

				MUE		
Ensemble	n	N	$\langle\|q\|\rangle$	1B	PA	3B
			EE-MB			
$[Cl(H_2O)_6]^-$	6	7	0.56	0.06	0.06	0.04
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	0.56	0.06	0.04	0.03
$(HF)_{4–5}$	16[b]	4–5[c]	0.438	0.004	0.003	0.002
$(HF)_3(H_2O)_{1–2}$	8[d]	4–5[e]	0.452	0.033	0.023	0.005
			MB			
$[Cl(H_2O)_6]^-$	6	7	0.56	0.10	0.04	0.04
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	0.56	0.09	0.03	0.03
$(HF)_{4–5}$	16	4–5	0.438	0.014	0.002	0.002
$(HF)_3(H_2O)_{1–2}$	8	4–5	0.452	0.036	0.014	0.009

[a] n is the number of configurations in the ensemble and N is the number of fragments.
[b] The average is over 148 partial atomic charges.
[c] Six cases with $N = 4$ and ten with $N = 5$.
[d] The average is over 78 partial atomic charges
[e] Six cases with $N = 4$ and two with $N = 5$.

Table 5.5. Average fragment charges and mean unsigned errors in fragment charges (in atomic units)[a]

Ensemble	*n*	*N*	$\langle \lvert Q \rvert \rangle$	$\langle \lvert Q^{\mathrm{MB}} - Q \rvert \rangle$ 1B	PA	3B
			EE-MB			
$[Cl(H_2O)_6]^-$	6	7	0.15	0.09	0.04	0.03
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	0.15	0.08	0.03	0.03
$(HF)_5$[b] and $(HF)_3(H_2O)_2$	3	5	0.036	0.036	0.023	0.006
			MB			
$[Cl(H_2O)_6]^-$	6	7	0.15	0.09	0.03	0.02
$[Cl{\bullet}H_2O(H_2O)_5]^-$	6	6	0.15	0.07	0.03	0.02
$(HF)_5$[b] and $(HF)_3(H_2O)_2$	3	5	0.036	0.036	0.010	0.012

[a] *n* is the number of configurations in the ensemble, *N* is the number of fragments, *Q* is the net fragment charge from a conventional calculation on the entire system, and Q^{MB} is the net fragment charge calculated by the EE-MB or MB method.
[b] Only one configuration of $(HF)_5$ is included in the calculation of these averages: hf5_fhhhf (see Figure 5.5).

Table 5.6. Average[a] water monomer dipole moment (in debye) calculated from CHelpG charges in $[Cl(H_2O)_6]^-$ clusters[b]

	Conventional	1B	PA	3B	EE-1B	EE-PA	EE-3B
int1	3.030	2.117	3.294	2.747	2.563	2.896	2.701
int2	2.897	2.102	3.266	2.744	2.525	2.955	2.740
int3	2.862	2.085	3.269	2.772	2.541	2.994	2.846
surf1	2.912	2.000	3.152	2.803	2.475	3.008	2.764
surf2	2.865	2.014	3.224	2.631	2.459	3.058	2.726
surf3	3.052	2.113	3.298	2.849	2.641	3.284	2.917
MUE[c]		0.864	0.314	0.178	0.402	0.141	0.154

[a]Averages are taken over the six water monomer dipole moments in each cluster.
[b]Fragmentation scheme 1 is used in all MB and EE-MB calculations shown in this table
[c]MUE = mean unsigned error in the average water monomer dipole moment from MB or EE-MB CHelpG charge distributions relative to the average water monomer dipole moment from the conventional CHelpG charge distribution using the M06-2X/cc-pV(T+d)Z+ method.

Figures

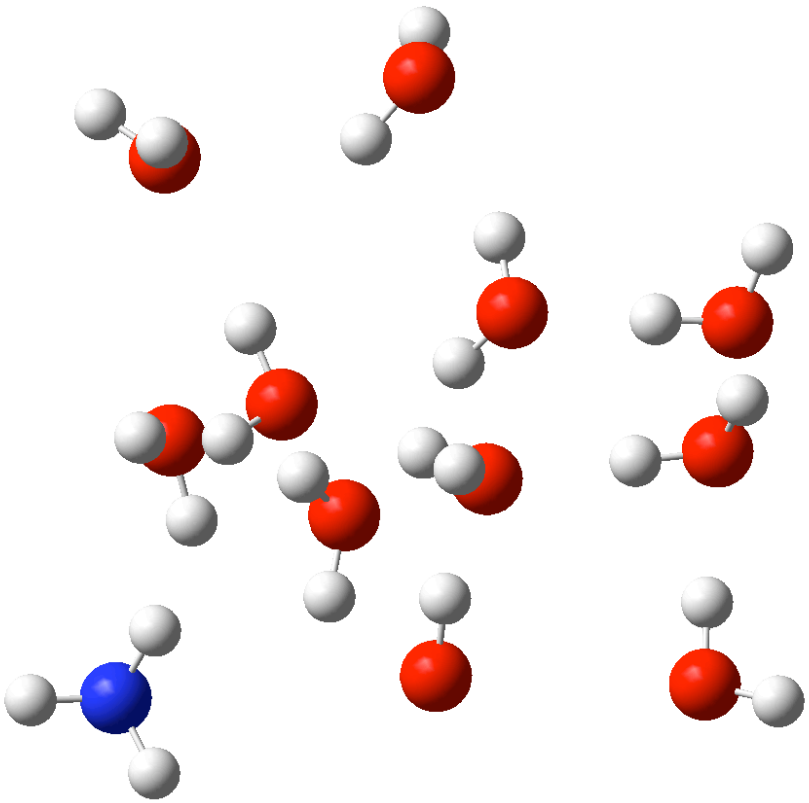


Figure 5.1

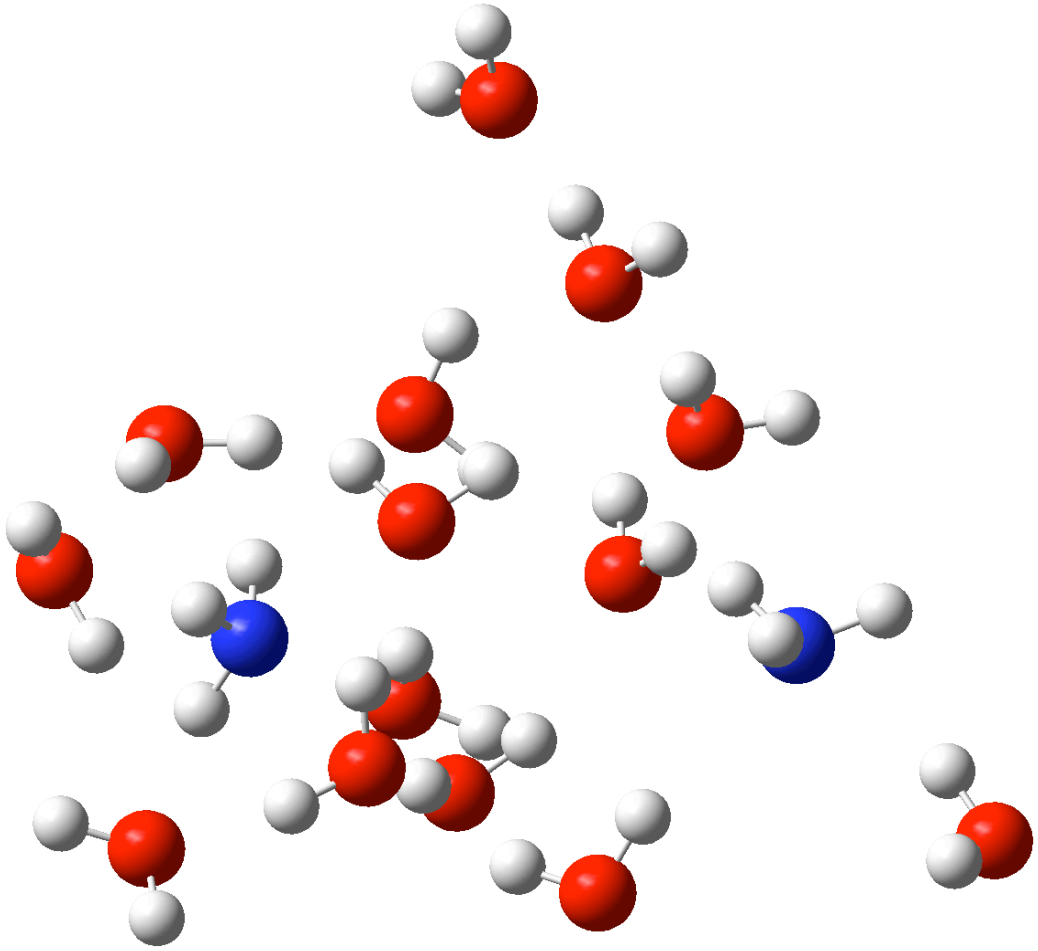
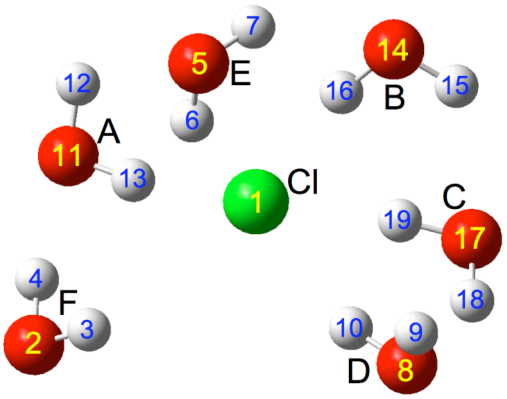
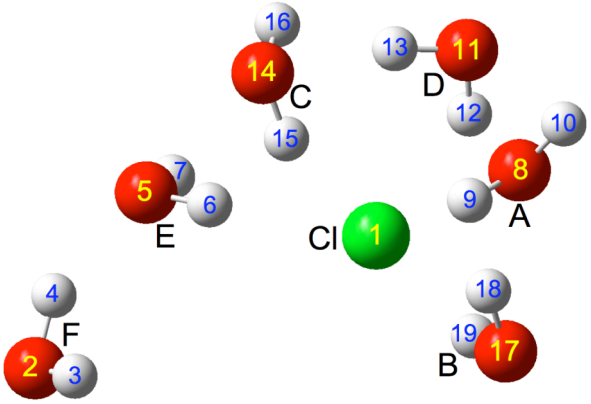


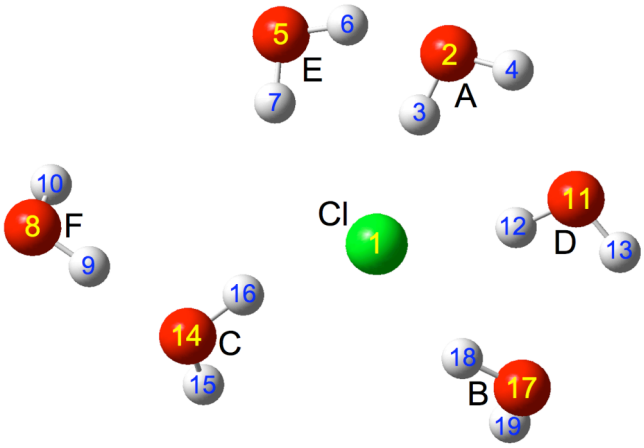
Figure 5.2



int1

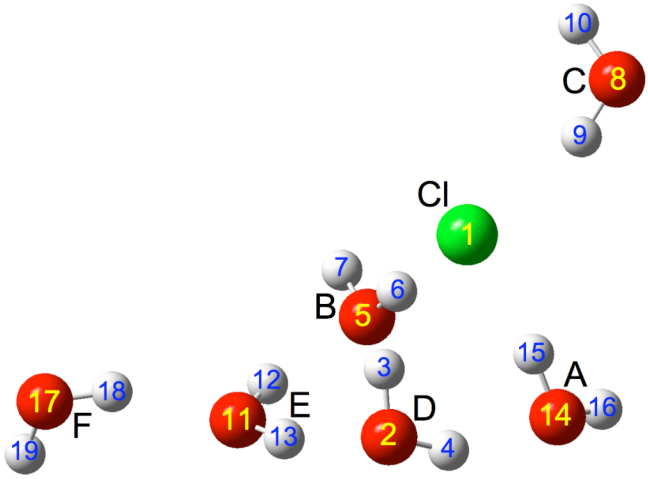


int2

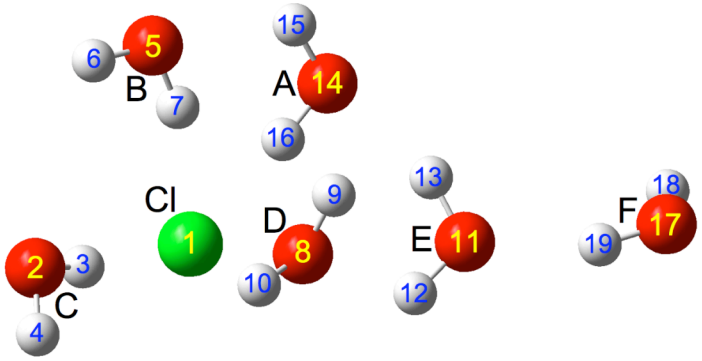


int3

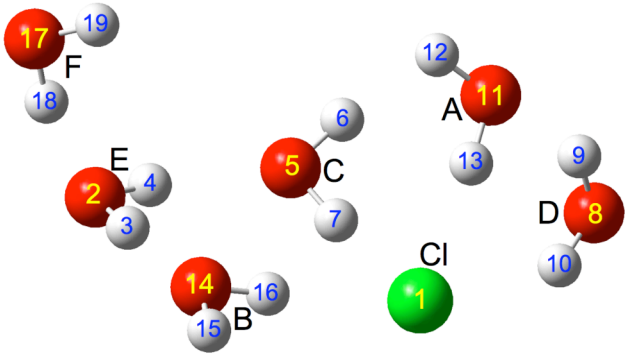
Figure 5.3



surf1



surf2



surf3

Figure 5.4

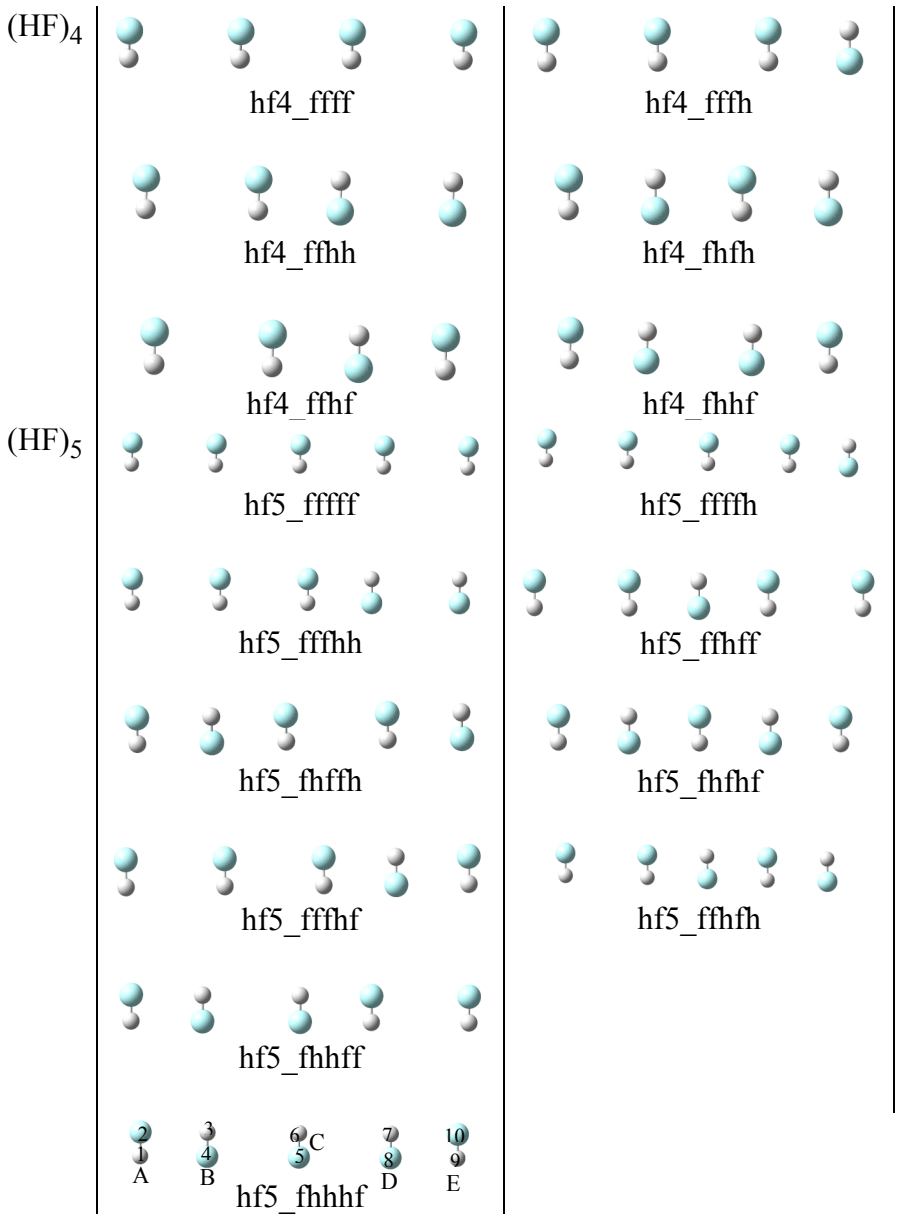

(HF)4
hf4_ffff
hf4_fffh
hf4_ffhh
hf4_fhfh
hf4_ffhf
hf4_fhhf
(HF)5
hf5_fffff
hf5_ffffh
hf5_fffhh
hf5_ffhff
hf5_fhffh
hf5_fhfhf
hf5_fffhf
hf5_ffhfh
hf5_fhhff
hf5_fhhhf
A
B
C
D
E

Figure 5.5

$(HF)_3(H_2O)$

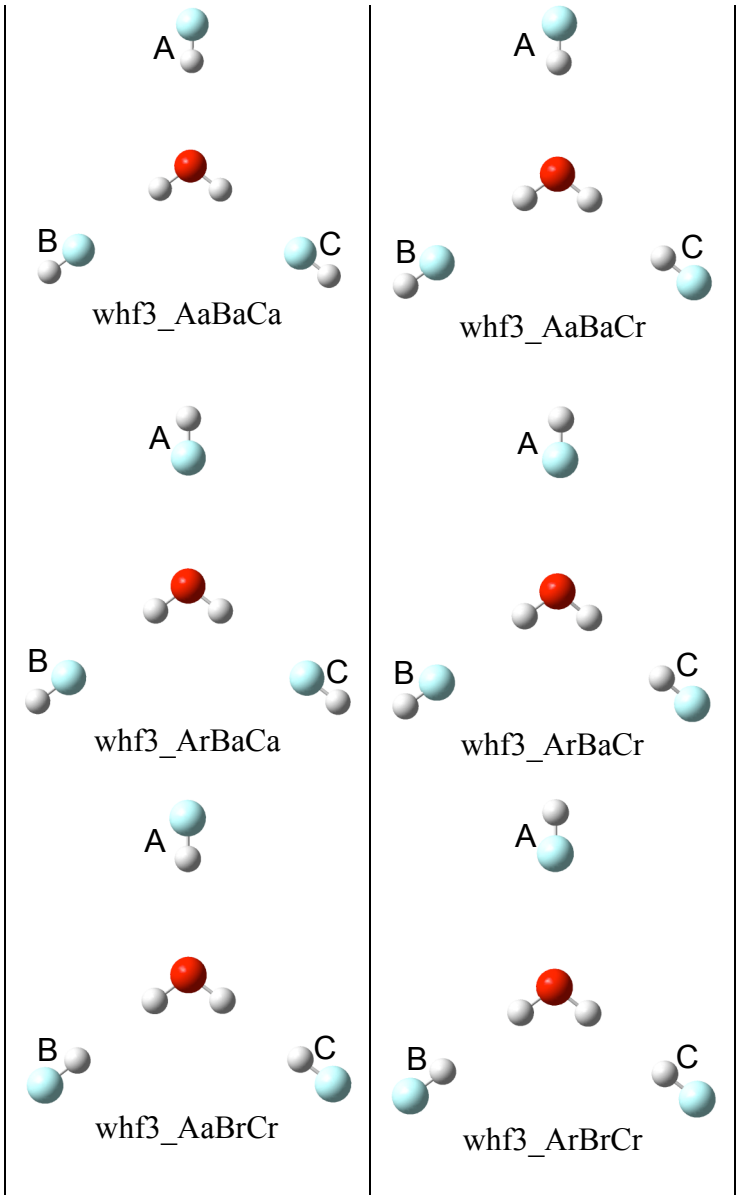
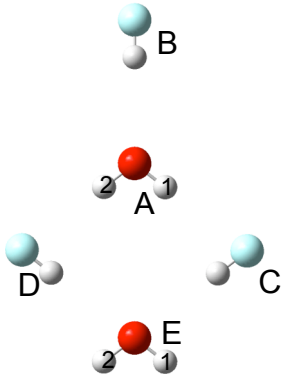
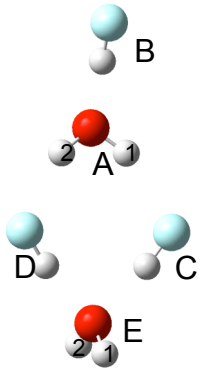


Figure 5.6

$(HF)_3(H_2O)_2$



w2hf3_a



w2hf3_opt

Figure 5.7

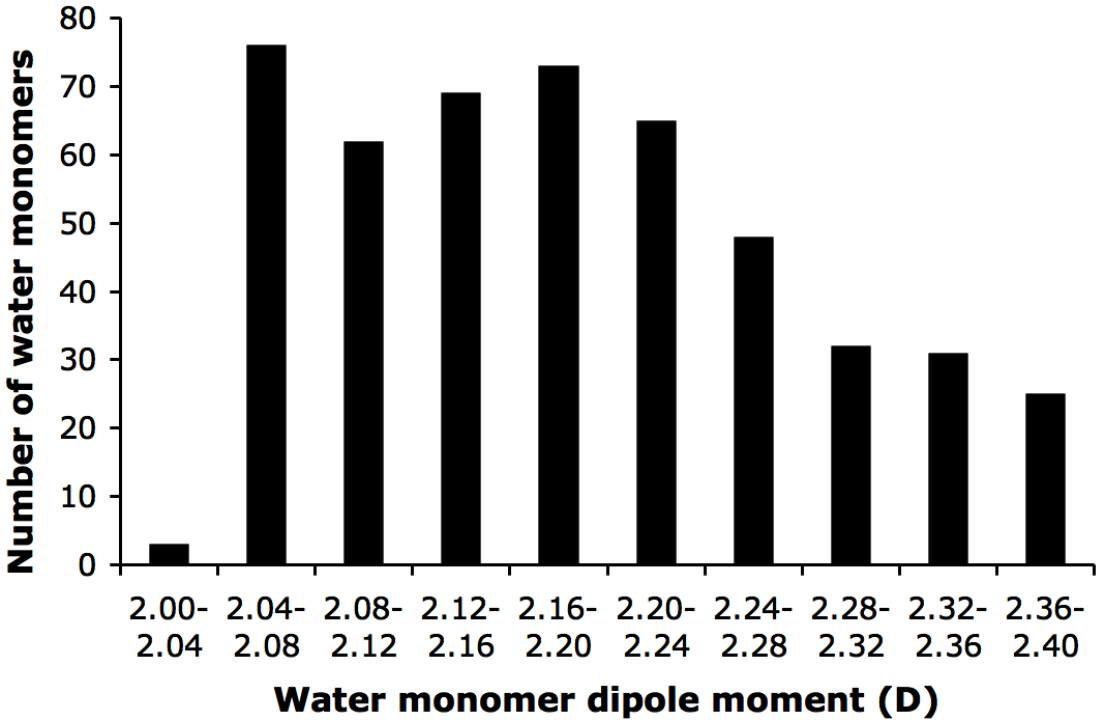

Number of water monomers
80
70
60
50
40
30
20
10
0
2.00-2.04
2.04-2.08
2.08-2.12
2.12-2.16
2.16-2.20
2.20-2.24
2.24-2.28
2.28-2.32
2.32-2.36
2.36-2.40
Water monomer dipole moment (D)

Figure 5.8

Bibliography

Addicoat, M. A.; Collins, M. A. *J. Chem. Phys*. **2009**, *131*, 104103.

Akama, T.; Kobayashi, M.; Nakai, H. *J. Comput. Chem*. **2007**, *28*, 2003.

Almlöf, J. *Chem. Phys. Lett*. **1991**, *181*, 319.

Amovilli, C.; Cacelli, I.; Campanile, S.; Prampolini, G. *J. Chem Phys*. **2002**, *117*, 3003.

Anderson, K. E.; Siepmann, J. I.; McMurry, P. H.; VandeVondele, J. *J. Am. Chem. Soc*. **2008**, *130*, 14144.

Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys*. **1992**, *96*, 1218.

Arunan, E.; Emilsson, T.; Gutowsky, H. S.; Fraser, G. T.; Oliveira, G. D.; Dykstra, C. E. *J. Chem. Phys*. **2002**, *117*, 9766-9776.

Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys*. **1999**, *110*, 3660.

Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221.

Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. *J. Chem. Phys*. **2011**, *135*, 044123.

Beglov, D.; Roux, B. *J. Chem. Phys*. **1994**, *100*, 9050.

Beran, G. J. O. *J. Chem. Phys*. **2009**, *130*, 164115.

Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem*. **1987**, *91*, 6269.

Berghoid, G.; Mundy, C. J.; Romero, A. H.; Hutter, J.; Parrinello, M. *Phys. Rev. B* **2000**, *61*, 10040.

Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem*. **1990**, *11*, 431.

Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.

Boys, S. F.; Bernardi, D. *Mol. Phys*. **1970**, *19*, 553.

Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem*. **1990**, *11*, 361.

Bytautas, L.; Matsunaga, N.; Nagata, T.; Gordon, M. S.; Reudenberg, K. *J. Chem. Phys*. **2007**, *127*, 204301.

Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Chem. Phys*. **2000**, *104*, 2378.

Chen, D. -L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010** *114*, 10225.

Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys*. **2006**, *125*, 44903.

Chen, X.; Zhang, Y.; Zhang, J. Z. H. *J. Chem. Phys*. **2005**, *122*, 184105.

Cizek, J. *J. Chem. Phys*. **1966**, *45*, 4256.

Clough, S. A.; Beers, Y.; Klein, G. P.; Rotham, L. S. *J. Chem. Phys*. **1973**, *59*, 2254.

Collins, M. A.; Deev, V. A. *J. Chem Phys*. **2006**, *125*, 104104.

Cramer, C. J. In *Essentials of Computational Chemistry. Theories and Models*; John Wiley and Sons Ltd: Chichester, 2004; .

Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys*. **1999**, *110*, 4703.

Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput*. **2008**, *4*, 33.

Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 1342.

Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 46.

Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput*. **2008**, *4*, 1.

Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys*. **2000**, *112*, 2063.
Deev, V.; Collins, M. A. *J. Chem. Phys*. **2005**, *122*, 154102.
Dellago, C.; Naor, M. M. *Computer Phys. Commun*. **2005**, *169*, 36.
Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc*. **1985**, *107*, 3902.
Dobrowolski, J.; Jamróz, M. J. Mol. Struc. **1993**, 293, 147.
Dunning, T. H. J.; Hay, P. J. In *Methods of Electronic Structure Theory*; Plenum: New York, 1977.
Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys*. **2001**, *114*, 9244.
Dyke, T. R.; Muenter, J. S. *J. Chem. Phys*. **1973**, *59*, 3125.
Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.
Edwards, T. H.; Moncur, N. K.; Snyder, L. E. *J. Chem. Phys*. **1967**, *46*, 2139.
Elrod, M. J.; Saykally, R. *J. Chem. Rev*. **1994**, *94*, 1975.
Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2004**, *108*, 4301.
Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys*. **2005**, *7*, 4061.
Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett*. **2006**, *433*, 182.
Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.
Ferratti, A.; Calzolari, A.; Bonferroni, B.; Di Felice, R. *J. Phys. Condens. Matter* **2007**, *19*, 036215.
Ferre, N.; Assfeld, X. *J. Mol. Struct. Theochem*. **2003**, *632*, 83.
Förner, W.; Ladik, J.; Otto, P.; Cízek, *J. Chem. Phys*. **1985**, *97*, 251.
Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc*. **1980**, *102*, 7211.
Gadre, S. G.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem*. **1994**, *98*, 9165.
Gaigeot, M. P.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 10344.
Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.
Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865.
Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev*. **2011**, ASAP, doi: 10.1021/cr200093j.
Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.
Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. *J. Chem. Phys*. **2004**, *121*, 10385.
Guo, W.; Wu, A.; Xu, X. *Chem. Phys. Lett*. **2010**, *498*, 203.
Hampel, C.; Werner, H.-J. *J. Chem. Phys*. **1996**, *104*, 6286.
Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett*. **1972**, *16*, 217.
He, X.; Zhang, J. Z. H. *J. Chem. Phys*. **2005**, *122*, 031103.
Head-Gordon, M.; Maslen, P. E.; White, C. A. *J. Chem. Phys*. **1998**, *108*, 616.
Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; .
Hirata, S. *J. Chem. Phys*. **2008**, *129*, 204104.
Hirata, S. *Mol. Phys*. **2010**, *108*, 3113.
Hirata, S.; Valiev, M.; Dupuis, M. X.,S. S.; Sugiki, S.; Sekino, H. *Mol. Phys*.

2005, *103*, 2255.
Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys*. **2005**, *103*, 2255.
Hopkins, B. W.; Tschumper, G. S. *J. Comput. Chem*. **2003**, *24*, 1563.
Hopkins, B. W.; Tschumper, G. S. *Mol. Phys*. **2005**, *103*, 309.
Hua, S.; Hua, W.; Li, S. *J. Phys. Chem. A* **2010**, *114*, 8126.
Hua, W.; Fang, T.; Li, W.; Yu, J.; and Li, S. *J. Phys. Chem. A* **2008**, *112*, 10864.
Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci*. **2006**, *103*, 1233.
Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem*. **2004**, *108*, 5692.
Isegawa, M.; Gao, J.; Truhlar, D. G. *J. Chem. Phys*. **2011**, *135*, 084107.
Iwahori, J.; Ueda, Y.; Nakagawa, K. *J. Mol. Spectrosc*. **1986**, *117*, 1.
Jacobson, L. D.; Herbert, J. M. *J. Chem Phys*. **2011**, *134*, 094118.
Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys*. **2006**, *124*, 114112.
Kang, D.; Dai, J.; Yuan, J. *J. Chem. Phys*. **2011**, *135*, 024505.
Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc*. **2005**, *113*, 133.
Kendall, R. A.; Dunning, T. H.,Jr.; Harrison, R. J. *J. Chem. Phys*. **1995**, *96*, 6796.
Kirchner, B.; Di Dio, P. G.; Hutter, J. *Top. Curr. Chem*. **2012**, *307*, 109.
Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* 1999, *313*, 701.
Klein, R. A. *J. Comp. Chem*. **2003**, *24*, 1120.
Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys*. **2007**, *127*, 074103.
Kobayashi, M.; Nakai, H. *Int. J. Quantum Chem*. **2009**, *109*, 2227.
Kohn, W. "Electronic Structure of Matter – Wave Functions and Density Functionals" Nobel Lecture, January 28, 1999. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/kohn-lecture.html. Accessed May 17, 2012.
Korth, H. G.; de Heer, M. I.; Mulder, P. *J. Phys. Chem. A* **2002**, *106*, 8779.
Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys*. **1980**, *72*, 4244.
Kulmala, M. *Science* **2003**, *302*, 1000.
Kulmala, M.; Riipinen, I.; Sipilä, M.; Manninen, H. E.; Petäjä, T.; Junninen, H.; Dal Maso, M.; Mordas, G.; Mirme, A.; Vana, M.; Hirsikko, A.; Laakso, L.; Harrison, R. M.; Hanson, I.; Leung, C.; Lehtinen, K. E. J.; Kerminen, V.-M. *Science* **2007**, *318*, 89.
Kurtén, T.; Vehkamäki, H. *Advances in Quantum Chemistry* **2008**, *55*, 407.
Larson, L. J.; Largent, A.; Tao, F. *J. Phys. Chem. A* **1999**, *103*, 6786.
Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009**, *113*, 10527.
Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput*. **2009**, *5*, 1573.
Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.
Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc*. **2005**, *127*, 7215.
Li, S.; Ma, J.; Jiang, Y. *J. Comput. Chem*. **2002**, *23*, 237.
Li, S.; Shen, J.; Li, W.; Jiang, Y. *J. Chem. Phys*. **2006**, *125*, 074109.
Li, W.; Li, S. *J. Chem. Phys*. **2005**, *122*, 194109.
Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.

Li, X. P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891.
Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
Maitland, G. C.; Rigby, M.; Smith, E. B.; Wakeham, A. In *Intermolecular Forces: Their Origin and Determination*; Pergamon Press: Oxford, 1987.
Mantina, M.; Valero, R.; Truhlar, D. G. *J. Phys. Chem. A* **2009**, *113*, 5806.
Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378.
Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 2011.
Marzari, N.; Vanderbilt, D. *Phys. Rev. B*, **1997**, *56*, 12847.
Mata, R. *Mol. Phys*. **2010**, *108*, 381.
Mayhall, N. J.; Raghavachari, K. *J. Chem.Theory Comput*. **2011**, *7*, 1336.
Mcgrath, M. J.; Siepmann, J. I.; Kuo, I. F. W.; Mundy, C. *J. Mol. Phys*. **2007**, *105*, 1411.
Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys*. **1953***, 21*, 1087.
Mignon, P.; Loverix, S.; Geerlings, P. *Chem. Phys. Lett*. **2005**, *401*, 40.
Millam, J. M.; Scuseria, G. E. *J. Chem. Phys*. **1997**, *106*, 5569.
Møller, C.; Plesset, M. S. *Phys. Rev*. **1934**, *46*, 618.
Mulliken, R. S. *J. Chem. Phys*. **1955**, *23*, 1833.
Mullin, J. M.; Roskop, L. B.; Pruitt, S. P.; Collins, M. A.; Gordon, M. S. *J. Phys. Chem. A* **2009**, *113*, 10040.
Myers, J. K.; Pace, C. N. *Biophys. J*. **1996**, *71*, 2033.
Nakao, Y.; Hirao, K. *J. Chem. Phys*. **2004**, *120*, 6375.
Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 2046.
Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput*. **2009**, *5*, 1197.
Pedulla, J. M.; Kim, K.; Jordan, K. D. *Chem. Phys. Lett*. **1998**, *291*, 78.
Perutz, M. F. *Phil. Trans. Phys. Sci. Eng*. 1993, *345*, 105.
Pulay, P. *Chem. Phys. Lett*. **1983**, *100*, 151.
Raghavachari, K.; Anderson, J. B. *Chem. Phys. Lett*. **1989**, *157*, 479.
Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett*. **1989**, *157*, 479.
Rappé, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys*. **1998**, *109*, 1223.
Rezac, J.; Salahub, D. R. *J. Chem. Theory Comput*. **2010**, *6*, 91.
Ringer, A. L.; Senenko, A.; Sherrill, C. D. *Prot. Sci*. **2007**, *16*, 2216.
Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc*. **1999**, *121*, 4827.
Rosenfeld, D.; Lohmann, U.; Raga, G. B.; O'Dowd, C. D.; Kulmala, M.; Fuzzi, S.; Reissell, A.; Andreae, M. O. *Science* **2008**, *321*, 1309.
Roux, B. *Biophys. J*., **1996**, *71*, 3177.
Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem*. **1993**, *44*, 213.
Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782.

Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys*. **1999**, *111*, 8330.
Sedo, G.; Schultz, J.; Leopold, K. R. *J. Mol. Spectrosc*. **2007**, *251*, 4.
Sharma, M.; Resta, R.; Car, R. *Phys. Rev. Lett*. **2005**, *95*, 187401.
Sharma, M.; Resta, R.; Car, R. *Phys. Rev. Lett*. **2007**, *98*, 247401.
Shavitt, I. In Schaefer, H. F. I., Ed.; *Methods of Electronic Structure Theory*; Plenum: New York, 1977; pp 189-275.
Shostak, S. L.; Ebenstein, W. L.; Muenter, J. S. *J. Chem. Phys*. **1991**, *94*, 5875.
Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys*. **1999**, *111*, 3572.
Singh, U. C.; Kollman, P. A. *J. Comput. Chem*. **1984**, *5*, 129.
Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*,11085.
Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput*. **2008**, *4*, 683.
Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G., in *Accurate Condensed-Phase Electronic Structure Theory*, Ed. Manby, F. R.; CRC Press, Boca Raton, FL, 2011, p. 105.
Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem*. **1994**, *98*, 11623.
Stoll, H.; Preuss, H. *Theor. Chem. Acc*. **1977**, *46*, 12.
Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput. -Aided Mol. Design* **1995**, *9*, 872.
Suárez, E.; Dias, N.; Suares, D. *J. Chem.Theory Comput*. **2009**, *5*, 1667.
Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys*. **2006**, *125*, 074116.
Sugiki, S. -I.; Kurita, N.; Sengoku, Y.; Sekino, H. *Chem. Phys. Lett*., **2003**, *382*, 611.
Tao, W.-K.; Chen, J.-P.; Li, Z.; Wang, C.; Zhang, C. *Rev. Geophys*. **2012**, *50*, RG2001.
Tauer, T. P.; Derrick, M. E.; Sherrill, C. D. *J. Phys. Chem*. **2005**, *109*, 191.
Tempkin, J. O. B.; Leverentz, H. R.; Wang, B.; Truhlar, D. G. *J. Phys. Chem. Lett*. **2011**, *2*, 2141.
Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. C. *J. Comput. Chem*. **1994**, *15*, 269.
Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem*. **2003**, *24*, 1291.
Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117.
Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z. H. *J. Am. Chem. Soc*. **2010**, *132*, 5137.
Turki, N.; Milet, A.; Ouameralli, O.; Moszynski, R.; Kockanski, E. *THEOCHEM* **2002**, *577*, 239.
Udier-Blagović , M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem*. **2004**, *25*, 1322.
Valiron, P.; Mayer, I. *Chem. Phys. Lett*. **1997**, *275*, 46.
Wang, Y.; Paulus, B. *Chem. Phys. Lett*. **2007**, *441*, 187.
Watanabe, T.; Manz, T. A.; Sholl, D. S. *J. Phys. Chem. C* **2011**, *115*, 4824.
Wen, W.; Beran, G. J. O. *J. Chem.Theory Comput*. **2011**, *7*, 3733.
Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys.*

Chem. A **2002**, *106*, 10707.
Woon, D. E.; Dunning, T. H.,Jr. *J. Chem. Phys*. **1993**, *98*, 1358.
Xantheas, S. S. *J. Chem. Phys*. **1994**, *100*, 7523.
Xie, W.; Gao, J. *J. Chem. Theory Comput*. **2007**, *3*, 1890.
Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys*. **2008**, *128*, 234108.
Yang, W. *Phys. Rev. Lett*. **1991**, *66*, 1438.
Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.
Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys*. **2005**, *123*, 161103.
Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comp*. **2006**, *213*, 364.
Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J Phys Chem B* **2005**, *109*, 19046.
Zhao, Y.; Truhlar, D. G. *J. Chem. Phys*. **2006**, *125*, 194101.
Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput*. **2006**, *2*, 1009.
Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput*. **2007**, *3*, 289.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 6624.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *100*, 5121.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126.
Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008** 112, 6860.
Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys*. **2005**, *7*, 2701.
Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215.
Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput*. **2009**, *5*, 808.

Appendices

Appendix for Chapter 3

Table A.3.1. Cartesian Coordinates (Å) for M06-2X/MG3S Optimized Geometries of the 1_H2O and 2_H2O Complexes

1_H2O			
S	0.766575	0.011671	-0.021124
O	0.036164	0.016361	1.234542
O	0.297149	-1.076632	-0.887655
O	0.839640	1.297030	-0.664024
O	2.261395	-0.419681	0.299534
H	2.789121	0.379418	0.408938
N	-2.101286	-1.304463	-0.005799
H	-1.108054	-1.359129	-0.463468
H	-2.351118	-0.296002	-0.061529
H	-2.791419	-1.896904	-0.451557
H	-1.984847	-1.547038	0.974388
O	-2.077310	1.487154	0.010667
H	-1.441212	1.419281	0.737817
H	-1.524974	1.811042	-0.710518
2_H2O			
S	1.021816	-0.095813	0.068483
O	1.408267	1.295691	-0.024119
O	0.523611	-0.505755	1.368031
O	0.137270	-0.495522	-1.018624
O	2.358068	-0.940920	-0.085420
H	2.912999	-0.521000	-0.752838
N	-2.047166	0.326381	1.121219
H	-1.118415	0.014369	1.485574
H	-1.848908	1.122639	0.468769
H	-2.413138	-0.471198	0.570021
H	-2.679140	0.600620	1.863777
O	-1.160632	2.222171	-0.683933
H	-0.215001	2.201038	-0.449989
H	-1.182715	1.736005	-1.514376
O	-2.248942	-1.732245	-0.688347
H	-2.288556	-2.688257	-0.646314
H	-1.327163	-1.493234	-0.909590

Table A.3.2. Cartesian Coordinates (Å) for the Geometries of Complexes Obtained from Molecular Dynamics Simulations

1A

O	2.4436391	-0.5052160	1.6190444
S	1.2293899	-0.3255168	0.8504504
O	0.1462842	0.3770494	1.5759144
O	1.5217851	0.4576051	-0.4717481
H	1.6411791	1.4873387	-0.2884964
O	0.7946155	-1.7076478	0.2937501
H	-0.1793150	-1.7156486	-0.1840520
H	2.5136698	2.6700869	0.9137617
O	1.9080388	2.8546460	0.0881586
H	1.1725267	3.3852857	0.4387917
O	3.7855367	1.9538762	2.0880870
H	3.4232141	1.0312779	2.1531407
H	4.6353584	1.9156659	1.6084468
H	-4.6022708	-0.7085454	1.6262382
O	-4.6011520	0.2605866	1.5314977
H	-3.9268127	0.5082108	2.2334072
H	-1.2044812	-2.3748944	-4.4344682
O	-0.7495258	-2.2527542	-3.6079615
H	0.1212130	-1.8204254	-3.8422817
O	-2.3915987	1.5412340	2.9893100
H	-1.4407150	1.5309918	2.7763959
H	-2.6826640	2.4664989	2.7954330
H	-1.3873540	-2.0106886	-1.8359292
O	-1.4468388	-2.0507791	-0.8609366
H	-2.1661012	-1.4105719	-0.7390745
H	2.4787226	-1.6880389	-4.6748486
N	1.9933583	-1.1277073	-4.0074981
H	2.2435786	-1.4012667	-3.0325998
H	2.2666527	-0.1533541	-4.1735050

1B

O	-0.9402896	0.9110457	2.1311876
S	-1.1455596	0.8720517	0.6949687
O	-1.8750852	2.0076961	0.1243356
O	0.2146043	0.7173706	-0.0432397
H	1.0583691	1.3099979	0.3519451
O	-1.9580370	-0.4535926	0.2682523
H	-2.0540787	-0.6026633	-0.7163108
H	3.0177382	1.9878972	0.5410816
O	2.1111506	2.4226707	0.7724770
H	2.1411177	2.6019217	1.7556771
O	4.5083604	1.4324468	-0.4161748
H	4.7368010	0.5745682	-0.0647367
H	5.3940671	1.8851148	-0.5433904
H	-6.1068220	1.9986706	-1.8536170
O	-6.3111889	2.3707148	-0.9997611
H	-5.4832058	2.8290549	-0.6582977
H	-0.8974833	-2.2468907	-4.8477415
O	-1.3570479	-1.3899380	-4.8867464
H	-0.9125737	-0.8852927	-5.6339910
O	-4.1020024	3.9883068	-0.5412614
H	-3.3285740	3.4149869	-0.3068118
H	-3.7254321	4.6902144	-1.0488821
H	-1.8321120	-0.9132061	-3.3368911
O	-2.2403904	-0.9420420	-2.3914250
H	-3.0293066	-1.4675771	-2.5266588
H	1.0675469	-0.0029918	-6.1945457
N	0.1438930	0.1860458	-6.6175591
H	-0.1918080	1.1426505	-6.5040113
H	0.2698349	0.0399716	-7.6259323

1C

O	-0.4031948	-0.7526565	1.5905288
S	-1.0590619	0.4219755	1.0363294
O	-2.4662184	0.6214548	1.1851682
O	-0.3530241	1.7225525	1.5907187
H	0.6682073	1.8095297	1.5830543
O	-0.6703606	0.4138737	-0.5046720
H	-1.2974529	0.7676055	-1.1632942
H	2.8579293	2.6315781	0.6806419
O	2.3085216	1.8798668	0.9673235
H	2.4010323	1.2157767	0.2693846
O	4.1610065	4.0914658	1.1283789
H	5.1234883	4.1855723	0.9276997
H	3.7839430	4.9813982	1.0993599
H	-5.5791022	4.1913127	0.8224824
O	-6.2631638	4.8584069	1.0170608
H	-6.0040351	5.6798935	0.6534611
H	-4.2810977	-0.6097891	-0.2325296
O	-4.3985251	-0.7577328	-1.1611894
H	-3.7599494	-1.5173061	-1.3396654
O	-4.0923217	3.3948120	0.1778589
H	-3.6414052	2.7955059	0.7568923
H	-4.2299723	3.0464219	-0.7111569
H	-3.5850848	0.4408801	-1.7648137
O	-2.9238506	1.1139509	-2.0285712
H	-3.0201407	1.0000562	-2.9890747
H	-2.8018749	-3.7995375	-1.4083962
N	-2.5795514	-2.8195988	-1.5004560
H	-2.1709721	-2.4996056	-0.6461697
H	-1.7629060	-2.8034442	-2.1446069

2A

S	-1.5670185	2.3941008	-1.4227842
O	-1.0207933	2.4931556	-0.0527035
H	-0.5444917	0.9816124	1.1814067
O	-2.7932064	3.5149398	-1.5901360
H	-2.4246905	4.3982296	-1.4734859
O	-2.3380619	1.0929489	-1.7143236
O	-0.5513314	2.6138995	-2.4664980
O	0.3978578	-2.8678215	-2.0507904
S	-0.0692625	-1.6163978	-1.5194215
O	-0.9781951	-1.7278195	-0.2794910
O	0.9430590	-0.6208539	-1.2758848
H	3.0298545	-1.8315097	-6.2871521
O	-1.0468885	-1.0682106	-2.6898102
H	-1.5182202	-0.3200704	-2.2582356
H	1.0963688	2.2980869	-3.2454694
O	1.9551366	2.3797980	-3.7691260
H	2.0133939	3.3312382	-3.9449555
O	6.1040225	0.8152107	-2.7810180
H	6.2578854	1.6239460	-2.2354886
H	7.0260355	0.5759673	-3.0220347
H	-4.6768476	-0.6378939	-2.0691645
O	-4.6338381	-0.0306061	-2.8464699
H	-3.8063977	0.4911571	-2.7366757
H	1.5193884	-1.9559100	-5.7485065
O	2.3689986	-1.3999113	-5.7485966
H	3.0589286	-0.3638663	-4.2286178
O	-0.3078290	0.0770134	1.5487746
H	-0.6589168	-0.9824926	0.3753496
H	0.6635772	0.0775849	1.7351319
H	-0.8953247	-2.3888070	-5.0086091
O	0.0028618	-2.8022977	-5.0036518
H	0.2481742	-2.8392783	-4.0577386
H	4.4967489	0.3115834	-3.5024559
N	3.4973753	0.2222373	-3.4649748
H	3.3950656	-0.2340258	-2.5570353
H	3.0282200	1.1820616	-3.4651395

2B

S	-2.5701088	0.5840872	0.2717406
O	-2.6956617	-0.6847224	0.9502530
H	-2.7147646	-3.5157293	0.9664486
O	-3.8181186	1.5735773	1.0504310
H	-3.4949234	2.3073253	1.5531621
O	-3.0636364	0.6377206	-1.1131986
O	-1.2637517	1.2212281	0.5523916
O	-0.8314983	-2.3772154	-4.0005625
S	-1.0602011	-2.2119802	-2.5577165
O	-1.2116685	-3.6565673	-1.8987872
O	-0.1242105	-1.4422741	-1.7139665
H	2.7698201	0.5865497	-5.4238047
O	-2.4842314	-1.5484259	-2.4267908
H	-2.5998932	-0.7114200	-1.9517061
H	-0.0760785	2.1391969	-0.4872390
O	0.6365504	2.4345981	-1.0959670
H	0.2971667	3.1124809	-1.6995732
O	4.9434019	-0.2409216	-1.4597930
H	5.8264034	0.1919564	-1.1929628
H	4.8980428	-1.1346866	-1.0322361
H	-5.0937326	1.6205578	-1.5521732
O	-5.9269001	1.6302637	-1.0635172
H	-5.5457884	1.7098657	-0.1671526
H	1.1705824	0.6274579	-5.1925889
O	2.0815289	0.8718763	-4.8050907
H	2.2806881	0.4170157	-3.0801437
O	-1.8023906	-3.1901084	0.6165618
H	-1.4592493	-3.5067726	-0.8939916
H	-1.8451928	-2.2255586	0.7414796
H	-0.5275591	-1.0320682	-5.4208497
O	-0.1051405	-0.4223784	-6.0277875
H	-0.6997723	-0.3764707	-6.8158730
H	3.0742082	0.2962283	-1.6904986
N	2.1700110	0.4134668	-2.0862867
H	1.5734289	-0.4150360	-1.8217375
H	1.6553164	1.2469447	-1.8019492

2C

S	-5.2389619	1.3411533	1.6894362
O	-4.5248866	1.1929701	2.9748528
H	-2.1544203	0.2197161	4.3219662
O	-6.5768048	2.3259752	1.9866377
H	-6.2988162	2.8605214	2.7778607
O	-5.9439188	0.1106325	1.2213497
O	-4.5594843	2.0102690	0.6342435
O	-3.9821344	-4.3177572	1.7031465
S	-3.9420368	-2.8977297	1.5783444
O	-3.1766199	-2.2994307	2.8900680
O	-3.2685728	-2.3490365	0.4377892
H	-0.9918924	-1.8570614	-1.5843179
O	-5.4223142	-2.3357352	1.6951283
H	-5.5790805	-1.2730556	1.5184256
H	-2.6715486	2.3970033	0.3629485
O	-1.7950509	2.6704003	0.1505729
H	-1.8907922	3.4923395	-0.3304461
O	2.3790105	0.8515926	1.6340157
H	3.0928905	0.1508973	1.7339624
H	2.9724422	1.5847319	1.2681108
H	-7.7942892	-0.3737331	1.5741226
O	-8.6506988	-0.6350276	2.0154379
H	-8.4121823	-1.3786127	2.5457446
H	-1.7377865	-1.8028333	-0.1964599
O	-0.9355322	-1.5124850	-0.6974749
H	-0.5492642	-0.2836123	0.4075198
O	-2.1262130	0.1669118	3.3519533
H	-2.9899672	-1.3523074	2.6813302
H	-2.9847343	0.6708462	3.0699546
H	-1.8822447	-2.0857818	-4.2563499
O	-1.0158203	-1.9265848	-3.8245784
H	-0.3347959	-2.5197881	-4.2804623
H	0.7852776	0.5578568	1.1428906
N	-0.2557994	0.5373634	1.0289769
H	-0.6557562	0.3335937	1.9656598
H	-0.6937965	1.3938970	0.6969329

Appendix for Chapter 5

Table A.5.1. Dipole Moments and Errors for Forty-Four Configurations of $(NH_3)(H_2O)_{11}$

Step	Conventional Dipole Moment (D)	Error in EE-1B Dipole Moment (D)	Error in EE-PA Dipole Moment (D)	Error in EE-3B Dipole Moment (D)
1	13.967	-1.085	-0.132	0.091
2	13.819	-1.065	-0.126	0.079
3	14.079	-1.056	-0.111	0.074
4	14.107	-1.060	-0.112	0.072
5	14.105	-1.058	-0.114	0.079
6	14.550	-1.354	-0.105	0.085
7	14.177	-1.088	-0.112	0.072
8	14.197	-1.099	-0.113	0.074
9	14.158	-1.097	-0.114	0.072
10	13.704	-1.107	-0.108	0.086
11	13.711	-1.114	-0.105	0.080
12	13.703	-1.111	-0.106	0.084
13	13.988	-1.113	-0.108	0.090
14	13.985	-1.112	-0.107	0.085
15	13.935	-1.090	-0.098	0.082
16	13.975	-1.112	-0.107	0.090
17	13.878	-1.074	-0.093	0.084
18	14.093	-1.065	-0.091	0.081
19	13.811	-1.062	-0.092	0.086
20	14.049	-1.186	-0.090	0.083
21	14.130	-1.039	-0.097	0.076
22	13.881	-1.078	-0.093	0.081
23	13.875	-1.079	-0.092	0.078
24	13.840	-1.063	-0.091	0.089
25	13.695	-0.975	-0.081	0.062
26	13.745	-0.989	-0.085	0.072
27	13.708	-0.974	-0.086	0.079
28	13.742	-0.972	-0.085	0.078
29	13.393	-0.864	-0.089	0.081
30	13.738	-0.969	-0.084	0.073
31	13.152	-0.831	-0.083	0.062
32	13.790	-0.969	-0.082	0.071
33	13.778	-0.962	-0.084	0.078
34	13.850	-0.969	-0.084	0.077
35	13.833	-0.961	-0.087	0.076
36	13.807	-0.954	-0.085	0.074
37	13.798	-0.944	-0.085	0.070
38	13.908	-0.913	-0.088	0.079

39	13.912	-0.915	-0.086	0.078
40	14.082	-0.921	-0.086	0.080
41	13.910	-0.821	-0.092	0.083
42	14.186	-0.786	-0.096	0.084
43	13.820	-0.825	-0.091	0.081
44	13.752	-0.800	-0.094	0.070
MSE		-1.015	-0.097	0.079
MUE		1.015	0.097	0.079
RMSE		1.021	0.097	0.079
MUPE		7.3	0.7	0.6

Table A.5.2. Dipole Moments and Errors for Forty-Four Configurations of $(NH_3)_2(H_2O)_{14}$

Step	Conventional Dipole Moment (D)	Error in EE-1B Dipole Moment (D)	Error in EE-PA Dipole Moment (D)	Error in EE-3B Dipole Moment (D)
1	18.354	-0.454	-0.205	0.081
2	18.372	-0.431	-0.206	0.080
3	18.345	-0.421	-0.208	0.095
4	18.423	-0.416	-0.211	0.093
5	18.355	-0.380	-0.228	0.091
6	18.375	-0.403	-0.208	0.070
7	18.354	-0.386	-0.220	0.090
8	17.902	-0.436	-0.198	0.084
9	17.951	-0.458	-0.194	0.089
10	18.095	-0.504	-0.181	0.075
11	18.333	-0.487	-0.189	0.082
12	18.163	-0.530	-0.191	0.056
13	18.085	-0.500	-0.181	0.061
14	18.148	-0.511	-0.176	0.068
15	18.131	-0.510	-0.179	0.089
16	18.501	-0.487	-0.180	0.081
17	18.132	-0.510	-0.176	0.066
18	18.147	-0.504	-0.177	0.081
19	18.208	-0.545	-0.186	0.088
20	18.043	-0.470	-0.192	0.090
21	17.999	-0.522	-0.183	0.063
22	18.008	-0.521	-0.184	0.061
23	18.071	-0.539	-0.182	0.068
24	17.949	-0.553	-0.185	0.086
25	18.124	-0.540	-0.187	0.079
26	18.240	-0.554	-0.186	0.059
27	18.254	-0.559	-0.186	0.072
28	18.188	-0.558	-0.186	0.075
29	18.403	-0.528	-0.193	0.070
30	17.729	-0.567	-0.178	0.092
31	17.733	-0.569	-0.176	0.081
32	17.514	-0.454	-0.181	0.096
33	17.774	-0.568	-0.174	0.065
34	17.633	-0.550	-0.171	0.069
35	17.534	-0.483	-0.159	0.060
36	16.801	-0.491	-0.168	0.069
37	17.542	-0.481	-0.161	0.075
38	17.540	-0.480	-0.160	0.075
39	17.499	-0.478	-0.162	0.058
40	17.678	-0.498	-0.147	0.058

41	17.333	-0.482	-0.162	0.067
42	17.617	-0.544	-0.163	0.086
43	17.444	-0.437	-0.170	0.073
44	17.542	-0.489	-0.158	0.064
MSE		-0.495	-0.183	0.076
MUE		0.495	0.183	0.076
RMSE		0.498	0.184	0.077
MUPE		2.8	1.0	0.4

Table A.5.3a. Dipole Moments and EE-MB Errors for Six Configurations of $[Cl(H_2O)_6]^-$

System	Conventional μ (D)	Error in EE-1B (S1) μ (D)	Error in EE-1B (S2) μ (D)	Error in EE-PA (S1) μ (D)	Error in EE-PA (S2) μ (D)	Error in EE-3B (S1) μ (D)	Error in EE-3B (S2) μ (D)
int1	1.905	-0.289	-0.098	-0.247	-0.241	-0.182	-0.202
int2	1.654	-0.042	-0.198	-0.107	-0.066	-0.010	0.014
int3	8.242	-0.186	-0.075	0.015	-0.006	0.018	-0.006
surf1	8.153	-0.872	-0.724	-0.075	-0.102	-0.011	-0.020
surf2	7.870	-0.352	-0.110	0.088	0.009	-0.003	0.005
surf3	10.330	-0.808	-0.796	0.076	-0.017	0.013	0.016
MSE		-0.425	-0.333	-0.042	-0.070	-0.029	-0.032
MUE		0.425	0.333	0.102	0.073	0.039	0.044
RMSE		0.526	0.452	0.124	0.110	0.075	0.083
MUPE		7.2	6.0	3.7	3.0	1.8	2.0

Table A.5.3b. Dipole Moments and MB Errors for Six Configurations of $[Cl(H_2O)_6]^-$

System	Conventional μ (D)	Error in 1B (S1) μ (D)	Error in 1B (S2) μ (D)	Error in PA (S1) μ (D)	Error in PA (S2) μ (D)	Error in 3B (S1) μ (D)	Error in 3B (S2) μ (D)
int1	1.905	-0.072	1.143	-0.256	-0.324	-0.251	-0.203
int2	1.654	0.503	-0.139	-0.257	-0.003	0.025	-0.007
int3	8.242	-1.163	-0.207	0.050	-0.170	-0.004	0.014
surf1	8.153	-2.346	-1.767	-0.212	-0.354	0.046	0.030
surf2	7.870	-2.113	-0.709	0.069	-0.130	0.063	0.019
surf3	10.330	-3.403	-2.790	0.191	-0.031	0.064	0.077
MSE		-1.432	-0.745	-0.069	-0.169	-0.010	-0.012
MUE		1.600	1.126	0.172	0.169	0.075	0.058
RMSE		1.965	1.460	0.191	0.215	0.111	0.090
MUPE		22.8	21.4	5.8	4.3	2.8	2.1

Table A.5.4. M06-2X/cc-pV(T+d)+ Dipole Moments (D) of $(HF)_m$ Clusters and Average MB and EE-MB Errors

Cluster	Full	1B	PA	3B	EE-1B	EE-PA	EE-3B
hf3_fff	5.523	5.684	5.522	5.523	5.527	5.523	5.523
hf3_ffh	1.821	1.897	1.821	1.820	1.824	1.821	1.820
hf3_fhf	1.874	1.896	1.871	1.874	1.873	1.872	1.874
hf4_ffff	7.329	7.578	7.327	7.329	7.333	7.329	7.329
hf4_fffh	3.621	3.784	3.621	3.621	3.627	3.623	3.621
hf4_ffhf	3.686	3.789	3.682	3.687	3.688	3.685	3.686
hf4_ffhh	0.004	0.004	0.004	0.004	0.004	0.004	0.004
hf4_fhfh	0.004	0.004	0.004	0.004	0.004	0.004	0.004
hf4_fhhf	0.069	0.005	0.067	0.069	0.063	0.067	0.069
hf5_fffff	9.133	9.472	9.130	9.133	9.139	9.134	9.133
hf5_ffffh	5.435	5.685	5.433	5.435	5.441	5.436	5.435
hf5_fffhf	5.493	5.682	5.488	5.493	5.496	5.492	5.493
hf5_ffhff	5.496	5.683	5.492	5.497	5.501	5.497	5.496
hf5_fffhh	1.805	1.890	1.804	1.805	1.806	1.805	1.805
hf5_ffhfh	1.816	1.897	1.815	1.816	1.819	1.817	1.816
hf5_fhffh	1.797	1.896	1.795	1.797	1.801	1.797	1.797
hf5_fhfhf	1.871	1.895	1.867	1.872	1.869	1.869	1.871
hf5_fhhff	1.883	1.889	1.882	1.884	1.882	1.883	1.884
hf5_fhhhf	1.730	1.891	1.729	1.730	1.736	1.732	1.730
MSE[a] (D)		0.112	-0.002	0.000	0.002	0.000	0.000
MUE[b] (D)		0.119	0.002	0.000	0.003	0.001	0.000
RMSE[c] (D)		0.151	0.002	0.000	0.004	0.001	0.000
MUPE[d] (%)		8.4	0.3	0.1	0.7	0.3	0.1

[a]MSE = mean signed error over all $(HF)_m$ clusters for m = 3–5.

[b]MUE = mean unsigned error over all $(HF)_m$ clusters for m = 3–5.

[c]RMSE = root mean squared error over all $(HF)_m$ clusters for m = 3–5.

[d]MUPE = mean unsigned percent error over all $(HF)_m$ clusters for m = 3–5.

Table A.5.5. M06-2X/cc-pV(T+d)+ Dipole Moments (D) of $(HF)_m(H_2O)_n$ Clusters and Average MB and EE-MB Errors

Cluster	Full	1B	PA	3B	EE-1B	EE-PA	EE-3B
whf2_BaCa	4.449	4.286	4.453	4.449	4.441	4.449	4.449
whf2_BaCr	3.838	3.609	3.834	3.838	3.820	3.838	3.838
whf2_BrCr	0.398	0.320	0.409	0.398	0.394	0.399	0.398
whf3_AaBaCa	6.663	6.183	6.649	6.663	6.601	6.663	6.663
whf3_AaBaCr	5.315	4.914	5.312	5.316	5.271	5.314	5.315
whf3_AaBrCr	1.748	1.577	1.741	1.748	1.723	1.746	1.748
whf3_ArBaCa	2.547	2.389	2.541	2.547	2.543	2.544	2.547
whf3_ArBaCr	3.271	3.016	3.263	3.271	3.253	3.271	3.271
whf3_ArBrCr	2.367	2.217	2.365	2.368	2.339	2.371	2.367
w2hf3_a	8.794	8.178	8.788	8.797	8.669	8.782	8.797
w2hf3_opt	11.392	8.717	11.254	11.425	10.382	11.342	11.409
MSE[a] (D)		-0.489	-0.016	0.003	-0.122	-0.006	0.002
MUE[b] (D)		0.489	0.018	0.004	0.122	0.007	0.002
RMSE[c] (D)		0.861	0.042	0.010	0.308	0.015	0.005
MUPE[d] (%)		9.5	0.5	0.0	1.5	0.1	0.0

[a]MSE = mean signed error over all $(HF)_m(H_2O)_n$ clusters shown in the table.

[b]MUE = mean unsigned error over all $(HF)_m(H_2O)_n$ clusters shown in the table.

[c]RMSE = root mean squared error over all $(HF)_m(H_2O)_n$ clusters shown in the table.

[d]MUPE = mean unsigned percent error over all $(HF)_m(H_2O)_n$ clusters shown in the table.

Table A.5.6a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int1

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.563	-0.437	-0.424	-0.289	-0.276	-0.223	-0.123
2	O	-0.788	-0.019	-0.020	-0.075	-0.058	0.053	0.000
3	H	0.344	0.040	0.029	0.074	0.055	-0.033	0.016
4	H	0.419	0.003	0.015	0.020	0.021	-0.013	0.012
5	O	-0.795	-0.027	-0.017	-0.065	-0.052	0.036	0.009
6	H	0.337	0.089	0.081	0.080	0.077	-0.034	-0.010
7	H	0.412	-0.016	-0.019	-0.002	-0.008	-0.020	-0.005
8	O	-0.712	-0.106	-0.099	-0.042	-0.037	-0.062	-0.025
9	H	0.346	0.021	0.022	-0.033	-0.028	0.041	0.018
10	H	0.302	0.148	0.141	0.108	0.107	0.063	0.018
11	O	-0.768	-0.095	-0.136	-0.146	-0.069	-0.053	-0.061
12	H	0.384	0.005	-0.003	0.000	0.019	0.025	0.016
13	H	0.264	0.210	0.246	0.267	0.144	0.132	0.084
14	O	-0.764	-0.094	-0.083	-0.073	-0.055	0.032	0.040
15	H	0.378	0.015	0.014	-0.008	-0.016	-0.011	-0.020
16	H	0.266	0.199	0.189	0.126	0.118	0.018	-0.002
17	O	-0.810	-0.036	-0.025	-0.116	-0.098	-0.006	-0.001
18	H	0.377	0.009	0.008	0.027	0.025	0.021	-0.002
19	H	0.369	0.090	0.081	0.145	0.131	0.033	0.035
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.087	0.087	0.089	0.073	0.048	0.026
	RMSE		0.135	0.135	0.119	0.096	0.069	0.041
	MUPE		21.7	21.7	21.6	17.7	11.1	6.1

Table A.5.6b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int1

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.563	-0.437	-0.350	-0.004	-0.049	-0.138	-0.139
2	O	-0.788	0.115	0.115	-0.047	-0.054	0.055	0.025
3	H	0.344	-0.010	-0.010	0.033	0.025	-0.040	-0.004
4	H	0.419	-0.080	-0.080	-0.015	0.030	0.014	-0.002
5	O	-0.795	0.106	0.106	-0.072	-0.054	0.039	-0.004
6	H	0.337	0.008	0.008	0.061	0.068	-0.040	-0.017
7	H	0.412	-0.068	-0.068	0.008	-0.005	-0.011	0.022
8	O	-0.712	0.001	0.001	-0.015	-0.014	-0.040	-0.010
9	H	0.346	0.009	0.009	-0.037	-0.025	0.046	0.010
10	H	0.302	0.054	0.054	0.046	0.054	0.019	0.011
11	O	-0.768	0.081	-0.090	-0.097	-0.002	-0.001	-0.076
12	H	0.384	-0.041	-0.059	-0.015	0.006	0.015	0.011
13	H	0.264	0.080	0.182	0.166	0.027	0.028	0.127
14	O	-0.764	0.058	0.058	-0.065	-0.043	0.021	0.007
15	H	0.378	-0.024	-0.024	-0.024	-0.027	0.022	0.016
16	H	0.266	0.086	0.086	0.090	0.081	-0.016	-0.011
17	O	-0.810	0.113	0.113	-0.064	-0.029	0.018	-0.013
18	H	0.377	-0.029	-0.029	0.002	0.000	0.021	0.014
19	H	0.369	-0.021	-0.021	0.049	0.012	-0.013	0.035
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.075	0.077	0.048	0.032	0.032	0.029
	RMSE		0.119	0.110	0.062	0.039	0.043	0.049
	MUPE		15.4	16.9	11.9	7.8	7.0	7.0

Table A.5.7a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int2

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.751	-0.249	-0.228	-0.061	0.001	-0.045	-0.102
2	O	-0.839	0.059	0.062	-0.023	-0.009	0.024	0.014
3	H	0.377	-0.008	-0.010	0.024	0.016	-0.010	-0.008
4	H	0.450	-0.038	-0.040	-0.007	-0.014	0.015	0.012
5	O	-0.831	-0.039	-0.032	-0.052	-0.021	-0.051	-0.080
6	H	0.368	0.113	0.105	0.097	0.068	0.076	0.089
7	H	0.400	-0.011	-0.010	0.000	-0.008	0.005	0.028
8	O	-0.840	0.005	-0.010	-0.010	0.066	0.034	-0.080
9	H	0.421	0.056	0.075	0.080	-0.066	-0.070	0.072
10	H	0.381	-0.023	-0.047	-0.046	-0.022	0.013	0.049
11	O	-0.816	-0.027	-0.013	-0.073	-0.055	0.061	0.038
12	H	0.319	0.122	0.113	0.054	0.049	-0.080	-0.065
13	H	0.379	0.022	0.017	0.024	0.013	0.003	0.013
14	O	-0.712	-0.105	-0.097	-0.024	0.001	0.026	0.019
15	H	0.389	0.062	0.053	0.004	-0.039	-0.022	0.003
16	H	0.335	0.032	0.032	-0.012	-0.008	-0.013	-0.020
17	O	-0.823	-0.010	0.001	-0.034	-0.019	-0.027	-0.027
18	H	0.363	0.006	0.004	-0.036	-0.037	0.030	0.030
19	H	0.432	0.033	0.024	0.096	0.084	0.029	0.018
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.054	0.051	0.040	0.031	0.033	0.040
	RMSE		0.079	0.074	0.050	0.040	0.041	0.050
	MUPE		10.9	10.5	8.4	6.9	6.9	8.1

Table A.5.7b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int2

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.751	-0.249	-0.171	0.129	0.097	-0.016	-0.040
2	O	-0.839	0.168	0.168	0.010	0.024	-0.007	-0.026
3	H	0.377	-0.041	-0.041	-0.008	-0.012	0.005	0.018
4	H	0.450	-0.114	-0.114	-0.032	-0.035	0.041	0.033
5	O	-0.831	0.132	0.132	-0.009	0.010	-0.017	-0.031
6	H	0.368	-0.019	-0.019	0.025	0.007	0.007	0.019
7	H	0.400	-0.050	-0.050	-0.020	-0.016	0.013	0.016
8	O	-0.840	0.129	-0.035	-0.018	0.063	0.045	-0.015
9	H	0.421	-0.065	0.034	0.008	-0.070	-0.069	0.022
10	H	0.381	-0.025	-0.038	-0.022	-0.010	0.011	0.000
11	O	-0.816	0.118	0.118	-0.096	-0.065	0.066	0.029
12	H	0.319	0.029	0.029	0.056	0.038	-0.070	-0.024
13	H	0.379	-0.029	-0.029	0.046	0.034	-0.006	-0.004
14	O	-0.712	0.017	0.017	-0.024	-0.006	0.033	0.017
15	H	0.389	-0.041	-0.041	0.008	-0.005	-0.047	-0.026
16	H	0.335	0.013	0.013	-0.023	-0.022	-0.006	-0.005
17	O	-0.823	0.122	0.122	-0.031	-0.003	0.013	-0.003
18	H	0.363	-0.012	-0.012	-0.014	-0.019	0.015	0.014
19	H	0.432	-0.081	-0.081	0.015	-0.008	-0.010	0.006
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.077	0.067	0.031	0.029	0.026	0.018
	RMSE		0.099	0.084	0.044	0.039	0.034	0.021
	MUPE		12.9	11.6	6.0	5.5	5.5	3.7

Table A.5.8a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int3

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.714	-0.286	-0.256	-0.098	-0.057	0.025	0.037
2	O	-0.844	-0.037	-0.048	-0.116	-0.040	-0.018	-0.010
3	H	0.356	0.114	0.106	0.147	0.014	-0.013	-0.003
4	H	0.448	-0.037	-0.047	0.024	0.030	0.037	0.006
5	O	-0.793	-0.048	-0.044	-0.121	-0.111	0.064	0.028
6	H	0.403	0.017	0.025	0.056	0.074	-0.010	-0.009
7	H	0.363	0.057	0.046	0.088	0.067	-0.043	-0.004
8	O	-0.856	0.055	0.059	-0.020	-0.002	0.118	0.075
9	H	0.485	-0.047	-0.049	-0.009	-0.024	-0.083	-0.049
10	H	0.373	-0.009	-0.010	0.015	0.009	-0.070	-0.049
11	O	-0.876	0.033	0.056	-0.079	-0.026	0.015	0.011
12	H	0.388	0.084	0.068	0.048	0.027	-0.041	-0.014
13	H	0.428	-0.056	-0.064	0.020	0.001	0.012	0.004
14	O	-0.865	-0.031	-0.020	-0.068	-0.047	0.036	0.013
15	H	0.398	0.006	0.004	0.004	-0.004	0.017	0.010
16	H	0.401	0.091	0.082	0.115	0.097	-0.072	-0.046
17	O	-0.798	-0.012	-0.005	-0.087	-0.059	0.020	0.017
18	H	0.341	0.131	0.122	0.051	0.028	0.013	-0.021
19	H	0.366	-0.027	-0.025	0.031	0.023	-0.006	0.002
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.062	0.060	0.063	0.039	0.038	0.021
	RMSE		0.088	0.082	0.076	0.050	0.048	0.029
	MUPE		12.8	12.2	12.3	7.6	7.3	4.1

Table A.5.8b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Int3

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.714	-0.286	-0.203	0.123	0.069	0.029	0.063
2	O	-0.844	0.146	0.015	-0.080	-0.015	0.000	-0.010
3	H	0.356	-0.007	0.073	0.082	0.014	0.003	0.008
4	H	0.448	-0.098	-0.131	0.006	0.026	0.039	-0.002
5	O	-0.793	0.100	0.100	-0.075	-0.062	0.109	0.050
6	H	0.403	-0.056	-0.056	0.004	0.030	-0.025	-0.016
7	H	0.363	-0.017	-0.017	0.042	0.021	-0.099	-0.041
8	O	-0.856	0.162	0.162	-0.021	0.003	0.124	0.055
9	H	0.485	-0.137	-0.137	-0.007	-0.026	-0.097	-0.042
10	H	0.373	-0.025	-0.025	0.009	0.005	-0.051	-0.026
11	O	-0.876	0.179	0.179	-0.065	0.008	0.006	0.005
12	H	0.388	-0.039	-0.039	-0.005	-0.066	-0.034	-0.017
13	H	0.428	-0.079	-0.079	0.011	-0.001	0.024	0.011
14	O	-0.865	0.163	0.163	-0.056	-0.035	0.092	0.049
15	H	0.398	-0.047	-0.047	-0.009	-0.002	-0.002	-0.005
16	H	0.401	-0.050	-0.050	0.087	0.063	-0.110	-0.072
17	O	-0.798	0.085	0.085	-0.048	-0.043	0.043	0.042
18	H	0.341	0.015	0.015	-0.003	-0.003	-0.052	-0.043
19	H	0.366	-0.008	-0.008	0.006	0.014	0.001	-0.009
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.089	0.083	0.039	0.027	0.049	0.030
	RMSE		0.115	0.103	0.053	0.035	0.064	0.037
	MUPE		14.5	14.4	6.9	5.2	9.7	5.8

Table A.5.9a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf1

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.738	-0.262	-0.169	-0.014	-0.039	-0.052	-0.034
2	O	-0.946	0.001	0.002	-0.016	0.000	0.096	0.020
3	H	0.439	0.027	0.014	0.084	0.054	-0.004	0.046
4	H	0.513	-0.034	-0.021	-0.030	-0.017	-0.049	0.020
5	O	-0.821	-0.067	-0.053	-0.095	-0.040	0.001	0.019
6	H	0.322	0.187	0.176	0.060	0.055	-0.006	-0.030
7	H	0.398	-0.018	-0.021	0.039	0.007	-0.005	-0.001
8	O	-0.870	-0.012	-0.001	-0.099	-0.046	0.057	-0.006
9	H	0.434	0.070	0.060	0.107	0.080	-0.073	-0.001
10	H	0.384	-0.005	-0.007	0.028	-0.003	0.004	0.016
11	O	-1.099	0.151	0.152	0.054	0.045	-0.024	0.025
12	H	0.477	-0.020	-0.021	-0.055	-0.055	0.035	-0.003
13	H	0.590	-0.099	-0.099	-0.049	-0.036	-0.027	-0.049
14	O	-0.868	-0.004	-0.002	0.030	0.139	0.050	-0.108
15	H	0.389	0.107	0.025	0.018	-0.058	-0.030	0.003
16	H	0.402	-0.026	-0.039	-0.035	-0.057	-0.003	0.066
17	O	-0.875	0.069	0.069	0.059	0.061	-0.044	-0.025
18	H	0.491	-0.043	-0.044	-0.070	-0.081	0.051	0.034
19	H	0.379	-0.021	-0.020	-0.018	-0.010	0.024	0.010
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.064	0.052	0.050	0.046	0.033	0.027
	RMSE		0.094	0.076	0.058	0.056	0.042	0.037
	MUPE		12.1	9.8	9.4	8.5	5.6	4.8

Table A.5.9b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf1

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.738	-0.262	-0.168	0.085	0.051	-0.043	-0.008
2	O	-0.946	0.225	0.225	-0.017	-0.001	0.098	0.042
3	H	0.439	-0.079	-0.079	0.052	0.029	-0.035	0.009
4	H	0.513	-0.151	-0.151	-0.024	0.005	-0.029	-0.016
5	O	-0.821	0.072	0.072	-0.036	0.004	-0.029	-0.013
6	H	0.322	0.053	0.053	0.042	0.042	0.001	-0.027
7	H	0.398	-0.023	-0.023	-0.006	-0.023	0.016	0.013
8	O	-0.870	0.122	0.122	-0.030	0.007	0.044	-0.002
9	H	0.434	-0.060	-0.060	0.008	-0.031	-0.069	-0.001
10	H	0.384	-0.010	-0.010	-0.008	-0.012	0.007	0.003
11	O	-1.099	0.371	0.371	0.018	0.006	-0.025	0.007
12	H	0.477	-0.113	-0.113	-0.036	-0.028	0.037	0.004
13	H	0.590	-0.226	-0.226	-0.037	-0.022	-0.013	-0.032
14	O	-0.868	0.162	0.026	0.037	0.065	0.005	-0.023
15	H	0.389	-0.036	0.050	0.014	-0.065	-0.032	-0.007
16	H	0.402	-0.050	-0.093	-0.050	-0.022	0.025	0.031
17	O	-0.875	0.153	0.153	0.024	0.023	-0.022	0.001
18	H	0.491	-0.130	-0.130	-0.027	-0.023	0.041	0.010
19	H	0.379	-0.018	-0.018	-0.009	-0.004	0.023	0.009
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.122	0.113	0.029	0.024	0.031	0.014
	RMSE		0.153	0.143	0.035	0.031	0.038	0.018
	MUPE		18.6	17.8	5.4	4.9	5.5	2.6

Table A.5.10a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf2

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.712	-0.288	-0.205	-0.053	-0.015	-0.009	-0.005
2	O	-0.810	-0.021	-0.013	-0.116	-0.047	0.080	0.016
3	H	0.365	0.118	0.109	0.109	0.010	-0.024	0.025
4	H	0.363	-0.014	-0.013	0.036	0.025	-0.037	-0.016
5	O	-0.903	0.015	0.035	-0.092	-0.039	0.139	0.053
6	H	0.446	-0.047	-0.054	0.061	0.027	-0.082	-0.027
7	H	0.384	0.104	0.092	-0.001	-0.011	-0.078	-0.034
8	O	-0.911	0.034	0.041	-0.099	-0.108	-0.008	0.042
9	H	0.462	-0.047	-0.040	0.015	0.071	0.039	-0.019
10	H	0.457	0.005	-0.009	0.117	0.088	-0.001	-0.022
11	O	-0.958	0.061	0.063	-0.059	-0.063	0.021	0.011
12	H	0.438	0.005	-0.004	0.060	0.044	-0.040	-0.023
13	H	0.441	0.013	0.020	-0.001	0.029	-0.009	-0.003
14	O	-0.794	-0.102	-0.100	-0.120	-0.036	0.067	-0.006
15	H	0.446	-0.032	-0.049	0.019	0.046	-0.001	-0.001
16	H	0.263	0.218	0.150	0.140	-0.011	-0.067	0.016
17	O	-0.941	0.092	0.090	0.013	-0.001	-0.013	0.019
18	H	0.420	-0.034	-0.032	0.004	0.014	0.019	-0.012
19	H	0.545	-0.082	-0.082	-0.032	-0.021	0.005	-0.014
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.070	0.063	0.060	0.037	0.039	0.019
	RMSE		0.101	0.081	0.075	0.046	0.053	0.023
	MUPE		14.7	12.8	11.9	6.7	7.7	3.7

Table A.5.10b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf2

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.712	-0.288	-0.202	0.079	0.058	-0.035	0.014
2	O	-0.810	0.094	0.094	-0.054	-0.044	0.046	0.036
3	H	0.365	-0.006	-0.006	-0.008	-0.009	0.016	0.004
4	H	0.363	-0.005	-0.005	0.016	0.024	-0.026	-0.024
5	O	-0.903	0.181	0.181	-0.082	-0.014	0.089	0.057
6	H	0.446	-0.086	-0.086	0.041	0.015	-0.045	-0.022
7	H	0.384	-0.023	-0.023	-0.011	-0.054	-0.030	-0.050
8	O	-0.911	0.192	0.192	-0.102	-0.063	0.080	0.052
9	H	0.462	-0.103	-0.103	0.016	0.017	-0.037	-0.028
10	H	0.457	-0.097	-0.097	0.061	0.016	-0.053	-0.029
11	O	-0.958	0.237	0.237	-0.138	-0.116	0.113	0.063
12	H	0.438	-0.077	-0.077	0.081	0.056	-0.079	-0.036
13	H	0.441	-0.081	-0.081	0.046	0.059	-0.048	-0.035
14	O	-0.794	0.088	-0.073	-0.104	0.006	0.102	-0.022
15	H	0.446	-0.093	-0.120	0.009	0.010	-0.036	0.011
16	H	0.263	0.090	0.192	0.137	0.020	-0.051	0.026
17	O	-0.941	0.197	0.197	-0.026	-0.032	0.041	0.041
18	H	0.420	-0.048	-0.048	0.018	0.023	-0.008	-0.016
19	H	0.545	-0.173	-0.173	0.020	0.028	-0.037	-0.041
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.114	0.115	0.055	0.035	0.051	0.032
	RMSE		0.136	0.135	0.069	0.044	0.058	0.036
	MUPE		18.7	20.4	10.2	6.2	9.1	5.8

Table A.5.11a. Conventional CHelpG Atomic Charges (in *e*) and EE-MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf3

#	Atom	Conventional Charge	Error in EE-1B (S1)	Error in EE-1B (S2)	Error in EE-PA (S1)	Error in EE-PA (S2)	Error in EE-3B (S1)	Error in EE-3B (S2)
1	Cl	-0.710	-0.290	-0.195	0.007	0.049	0.015	-0.028
2	O	-0.904	-0.016	-0.013	0.003	0.017	0.050	-0.004
3	H	0.471	-0.008	-0.011	0.004	-0.016	-0.037	-0.019
4	H	0.473	-0.016	-0.017	0.024	0.026	-0.003	-0.003
5	O	-0.853	0.018	0.018	-0.120	-0.128	-0.029	0.034
6	H	0.400	-0.026	-0.017	0.013	0.048	0.041	-0.008
7	H	0.380	0.081	0.072	0.049	0.032	0.005	-0.011
8	O	-0.847	0.004	0.003	-0.052	-0.074	0.083	0.063
9	H	0.436	-0.023	-0.009	-0.023	0.031	-0.016	-0.021
10	H	0.393	0.037	0.025	0.090	0.073	-0.088	-0.067
11	O	-0.820	-0.049	-0.050	-0.049	-0.008	0.005	0.004
12	H	0.398	-0.023	-0.044	-0.031	0.045	0.053	-0.049
13	H	0.304	0.190	0.118	0.094	-0.068	-0.095	0.061
14	O	-0.875	-0.016	-0.005	-0.152	-0.085	0.078	0.070
15	H	0.405	-0.012	-0.013	0.076	0.048	-0.030	-0.032
16	H	0.390	0.108	0.098	0.087	0.022	-0.043	-0.013
17	O	-0.855	0.050	0.050	0.026	0.018	-0.016	-0.060
18	H	0.471	-0.023	-0.024	-0.023	-0.025	0.019	0.062
19	H	0.343	0.014	0.015	-0.020	-0.004	0.008	0.021
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.053	0.042	0.050	0.043	0.038	0.033
	RMSE		0.088	0.064	0.065	0.053	0.047	0.041
	MUPE		11.3	8.9	9.9	8.3	7.8	6.6

Table A.5.11b. Conventional CHelpG Atomic Charges (in *e*) and MB Errors (in *e*) for $[Cl(H_2O)_6]^-$ Configuration Surf3

#	Atom	Conventional Charge	Error in 1B (S1)	Error in 1B (S2)	Error in PA (S1)	Error in PA (S2)	Error in 3B (S1)	Error in 3B (S2)
1	Cl	-0.710	-0.290	-0.194	0.070	0.061	-0.037	-0.034
2	O	-0.904	0.198	0.198	-0.009	-0.004	0.009	-0.025
3	H	0.471	-0.118	-0.118	0.015	0.004	0.000	0.021
4	H	0.473	-0.119	-0.119	0.020	0.033	-0.001	-0.002
5	O	-0.853	0.172	0.172	-0.040	-0.058	-0.025	0.011
6	H	0.400	-0.059	-0.059	-0.024	0.006	0.045	0.006
7	H	0.380	-0.039	-0.039	0.018	0.021	-0.009	-0.013
8	O	-0.847	0.165	0.165	-0.039	-0.013	0.034	0.001
9	H	0.436	-0.095	-0.095	-0.007	-0.003	-0.015	0.003
10	H	0.393	-0.052	-0.052	0.038	0.022	-0.037	-0.009
11	O	-0.820	0.127	-0.007	-0.022	0.025	0.036	-0.023
12	H	0.398	-0.051	-0.099	-0.046	0.032	0.032	-0.039
13	H	0.304	0.043	0.129	0.076	-0.078	-0.089	0.068
14	O	-0.875	0.185	0.185	-0.106	-0.064	0.078	0.033
15	H	0.405	-0.060	-0.060	0.038	0.020	-0.041	-0.015
16	H	0.390	-0.045	-0.045	0.029	0.002	0.001	0.001
17	O	-0.855	0.147	0.147	0.021	0.019	-0.031	-0.041
18	H	0.471	-0.117	-0.117	-0.019	-0.019	0.028	0.039
19	H	0.343	0.011	0.011	-0.014	-0.006	0.023	0.018
	MSE		0.000	0.000	0.000	0.000	0.000	0.000
	MUE		0.110	0.106	0.034	0.026	0.030	0.021
	RMSE		0.130	0.121	0.042	0.034	0.038	0.027
	MUPE		18.3	19.0	6.8	5.0	6.2	4.4

Table A.5.12. CHelpG Charges (in *e*) on Atoms of $(HF)_5$ Configuration hf5_fhhhf

Atom[a]	Conv.[b]	1B	PA	3B	EE-1B	EE-PA	EE-3B
H(A)	0.450	0.433	0.453	0.447	0.455	0.452	0.447
F(A)	-0.451	-0.433	-0.453	-0.446	-0.455	-0.452	-0.447
H(B)	0.441	0.433	0.441	0.438	0.444	0.442	0.438
F(B)	-0.438	-0.433	-0.441	-0.435	-0.444	-0.443	-0.435
F(C)	-0.420	-0.433	-0.419	-0.422	-0.420	-0.414	-0.421
H(C)	0.416	0.433	0.419	0.414	0.420	0.416	0.414
H(D)	0.441	0.433	0.441	0.438	0.444	0.441	0.439
F(D)	-0.438	-0.433	-0.441	-0.436	-0.444	-0.443	-0.436
H(E)	0.450	0.433	0.454	0.447	0.455	0.453	0.447
F(E)	-0.451	-0.433	-0.454	-0.446	-0.455	-0.452	-0.446
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.013	0.002	0.003	0.004	0.002	0.003
RMSE		0.014	0.003	0.003	0.004	0.003	0.003
MUPE		2.9	0.5	0.7	0.9	0.6	0.6

[a]See Figure 5.5 of the paper. For example, the atom listed as H(B) is the hydrogen atom of the HF molecule labeled "B" in Figure 5.5 of configuration hf5_fhhhf.
[b]Conv. = Conventional

Table A.5.13. CHelpG Charges (in *e*) on Atoms of $(HF)_3(H_2O)_2$ Configuration w2hf3_a

Atom[a]	Conv.[b]	1B	PA	3B	EE-1B	EE-PA	EE-3B
O(A)	-0.746	-0.725	-0.792	-0.775	-0.757	-0.798	-0.774
H1(A)	0.361	0.363	0.390	0.376	0.379	0.394	0.375
H2(A)	0.364	0.362	0.393	0.366	0.378	0.397	0.366
H(B)	0.447	0.434	0.440	0.462	0.456	0.436	0.461
F(B)	-0.454	-0.434	-0.450	-0.461	-0.456	-0.446	-0.461
F(C)	-0.401	-0.434	-0.398	-0.402	-0.434	-0.398	-0.402
F(D)	-0.403	-0.434	-0.397	-0.396	-0.434	-0.398	-0.395
H(C)	0.418	0.434	0.412	0.420	0.434	0.410	0.421
H(D)	0.420	0.434	0.412	0.414	0.434	0.410	0.414
O(E)	-0.838	-0.725	-0.859	-0.833	-0.805	-0.870	-0.833
H1(E)	0.417	0.363	0.422	0.416	0.403	0.429	0.415
H2(E)	0.416	0.362	0.427	0.413	0.402	0.433	0.413
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.031	0.015	0.008	0.018	0.019	0.008
RMSE		0.043	0.019	0.011	0.020	0.024	0.011
MUPE		6.2	3.0	1.6	3.9	3.9	1.6

[a]See Figure 5.7 of the paper. For example, the atom listed as H2(A) is the hydrogen atom labeled "2" on the water molecule labeled "A" in Figure 5.7.
[b]Conv. = Conventional

Table A.5.14. CHelpG Charges (in *e*) on Atoms of $(HF)_3(H_2O)_2$ Configuration w2hf3_opt

Atom[a]	Conv.[b]	1B	PA	3B	EE-1B	EE-PA	EE-3B
O(A)	-0.731	-0.721	-0.741	-0.728	-0.872	-0.735	-0.735
H1(A)	0.388	0.361	0.404	0.379	0.436	0.395	0.385
H2(A)	0.388	0.361	0.403	0.381	0.436	0.394	0.386
H(B)	0.418	0.433	0.396	0.435	0.491	0.387	0.435
F(B)	-0.501	-0.433	-0.485	-0.503	-0.491	-0.479	-0.504
F(C)	-0.398	-0.433	-0.422	-0.434	-0.473	-0.466	-0.413
F(D)	-0.401	-0.433	-0.427	-0.431	-0.474	-0.472	-0.408
H(C)	0.323	0.433	0.366	0.388	0.473	0.462	0.346
H(D)	0.326	0.433	0.373	0.384	0.474	0.471	0.340
O(E)	-0.592	-0.730	-0.679	-0.737	-0.927	-0.927	-0.631
H1(E)	0.388	0.365	0.405	0.431	0.464	0.487	0.398
H2(E)	0.391	0.365	0.408	0.436	0.464	0.481	0.400
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.052	0.028	0.038	0.104	0.085	0.012
RMSE		0.066	0.035	0.054	0.132	0.123	0.016
MUPE		12.7	6.8	9.2	23.9	20.3	2.9

[a]See Figure 5.7 of the paper. For example, the atom listed as H2(A) is the hydrogen atom labeled "2" on the water molecule labeled "A" in Figure 5.7.
[b]Conv. = Conventional

Table A.5.15a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int1

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.563	-1.000	-0.987	-0.851	-0.838	-0.785	-0.686
A H_2O	-0.120	0.000	-0.013	0.002	-0.026	-0.016	-0.080
B H_2O	-0.120	0.000	0.000	-0.074	-0.073	-0.081	-0.102
C H_2O	-0.064	0.000	0.000	-0.007	-0.006	-0.016	-0.031
D H_2O	-0.064	0.000	0.000	-0.030	-0.022	-0.022	-0.054
E H_2O	-0.046	0.000	0.000	-0.033	-0.029	-0.063	-0.052
F H_2O	-0.025	0.000	0.000	-0.005	-0.006	-0.017	0.004
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.125	0.121	0.082	0.079	0.069	0.037
RMSE		0.182	0.176	0.122	0.115	0.097	0.052
MUPE		96.8	94.9	62.5	62.3	52.4	38.2

Table A.5.15b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int1

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.563	-1.000	-0.913	-0.567	-0.612	-0.701	-0.702
A H_2O	-0.120	0.000	-0.087	-0.066	-0.089	-0.077	-0.059
B H_2O	-0.120	0.000	0.000	-0.118	-0.109	-0.092	-0.108
C H_2O	-0.064	0.000	0.000	-0.076	-0.081	-0.037	-0.028
D H_2O	-0.064	0.000	0.000	-0.070	-0.049	-0.039	-0.053
E H_2O	-0.046	0.000	0.000	-0.049	-0.036	-0.058	-0.045
F H_2O	-0.025	0.000	0.000	-0.054	-0.024	0.004	-0.006
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.125	0.100	0.016	0.019	0.043	0.040
RMSE		0.182	0.146	0.024	0.024	0.059	0.060
MUPE		96.8	84.2	28.9	17.0	43.8	34.0

Table A.5.16a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int2

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.751	-1.000	-0.979	-0.812	-0.750	-0.796	-0.854
A H_2O	-0.039	0.000	-0.021	-0.016	-0.061	-0.061	0.002
B H_2O	-0.028	0.000	0.000	-0.002	0.000	0.005	-0.007
C H_2O	0.011	0.000	0.000	-0.021	-0.035	0.002	0.013
D H_2O	-0.118	0.000	0.000	-0.113	-0.110	-0.134	-0.132
E H_2O	-0.063	0.000	0.000	-0.017	-0.024	-0.033	-0.027
F H_2O	-0.013	0.000	0.000	-0.019	-0.019	0.016	0.005
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.074	0.068	0.028	0.021	0.026	0.033
RMSE		0.108	0.101	0.034	0.027	0.028	0.045
MUPE		90.4	82.4	81.9	97.7	78.5	59.6

Table A.5.16b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int2

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.751	-1.000	-0.922	-0.622	-0.654	-0.767	-0.791
A H_2O	-0.039	0.000	-0.078	-0.071	-0.057	-0.052	-0.032
B H_2O	-0.028	0.000	0.000	-0.058	-0.058	-0.011	-0.011
C H_2O	0.011	0.000	0.000	-0.028	-0.021	-0.009	-0.003
D H_2O	-0.118	0.000	0.000	-0.112	-0.111	-0.129	-0.118
E H_2O	-0.063	0.000	0.000	-0.067	-0.063	-0.059	-0.058
F H_2O	-0.013	0.000	0.000	-0.042	-0.036	0.027	0.013
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.074	0.063	0.038	0.030	0.017	0.015
RMSE		0.108	0.084	0.055	0.042	0.020	0.020
MUPE		90.4	89.1	114.3	92.9	86.1	59.6

Table A.5.17a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int3

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.714	-1.000	-0.970	-0.812	-0.771	-0.689	-0.677
A H_2O	-0.041	0.000	-0.030	0.015	-0.037	-0.034	-0.048
B H_2O	-0.092	0.000	0.000	-0.096	-0.099	-0.065	-0.093
C H_2O	-0.067	0.000	0.000	-0.016	-0.021	-0.086	-0.089
D H_2O	-0.061	0.000	0.000	-0.072	-0.060	-0.075	-0.060
E H_2O	-0.027	0.000	0.000	-0.004	0.003	-0.017	-0.012
F H_2O	0.001	0.000	0.000	-0.014	-0.015	-0.034	-0.021
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.082	0.073	0.037	0.023	0.019	0.015
RMSE		0.120	0.109	0.048	0.031	0.022	0.020
MUPE		91.4	80.5	261.4	261.8	524.4	333.1

Table A.5.17b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Int3

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.714	-1.000	-0.917	-0.591	-0.645	-0.685	-0.651
A H_2O	-0.041	0.000	-0.083	-0.032	-0.015	0.001	-0.045
B H_2O	-0.092	0.000	0.000	-0.137	-0.124	-0.101	-0.102
C H_2O	-0.067	0.000	0.000	-0.045	-0.040	-0.087	-0.095
D H_2O	-0.061	0.000	0.000	-0.120	-0.120	-0.065	-0.062
E H_2O	-0.027	0.000	0.000	-0.056	-0.038	-0.042	-0.033
F H_2O	0.001	0.000	0.000	-0.019	-0.017	-0.022	-0.013
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.082	0.070	0.044	0.034	0.020	0.018
RMSE		0.120	0.093	0.057	0.040	0.024	0.027
MUPE		91.4	90.3	326.5	292.7	364.0	214.1

Table A.5.18a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf1

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.738	-1.000	-0.907	-0.752	-0.777	-0.790	-0.772
A H_2O	-0.077	0.000	-0.093	-0.063	-0.052	-0.060	-0.116
B H_2O	-0.102	0.000	0.000	-0.097	-0.080	-0.112	-0.114
C H_2O	-0.053	0.000	0.000	-0.016	-0.021	-0.065	-0.044
D H_2O	0.006	0.000	0.000	0.044	0.043	0.048	0.091
E H_2O	-0.032	0.000	0.000	-0.082	-0.078	-0.048	-0.059
F H_2O	-0.005	0.000	0.000	-0.033	-0.035	0.026	0.014
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.076	0.054	0.026	0.033	0.026	0.032
RMSE		0.112	0.078	0.030	0.034	0.030	0.040
MUPE		90.8	77.7	219.0	224.1	217.0	295.2

Table A.5.18b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf1

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.738	-1.000	-0.906	-0.653	-0.688	-0.781	-0.747
A H_2O	-0.077	0.000	-0.094	-0.076	-0.100	-0.078	-0.076
B H_2O	-0.102	0.000	0.000	-0.101	-0.079	-0.113	-0.128
C H_2O	-0.053	0.000	0.000	-0.083	-0.089	-0.070	-0.052
D H_2O	0.006	0.000	0.000	0.017	0.039	0.039	0.041
E H_2O	-0.032	0.000	0.000	-0.087	-0.075	-0.034	-0.053
F H_2O	-0.005	0.000	0.000	-0.016	-0.009	0.038	0.015
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.076	0.054	0.028	0.030	0.022	0.016
RMSE		0.112	0.078	0.040	0.034	0.027	0.020
MUPE		90.8	77.9	98.8	134.2	221.3	161.5

Table A.5.19a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf2

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.712	-1.000	-0.916	-0.765	-0.727	-0.721	-0.717
A H_2O	-0.085	0.000	-0.084	-0.046	-0.085	-0.086	-0.076
B H_2O	-0.073	0.000	0.000	-0.105	-0.097	-0.093	-0.081
C H_2O	-0.083	0.000	0.000	-0.055	-0.096	-0.065	-0.058
D H_2O	0.008	0.000	0.000	0.041	0.059	0.038	0.008
E H_2O	-0.079	0.000	0.000	-0.079	-0.069	-0.107	-0.093
F H_2O	0.024	0.000	0.000	0.009	0.016	0.034	0.017
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.091	0.068	0.029	0.017	0.017	0.010
RMSE		0.125	0.093	0.033	0.023	0.019	0.012
MUPE		91.5	75.7	87.1	106.1	73.5	14.3

Table A.5.19b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf2

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.712	-1.000	-0.914	-0.633	-0.654	-0.747	-0.698
A H_2O	-0.085	0.000	-0.086	-0.043	-0.049	-0.070	-0.070
B H_2O	-0.073	0.000	0.000	-0.125	-0.126	-0.060	-0.089
C H_2O	-0.083	0.000	0.000	-0.128	-0.112	-0.048	-0.067
D H_2O	0.008	0.000	0.000	-0.017	-0.022	-0.003	0.002
E H_2O	-0.079	0.000	0.000	-0.090	-0.080	-0.093	-0.087
F H_2O	0.024	0.000	0.000	0.036	0.043	0.020	0.009
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.091	0.067	0.038	0.032	0.018	0.013
RMSE		0.125	0.093	0.044	0.037	0.021	0.013
MUPE		91.5	75.7	81.3	88.0	36.4	29.6

Table A.5.20a. Net Fragment Charges (in *e*) and Average Errors in EE-MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf3

Fragment	Conventional	EE-1B (S1)	EE-1B (S2)	EE-PA (S1)	EE-PA (S2)	EE-3B (S1)	EE-3B (S2)
Cl	-0.710	-1.000	-0.905	-0.703	-0.661	-0.695	-0.738
A H_2O	-0.118	0.000	-0.095	-0.104	-0.149	-0.154	-0.102
B H_2O	-0.080	0.000	0.000	-0.069	-0.095	-0.075	-0.054
C H_2O	-0.073	0.000	0.000	-0.132	-0.122	-0.057	-0.058
D H_2O	-0.018	0.000	0.000	-0.004	0.012	-0.039	-0.043
E H_2O	0.040	0.000	0.000	0.071	0.067	0.051	0.014
F H_2O	-0.041	0.000	0.000	-0.059	-0.051	-0.030	-0.018
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.094	0.067	0.022	0.030	0.016	0.023
RMSE		0.127	0.088	0.027	0.033	0.019	0.023
MUPE		91.5	78.2	43.4	53.5	32.8	47.2

Table A.5.20b. Net Fragment Charges (in *e*) and Average Errors in MB Approximations for $[Cl(H_2O)_6]^-$ Configuration Surf3

Fragment	Conventional	1B (S1)	1B (S2)	PA (S1)	PA (S2)	3B (S1)	3B (S2)
Cl	-0.710	-1.000	-0.905	-0.640	-0.649	-0.748	-0.745
A H_2O	-0.118	0.000	-0.095	-0.110	-0.140	-0.140	-0.112
B H_2O	-0.080	0.000	0.000	-0.118	-0.122	-0.042	-0.061
C H_2O	-0.073	0.000	0.000	-0.120	-0.103	-0.062	-0.069
D H_2O	-0.018	0.000	0.000	-0.026	-0.012	-0.036	-0.023
E H_2O	0.040	0.000	0.000	0.066	0.072	0.048	0.034
F H_2O	-0.041	0.000	0.000	-0.053	-0.047	-0.020	-0.025
MSE		0.000	0.000	0.000	0.000	0.000	0.000
MUE		0.094	0.067	0.030	0.028	0.022	0.013
RMSE		0.127	0.087	0.037	0.034	0.024	0.017
MUPE		91.5	78.1	37.8	35.4	36.4	17.4

Table A.5.21. Net Charges[a] (in *e*) on Fragments of Two Configurations of $(HF)_3(H_2O)_2$ and on One Configuration of $(HF)_5$

Frag.[b]	Conv.[c]	1B	PA	3B	EE-1B	EE-PA	EE-3B
hf5_fhhhf							
A	-0.001	0.000	0.000	0.001	0.000	0.000	0.001
B	0.003	0.000	0.000	0.003	0.000	-0.001	0.003
C	-0.004	0.000	0.000	-0.008	0.000	0.002	-0.007
D	0.003	0.000	0.000	0.003	0.000	-0.001	0.003
E	-0.001	0.000	0.000	0.001	0.000	0.000	0.001
w2hf3_a							
A	-0.021	0.000	-0.010	-0.033	0.000	-0.007	-0.033
B	-0.007	0.000	-0.009	0.001	0.000	-0.010	0.001
C	0.017	0.000	0.015	0.018	0.000	0.012	0.019
D	0.016	0.000	0.015	0.018	0.000	0.012	0.018
E	-0.005	0.000	-0.010	-0.004	0.000	-0.008	-0.004
w2hf3_opt							
A	0.046	0.000	0.066	0.032	0.000	0.055	0.036
B	-0.083	0.000	-0.089	-0.069	0.000	-0.091	-0.069
C	-0.076	0.000	-0.056	-0.046	0.000	-0.004	-0.066
D	-0.075	0.000	-0.054	-0.047	0.000	-0.001	-0.068
E	0.188	0.000	0.134	0.130	0.000	0.042	0.168
MSE[d]		0.000	0.000	0.000	0.000	0.000	0.000
MUE[d]		0.036	0.010	0.012	0.036	0.023	0.006
RMSE[d]		0.061	0.017	0.019	0.061	0.046	0.008
MUPE[d]		100.0	55.8	72.6	100.0	79.7	52.0

[a]The net charge on a fragment is defined as the sum of the CHelpG charges on the atoms that compose the given fragment.

[b]Frag. = Fragment. See Figures 5.5 and 5.7 of the paper in order to understand the fragment labels.

[c]Conv. = Conventional.

[d]MSE = mean signed error, MUE = mean unsigned error, RMSE = root mean squared error, MUPE = mean unsigned percent error. The averages are taken over all fragments of all three systems shown in the table (i.e., hf5_fhhhf, w2hf3_a, and w2hf3_opt).

Table A.5.22. Net Charges[a] (in *e*) on Fragments of Six Configurations of $(HF)_3(H_2O)$

Frag.[b]	Conv.[c]	1B	PA	3B	EE-1B	EE-PA	EE-3B
whf3_AaBaCa							
H_2O	0.036	0.000	0.028	0.037	0.000	0.032	0.037
A	-0.009	0.000	-0.003	-0.010	0.000	-0.004	-0.009
B	-0.013	0.000	-0.012	-0.014	0.000	-0.014	-0.014
C	-0.014	0.000	-0.012	-0.014	0.000	-0.014	-0.014
whf3_AaBaCr							
H_2O	0.033	0.000	0.024	0.033	0.000	0.026	0.033
A	-0.009	0.000	-0.004	-0.010	0.000	-0.004	-0.010
B	-0.013	0.000	-0.013	-0.012	0.000	-0.013	-0.013
C	-0.010	0.000	-0.007	-0.011	0.000	-0.009	-0.010
whf3_AaBrCr							
H_2O	0.029	0.000	0.019	0.032	0.000	0.022	0.033
A	-0.010	0.000	-0.005	-0.010	0.000	-0.004	-0.010
B	-0.009	0.000	-0.007	-0.011	0.000	-0.009	-0.011
C	-0.010	0.000	-0.007	-0.012	0.000	-0.009	-0.012
whf3_ArBaCa							
H_2O	0.035	0.000	0.031	0.036	0.000	0.031	0.036
A	-0.011	0.000	-0.008	-0.011	0.000	-0.009	-0.010
B	-0.013	0.000	-0.011	-0.013	0.000	-0.011	-0.013
C	-0.011	0.000	-0.011	-0.013	0.000	-0.011	-0.013
whf3_ArBaCr							
H_2O	0.028	0.000	0.027	0.030	0.000	0.025	0.030
A	-0.012	0.000	-0.009	-0.012	0.000	-0.009	-0.012
B	-0.010	0.000	-0.012	-0.011	0.000	-0.010	-0.011
C	-0.006	0.000	-0.006	-0.007	0.000	-0.005	-0.007
whf3_ArBrCr							
H_2O	0.023	0.000	0.022	0.028	0.000	0.019	0.028
A	-0.013	0.000	-0.009	-0.012	0.000	-0.009	-0.011
B	-0.005	0.000	-0.007	-0.007	0.000	-0.005	-0.008
C	-0.005	0.000	-0.007	-0.009	0.000	-0.005	-0.009
MSE[d]		0.000	0.000	0.000	0.000	0.000	0.000
MUE[d]		0.015	0.003	0.001	0.015	0.002	0.001
RMSE[d]		0.018	0.004	0.002	0.018	0.003	0.002
MUPE[d]		100.0	23.0	12.4	100.0	16.7	13.1

[a]See footnote *a* of Table A.5.21.

[b]Frag. = Fragment. See Figure 5.6 of the paper in order to understand the fragment labels.
[c]Conv. = Conventional.
[d]MSE = mean signed error, MUE = mean unsigned error, RMSE = root mean squared error, MUPE = mean unsigned percent error. The averages are taken over all fragments of all six systems shown in the table.