IOWA STATE UNIVERSITY
Digital Repository

Graduate Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

2012

Speed and accuracy: Having your cake and eating it too

Spencer Pruitt
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

 Part of the Physical Chemistry Commons

Recommended Citation
Pruitt, Spencer, "Speed and accuracy: Having your cake and eating it too" (2012). *Graduate Theses and Dissertations*. 12441.
https://lib.dr.iastate.edu/etd/12441

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

www.manaraa.com

Speed and accuracy: Having your cake and eating it too

by

Spencer Rian Pruitt

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Physical Chemistry

Program of Study Committee
Mark S. Gordon, Major Professor
Theresa L. Windus
Brett M. Bode
Malika Jefferies-EL
Aaron D. Sadow

Iowa State University

Ames, Iowa

2012

Copyright © Spencer Rian Pruitt, 2012. All rights reserved

www.manaraa.com

Table of Contents

www.manaraa.com

Chapter 1. Introduction

1. Overview

Since the first *ab initio* methods were developed, the ultimate goal of quantum chemistry has been to provide insights, not readily accessible through experiment, into chemical phenomena. Over the years, two different paths to this end have been taken. The first path provides as accurate a description of relatively small systems as modern computer hardware will allow[1]. The second path follows the desire to perform simulations on systems of physically relevant sizes while sacrificing a certain level of accuracy[2]. The merging of these two paths has allowed for the accurate modeling of large molecular systems through the use of novel theoretical methods[3-9]. The largest barrier to achieving the goal of accurate calculations on large systems has been the computational requirements of many modern theoretical methods. While these methods are capable of providing the desired level of accuracy, the prohibitive computational requirements can limit system sizes to tens of atoms. By decomposing large chemical systems into more computationally tractable pieces, fragmentation methods have the capability to reduce this barrier and allow for highly accurate descriptions of large molecular systems such as proteins, bulk phase solutions and polymers and nano-scale systems.

2. Dissertation Organization

The introductory section of this thesis provides the foundation of *ab initio* quantum chemistry relevant to the theoretical methods presented in the following chapters. Chapter 2 provides a detailed history of fragmentation methods, including rigorous formalism and applications of the most important methods. Chapter 3 details new developments and a rigorous comparison of two modern fragmentation methods, the FMO method and the systematic molecular fragmentation (SMF) method, for water clusters. Chapter 4 describes the development of the restricted open-shell formulation of the FMO method (FMO-ROHF). Chapter 5 builds upon the initial FMO-ROHF method through the addition of gradients. Chapter 6 describes the development and extension of the effective fragment molecular orbital (EFMO) method to model all relevant intermolecular forces.

3. Theoretical Background

Modern *ab initio* quantum chemical methods are derived from the first principles of quantum mechanics. The time-dependent Schrödinger equation[10]

$$\frac{\partial \Psi(r,t)}{\partial t} = -\frac{i}{\hbar} H \Psi(r,t) \tag{1}$$

www.manaraa.com

describes a system of particles at any time t. $\Psi(r,t)$ is a state function that describes the position r of the particles at time t, i is $\sqrt{-1}$, $\hbar$ is Planck's constant divided by 2π and H is the Hamiltonian operator. This equation is the fundamental expression upon which all of *ab initio* quantum chemistry is based.

To arrive at the formalism of many modern quantum chemistry methods, a variety of approximations to this fundamental expression must be made. The first of these approximations is the removal of time dependence by focusing on the stationary state solutions to Eq. (1). This simplification allows Eq. (1) to be reduced to an eigenvalue problem:

$$H\Psi(r) = E\Psi(r) \tag{2}$$

The total energy of the system, E, is determined by the application of the Hamiltonian operator to the eigenfunction $\Psi(r)$. The complete form of the Hamiltonian operator for a system composed of n nuclei and e electrons is

$$H = -\frac{1}{2}\sum_{i=1}^{e}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{n}\frac{\nabla_A^2}{m_A} + \sum_{A=1}^{n}\sum_{i=1}^{e}\frac{Z_A}{|R_A - r_i|} + \sum_{i>j}^{e}\frac{1}{|r_i - r_j|} + \sum_{A>B}^{n}\frac{Z_A Z_B}{|R_A - R_B|} \tag{3}$$

where ∇ is the Laplacian operator, m_A is the mass ratio of a nucleus and an electron, Z is the nuclear charge, R represents the coordinates of nuclei and r represents the coordinates of electrons. Each of the terms in Eq. (2) describes a different interaction between electrons and nuclei, including the electronic and nuclear kinetic energy (terms 1 and 2), nuclear-electron attraction (term 3), electron-electron repulsion (term 4) and nuclear-nuclear repulsion (term 5).

From the form of H in Eq. (3) arises the second approximation, due to the difference between electrons and nuclei. Proportionally, nuclei are significantly more massive (~3 orders of magnitude) than electrons, and therefore move with much slower velocity. From the perspective of the much lighter and faster moving electrons, the nuclei appear to be at fixed positions in space. Conversely, the nuclei only "feel" the averaged potential of the electron cloud, since the electrons rearrange instantaneously with changes in the positions of the nuclei. This allows for the electronic distribution to be calculated based upon the nuclei being fixed in space. This separation of electronic and nuclear motion is called the Born-Oppenheimer approximation[11]. Invoking this approximation effectively removes terms 2 and 5 from Eq. (3). The resulting expression, called the electronic Hamiltonian, is written as:

$$H_{elec} = -\frac{1}{2}\sum_{i=1}^{e}\nabla_i^2 - \sum_{A=1}^{n}\sum_{i=1}^{e}\frac{Z_A}{|R_A - r_i|} + \sum_{i>j}^{e}\frac{1}{|r_i - r_j|} \tag{4}$$

www.manaraa.com

This expression contains only terms that do not depend on the nuclear coordinates exclusively. Applying both of these approximations to Eq. (1) results in:

$$H_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \tag{5}$$

Eq. (5) can now be solved to obtain the electronic energy, E_{elec}, of the system at a given set of nuclear positions, with the nuclear-nuclear repulsion energy (term 5 in Eq. (3)) solved for separately.

While these first two approximations have greatly simplified the initial equation to be solved, the resultant equations still cannot be solved for systems containing more than one electron due to the inseparability of the problem. The trouble arises specifically from term 3 in Eq. (4), which represents the repulsive electron-electron interaction. In an effort to circumvent this problem, the Hartree-Fock (HF) approximation[12] is applied to Eq. (4) for many electron systems. By replacing term 5 in Eq. (4) with a one-electron potential $v^{HF}(i)$, the *i*th electron experiences a smeared out average of all other electrons in the system, effectively avoiding the calculation of explicit electron-electron interactions. Eq. (4) can now be rewritten as:

$$F = -\frac{1}{2}\sum_{i=1}^{e}\nabla_i^2 - \sum_{A=1}^{n}\sum_{i=1}^{e}\frac{Z_A}{|R_A - r_i|} + \sum_{i=1}^{e}v^{HF}(i) \tag{6}$$

Eq. (6) is the so-called one-electron Fock operator. Due to the one-electron nature of the Fock operator, a simple product of one-electron orbitals representing each electron, called a Hartree product, may be used to describe a many-electron system. Each of these orbitals can be constructed from a linear combination of atomic orbitals (LCAO), commonly referred to as basis functions:

$$\psi_i = \sum_{\mu} C_{\mu i}\phi_{\mu} \tag{7}$$

Eq. (7) shows the expansion of the *i*th molecular orbital as a sum of μ basis functions, where the $C_{\mu i}$ represent the LCAO coefficients. To obtain the exact expansion a complete set of basis functions is necessary, however a complete set would be infinite, thereby making such calculations intractable. In practice, a finite number of basis functions are used to perform the expansion in Eq. (7), with a larger number of basis functions providing solutions closer to the exact one.

The Hartree product wavefunction for a system of *N* electrons that is produced by the aforementioned procedure

$$\Psi^{\mathrm{HP}} = \psi_1\psi_2\psi_3\psi_4\ldots\psi_N \tag{8}$$

www.manaraa.com

still has a fundamental deficiency, since it does not take into account the indistinguishability of electrons, specifically distinguishing each electron as occupying a particular orbital. This violates the anti-symmetry, or Pauli exclusion, principle that requires an electronic wavefunction to be anti-symmetric with respect to the interchange of any two electrons. However, this requirement can be enforced through the use of Slater determinants[13], giving the final form of a Hartree Fock wavefunction as:

$$\Psi^{\mathrm{HF}} = \left|\psi_1\psi_2\psi_3\psi_4...\psi_N\right| \tag{9}$$

Since the solutions to Eq. (7) depend upon knowledge of the HF orbitals, an iterative method must be used, called the self-consistent field (SCF) procedure. The SCF procedure begins with a guess at the HF orbitals, followed by the calculation of the electronic density for these orbitals. From the density the Fock matrix can be assembled and diagonalized, producing a new set of molecular orbitals. These new orbitals can then be used as a subsequent guess at the HF orbitals. This process is then repeated until the density no longer changes within a predetermined amount.

The HF energy obtained from this procedure, E_{HF}, is guaranteed to be greater than the exact energy, E_{exact}, by the variational principle. The variational principle states that the energy obtained from any approximate wavefunction will be an upper bound on the exact non-relativistic energy. This means that for each subsequent SCF iteration the energy obtained will approach the exact energy.

Depending on the chemical system of interest, there are different representations of the HF wavefunction. Closed shell systems with no unpaired electrons may be represented using spin restricted Hartree Fock (RHF). This method restricts spatial orbital occupation numbers to two electrons of opposite spin and is used exclusively for closed-shell systems. When a chemical system contains any number of unpaired electrons, the restriction on spatial orbital occupation numbers must be loosened to include occupation of a spatial orbital by either one or two electrons. There are two formalisms used to describe such systems, spin restricted open-shell Hartree Fock (ROHF)[14] and spin unrestricted Hartree Fock (UHF)[15]. While ROHF allows for orbitals to be occupied by a single electron, two electrons of opposite spin occupy spatial orbitals containing an electron pair just as they would in the RHF case. In the UHF formalism, each electron is placed in a different spatial orbital depending upon its spin state. The relaxation of the restriction on spatial orbitals being doubly occupied leads to lower energies for UHF wavefunctions compared to ROHF wavefunctions. However, since a UHF wavefunction is not an eigenfunction of the spin operator $\hat{S}^2$, UHF wavefunctions often suffer from spin contamination by higher multiplicity states.

While these different formulations of the HF method can provide descriptions of a variety of chemical systems, the approximation made in obtaining Eq. 6 prevents the inclusion of electron-electron correlation. Although small, the energy obtained from this dynamic correlation plays an important role in the relative energetics of chemical systems. In an effort to regain the dynamic correlation energy lost through the use of the

www.manaraa.com

HF approximation, higher levels of theory can be appended onto the standard HF methodology. Using the HF wavefunction as a starting point, theoretical methods such as second order Møller-Plesset perturbation theory (MP2)[16] and coupled cluster theory (CC)[1], commonly referred to as post-Hartree Fock methods, are capable of recovering a portion of the dynamic correlation energy.

Perturbation theory is derived from the addition of a small perturbation, *V*, to the HF Hamiltonian:

$$H = H_0 + \lambda V \tag{10}$$

Expansion of the energy and wavefunction in a Taylor series yields:

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots \tag{11}$$

$$\left|\Phi_i\right\rangle = \left|i\right\rangle + \lambda\left|\Psi_i^{(1)}\right\rangle + \lambda^2\left|\Psi_i^{(2)}\right\rangle + \ldots \tag{12}$$

In Eq. (12) |i> represents the zeroth order wavefunction. Solving for $\langle i|V|j\rangle$, we choose

$$\langle i|j\rangle = \delta_{ij} \tag{13}$$

$$\langle i|\Phi_i\rangle = 1 \tag{14}$$

where δ_{ij} is the Kronecker delta. Using Eqs. (11) - (14), expressions for the zeroth, first and second order energy and wavefunction can be derived. Since the Hartree Fock energy (E_{HF}) contains the zeroth and first order energies, the second order energy

$$E_i^{(2)} = \left\langle i\middle|V\middle|\Psi_i^{(1)}\right\rangle \tag{15}$$

is added to E_{HF}. Further rearrangement of Eq. (15) to express the second order energy in terms of the zeroth order wavefunction gives

$$E_i^{(2)} = \sum_n \frac{\left|\langle i|V|n\rangle\right|^2}{E_i^{(0)} - E_n^{(0)}} \tag{16}$$

where the summation includes all states except the ground state of the system. However, the states represented by $|n\rangle$ cannot be single excitations since their contributions are zero. This leaves double excitations as the sole contribution to the second order energy, leading to the expression for the MP2 energy (E_{MP2})

www.manaraa.com

$$E_{\mathrm{MP2}} = E_{\mathrm{HF}} + \sum_{\substack{a<b \\ r<s}} \frac{\left|\left\langle i\left|V\right|\Psi_{ab}^{rs}\right\rangle\right|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s} \tag{17}$$

where the summation in Eq. (17) is over all possible double excitations from occupied orbitals (*a* and *b*) to virtual orbitals (*r* and *s*). The application of MP2 can recover between 80 and 90% of the total dynamic correlation energy. In addition to the recovery of dynamic correlation energy, MP2 has other advantages, including being size consistent and size extensive. Size consistent methods are those in which the energy of two infinitely separated, identical molecules is equal to double the energy of a single molecule. Size extensive methods produce a total energy that scales linearly with the number of particles. The main disadvantage of MP2 is that it is not a variational method, since the exact Hamiltonian is not used. Additionally, higher order energy corrections (third, fourth and higher orders of perturbation theory) have been derived; yet only MP2 has become a reliable method due to convergence issues inherent with higher order corrections.

Another post-Hartree Fock method, CC theory recovers more dynamic correlation energy than MP2 theory but at a considerably higher cost. The CC formalism is based upon the use of a cluster operator

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots + \hat{T}_n \tag{17}$$

where $\hat{T}_1$ represents the single electron excitation operator, $\hat{T}_2$ represents the two electron excitation operator and so on. The cluster operator can be expressed in exponential form as a Taylor series expansion:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \ldots \tag{18}$$

In practice this expansion is truncated at single and double (CCSD) or single, double and triple (CCSDT) excitations. However, due to increasing computational cost, the triple excitations are commonly obtained using perturbation theory (CCSD(T)). The truncated set of cluster operators then operates on the Hartree Fock wavefunction to produce the coupled cluster energy.

Due to the importance of dynamic electron correlation in accurate modeling of chemical processes, it is desirable to use either MP2 or CC when computationally feasible. Unfortunately, while both the MP2 and CC methods provide an effective way to recover the dynamic correlation energy from a HF reference, these methods are limited by their treatable system size. The computational cost for HF theory scales modestly as $O(n^4)$, where *n* measures the size of the system, for example in the number of basis functions or number of atoms. With the use of MP2 or CC theory, the computational cost increases significantly, with MP2 and CC scaling as $O(n^5)$ and $O(n^7)$ respectively. Beyond the initial computational scaling of these methods, the memory

www.manaraa.com

requirements for CC scale as O(n^4), adding an additional level of difficulty in modeling systems larger than ~10 heavy atoms using a moderately sized basis set.

In an effort to achieve the desired level of accuracy obtained through the use of MP2 or CC methods without limiting the treatable system size, a novel approach must be taken. The most common approach to this end is the decomposition of molecular systems into smaller pieces or "fragments" of a more tractable size. This allows the overall system size to increase without significantly increasing the computational cost beyond that of the method employed. There have been many such methods proposed, however the main focus of this dissertation will be one such method, the FMO method[17].

The FMO method is based upon the assumption that electron exchange, charge transfer and self-consistency are local effects. Breaking a molecular system into fragments and treating far separated fragments using only the one-electron Coloumb operator can achieve significant reductions in computational cost. This reduction in computational cost is further enhanced by the inherent parallelizability of the method, since each fragment calculation is essentially autonomous from all others. Computational algorithms, specifically the generalized distributed data interface (GDDI), have been implemented to take advantage of this by performing each fragment calculation in separate groups of compute nodes (coarse-grained parallel), with each group performing the assigned fragment calculation internally in parallel (fine-grained parallel). This two level parallelization scheme allows for nearly linear scaling of the FMO method with system size.

The general formalism of the FMO method begins with the expression for the *n*-body FMO*n* energy of a system of *N* fragments

$$E^{\mathrm{FMO2}} = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \left(E_{IJ} - E_I - E_J \right) \tag{15}$$

$$E^{\mathrm{FMO3}} = E^{\mathrm{FMO2}} + \sum_{I>J>K}^{N} \left(E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK} \right) \tag{16}$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \tag{17}$$

and E_I, E_{IJ} and E_{IJK} are the monomer (single fragment), dimer (fragment pair) and trimer (fragment triple) energies, respectively, computed in the electrostatic field of other fragments. Additional details of the FMO method as well as many other modern fragmentation methods will be presented in Chapter 2.

www.manaraa.com

References

1. (a) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989,** *157*, 479. 10. (b) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2000,** *149*, 71. (c) Włoch, M.; Gour, J.R.; Piecuch, P. *J. Phys. Chem. A*, **2007**, *111*, 11359.

2. Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227. (b) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem* **2002**, *53*, 467. (c) Friesner, R. A.; Guallar, Y. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389. (d) Damjanovic, A.; Kosztin, I.; Kleinekathoefer, U.; Schulten, K. *Phys. Rev. E* **2002**, *65*, 031919. (e) Chen, J.; Martínez, T. J. *Chem. Phys. Lett.* **2007**, *438*, 315. (f) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700. (g) Stanton, R. V.; Little, L. R.; Merz, K. M. *J. Phys. Chem.* **1995**, *99*, 17344. (h) Froese, R. D. J.; Morokuma, K. *Chem. Phys. Lett.* **1996**, *263*, 393. (i) Gao, J. L. *Acc. Chem. Res.* **1996**, *29*, 298. (j) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718. (k) Gao, J. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, **1996**; Vol. 7; p 119. (l) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718. (m) Mordasini, T. Z.; Thiel, W. *Chimia* **1998**, *52*, 288.

3. Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.

4. Pomogaev, V.; Pomogaeva, A.; Aoki, Y. *J. Phys. Chem. A* **2009**, *113*, 1429.

5. Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comp.* **2009**, *5*, 459.

6. Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comp.* **2009**, *5*, 1573.

7. Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comp.* **2009**, *5*, 1667.

8. 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.

9. Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927

10. (a) Schrödinger, E. *Ann. Phys.* **1926**, *79***,** 361. (b) Schrödinger, E. *Ann. Phys.*, **1926**, *79*, 489. (c) Schrödinger, E. *Ann. Phys.*, **1926**, *79*, 734. (d) Schrödinger, E. *Ann. Phys.*, **1926**, *80*, 437. (e) Schrödinger, E. *Ann. Phys.*, **1926**, *81*, 109. (f) Schrödinger, E. *Die Naturwissenschaften*, **1926**, *14*, 664.

11. Born, M; Oppenheimer, R. *Ann. Phys.*, **1927**, *84*, 457.

12. (a) Hartree, D. R. *Proc. Cambridge Philos. Soc.*, **1928**, *24*, 89. (b) Hartree, D. R. *Proc. Cambridge Philos. Soc.*, **1928**, *24*, 111. (c) Hartree, D. R. *Proc. Cambridge Philos. Soc.*, **1928**, *24*, 426. (d) Fock, V. *Z. Phys.*, **1930**, *61*, 126. (e) Fock, V. *Z. Phys.*, **1930**, *62*, 795.

13. Slater, J. C. *Phys. Rev.*, **1929**, *34*, 1293.

14. Binkley, J. S.; Pople, J. A.; Dobosh, P. A. *Mol. Phys.*, **1974,** *6*, 1423.

15. Pople, J. A,; Nesbet, R. K. *J. Chem. Phys.*, **1954,** *22*, 571.

www.manaraa.com

16. Møller, C.; Plesset, M. S. *Phys. Rev*. **1934,** *46*, 618.

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

www.manaraa.com

Chapter 2. Fragmentation Methods: A Route to Accurate Calculations on Large Systems

A paper published in *Chemical Reviews*

Mark S. Gordon, Dmitri G. Fedorov, Spencer R. Pruitt, Lyuda V. Slipchenko

Abstract

The development history of fragment-based methods including quantum mechanics derived force fields is reviewed in the context of other reduced scaling methods such as the interaction analyses. For some important methods a mathematical description is provided in some detail, and the applications of all methods are introduced according to their field. Some attention is paid to available software to perform practical calculations.

1. Introduction

Theoretical chemists have always strived to perform quantum mechanics (QM) calculations on larger and larger molecules and molecular systems, as well as condensed phase species, that are frequently much larger than the current state-of-the-art would suggest is possible. The desire to study species (with acceptable accuracy) that are larger than appears to be feasible has naturally led to the development of novel methods, including semi-empirical approaches, reduced scaling methods, and fragmentation methods. The focus of the present review is on fragmentation methods, in which a large molecule or molecular system is made more computationally tractable by explicitly considering only one part (fragment) of the whole in any particular calculation. If one can divide a species of interest into fragments, employ some level of *ab initio* QM to calculate the wavefunction, energy, and properties of each fragment, and then combine the results from the fragment calculations to predict the same properties for the whole, the *possibility* exists that the accuracy of the outcome can approach that which would be obtained from a full (non-fragmented) calculation. It is this goal that drives the development of fragmentation methods.

An additional potential positive aspect of fragmentation methods is their ability to take advantage of massively parallel computers. If one can calculate the energy of each fragment in such a manner that the fragment calculation is essentially independent of the calculations for all of the other fragments, then each fragment calculation can be performed on a separate compute node. This is sometimes called coarse-grained parallel computing, and it can be very efficient. If, in addition, the QM algorithm one is using (e.g., for second order perturbation theory, MP2) has itself been developed for parallel hardware, one can implement a multi-level parallel approach, with fine-grained parallelism employed among the core within each node. This further improves the parallel efficiency of the calculation. [1]

Aiming at achieving computational efficiency, a variety of methods based on localized orbitals was developed starting with Adams in 1961.[2] These methods are only briefly mentioned in this review, as they

www.manaraa.com

typically deal with orbitals in the full system rather than molecular fragments. Klessinger and McWeeny[3] in 1965 in their Group SCF (self-consistent field) method suggested defining molecular orbital groups to reduce the scaling of SCF. Christoffersen and co-workers[4,5,6,7,8,9,10,11,12,13,14] employed localized molecular orbitals (LMOs) and floating spherical Gaussian orbitals (FSGO)[15] to separate the sub-density matrices of fragments that are then summed to obtain the total density and (thereby) desired properties. The method has been applied with success to many species, including molecules of biological importance. As another important example, Stoll and Preuß[16] in 1977 used a subsystem-based approximation to define the total density, and suggested many-body corrections to the energy from the density computed at these levels. Stoll [17] in 1992 suggested many-body incremental corrections to the correlation energy, based on LMOs. In a group of methods based on the divide-and-conquer (DC) idea, [18] an ad hoc total density from fragment calculations is constructed, followed by a single energy evaluation using this density. Divide and conquer methods are still an active area of research.[19]

The elongation method (usually abbreviated as ELG), based upon localized orbitals, was developed by Imamura[20] in 1991 (related to the earlier work by Demelo et al.[21]) to enable calculations on large polymeric chains at the computational cost of a much smaller system. A good example of an application for which the ELG method was developed is that of a polymerization/copolymerization reaction. Reactions of this type begin with an initiation step, followed by some number of propagation steps, and ultimately complete with a termination step. This is the series of steps that are employed in the elongation methodology. Initially, the applicability of the elongation method was limited to periodic polymer chains,[20] however the method has since been extended to aperiodic systems[22] and hydrogen bonded systems such as water clusters.[23] Since the original formulation, a new localization scheme[24] (a central aspect of the formalism) as well as extensions have been proposed. [25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43] The elongation method is fundamentally an LMO approach, rather than a fragment-based approach, so it is discussed only briefly here.

There have been many fragment-like approaches to electronic structure theory, in which fragments appear as groups of atoms and the electronic state of the full system is computed. Morokuma[44] in 1971 systematically considered both the Hartree product wave function and the fully antisymmetrized total wave functions constructed fom wave functions of two weakly interacting fragments. This approach was extended to an arbitrary number of fragments by Ohno and Inokuchi[45] for an antisymmetrized product in 1972, and by Gao[46] for a Hartree product in 1997. Some theoretical foundations for fragment-based methods were given by Kutzelnigg and Maeder in 1978. [47]

The idea of performing fragment calculations in the Coulomb field of other fragments with or without the corresponding exchange has been reinvented and reformulated over the years in various forms, and given names such as mutual consistent field (MCF), [48] structural SCF, [49] self-consistent embedded ions (SCEI), [50,51] double self-consistent field (DSCF), [52] self-consistent charge (SCC)[53] and monomer SCF. [54] One of the early methods to accomplish this self-consistent cycle is due to Morokuma[44] in 1971; The Morokuma development was followed

www.manaraa.com

by the systematic work of Ohno[45] in 1972 and the energy decomposition (EDA) method by Kitaura and Morokuma [55] in 1976. The EDA method was later extended to more than two fragments by Chen and Gordon[56] in 1996. While these earlier approaches were not intended for large-scale calculations, the MCF method by Otto and Ladik[48] is the prototypical approach for self-consistent fragment calculations that are used today for large systems. Later it was applied to periodic systems by Bohm[57] in 1982, Kubota et al.[49] in 1994 and Pascual and Seijo[51] in 1995. Gao employed this MCF approach in his molecular orbital-derived empirical potential for liquid simulations (MODEL) [46] in 1997 (later renamed explicit polarization, X-Pol) and by Kitaura et al.[58] in the fragment molecular orbital (FMO) method in 1999.

An important advance in the development of fragment-based methods was made by the explicit division of the equations into blocks derived for individual fragments. While the earlier methods had to diagonalize the full Fock matrix,[44,45,55] the introduction of strong orthogonality between fragment wave functions made it possible to block the fragment Fock matrices[48,49,50,52,58] although they are still coupled in many methods via the electrostatic field. Here again, note the paramount importance of the MCF approach, [48] which with modifications has been integrated as a part of modern methods such as X-Pol and FMO.

Based on the starting point of computing the electronic state of individual fragments, two main branches of methods emerged for the consequent refinement of properties. The first group constitutes a perturbation treatment of fragment interactions, a typical example being the MCF method,[48,59,60,61,62,63,64,65,66,67,68,69] in which the mutually polarized fragment wave functions are used in a perturbative manner to obtain the exchange and charge transfer interactions between fragments. The other group of methods was inspired by the desire to decompose intermolecular interactions into conceptually familiar components. In this group, pairs or larger conglomerates of fragments are computed, and the important concept is the treatment of many-body effects other than the electrostatics. One of the early methods was developed by Hankins et al.,[70] who considered many-body corrections to intermolecular interactions in vacuum. By a clever introduction of the electrostatic potential (ESP) in the many-body expansion, Kitaura et al.[58] in 1999 were able to incorporate many-body effects in the framework of a two-body expansion, conceptually reminiscent of some earlier methods[70,16,17]. The systematic cancellation of the double counting of the Coulomb interaction in many-body calculations was shown diagrammatically by Fedorov and Kitaura in 2004.[54] Suárez et al.[71] in 2009 discussed some general aspects of many-body expansions. Gao et al.[72] suggested a means for including many-body effects using the total wave function. Many methods discussed in detail below partially consider many-body effects implicitly by computing large conglomerations of atoms (fragments).

In the original 1975 implementation of MCF by Otto and Ladik[48], the wave function of a fragment was polarized by a partner fragment through modifying the Fock operator by the Coulomb potential of a partner. Simple monopole approximation was used originally to represent the electrostatic potential; HF equations for a pair of fragments were solved until self-consistency. This strategy automatically takes care of electrostatic and polarization effects. Exchange and charge-transfer energies were then obtained perturbatively based on the

www.manaraa.com

polarized wave functions of the fragments. Later on, the method was extended by including a local exchange potential (using the original Slater expression) into the Fock operator and substituting the monopole approximation by distributed charges.[59] In 1985 implementation of the pseudo-polarization tensor (PPT) MCF,[65,67] point charges and pseudo-polarizabilities were calculated for each individual fragment, and the Coulomb and polarization interaction energies were computed using these charges and polarizabilities. So, PPT-MCF is related to universal force fields, like EFP[73,74,75] SIBFA[76], and AMOEBA. [77]

The idea of a perturbative treatment of fragment interactions has been further extended and developed in the symmetry-adapted perturbation theory (SAPT)[78,79] by Moszynski, Jeziorski, and Szalewicz in which the interaction energy is expressed in orders of an intermolecular interaction operator *V* and a many body perturbation theory (MBPT) operator *W*. The polarization energies are obtained from a regular Rayleigh-Schrödinger perturbation theory; e.g., the Coulomb energy appears in the first order, induction and dispersion energies in the second order, etc. Additionally, the exchange corrections arise from the use of a global antisymmetrizer to force the correct permutational symmetry of the dimer wave function in each order. In this way, the exchange-repulsion energy appears in the first order, and exchange-induction and exchange-dispersion contribute to the second order of the perturbation theory. The supermolecular HF energy in SAPT can be represented by corrections in zero order in *W*, namely, Coulomb, induction, exchange, and exchange-induction. The SAPT2 level is roughly equivalent to supermolecular second order MBPT calculations. The new contributions to the interaction energy in the second order theory are Coulomb, induction, and exchange-induction corrections that are second order in *W*, in addition to first and second order corrections to *W* for exchange, and zeroth-order dispersion and exchange-dispersion corrections. The highest routinely used level of SAPT is equivalent to fourth order supermolecular MBPT and includes third-order corrections to *W* for the Coulomb energy and higher orders of the exchange and dispersion energies.

SAPT was extended to enable its use with density functional theory (DFT)[80]. In this SAPT(DFT) scheme, Kohn-Sham (KS) orbitals and orbital energies are used to obtain Coulomb and exchange in zero order of *W*, similarly to how it is done in the wave function-based SAPT (essentially, by replacing HF orbitals with KS orbitals). Induction, dispersion, exchange-induction, and exchange-dispersion are calculated by using the frequency-dependent density susceptibility functions obtained from the time-dependent DFT theory at the coupled Kohn-Sham level of theory.[80,81,82] The scaling of SAPT(DFT) is $O(N^6)$ and becomes $O(N^5)$ if density fitting is used.[83,84] This is significantly better than the $O(N^7)$ scaling of the wave function-based SAPT.

In variational (i.e., HF) energy decomposition schemes, the total energy is typically represented as a sum of a frozen density interaction energy, a polarization energy, and a charge-transfer energy. The frozen density term is calculated as the interaction of the unrelaxed electron densities on the interacting molecules and consists of Coulomb and exchange contributions. The polarization term originates from the deformation of the electron clouds of the interacting molecules in the fields of each other. The charge transfer term arises due to the electron flow between the molecules in the system. Quantum mechanically, polarization and charge transfer terms can

www.manaraa.com

be described as energy lowerings due to the intramolecular and intermolecular relaxation of the molecular orbitals, respectively. The main differences in the variational EDA schemes come from the manner in which the intermediate self-consistent energies, corresponding to the variationally optimized antisymmetrized wave functions constructed from MOs localized on the individual molecules, are determined.

The original EDA method of Kitaura and Morokuma[55,85,86] (KM), which has become prototypical for many subsequent methods, lacks the antisymmetrization of the intermediate wave function, resulting in numerical instabilities of the polarization and charge-transfer components at short distances and with large basis sets. It is, however, not entirely artificial, but is connected to the intrinsic similarity between the intramolecular (polarization) and intermolecular charge transfers. The concept of the fragment (monomer) polarization is apparently not well defined with large basis sets, especially those that contain diffuse functions and are therefore not obviously localized on a given fragment. In this sense, when the polarization of fragments as an important physical concept is of interest, one may have to employ smaller basis sets.

The restricted variational space (RVS) analysis[87,56] and the constrained space orbital variations (CSOV) method[88,89] improve on the KM scheme by employing fully antisymmetrized intermediate wave functions. The main deficiency of both methods is that they do not produce self-consistent polarization energies (which results, for example, in a dependence on the order in which the fragments are treated) and do not completely separate charge-transfer from polarization. In the natural energy decomposition analysis (NEDA)[90,91,92] the intermediate wave function is not variational and the resulting polarization and charge-transfer energies may be under- and overestimated, respectively. In absolutely localized molecular orbital (ALMO) approach[93] by Head-Gordon, all energy terms are calculated variationally and the polarization and charge-transfer terms are naturally separated. Additionally, the charge-transfer energy can be decomposed into forward donation and back-bonding contributions. An earlier implementation of EDA based on the block localized wave function developed by Mo et al. is similar in spirit.[94]

Fedorov and Kitaura have extended the EDA method to covalently bound fragments by introducing the pair interaction energy decomposition analysis (PIEDA) [95] within the FMO framework[58] and showed the equivalency of the fragment SCF in FMO and the polarized state of monomers in EDA. In the FMO method, one performs fragment calculations in the electrostatic field of other fragments, mutually self-consistent with each other. This approach provides the polarized state of the fragments including many-body polarization. In the EDA method, the same thing is accomplished by the restriction of orbital rotations, which only allows many-body polarization to take place for the interfragment interactions. The exact agreement of the resultant many-body polarization, shown numerically with the aid of PIEDA for water clusters, clearly demonstrated the equivalence of the self-consistent fragment polarizations in FMO and EDA, in this regard.

In the EDA approach developed by Su and Li,[96] the interaction energy is separated into Coulomb, exchange, repulsion, polarization, and dispersion terms. In contrast with the KM and other EDA schemes, the exchange

www.manaraa.com

and repulsion energies in this scheme are separated according to the method of Hayes and Stone.[97] Dispersion is obtained using a supermolecular approach and size-consistent correlation methods (MP2 or CCSD(T)). Polarization is defined as the orbital relaxation energy that occurs on going from the monomer to the supermolecular orbitals. That is, the polarization interaction in the Su-Li EDA method includes both the polarization and charge-transfer contributions of the KM scheme. A DFT version of the Su-Li EDA scheme was also developed.

Wu et al developed a density-based EDA[98], in which the energies of the intermediate states are calculated using the densities of the fragments, rather than their wave functions. Similarly to the KM EDA method, the total energy is separated into a frozen density Coulomb plus exchange repulsion terms, polarization, and charge-transfer. The frozen density energy is obtained using the constrained search technique developed by Wu and Yang.[99] The polarization and charge-transfer terms are separated from each other by constructing an intermediate state in which the density is relaxed without charge transfer, using the constrained DFT method of Wu and Van Voorhis.[100] The EDA approach is a useful tool for the development of reliable force fields for condensed phase molecular simulations.[101,102,103,104,105,106,198,199,197]

Related to EDA schemes and ideas is the self-consistent-field method for molecular interactions (SCF-MI) first introduced by Gianinetti et al.[107 108] and further reformulated and extended by Nagata et al.[109] In the SCF-MI method one expands molecular orbitals of a given fragment in terms of only the atomic orbitals belonging to atoms of that fragment. This leads to absolutely localized MOs that are free from basis set superposition error (BSSE), but also prevents charge-transfer between the fragments. The charge-transfer interactions between the fragments can be added by means of the single excitation second order perturbation theory, as suggested by Nagata and Iwata.[110] Khaliullin et al. showed that the SCF-MI method can result in significant computational savings (e.g., $(N/O)^2$ speedups for the diagonalization step compared to the conventional SCF method) and thus is applicable to systems containing hundreds of molecules.[111]

Murrell and co-workers developed a method called diatomics-in-molecules,[112] that relied on a power expansion in the intermolecular overlap integral, in order to obtain a general expression for the intermolecular exchange repulsion. Jensen and Gordon[113] built upon the Murrell approach using LMOs to derive a general expression for the intermolecular exchange repulsion interaction energy. [114,115]

A number of fragment methods have been proposed based on the so called thermochemical[116] analogy, i.e., by capping the fragments of interest and eliminating or subtracting the effect of the caps. These include the molecular fractionation with conjugate caps (MFCC) method,[117] its later extension, the generalized energy-based fragmentation (GEBF) method,[118] the molecular tailoring approach (MTA),[119] the kernel energy method (KEM) and the fragment energy method (FEM).[71] The relation between these has been discussed by Suárez et al.[71] Collins and co-workers[120] have developed the systematic molecular fragmentation (SMF) method in order to describe large molecular systems with accurate QM methods. Smaller subsystems are treated with a high

www.manaraa.com

level of accuracy. This accuracy is, in principle, retained by incorporating non-bonded interactions between the fragments using model potentials.[121]

A related, but somewhat different approach has been developed and implemented by several groups in order to expand the size of accessible species and active species in multi-configurational self-consistent field (MCSCF) methods, in particular the complete active space (CAS) SCF approach. Among the most well known of these methods are the restricted active space (RAS) SCF method, developed by Roos and co-workers,[122] the quasi (Q) CAS method of the Hirao group,[123] and the occupation restricted multiple active space (ORMAS) method developed by Ivanic.[124] The general philosophy of these methods is to divide a large (possibly computationally intractable) CAS active space into logically determined sub-spaces, so that each subspace is amenable to MCSCF calculations. ORMAS is the most general of these methods and subsumes the others. Because MCSCF calculations account only for static correlation, it is necessary to add dynamic correlation, either variationally via configuration interaction (multi-reference CI = MRCI) or perturbatively (MRPT). MRPT is the computationally more efficient approach. Therefore, the RASPT2 method[125] and ORMAS-PT2 method[126] have recently been introduced.

Another class of methods that is related to fragmentation is the use of LMOs to reduce the scaling of (most commonly) correlated electronic structure methods. This is accomplished by first performing a Hartree-Fock (HF) calculation on the entire system and localizing the orbitals. In this sense the "local orbital" approach is distinct from most fragmentation methods in that the latter first separate the system into separate physical collections of atoms, and subsequently perform explicit calculations on only one fragment at a time, not the entire system. Most local orbital methods[127] have built upon the beautiful early work of Pulay and co-workers[128,129] whose work primarily addressed second order perturbation theory (MP2). Many others have contributed to this field, most notably Werner and co-workers[130] (MP2 and coupled cluster (CC) methods), Head-Gordon *et al.*[131] (MP2 and CC methods), and Carter and co-workers[132] (multi-reference configuration interaction). A very recent method developed by the Piecuch group[133] called cluster in molecule (CIM) is generally applicable and can be used as a multi-level method, in which different parts of a system can be treated with different levels of theory (Werner and co-workers[134] used MP2 for small contributions in their local CC method).

Another important example is the incremental method,[17,135] which is mainly used to estimate the correlation energy in periodic systems, similar to other local orbital methods, with the important distinction that a many-body expansion in terms of orbital contributions is used. This is conceptually similar to the expansion used in some fragment methods, although molecular applications have also been reported.[136,137,138] Manby et al. [139] suggested a hierachical method for the additive calculation of the properties of clusters using edge, corner, surface and bulk unit energies, and compared their method to the incremental method.[140]

www.manaraa.com

A number of fragment-based methods represent a important class of linear scaling methods (sometimes called order *N*, *O(N)* methods), which are a thriving field of current research.[141,142] Strictly speaking, it is often not entirely clear that a particular method really scales linearly, and a description like "nearly linear scaling" may be more appropriate (as discussed by Nagata et al. in chapter 2 of Ref. 141)

Another group of approaches, only briefly mentioned is related to the integrated MO and molecular mechanics (MM), IMOMM method,[143] its later extension our own *n*-layered integrated molecular orbital and molecular mechanics (ONIOM),[144] methods known as integrated MO and MO (IMOMO),[145] QM:QM,[146] and multiple area QM/MM.[147,148] Most of these methods compute the whole system at a low level and add higher level results for a selected part of the system. Some of the methods such as multiple area QM/MM[147,148] and extended ONIOM[149] can be considered fragment-based approaches.

2. Methodologies

One can classify fragment-based methods in various ways. Fedorov and Kitaura[223] suggested three categories: divide-and-conquer (DC),[18] transferable approaches (SMF[120]) and methods based on many-body molecular interactions (FMO,[58] MFCC[117]). The distinction between the latter two groups is in the way in which fragment interactions are added: the latter category typically has a simple many-body expansion inspired by the theory of molecular interactions;[70] while the former group has a different recipe for taking into account these interactions. Fedorov and Kitaura also proposed the name of *e pluribus unum* for those methods in which the total properties are obtained by fragment calculations under the influence (for example, in the electrostatic field) of the whole system.

Li et al.[118] grouped fragment methods into density matrix (DC, [18] ELG,[20] MFCC[117]) or energy based approaches (IMiCMO, [150] MFCC, [117] SMF, [120] MTA[119]). The density based methods compute the density of the whole system using fragments, followed by the calculation of the energy; the energy based methods compute the energy directly from fragment energies avoiding the expensive step of calculating the energy from the density of the whole system. For some methods, like MFCC, it is possible to either compute the energy directly or to first obtain the density and then the energy; thus MFCC belongs to both categories.

Suárez et al.[71] classified fragment-based methods into two main groups, those using (a) overlapping fragments (MTA, [119] MFCC, [117] FEM[71] and SMF[120]) and (b) disjoint fragments (FMO, KEM). The main idea of this classification is in the fragment definition: whether an atom is assigned to a single fragment as in (a) or can be found in several, as in (b).

Figure 1 illustrates an elaboration and generalization of the classification introduced by Li et al.[118] The purpose of this modification is to introduce a more elaborate division than the two groups that were used in the original classification, and to extend it to apply when a property other than the energy is directly computed from the fragment calculations. The energy-based group is renamed the *one-step* group, which refers to the direct

www.manaraa.com

calculation of total properties from the fragment-derived values. This approach allows one to include properties, other than energy, such as the density and properties linear in the density (i.e., one-electron properties) to be included in the classification. The distinctive feature of the one-step group is that the total properties are obtainable directly from fragment calculations. That is, the one-electron properties from the density can be rewritten as an appropriate combination of fragment properties.

The *two-step* group, which corresponds to the density based method of Li et al.,[118] includes methods which consist of two steps: (1) computing some total property from piecewise values (usually, the electron density); (2) calculating a related total property dependent on the one determined in the first step in a nonlinear way (usually, the total energy). For methods in this group it is necessary to perform calculations for the whole system at once, which imposes various limitations and leads to complications.

Each of the one and two-step groups can be further divided to reflect the manner in which the QM calculations on the combinations of fragments are performed. In the 1-body methods, where "1-body" refers to fragments, there are no QM calculations of the conglomerates of fragments. Instead, often either force field derived terms are added (e.g., X-Pol[46]) or perturbation theory is used based on the wave functions of fragments as the starting point (MCF[48]). In the many-body subgroup pairs, triples or large unions of fragments are computed, whereas in the conglomerate subgroup different principles are used to combine fragments into unions, e.g., based on the cardinality (MTA method) or buffer zone (DC method) to include a certain number of atoms around each fragment. The many-body group eventually can be said to be inspired by the theory of molecular interactions[70] with its many-body expansion and typically contains many-body energy corrections such as E_{IJ}-E_I-E_J, for dimer *IJ* corrections to monomers *I* and *J*. On the other hand, the conglomerate group uses different recipes for the construction of groups of atoms to be computed.

Several families of methods, namely, FMO, MFCC, and KEM have variants that fall into different categories. For instance, in the FMO-MO method (introduced below), an FMO calculation is performed in the first step, whereas the second step is performed as in a full ab initio QM calculation for the whole system, corresponding to one SCF iteration. In FMO-based NMR calculations, denoted FMO1 (merged), one first performs self-consistent calculations on fragments (FMO1). This is followed by the calculation of NMR shifts for a large "superfragment" composed of a central fragment *I* and all fragments within a desired radius around it (*I* consequently goes over all fragments). This is the reason that some methods appear in several different groups.

2.1 QM-based Force Fields

Quantum mechanics calculations are often used to assist the development and parametrization of modern polarizable force fields. However, while it is straightforward to extract the electrostatic point charges or multipoles from electronic structure calculations on small fragments or molecules, rigorous formulation of

www.manaraa.com

accurate but computationally inexpensive ways to model exchange-repulsion self-consistent induction, dispersion, and charge-transfer terms can be quite intricate, especially in the interfacial region between QM and the force field. Therefore, various approaches have been suggested and tried over years, often combining quantum-mechanically-based terms (typically Coulomb and polarization) and parameterizing the rest of the potential.

For example, in AMOEBA (atomic multipole optimized energetics for biomolecular applications) force field, multipoles on atoms (up to quadrupoles)[77] are obtained from the distributed multipolar analysis (DMA) and (experimental) isotropic atomic polarizabilities with Thole's damping functions are used.[151] Also, in AMOEBA the Van der Waals R^{-7}-R^{-14} term is parameterized to reproduce experimental gas phase and condense phase data. Parameters of bonded terms in AMOEBA are also obtained by fitting to experimental data. In the NEMO (non-empirical molecular orbital) force field for water[152], multipoles up to quadrupoles are also used to calculate Coulomb interactions, while atomic polarizabilities are obtained by solving coupled HF equations. Dispersion is fitted to an analytic R^{-6} expression with exponential damping, while the exchange-repulsion is represented by a sum of exponential and R^{-14} terms.

The essence of the direct reaction field (DRF) approach[153] of van Duijnen is an explicit evaluation of the polarization (induction) term using atomic polarizabilities and Thole's damping functions. The Coulomb term is evaluated using screened charges; dispersion energy is calculated using the Slater-Kirkwood expression[154] with atomic polarizabilities, and the repulsion component is taken to be proportional to the dispersion scaled by Van der Waals atomic radii.

In the SIBFA (sum of interactions between fragments ab initio) force field, [76,155,156,157,158,159] developed by Gresh and coworkers, the interaction energy between fragments is a sum of Coulomb, polarization, repulsion, dispersion, and charge-transfer components. The electrostatic term is computed using distributed multipoles up to quadrupoles centered at atoms and bond barycenters. The multipoles are obtained using the procedure developed by Vigné-Maeder and Claverie.[160] The Coulomb term is augmented by an explicit penetration contribution.[161,162] The repulsion term is formulated as a sum of bond-bond, bond-lone pair, and lone pair-lone pair interactions expressed as S^2/R terms. Here, S is an approximation of the overlap between LMOs of the interacting partners; R is the distance between the LMO centroids. The S^2/R^2 term was added to improve the accuracy of the repulsion term.[162] Polarization energies are obtained by using permanent multipoles (the same as those that appear in the Coulomb term) and distributed polarizability tensors computed at the bond barycenters and on the heteroatom lone pairs. Polarization interactions are screened by Gaussian functions that depend on the distance between the interaction centers. The dispersion term is described as a sum of R^{-6}, R^{-8}, and R^{-10} contributions,[163] calibrated on the basis of a SAPT analysis. The uniqueness of SIBFA is in the treatment of the charge-transfer interactions that are modeled following the Murrell formulation[164] using the ionization potential of the electron donor and the electron affinity and 'self-potential' of the electron acceptor.[157,156] The charge transfer term is essential for describing polycoordinated complexes of cations.[165,159]

www.manaraa.com

2.1.1 Effective Fragment Potential Method

The original idea of the effective fragment potential method[73,166,167] (EFP) was to describe aqueous solvent effects on molecules of biological interest, in which the solute was described by some QM level (usually HF) and the solvent molecules were represented by EFPs. The EFP potential was represented as series deduced from the long-range (in powers of (1/*R*) Coulomb operator) and short-range (in powers of intermolecular overlap *S*) perturbation theory. In this formulation each water molecule is represented as a fragment of fixed geometry with a set of parameters deduced from ab initio calculations. In this original implementation,[166,167] now called EFP1, the interaction energy between water molecules consists of Coulomb, polarization, and repulsion terms:

$$E^{\mathrm{EFP1}} = E^{\mathrm{Coul}} + E^{\mathrm{pol}} + E^{\mathrm{rep}} \tag{1}$$

In the presence of a quantum region, the total energy of the QM/EFP1 system can be written as:

$$E^{\mathrm{QM-EFP1}} = \left\langle \Psi \middle| H^{\mathrm{QM}} + V^{\mathrm{Coul}} + V^{\mathrm{pol}} + V^{\mathrm{rep}} \middle| \Psi \right\rangle + E^{\mathrm{Coul}} + E^{\mathrm{pol}} + E^{\mathrm{rep}} \tag{2}$$

V^{Coul}, V^{pol}, and V^{rep} are one-electron contributions to the (unperturbed) quantum Hamiltonian H^{QM} due to Coulomb, polarization, and repulsion terms of EFP1 water fragments. E^{Coul}, E^{pol}, and E^{rep} are fragment-fragment Coulomb, polarization, and exchange-repulsion energies, respectively.

The Coulomb E^{Coul} term is evaluated using the classical multipoles up to octopoles centered at each atom and bond midpoint. The distributed multipole moments are obtained using the Stone distributed multipole analysis;[168,169] each water fragment has five points with distributed multipoles. The fragment-fragment Coulomb interactions consist of charge-charge, charge-dipole, charge-quadrupole, charge-octopole, dipole-dipole, dipole-quadrupole, and quadrupole-quadrupole terms. The Coulomb contribution V^{Coul} from a multipole point *k* to the *ab initio* Hamiltonian consists of four terms due to EFP charges *q*, dipoles μ, quadrupoles Θ, and octopoles Ω:

$$V_k^{\mathrm{Coul}}(x) = q_k T(r_{kx}) - \sum_{\alpha}^{x,y,z} \mu_\alpha^k T_\alpha(r_{kx}) + \frac{1}{3}\sum_{\alpha,\beta}^{x,y,z} \Theta_{\alpha\beta}^k T_{\alpha\beta}(r_{kx}) - \frac{1}{15}\sum_{\alpha,\beta,\gamma}^{x,y,z} \Omega_{\alpha\beta\gamma}^k T_{\alpha\beta\gamma}(r_{kx}). \tag{3}$$

T, T_α, $T_{\alpha\beta}$, and $T_{\alpha\beta\gamma}$ are electrostatic tensors of zero, first, second, and third rank, respectively; *x* is the electron coordinate; r_{kx} is the distance between the position of electron *x* and multipole point *k*.

To account for short-range charge-penetration effects, the charge-charge fragment-fragment terms and the charge-based term in the Hamiltonian are augmented by Gaussian like damping functions of the form $1-\beta\exp\left(-\alpha R^2\right)$. The α and β parameters were determined from a fit of damped multipole potential to the quantum Hartree-Fock potential on a set of points around the fragment. *R* is the distance between two multipole points or between a multipole point and an electron.

www.manaraa.com

The fragment polarization energy E^{pol} is evaluated as an interaction of induced dipoles of each fragment with the static field due to the Coulomb multipoles and the induced field due to the induced dipoles of the other fragments. The induced dipoles originate at the centroids of localized molecular orbitals (LMO), where (anisotropic) distributed polarizability tensors are placed. Each water fragment has five distributed polarizability points: at the oxygen (inner shell), the centers of the two O-H bonds, and at the centroids of the two lone pairs. The polarization energy of the fragments is calculated self-consistently using an iterative procedure. Thus, polarization accounts for some many-body effects that are important in aqueous systems.[170]

When a quantum region is present in the system, the induced dipoles of the fragments also interact with the electron density and nuclei of the quantum part, by means of V^{pol} a one-electron contribution to the quantum Hamiltonian:

$$V_k^{\text{pol}}(x) = \frac{1}{2}\sum_{\alpha}^{x,y,z}\left(\mu_\alpha^{\text{ind},k} + \tilde{\mu}_\alpha^{\text{ind},k}\right)T_\alpha(r_{kx}), \tag{4}$$

where μ^{ind} and $\tilde{\mu}^{\text{ind}}$ are the induced and the conjugate induced dipoles at the polarization point *k*. The total electric field acting on each fragment now consists of static and induced fields (due to the multipoles and induced dipoles of the other fragments), as well as fields due to the electron density and nuclei of the *ab initio* part. The total polarization energy of the QM/EFP1 system is:

$$E^{\text{pol}} = -\frac{1}{2}\sum_k\sum_{\alpha}^{x,y,z}\mu_\alpha^k\left(F_\alpha^{\text{mult},k} + F_\alpha^{\text{nuc},k}\right) + \frac{1}{2}\sum_k\sum_{\alpha}^{x,y,z}\tilde{\mu}_\alpha^k F_\alpha^{\text{ai},k}, \tag{5}$$

where F^{mult} is the field due to the static multipoles on the EFP fragments, and F^{ai} and F^{nuc} are fields due to the density and the nuclei of the quantum region. Polarization in a QM/EFP system is computed self-consistently using a two-level iterative procedure. The lower level, which takes care of the convergence of the induced dipoles, is identical to the fragment-only system. The wave function is kept frozen at this level. The lower-level iterative procedure exits when the induced dipoles are self-consistent and are consistent with the frozen *ab initio* wave function. At the higher level, the wave function is updated based on the converged values of the induced dipoles from the lower level. Convergence of the upper level is determined by convergence of wave function parameters (molecular orbital coefficients). Convergence of the two-level iterative procedure yields self-consistent induced dipoles and the *ab initio* wave function.

The remaining contribution to the *ab initio*-EFP interactions accounts for exchange-repulsion and charge transfer effects. This V^{rep} term is modeled in the form of Gaussian functions:

$$V_k^{\text{rep}}(x) = \sum_{i=1}^{2}\beta_{i,k}\exp(-\alpha_{i,k}r_{kx}^2) \tag{6}$$

www.manaraa.com

where *k* represents repulsion centers (atoms and center of mass), r_{kx} is the distance between a repulsion center and an electron. Parameters α and β are obtained by fitting the repulsion term to the difference between the HF energies of water dimers and a sum of electrostatic and polarization terms. A total of 192 water dimers were used in the fitting procedure. Fragment-fragment repulsion energy E^{rep} is modeled similarly to the *ab initio*-EFP1 repulsion. However, a single exponential function has been used instead of Gaussians.

The EFP1 water model has been combined with various wave functions in the quantum region, including HF, DFT, MCSCF and MP2. Excited state methods such as configuration interaction with single excitations (CIS[171]), TDDFT[172], and MRPT[173] have been interfaced with EFP1 as well. The general EFP2 model (see below) is interfaced with CIS(D)[174] (CIS with perturbative double excitations) and equations of motion (EOM)-CCSD.[175,176] Additionally, the EFP1 model has been combined with the FMO method.[177, 178] The interface between the EFP1 method and the polarized continuum model (PCM) has been also developed.[179,180,181,182] In order to describe heterogeneous catalytic systems in the presence of a solvent (the liquid-surface interface), an interface between the EFP method and the universal force field has been developed.[183] The implementations of the EFP1 and QM/EFP1 methods have been efficiently parallelized.[184]

Analytic gradients have been developed for the EFP1 and QM/EFP1 models[166,167] and the gradients have been used extensively in studies of water clusters and bulk,[185,186,187,188] aqueous reactions,[189,190] amino acid neutral-zwitterion equilibria,[179,191] and photochemistry in water.[171,172]

To summarize, the EFP1 water potential described above has proven to be a robust and useful model for treating interactions in water. However, the main drawback of the original version of the EFP1 model is that it is fitted to HF or DFT energies of water dimers. Therefore, this potential does not include long-range correlation effects such as dispersion. In order to improve the accuracy of the water model, two more variants of EFP1 have been developed. In one, referred to as EFP1/DFT[192], the B3LYP functional was used to obtain EFP Coulomb and polarization parameters, and the repulsion term was fitted to reproduce the B3LYP/DH energies in dimers. In the EFP1/MP2 variant[193], the potential is based on the MP2 method, i.e., the multipoles and polarizability tensors are obtained from MP2 calculations on a water monomer, and the repulsion is fitted to HF dimer energies. In order to incorporate correlation effects, an additional dispersion term of form $E^{\text{disp}} = \frac{C_6}{R^6} + \frac{C_8}{R^8}$ is added to fragment-fragment interaction energies. The C_6 and C_8 coefficients were obtained from a fit to MP2 correlation energies of the dimers. The original HF-based potential is now referred to as EFP1/HF.

To extend the EFP idea to a general solvent, one needs to design a way to accurately model exchange-repulsion and charge-transfer interactions without extensive fitting. Moreover, the dispersion term must be added to the model for a balanced description of π-stacking, Van der Waals complexes, and other weak

www.manaraa.com

interactions. This has been realized in a general formulation of the EFP method, often called EFP2[194,195,196]. The fragment-fragment energy in the EFP2 method consists of the following terms:

$$E^{\mathrm{EFP2}} = E^{\mathrm{Coul}} + E^{\mathrm{pol}} + E^{\mathrm{exrep}} + E^{\mathrm{disp}} + E^{\mathrm{ct}}, \tag{7}$$

where E^{exrep} is a general (not-fitted) exchange-repulsion energy, E^{disp} and E^{ct} are dispersion and charge-transfer energies, respectively. The EFP2 potential can be built for any solvent molecule in a MAKEFP run in GAMESS. However, solvents with flexible degrees of freedom (for example, long alcohols or chain hydrocarbons) should be treated with care since EFP is inherently a rigid-body model.

The Coulomb E^{Coul} and polarization E^{pol} terms in the EFP2 method are obtained similarly to those in the EFP1 models. That is, the multipoles at the atoms and bond-mid-points are generated by the DMA, and distributed polarizability tensors are calculated at the LMO centroids by using the coupled perturbed HF (CPHF) approach. Since the general EFP2 potential is intended to be used for various species, including biologically relevant highly polar and charged fragments, such as NH_4^+, H_3O^+, OH^-, F^-, Cl^-, metal cations, etc., special care should be taken for short-range and quantum effects. Therefore, several damping formulas have been developed and investigated in combination with the Coulomb and polarization energies.[197,198,199] In particular, Coulomb interactions at short distances can be damped either by exponential damping function applied to charge-charge term only[197]:

$$f_{kl}^{\mathrm{ch-ch}} = 1 - \frac{\alpha_l^2}{\alpha_l^2 - \alpha_k^2}\exp\left(-\alpha_k R_{kl}\right) - \frac{\alpha_k^2}{\alpha_k^2 - \alpha_l^2}\exp\left(-\alpha_l R_{kl}\right), \tag{8}$$

or by using a set of exponential-based damping functions applied to higher order Coulomb terms as well (charge-dipole, charge-quadrupole, dipole-dipole, dipole-quadrupole)[198,199]. Damping parameters α at each multipole distributed point (*k* or *l* in the above equation) are obtained in a MAKEFP run by fitting the damped classical Coulomb potential to the quantum (Hartree-Fock) potential on a set of points around the fragment. Another option is to estimate the short-range charge penetration energy using the spherical Gaussian overlap (SGO) approximation[200,199]:

$$E_{kl}^{\mathrm{pen}} = -2\left(\frac{1}{-2\ln\left|S_{kl}\right|}\right)^{1/2}\frac{S_{kl}^2}{R_{kl}} \tag{9}$$

Damping of the polarization energy is extremely important in order to avoid the “polarization catastrophe” that may happen due to breaking the multipole approximation at short separations between the fragments. Exponential and Gaussian damping functions have been investigated[199]; damping parameters for these functions (one parameter per polarization point, i.e., a LMO centroid) have been set to predefined values that were picked to be the same for all fragment types and all polarization points.

www.manaraa.com

The dispersion energy between fragments is calculated using the leading induced dipole-induced dipole $1/R^6$ term.[201] An empirical correction for the next $1/R^8$ term is added as one-third of the $1/R^6$ term. The resulting formula for the fragment-fragment dispersion energy is:

$$E^{\text{disp}} = \frac{4}{3}\sum_{k,l}\frac{C_{6,kl}}{R_{kl}^6} \tag{10}$$

where k and l are distributed (LMO) dispersion points on fragments A and B, respectively ($A \neq B$), R_{kl} is the distance between these points, and C_6 is obtained using the following integration:[201,82]

$$C_{6,kl} = \int_0^\infty \bar{\alpha}_k(i\nu)\bar{\alpha}_l(i\nu)d\nu\,, \tag{11}$$

where $\bar{\alpha}_{k(l)}$ is the 1/3 of the trace of dynamic (frequency dependent) polarizability tensor at the point k (or l), respectively. The integration is performed on the fly between all pairs of dispersion points of all fragments, using a 12-point Gauss-Legendre quadrature. Distributed dynamic polarizability tensors on each fragment are obtained from time-dependent HF calculations within the MAKEFP run in GAMESS.

Dispersion interactions are corrected for charge penetration effects. Two variants of damping functions have been employed within the EFP2 method. One is the Tang-Toennies damping formula[202] with damping parameter set to β=1.5[201]:

$$f_{kl}^{\text{TT}} = \left(1 - \exp(-\beta R_{kl})\sum_{n=0}^{6}\frac{(\beta R_{kl})^n}{n!}\right). \tag{12}$$

Another damping formula[199] uses the intermolecular overlap integrals S_{kl} between LMOs k and l on fragments A and B:

$$f_{kl}^{S} = 1 - S_{kl}^2\left(1 - 2\ln|S_{kl}| + 2\ln^2|S_{kl}|\right) \tag{13}$$

The latter formulation is parameter free and was shown to provide accurate dispersion energies up to very short (almost valence) separations between the fragments.

The exchange-repulsion energy in EFP2 is derived from the exact HF expression for the exchange-repulsion energy of two closed-shell molecules. Truncating the sequence at the quadratic term in the intermolecular overlap and applying an infinite basis set and the spherical gaussian overlap approximation[203] lead to the following expression for the exchange-repulsion energy:[204,205]

www.manaraa.com

$$E^{\text{exrep}} = \sum_{k,l} \left[\begin{array}{l} -4\sqrt{\dfrac{-2\ln|S_{kl}|}{\pi}}\,\dfrac{S_{kl}^2}{R_{kl}} - 2S_{kl}\left(\sum_{i\in A} F_{ki}^A S_{il} + \sum_{j\in B} F_{lj}^B S_{jk} - 2T_{kl}\right) \\ +2S_{kl}^2\left(-\sum_{J\in B}\dfrac{Z_J}{R_{kJ}} - \sum_{I\in A}\dfrac{Z_I}{R_{lI}} + 2\sum_{j\in B}\dfrac{1}{R_{kj}} + 2\sum_{i\in A}\dfrac{1}{R_{il}} - \dfrac{1}{R_{kl}}\right) \end{array} \right], \tag{14}$$

where *i*, *j*, *k*, and *l* are LMOs; *I* and *J* are nuclei, S and T are the intermolecular overlap and kinetic energy integrals, respectively; and F is the intramolecular Fock matrix. The overlap and kinetic energy integrals are calculated on the fly between each pair of LMOs on different fragments A and B. The intramolecular Fock matrix elements are pre-calculated as part of the MAKEFP run.

The charge transfer fragment-fragment energy E^{ct} is derived by considering the interactions between the occupied valence molecular orbitals on one fragment with the virtual orbitals on another fragment.[206,207] The charge transfer term results in significant energy-lowering in polar or ionic species. An approximate formula, based on a second order perturbative treatment of the intermolecular interactions, uses canonical HF orbitals of individual fragments and a multipolar expansion of the electrostatic potential *V* of the fragment. The charge transfer energy of fragment A induced by fragment B is approximated as:

$$E_{A(B)}^{\text{ct}} = 2\sum_{i}^{\text{occ}A}\sum_{n}^{\text{vir}B} \frac{V_{in}^B - \sum_{m}^{\text{all}A} S_{nm} V_{im}^B}{\left(1 - \sum_{m}^{\text{all}A} S_{nm}^2\right)\left(F_{ii}^A - T_{nn}\right)} \left[V_{in}^B - \sum_{m}^{\text{all}A} S_{nm} V_{im}^B + \sum_{j}^{\text{occ}B} S_{ij}\left(T_{nj} - \sum_{m}^{\text{all}A} S_{nm} T_{mj}\right)\right], \tag{15}$$

where V^{B} is the electrostatic potential on fragment B; *i*, *j*, *n*, and *m* are canonical orbitals. A similar expression for the energy $E_{B(A)}^{\text{ct}}$ is obtained when fragment A induces charge transfer in fragment B. Charge transfer is treated as an additive pairwise interaction, so the total charge transfer energy is a sum of the $E_{B(A)}^{\text{ct}}$ and $E_{A(B)}^{\text{ct}}$ terms of all pairs of fragments A and B. The charge transfer term is the most computationally expensive term in EFP2; therefore, since it is relatively small in non-polar or weakly polar systems, the charge transfer interaction is often omitted in EFP2 calculations.

Exact analytic gradients are available for all EFP2 fragment-fragment terms.[208,209] Similar to the EFP1 models, the general EFP2 potential can be used in Monte Carlo and molecular dynamics simulations. It was also shown that in Monte Carlo simulations, EFP can serve as an accurate importance function for a QM/EFP potential.[210] The general EFP2 model is also used in combination with the SMF method, as discussed below. A hybrid EFP-FMO method, called EFMO, is also described below.

www.manaraa.com

A general EFP2 ab initio-fragment interface is under development. The Coulomb and polarization quantum-EFP terms are treated similarly to the QM-EFP1 scheme, through inclusion of one-electron EFP terms into the QM Hamiltonian. Recently, a formula for the exchange-repulsion energy between quantum and EFP regions has been derived and implemented[211]. Implementation of the exchange-repulsion gradients is in progress, as is the formulation of the *ab initio*-EFP dispersion interaction.

In order to extend the general EFP model to covalently bound systems, a covalently bound *ab initio*/EFP interface has been developed[212]. The method is similar in spirit to that of Assfeld and Rivail[213] and is based on defining a buffer region that separates the QM and EFP parts. The buffer region consists of several LMOs that are kept frozen in the EFP calculations. In order to avoid variational collapse of the *ab initio* wave function into the buffer region, MOs of the *ab initio* part are kept orthogonal to the buffer LMOs. The buffer region creates a necessary separation between the *ab initio* and EFP regions such that the QM-EFP interactions can be considered to be non-bonded interactions and can therefore be treated in the same way as other EFP-QM interactions.

A variant of a flexible EFP potential has been developed by Nemukhin and co-workers[214,215] in which a covalently bound system (such as a polypeptide) is divided into small EFP fragments. Ab initio-EFP interactions are described in a similar manner to those in a QM-EFP1 model (with Coulomb, polarization, and repulsion terms), while fragment-fragment interactions are described at the level of a customary force field. This allows flexibility in a covalently bound molecule and optimization of internal degrees of freedom.

2.1.2 Explicit Polarization Potential

The original formulation of X-Pol by Gao[46,52] was called MODEL and formulated for homogenous clusters representing liquids. In 2007, MODEL was extended by Xie and Gao[216] to covalently connected fragments with the use of the generalized hybrid orbital method (GHO).[217] From the beginning, X-Pol was envisioned as the next generation force field, based on QM calculations of fragments, and it was suggested that some parameters are needed for the implementation. X-Pol is typically used with semiempirical methods that describe fragment wave functions. The distinct feature of the X-Pol method is the incorporation of many-body polarization through performing a double SCF calculation; that is, self-consistent calculations of fragments in the electrostatic field of the other fragments (similar to the EFP approach). In the X-Pol method, the electrostatic field `is usually represented by the field of Mulliken charges (often scaled by an empirical factor).

The basic equations of X-Pol can be summarized as follows. The total energy of the system, E_{tot}, divided into N fragments, is defined relative to the sum of the energies of non-interacting fragments E_I^0

$$E_{\text{tot}} = \langle \Phi | \hat{H} | \Phi \rangle - \sum_{I=1}^{N} E_I^0 \tag{16}$$

www.manaraa.com

where the total Hamiltonian is given by

$$\hat{H} = \sum_{I=1}^{N} \hat{H}_I^0 + \frac{1}{2} \sum_{I=1}^{N} \sum_{J \neq I} \hat{H}_{IJ} \tag{17}$$

and the total wave function F is the Hartree product of fragment wave functions Y_I.

$$\Phi = \prod_{I=1}^{N} \Psi_I \tag{18}$$

$\hat{H}_I^0$ is the electronic Hamiltonian of isolated fragment *I*, and the interaction potential is expressed as

$$\hat{H}_{IJ} = -\sum_{i=1}^{2M} \sum_{\beta=1}^{B} \frac{q_\beta^J}{\left| \mathbf{r}_i^I - \mathbf{r}_\beta^J \right|} + \sum_{\alpha=1}^{A} \sum_{\beta=1}^{B} \frac{Z_\alpha^I q_\beta^J}{\left| \mathbf{r}_\varepsilon^I - \mathbf{r}_\beta^J \right|} + E_{IJ}^{\mathrm{vdW}} \tag{19}$$

The first term in Eq. (19) describes the interaction of electrons *i* in fragment *I* with partial charges q on atoms b in fragment *J*. The second term in Eq. (19) describes the interaction of nuclear charges *Z* on atoms a in fragment *I* with partial charges on atoms b in fragment *J*. The third term describes the van der Waals (vdW) interaction energy usually represented by the Lennard-Jones potential between atoms in fragments *I* and *J*.

In applying the X-Pol method, one performs SCF calculations on the individual fragments in the electrostatic field of the point charges due to the other fragments, determined using the electron density of the other fragments, until full self-consistency is reached. The fragment boundaries are treated with frozen orbitals, whereby the four bonds on the typical boundary carbon atom are divided equally (meaning two bonds to each fragment) between two fragments. The first group of orbitals, which is inside the fragment, is fully SCF-optimized, while the other group of orbitals that belong to other fragments is frozen and contributes to the total Fock matrix and density, while remaining frozen. Polypeptides are usually divided at Ca atoms, which are the boundary atoms in X-Pol. Thus, the fragments correspond to peptide units as defined by IUPAC[218] rather than residues.

In 2008, Xie et al.[219] modified the formulation of X-Pol, in order to obtain rigorously analytic gradients. The modification was necessitated by the use of the Mulliken charges to represent the external field. In this method, individual fragment Fock operators are modified with additional terms derived from the variation of the total electronic energy with respect to the electronic states of all fragments. In 2009, Song et al.[220] extended the RHF version of X-Pol to DFT.

In 2010, Cembran et al.[221] proposed the X-Pol-X method, which adds inter-fragment exchange interactions. Different from the earlier X-Pol, this new formulation appears to belong to the two-step category in the

www.manaraa.com

classification of fragment-based methods. In the X-Pol-X formalism, the fragments in X-Pol are replaced by blocks, which may or may not be associated with specific groups of atoms, although the practical implementation still uses fragments as blocks. The blocks are described by block-localized wave functions (BLW). The total X-Pol-X energy is given by:

$$E^{\mathrm{X-Pol-X}} = \sum_{\mu\nu}^{K} D_{\mu\nu}\left(H_{\mu\nu} + F_{\mu\nu}\right) \tag{20}$$

where K is the total number of basis functions in the whole system and **H** and **F** are the one-electron and Fock matrices, respectively, as in an ab initio calculation for the whole system. The density matrix is defined as

$$\mathbf{D} = \tilde{\mathbf{C}}\left(\tilde{\mathbf{C}}^{T}\mathbf{R}\tilde{\mathbf{C}}\right)^{-1}\tilde{\mathbf{C}}^{T} \tag{21}$$

In Eq. (21) $\tilde{\mathbf{C}}$ is the matrix of occupied MOs, with non-zero blocks on the diagonal. Each block in the **D** matrix corresponds to the respective block in X-Pol-X. **R** is the matrix of the interblock AO-based overlap integrals (*a* and *b* denote blocks):

$$R_{\mu\nu}^{ab} = \left\langle \chi_{\mu}^{a} \middle| \chi_{\nu}^{b} \right\rangle \tag{22}$$

The interfragment exchange is apparently accounted for through the block-orthogonalization constraint $\left(\tilde{\mathbf{C}}^{T}\mathbf{R}\tilde{\mathbf{C}}\right)^{-1}$ although the interblock potential is described with Coulomb terms.

Practically, one still has to solve Hartree-Fock equations for blocks, with the appropriately constructed Fock matrix.[221] In X-Pol-X, in contrast to X-Pol, the free block energies are not subtracted from the QM energy as in Eq. (16). This suggests that X-Pol-X has become more sophisticated than a “next generation force field”.

In 2010, Gao et al.[72] proposed the generalized X-Pol method, GX-Pol. This approach has two formulations, based on (a) consistent diabatic configurations (CDC) and (b) variational diabatic configurations (VDC). The two approaches have some similarity with (a) the valence bond (VB) (or MCSCF) and (b) CI theories, respectively. The analogy is not complete, since the VB, MCSCF and CI methods all account for the occupation of virtual orbitals and describe electron correlation, whereas GX-Pol effectively accounts for quantum effects (such as charge transfer) between fragments in the RHF fashion. This GX-Pol approach is somewhat similar to the MP2 and CI in vibrational SCF (VSCF),[222] which accounts for the anharmonicity rather than the electron correlation.

The CDC approach is based on pairs of blocks X2; the energy is obtained by

www.manaraa.com

$$E_{\mathrm{X2}} = \left\langle \Theta_{\mathrm{X2}} \middle| \hat{H} \middle| \Theta_{\mathrm{X2}} \right\rangle \tag{23}$$

where the wave function for *M* blocks is given by

$$\Theta_{\mathrm{X2}} = \sum_{a=1}^{M} \sum_{b=a+1}^{M} c_{ab} \Psi_{ab}^{A} \tag{24}$$

The pair wave functions in Eq. (24) are

$$\Psi_{ab}^{A} = R_{ab}^{A} \hat{A}\left(\Phi_1 \ldots \Phi_{ab} \ldots \Phi_M\right), \tag{25}$$

where $\hat{A}$ is the antisymmetrizer and R_{ab}^{A} is the normalization constant. Φ_{ab} is the wavefunction of the pair of blocks (*ab*). The coefficients c_{ab} and other variables in the block wave functions are determined self-consistently. An extension to an arbitrary combination of blocks (i.e., more than pairs) was also proposed.

In the VDC approach, the energy *E* is obtained by diagonalizing a CI-like secular equations with each matrix element such as $H_{I,J}$ corresponding to the combination of pairs *I*=(*ab*) and *J*=(*rs*):

$$\begin{vmatrix} H_{11} - ES_{11} & \ldots & H_{1,M2} - ES_{1,M2} \\ \ldots & \ldots & \ldots \\ H_{M2,1} - ES_{M2,1} & \ldots & H_{M2,M2} - ES_{M2,M2} \end{vmatrix} = 0 \tag{26}$$

In Eq. (26) *M2=M(M-1)/2* is the total number of dimer configurations Ψ_{ab}^{A}, and the corresponding overlaps and Hamiltonian matrix elements are given, respectively, by

$$S_{ab,st} = \left\langle \Psi_{ab}^{A} \middle| \Psi_{st}^{A} \right\rangle \tag{27}$$

$$H_{ab,st} = \left\langle \Psi_{ab}^{A} \middle| \hat{H} \middle| \Psi_{st}^{A} \right\rangle \tag{28}$$

It should be noted that CDC and VDC give different energies, like CI and MCSCF do for electron correlation. This new formulation of GX-Pol can describe charge transfer between blocks, previously not considered in X-Pol. As for X-Pol-X, GX-Pol is a more general method than the original X-Pol.

2.2 Fragment Molecular Orbital Methods

Since the original formulation[58] during the last century, the FMO method[223,224] has undergone considerable development extending the applicability of the method to a wide variety of types of systems. A review of the

www.manaraa.com

FMO method in 2007 by Fedorov and Kitaura[225] was followed by a book[226,227] summarizing the activity of a number of research groups developing and applying FMO.

The gist of the FMO method is to incorporate high order interactions into low order expansions in terms of fragments, inspired by molecular interaction models.[70] In particular, the electrostatic interaction is treated at the full N-body order (M=N), where N is the total number of fragments; this is accomplished by incorporating the Coulomb field of all N fragments into the self-consistent monomer SCF cycle (see introduction for a more detailed discussion), while non-electrostatic interactions (exchange-repulsion, charge transfer and dispersion) are treated at a lower order, typically M=2 or 3. The fragments, their pairs (M=2: FMO2) and triples (M=3: FMO3) are explicitly treated quantum mechanically in the presence of the Coulomb (i.e., electrostatic) field of the whole system. In this manner, the low order non-electrostatic interactions are coupled to the electrostatics. In most systems, this is a reasonable approximation, because the non-electrostatic interactions are typically shorter-range.

Beginning with this assumption, it becomes acceptable to break a system into many smaller, localized pieces, treating the long-range effects of the full system using only a Coulomb operator. The driving motivation for this approximation and subsequent fragmentation is to find a route to drastically reduce the computational cost required for calculations. Even when the most efficient algorithms are used, for example, with MP2 or CCSD(T), the system size in terms of basis functions is the limiting factor. It is this roadblock in current *ab initio* methods that the FMO method attempts to overcome, first by reducing computational cost through fragmentation as mentioned above, and second by employing multi-level parallelization through the use of the generalized distributed data interface (GDDI).[1] It is this combination of theoretical approximations and exploitation of modern computational resources that has allowed the FMO method to perform all electron calculations on over 20,000 atoms.[228]

Fragment creation in the FMO method involves the detaching of bonds electrostatically, assigning two electrons from the broken bond to one fragment and none to the other. Where these bond detachments occur is left up to the user, relying on their own chemical intuition and knowledge of the system being investigated. A general guideline is to avoid detaching bonds involved in a considerable electron delocalization, e.g., within benzene rings. There are well-established standard fragments for polypeptides (fragmented at Ca atoms), nucleotides and saccharides. Metal ions have very considerable charge transfer to solvent[229] and represent an example of the need to define large fragments. When fragmenting a chemical system in this manner, the charge on fragments is not affected by bond detachment, so a negatively charged residue retains this same negative charge after fragmentation. The technical details of this are well illustrated elsewhere.[226,373] Note that the FMO, ELG and X-Pol methods appear to be among the few fragment methods that do not add caps (hydrogen or more involved) to detached bonds.

www.manaraa.com

The detachment of bonds in the FMO method can be done in two ways. In the original method, hybrid orbital projection (HOP) operators[230] are used to properly divide the variational space for the atom from which a bond is detached, because this atom has to be present in two fragments: This atom is present in one fragment just to describe the detached bond; the same atom is present in the second fragment since it contains the rest of the electrons that "belong" to that atom. Of course, when a bond is fragmented, there are two atoms (one at each end of the bond) that must be treated in this manner. This procedure is readily accomplished with projection operators, based on precomputed hybridized orbitals (e.g., sp^3) for the atoms at the ends of the fragmented bond. In this method, often referred to as the HOP method, there are no restrictions on the variational optimization of the occupied orbitals of the fragments, in contrast to the second approach described in the next paragraph. This HOP method is considered to be a good approach for very polar systems such as proteins with charged residues.

The second detachment method is based on adaptive frozen orbitals (AFO), [231] in which the electron density distribution is precomputed for an approptiate MO in a model system that represents the system of interest. This pre-computed electron density is subsequently frozen in the FMO calculation. The treatment of the variational space division for each bond-detached atom in a fragmentation process is very similar to the HOP procedure, although the AFO procedure is achieved in practie in a different manner by Fock matrix transformations. However, an important distinction in the AFO procedure is that there is a restriction on the variation of the occupied fragment orbitals on the atoms that are involved in the fragmentation (the detached bond orbitals). This is particularly important for systems in which the detached bonds are inevitably close to one another. This happens in covalent crystals and related systems such as nanowires, for which the AFO procedure is preferred. The HOP method is based on projection operators in the general form $|\phi\rangle\langle\phi|$. Such projection operators are used in many methods, for instance, in the core operator of model core potentials. [232] The FMO AFO approach was adapted from the EFP method, [212,233] which is in turn related to other methods, such as the GHO approach of Assfeld and Rivail.[217] The AFO method is similar to these other methods in that the density of some occupied orbitals is frozen, but differs in the variational space division, as well in other technical details such as the method to obtain the bond orbitals, and their orthogonality.

HOP-based approaches require a separate preliminary generation of hybrid orbitals. Although this is straightforward, it does add an extra step. To date, bonds have been detached at carbon and silicon[234] atoms. On the other hand, in the AFO approach the frozen orbitals are generated on the fly. Therefore, in principle any system can be computed, although a practical implementation may require an elaborate algorithm to properly build a model system in complicated cases. The AFO-based FMO approach has been applied to bonds detached at carbon and silicon.

Now, consider a summary of the basic FMO formulation. For a more detailed mathematical description the reader is referred to chapter 2 of Ref 141. The basic algorithm for calculating the energy of a system using the FMO method is as follows (see Fig. 2):

www.manaraa.com

(1) The initial electron density distribution is calculated for each monomer.

(2) The monomer Fock operators are then constructed using these densities and the energy of each monomer is calculated in the Coulomb bath of the rest of the system.

(3) Each of the monomer energies is iterated to self-consistency, leading to a converged ESP. This step in FMO development is usually called the monomer SCF or SCC.

(4) Fragment dimer calculations (FMO2) are then performed in the converged ESP of the rest of the system. Each dimer calculation is only performed once (not self-consistently).

(5) Optionally, fragment trimer calculations (FMO3) are performed next in the converged ESP of the rest of the system. Each trimer calculation is only performed once.

Within the two- and three-body FMO methods (FMO2 and FMO3, respectively), the total energy of the system can be written as:

$$E^{\mathrm{FMO2}} = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) \qquad (29)$$

$$E^{\mathrm{FMO3}} = E^{\mathrm{FMO2}} + \sum_{I>J>K}^{N} \left\{ (E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{IK} - E_I - E_K) - (E_{JK} - E_J - E_K) \right\} \qquad (30)$$

with monomer (*I*), dimer (*IJ*) and trimer (*IJK*) energies being obtained as described above. The beauty of the seemingly simplistic form of Eq. (29) is its ability to encapsulate the concepts of properly handling many-body effects, shown in the diagrammatic treatment[54] and the Green's function formalism.[235]

The Fock equation used in the FMO method;

$$\tilde{\mathbf{F}}^x \mathbf{C}^x = \mathbf{S}^x \mathbf{C}^x \tilde{\varepsilon}^x \quad x = I, IJ, IJK \qquad (31)$$

$$\tilde{\mathbf{F}}^x = \tilde{\mathbf{H}}^x + \mathbf{G}^x \qquad (32)$$

is a modified version of the standard form. The modification adds a term, $V_{\mu\nu}^x$, to the one-electron Hamiltonian, $\tilde{\mathbf{H}}^x$, that represents the ESP calculated during the monomer SCF:

$$\tilde{H}_{\mu\nu}^x = H_{\mu\nu}^x + V_{\mu\nu}^x + B \sum_i \left\langle \mu \middle| \phi_i^h \right\rangle \left\langle \phi_i^h \middle| \nu \right\rangle \qquad (33)$$

www.manaraa.com

In order to properly divide basis functions across fractioned bonds, a projection operator (HOP), $B\sum_i \langle \mu | \varphi_i^h \rangle \langle \varphi_i^h | \nu \rangle$, is also added to the Hamiltonian, where *i* runs over hybrid orbitals φ_i^h on all bond detached atoms in fragment *x*. The constant *B* is chosen to be sufficiently large to remove the corresponding orbitals out of the variational space. Typically, $B=10^6$ a.u..

One of the most important components of the FMO approach is the ESP, calculated during the monomer SCF process. The ESP of the full system takes the form

$$V_{\mu\nu}^x = \sum_{K(\neq x)} (u_{\mu\nu}^K + v_{\mu\nu}^K) \tag{34}$$

$$u_{\mu\nu}^K = \sum_{A \in K} \langle \mu | (-Z_A / | \mathbf{r} - \mathbf{r_A} |) | \nu \rangle \tag{35}$$

$$v_{\mu\nu}^K = \sum_{\lambda\sigma \in K} D_{\lambda\sigma}^K (\mu\nu | \lambda\sigma) \tag{36}$$

The term $u_{\mu\nu}^K$ represents the nuclear attraction contribution to the energy, with the two-electron contribution represented by the second term $v_{\mu\nu}^K$, expressed in terms of AOs m and n. Both of these terms are calculated for each of the surrounding monomers *K* with electron density D^K.

One might wonder if the addition of exchange integrals to the ESP in Eq. (36) could improve the results. Following an initial analysis[235] in 2005, it was later concluded[236] that due to the lack of orthogonality between fragment wave functions, the addition of exchange to ESPs in the SCC procedure actually decreases the accuracy. It is also possible to use point charges for all ESPs. This was shown to work particularly well for large basis sets by Fedorov et al.[237] These authors also proposed to screen point charges by introducing an exponent prefactor to the point charge potentials in Eq. (35). Such screening was found useful in particular for the AFO scheme when ESP are represented by point charges, and for molecular clusters the effects of screening were found to be small.

Even using the formulation described thus far, computational cost can still become excessive due to the increasing cost of the two-electron term, $v_{\mu\nu}^K$, contained in the ESP. Additional approximations can be used to reduce this cost by creating a cut-off value R_{app}. Two-electron terms outside of this boundary can be treated in a more approximate manner; however, this additional layer of complexity can decrease the accuracy of the method. This is due to the delicate balance among the approximations in different FMO terms.

www.manaraa.com

To understand this, consider a dimer *IJ* separated from a fragment *L*, which exerts an ESP upon *I*, *J* or *IJ*, schematically represented as *IJ*…*L*. If the distance between *J* and *L* is short (no ESP approximation), but that between *I* and *L* is long enough to apply the ESP approximation, there is a loss of balance, because in dimer *IJ* and monomer *J* the ESP due to *L* is computed without approximations, but in monomer *I* it is approximated. Consequently, pair corrections of the form $E_{IJ} - E_I - E_J$ include potentials due to *L* both with and without approximations. Due to the rapidly increasing number of dimers with system size, the error accumulates, creating a significant loss of accuracy in the total energy of the system.

To avoid the problems created by using distance based approximations to ESP, a reformulation of the energy expression is used[238] that is equivalent to Eq. (29) but more accurate when using approximations to the ESP:

$$E^{\mathrm{FMO2}} = \sum_{I}^{N} E'_I + \sum_{I>J}^{N} \Delta E'_{IJ} + \sum_{I>J}^{N} Tr(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \tag{37}$$

with the analogous expression[239] for FMO3 being;

$$\begin{aligned} E^{\mathrm{FMO3}} = & \sum_{I}^{N} E'_I + \sum_{I>J}^{N} \Delta E'_{IJ} + \sum_{I>J>K}^{N} (E'_{IJK} - E'_I - E'_J - E'_K - \Delta E'_{IJ} - \Delta E'_{IK} - \Delta E'_{JK}) + \sum_{I>J}^{N} Tr(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) \\ & + \sum_{I>J>K}^{N} [Tr(\Delta \mathbf{D}^{IJK} \mathbf{V}^{IJK}) - Tr(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}) - Tr(\Delta \mathbf{D}^{IK} \mathbf{V}^{IK}) - Tr(\Delta \mathbf{D}^{JK} \mathbf{V}^{JK})] \end{aligned} \tag{38}$$

where the new energy terms E'_x are defined as the internal *n*-mer energies with the ESP contributions subtracted out;

$$E'_x = E_x - Tr(\mathbf{D}^x \mathbf{V}^x) \quad x = I, J, IJ \tag{39}$$

and $\Delta \mathbf{D}^x$ is the difference density matrix, defined as;

$$\Delta \mathbf{D}^{IJ} = \mathbf{D}^{IJ} - \mathbf{D}^I \oplus \mathbf{D}^J = \mathbf{D}^{IJ} - \begin{bmatrix} \mathbf{D}^I & 0 \\ 0 & \mathbf{D}^J \end{bmatrix} \tag{40}$$

The above formulation in Eq. (37) allows the total energy to be calculated explicitly from only the dimer ESP $\mathbf{V}^{IJ}$. Eq. (38) contains both dimer and trimer ESPs, resulting in a more considerable effect of approximations on FMO3. Approximations can be applied to monomers and dimers separately, with the dimer ESP directly contributing to the total energy. The monomer ESP determines the monomer electron densities, and therefore only contributes to the total energy indirectly.

www.manaraa.com

The total density can be calculated in the same manner as the total energy (cf. Eq. (29)), and one-electron properties such as multipole moments are straightforward to calculate once the total density is obtained. For example, Sekino et al. computed moments up to quadrupoles. [240]

Eq. (38) includes both the two-body $\Delta\mathbf{D}^{IJ}$ and three-body $\Delta\mathbf{D}^{IJK} = \mathbf{D}^{IJK} - \mathbf{D}^{I} \oplus \mathbf{D}^{J} \oplus \mathbf{D}^{K}$ difference density matrices. This form of the energy equation, when applying three-body approximations to the ESP, partially retains the problem of ESP imbalance that the reformulation attempts to avoid. This retention of error is due to the uniform application of the distance definition R_{app}. To avoid this the distance based approximation was also reformulated[239] to treat the matrix elements of $V^{x}_{\mu\nu}$ block-wise so that;

$$\tilde{R}_{IJ,L}(\mu,\nu) = \begin{Bmatrix} R_{I,L}\ for \mu,\nu \in I \\ R_{J,L}\ for \mu,\nu \in J \\ R_{IJ,L}\ for \mu \in I, \nu \in J \text{ or } \mu \in J, \nu \in I \end{Bmatrix} \tag{41}$$

This allows for the application of approximations, for example in the I block of $V^{x}_{\mu\nu}$, to use the distance $R_{I,L}$ while using the uniform distance definition $R_{IJ,L}$ for the off-diagonal elements corresponding to both fragments I and J.

It is now appropriate to discuss the types of approximations applied to the ESP using the previously described energy expressions. Two levels of approximation can be used, both of which are applied to the two-electron integral term in Eq. (34). The first level of approximation, applied at intermediate distances, is the Mulliken atomic orbital population to the two-electron integrals.[241,242] Eq. (36) can be rewritten as;

$$v^{K}_{\mu\nu} \cong \sum_{\lambda \in K} (\mathbf{D}^{K}\mathbf{S}^{K})_{\lambda\lambda} (\mu\nu \mid \lambda\lambda) \tag{42}$$

The application of this approximation can reduce the computational cost of the two-electron integrals by a factor of N_B (number of basis functions).

The second level of approximation uses fractional atomic charges Q_A derived from the Mulliken atomic populations of the monomers. This approximation is applied at long distances and simplifies Eq. (36) as

$$v^{K}_{\mu\nu} \cong \sum_{A \in K} \left\langle \mu \middle| (Q_A / |\mathbf{r} - \mathbf{r}_{\mathbf{A}}|) \middle| \nu \right\rangle \tag{43}$$

effectively reducing the computational cost of the two-electron integrals by another factor of N_B.

www.manaraa.com

For far separated monomers the corresponding dimers do not need to be computed with the SCF procedure, as the monomer states already polarized by the ESP are very little perturbed by an explicit dimer SCF. However, there is a Coulomb interaction between such monomers that should be computed and added to the total energy,

$$E'_{IJ} \cong E'_I + E'_J + Tr(D^I u^{1,I(J)}) + Tr(D^J u^{1,J(I)}) + \sum_{\mu\nu\in I}\sum_{\rho\sigma\in J} D^I_{\mu\nu} D^J_{\rho\sigma} (\mu\nu \mid \rho\sigma) + \Delta E_{\mathrm{NR}} \qquad (44)$$

where $u^{1,I(J)}$ and $u^{1,J(I)}$ (cf. Eq. (35)) are the one-electron Coulomb potentials of the force exerted by fragment *J* on fragment *I*, and vise-versa. The electron-electron interaction (cf. Eq. (36)) and the nuclear repulsion are also added.

The final equations for the basic energy evaluation in FMO2 and FMO3 for RHF wave functions are Eqs. (37) and (38), respectively. These equations can be readily extended to other wave functions. A presentation of the application to other methods can be found in chapter 2 of Ref 141.

The original energy gradient[243] based on the HOP approach was proposed for FMO2 in 2001 and contained several approximations. The derivative of Mulliken point charges in ESP was added by Nagata et al.[244] in 2009, and the derivative of HOP terms was developed by Nagata et al. [245] in 2010. The FMO3 gradient was introduced in 2010 by Komeiji et al.[246] The FMO2 gradient for the AFO approach was proposed in 2009 by Fedorov et al. [247] A fully analytic FMO gradient without approximation was developed by Nagata et al.[248] in 2011 with the introduction of the self-consistent *Z*-vector (SCZV) procedure to obtain the exact derivatives of the dimer terms coupled to the electrostatic potential.

In 2002, Inadomi et al. [249] proposed the FMO-MO method, which according to the classification of fragment methods as discussed above, belongs to a different category from the rest of the FMO methods. While most FMO approaches are one-step approaches, the FMO-MO method is two-step, in which the density is obtained from the FMO calculation and used to construct the Fock matrix of the whole system in the exact ab initio fashion. Watanabe et al.[250] benchmarked the FMO-MO method for DNA, and Umeda et al.[251] proposed an efficient parallel algorithm for this method. It should be noted that the FMO-MO method (so far limited to RHF) can be used as a traditional ab initio method, because it produces the energy, MOs and other properties from the total Fock matrix.

Sekino et al.[240] in 2003 suggested taking the union of fragment MOs as a representation of the MOs of the whole system. Tsuneyuki et al.[252] in 2009 developed FMO-LCMO method, which has the limited but important aim of producing MOs near the HOMO and LUMO. This is accomplished by constructing a small MO-based Fock matrix collecting monomer and dimer contributions for a few orbitals of interest. For a similar motivation, Fedorov and Kitaura[236] in 2009 proposed to build the total AO-based Fock matrix for the whole system (FMO/F) from monomer and dimer Fock matrices. This is in contrast to the FMO-MO method, in which a similar matrix is computed in the ab initio fashion using the total density. Adding the exchange to this total

www.manaraa.com

Fock matrix (FMO/FX) considerably improves the description of the virtual MOs. All of these methods are FMO extensions for generating a specific property (the MOs), rather than a density-based method like FMO-MO, aimed at delivering the energy, MOs and other properties.

Sugiki et al.[53d] in 2003 began the process of extending FMO to wave functions other than RHF, by developing FMO2-based DFT, extended to FMO3 by Fedorov and Kitaura [253] in 2004; Shimodo et al.[254] in 2006 reported some accuracy benchmarks for FMO2-DFT. In 2004, Møller-Plesset perturbation theory was interfaced with FMO by Fedorov and Kitaura,[255] and an alternative implementation was reported by Mochizuki et al. [256,257] In 2009, RI-based FMO2-MP2 was proposed by Ishikawa and Kuwata [258] and Cholesky decomposition based FMO2-MP2 by Okiyama et al.[259] Mochizuki et al. developed an efficient parallel algorithm on the Earth Simulator for FMO-based MP2 in 2008 [260] and MP3 in 2010.[261] The FMO3-based MP2 method was developed for the energy by Fedorov and Kitaura[262] in 2009 and gradient by Mochizuki et al.[263] in 2011.

The first MD simulations using FMO were reported by Komeiji et al.[264] in 2003, indicating that the gradient accuracy problem (which has subsequently been solved by Nagata et al.[248]) requires larger fragments. Ishimoto et al.[265,266] in 2004-2005 used their Hamiltonian algorithm for FMO-MD. Dynamic fragmentation was proposed by Komeiji et al.[267] in 2009 to cope with the problems of bond breaking in FMO-MD. A review of FMO-MD methods was published by Komeiji et al.[268] in 2009. A path integral based FMO-MD was developed by Fujita et al.[269] in 2009, and an interface with the nuclear electron orbital (NEO) approach we developed by Pak and Hammes-Schiffer. A FMO-MD method with periodic boundary conditions was developed by Fujita et al.[270] in 2011.

Fedorov and Kitaura[271] in 2005 developed FMO-based MCSCF and coupled cluster[272] methods. As an important step to improve the efficiency, Fedorov et al.[273] introduced the multilayer FMO (MFMO) method, in which the system can be divided into several layers, and each layer is allowed to have its own basis set and level of theory. The original MFMO formulation[273] suggested that layers can have a different many-body expansion, which was later realized by Mochizuki et al.[263] under the name of FMO(3)-MP2, which should also be called FMO3-RHF:FMO2-MP2 for the case when the two layers include all fragments. In a quest for the exact solution, Maezono et al.[274,275] in 2006-2007 proposed the FMO-based quantum Monte-Carlo method.

One of the main targets of the FMO method is biochemical simulations, for which it is very important to properly consider solvent effects. The FMO-based PCM was developed for both the energy by Fedorov et al.[276] in 2006 and the gradient by Li et al.[277] in 2010. A Poisson-Boltzmann model was incorporated into FMO by Watanabe et al.[278] in 2010. For discrete models of explicit solvation, water can be described either as FMO or EFP fragments. The latter is due to the work of Nagata et al. for the FMO/EFP energy[177] in 2009 and gradient[178] in 2011.

www.manaraa.com

Complementing the ground state methods, FMO-based MCSCF was the first excited state method, reporting accuracy for singlet and triplet calculations of solvated phenol.[271] Mochizuki et al. developed FMO-based CI with singles (CIS) [279] in 2005 and added the perturbative treatment of doubles in 2007.[280] The FMO2-based TDDFT method was proposed by Chiba et al. [281,282] in gas phase in 2007 and solution[283] in 2008; the gas-phase gradient was developed[284] in 2009 and the FMO3-based TDDFT method was proposed[285] in 2010. For open shells, the ROHF-based MP2 and CC methods were developed by Pruitt et al.[286] in 2010.

All of these excited state methods are built around the central excited state fragment, which naturally contains the chromophore. In the FMO2 model, dimers that include the central fragment add explicit quantum effects, while other fragments only exert their electrostatic potential (ESP). In contrast to ground state FMO, FMO1 for excited states is quite useful as an approximation to full FMO2, if one is interested in the excitation energies. The working equation is very similar to Eq. (29), with the difference that the energies of the central monomer and dimers that contain the central monomer correspond to the excited state, while other energies are for the ground state.

There is an important difference between MCSCF on the one hand, and CI or TDDFT on the other. In the former, ESPs are computed for the central excited state (described by MCSCF), while in the latter the ESPs correspond to the ground state (RHF or DFT, respectively). The FMO1-based CI method has been described as a multilayer approach with a single chromophore fragment in the second layer,[279] but it can also be thought of as a single layer FMO1-CI, because these two descriptions are identical. On the other hand, for TDDFT, one can define FMO1-RHF:TDDFT, which corresponds to performing TDDFT calculations for the chromophore in the field of the ESP computed at the RHF level. This is different from FMO1-TDDFT, in which the ESPs are computed with DFT.

To analyze the effect of isotopic substitutions by using wave functions for nuclei, Ishimoto et al. [287] in 2006 developed a multicomponent method. Auer et al.[288] in 2009 proposed the FMO-based nuclear-electronic orbital (NEO) method. Mochizuki et al.[289] in 2006 developed a method for calculating the dynamic polarizability in the FMO framework. An interface to model core potentials (MCP) for the treatment of heavy atoms was added in 2006 by Ishikawa et al.[290] The total electrostatic potential was used to obtain atomic charges fitted to it by Okiyama et al. in 2007 [291] and 2009.[292] Sekino et al.[293] in 2007 implemented the calculation of NMR chemical shifts with FMO2. Gao et al.[294] in 2007 proposed an approach for evaluating NMR shifts from dimers, by using two fragmentations shifted with respect to each other, extended [295] later in 2010 to larger conglomerates of fragments but with a single fragmentation.

The FMO method is naturally suited to various analyses, as it provides information on fragments and their interactions that are naturally built into the method. In 2006, Amari et al.[296] developed the visualized cluster analysis of protein-ligand interactions based on the FMO method. Du and Sakurai[297] in 2010 proposed multivariate analysis of properties of amino acid residues.

www.manaraa.com

An important goal is to obtain more information than simple pair interaction energies (PIEs), also called interfragment interaction energies, IFIEs. Mochizuki et al.[298] in 2005 developed the configuration analysis for fragment interaction (CAFI), providing a means to extract for each orbital the stabilization component of the polarization and the charge transfer for IFIEs. Fedorov and Kitaura built EDA into the FMO method, developing the pair interaction energy decomposition analysis (PIEDA),[95] which decomposes PIEs into their electrostatic, exchange-repulsion, charge transfer and dispersion components. A fragment interaction analysis based on local MP2 (FILM) was proposed by Ishikawa et al.[299] in 2007. FILM allows extracting orbital-based contributions to the electron correlation component of the interaction energy.

The counterpoise correction[300] for the basis set superposition error (BSSE) conceptually is a two-fragment method, so it is natural to attempt to include it into fragment methods. Within the FMO framework, after the initial studies by others,[290,299] Ishikawa et al.[301] complemented the FMO energy in Eq. (29) with BSSE corrections. The validity of these corrections remains to be seen, as it has not been shown that they indeed bring the results closer to the complete basis set limit. One should also be aware that careful tests[302] indicate that in particular for MP2 the CP correction does not improve the results in this sense, especially for small basis sets.

Geometry optimizations for the FMO method were reported by Fedorov et al.[303] in 2007. Consequently, Ishikawa et al.[304] in 2010 suggested the partial energy gradient (PEG) method and Fedorov et al.[305] in 2011 introduced the frozen domain (FD) concept in their FMO/FD method. The focus of both of these methods is to perform efficient geometry optimization of a part of the system. In the FMO/FD approach (Fig. 3), the system is divided into several domains: frozen, polarizable and active. The electronic state of the frozen domain is computed only at the initial geometry, and the polarizable domain is recomputed for new geometries during geometry optimizations of atoms in the active domain. The main purpose of FMO/FD and its faster derivative FMO/FDD (in which the number of dimer calculations in the polarizable domain is reduced by omitting pairs of fragments except those in the active domain) is to efficiently perform geometry optimizations of the active sites in large molecular systems. To demonstrate the efficiency of the FMO/FDD approach, Fedorov et al.[305] optimized the geometry of a large protein-ligand complex with 19471 atoms at the B3LYP/6-31G(d) level of theory for the polarizable domain. This calculation took 32 h on six dual-CPU quad-core 2.83 GHz Xeon nodes.

There are several fragment methods that are closely related to the FMO method. The electrostatically embedded many-body expansion (EEMB) method [306,307,308,309] corresponds to FMO with constant ESP, which is not updated in the SCC fashion, and is often taken to be represented by point charges from molecular mechanics. Hirata et al. [310] in 2005 proposed representing the ESP by dipoles and in 2008[311] by point charges fitted to reproduce the fragment electrostatic potential.

In the effective fragment molecular orbital (EFMO) method Steinmann et al.[312] merged the ideas of EFP and FMO (Scheme I. In their method, no SCC calculation is performed, and all fragments are computed in vacuum (i.e., in the absence of the ESP). The many-body polarization is added, computed from fragment polarizabilities

www.manaraa.com

(as in the EFP method). The long range Coulomb interaction in Eq. (44) is computed using multipoles, also following the EFP method. Short range dimers (for which the interfragment distance R_{IJ} is smaller than a threshold R_{ESDIM}) are computed self-consistently, in vacuum. The EFMO energy expression is given by

$$E = \sum_{I=1}^{N} E_I^0 + \sum_{\substack{I>J \\ R_{IJ} \le R_{ESDIM}}}^{N} \left(E_{IJ}^0 - E_I^0 - E_J^0 - E_{IJ}^{\mathrm{ind}} \right) + \sum_{\substack{I>J \\ R_{IJ} > R_{ESDIM}}}^{N} \Delta E_{IJ}^{\prime \mathrm{ESM}} + E_{\mathrm{total}}^{\mathrm{ind}} \tag{45}$$

The superscript 0 indicates that the corresponding calculations are calculated in vacuum. The total induction energy $E_{\mathrm{total}}^{\mathrm{ind}}$ is corrected for double counting by subtracting SCF dimer contributions E_{IJ}^{ind} (because SCF dimers have implicit polarization). The long range electrostatic term $\Delta E_{IJ}^{\prime \mathrm{ESM}}$ of dimers, for which the interfragment separation is larger than R_{ESDIM}, is computed with multipoles, consistent with the approximate method of computing the energy of far separated dimers in FMO, Eq. (44).

In the EFP method, multipoles up to octopoles are used. In the EFMO method, the multipole expansion is truncated at the level of quadrupoles since higher multipoles become less important at large inter-fragment separations. Distributed multipoles and polarizability tensors for computing polarization terms are calculated on-the-fly for each individual fragment. The main differences between EFMO, FMO, and EFP are summarized in Scheme I. The computational cost of the EFMO method is significantly less than the cost of a corresponding FMO calculation. The EFMO method employs explicit *ab initio* calculations on the closely-lying dimers where quantum effects such as charge penetration and charge-transfer are important and the semi-classical EFP terms could become less accurate.

Several methods, which appeared earlier than EFMO, are related and differ in the details of computing EFP-related terms in EFMO. These methods include the polarizable multipole interaction with supermolecular pairs (PMISP)[313,314,315] and the approach by Beran.[316,317] PMISP involves capping fragments so it may be closer to methods like KEM rather than FMO, modified with the addition of the explicit polarization.

2.3 Molecular Tailoring Approach

Since the original formulation of the molecular tailoring approach (MTA)[119,318] the fragmentation scheme has undergone significant changes and improvements[319,320]. The general idea of the MTA is similar in nature to both MFCC and SMF, with aspects of the divide and conquer method incorporated as well. The MTA has been tested on a number of systems[321,322,323,324] including large p-conjugated systems, e.g., graphene sheets.[325] Further refinements of the method allow for geometry optimizations based on the adjustment of the cardinality of the fragments and their overlapping sections. This newest implementation of the MTA is appropriately named the cardinality guided molecular tailoring approach (CG-MTA).[319]

www.manaraa.com

The current implementation of the MTA is highly automated, with the user only needing to specify two values in addition to providing the coordinates of the target system. The first value specified is the maximum fragment size in terms of atoms per fragment. The second value, *R*-goodness (R_g), is central to the MTA fragmentation scheme and will be discussed in greater detail below.

The general fragmentation process for the MTA begins with the specification of the maximum fragment size and R_g and proceeds as follows. (1) Create an initial set of fragments by centering a sphere of radius R_g at each atom and assigning all atoms falling within the sphere to the fragment. During this process care is taken not to break aromatic rings or double bonds. Additional atoms are included or excluded based on the maximum fragment size. (2) Fragments created in step one are merged according to their proximity, while staying within the maximum fragment size set initially. (3) Merging of fragments is performed recursively depending on the maximum overlap of nearest neighbor fragments. This recursive merging stops once all fragments reach the maximum fragment size. (4) The final set of fragments is checked for the respective R_g value of the included atoms. (5) Broken bonds are capped using hydrogen atoms positioned along the appropriate bond vectors. (6) All intersecting portions of the merged fragments are computed, with the sign of each contribution being set according to $(-1)^{K-1}$ where K=number of fragments involved in the intersection. (7) The energy expression for the fragmentation scheme is created and the total energy of the system is calculated.

The value R_g used for fragment definition defines the radius of the sphere centered at atom *i* in the molecule of interest. All atoms that fall within this sphere are chosen to be part of fragment F_i. Since a sphere of radius R_g is created around all atoms in the system of interest at the beginning of the fragmentation process, each atom is bound to appear in more than one fragment even after the merging process in step 2. When this occurs, such atoms are evaluated for how accurately the local environment is represented by measuring the R_g value of the atom in each fragment. The largest of these values is chosen to represent the R_g of the atom in that particular fragmentation scheme. This is important since the value of R_g measures the quality of the fragmentation scheme based on how well the chemical environment around atom *i* is represented. A larger value of R_g will give a better representation of the chemical environment, providing more accurate results. It has been shown that values of R_g in the range 3-4 Å is necessary to achieve accurate results. While the minimum value of R_g ensures that each atom will have a certain representation of the surrounding chemical environment, the maximum fragment size assures that fragments do not become too large based on the computational resources available. The use of spheres centered around atoms also allows for an inherent treatment of non-bonded effects. For example, in a system such as a protein with a three dimensional structure, one can imagine atoms in parts of the system that are not connected to the atom center i falling within the sphere created by R_g. In such cases such, it is possible for only hydrogen atoms in the unconnected region to fall into the sphere. When this occurs, atoms bonded to these hydrogens are added to the fragment as well. In general, the fragmentation process is automated to account for such systems.

www.manaraa.com

Consider a general molecule *M* that, after the complete fragmentation procedure, is broken into two fragments F_1 and F_2. The overlap of these two fragments, $F_1 \bigcap F_2$, must be subtracted from the energy, giving the total energy expression for the system as:

$$E_M = E_{F_1} + E_{F_2} - E_{F_1 \cap F_2} \tag{46}$$

This expression for the total energy holds true for any system with only single overlaps between fragments. In the case of non-linear systems, where more than two fragments overlap in the same region, additional terms must be added or subtracted from the total energy. Taking these cases into consideration, a general expression for the total energy of a chemical system containing *K* fragments can be derived as:

$$E_M = \sum_i^K E_{F_i} - \sum_{i>j}^K E_{F_i \cap F_j} + \ldots + (-1)^{K-1} \sum_{i>j>k}^K E_{F_i \cap F_j \cap \ldots \cap F_K} \tag{47}$$

During the course of energy calculations for all fragments, the individual fragment densities are obtained during the standard SCF procedure. The complete density matrix for the full system can be constructed using the density matrix from each fragment. To obtain the most accurate density matrix for the system, individual density matrix elements are chosen from fragments based upon the quality of the R_g value calculated for each atom during the fragmentation process.

The derivative of the energy expression Eq. (47) with respect to coordinates X^m can be easily derived as:

$$\frac{\partial E_M}{\partial X^\mu} = \sum_i^K \frac{\partial E_{F_i}}{\partial X^\mu_{F_i}} - \sum_i^K \frac{\partial E_{F_i \cap F_j}}{\partial X^\mu_{F_i \cap F_j}} + \ldots + (-1)^{K-1} \sum_i^K \frac{\partial E_{F_i \cap F_j \cap \ldots \cap F_K}}{\partial X^\mu_{F_i \cap F_j \cap \ldots \cap F_K}} \tag{48}$$

where $X^\mu_{F_i}$ refers to the nuclear coordinates of atom μ in fragment F_i and $X^\mu_{F_i \mid F_j}$ refers to the coordinates of the overlapping section of fragments F_i and F_j. An analogous expression for the Hessian can be written as:

$$\mathbf{H} = \sum_i^K \mathbf{H}_{F_i} - \sum_i^K \mathbf{H}_{F_i \cap F_j} + \ldots + (-1)^{K-1} \sum_i^K \mathbf{H}_{F_i \cap F_j \cap \ldots \cap F_K} \tag{49}$$

Additional capabilities have been added to the CG-MTA including correlated methods such as MP2 and RI-MP2, [326] course grained parallelization[319] and the calculation of vibrational frequencies.[327]

2.4 Kernel Energy Method

The original formulation of the kernel energy method (KEM) [328] was tested on a number of systems.[329,330,331,332] Further improvement to the method[333] allowed the KEM to be applied to systems beyond

www.manaraa.com

its original limitation to peptides and polymers. Based loosely on the energy decomposition analysis, the kernel energy method takes into account fragment or "kernel" interactions up to fourth order terms.[334] More recent improvements have enabled the KEM to accurately model p-conjugated systems, most notably graphene.[335]

The purpose for developing the KEM was to perform quantum calculations on systems with biological importance such as proteins. By considering a system to be composed of individual "kernels" with known atomic coordinates, the contributions of all kernels in a system can be combined to provide properties for the full system. In this way the energy and properties of large systems can be obtained even if calculations on the full system are impossible. A few simple rules govern a calculation using the KEM, the most important being the requirement that each atom must be present in some kernel once and only once.

Once a system has been divided into separate kernels, care must be taken to remove any dangling bonds at the periphery of the kernels through the use of hydrogen caps. After all dangling bonds are capped, each kernel energy is evaluated, followed by all "double kernel" calculations of nearest neighbor kernels. In the original method, only those kernels covalently bonded to one another were considered during double kernel calculations. However, during subsequent development, the method was modified to include separated kernels during the double kernel evaluation, providing a more accurate description of the systems tested.

The sum of kernel contributions to the total energy of a system can be written mathematically as:

$$E_{\text{total}} = \sum_{i=1;\,j=i+1}^{n-1} E_{ij} - \sum_{i=2}^{n-1} E_i \tag{50}$$

E_{ij} represents the energy of two covalently bonded kernels and E_i is the energy of a single kernel. During further development, the definition of double kernel calculations was extended to include all double kernels, not only those connected by a covalent bond. The modified formula for the total energy is then written as:

$$E_{\text{total}} = \sum_{i=1;\,j=i+1}^{n-1} \left(\sum_{\substack{i=1 \\ j=i+m}}^{n-m} E_{ij} \right) - (n-2) \sum_{i=2}^{n-1} E_i \tag{51}$$

or, written more intuitively:

$$E_{\text{total}} = \sum_{1 \le i \le n} E_i + \sum_{1 \le i < j \le n} \Delta E_{ij} \tag{52}$$

www.manaraa.com

ΔE_{ij} is defined as the interaction energy of two kernels ($\Delta E_{ij} = E_{ij} - (E_i + E_j)$). To improve the accuracy of the method beyond the inclusion of disconnected double kernel calculations, both triple and quadruple kernel calculations were implemented. The total energy including triple kernel contributions is:

$$E_{\text{total}} = \sum_{1 \le i \le n} E_i + \sum_{1 \le i < j \le n} \Delta E_{ij} + \sum_{1 \le i < j < k \le n} \Delta E_{ijk} \tag{53}$$

with the expression for ΔE_{ijk} following that of ΔE_{ij} ($\Delta E_{ijk} = E_{ijk} - (E_i - E_j - E_k) - (\Delta E_{ij} - \Delta E_{ik} - \Delta E_{jk})$). Similarly, inclusion of quadruple kernel energies gives the expression:

$$E_{\text{total}} = \sum_{1 \le i \le n} E_i + \sum_{1 \le i < j \le n} \Delta E_{ij} + \sum_{1 \le i < j < k \le n} \Delta E_{ijk} + \sum_{1 \le i < j < k < l \le n} \Delta E_{ijkl} \tag{54}$$

and ΔE_{ijkl} follows from the expressions for ΔE_{ij} and ΔE_{ijk}. Although the use of quadruple kernels is not typically required, addition of these terms can be advantageous for large systems or if a computer with many nodes and cores is available. In most cases, inclusion of double and triple kernel interactions is sufficient to achieve a high level of accuracy compared to fully ab initio calculations.

Applications of the KEM have included proteins[328], as well as other biologically relevant systems such as DNA[329] More recently the KEM has applied successfully to systems containing extended aromatic character[335]. Application to systems with such diffuse electrons is typically a failing point for most fragmentation methods, however the KEM overcomes this deficiency through the use of a new bond fractioning scheme. Instead of fractioning single bonds perpendicular to the direction of bonding, the KEM method employs a "fissioning" process where the aromatic bonds are divided in half parallel to the direction of bonding. This creates two aromatic bonds, one in each of the two kernels created. Breaking a conjugated system such as graphene into kernels using this process has produced results accurate to within 1 kcal/mol of full ab initio HF and MP2 calculations. Recent development of the KEM includes a generalized fragmentation scheme[336] based on the approach of Deev and Collins[120] aims to increase the computational efficiency through elimination of extraneous double, triple and quadruple kernel calculations.

2.5 Molecular Fractionation with Conjugated Caps and Related Fragmentation Methods

2.5.1 MFCC

As with many fragmentation methods, the molecular fractionation with conjugate caps (MFCC) approach[117,337,338,339,340,341,342,343,344] attempts to reduce computational costs and provide a means to calculate interaction energies, but specifically for protein-ligand systems. The original formulation of the MFCC approach[337] fractioned only peptide bonds to enable the calculation of protein-ligand binding energies. The

www.manaraa.com

fractioned bonds are then capped with so called “concaps” that resemble the local environment of the fragment. By adding together the individual contributions of the fragments and subtracting the contributions from the merged concaps, the total interaction energy of the protein-ligand system can be calculated. An important difference between the MFCC method and other fragmentation methods that employ capping of fractioned bonds is the nature of the caps used. Instead of simple hydrogen caps, the caps in the MFCC approach are formed using portions of the neighboring sections of the molecule. This provides both an efficient method for choosing caps as well as including a representation of the local environment during individual fragment calculations. Further developments have provided the ability to fraction disulfide bonds and allowed for the inclusion of non-bonded interactions in globular proteins.

As mentioned previously, the focus of the original formulation of the MFCC approach was to break a protein into its constituent amino acids and calculate the interaction energies between individual protein fragments and the ligand of interest. The simplest example is a protein *P* composed of *N* amino acids;

$$P = nA_1A_2A_3A_4...A_N \tag{55}$$

where A_i represent the individual amino acids and *n* distinguishes the N terminal end of the protein

$$n = \mathrm{NH_3^+(NH_2)} \tag{56}$$

with the opposite end, A_N, representing the C terminal tail;

$$A_N = R_N\mathrm{CHCOO^-}(R_N\mathrm{CHCOOH}) \tag{57}$$

To calculate the interaction between protein *P* and an arbitrary molecule *M*, the protein is divided into single amino acid fragments across the homolytically broken N-C peptide bonds (Figure 4). This creates *N* fragments with either one or two unpaired electrons located where the N-C bond(s) used to be. To avoid this unnatural electronic state, each fragment is assigned either one or two concaps, C_{ap}^1 and C_{ap}^{1*}. The main purpose of the caps[345] is to complete the valency requirements of the “dangling” bonds left over after fractionation.

Consider the fractioning of the bond in a simple two amino acid (dipeptide) system. The bond fractioning and subsequent capping gives;

$$A_1A_2 = A_1C_{ap}^1 + C_{ap}^{1*}A_2 - C_{ap}^{1*}C_{ap}^1 \tag{58}$$

The energy contained in A_1A_2 can be represented by the sum of the individual amino acid fragments minus the artificial molecule create by joining the two concaps. One can similarly break a larger tripeptide system of three amino acids to give three fragments and the corresponding concaps:

www.manaraa.com

$$A_1A_2A_3 = A_1C_{ap}^1 + C_{ap}^{1*}A_2C_{ap}^2 + C_{ap}^{2*}A_3 - C_{ap}^{1*}C_{ap}^1 - C_{ap}^{2*}C_{ap}^2 \quad (59)$$

These two simple examples illustrate the two cases of singly and doubly capped fragments, with the general systems of interest consisting of both the protein and the general ligand *M*. Note that the original MFCC formulation was only concerned with the interaction energy between a rigid protein and ligand, foregoing the subsequent intramolecular energy calculation for the full system. The total interaction energy can be represented by;

$$E(M-P) = \sum_i^N E(M - C_{ap}^{i-1*}A_iC_{ap}^i) - \sum_i^{N-1} E(M - C_{ap}^{i*}C_{ap}^i) \quad (60)$$

The term $E(M - C_{ap}^{i-1*}A_iC_{ap}^i)$ represents the interaction energy between molecule *M* and the capped protein fragment $C_{ap}^{i-1*}A_iC_{ap}^i$. The second sum of terms represents the interaction of molecule *M* with the artificial molecule formed by connecting concaps, $C_{ap}^{i*}C_{ap}^i$. The regularity of amino acid bonding provides a simple choice for the concaps. Since only peptide bonds are being broken, the N-terminus side of each amino acid can be capped with an NH_2 group and the C-terminus side can be capped with $R_{i+1}C_\alpha H_2$ (see Fig. 4). This choice of caps ensures that the valence requirements of the broken bonds are complete and the approximate chemical environment around the fragment is being properly represented. Another benefit of choosing caps in this way is the nature of the artificial molecules formed when joining caps, creating reasonable molecular species such as $H_2NR_{i+1}C_\alpha H_2$. The most recent implementation of the MFCC method employs two different sizes of concaps,[346] a "small" and a "large" version. The small concap only extends across the nearest neighbor amino acid, while the large concap extends across the two nearest neighbor amino acids. Use of the large concap provides the obvious benefit of increased accuracy, with the tradeoff being an increase in computational cost.

In the special case where disulfide bonds are fractioned,[347] for example a bond between two cystines, an additional term must be added to Eq. (60);

$$\begin{aligned} E(M-P) &= \sum_i^N E(M - C_{\mathrm{ap}}^{i-1*}A_iC_{\mathrm{ap}}^i) \\ &- \sum_i^{N-1} E(M - C_{\mathrm{ap}}^{i*}C_{\mathrm{ap}}^i) - \sum_i^{N-1} E(M - DC_{\mathrm{ap}}^{i*}DC_{\mathrm{ap}}^i) \end{aligned} \quad (61)$$

DC_{ap}^{i*} and DC_{ap}^i can be either MeS or HS caps, with the MeS caps shown to give slightly more accurate results.

www.manaraa.com

Further development of the MFCC approach[348,349] addressed the limitation of calculating only the interaction energy by allowing for the calculation of the total electron density, electrostatic potential and dipole moment of proteins. Using the basic dipeptide example in Eq. (58), the total electron density of the system, ρ, can be expressed as:

$$\rho = \rho_{A_1} + \rho_{A_2} - \rho^{cc} \tag{62}$$

where ρ_{A_1} and ρ_{A_2} represent the densities of the capped amino acid fragments after fractionation and ρ^{cc} representing the density of the merged concap. By calculating the densities of the separate fragments, the total density of the system can be determined. One advantage of this formulation lies in the choice of concaps after fractionation. If the cap chosen for fragment A_1 includes the entirety of fragment A_2, then Eq. (62) is exact, giving the correct limiting behavior if the caps are chosen to be sufficiently large. For a system of N amino acids, Eq. (62) can be generalized as;

$$\rho = \sum_{i=1}^{N} \rho_i - \sum_{i=1}^{N-1} \rho_i^{cc} - \sum_{i=1}^{N_d} \rho_i^{dc} \tag{63}$$

where the third summation over ρ_i^{dc} is only included when disulfide bonds are fractioned. Using this representation of the total density of the system, the electrostatic potential can easily be obtained as

$$\phi(\mathbf{r}) = -\int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}' \tag{64}$$

with the total electrostatic potential derived through the combination of individual electrostatic potentials of the fragments:

$$\phi = \sum_{i=1}^{N} \phi_i - \sum_{i=1}^{N-1} \phi_i^{cc} - \sum_{i=1}^{N_d} \phi_i^{dc} \tag{65}$$

In a similar fashion, the dipole moment, μ, of the protein can be calculated as

$$\mu = \sum_{i=1}^{N} \mu_i - \sum_{i=1}^{N-1} \mu_i^{cc} - \sum_{i=1}^{N_d} \mu_i^{dc} \tag{66}$$

Obtaining the total electron density of the system provides a route to calculating the total energy of the system,[346] either through the standard SCF method or by using Kohn-Sham orbitals and DFT formalism. The

www.manaraa.com

implementation of the MFCC approach that allows for the calculation of the total energy using the density matrix (DM) has been termed the MFCC-DM approach.

Two versions of the MFCC-DM approach,[349] the "simple" approach (MFCC-SDM) and the "ghost" approach (MFCC-GDM), are available. These two variants are the result of using extra hydrogens as part of the conjugate caps. If the MFCC approach were exact, then the density matrix elements associated with the hydrogen atomic orbitals (AOs) would be effectively zero. Since the MFCC approach is not exact in practice, these matrix elements add small contributions to the density matrix. The two implementations, MFCC-SDM and MFCC-GDM, differ in how these extraneous hydrogen AOs are handled. In the MFCC-SDM approach the matrix elements are simply neglected, whereas in the MFCC-GDM approach the extra hydrogens are treated as ghost atoms and the matrix elements are accounted for explicitly.

Other than how the extraneous hydrogen atoms are handled, the most important difference between the two MFCC-DM implementations is the structure of the density matrix. In the MFCC-SDM approach the number of electrons is not exactly conserved since the contributions from the extra hydrogen atoms are simple ignored. In the MFCC-GDM approach this deficiency is addressed, since the extraneous hydrogen atoms are treated explicitly as ghost atoms in the system. The trade-off for the exact treatment of the density matrix in the MFCC-GDM approach is an increase in computational cost compared to the MFCC-SDM approach. In practice it has been shown that both approaches produce sufficiently accurate results when compared to full *ab initio* HF calculations.[349]

To improve the description of globular macromolecules with two- and three-dimensional structures, the addition of non-bonded or "through-space" interactions was implemented.[350] Currently this includes a description of two-body interactions only due to the negligible contribution of higher order interactions in proteins. In general, if two fragments are separated by more than one fractioned bond, they may still be adjacent to one another in the structure of the macromolecule. Following the EDA, the interaction of these two fragments can be calculated by;

$$\Delta E^{(2)} = \sum_i \sum_j [E(A_i A_j) - E(A_i) - E(A_j)] \tag{67}$$

Instead of capping these fragments in the same fashion as described earlier, these separated fragments are capped only with hydrogen atoms. This avoids over-counting the two-body correction as well as simplifying the expression needed to calculate the "through-space" two-body correction. This correction can be added to the previous representation for the full energy of the system, giving

$$E = \sum_{i=1}^{N} E(A_i) - \sum_{i=1}^{N-1} E(A_i^{\mathrm{cc}}) - \sum_{i=1}^{N_d} E(A_i^{\mathrm{dc}}) + \Delta E^{(2)} \tag{68}$$

www.manaraa.com

This new expression for the total energy of the system is termed the energy-corrected MFCC (EC-MFCC) approach.

In addition to the MFCC formalism described above, a number of other capabilities have been added recently. These capabilities include gradients for geometry optimizations[351,350] and the hybrid generalized molecular fractionation with conjugate caps/molecular mechanics (GMFCC/MM) approach[352] that uses MM for long range interactions. The addition of the conductorlike polarizable continuum model (MFCC-CPCM)[353] and the pairwise interaction correction to the density matrix formulation of the MFCC approach (MFCC-DM-PIC),[354] allow for the calculation of the electrostatic solvation energy of macromolecules and the treatment of short range polarization interactions, such as hydrogen bonding, respectively. Additionally, the electrostatic field-adapted molecular fractionation with conjugate caps (EFA-MFCC)[355] approach improves the ability to treat charged systems by adding a description of the surrounding environment using point charges. All of these improvements to the original formulation provide a means to perform calculations on a variety of large macromolecular systems while including a number of important intermolecular interactions.

2.5.2 Generalized Energy-Based Fragmentation Approach

The generalized energy-based fragmentation (GEBF) approach has been proposed[118] as a reformulation of the EFA-MFCC approach. This method builds upon the fragmentation scheme of the MFCC approach, but improves upon the description of the environmental electrostatic field added in the EFA-MFCC approach. The atoms far separated from the fragment of interest are represented as point charges and included in the *ab initio* calculation on the fragment. The main improvement over the EFA-MFCC approach is the inclusion of electrostatic interactions with polar groups, improving the ability of the method to calculate dipole moments and static polarizabilities.

A new fragmentation scheme is also introduced for the GEBF approach, distinguishing the method from previous MFCC implementations. The general fragmentation scheme consists of the same ideas, dividing a large system into smaller fragments and capping each fragment with the neighboring fragment to conserve the valence requirements of the broken bonds. Any dangling bonds left, such as those on the capping fragment, are capped with hydrogen atoms. However, the inclusion of neighboring fragments is not restricted to those exclusively covalently bonded to the fragment; non-bonded fragments are included as well. By choosing the number of fragments to be included in each quantum calculation using a distance based cut-off, the size of fragments and therefore the accuracy of the method can be easily varied depending on computational resources.

Consider a linear system, *M*, of six fragments;

$$M = m_1 m_2 m_3 m_4 m_5 m_6 \tag{69}$$

www.manaraa.com

with connectivity represented in Fig. 5. Starting with m_1, derivative subsystems are formed including the central fragment (m_1) and all fragments covalently and, in this case, hydrogen bonded to m_1. This process is then repeated for all m_i (i = total number of fragments) fragments in the full system M, producing i subsystems in the general case. Subsystems that may be present in a larger subsystem are eliminated to avoid double counting. For example, the subsystem $m_1m_2m_3m_4$ appears in the subsystem $m_1m_2m_3m_4m_6$ and should therefore be eliminated from consideration. In this case, two subsystems are eliminated, leaving four unique subsystems.

The remaining subsystems are then checked for double counting of fragment interactions. For example, if the four-body interaction $m_1m_3m_4m_6$ occurs more than once in the four derivative subsystems, a complementary subsystem must be built and subtracted from the sum of the subsystems. This process is followed through for all n-body interactions down to single fragment terms, with each subsystem being assigned a coefficient C_i of either 1 or -1. The total energy of the system M can then be represented as;

$$E = \sum_{i}^{N} C_i E_i \tag{70}$$

where N is the total number of subsystems. This fragmentation scheme provides nearly all of the three- and four-body interactions, with a small number being neglected. However, these neglected terms were shown to be a small source of error.[118] The final requirement of the fragmentation process in the GEBF approach is for the net number of hydrogen atoms used for capping to be zero.

After the fragmentation and construction of derivative subsystems is complete, each subsystem calculation is performed in the field of point charges on all other atoms. The partial charges Q_A used in the GEBF approach are derived from the natural population analysis (NPA)[356,357] and computed for the central fragment only during an initial HF or DFT calculation on each subsystem. These point charges are then incorporated into a second HF or DFT calculation for all subsystems. Including the point charges in the energy expression, the total energy of the system is

$$E = \sum_{i}^{N} C_i \tilde{E}_i - (\sum_{i}^{N} C_i - 1) \sum_{A} \sum_{B>A} \frac{Q_A Q_B}{R_{AB}} \tag{71}$$

$\tilde{E}_i$ is the total energy of the i-th subsystem that includes the effects of the electrostatic field. The derivatives of this equation have also been derived,[358] allowing for geometry optimizations and vibrational frequency calculations. The most recent improvements to the method include an algorithm for automatic fragmentation and derivative subsystem construction for general molecules.[359]

2.5.3 Other MFCC-related methods

www.manaraa.com

A number of more recent approaches have been proposed based on the general fractionation scheme suggested in the MFCC method. An extension of the frozen-density embedding (FDE)[360] scheme to the MFCC approach uses overlapping electron densities of different subsystems to provide a more accurate representation of the surrounding environment, as well as a better representation of the total electron density. Another approach[361] rigorously derives the additivity rule used in the MFCC approach within the semilocal DFT formalism. Finally, the multilevel fragment-based approach (MFBA)[362] employs a fragmentation scheme that is similar to the MFCC formulation (using hydrogen for caps instead of nearest neighbor fragments), while using different levels of *ab initio* theory for non-bonded fragment calculations depending on the fragment separation.

The polarized protein-specific charge (PPC) method proposed by Ji et al.[363] in 2008 is based on fitting atomic charges to MFCC-derived electrostatic potentials, analogously to the FMO-derived charges[291,292] developed in 2007. The PPC method was applied[364,365,366,367,368,369,370,371] to a number of studies showing the importance of the polarization missing in commonly used force fields.

2.6 The Systematic Molecular Fragmentation Method

A more recent approach[120, 372] to fragmenting large chemical systems is the systematic molecular fragmentation (SMF) method. Originally designed to treat large proteins, polymers and surfaces, the SMF method takes a unique approach to fragmenting systems. In the same vein as the other fragmentation methods, SMF employs highly accurate methods, such as MP2 or CC, to perform calculations on the smaller fragments, obtaining the total energy of the system. However, instead of having uniquely defined fragments, the SMF method uses overlapping sections of the molecule to account for interactions between fragments. By performing the fragmentation in this way, each atom "feels" the presence of the full system without the use of an externally applied field. Non-bonded interactions are also accounted for between far separated regions of the system, originally using simple electrostatics, but more recently by using the EFP method to obtain more accurate results.[373] The SMF method has been used to describe the isomerization of DNA helices[373] and retinal,[374] as well as to give a proper description of potential energy surfaces of chemical reactions.[375]

The underlying premise behind the SMF method is in thinking of a chemical system as a collection of single bonded functional groups. The definition of a functional group is central to the fragmentation, allowing for different levels of fragmentation (larger individual fragments) to increase the accuracy of the individual calculations. This also allows for a better description of many-body effects, as any atom in the fragment would be influenced by more of the system during the full *ab initio* calculations. The general fragmentation scheme is best described using a linear molecule, *M*, of arbitrary length *K*;

$$M = G_1 G_2 G_3 ... G_K \tag{72}$$

www.manaraa.com

The "super" molecule can be broken into two distinct molecules by stretching the bond between G_{n-1} and G_n to infinity. The bond is broken homolytically, assigning one electron from the broken bond to G_{n-1} and the other to G_n. To avoid charged fragments created by such a scheme, hydrogen atoms are used as "caps" on each of the fragments. The overall fragmentation creates two molecules:

$$M \rightarrow M_1 + M_2 \tag{73}$$

which are composed of;

$$M_1 = G_1 G_2 G_3 ... G_{n-1} H^{(n-1)} \tag{74}$$

$$M_2 = H^{(n)} G_n G_{n+1} ... G_K \tag{75}$$

where $H^{(n-1)}$ and $H^{(n)}$ are the hydrogen caps for their respective fragments, located along the direction of the broken bond at a chemically sensible distance for the specific *GH* bond.[120]

The energies of these fragments can then be calculated and are related by:

$$E(M) = E(M_1) + E(M_2) + dE_1 \tag{76}$$

where dE_1 represents the energy change created by the bond breakage. To allow for overlapping fragments, it is acknowledged that the fragmentation choice just described is not the only possible fragmentation scheme. The complete molecule *M* could also be broken at some other single bond, creating different fragments, M_3 and M_4, giving the following energy expression:

$$E(M) = E(M_3) + E(M_4) + dE_2 \tag{77}$$

where

$$M_3 = G_1 G_2 G_3 ... G_{i-1} H^{(i-1)} \tag{78}$$

$$M_4 = H^{(i)} G_i G_{i+1} ... G_K \tag{79}$$

By performing both fragmentations at the same time, one can represent the total molecule *M* as;

$$M \rightarrow G_1 G_2 G_3 ... G_{n-1} H^{(n-1)} + H^{(n)} G_n G_{n+1} ... G_{i-1} H^{(i-1)} + H^{(i)} G_i G_{i+1} ... G_K \tag{80}$$

This double fragmentation creates a new energy expression:

$$E(M) = E(M_1) + E(M_5) + E(M_4) + dE_3 \tag{81}$$

www.manaraa.com

The new fragment M_5 is the result of fragment M_2 being broken into two pieces by the second fragmentation. Specifically,

$$M_5 = H^{(n)} G_n G_{n+1} \ldots G_{i-1} H^{(i-1)} \qquad (82)$$

The new term dE_3 is the result of an approximation; namely, if the $G_{n-1}G_n$ bond is separated from the $G_{i-1}G_i$ bond by a great enough distance, the energy difference from simultaneous fragmentation will equal the sum of the energy changes of each fragmentation performed separately;

$$dE_3 \approx dE_1 + dE_2 \qquad (83)$$

As the distance between the two fragmentation sites increases, the approximation in Eq. (83) becomes more and more reliable. Rearranging Eqs. (77) and (81), while using the equality in Eq. (83), it can be shown that;

$$\begin{aligned} dE_3 &= E(M) - E(M_1) - E(M_5) - E(M_4) \\ &\approx E(M) - E(M_1) - E(M_2) + E(M) - E(M_3) - E(M_4) \end{aligned} \qquad (84)$$

or more simply;

$$E(M) \approx E(M_2) + E(M_3) - E(M_5) \qquad (85)$$

Using the molecular definitions, it becomes apparent that M_5 is simply the overlapping or "double counted" region common to both M_2 and M_3.

Within the foregoing formulation, there are different "levels" of fragmentation. Fragmentation level 1 consists of fragmentation sites separated by one functional group, level 2 has fragmentation sites separated by two functional groups and so on. The current implementation of the SMF method allows for up to fragmentation level 3, with the obvious possibility for extension to higher levels. As the fragmentation sites become farther separated, the approximation in Eq. (83) becomes more reliable and the total energy of the system approaches that of the full *ab initio* calculation.

To illustrate the exhaustive fragmentation of a system using the different fragmentation levels, consider the acyclic molecule *M* again:

$$M = G_1 G_2 G_3 G_4 G_5 G_6 G_7 G_8 \qquad (86)$$

In this case, the fixed length K=8. Under the level 1 fragmentation scheme the two broken bonds are separated by only one functional group. The first fragmentation site is chosen between G_1 and G_2 and the second site between G_2 and G_3. The first fragmentation creates

www.manaraa.com

$$M = G_1 + G_2G_3G_4G_5G_6G_7G_8 \qquad (87)$$

while the second creates;

$$M = G_1G_2 + G_3G_4G_5G_6G_7G_8 \qquad (88)$$

Subtracting group G_2 to avoid double counting, we get the following representation of molecule M:

$$M = G_1G_2 + G_2G_3G_4G_5G_6G_7G_8 - G_2 \qquad (89)$$

Following this scheme through to all possible fragmentation sites allowed by level 1, until no fragment larger than two functional groups remains, the bonded energy of molecule M can be represented by the sum of fragment energies as follows:

$$\begin{aligned} E^{\text{bonded}}_{\text{level1}}(M) &= E(G_1G_2) + E(G_2G_3) + E(G_3G_4) + \\ &E(G_4G_5) + E(G_5G_6) + E(G_6G_7) + E(G_7G_8) - E(G_2) - \\ &E(G_3) - E(G_4) - E(G_5) - E(G_6) - E(G_7) \end{aligned} \qquad (90)$$

Using the same methodology for level 2, with fragmentation sites instead separated by two functional groups, the bonded energy of M is represented by:

$$\begin{aligned} E^{\text{bonded}}_{\text{level2}}(M) &= E(G_1G_2G_3) + E(G_2G_3G_4) + E(G_3G_4G_5) + \\ &E(G_4G_5G_6) + E(G_5G_6G_7) + E(G_6G_7G_8) - E(G_2G_3) - \\ &E(G_3G_4) - E(G_4G_5) - E(G_5G_6) - E(G_6G_7) \end{aligned} \qquad (91)$$

and using level 3, with fragmentation sites separated by three functional groups, gives a representation of the bonded energy of M:

$$\begin{aligned} E^{\text{bonded}}_{\text{level3}}(M) &= E(G_1G_2G_3G_4) + E(G_2G_3G_4G_5) + \\ &E(G_3G_4G_5G_6) + E(G_4G_5G_6G_7) + E(G_5G_6G_7G_8) - \\ &E(G_2G_3G_4) - E(G_3G_4G_5) - E(G_4G_5G_6) - E(G_5G_6G_7) \end{aligned} \qquad (92)$$

Carrying this fragmentation scheme out to level n, where n is the number of groups in the system, one would be left with the unfragmented system. By using higher levels of SMF, and consequently larger fragments, the total energy will approach that of the exact system.

So far, only the bonded energy of the example system has been discussed, leaving out a great deal of important contributions to the total energy contained in the non-bonded interactions of separated functional groups. Recently, the calculation of these interactions has moved from a simple electrostatic model in the

www.manaraa.com

original formulation, to a more sophisticated approach combining *ab initio* and EFP method calculations.[374] The choice of performing either full *ab initio* calculations or EFP calculations is based on the shortest atom-atom distance between the interacting fragments. At short ranges (<2.7Å) the non-bonded interactions are calculated with the full *ab initio* method being employed. For intermediate distances (2.7-4.5Å) the EFP method is used, effectively reducing the number of *ab initio* calculations required and increasing the computational efficiency.

The simplest case of non-bonded interactions occurs between just two separated functional groups, for example G_1 and G_4. The non-bonded energy contained in the "super-group" G_1G_4 can be given by:

$$E_{\mathrm{nb}}^{(1,1)}[G_1;G_4] = E(G_1G_4) - E(G_1) - E(G_4) \qquad (93)$$

$E(G_1G_4)$ is the supermolecular energy of the two separated functional groups and $E(G_1)$ and $E(G_4)$ are the one-body fragment energies. During the calculation of the supermolecular energy $E(G_1G_4)$, the two functional groups retain their original positions from the complete system *M*. This procedure can be carried out to its eventual conclusion of all possible pairs of separated functional groups, providing the total two-body non-bonded interaction energy of the entire system *M*.

In some instances, three-body interactions can play a very important role in chemical systems.[376,377,378,379,380] To account for these interactions a similar approach can be taken by considering the interaction of three functional groups, G_1, G_2 and G_3. The interactions of these three groups is ignored unless two of the three are bonded in the full system *M*. To give a specific example, consider the case in which G_2 is directly bonded to G_3, giving the three body interaction energy as;

$$E_{\mathrm{nb}}^{(1,2)}[G_1;G_2G_3] = E(G_1G_2G_3) - E(G_1) - E(G_2G_3) - E_{\mathrm{nb}}^{(1,1)}[G_1;G_2] - E_{\mathrm{nb}}^{(1,1)}[G_1;G_3] \qquad (94)$$

The three body energy is simply the supermolecular energy $E(G_1G_2G_3)$ minus the one-body energy $E(G_1)$, the bonded energy of $E(G_2G_3)$ and the two-body non-bonded energies $E_{nb}^{(1,1)}[G_1;G_2]$ and $E_{nb}^{(1,1)}[G_1;G_3]$.

Using all three of these expressions for the bonded and non-bonded energies, the total energy of the entire system, E_{SMF}, can be expressed as a sum of bonded and non-bonded energies:

$$E_{\mathrm{SFM}} = E^{\mathrm{bonded}} + E^{\mathrm{non-bonded}} \qquad (95)$$

The term $E^{\text{non-bonded}}$ includes all of the terms up to *n*th order. In practice it has been shown that the inclusion of non-bonded terms in the level 3 fragmentation scheme provides the best combination of accuracy and

www.manaraa.com

computational efficiency. The choice of including three-body non-bonded interactions depends on the system of interest, with molecular clusters such as water necessitating the inclusion of three-body interactions.

The SMF method has some limitations, the most important of which is the inability to break bonds more complex than single bonds. There is also a limitation in the fragmentation of cyclic molecules such as hexane. During fragmentation the capping hydrogens may become too close, causing a steric interaction not present in the complete system and violating the approximation that fragmentation sites are energetically independent. This problem can be alleviated through the use of the so-called ring repair rule[120] to effectively avoid such non-physical interactions. Active development of the SMF method currently includes a reformulation of the polarization interactions to give a more accurate description of highly polar molecular clusters.

2.7 Divide-and-Conquer Methods

2.7.1 Original divide and conquer approach

The original formulation of the divided and conquer (DC) approach was proposed by Yang[18] in 1991. Based on the Kohn-Sham (KS) formalism, the DC approach aims to avoid the use of the *N*/2 KS orbitals and instead divide the density of the system of interest into the sum of the densities of the subsystems. The DC formalism was generalized to a more efficient one-electron density matrix approach that is generally applicable to *ab initio*[381,382,383,384] and semiempirical methods.[385,386] As with other fragmentation methods, the local environment of each subsystem is taken into account, in this case through the use of buffer regions surrounding each fragment. Since the method was initially proposed, it has been expanded electron correlation methods as well. [387] Nakai's group [388,389,390,391,392,393,394,395,396,397] and others[398,399] made many contributions to further development. The main contributors to these enhancements of the DC approach will be discussed, beginning with both the original and improved density matrix formulations proposed by Yang.

In the KS formalism, the total energy of a system of *N* electrons in an external field $\upsilon(\mathbf{r})$ can be represented using the electron density $\rho(\mathbf{r})$ as:

$$E[\rho] = T_s[\rho] + \int \upsilon(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + E_{xc}[\rho] + \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + \sum_{a,b}\frac{Z_a Z_b}{R_{ab}} \qquad (96)$$

$T_s[\rho]$ represents the kinetic energy of the noninteracting electron gas with density ρ in the ground state. The term $E_{xc}[\rho]$ is the exchange-correlation energy. The final term in Eq. (96) is the nuclear repulsion. In the conventional KS formalism, the energy functional $E[\rho]$ is minimized with respect to the density by satisfying the KS equation:

$$\hat{H}\psi_i(\mathbf{r}) = \left[-\tfrac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \qquad (97)$$

www.manaraa.com

$V_{\text{eff}}(\mathbf{r})$ is the KS effective local potential and $\hat{H}$ is the KS Hamiltonian. As in the KS formalism, the DC approach uses the electron density as the main variable. However, the DC approach represents the total electron density of the full system as a sum of subsystem contributions. This is accomplished through the use of normalized "partition functions",

$$\sum_{\alpha} p^{\alpha}(\mathbf{r}) = 1 \qquad (98)$$

$p^{\alpha}(\mathbf{r})$ is a positive weighting function for subsystem α. More specifically, $p^{\alpha}(\mathbf{r})$ is large for the subspace of subsystem α and small otherwise. This leads to an expression for the total electron density of the system:

$$\rho(\mathbf{r}) = \sum_{\alpha} p^{\alpha}(\mathbf{r})\rho(\mathbf{r}) = \sum_{\alpha} \rho^{\alpha}(\mathbf{r}) \qquad (99)$$

The partition functions have been determined previously,[18] and apparently the density and energy do not depend significantly upon the specific form of the partition function. The density of an individual subsystem can now be defined using the Fermi function[400] $f_{\beta}(x)$ as:

$$\rho^{\alpha}(\mathbf{r}) = 2\,p^{\alpha}(\mathbf{r}) \sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha}) \left|\psi_m^{\alpha}(\mathbf{r})\right|^2 \qquad (100)$$

where

$$f_{\beta}(x) = \left[1 + \exp(-\beta x)\right]^{-1} \qquad (101)$$

ψ_m^{α} represents an eigenfunction that is localized on the particular subsystem α, and ε_m^{α} is the corresponding eigenvalue of the subsystem α. The eigenfunctions ψ_m^{α} are linear combinations of a set of local basis functions.

$$\psi_m^{\alpha}(\mathbf{r}) = \sum_{j} C_{jm}^{\alpha} \phi_j^{\alpha}(\mathbf{r}) \qquad (102)$$

It is the use of localized basis functions for a particular subsystem that allows the DC method to scale nearly linearly with system size. The coefficients C_{jm}^{α} are solutions to the generalized eigenvalue equation, derived from the Rayleigh-Ritz variational principle:

$$(\mathbf{H}^{\alpha} - \varepsilon_m^{\alpha}\mathbf{S}^{\alpha})\mathbf{C}_m^{\alpha} = 0 \qquad (103)$$

www.manaraa.com

The Fermi energy[400] ε_F from Eq. (98) is determined from

$$N = \int \rho(\mathbf{r}) d\mathbf{r} = 2\sum_{\alpha}\sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha})\left\langle \psi_m^{\alpha} \middle| \rho^{\alpha}(\mathbf{r}) \middle| \psi_m^{\alpha} \right\rangle \qquad (104)$$

N is simply the normalization condition for the electron density. After solving the preceding equations self consistently, the energy expression from Eq. (96) can be expressed in terms of the eigenvalues,

$$\tilde{E}[\rho] = \tilde{\varepsilon} + \int \rho\left[-\varphi(\mathbf{r})/2 - V_{\mathrm{xc}}(\mathbf{r})\right] d\mathbf{r} + E_{xc}[\rho] + \sum_{a,b} \frac{Z_a Z_b}{R_{ab}} \qquad (105)$$

$\tilde{\varepsilon}$ is an approximation to the KS eigenfunctions, represented by:

$$\tilde{\varepsilon} = 2\sum_{\alpha}\sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha})\left\langle \psi_m^{\alpha} \middle| \rho^{\alpha}(\mathbf{r}) \middle| \psi_m^{\alpha} \right\rangle \qquad (106)$$

In an effort to improve upon the original formulation, Yang proposed a reformulation of the DC formalism based on the one-electron density matrix[381]. Defined in terms of the KS orbitals, the one-electron density matrix can be expressed as:

$$\rho(\mathbf{r}, \mathbf{r}') = 2\sum_{m}^{N/2} \psi_m(\mathbf{r})\psi_m(\mathbf{r}') = \sum_{ij} \rho_{ij}\varphi_j(\mathbf{r})\varphi_j(\mathbf{r}') \qquad (107)$$

The density matrix ρ_{ij} is given by the linear coefficients in the expansion of the KS orbitals:

$$\rho_{ij} = 2\sum_{m}^{N/2} C_{im} C_{jm} \qquad (108)$$

The partition matrix for each subsystem can now be defined in the atomic orbital space, with a normalization condition that is similar to the original formulation in Eq. (98),

$$\sum_{\alpha} \mathbf{p}_{ij}^{\alpha} = 1 \qquad (109)$$

and constructed using the following rules:

$$\mathbf{p}_{ij}^{\alpha} = \begin{cases} 1 & \text{if } i \in \alpha \text{ and } j \in \alpha \\ \tfrac{1}{2} & \text{if } i \in \alpha \text{ and } j \notin \alpha \\ 0 & \text{if } i \notin \alpha \text{ and } j \notin \alpha \end{cases} \qquad (110)$$

www.manaraa.com

Subsystem contributions to the density matrix can now be written as:

$$\rho_{ij} = \sum_{\alpha} \mathbf{p}_{ij}^{\alpha} \rho_{ij} = \sum_{\alpha} \rho_{ij}^{\alpha} \qquad (111)$$

The expression in Eq. (111) is equivalent to the expression from the original formulation in Eq. (99). A corresponding approximation to the original formulation, using a set of local eigenvectors to approximate the density matrix of a subsystem, can be applied to give:

$$\rho_{ij}^{\alpha} = 2\mathbf{p}_{ij}^{\alpha} \sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha}) C_{im}^{\alpha} C_{jm}^{\alpha} \qquad (112)$$

An analogous expression for the Fermi energy, determined by the normalization, can then be expressed as:

$$N = \sum_{ij} \rho_{ij} \mathbf{S}_{ij} = \sum_{ij} \left(2 \sum_{\alpha} \mathbf{p}_{ij}^{\alpha} \sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha}) C_{im}^{\alpha} C_{jm}^{\alpha} \right) \mathbf{S}_{ij} \qquad (113)$$

Finally, the analogous expression for the sum of the eigenvalues is written as:

$$\begin{aligned} \varepsilon &= 2 \sum \sum f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha}) \varepsilon_m^{\alpha} \sum \mathbf{p}_{ij}^{\alpha} C_{im}^{\alpha} C_{jm}^{\alpha} \mathbf{S}_{ij} \\ &= \sum_{ij} \left(2 \sum_{\alpha} \mathbf{P}_{ij}^{\alpha} \sum_{m} f_{\beta}(\varepsilon_F - \varepsilon_m^{\alpha}) C_{im}^{\alpha} C_{jm}^{\alpha} \right) \mathbf{H}_{ij} \end{aligned} \qquad (114)$$

The main advantages of the new DC formulation include the removal of the time consuming partition function integrals, as well as the general applicability the approach now has to *ab initio* methods such as Hartree Fock. Analogous expressions for the gradients have been derived from the original formulation as well.

The implementation of the DC method described divides a molecular system into various subsystems, each described by a set of local orbitals. To aid in the description of each subsystem, the basis functions of neighboring subsystems, or "buffer" regions, are included in each subsystem calculation. The inclusion of buffer regions is determined by a distance based cut-off, R_b. Any atom that falls inside the sphere created by R_b is included as a buffer atom in the subsystem calculation. By including the atomic basis functions from both the subsystem and the buffer region, computational requirements scale by a factor of N_{α}^3, where N_{α} is the number of basis functions contained in subsystem α and the corresponding buffer region. Yang determined that the buffer region required to achieve a certain level of accuracy remains constant regardless of the system size.[382] Linear scaling is then possible by fixing the size of the buffer region for each subsystem.

www.manaraa.com

In addition to linear scaling in terms of the computational time required, Yang also proposed a reduction in memory requirements by storage of a sparse matrix for the total system.[382] This is accomplished using a distance based cut-off for matrix elements between atom pairs with an interatomic distance less than R_h. By applying this cut-off the storage requirements of the density matrix become proportional to the size of the molecule. The amount of CPU time required also decreases due to the reduction in matrix element evaluations. In addition to the density matrix, both the one-electron core Hamiltonian and Fock matrices are treated in this way.

Following the two formulations of the DC method in 1995 by Yang, a number of other groups proposed extensions to the method to enhance the functionality. The first to propose such an extension was Merz in 1996 with his semiempirical MO implementation.[385,386] A notable improvement pointed out by Merz was the need for overlap between adjacent subsystems. To illustrate why this is necessary, consider the closed shell Fock matrix:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda=1}^{M}\sum_{\sigma=1}^{M}\left[(\mu\nu\,|\,\lambda\sigma) - \frac{1}{2}(\mu\sigma\,|\,\lambda\nu)\right]P_{\lambda\sigma} \qquad (115)$$

where the two electron integrals are represented as:

$$(\mu\nu\,|\,\lambda\sigma) = \int\int \chi_\mu^*(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)\times\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\chi_\lambda^*(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \qquad (116)$$

The density matrix for a particular subsystem, $\mathbf{P}^\alpha$, depends on how the subsystem is defined. Buffer regions are included to reduce the truncation effects of fragmentation, however their basis functions are only included during the solution of the SCF equations. This leads to their exclusion during construction of the density matrix, since only basis functions that are part of the subsystem directly, and not contributed by a buffer region, are included. Since the density matrix elements contributed exclusively from buffer region basis functions are zero, overlap between subsystems is required to provide these contributions to the density matrix.

A number of improvements to the DC formulation of Yang were implemented by the Nakai group, including the ability to perform MP2 calculations,[388,391] application to delocalized systems,[389] the addition of Hartree-Fock exchange,[390] CC calculations[392] and a two level hierarchical scheme.[395] The problem of overlapping subsystems pointed out by Merz was also overcome in the Nakai implementation by treating each atom as a subsystem. Detailed test calculations were performed and showed that the use of a sufficiently large buffer region is adequate to overcome such a small subsystem division. Even in the case of delocalized systems, subsystems consisting of a single atom were capable of producing reasonably accurate results when a large buffer region (greater than 10 Å) was considered.[389] It was noted, however, that the accuracy of any DC calculation can obviously be improved through the use of larger subsystems if desired.

www.manaraa.com

Correlated calculations were improved through the use of a "dual buffer" scheme.[391] The first buffer region surrounding each subsystem includes the calculation of the correlation energy, while the second layer buffer only performs a Hartree-Fock calculation. This scheme effectively exploits the local nature of electron correlation and provides additional reductions in the computational effort for correlated calculations.

A dual-level hierarchical scheme[395] is built upon the foundation of a dual buffer scheme. Following the same procedure for calculating the correlation energy as in the dual buffer scheme, a second level of approximation is introduced by using a small and large basis set for HF and correlated calculations respectively. The use of a larger basis set for the correlated buffer calculations provides more accurate energies, while the use of a smaller basis set for the HF buffer regions reduces the computational cost for each subsystem calculation.

2.7.2 The adjustable density matrix assembler approach (ADMA)

Developed contemporarily with the divide and conquer method, the adjustable density matrix assembler approach divides the density of a molecular system using a density matrix approach. Originating from a similar approach called the molecular electron density lego approach[401,402] (MEDLA), and inspired by earlier work by Michl[403] and Stoddart[404], the ADMA[405,406,407,408,409,410] method uses the idea of a fuzzy-set to remove any discontinuities between fragments, circumventing the need for explicit capping or other such bond conserving approaches. Using a density matrix formalism allows the ADMA approach to be applied to many modern *ab initio* methods. The most recent formulation of the ADMA approach will be discussed, as well as recent improvements such as the inclusion of point charges during individual fragment calculations.

The ADMA approach expresses the electron density of a molecular system in terms of a sum of contributions from individual fragments. Consider the exact total density of a molecule at a given nuclear configuration:

$$\rho(\mathbf{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) \qquad (117)$$

where $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are atomic orbitals (AOs). To rewrite this expression in terms of a summation of smaller contributions, the molecule of interest must be divided into a set of mutually exclusive "families" f^k where $k=1,\ldots,m$ and m is the total number of families. Division of AOs among families is accomplished through the use of a membership function $m^k(i)$ such that:

$$m^k(i) = \begin{cases} 1 & \text{if AO } \phi(\mathbf{r}) \text{ is centered on one nuclei of set } f^k \\ 0 & \text{otherwise} \end{cases} \qquad (118)$$

Each fragment k then has a density matrix defined by:

www.manaraa.com

$$P_{ij}^{k} = \begin{cases} P_{ij} & \text{if both } \varphi_i \text{ and } \varphi_j \text{ are centered on a nucleus of } f_k \\ \tfrac{1}{2} & \text{if one of } \varphi_i \text{ and } \varphi_j \text{ are centered on a nucleus of } f_k \\ 0 & \text{otherwise} \end{cases} \quad (119)$$

This scheme allows for an exactly additive decomposition of the total density matrix. To illustrate this, consider a molecular system *M* containing nine families:

$$M = f_1 f_2 f_3 .... f_9 \quad (120)$$

The interactions of these nine families can be artificially represented in matrix form:

$$\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{matrix} \quad (121)$$

Any nearest neighbor families are taken to have a strong interaction, for example, family 4 has a strong interaction with family 5, but not family 6. Building the total interactions from family 4, a reduced form of the interaction matrix can be written as:

$$\begin{matrix} 1 & 2 \\ \mathbf{4} & 5 \\ 7 & 8 \end{matrix} \quad (122)$$

which includes all families with a strong interaction with the central family 4. Of the other families considered in this reduced system, families 2, 5 and 8 contain dangling bonds that were previously connected to families 3, 6 and 9 respectively. These bonds are capped with hydrogen atoms, however, the size of each family (in this case the family represented by Eq (122)) is chosen to be large enough that the corresponding orbitals from the capping hydrogen have negligible contributions to the orbitals on the central family 4. This allows the capping hydrogen orbitals to be excluded from construction of the density matrix. The contribution of the reduced system in Eq (122) to the total density matrix can then be written as:

www.manaraa.com

$$
\begin{array}{cccccccccc}
1 & & & * & & & & & & \\
 & 2 & & * & & & & & & \\
 & & 3 & 0 & & & & & & \\
* & * & 0 & 4 & * & 0 & * & * & 0 & \\
 & & & * & 5 & & & & & \\
 & & & 0 & & 6 & & & & \\
 & & & * & & & 7 & & & \\
 & & & * & & & & 8 & & \\
 & & & 0 & & & & & 9 &
\end{array}
\qquad (123)
$$

Each asterisk represents a contribution to the total density matrix created by interactions contained in the reduced subsystem of Eq (122). Every interaction is scaled by 0.5 and all other off diagonal contributions are zero. This procedure can then be repeated, assigning each family as the central family and building the corresponding reduced systems. Following this procedure through to completion, a matrix of fragment contributions to the total density matrix is obtained:

$$
\begin{array}{ccccccccc}
1 & * & 0 & * & * & 0 & 0 & 0 & 0 \\
* & 2 & * & * & * & * & 0 & 0 & 0 \\
0 & * & 3 & 0 & * & * & 0 & 0 & 0 \\
* & * & 0 & 4 & * & 0 & * & * & 0 \\
* & * & * & * & 5 & * & * & * & * \\
0 & * & * & 0 & * & 6 & 0 & * & * \\
0 & 0 & 0 & * & * & 0 & 7 & * & 0 \\
0 & 0 & 0 & * & * & * & * & 8 & * \\
0 & 0 & 0 & 0 & * & * & 0 & * & 9
\end{array}
\qquad (124)
$$

Each asterisk now represents a sum of two scaled contributions. This relatively sparse fragment interaction matrix is then used to construct a good approximation to the total density matrix. The example shown is overly simplified; however, it provides a reasonable illustration of how the ADMA method can reduce the computational cost required to compute the total density matrix.

In general the fragment choice in the ADMA approach is unimportant, however in practice chemically reasonable fragments are chosen. For example, when fragmenting a protein, a set of chemically sensible rules are followed: (i) all heavy atoms double bonded to an oxygen atom are considered to be fragments; (ii) if a heavy atom is part of an aromatic moiety, the entire aromatic region defines a fragment; (iii) specific chemically important substituents such as the $CONH_2$ groups of asparagine and gluatamine are defined as a fragment. After the initial fragment choice, in a manner similar to the MTA method, a sphere of a user defined radius is used to include neighboring families. If a single atom belonging to another nuclear family falls within the radius around the fragment of interest, then all atoms in that family are included. This includes any atoms that

www.manaraa.com

are not directly bonded to the fragment of interest. All dangling bonds are considered at this point and capped using hydrogen atoms.

Once the fragment of interest and the included surroundings are defined, the density matrix contribution can be calculated for all defined fragments. The accuracy of the ADMA approach may be improved either by choosing larger fragments or by increasing the radius of inclusion for neighboring nuclear families. The composite fragment density matrix can then be used to calculate a number or properties including the total energy, dipole moment, electrostatic potential and total density matrix for the molecular system.

The accuracy of the ADMA approach depends on how well the local environment, determined by the radius of inclusion, is described during each fragment calculation. The most obvious way to improve the description of the environment is to increase the radius of inclusion of neighboring atoms. It was shown that a radius of up to 10 Å is needed to obtain sufficiently accurate results (errors of less than 1 kcal/mol).[411] Unfortunately, this decreases the efficiency of the approach, particularly for large molecules. The solution was implemented by Mezey and Exner in 2006, termed the field-adapted ADMA (FA-ADMA)[412] approach. By using a smaller radius of inclusion coupled with a more approximate description of far separated atoms, the FA-ADMA approach effectively increases the accuracy of the original ADMA approach without any significant increase in computational cost.

The most effective approximation to the surroundings of fragments in the FA-ADMA was determined to be partial charges based on Mulliken or Löwdin charges. The computational scheme employed begins with a standard ADMA calculation which is used to obtain the partial charges of all the atoms present in the fragment calculation. During subsequent fragment calculations, these partial charges are included, gradually increasing the number or partial charges until the full system is represented. The entire process is then started anew, using all of the partial charges obtained during the previous step, and continued in an iterative process until the charges and densities for all fragments do not change up to a certain threshold. Atoms present in the point charge description of the environment that have been replaced by capping hydrogens in the fragment (so-called junction atoms) are represented by a 60% scaling of the total Mulliken charge or an 80% scaling of the total Löwdin charge, aiding in the reduction of error incurred by over-counting these atoms during the ab initio portion of the FA-ADMA calculations.

Additional improvements to the ADMA approach have been implemented in recent years. These include the use of an alternative, more general fragmentation scheme[413], the use of the ADMA density matrix as an initial guess for ab initio SCF calculations on large molecules[414] and an improved description of the junction atoms at the edges of the surroundings encompassed by the radius of inclusion.[415]

2.8 Other Methods

www.manaraa.com

Das et al.[416] in 2003 proposed the ab initio fragment orbital-based theory (AFOT), in which the total wave function is constructed from orbitals of fragments, which are formed and computed without capping.

The integrated multi-center molecular orbital method (IMiCMO) formulated by Sakai and Morita [150,417] for molecular clusters relies on the addition of the properties of target molecules computed in the buffer region of adjacent molecules, while the effect of the rest of the system is computed with point charge interactions. In addition to the total energies, forces and second derivatives have also been developed.

Mata et al.[418] in 2009 suggested a model to estimate the excitation energies of large systems by diagonalizing the total Hamiltonian H_{ij}, $i,j=1,\ldots,N$ where N is the number of fragments. In this Hamiltonian only the diagonal elements H_{ii} are computed quantum-mechanically (for each fragment, electrostatically embedded in the field of others) while off-diagonal elements are estimated semi-classically with the dipole-dipole interaction model. This model takes into account the excitonic coupling in molecular clusters and it has been used to compute the first electronic absorption band of water.

Following the same methodology as the SMF method, Bettens et al., [419,420,421] have developed an energy based fragmentation method based on the idea of isodesmic reactions**.** [422] Molecules are divided into groups in a similar fashion to the SMF method and capped with hydrogens. The main difference between the two methods is how the groups are recombined into fragments. In many cases the two fragmentation schemes produce the exact same fragments. However, for higher levels of fragmentation such as SMF level 3, the Bettens method can produce a different fragmentation scheme. The fragments produced in these cases are typically smaller in size than those produced by the SMF method, but many more fragments are formed. In a specialized case provided by Bettens[421], fragments formed are one group size smaller than the SMF method fragments. In this instance SMF produces three fragments, while the Bettens method produces eleven fragments.

In the linear-scaling three-dimensional fragment method for large-scale electronic structure calculations (LS3DF) developed by Wang et al.[423] in 2008, the energy is constructed as a sum of additive contributions of small conglomerates of fragments in the self-consistent Coulomb potential determined at each iteration by solving the Poisson equation for the total density; the conglomerates are used to reduce the error caused by the caps added to dangling bonds. This method has been applied to CdSe quantum dots that contain up to 2616 atoms.

3. Software and Parallel Computing

Developing efficient software is a very important aspect of computational and theoretical chemistry. Many research groups concentrate on method development using some “in-house” programs or locally modified versions of commonly used software. Although this often results in important advances in method development, such results are often not reproducible by other scientists without reproducing the code, an arduous process. Moreover, there is a growing need by researchers in adhacent fields, such as biochemistry material science and

www.manaraa.com

engineering, computational biology and others to use ab initio methods for practical calculations. It is clear that there is a demand for easy-to-use software that can be run by users on their own. This need is at present is only partially filled by software development of fragment-based methods.

There are two types of software: programs to perform calculations, and software to aid in input file preparation and to analyze and visualize results. For the first group, the most fragment-friendly program at present is the general atomic and molecular electronic structure system (GAMESS),[424,425] which has EFP, FMO, ELG and DC in the production version. Q-Chem supports EFP, while ABINIT-MP[426,427] and PAICS[428] are FMO programs. PEACH[426] is an MD program used in conjunction with ABINIT-MP to run FMO-MD simulations. MTA is based on locally modified versions of GAMESS, and it has a Web interfaces.[429,430] Some recent X-Pol developments use a locally modified version of GAMESS.

GAMESS has MacMolPlt[431] as its graphical interface, although it has limited capability specifically for fragment methods. Facio[432] has a very elaborate Windows GAMESS/FMO interface for input file preparation and result visualization; it has an automatic fragmentation engine for dividing polypeptides (including proteins), nucleotides (including DNA), saccharides or any combination thereof into standard fragments. Biostation[427] is a graphical interface developed in conjunction with ABINIT-MP for FMO calculations.

One of the important advantages of fragmentation methods is that they are frequently inherently scalable to many compute nodes and cores. The FMO method in GAMESS presents a nice example of this, since it has been implemented with the associated distributed data interface (DDI)[433,434] and its generalized, multi-level partner GDDI.1 DDI facilitates the distribution of large arrays across many compute nodes, while GDDI makes it possible to use both coarse graining and fine graining. The coarse-grained parallelism allows the calculation (for example) of the contribution from each fragment to the total energy on a different node, while the fine-grained parallelism occurs among the cores within each node. This approach has enabled FMO calculations to take advantage of virtually perfect scaling on tens of thousands of cores.[435] While the example presented here is for FMO in GAMESS, most of the fragmentation methods discussed here have the potential for similar scaling.

4. Applications

The applications below are described in subsections according to the application field.

4.1 Homogeneous clusters and explicit solvent treatments

Fragment methods are a powerful alternative to force fields in running molecular dynamics simulations (in fact, some ab initio based methods are thought of as new generation force fields[46,73,312]). Of course, MD simulations require very accurate gradients, and in some cases the gradient used in the initial MD simulations had questionable accuracy.[264] Also, during QM MD simulations, there is no enforced bond definition, and bonds can be broken along the trajectory. If this happens, the pieces of the broken fragment can move away from each

www.manaraa.com

other and attach themselves to other fragments. If this happens, the fragment definition becomes inappropriate. The dynamic fragmentation[267] method is designed to address this issue: When a fragment breaks into pieces, the fragment definition is changed on the fly; this can, however, result in a discontinuity on the potential energy surface. Some fragment methods like EFP have frozen fragment geometries, and do not suffer from this problem. On the other hand, rigid fragments cannot adjust their structure to the environment, which limits their accuracy.

Many methods assume that fragments do not interact very strongly. But a small inter-fragment distance can cause large charge transfer and other strong interaction. In the methods that use multipole models to describe the electrostatics, the short-range multipole interaction can break down due to the neglect of charge penetration; multipole screening [436,237] can be used as a remedy.

Nevertheless, fragment based methods can take into account many-body polarization and charge transfer, and thus are new generation force fields. In charged solutes these effects of the polarization and charge transfer between the solute and solvent are large and difficult to neglect.

One important application of fragment methods is the investigation of chemical reactions of small molecules in solution with MD, with solvent molecules represented as fragments, on a par with the solute. Sato et al.[437] applied the FMO method to study a S_N2 reaction of the hydrolysis of the methyl diazonium ion, and found two types of reaction pathway. Solute-solvent charge transfer plays a very important role in this reaction and the real trajectories are quite different from the textbook images of the S_N2 reaction. Sato et al.[438] used the FMO method to study the amination of formaldehyde in water and found that this reaction proceeds via a stepwise mechanism through a zwitterionic intermediate, not by a concerted mechanism.

Pomogaev et al. [439] employed ELG to compute the absorption spectra for explicitly solvated estradiol and tryptophan in the Trp-cage protein, using snapshots from classical MD simulations. Kistler and Matsika[440] applied the multi-layer FMO-MCSCF approach to study solvatochromic shifts of uracil and cytosine, using snapshots of classical MD simulations. They showed by comparison of various methods that explicit solvent methods including the use of QM or MM to describe the solvent produce similar results, whereas the continuum-based PCM for some excited states gives significantly different and parameter dependent results.

Fujiwara et al.[441] applied the FMO method to MD simulations of Zn^{2+} solvated by a droplet with 64 water molecules and reproduced the experimental value of the first peak in the radial distribution function to within 0.01 Å.

Several EFP1 and QM/EFP1 studies have been devoted to the investigation of structures and energetics of small to medium water clusters[185,187,442]. Structural properties of bulk water[188] and origins of enhancement of the dipole moment in bulk water [443] were also investigated. Another important area of application of QM/EFP1

www.manaraa.com

methodology is hydration of various ions, including halogen[444] and molecular anions[445 446], as well as metal[447 448] and molecular[449] cations.

Hydration and dissociation of ionic species were investigated by Peterson and Gordon[450] for NaCl and by Yoshikawa and Morales [451] for LiOH. Solvation effects on the electronic properties of a solute, such as electronegativity, hardness, HOMO-LUMO gaps, were investigated for ammonia[452] and a set of polyatomic molecules and ions.[453]

Solvent effects on neutral-zwitterionic equilibrium of aminoacids were also studied with QM/EFP1. Day et al. [454] investigated small (up to ten waters) clusters of glutamic acid and water. Solvation of glycine was studied by Bandyopadhyay and Gordon[191]. Solvation of alanine by increasing number of water molecules was studied by Mullin and Gordon.[455,456] Song et al. investigated solvent effects on the conformational potential energy surfaces (gauche versus trans conformations) of acetylcholine and acetylthiocholine ACh and ATCh.[457]

The EFP2 model has been used to investigate intermolecular interactions in complex molecular clusters. Adamovic and Gordon showed that water-methanol mixtures are heterogeneic at the microscopic level [458]. The same authors investigated interactions in styrene clusters[459], where both H-bonding and π-bonding structures take place. π-stacking interactions were investigated in benzene dimer[460], substituted benzene dimers[461], and benzene-pyridine dimers[462]; an intriguing competition between H-bonding and π-π bonding was observed in water-benzene clusters[463]. Interactions in π-stacks of DNA base-pairs were studied by Ghosh et al[196]. Shortening of the B-N bond in H_3BNH_3 on going from the gaseous to the solid state were explained using the EFP2 method.[464]

The QM/EFP method has also been used to investigate the electronic excited states of chromophores in solution. The spectroscopy of enzyme active sites in the presence of several EFP1 waters was studied by Krauss[465,466]. Aqueous olvation of the lowest singlet excited state of Coumarin 151 was investigated with the CIS/EFP1 method.[171] The absorption spectrum of acetone in water was modeled with TD-DFT/EFP1 by Yoo et al.[172] Ab initio molecular dynamics with CASSCF/EFP1 was used to investigate the dynamics of solvated excited states in coumarin 151 [467] and excited state hydrogen transfer in 7-azaindole.[468] The EFP2 model combined with CIS(D) was used to investigate solvatochromic shifts of singlet and triplet electronic states of para-nitroaniline in water, dioxane, and cyclohexane by Kosenkov and Slipchenko.[174] A combination of the EOM-CCSD for ionization potentials (IP) method with EFP allowed accurate description of vertical ionization energy (VIE) of thymine in bulk water.[176] Intrestingly, the first hydration shell increases VIE by ~ 0.1 eV, while subsequent solvation lowers ionization energy and the bulk value of the solvent-induced shift of thymine's VIE is approximately -0.9 eV.

Influence of water environment on energetics and pathways of various reactions were investigated with QM/EFP1 methods. The energetics of the Menshutkin reaction between ammonia and methyl bromide in a

www.manaraa.com

presence of increasing number of water molecules were studied by HF/EFP1[189]. Adamovic and Gordon showed that MP2/EFP1 provides accurate description of hydration effects on structures and barriers in S_N2 reaction between Cl^- and CH_3Br.[190] The EFP solvent model was also used to study the kinetics of the hydrogen abstraction from H_2O_2 by $OHO\cdot$. [469] The reaction mechanism of phosphate monoester aminolysis in aqueous solution was investigated with DFT/EFP1 by Ferreira et al. [470] The role of solvation on the relative thermodynamic stabilities of cis- and trans-platinum dichloride by in aqueous solution was investigated by Hush et. al. [471] The energy profiles for the reaction $OH^- + CO_2 \rightarrow HCO_3^-$ in the presence of 30 water molecules were investigated by Nemukhin et al. [472] In another study by the same authors,[473] The QM/EFP method was used to model the reactions of hydroxymethyl radical (CH_2OH)-C with glutathione tripeptide (GSH) and with methylthiol(CH_3SH) in water.

Jose and Gadre[474] applied the MTA method to investigate the properties of Li clusters and CO_2 clusters.[475]. Mahadevi et al.[476] studied benzene clusters using the MTA method. Yang et al.[477] employed GEBF to study water clusters.

4.2 Biochemical Systems

Biochemical systems are a major application field for fragment methods, because the molecules in biology are large, often containing thousands of atoms. Fragment methods have several competitors, one is force fields; another is electronic structure approaches that rely either on semi-empirical approximations[478,479] or multi-layer approaches, like ONIOM[144] and QM/MM. [480] In addition, purely ab initio methods are also becoming increasingly efficient.[481,482]

Force fields are highly tuned for biochemical calculations, and are ubiquitously used. Their drawback is that frequently they are not polarizable and do not account for charge transfer (EFP and SIBFA are examples in which both are considered), and in addition, they are often difficult to use in practice by non-experts if the system contains non-standard parts for which no precomputed parameters are available. Force fields do have a major advantage of their speed and the ability to use them for molecular dynamics simulations, often essential to describe biological processes. On the other hand, ab initio based methods, including fragment-based approaches can incorporate full many-body polarization and charge transfer, and they do not rely on the need for parameters (aside from basis sets); of course the associated cost is an increased amount of computations.

4.2.1 Polypeptides, proteins, saccharides and oligoamides

Jensen and coworkers[483,484, 485, 486, 487, 488, 489] have successfully applied the QM/EFP approach with a buffer region to predicting pK_a values (Fig. 6) of ionizable residues in proteins. Xie et al.[490] employed the X-Pol method at the AM1 level to perform a 50 ps MD simulation of bovine pancreatic trypsin inhibitor in water with periodic boundary conditions.

www.manaraa.com

Komeiji et al.[491] applied the FMO/MP2/6-31G* method to analyze the dependence of the change in the electronic structure of ubiquitin (PDB: 1UBQ) on the thickness of solvating water layers. Five configurations corresponding to local minima were averaged. He et al.[492] applied the FMO/MP2/PCM/6-31G* method to study the ability of FMO and empirical dispersion to discriminate between the native and decoy structures for the Pin1 WW domain (PDB: 1I6C) and the Co repressor protein (PDB: 1ORC). Sawada et al.[493] used the FMO/RHF/PCM/6-31G* method to optimize the structure of helical heparin oligosaccharides (PDB: 1HPN). They found that the optimized structure is in good agreement with the NMR experiment. In addition, a comparison of the FMO predictions with those of the force field revealed the differences pointing to possible deficiencies in the force field model.

Huang et al.[494] applied the KEM with MP2/6-31G** to analyze the interactions in vesicular stomatitis virus nucleoprotein (PDB: 2QVJ) containing 33,175 atoms. Duan et al.[495] used the MFCC method at the HF, B3LYP, and MP2 levels with the 6-31+G* basis set to investigate the interaction of HIV-1 protease with the water molecule that bridges the flaps of the protease with the inhibitors. Dong et al.[496] applied the GEBF method at the DFT/6-311+G** level to study the formation of single and double helices of aromatic oligoamides. Deshmukh et al. applied the MTA method to study intramolecular hydrogen bonding in sugars[497] at the MP2/6-311++G(2d,2p) level and polypeptides[498] with DFT/6-311++G** level.

4.2.2 Protein-ligand binding

A number of applications of the FMO method to various protein-ligand complexes have been performed. Fukuzawa et al. [499,500,501] studied human estrogen receptor and showed good correlation with the experimental binding energies. Sawada et al. [502,503,504,505] performed a number of FMO applications to influenza virus hemagglutinin and showed the importance of considering the full-sized hemagglutinin upon the binding energies. A very detailed analysis[506] of the recognition patterns of sialosides by avian and human influenza hemagglutinins was performed using MP2/PCM/6-31G*, resulting in good agreement with the experimental binding energies. In another study[507] at the same level it was shown that binding of influenza A virus hemagglutinin to the sialoside receptor is not controlled by the homotropic allosteric effect (i.e., hemagglutinin trimerization does not not increase the binding energy per ligand). Future mutations[508] of the influenza virus were predicted by a combination of hemadsorption experiment and quantum chemical calculations for antibody binding, and details of the ligand binding were investigated. [509,261]

Yamagishi et al. [510,511,512] performed an analysis of the functions of key residues in the ligand-binding pocket of vitamin D. Ito et al. [513,514,515] analyzed the role of the functional groups in retinoid X receptor and studied the influence of mutations upon the transcriptional activation. Nakanishi et al.[516] applied FMO/MP2/6-31G* to elucidate the molecular recognition mechanism in the FK506 binding protein. A number of other protein-ligand complexes[517,518,519,520,521,522,523,524,525,526,527] as well as systems involving nucleic acids[528,529,530,531] were also studied with the FMO method.

www.manaraa.com

Taking advantage of the FMO method in the GAMESS and ABINIT-MP packages, several pharmaceutical companies have used it for drug-design related research. Ozawa et al. applied the FMO MP2/6-31G* approach [532] to demonstrate that CH/p hydrogen bonds determine the selectivity of the Src homology 2 domain to tyrosine phosphotyrosyl peptides. This was followed by a study[533] that showed the importance of CH/p hydrogen bonds in rational drug design as exemplified by leukocyte-specific protein tyrosine kinase. Fujimura and Sasabuchi[534] applied FMO/MP2/6-31G* to elucidate the role of fluorine atoms in a fluorinated prostaglandin agonist. Ohno et al.[535] employed FMO MP2/6-31G and discussed the strong correlation of pair interaction energies (PIEs) with the drug's potency.

The MFCC method has been applied to a number of ligand binding studies.[536,537,538,539,540] Huang et al.[541] employed KEM to study the interaction of aminoglycoside drugs and ribosomal A site RNA targets. Using the ELG method, Orimoto et al. [542] investigated the electronic structure of B-type poly(dG)·poly(dC) DNA.

4.2.3 Quantitative structure-activity relationship (QSAR)

It is very difficult to evaluate the free energies of binding from first principles with the accuracy of 1 kcal/mol, which is often the difference between several ligands in protein-ligand binding. In practical ab initio calculations, there are not only basis set and wave function limitations and the difficulty in describing solvent and counter-ions, but also the entropic contribution at room temperature requires proper configurational sampling, which at present is usually done with MD and can require long-time trajectories.

In QSAR studies, one introduces empirical factors by taking computed physical quantities, $\{x_i\}$called descriptors as arguments of a function $f(\{x_i\})$ (often, *f* is the binding energy), the coefficients in which are optimized for some training set of systems with $f(\{x_i\})$ known experimentally. The obtained relation is used to predict properties of compounds outside the training set, for which $\{x_i\}$ are computed. The underlying condition for successful QSAR studies is that descriptors should correlate with the desired property. Because of the fitting nature of QSAR, various completely unrelated methods can be used to compute descriptors. [543] For example, some descriptors can be computed from gas phase ab initio based calculations, the solvation effects can be estimated with the Poisson-Boltzmann model, and the entropic factor descriptor can be obtained from other simple models such as the number of rotatable bonds. [544]

Fragment-based methods seem to be a perfect match for fragment-based drug discovery (FBDD)[545] and they can be utilized to provide descriptors in QSAR. In particular, a very useful set of descriptors is given by the pair interactions between ligands and constituent parts of the protein (residues or residue fragments). Because the FMO method provides PIEs as a by-product of the calculations, it is straightforward to apply FMO to QSAR

www.manaraa.com

studies, as reviewed by Yoshida et al. [546] In other one-step methods such as KEM, MFCC or PMISP methods one can have similar pair interaction energies, whereas two-step methods operate with the properties of the whole system. Ishikawa et al. [547] discussed the basis sets effects upon the PIEs.

Yoshida et al. [548] employed the FMO RHF/6-31G method in QSAR studies of cyclic urea type HIV-1 PR inhibitors, using the sum of the residue fragment – ligand PIEs and the charge transfer between the ligand and protein from FMO calculations as descriptors. They found a strong correlation between the FMO binding energy in vacuum and the sum of protein-ligand PIEs, thus only the former was used.

Because the foregoing strategy was also used in further studies, it is instructive to comment on the relation between the sum of the protein-ligand PIEs ΔE_{PL} in the protein-ligand complex PL and the binding energy ΔE_{b} computed as the difference between the energies of the complex PL and the isolated protein (P) and ligand (L). Both ΔE_{PL} and ΔE_{b} can be used as descriptors; the former is often used as an approximation and replacement of the latter. For simplicity, the deformation energy, i.e., the change in the geometry of the protein and the ligand, when comparing their isolated and complexed minima, is implicitly included in the following derivations.[516]

$$\Delta E_{\mathrm{b}} = E_{\mathrm{PL}} - E_{\mathrm{P}} - E_{\mathrm{L}} =$$
$$\left[\sum_{I \in P} E'^{\mathrm{P(PL)}}_{I} + \sum_{I \in L} E'^{\mathrm{L(PL)}}_{I} + \sum_{\substack{I>J \\ I,J \in P}} \Delta E^{\mathrm{P(PL)}}_{IJ} + \sum_{\substack{I>J \\ I,J \in L}} \Delta E^{\mathrm{L(PL)}}_{IJ} + \sum_{I \in P, J \in L} \Delta E^{\mathrm{PL(PL)}}_{IJ} \right] - \qquad (125)$$
$$\left[\sum_{I \in P} E'^{\mathrm{P}}_{I} + \sum_{\substack{I>J \\ I,J \in P}} \Delta E^{\mathrm{P}}_{IJ} \right] - \left[\sum_{I \in L} E'^{\mathrm{L}}_{I} + \sum_{\substack{I>J \\ I,J \in L}} \Delta E^{\mathrm{L}}_{IJ} \right]$$

$$\Delta E_{\mathrm{PL}} = \sum_{I \in P, J \in L} \Delta E^{\mathrm{PL(PL)}}_{IJ}$$

P(PL) indicates the partial properties of P in the complex PL (and, similarly, for L). The difference is then

$$\Delta E_{\mathrm{b}} - \Delta E_{\mathrm{PL}} = \sum_{I \in P} \Delta E'^{\mathrm{P(PL)}}_{I} + \sum_{I \in L} \Delta E'^{\mathrm{L(PL)}}_{I} + \sum_{\substack{I>J \\ I,J \in P}} \Delta\Delta E^{\mathrm{P(PL)}}_{IJ} + \sum_{\substack{I>J \\ I,J \in L}} \Delta\Delta E^{\mathrm{L(PL)}}_{IJ} \qquad (126)$$

where (A = P or L)

www.manaraa.com

$$\Delta E_I'^{A(\mathrm{PL})} = E_I'^{A(\mathrm{PL})} - E_I'^{A}$$
$$\Delta\Delta E_{IJ}'^{A(\mathrm{PL})} = \Delta E_{IJ}'^{A(\mathrm{PL})} - \Delta E_{IJ}'^{A} \qquad (127)$$

$\Delta E_I'^{A(\mathrm{PL})}$ and $\Delta\Delta E_{IJ}'^{A(\mathrm{PL})}$ are the internal fragment and pair interaction energies, respectively, and are affected by the polarization and deformation in the complex formation. In other words,

$$\Delta E_{\mathrm{b}} - \Delta E_{\mathrm{PL}} = \Delta E_{\mathrm{P}}^{\mathrm{pold+def}} + \Delta E_{\mathrm{L}}^{\mathrm{pold+def}} \qquad (128)$$

$$\Delta E_A^{\mathrm{pol+def}} = \sum_{I \in A} \Delta E_I'^{A(\mathrm{PL})} + \sum_{\substack{I>J \\ I,J \in A}} \Delta\Delta E_{IJ}^{A(\mathrm{PL})} \qquad (129)$$

The difference between the binding energy ΔE_{b} and the sum of the pair interaction energies ΔE_{PL} is simply the destabilization polarization (denoted as "pold") plus the deformation energy of the protein and ligand. $\Delta E_A^{\mathrm{pold+def}}$ does not include all terms in the polarization. The polarization process (see PIEDA[297]) between isolated systems (P and L here) is divided into two contributions: first, each of the two interacting systems is destabilized relative to its own lowest energy state due to the polarization by the other (this gives the destabilization polarization). Consequently, the polarized systems interact, and this interaction energy includes the electrostatic energy, whose role is the stabilizing polarization energy, exchange-repulsion, charge transfer, dispersion and higher order terms. These terms are separable in the EDA[55] or its FMO extension, PIEDA[297] when *A* are individual fragments.

When the polarization and deformation energies vary considerably from ligand to ligand (for instance, when comparing ligands with different charges), the use of ΔE_{PL} in place of ΔE_{b} may lead to errors in consequent QSAR, and care should be taken to use a proper training set. Finally, for completeness, it is useful to separate[516] the sum of the polarization and deformations energies given by the difference in the energy of the protein or ligand in the complexed and isolated states.

$$\Delta E_A^{\mathrm{pold+def}} = E^{A(\mathrm{PL})} - E^{A} = E^{A(\mathrm{PL})} - E^{A} - \tilde{E}^{A} + \tilde{E}^{A} = \Delta E_A^{\mathrm{pold}} + \Delta E_A^{\mathrm{def}}$$
$$\Delta E_A^{\mathrm{pold}} = E^{A(\mathrm{PL})} - \tilde{E}^{A} \qquad (130)$$
$$\Delta E_A^{\mathrm{def}} = \tilde{E}^{A} - E^{A}$$

$\tilde{E}^{A}$ is the energy of isolated protein or ligand at the geometry in the complex, whereas E^{A} is computed at the geometry of A in its minimum. This analysis can be used for any type of interactions, not just those between a protein and a ligand, and more subsystems can be defined. In addition, the FMO method provides all details of

www.manaraa.com

individual fragment contributions in the total sums, for further insight regarding drug design and other applications. Moreover, this analysis defines the protein and ligand destabilization polarization energy ΔE_A^{pold} (often neglected in MM studies), as well as the deformation energy ΔE_A^{def} (typically included). The stabilizing polarization (pols) and charge transfer, as well as other types of interactions are included in the sum of the protein-ligand PIEs ΔE_{PL} . The former can be estimated[516] using the simple response relation $\Delta E_A^{\text{pols}} \approx -2\Delta E_A^{\text{pold}}$ (see actual calculations[297] for numerical justification), which leads to the total polarization

$$E_A^{\text{pol}} = \Delta E_A^{\text{pold}} + \Delta E_A^{\text{pols}} = -\Delta E_A^{\text{pold}} \qquad (131)$$

This way, the total polarization energy of the protein and ligand in their complex can be estimated, which requires 5 calculations (complex, plus protein and ligand, both at the geometry in the complexed and isolated states).

Fischer et al. [549] improved the scoring functions representing the binding energy for human estrogen receptor subtype a and human retinoic acid receptor of isotype g, using atomic charges from FMO calculations at the RHF/STO-3G level and concluded that such quantum scoring functions (QSF) describe the electrostatics accurately, and that QSF performs better than force field analogues. Yoshida et al. [550] applied FMO/RHF/6-31G to QSAR studies of the binding affinity of substituted benzenesulfonamides with carbonic anhydrase, noting the difficulties in modeling Zn^{2+}-containing systems. Hitaoka et al. [551] used FMO/MP2/6-31G in QSAR studies of the binding affinity of sialic acid analogues with influenza virus neuraminidase-1, where they divided the protein into three binding pockets and used the sums of PIEs for these subsystems as separate descriptors.

Munei et al. [552] used FMO/RHF/6-31G in QSAR studies of the binding affinity of substituted benzenesulfonamides with carbonic anhydrase. Mazanetz et al.[543] used FMO/MP2/6-31G* in QSAR studies of cyclin-dependent kinase 2 inhibitor potency and did a careful comparison of FMO and MM derived QSAR models popular in drug design industry. They found that the FMO outperformed all three MM based QSAR models.

4.2.4 Excited states and chemical reactions

Excited states and chemical reactions are an attractive application field of ab initio based methods. Ishida et al.[553] using the FMO method at the RHF and MP2 levels with the 6-31G(d) basis set analyzed the mechanism of the chorismate mutase reaction. Nakamura et al.[554] applied the FMO/MP2/6-31G method to clarify the role of K151 and D180 in L-2-haloacid dehalogenase from Pseudomonas sp. YL (PDB: 1ZRM). Pruitt et al.[286] using the multi-layer open shell (MP2/6-311G**: CR-CC(2,3)/3-21G) FMO method evaluated the enthalpy of the reversible addition-fragmentation chain transfer (RAFT) reaction involving radical species.

www.manaraa.com

In treating excited states with the FMO method, a frequently used approach is to apply an excited state calculation such as CI (or TDDFT) to a single fragment (chromophore), computed in the electrostatic field due to the other fragments, evaluated using RHF (or DFT). In this method (called FMO1, indicating that only monomers are computed) the electrostatic effect upon excited states is considered, which is often the major effect of the environment (i.e., the rest of the system excluding the chromophore), and higher order effects such as charge transfer can be considered with including dimer calculations of excited states (FMO2).

Mochizuki et al. did several FMO1 studies. They computed [555] red fluorescent protein with CIS(D)/6-31G(d), they simulated emission spectra of bioluminescent luciferases[556] with CIS(D)/6-31G, and they also performed CIS(D)/6-31G(d) calculations on the family of red [557] as well as blue and yellow[558] fluorescent proteins. Chiba et al. computed the yellow photoactive protein (PDB: 2PHY) both in gas phase[281] and solution[281] using FMO/TDDFT/6-31G(d), considering both FMO1 and FMO2 types of excitations. Ikegami et al.[559] analyzed the asymmetric excitations in the left and right branches of the photosynthetic reaction center of *Blastochloris viridis* (PDB: 1PRC) using the FMO method, where the asymmetry was studied for monomers and selected dimers at the CIS/6-31G(d) level. Milne et al.[560] applied FMO at the FMO1-TDDFT/6-311G(d,p) and FMO2-MP2/6-311G(d,p) levels to investigate the role of AMP protonation in firefly luciferase pH-sensitivity.

4.3 Solid State Applications

For systems with perfect periodicity, cluster based approaches compete with the methods employing periodic boundary conditions, which often use plane waves as a natural basis set for these systems. The latter group of methods is ultimately better, however, given the practical limitations of the existing theories and computer programs, and the difficulty in deriving and implementing high level of calculations for them often drives users to use cluster based approaches.

An important question is how relevant is the electron delocalization for a property one is interested in? Some properties, such as band gaps and the density of states, appear to require a consideration of the whole system, and fragment-based methods are at a serious disadvantage. However, there are ways to account for the delocalization at the final step by building the Fock matrices for the whole system, as is done in FMO-MO, FMO-LCMO and FMO/F, or in a different way in ELG. On the other hand, one is often interested in some local properties in periodic systems, such as the interactions determining their global properties, and for this, pair interactions in the FMO and KEM methods were found to be useful as described below. Applications of fragment-based methods to p-conjugated systems[325] and graphene[335] suggest their usefulness even for such delocalized systems.

4.3.1 Crystals, surfaces and nanomaterials

Huang et al.[561] applied the KEM/MP2/6-31G(d,p) method to investigate the interactions in the crystal of two molecules TDA1 and RangDP52. Fukunaga et al.[562] applied the FMO/TDDFT/6-31G(d,p) method to investigate

www.manaraa.com

the role of intermolecular interactions upon the excitations energies in three isomers of quinacridone crystals. To facilitate calculations and make them more realistic, an embedding model was used, in which a cluster of quinacridone molecules was immersed in the field of a large number of atomic charges, computed with the BLYP functional and periodic boundary conditions, with the 6-31G(d,p) basis set.

Faujasite zeolites were modeled by Fedorov et al.[231] using the FMO/RHF/6-31G(d) method. It was found that the adsorption energies can be quite accurately modeled with the FMO method, despite a large number of detached bonds. The fragmentation scheme is shown in Fig. 7. The phenol molecule occupies a place in the zeolite pore near the aluminum-containing fragment with acidic hydrogens.

Zhang et al.[563] applied the ELG RHF/STO-3G method to study the adsorption of Si and C chains onto unfaulted and faulted Si(111) surfaces. Chen et al.[564] calculated the electronic structure of the single-wall pristine boron nitride (BN) and boron nitride-carbon (BN/C) heterostructured nanotubes using ELG at the RHF/6-31G level.

Fedorov et al.[565] applied the FMO method at the B3LYP/3-21G(d) level to optimize the structure of silicon nanowires of diameter 1.2 and length 4.8 (nm), and showed that FMO optimized structures closely agree with those by PBC methods and experiment, suggesting that geometry can be optimized with reasonable results using fragment methods.

4.3.2 Polymers

Many applications of fragment-based methods are performed with the ELG method, which is particularly suited for (although not limited to) linear polymer calculations. Orimoto et al.[566] applied the ELG/PM3 level to optimize structures of polysilane derivatives, poly[bis(4-propoxybutyl)silylene].

The ELG method was used extensively to calculate polarizabilities and hyperpolarizabilities of important nonlinear optics materials. These two properties were studied using ELG by a number of researchers. Ohnishi et al.[567] computed donor/acceptor substituted polydiacetylenes at the RHF/6-31G level. Yu et al. using RHF/6-31G calculated polyimides[568] and $[Li^{+}[calix[4]pyrrole]Li^{-}]_{n}$, up to 15 units.[569] Pomogaeva et al.[570] at the RHF/6-311G level with ECP/VDZ for chalcogen atoms (S, Se and Te) studied series of benzo-2,1,3-chalcogendiasoles ribbon oligomers (up to 15 units). At the unrestricted PM3 level, Orimoto et al.[571] calculated a pyrrole-based spin-polarized molecular wire containing 1-pyrrolylphenyl nitronyl nitroxide with oligothiophene units under the influence of an applied electric field. Yan et al.[572] computed meso-mesolinked metalloporphyrin oligomers up to 22 units at RHF/6-31G (ECP/VDZ for metals Mg, Zn and Ni).

Pomogaev et al. using ELG at the CIS/INDO level computed absorption spectra for aromatic molecules (benzene, anthracene, 4-dicyanomethylene-4H-pyran, tryptophan and estradiol) bound to polyethylene[573] and explicitly solvated estradiol and tryptophan.

www.manaraa.com

5. Conclusions

During several decades, fragment-based methods have come a long way from the initial stage of method development to large scale applications, which span many types of systems: molecular clusters, proteins, DNA, oligosaccharides, zeolites, quantum dots, nanowires and others. Despite the very considerable progress, they remain underused; there may be several reasons for this. Some of the software developments are only locally implemented, making it difficult for most interested users to utilize the methods. Many, perhaps most methods are specific to one specific program; users who are unfamiliar with that program may have an inertial barrier to using it. Second, many applications so far have been performed in what should be considered demonstrative fashion, with low level wave functions and basis sets. Third, in some cases the applications are not conducted with properly considering all necessary effects and factors; the most conspicuous example is the need to incorporate solvent effects and entropy in biochemical applications.

Nevertheless, fragment-based methods also offer many advantages. One is the efficiency and the ability to compute realistic systems. Another is the additional information which they can deliver, such as the intrinsic details of the physical picture of the interactions in the system. As shown above, the application field is very broad encompassing most systems of finite size, which chemists and physicists are interested in. We expect that in near future with the revolutionary progress in the computer technology and the advent of multicore CPUs and GPUs, increasing level of calculations and the ease of their performance, the fragment-based methods will grow more popular in computational community.

www.manaraa.com

References

1 Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A new hierarchical parallelization scheme: generalized distributed data interface (GDDI), and an application to the fragment molecular orbital method (FMO), *J. Comput. Chem.* **2004**, *25*, 872.

2 Adams, W. H. On the Solution of the Hartree-Fock Equation in Terms of Localized Orbitals. *J. Chem. Phys.* **1961**, *34*, 89–102.

3 Klessinger, M.; McWeeny, R. Self-consistent group calculations on polyatomic molecules. I. Basis theory with an application to methane. *J. Chem. Phys.* **1965**, *42*, 3343–3354.

4 Christoffersen. R. E.; Maggiora, G. M. Ab initio calculations on large molecules using molecular fragments. Preliminary investigations. *Chem. Phys. Lett.* **1969**, *3*, 419.

5 Christoffersen, R. E.; Cheney, V.; Maggiora, G. M. Ab initio calculations on large molecules using molecular fragments. Hydrocarbon Characterizations. *J. Chem. Phys.* **1971**, *54*, 239.

6 Shipman, L. L.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Initial P-matrices for SCF calculations. *Chem. Phys. Lett.* **1972**, *15*, 469.

7 Shipman, L. L.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Evidence of hydrogen bonding in polypeptides of glycine. *Proc. Nat. Acad. Sci U.S.A.* **1972**, *69*, 3301.

8 Cheney, B. V.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Oxygen-containing molecules. *J. Chem. Phys.* **1972**, *56*, 3503.

9 Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Characterization of anthracene and phenanthrene. *Int. J. Quantum Chem.* **1973**, *7*, 169.

10 Shipman, L. L.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Characterization of the zwitterion of glycine. *Theor. Chim. Acta* **1973**, *31*, 75.

11 Shipman, L. L.; Christoffersen, R. E.; Cheney, V. Ab initio calculations on large molecules using molecular fragments. Lyncomycin model studies. *J. Med. Chem.* **1974**, *17*, 583.

12 Oie, T.; Maggiora, G. M.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Comparison of charge distribution and molecular electrostatic potential for ethyl chlorophyllide a and related molecules. *Int. J. Quantum Chem.* **1976**, *10*, 119.

13 Petke, J. D.; Christoffersen, R.E.; Maggiora, G. M.; Shipman, L. L. Ab initio calculations on large molecules using molecular fragments. SCF-MO-CI studies on low-lying triplet states of pyrazine. *Int. J. Quantum Chem.* **1977**, *12*, 343.

14 Murk, D.; Nietzsche, L. E.; Christoffersen, R. E. Ab initio calculations on large molecules using molecular fragments. Configuration interaction studies on the ground and lower excited states of N-methyl carbazole. *J. Am. Chem. Soc.* **1978**, *100*, 1371.

15 Frost, A. A. *J. Chem. Phys.* **1967**, *47*, 3707.

16 Stoll, H.; Preuß, H. On the direct calculation of localized HF orbitals in molecule clusters, layers and solids. *Theor. Chim. Acta* **1977**, *46*, 11–21.

www.manaraa.com

17 Stoll, H. *Phys. Rev. B* **1992**, *46*, 6700.

18 Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.

19 Szekeres, Z.; Surján, P. R. Direct determination of fragment localized molecular orbitals and the orthogonality constraint. *Chem. Phys. Lett.* **2003**, *369*, 125–130.

20 Imamura, A.; Aoki, Y.; Maekawa, K. A theoretical synthesis of polymers by using uniform localization of molecular-orbitals - proposal of an elongation method. *J. Chem. Phys.* **1991**, *95*, 5419–5431.

21 Demelo, C. P.; Dossantos, M. C.; Matos, M.; Kirtman, B., Local-space approximation for treatment of chemisorption – application to a model transition-metal system. *Phys. Rev. B* **1987**, *35*, 7847–7856.

22 Aoki, Y.; Imamura, A. Local density of states of aperiodic polymers using the localized orbitals from an abinitio elongation method. *J. Chem. Phys.* **1992**, *97*, 8432–8440.

23 Korchowiec, J.; Lewandowski, J.; Makowski, M.; Gu, F. L.; Aoki, Y. Elongation Cutoff Technique Armed with Quantum Fast Multipole Method for Linear Scaling. *J. Comput. Chem.* **2009**, *30*, 2515–2525.

24 Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. A new localization scheme for the elongation method. *J. Chem. Phys.* **2004**, *121*, 10385–10391.

25 Mitani, M.; Aoki, Y.; Imamura, A. A novel molecular-orbital method for the calculations of polymer systems with local aperiodic part – the combination of the elongation method with the supercell method. *J. Chem. Phys.* **1994**, *100*, 2346–2358.

26 Aoki, Y.; Suhai, S.; Imamura, A. An efficient cluster elongation method in density-functional theory and its application to poly-hydrogen-bonding molecules. *J. Chem. Phys.* **1994**, *101*, 10808–10823.

27 Aoki, Y.; Suhai, S.; Imamura, A., A density-functional elongation method for the theoretical synthesis of aperiodic polymers. *Int. J. Quantum Chem.* **1994**, *52*, 267–280.

28 Imamura, A.; Aoki, Y.; Nishimoto, K.; Kurihara, Y.; Nagao, A. Calculations of the electronic-structure of various aperiodic polymers by an elongation method. *Int. J. Quantum Chem.* **1994**, *52*, 309–319.

29 Mitani, M.; Aoki, Y.; Imamura, A. Electronic-structures of large, extended, nonperiodic systems by using the elongation method – model-calculations for the cluster series of a polymer and the molecular stacking on a surface. *Int. J. Quantum Chem.* **1995**, *54*, 167–196.

30 Kurihara, Y.; Aoki, Y.; Imamura, A. Calculations of the excitation energies of all-trans and 11,12s-dicis retinals using localized molecular orbitals obtained by the elongation method. *J. Chem. Phys.* **1997**, *107*, 3569–3575.

31 Mitani, M.; Aoki, Y.; Imamura, A. Geometry optimization of polymers by the elongation method. *Int. J. Quantum Chem.* **1997**, *64*, 301–323.

32 Imamura, A.; Aoki, Y. An elongation method to calculate the electronic structure of non-periodical and periodical polymers. *Adv. Colloid Interface Science* **1997**, *71-2*, 147–164.

33 Kurihara, Y.; Aoki, Y.; Imamura, A. Calculations of phase transition of polydiacetylenes using localized molecular orbitals by elongation method. *J. Chem. Phys.* **1998**, *108*, 10303–10308.

www.manaraa.com

34 Ladik, J.; Imamura, A.; Aoki, Y.; Ruiz, M.; Otto, P. Approximate methods of the calculation of effective energies in disordered chains I. Comparative theoretical study of the energy properties of structural isomers of polypropene. *J. Mol. Str. (THEOCHEM)* **1999**, *491*, 49–55.

35 Rather, G.; Aoki, Y.; Imamura, A. Performance of the elongation method with larger basis sets. *Int. J. Quantum Chem.* **1999**, *74*, 35–47.

36 Kim, J. T.; Lee, M. J.; Kim, U. R.; Kimura, M.; Aoki, Y.; Imamura, A. Theoretical synthesis of poly-(2-hydroxyethylmethacrylate) by uniform localization of molecular orbitals calculation. *J. Polymer Science Part a - Polymer Chem.* **2001**, *39*, 2677–2682.

37 Gu, F. L.; Aoki, Y.; Imamura, A.; Bishop, D. M.; Kirtman, B. Application of the elongation method to nonlinear optical properties: finite field approach for calculating static electric (hyper)polarizabilities. *Mol. Phys.* **2003**, *101*, 1487–1494.

38 Korchowiec, J.; Gu, F. L.; Imamura, A.; Kirtman, B.; Aoki, Y. Elongation method with cutoff technique for linear SCF scaling. *Int. J. Quantum Chem.* **2005**, *102*, 785–794.

39 Korchowiec, J.; Gu, F. L.; Aoki, Y. Elongation method at restricted open-shell Hartree-Fock level of theory. *Int. J. Quantum Chem.* **2005**, *105*, 875–882.

40 Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. Efficiency and accuracy of the elongation method as applied to the electronic structures of large systems. *J. Comput. Chem.* **2006**, *27*, 1603–1619.

41 Pomogaeva, A.; Springborg, M.; Kirtman, B.; Gu, F. L.; Aoki, Y. Band structures built by the elongation method. *J. Chem. Phys.* **2009**, *130*, 194106.

42 Korchowiec, J.; De Silva, P.; Makowski, M.; Gu, F. L.; Aoki, Y. Elongation Cutoff Technique at Kohn-Sham Level of Theory. *Int. J. Quantum Chem.* **2010**, *110*, 2130–2139.

43 Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. Describing Electron Correlation Effects in the Framework of the Elongation Method–Elongation-MP2: Formalism, Implementation and Efficiency. *J. Comput. Chem.* **2010**, *31*, 1733–1740.

44 Morokuma, K. *J. Chem. Phys.* **1971**, 55, 1236–1244.

45 Ohno, K.; Inokuchi, H. *Theor. Chim. Acta (Berl.)* **1972**, 26, 331–350.

46 Gao, J. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997,** *101*, 657–663.

47 Kutzelnigg, W.; Maeder, F. Natural states of interacting systems and their use for the calculation of intermolecular forces.: I. General theory of the natural states of interacting systems. *Chem. Phys.* **1978**, *32*, 451–455.

48 Otto, P.; Ladik, J. Investigation of the interaction between molecules at medium distances: I. SCF LCAO MO supermolecule, perturbational and mutually consistent calculations for two interacting HF and CH2O molecules. *Chem. Phys.* **1975**, *8*, 192–200.

49 Kubota, H.; Aoki, Y.; Imamura, A. *Bull. Chem. Jpn.* **1994**, 67, 13–20.

50 Barandiaran, Z.; Seijo, L. *J. Chem. Phys.* **1988**, 89, 5739.

www.manaraa.com

51 Pascual, J. L.; Seijo, L. Ab initio model potential embedded cluster calculations including lattice relaxation and polarization: Local distortions on Mn2+ -doped CaF2. *J. Chem. Phys.* **102**, *1995*, 5368–5376.

52 Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346–2354.

53 The term "self-consistent-charge" was used in various contexts with quite different exact mathematical meaning but alluding to the same physical concept. Some examples of the usages are (only the last one is for a fragment-based method): (a) Bardeen, J. Theory of the Work Function. II. The Surface Double Layer. *Phys. Rev.* **1936**, *49*, 653–663; (b) Rosén, A.; Ellis, D. E.; Adachi, H.; Averill, F. W. Calculations of molecular ionization energies using a self-consistent-charge Hartree-Fock-Slater method. *J. Chem. Phys.* **1976**, *65*, 3629; (c) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **2001**, *105*, 569–585. (d) Sugiki, S.-I. Kurita, N.; Sengoku, Y.; Sekino, H. Fragment molecular orbital method with density functional theory and DIIS convergence acceleration, *Chem. Phys. Lett.* **2003,** *382*, 611.

54 Fedorov, D. G.; Kitaura, K. The importance of three-body terms in the fragment molecular orbital method *J. Chem. Phys.* **2004,** *120***,** 6832.

55 Kitaura, K.; Morokuma, K. *Int. J. Quant. Chem.* **1976**, *10*, 325.

56 Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *100*, 14316.

57 Bohm, M. C. A simple self-consistent electrostatic field approximation for neighbour strand interactions in band structure calculations. *Chem. Phys. Lett.* **1982**, *89*, 126–130.

58 Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules, *Chem. Phys. Lett.* **1999,** *313***,** 701.

59 Otto, P.; Ladik, J. Investigation of the interaction between molecules at medium distances : II. Perturbational and MCF calculations with directly integrated potentials and in the monopole approximation. *Chem. Phys.* **1977**, *19*, 209–216.

60 Otto, P. Investigation of the interaction between molecules at medium distances. III. SCF LCAO MO supermolecule, perturbational and MCF calculations for two and three interacting molecules. *Chem. Phys.* **1978**, *33*, 407–414.

61 Otto, P. Ab intio calculationns of the vertical interaction between two and three cytosine molecules. *Chem. Phys. Lett.* **1979**, *62*, 538–541.

62 Otto, P.; Ladik, J. Concept of charge transfer: Investigation of charge transfer in the glyoxal-formamide, water dimer, and water-lithium cation systems. *Int. J. Quant. Chem.* **1980**, *18*, 1143–1155.

63 Förner, W.; Otto, P.; Bernhardt, J.; Ladik, J. A model study of the intermolecular interactions of amino acids in aqueous solution: The glycine-water system. *Theor. Chem. Acc. (Theor. Chim. Acta)*, **1981**, *60*, 269–281.

64 Zyss, J.; Berthier, G. Nonlinear optical properties of organic crystals with hydrogen-bonded molecular units: The case of urea. *J. Chem. Phys.* **1982**, *77*, 3635.

65 Otto, P. The pseudo-polarization tensor mutually consistent-field (PPT-MCF) method – a new approach to study intermolecular interactions and its application to dimeric and trimeric water configurations. *Int. J. Quant. Chem.* **1985**, *28*, 895–911.

www.manaraa.com

66 Otto, P.; Ladik, J.; Liu, S. C. Ab initio and semiempirical calculations on adducts of guanine and the metabolites of different carcinogens. *J. Mol. Str. (THEOCHEM)* **1985**, *123*, 129–140.

67 Otto, P. On the stability of single- and double-stranded DNA helices: The application of the PPT-MCF method on large fragments of DNA. *Int. J. Quant. Chem.* **1986**, *30*, 275–288.

68 Ladik, J. Ab-initio self-consistent field theory for the treatment of interface between two different quasi-one-dimensional chains. *Progress in Surface Science*, **1987**, *26*, 135–143.

69 Otto, P. Investigation of the stability of oligonucleotides and oligodinucleotides. *J. Mol. Str. (THEOCHEM)* **1989**, *188*, 277–288.

70 Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. *J. Chem. Phys.* **1970**, *53*, 4544–4554.

71 Suárez, E.; Díaz, N.; Suárez, D. Thermochemical Fragment Energy Method for Biomolecules: Application to a Collagen Model Peptide. *J. Chem. Theory Comput.* **2009**, *5*, 1667.

72 Gao, J.; Cembran, A.; Mo, Y., Generalized X-Pol Theory and Charge Delocalization States. *J. Chem. Theory Comput.* **2010**, *6*, 2402–2410.

73 Jensen, J.H.; Day, P.N.; Gordon, M.S.; Basch, H.; Cohen, D.; Garmer, D.R.; Kraus, M.; Stevens, W.J. Effective Fragment Method for Modeling Intermolecular Hydrogen-Bonding Effects on Quantum-Mechanical Calculations. *Modeling the Hydrogen Bond*, **1994**, *569*, 139–151.

74 Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W.J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modelling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293.

75 Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. The Effective Fragment Potential: A General Method for Predicting Intermolecular Forces. *Ann. Rep. Comp. Chem.* **2007**, *3*, 177.

76 Gresh, N.; Cisneros, G.A.; Darden, T.A.; Piquemal, J.P. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.

77 Ponder, J.W.; Wu, C.; Ren, P.; Pande, V.S.; Chodera, J.D.; Schnieders, M.J.; Haque, I.; Mobley, D.L.; Lambrecht, D.S.; DiStasio, R.A.; Head-Gordon, M.; Clark, G.N.I.; Johnson, M.E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, **2010**, *114*, 2549–2564.

78 Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.

79 Moszynski, R. Symmetry-adapted perturbation theory for the calculation of Hartree-Fock interaction energies. *Mol. Phys.* **1996**, *88*, 741–758.

80 Misquitta, A.J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005**, *123*, 214103–214114.

81 Bukowski, R.; Podeszwa, R.; Szalewicz, K. Efficient calculation of coupled Kohn-Sham dynamic susceptibility functions and dispersion energies with density fitting. *Chem. Phys. Lett.* **2005**, *414*, 111–116.

www.manaraa.com

82 Amos, R.D.; Handy, N.C.; Knowles, P.J.; Rice, J.E.; Stone, A.J. *Ab initio* Prediction of Properties of CO2, NH3, and CO2-NH3. *Journal of Physical Chemistry*, **1985**, *89*, 2186-2192.

83 Hesselmann, A.; Jansen, G.; Schutz, M. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103.

84 Podeszwa, R.; Bukowski, R.; Szalewicz, K. Density-fitting method in symmetry-adapted perturbation theory based on Kohn-Sham description of monomers. *J. Chem. Theory Comput.* **2006**, *2*, 400–412.

85 Kitaura, K.; Morokuma, K. Variational approach (SCF ab-initio calculations) to the study of molecular interactions: the origin of molecular interactions, in "Molecular Interactions", Vol. 1, edited by H. Ratajczak and W. J. Orville-Thomas, John Wiley and Sons, Chichester, 1980; p. 21.

86 Morokuma, K.; Kitaura, K. Chemical Applications of Atomic and Molecular Electrostatic Potentials; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981; p 215.

87 Stevens, W.J.; Fink, W.H. Frozen Fragment Reduced Variational Space Analysis of Hydrogen-Bonding Interactions – Application to the Water Dimer. *Chem. Phys. Lett.* **1987**, *139*, 15–22.

88 Bagus, P.S.; Hermann, K.; Bauschlicher, J.C.W. A new analysis of charge transfer and polarization for ligand--metal bonding: Model studies of Al[sub 4]CO and Al[sub 4]NH[sub 3]. *J. Chem. Phys.* **1984**, *80*, 4378–4386.

89 Bagus, P.S.; Illas, F. Decomposition of the chemisorption bond by constrained variations: Order of the variations and construction of the variational spaces. *J. Chem. Phys.* **1992**, *96*, 8962–8970.

90 Glendening, E.D.; Streitwieser, A. Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding, strong ionic, and moderate donor-acceptor interactions. *J. Chem. Phys.* **1994**, *100*, 2900–2909.

91 Schenter, G.K.; Glendening, E.D. Natural Energy Decomposition Analysis: The Linear Response Electrical Self Energy. *J. Phys. Chem.* **1996**, *100*, 17152–17156.

92 Glendening, E.D. Natural Energy Decomposition Analysis: Extension to Density Functional Methods and Analysis of Cooperative Effects in Water Clusters. *J. Phys. Chem. A*, **2005**, *109*, 11936–11940.

93 Khaliullin, R.Z.; Cobar, E.A.; Lochan, R.C.; Bell, A.T.; Head-Gordon, M. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A*, **2007**, *111*, 8753–8765.

94 Mo, Y.R.; Gao, J.L.; Peyerimhoff, S.D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530–5538.

95 Fedorov, D. G.; Kitaura, K. Pair interaction energy decomposition analysis. *J. Comput. Chem.* **2007**, *28*, 222–237.

96 Su, P.F.; Li, H. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.* **2009**, *131*, 014102.

97 Hayes, I.C.; Stone, A.J. An intermolecular perturbation theory for the region of moderate overlap. *Mol. Phys.* **1984**, *53*, 83–105.

www.manaraa.com

98 Wu, Q.; Ayers, P.W.; Zhang, Y.K. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.

99 Wu, Q.; Yang, W. A direct optimization method for calculating density functionals and exchange-correlation potentials from electron densities. *J. Chem. Phys.* **2003**, *118*, 2498–2509.

100 Wu, Q.; Van Voorhis, T. Direct optimization method to study constrained systems within density-functional theory. *Phys. Rev. A*, **2005**, *72*, 024502.

101 Curutchet, C.; Bofill, J.M.; Hernández, B.; Orozco, M.; Luque, F.J. Energy decomposition in molecular complexes: Implications for the treatment of polarization in molecular simulations. *J. Comput. Chem.* **2003**, *24*, 1263–1275.

102 Ponder, J.W.; Wu, C.; Ren, P.; Pande, V.S.; Chodera, J.D.; Schnieders, M.J.; Haque, I.; Mobley, D.L.; Lambrecht, D.S.; DiStasio, R.A.; Head-Gordon, M.; Clark, G.N.I.; Johnson, M.E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, **2010**, *114*, 2549–2564.

103 Cisneros, G.A.; Darden, T.A.; Gresh, N.; Pilmé, J.; Reinhardt, P.; Parisel, O.; Piquemal, J.P., *Design Of Next Generation Force Fields From AB Initio Computations: Beyond Point Charges Electrostatics*, in *Multi-scale Quantum Models for Biocatalysis*, York, D.M. and Lee, T.-S., Editors. 2009, Springer Netherlands. p. 137–172.

104 Piquemal, J.P.; Cisneros, G.A.; Reinhardt, P.; Gresh, N.; Darden, T.A. Towards a force field based on density fitting. *J. Chem. Phys.* **2006**, *124*, 104101.

105 Piquemal, J.P.; Perera, L.; Cisneros, G.A.; Ren, P.Y.; Pedersen, L.G.; Darden, T.A. Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J. Chem. Phys.* **2006**, *125*, 054511.

106 Halgren, T.A.; Damm, W. Polarizable force fields. *Current Opinion in Structural Biology*, **2001**, *11*, 236–242.

107 Gianinetti, E.; Raimondi, M.; Tornaghi, E. Modification of the Roothaan equations to exclude BSSE from molecular interaction calculations. *Int. J. Quant. Chem.* **1996**, *60*, 157-166.

108 Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M., *Extension of the SCF-MI Method to the Case of K Fragments one of which is an Open-Shell System*, in *Advances in Quantum Chemistry*. 1998, Academic Press. p. 251-266.

109 Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism. *The J. Chem. Phys.* **2001**, *115*, 3553-3560.

110 Nagata, T.; Iwata, S. Perturbation expansion theory corrected from basis set superposition error. I. Locally projected excited orbitals and single excitations. *J. Chem. Phys.* **2004**, *120*, 3555-3562.

111 Khaliullin, R.Z.; Head-Gordon, M.; Bell, A.T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105-204111.

112 F. Y. Nemukin, P.J. Knowles, and J.N. Murrell, Towards reliable modelling of large clusters: on the overall accuracy of the diatomics-in-molecule method for rare gas cluster ions, *Chem. Phys.* **1995**, 193, 27.

www.manaraa.com

113 Jensen, J. H.; Gordon, M. S. An Approximate Formula For the Intermolecular Pauli Repulsion Between Closed Shell Molecules. *Mol. Phys.* **1996**, *89*, 1313.

114 Huxley, P.; Knowles, D. B.; Murrell, J. N.; Watts, J. D. Ground-state diatomic potentials. Part 2.-Van der Waals molecules. *J. Chem. Soc., Faraday Trans. 2*, **1984**, *80*, 1349–1361.

115 Brobjera, J. T.; Murrell, J. N. A comparison of the multipole-expansion and point-charge models of the electrostatic energy between polar diatomic molecules. *Chem. Phys. Lett.* **1981**, *77*, 601.

116 Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. Additivity rules for the estimation of thermochemical properties. *Chem. Rev.* **1969**, *69*, 279–324.

117 Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New advance in computational chemistry: Full quantum mechanical ab initio computation of streptavidin-biotin interaction energy. *J. Phys. Chem. B* **2003**, *107*, 12039-12041.

118 Li, W.; Li, S. H.; Jiang, Y. S., Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193-2199.

119 Babu, K.; Gadre, S. R. Ab initio quality one-electron properties of large molecules: Development and testing of molecular tailoring approach. *J. Comput. Chem.* **2003**, *24*, 484-495.

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

121 Deev, V.; Collins, M.A. *J. Chem. Phys.* **2005**, *122*, 154102.

122 Olsen, J.; Roos, B. O.; Jøgensen, P.; Jensen, H. J. Aa. *J. Chem. Phys.* **1988**, *89*, 2185.

123 Nakano, H.; Hirao, K. *Chem. Phys. Lett.* **2000**, *317*, 90.

124 Ivanic J. *J. Chem. Phys.* **2003**, *119*, 9364.

125 Sauri, V.; Serrano-Andres, L.; Shahi, A. R. M.; Gagliardi, L.; Vancoilie, S.; Pierloot, K. *J. Chem. Theory Comput.* **2010**, *7*, 153.

126 Roskop, L.; Gordon, M. S. *J. Chem. Phys.*, submitted.

127 Werner, H.-J.; Pfluger, K. On the selection of domains and orbital pairs in local correlation treatments. *Ann. Reports in Comput. Chem.* **2006**, *2*, 53.

128 Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151.

129 Saebo, S.; Pulay, P. *Chem. Phys. Lett.* **1985**, *113*, 13.

130 Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 241101, and references cited therein.

131 Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 034103.

132 Venkatnathan, A.; Szilva, A. B.; Walter, D.; Gdanitz, R. J.; Carter, E. A. *J. Chem. Phys.* **2004**, *120*, 1693.

133 Li, W.; Piecuch, P. *J. Phys. Chem. A* **2010**, *114*, 6721.

134 Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2000**, *114*, 661.

www.manaraa.com

135 Paulus, B. *Phys. Rep.* **2006**, *428*, 1.

136 Friedrich, J.; Hanrath, M.; Dolg, M. Energy screening for the incremental scheme: application to intermolecular interactions. *J. Phys. Chem. A* **2007**, 111, 9830-9837.

137 Friedrich, J.; Hanrath, M.; Dolg, M. Using symmetry in the framework of the incremental scheme: molecular applications. *Chem. Phys.* **2008**, 346, 266-274.

138 Mata, R.A.; Stoll, H. Incremental expansions for SCF interaction energies: A comparison for hydrogen-bonded clusters. *Chem. Phys. Lett.* **2008**, *465*, 136-141.

139 Manby, F. R.; Alfè, D.; Gillan, M. J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5178.

140 Nolan, S. J.; Bygrave, P. J.; Allan, N. L.; Manby, F. R. Comparison of the incremental and hierarchical methods for crystalline neon. *J. Phys.: Condens. Matter* **2010**, *22*, 074201.

141 Zalesny, R.; Papadopoulos, M.G.; Mezey, P.G.; Leszczynski, J. (Eds.), *Linear-Scaling Techniques in Computational Chemistry and Physics*, Springer, Berlin, 2011.

142 Reimers, J. R., Ed. *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, Wiley, New York, 2011.

143 Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

144 Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J Phys. Chem.* **1996**, *100*, 19357–19363.

145 Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for *n*-butane conformation and SN2 reaction: RCl+Cl-. *J. Chem. Phys.* **1996**, *105*, 1959-1967.

146 Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. Analytic gradients for the multicentred integrated QM : QM method for weakly bound clusters: efficient and accurate 2-body : many-body geometry optimizations, *Mol. Phys.* **2007**, *105*, 2777-2782.

147 Li, H.; Li, W.; Li, S. H.; Ma, J. Fragmentation-based QM/MM simulations: Length dependence of chain dynamics and hydrogen bonding of polyethylene oxide and polyethylene in aqueous solutions. *J. Phys. Chem. B* **2008,** *112*, 7061-7070.

148 He, X.; Wang, B.; Merz, K. M. *J. Phys. Chem. B* **2009**, *113*, 10380–10388.

149 Guo, W.; Wu, A.; Xu, X. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203-208.

(150) Morita, S.; Sakai, S. *J. Comput. Chem.* **2001**, *22*, 1107.

151 Thole, B.T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341-350.

www.manaraa.com

152 Åstrand, P.O.; Linse, P.; Karlström, G. Molecular dynamics study of water adopting a potential function with explicit atomic dipole moments and anisotropic polarizabilities. *Chem. Phys.* **1995**, *191*, 195-202.

153 Swart, M.; van Duijnen, P.T. DRF90: a polarizable force field. *Mol. Sim.* **2006**, *32*, 471-484.

154 Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988** *38*, 3098.

155 Gresh, N.; Claverie, P.; Pullman, A. *Theor. Chim Acta* **1984**, *66*, 1.

156 Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1986**, *29*, 101.

157 Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856.

158 Gresh, N.; Guo, H.; Kafafi, S. A.; Salahub, D. R.; Roques, B. P. *J. Am. Chem. Soc.* **1999**, *121*, 7885.

159 Gresh, N.; Piquemal, J.P.; Krauss, M. Representation of Zn(II) complexes in polarizable molecular mechanics. Further refinements of the electrostatic and short-range contributions. Comparisons with parallel ab initio computations. *J. Comput. Chem.* **2005**, *26*, 1113-1130.

160 Vigné-Maeder, F.; Claverie, P. *J. Chem Phys.* **1988**, *88*, 4934.

161 Piquemal, J.P.; Gresh, N.; Giessner-Prettre, C. Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A*, **2003**, *107*, 10353-10359.

162 Piquemal, J.P.; Chevreau, H.; Gresh, N. Toward a separate reproduction of the contributions to the Hartree-Fock and DFT intermolecular interaction energies by polarizable molecular mechanics with the SIBFA potential. *J. Chem. Theory Comput.* **2007**, *3*, 824-837.

163 Creuzet, S.; Langlet, J.; Gresh, N. *J. Chim.-Phys. Phys. Chim.Biol.* **1991**, *88*, 2399.

164 Murrell, J. N.; Randic, M.; Williams, D. R. *Proc. R. Soc.London, Ser. A* **1966**, *284*, 566.

165 Chaudret, R.; Ulmer, S.; van Severen, M.C.; Gresh, N.; Parisel, O.; Cisneros, G.A.; Darden, T.A.; Piquemal, J.P. Progress Towards Accurate Molecular Modeling of Metal Complexes Using Polarizable Force Fields. *Theory and Applications of Computational Chemistry - 2008*, **2009**, *1102*, 185-192

166 Day, P.N.; Jensen, J.H.; Gordon, M.S.; Webb, S.P.; Stevens, W.J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968-1986.

167 Gordon, M.S.; Freitag, M.A.; Bandyopadhyay, P.; Jensen, J.H.; Kairys, V.; Stevens, W.J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293-307.

168 Stone, A.J. Distributed Multipole Analysis, or How to Describe a Molecular Charge-Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233-239.

169 Stone, A.J., *The Theory of Intermolecular Forces*. 1996: Oxford University Press: Oxford.

170 Pruitt, S. R.; Sok, S.; Xu, P.; Fedorov, D. G.; Gordon, M. S. *J. Am. Chem. Soc.*, to be submitted.

www.manaraa.com

171 Arora, P.; Slipchenko, L.V.; Webb, S.P.; Defusco, A.; Gordon, M.S. Solvent-Induced frequency shifts: Configuration Interaction Singles combined with the Effective Fragment Potential Method. *J. Phys. Chem. A* **2010**, *114*, 6742-6750.

172 Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M.S. Solvent effects on optical properties of molecules: A combined time-dependent density functional theory/effective fragment potential approach. *J. Chem. Phys.* **2008**, *129*, 144112-144118.

173 Defusco, A.; Gordon, M.S. *unpublished*.

174 Kosenkov, D.; Slipchenko, L.V. Solvent Effects on the Electronic Transitions of p-Nitroaniline: A QM/EFP Study. *J. Phys. Chem. A*, **2010**, *115*, 392-401.

175 Slipchenko, L.V. Solvation of the Excited States of Chromophores in Polarizable Environment: Orbital Relaxation versus Polarization. *J. Phys. Chem. A*, **2010**, *114*, 8824-8830.

176 Ghosh, D.; Isayev, O.; Slipchenko, L.V.; Krylov, A.I. The effect of solvation on vertical ionization energy of thymine: from microhydration to bulk. *J. Phys. Chem. A*, **2011**, *in press*.

177 Nagata, T.; Fedorov, D.G.; Kitaura, K.; Gordon, M.S. A combined effective fragment potential-fragment molecular orbital method. I. The energy expression and initial applications. *J. Chem. Phys.* **2009**, *131*, 024101.

178 Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K.; Gordon, M. S. A combined effective fragment potentialfragment molecular orbital method. II. Analytic gradient and application to the geometry optimization of solvated tetraglycine and chignolin. *J. Chem. Phys.* **2011**, *134*, 034110.

179 Li, H.; Pomelli, C.S.; Jensen, J.H. Continuum solvation of large molecules described by QM/MM: a semi-iterative implementation of the PCM/EFP interface. *Theo. Chem. Acc.* **2003**, *109*, 71-84.

180 Bandyopadhyay, P.; Gordon, M.S.; Mennucci, B.; Tomasi, J. An integrated effective fragment---polarizable continuum approach to solvation: Theory and application to glycine. *J. Chem. Phys.* **2002**, *116*, 5023-5032.

181 Li, H.; Gordon, M.S. Polarization energy gradients in combined quantum mechanics, effective fragment potential, and polarizable continuum model calculations. *J. Chem. Phys.* **2007**, *126*, 124112-124110.

182 Li, H. Quantum mechanical/molecular mechanical/continuum style solvation model: Linear response theory, variational treatment, and nuclear gradients. *J. Chem. Phys.* **2009**, *131*, 184103-184108.

183 Zorn, D.; Lin, V.S.Y.; Pruski, M.; Gordon, M.S. An interface between the universal force field and the effective fragment potential method. *J. Phys. Chem. B*, **2008**, *112*, 12753-12760.

184 Netzloff, H.M.; Gordon, M.S. Fast fragments: The development of a parallel effective fragment potential method. *J. Comput. Chem.* **2004**, *25*, 1926-1935.

185 Day, P.N.; Pachter, R.; Gordon, M.S.; Merrill, G.N. A study of water clusters using the effective fragment potential and Monte Carlo simulated annealing. *J. Chem. Phys.* **2000**, *112*, 2063-2073.

186 Chen, W.; Gordon, M.S. The effective fragment model for solvation: Internal rotation in formamide. *J. Chem. Phys.* **1996**, *105*, 11081-11090.

187 Merrill, G.N.; Gordon, M.S. Study of Small Water Clusters Using the Effective Fragment Potential Model. *J. Phys. Chem. A* **1998**, *102*, 2650-2657.

www.manaraa.com

188 Netzloff, H.M.; Gordon, M.S. The effective fragment potential: Small clusters and radial distribution functions. *J. Chem. Phys.* **2004**, *121*, 2711-2714.

189 Webb, S.P.; Gordon, M.S. Solvation of the Menshutkin Reaction: A Rigorous Test of the Effective Fragment Method. *J. Phys. Chem. A* **1999**, *103*, 1265-1273.

190 Adamovic, I.; Gordon, M.S. Solvent Effects on the SN2 Reaction: Application of the Density Functional Theory-Based Effective Fragment Potential Method. *J. Phys. Chem. A* **2005**, *109*, 1629-1636.

191 Bandyopadhyay, P.; Gordon, M.S. A combined discrete/continuum solvation model: Application to glycine. *J. Chem. Phys.* **2000**, *113*, 1104-1109.

192 Adamovic, I.; Freitag, M.A.; Gordon, M.S. Density functional theory based effective fragment potential method. *J. Chem. Phys.* **2003**, *118*, 6725-6732.

193 Song, J.; Gordon, M.S. *unpublished*.

194 Gordon, M.S.; Slipchenko, L.V.; Li, H.; Jensen, J.H. The Effective Fragment Potential: A General Method for Predicting Intermolecular Forces. *Annual Reports in Computational Chemistry*, **2007**, *3*, 177-193.

195 Gordon, M.S.; Mullin, J.M.; Pruitt, S.R.; Roskop, L.B.; Slipchenko, L.V.; Boatz, J.A. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646-9663.

196 Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C.F.; Herbert, J.M.; Gordon, M.S.; Schmidt, M.W.; Slipchenko, L.V.; Krylov, A.I. Noncovalent Interactions in Extended Systems Described by the Effective Fragment Potential Method: Theory and Application to Nucleobase Oligomers. *J. Phys. Chem. A* **2010**, *114*, 12739-12754.

197 Freitag, M.A.; Gordon, M.S.; Jensen, J.H.; Stevens, W.J. Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000**, *112*, 7300-7306.

198 Slipchenko, L.V.; Gordon, M.S. Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer. *J. Comput. Chem.* **2007**, *28*, 276-291.

199 Slipchenko, L.V.; Gordon, M.S. Damping functions in the effective fragment potential method. *Mol. Phys.* **2009**, *107*, 999-1016.

200 Kairys, V.; Jensen, J.H. Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chem. Phys. Lett.* **1999**, *315*, 140-144.

201 Adamovic, I.; Gordon, M.S. Dynamic polarizability, dispersion coefficient C-6 and dispersion energy in the effective fragment potential method. *Mol. Phys.* **2005**, *103*, 379-387.

202 Tang, K.T.; Toennies, J.P. An Improved Simple-Model for the Vanderwaals Potential Based on Universal Damping Functions for the Dispersion Coefficients. *J. Chem. Phys.* **1984**, *80*, 3726-3741.

203 Jensen, J.H. Modeling intermolecular exchange integrals between nonorthogonal molecular orbitals. *J. Chem. Phys.* **1996**, *104*, 7795-7796.

204 Jensen, J.H.; Gordon, M.S. An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. *Mol. Phys.* **1996**, *89*, 1313-1325.

www.manaraa.com

205 Jensen, J.H.; Gordon, M.S. An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. II. Application to the effective fragment potential method. *J. Chem. Phys.* **1998**, *108*, 4772-4782.

206 Jensen, J.H. Intermolecular exchange-induction and charge transfer: Derivation of approximate formulas using nonorthogonal localized molecular orbitals. *J. Chem. Phys.* **2001**, *114*, 8775-8783.

207 Li, H.; Gordon, M.S.; Jensen, J.H. Charge transfer interaction in the effective fragment potential method. *J. Chem. Phys.* **2006**, *124*, 214108.

208 Li, H.; Netzloff, H.M.; Gordon, M.S. Gradients of the polarization energy in the effective fragment potential method. *J. Chem. Phys.* **2006**, *125*, 194103.

209 Li, H.; Gordon, M.S. Gradients of the exchange-repulsion energy in the general effective fragment potential method. *Theo. Chem. Acc.* **2006**, *115*, 385-390.

210 Bandyopadhyay, P. Accelerating quantum mechanical/molecular mechanical sampling using pure molecular mechanical potential as an importance function: The case of effective fragment potential. *J. Chem. Phys.* **2005**, *122*, 091102.

211 Kemp, D.; Rintelman, J.; Gordon, M.; Jensen, J. Exchange repulsion between effective fragment potentials and ab initio molecules. *Theor. Chem. Acc. (Theor. Chim. Acta)*, **2010**, *125*, 481-491.

212 Kairys, V.; Jensen, J.H. QM/MM boundaries across covalent bonds: A frozen localized molecular orbital-based approach for the effective fragment potential method. *J. Phys. Chem. A* **2000**, *104*, 6656-6665.

213 Assfeld, X.; Rivail, J.-L. Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method. *Chem. Phys. Lett.* **1996**, *263*, 100-106.

214 Nemukhin, A.V.; Grigorenko, B.L.; Topol, I.A.; Burt, S.K. Flexible effective fragment QM/MM method: Validation through the challenging tests. *J. Comput. Chem.* **2003**, *24*, 1410-1420.

215 Grigorenko, B.L.; Nemukhin, A.V.; Topol, I.A.; Burt, S.K. Modeling of biomolecular systems with the quantum mechanical and molecular mechanical method based on the effective fragment potential technique: Proposal of flexible fragments. *J. Phys. Chem. A*, **2002**, *106*, 10663-10672.

216 Xie, W.; Gao, J. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comp.* **2007**, *3*, 1890-1900.

217 Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102*, 4714-4721.

218 *Pure Appl. Chem.* **1974**, *40*, 291-308.

219 Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. *J. Chem. Phys.* **2008**, *128*, 234108.

220 Song, L.; Han, J.; Lin, Y.; Xie, W.; Gao, J. Explicit Polarization (X-Pol) Potential Using ab Initio Molecular Orbital Theory and Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 11656-11664.

221 Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. On the Interfragment Exchange in the X-Pol Method. *J. Chem. Theory Comput.* **2010**, *6*, 2469-2476.

www.manaraa.com

222 Matsunaga, N.; Chaban, G. M.; Gerber, R. B. Degenerate perturbation theory corrections for the vibrational self-consistent field approximation: Method and applications. *J. Chem. Phys.* **2002**, *117*, 3541-3547.

223 Fedorov, D. G.; Kitaura, K.; Theoretical development of the fragment molecular orbital (FMO) method, in "Modern methods for theoretical physical chemistry of biopolymers", E. B. Starikov, J. P. Lewis, Tanaka, S.; Eds., pp 3-38, Elsevier, Amsterdam, 2006.

224 Nakano, T.; Mochizuki, Y.; Fukuzawa, K.; Amari, S.; Tanaka, S. Developments and applications of ABINIT-MP software based on the fragment molecular orbital method, in "Modern methods for theoretical physical chemistry of biopolymers", E. B. Starikov, J. P. Lewis, Tanaka, S.; Eds., pp 39-52, Elsevier, Amsterdam,

225 Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2007**, *111*, 6904-6914.

226 Fedorov, D. G.; Kitaura, K.; Eds., The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems, CRC Press, Boca Raton, FL, 2009.

227 Massa, L. Book Review. *Int. J. Quantum Chem.*, in press. DOI: 10.1002/qua.22892

228 Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. Full Electron Calculation Beyond 20,000 Atoms: Ground Electronic State of Photosynthetic Proteins. *Proc. of Supercomputing 2005*, IEEE Computer Society, 2005.

229 Fujita, T.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Accuracy of fragmentation in ab initio calculations of hydrated sodium cation. *Chem. Phys. Lett.* **2009**, *478*, 295-300.

230 Nakano, T.; Kaminuma, T.: Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: application to polypeptides, *Chem. Phys. Lett.* **2000**, *318*, 614.

231 Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. Covalent Bond Fragmentation Suitable To Describe Solids in the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2008**, *112*, 11808-11816.

232 Klobukowski, M.; Huzinaga, S.; Sakai, Y. Model core potentials: Theory and aplications, in Computational Chemistry: Reviews of Current Trends, edited by J. Leszczynski, volume 3, World Scientific, Singapore, 1999.

(233) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.

234 Ishikawa, T.; Mochizuki, Y.; Imamura, K.; Nakano, T.; Mori, H.; Tokiwa, H.; Tanaka, K.; Miyoshi, E.; Tanaka, S. Application of fragment molecular orbital scheme to silicon-containing systems, *Chem. Phys. Lett.* **2006**, *430*, 361-366.

235 Yasuda, K.; Yamaki, D. The extension of the fragment molecular orbital method with the many-particle Green's function. *J. Chem. Phys.* **2006**, *125*, 154101.

236 Fedorov, D. G.; Kitaura, K. The role of the exchange in the embedding electrostatic potential for the fragment molecular orbital method. *J. Chem. Phys.* **2009**, *131*, 171106.

237 Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic Study of the Embedding Potential Description in the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2010**, *114*, 8742-8753.

238 Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: use of approximate electrostatic potential, *Chem. Phys. Lett.* **2002**, *351*, 475.

www.manaraa.com

239 Fedorov, D. G.; Kitaura, K. The three-body fragment molecular orbital method for accurate calculations of large systems. *Chem. Phys. Lett.* **2006**, *433*, 182-187.

240 Sekino, H.; Y. Sengoku, Sugiki, S.-I. Kurita, N. Molecular orbital analysis based on fragment molecular orbital scheme, *Chem. Phys. Lett.* **2003**, *378*, 589.

241 K. Ruedenberg, *J. Chem. Phys.* **1951**, *19*, 1433.

242 Companion, A. L.; Parr, R. G. Remark on the Mulliken Approximation for Two-Center Electron Distributions. *J. Chem. Phys.* **1961**, *35*, 2268.

243 Kitaura,K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. Fragment molecular orbital method: analytical energy gradients *Chem. Phys. Lett.* **2001**, *336*, 163.

244 Nagata, T.; Fedorov, D. G.; Kitaura, K.; Derivatives of the approximated electrostatic potentials in the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *475*, 124-131.

245 Nagata, T.; Fedorov, D. G.; Kitaura, K. Importance of the hybrid orbital operator derivative term for the energy gradient in the fragment molecular orbital method. *Chem. Phys. Lett.* **2010**, *492*, 302-308.

246 Komeiji, Y.; Mochizuki, Y.; Nakano, T. Three-body expansion and generalized dynamic fragmentation improve the fragment molecular orbital-based molecular dynamics (FMO-MD). *Chem. Phys. Lett.* **2010**, *484*, 380-386.

247 Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. Analytic gradient for the adaptive frozen orbital bond detachment in the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *477*, 169-175.

248 Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.*, in press.

249 Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. Definition of molecular orbitals in fragment molecular orbital method, *Chem. Phys. Lett.* **2002**, *364*, 139**.**

250 Watanabe, T.; Inadomi, Y.; Umeda, H.; Fukuzawa, K.; Tanaka, S.; Nakano, T.; Nagashima, U. Fragment Molecular Orbital (FMO) and FMO-MO Calculations of DNA: Accuracy Validation of Energy and Interfragment Interaction Energy. *J. Comput. Theor. Nanosc.* **2009**, *6*, 1328-1337.

251 Umeda, H.; Inadomi, Y.; Watanabe, T.; Yagi, T.; Ishimoto, T.; Ikegami, T.; Tadano, H.; Sakurai, T.; Nagashima, U. Parallel Fock matrix construction with distributed shared memory model for the FMO-MO method. *J. Comput. Chem.* **2010**, *31*, 2381-2388.

252 Tsuneyuki, S.; Kobori, T.; Akagi, K.; Sodeyama, K.; Terakura, K.; Fukuyama, H. Molecular orbital calculation of biomolecules with fragment molecular orbitals. *Chem. Phys. Lett.* **2009**, *476*, 104-108.

253 Fedorov, D. G.; Kitaura, K. On the accuracy of the 3-body fragment molecular orbital method (FMO) applied to density functional theory, *Chem. Phys. Lett.* **2004**, *389*, 129-134.

254 Shimodo, Y.; Morihashi, K.; Nakano, T. Examination of numerical accuracy on fragment-DFT calculations with integral values of total electron density functions, *J. Mol. Str. (THEOCHEM)* **2006**, *770*, 163-168.

255 Fedorov, D. G., Kitaura, K. Second order Møller-Plesset perturbation theory based upon the fragment molecular orbital method, *J. Chem. Phys.* **2004,** *121*, 2483-2490.

www.manaraa.com

256 Mochizuki, Y.; Koikegami, S.; Nakano, T.; Amari, S.; Kitaura, K. Large scale MP2 calculations with fragment molecular orbital scheme. *Chem. Phys. Lett.* **2004**, *396*, 473-479.

257 Mochizuki, Y.; Nakano, T.; Koikegami, S.; Tanimori, S.; Abe, Y.; Nagashima, U.; Kitaura, K. A parallelized integral-direct second-order Møller-Plesset perturbation theory method with a fragment molecular orbital scheme, *Theor. Chem. Acc.* **2004**, *112*, 442-452.

258 Ishikawa, T.; Kuwata, K. Fragment molecular orbital calculation using the RI-MP2 method. *Chem. Phys. Lett.* **2009**, *474*, 195-198.

259 Okiyama, Y.; Nakano, T.; Yamashita, K.; Mochizuki, Y.; Taguchi, N.; Tanaka, S. Acceleration of fragment molecular orbital calculations with Cholesky decomposition approach. *Chem. Phys. Lett.* **2010**, 490, 84-89.

260 Mochizuki, Y.; Yamashita, K.; Murase, T.; Nakano, T.; Fukuzawa, K.; Takematsu, K.; Watanabe, H.; Tanaka, S. Large scale FMO-MP2 calculations on a massively parallel-vector computer. *Chem. Phys. Lett.* **2008**, *457*, 396-403.

261 Mochizuki, Y.; K. Yamashita, Fukuzawa, K.; K. Takematsu, H. Watanabe, N. Taguchi, Y. Okiyama, M. Tsuboi, Nakano, T.; Tanaka, S. Large-scale FMO-MP3 calculations on the surface proteins of influenza virus, hemagglutinin (HA) and neuraminidase (NA). *Chem. Phys. Lett.* **2010**, *493*, 346-352.

262 Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. Accuracy of the three-body fragment molecular orbital method applied to Møller-Plesset perturbation theory. *J. Comput. Chem.* **2007**, *28*, 1476-1484.

263 Mochizuki, Y.; Nakano, T.; Komeiji, Y.; Yamashita, K.; Okiyama, Y.; Yoshikawa, H.; Yamataka, H. Fragment molecular orbital-based molecular dynamics (FMO-MD) method with MP2 gradient. *Chem. Phys. Lett.* **2011**, *504*, 95-99.

264 Komeiji, Y.; Nakano, T.; Fukuzawa, K.; Ueno, Y.; Inadomi, Y.; Nemoto, T.; Uebayasi, M.; Fedorov, D. G.; Kitaura, K. Fragment molecular orbital method: application to molecular dynamics simulation, 'ab initio FMO-MD', *Chem. Phys. Lett.* **2003**, *372*, 342.

265 Ishimoto, T.; Tokiwa, H.; Teramae, H.; Nagashima, U. Development of an ab initio MO-MD program based on fragment MO method: an attempt to analyze the fluctuation of protein, *Chem. Phys. Lett.* **2004**, *387*, 460-465.

266 Ishimoto, T.; Tokiwa, H.; Teramae, H.; Nagashima, U. Theoretical study of intramolecular interaction energies during dynamics simulations of oligopeptides by the fragment molecular orbital-Hamiltonian algorithm method, *J. Chem. Phys.* **2005**, *122*, 094905.

267 Komeiji, Y.; Ishikawa, T.; Mochizuki, Y.; Yamataka, H.; Nakano, T. Fragment Molecular Orbital method-based Molecular Dynamics (FMO-MD) as a simulator for chemical reactions in explicit solvation. *J. Comput. Chem.* **2009**, *30*, 40-50.

268 Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Fedorov, D. G. Fragment molecular orbital-based molecular dynamics (FMO-MD), a quantum simulation tool for large molecular systems. *J. Mol. Str. (THEOCHEM)* **2009**, *898*, 2-7.

269 Fujita, T.; Watanabe, H.; Tanaka, S. Ab initio path integral molecular dynamics based on fragment molecular orbital method. *J. Phys. Soc. Jap.* **2009**, *78*, 104723.

www.manaraa.com

270 Fujita, T.; Nakano, T.; Tanaka, S. Fragment molecular orbital calculations under periodic boundary condition. *Chem. Phys. Lett.*, in press.

271 Fedorov, D. G.; Kitaura, K. Multiconfiguration self-consistent-field theory based upon the fragment molecular orbital method, *J. Chem. Phys.* **2005**, *122*, 054108.

272 Fedorov, D. G.; Kitaura, K. Coupled-cluster theory based upon the fragment molecular-orbital method. *J. Chem. Phys.* **2005**, *123*, 134103.

273 Fedorov, D. G.; Ishida, T.; Kitaura, K. Multilayer Formulation of the Fragment Molecular Orbital Method (FMO). *J. Phys. Chem. A* **2005**, *109*, 2638-2646.

274 Maezono, R.; Watanabe, H.; Tanaka, S. Ab initio biomolecular calculations using quantum Monte Carlo combined with the fragment molecular orbital method. in Advances in Quantum Monte Carlo, J. B. Anderson, Rothstein, S. M.; Eds., ACS Symposium Series 953, pp 141-146, American Chemical Society: Washington, DC, 2006.

275 Maezono, R.; Watanabe, H.; Tanaka, S.; Towler, M. D.; Needs, R. J. Fragmentation Method Combined with Quantum Monte Carlo Calculations. *J. Phys. Soc. Jap.* **2007**, *76*, 064301.

276 Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (FMO), *J. Comput. Chem.* **2006**, *27*, 976-985.

277 Li, H.; Fedorov, D. G.; Nagata, T.; Kitaura, K.; Jensen, J. H.; Gordon, M. S. Energy gradients in combined fragment molecular orbital and polarizable continuum model (FMO/PCM) calculation. *J. Comput. Chem.* **2010**, *31*, 778-790.

278 Watanabe, H.; Okiyama, Y.; Nakano, T.; Tanaka, S. Incorporation of solvation effects into the fragment molecular orbital calculations with the Poisson-Boltzmann equation. *Chem. Phys. Lett.* **2010**, *500*, 116-119.

279 Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. Configuration interaction singles method with multilayer fragment molecular orbital scheme, *Chem. Phys. Lett.* **2005**, *406*, 283-288.

280 Mochizuki, Y.; Tanaka, K.; Yamashita, K.; Ishikawa, T.; Nakano, T.; Amari, S.; Segawa, K.; Murase, T.; Tokiwa, H.; Sakurai, M.; Parallelized integral-direct CIS(D) calculations with multilayer fragment molecular orbital scheme. *Theor. Chem. Acc.* **2007**, *117*, 541-553.

281 Chiba, M.; Fedorov, D. G.; Kitaura, K. Time-dependent density functional theory with the multilayer fragment molecular orbital method. *Chem. Phys. Lett.* **2007**, *444*, 346-350.

282 Chiba, M.; Fedorov, D. G.; Kitaura, K. Time-dependent density functional theory based upon the fragment molecular orbital method. *J. Chem. Phys.* **2007**, *127*, 104108.

283 Chiba, M.; Fedorov, D. G.; Kitaura, K. Polarizable continuum model with the fragment molecular orbital-based time-dependent density functional theory. *J. Comput. Chem.* **2008**, *29*, 2667-2676.

284 Chiba, M.; Fedorov, D. G.; Nagata, T.; Kitaura, K. Excited state geometry optimizations by time-dependent density functional theory based on the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *474*, 227-232.

285 Chiba, M.; Koido, T. Electronic excitation energy calculation by the fragment molecular orbital method with three-body effects. *J. Chem. Phys.* **2010**, *133*, 044113.

www.manaraa.com

286 Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Open-Shell Formulation of the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2010**, *6*, 1-5.

287 Ishimoto, T.; Tachikawa, M.; Nagashima, U. A fragment molecular-orbital-multicomponent molecular-orbital method for analyzing H/D isotope effects in large molecules, *J. Chem. Phys.* **2006**, *124*, 014112.

288 Auer, B.; Pak, M. V.; Hammes-Schiffer, S. Nuclear-Electronic Orbital Method within the Fragment Molecular Orbital Approach. *J. Phys. Chem. C* **2010**, *114*, 5582-5588.

289 Mochizuki, Y.; Ishikawa, T.; Tanaka, K.; Tokiwa, H.; Nakano, T.; Tanaka, S. Dynamic polarizability calculation with fragment molecular orbital scheme. *Chem. Phys. Lett.* **2006**, *418*, 418-422.

290 Ishikawa, T.; Mochizuki, Y.; Nakano, T.; Amari, S.; Mori, H.; Honda, H.; Fujita, T.; Tokiwa, H.; Tanaka, S.; Komeiji, Y.; Fukuzawa, K.; Tanaka, K.; Miyoshi, E. Fragment molecular orbital calculations on large scale systems containing heavy metal atom. *Chem. Phys. Lett.* **2006**, *427*, 159-165.

291 Okiyama, Y.; Watanabe, H.; Fukuzawa, K.; Nakano, T.; Mochizuki, Y.; Ishikawa, T.; Tanaka, S.; Ebina, K. Application of the fragment molecular orbital method for determination of atomic charges on polypeptides. *Chem. Phys. Lett.* **2007**, *449*, 329-335.

292 Okiyama, Y.; Watanabe, H.; Fukuzawa, K.; Nakano, T.; Mochizuki, Y.; Ishikawa, T.; Ebina, K.; Tanaka, S. Application of the fragment molecular orbital method for determination of atomic charges on polypeptides. II. Towards an improvement of force fields used for classical molecular dynamics simulations. *Chem. Phys. Lett.* **2009**, *467*, 417-423.

293 Sekino, H.; Matsumura, N.; Sengoku, Y. Evaluation of NMR Chemical Shift by Fragment Molecular Orbital Method. *Comput. Lett.* **2007**, *3*, 423-430.

294 Gao, Q.; Yokojima, S.; Kohno, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Fujihira, M.; Nakamura, S. Ab initio NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment. *Chem. Phys. Lett.* **2007**, *445*, 331-339.

295 Q. Gao, S. Yokojima, Fedorov, D. G.; Kitaura, K.; M. Sakurai, S. Nakamura, Fragment-Molecular-Orbital-Method-Based ab Initio NMR Chemical-Shift Calculations for Large Molecular Systems. *J. Chem. Theory Comput.* **2010**, *6*, 1428-1444.

296 Amari, S.; Aizawa, M.; Zhang, J.; Fukuzawa, K.; Mochizuki, Y.; Iwasawa, Y.; Nakata, K.; Chuman, H.; Nakano, T. VISCANA: Visualized Cluster Analysis of Protein-Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening, *J. Chem. Inf. Comput. Sci.* **2006**, *46*, 221-230.

297 Du, S.; Sakurai, M. Multivariate analysis of properties of amino acid residues in proteins from a viewpoint of functional site prediction. *Chem. Phys. Lett.* **2010**, *488*, 81-85.

298 Mochizuki, Y.; Fukuzawa, K.; Kato, A.; Tanaka, S.; Kitaura, K.; Nakano, T. A configuration analysis for fragment interaction, *Chem. Phys. Lett.* **2005**, *410*, 247-253.

299 Ishikawa, T.; Mochizuki, Y.; Amari, S.; Nakano, T.; Tokiwa, H.; Tanaka, S.; Tanaka, K.Fragment interaction analysis based on local MP2. *Theor. Chem. Acc.* **2007**, *118*, 937-945.

300 Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 533.

www.manaraa.com

301 Ishikawa, T.; Ishikura, T.; Kuwata, K. Theoretical study of the prion protein based on the fragment molecular orbital method. *J. Comput. Chem.* **2009**, *30*, 2594-2601.

302 Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418; **1986**, *84*, 4113; **1987**, *86*, 3760.

303 Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. The Fragment Molecular Orbital Method for Geometry Optimizations of Polypeptides and Proteins. *J. Phys. Chem. A* **2007**, *111*, 2722-2732.

304 Ishikawa, T.; Yamamoto, N.; Kuwata, K. Partial energy gradient based on the fragment molecular orbital method: Application to geometry optimization. *Chem. Phys. Lett.* **2010**, *500*, 149-154.

305 Fedorov, D. G.; Alexeev, Y.; Kitaura, K. Geometry Optimization of the Active Site of a Large System with the Fragment Molecular Orbital Method. *J. Phys. Chem. Lett.* **2011**, *2*, 282-288.

306 Dahlke, E.E., Truhlar, D.G. Electrostatically embedded many-body expansion for simulations. *J. Chem. Theor. Comput.* **2008**, *4*, 1-6.

307 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. Theor. Comput.* **2008**, *4*, 33-41.

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

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

310 Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.

311 Kamiya, M.; Hirata, S.; Valiev, M. J. *Chem. Phys.* **2008**, *128*, 074103.

312 Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **2010**, *114*, 8705-8712.

313 Söderhjelm, P.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 617.

314 Söderhjelm, P.; Aquilante, F.; Ryde, U., Calculation of Protein-Ligand Interaction Energies by a Fragmentation Approach Combining High-Level Quantum Chemistry with Classical Many-Body Effects. *J. Phys. Chem. B* **2009**, *113*, 11085-11094.

315 Söderhjelm, P.; Kongsted, J.; Ryde, U., Ligand Affinities Estimated by Quantum Chemical Calculations. *J. Chem. Theory .Comput.* **2010**, *6*, 1726-1737.

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

317 Sebetci, A.; Beran, G.J.O. *J. Chem. Theory Comput.* **2010**, *6*, 155–167.

318 Babu, K.; Ganesh, V.; Gadre, S. R.; Ghermani, N. E. Tailoring approach for exploring electron densities and electrostatic potentials of molecular crystals. *Theor. Chem. Acc.* **2004**, *111*, 255-263.

319 Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.* **2006**, *125*, 104109.

www.manaraa.com

320 Gadre, S. R.; Ganesh, V. Molecular tailoring approach: Towards PC-based ab initio treatment of large molecules. *J. Theor. Comput. Chem.* **2006**, *5*, 835-855.

321 Deshmukh, M. M.; Gadre, S. R.; Bartolotti, L. J. Estimation of intramolecular hydrogen bond energy via molecular tailoring approach. *J. Phys. Chem. A* **2006**, *110*, 12519-12523.

322 Deshmukh, M. M.; Suresh, C. H.; Gadre, S. R. Intramolecular hydrogen bond energy in polyhydroxy systems: A critical comparison of molecular tailoring and isodesmic approaches. *J. Phys. Chem. A* **2007**, *111*, 6472-6480.

323 Elango, M.; Subramanian, V.; Rahalkar, A. P.; Gadre, S. R.; Sathyamurthy, N. Structure, energetics, and reactivity of boric acid nanotubes: A molecular tailoring approach. *J. Phys. Chem. A* **2008**, *112*, 7699-7704.

324 Gadre, S. R.; Jose, K. V. J.; Rahalkar, A. P. Molecular tailoring approach for exploring structures, energetics and properties of clusters. *J. Chem. Sciences* **2010**, *122*, 47-56.

325 Yeole, S. D.; Gadre, S. R. On the applicability of fragmentation methods to conjugated p systems within density functional framework. *J. Chem. Phys.* **2010**, *132*, 094102.

326 Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. Molecular Tailoring Approach in Conjunction with MP2 and RI-MP2 Codes: A Comparison with Fragment Molecular Orbital Method. *J. Comput. Chem.* **2010**, *31*, 2405-2418.

327 Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. Enabling ab initio Hessian and frequency calculations of large molecules. *J. Chem. Phys.* **2008**, *129*, 234101.

328 Huang, L. L.; Massa, L.; Karle, J. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808-817.

329 Huang, L. L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochemistry* **2005**, *44*, 16747-16752.

330 Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to insulin. *Proc. Nat. Acad. Sciences U.S.A.* **2005**, *102*, 12690-12693.

331 Huang, L.; Massa, L.; Karle, J. The Kernel Energy Method: Application to a tRNA. *Proc. Nat. Acad. Sciences U.S.A.* **2006**, *103*, 1233-1237.

332 Huang, L.; Massa, L.; Karle, J. Kernel energy method: The interaction energy of the collagen triple helix. *J. Chem. Theory Comp.* **2007**, *3*, 1337-1341.

333 Huang, L. L.; Massa, L.; Karle, J. Kernel energy method: Basis functions and quantum methods. *Int. J. Quantum Chem.* **2006**, *106*, 447-457.

334 Huang, L.; Massa, L.; Karle, J. The kernel energy method of quantum mechanical approximation carried to fourth-order terms. *Proc. Nat. Acad. Sciences U.S.A.* **2008**, *105*, 1849-1854.

335 Huang, L.; Bohorquez, H. J.; Matta, C. F.; Massa, L. The Kernel energy method: Application to graphene and extended aromatics. *Int. J. Quantum Chem., in press. DOI: 10.1002/qua.22975*

336 Weiss, S. N.; Huang, L.; Massa, L. A generalized higher order kernel energy approximation method. *J. Comput. Chem.* **2010**, *31*, 2889–2899.

www.manaraa.com

337 Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599-3605.

338 Zhang, D. W.; Chen, X. H.; Zhang, J. Z. H. Molecular caps for full quantum mechanical computation of peptide-water interaction energy. *J. Comput. Chem.* **2003**, *24*, 1846-1852.

339 Zhang, D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H. Quantum mechanical map for protein-ligand binding with application to beta-trypsin/benzamidine complex. *J. Chem. Phys.* **2004**, *120*, 1145-1148.

340 Zhang, D. W.; Zhang, J. Z. H. Full ab initio computation of protein-water interaction energies. *J. Theor. Comput. Chem.* **2004**, *3*, 43-49.

341 Chen, X. H.; Zhang, J. Z. H. MFCC-downhill simplex method for molecular structure optimization. *J. Theor. Comput. Chem.* **2004**, *3*, 277-289.

342 Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. New method for direct linear-scaling calculation of electron density of proteins. *J. Phys. Chem. A* **2005**, *109*, 2-5.

343 Li, W.; Fang, T.; Li, S. A fragment energy assembler method for Hartree-Fock calculations of large molecules, *J. Chem. Phys.* **2006**, 124, 154102.

344 Mei, Y.; Wu, E. L.; Han, K. L.; Zhang, J. Z. H. Treating hydrogen bonding in ab initio calculation of biopolymers. *Int. J. Quantum Chem.* **2006**, *106*, 1267-1276.

345 Chen, X. H.; Zhang, J. Z. H. Theoretical method for full ab initio calculation of DNA/RNA-ligand interaction energy. *J. Chem. Phys.* **2004**, *120*, 11386-11391.

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

347 Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. Fractionation of peptide with disulfide bond for quantum mechanical calculation of interaction energy with molecules. *J. Chem. Phys.* **2004**, *120*, 839-844.

348 Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. K., An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* **2004**, *394*, 293-297.

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

350 Li, S. H.; Li, W.; Fang, T. An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7215-7226.

351 Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Fully quantum mechanical energy optimization for protein-ligand structure. *J. Comput. Chem.* **2004**, *25*, 1431-1437.

352 He, X.; Zhang, J. Z. H. The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy. *J. Chem. Phys.* **2006**, 124, 184703.

353 Mei, Y.; Ji, C. G.; Zhang, J. Z. H. A new quantum method for electrostatic solvation energy of protein. *J. Chem. Phys.* **2006**, *125*, 094906.

354 Chen, X. H.; 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.

www.manaraa.com

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

356 Foster, J.P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211

357 Reed, A.E.; Weinstock, R.B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735

358 Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. Geometry Optimizations and Vibrational Spectra of Large Molecules from a Generalized Energy-Based Fragmentation Approach. *J. Phys. Chem. A* **2008**, *112*, 10864-10872.

359 Hua, S. G.; Hua, W. J.; Li, S. H. An Efficient Implementation of the Generalized Energy-Based Fragmentation Approach for General Large Molecules. *J. Phys. Chem. A* **2010**, *114*, 8126-8134.

360 Jacob, C. R., Visscher, L. A subsystem density-functional theory approach for the quantum chemical treatment of proteins. *J. Chem. Phys.* **2008**, 128, 155102.

361 He, J.; Di Paola, C.; Kantorovich, L. Partitioning scheme for density functional calculations of extended systems, *J. Chem. Phys.* **2009**, *130*, 144104.

362 Řezáč, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *6*, 91-99.

363 Ji, C. G.; Mei, Y.; Zhang, J. Z. H. *Biophys. J.* **2008**, *95*, 1080.

364 Ji, C. G.; Zhang, J. Z. H. Protein Polarization Is Critical to Stabilizing AF-2 and Helix-2' Domains in Ligand Binding to PPAR-g. *J. Am. Chem. Soc.* **2008**, *130*, 17129.

365 Duan, L. L.; Mei, Y.; Zhang, Q. G.; Zhang, J. Z. H. Intra-protein hydrogen bonding is dynamically stabilized by electronic polarization. *J. Chem. Phys.* **2009**, *130*, 115102.

366 Ji, C. G.; Zhang, J. Z. H. NMR Scalar Coupling Constant Reveals That Intraprotein Hydrogen Bonds Are Dynamically Stabilized by Electronic Polarization. *J. Phys. Chem. B* **2009**, *113*, 13898–13900.

367 Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Simulation of NMR Data Reveals That Proteins' Local Structures Are Stabilized by Electronic Polarization. *J. Am. Chem. Soc.* **2009**, *131*, 8636–8641.

368 Ji, C. G.; Zhang, J. Z. H. Electronic Polarization Is Important in Stabilizing the Native Structures of Proteins. *J. Phys. Chem. B* **2009**, *113*, 16059–16064.

369 Lu, Y.; Mei, Y.; Zhang, J. Z. H.; Zhang, D. Electron polarization critically stabilizes the Mg2+ complex in the catalytic core domain of HIV-1 integrase. *J. Chem. Phys.* **2010**, *132*, 131101.

370 Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang John, Z. H. Electrostatic Polarization Makes a Substantial Contribution to the Free Energy of Avidin−Biotin Binding. *J. Am. Chem. Soc.* **2010**, *132*, 5137–5142.

371 Duan, L. L.; Mei, Y.; Zhang, D.; Zhang, Q. G.; Zhang, J. Z. H. Folding of a Helix at Room Temperature Is Critically Aided by Electrostatic Polarization of Intraprotein Hydrogen Bonds. *J. Am. Chem. Soc.* **2010**, *132*, 11159–11164.

372 Netzloff, H. M.; Collins, M.A. *J. Chem. Phys.* **2007**, *127*, 134113.

www.manaraa.com

373 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.

374 Mullin, J. M.; Roskop, L. B.; Pruitt, S. R.; Collins, M. A.; Gordon, M. S. *J. Phys. Chem. A* **2009**, *113*, 10040.

375 Collins, M. A. Molecular potential energy surfaces constructed from interpolation of systematic fragment surfaces. *J. Chem. Phys.* **2007**, 127, 024104.

376 Xantheas, S. S.; Dunning Jr., T. H. *J. Chem. Phys.* **1993**, *98*, 8037.

377 Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.

378 Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163.

379 Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225.

380 Xantheas, S. S. *Struct. Bonding* **2005**, *116*, 119.

381 Yang, W.; Lee, T.S. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *J. Chem. Phys.* **1995**, *103*, 5674.

382 Lee, T.S.; York, D. M.; Yang, W. Linear-scaling semiempirical quantum calculations for macromolecules. *J. Chem. Phys.*, **1996**, *105*, 2744.

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

384 He, X.; Merz, K. M. Jr. *J. Chem. Theory Comput.* **2010**, *6*, 405.

385 Dixon, S. L.; Merz. K.M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.*, **1996**, *104*, 6643.

386 Dixon, S. L.; Merz, K. M. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.*, **1997**, *107*, 879.

387 Li, W.; Li, S. Divide-and-conquer local correlation approach to the correlation energy of large molecules. *J. Chem. Phys.* **2004**, *121*, 6649.

388 Kobayashi, M.; Akama, T.; Nakai, H. Second-order Moller-Plesset perturbation energy obtained from divide-and-conquer Hartree-Fock density matrix. *J. Chem. Phys.* **2006**, *125*, 204106.

389 Akama, T.; Fujii, A.; Kobayashi, M.; Nakai, H. Is the divide-and-conquer Hartree-Fock method valid for calculations of delocalized systems? *Mol. Phys.* **2007**, *105*, 2799-2804.

390 Akama, T.; Kobayashi, M.; Nakai, H. Implementation of divide-and-conquer method including Hartree-Fock exchange interaction. *J. Comput. Chem.* **2007**, *28*, 2003-2012.

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

www.manaraa.com

392 Kobayashi, M.; Nakai, H. Extension of linear-scaling divide-and-conquer-based correlation method to coupled cluster theory with singles and doubles excitations. *J. Chem. Phys.* **2008**, *129*, 044103.

393 Akama, T.; Kobayashi, M.; Nakai, H. Electronic Temperature in Divide-and-Conquer Electronic Structure Calculation Revisited: Assessment and Improvement of Self-Consistent Field Convergence. *Int. J. Quantum Chem.* **2009**, *109*, 2706-2713.

394 Kobayashi, M.; Nakai, H. Divide-and-conquer-based linear-scaling approach for traditional and renormalized coupled cluster methods with single, double, and noniterative triple excitations. *J. Chem. Phys.* **2009**, *131*, 114108.

395 Kobayashi, M.; Nakai, H. Dual-Level Hierarchical Scheme for Linear-Scaling Divide-and-Conquer Correlation Theory. *Int. J. Quantum Chem.* **2009**, *109*, 2227-2237.

396 Touma, T.; Kobayashi, M.; Nakai, H. Time-dependent Hartree-Fock frequency-dependent polarizability calculation applied to divide-and-conquer electronic structure method. *Chem. Phys. Lett.* **2010**, *485*, 247-252.

397 Kobayashi, M.; Kunisada, T.; Akama, T.; Sakura, D.; Nakai, H. Reconsidering an analytical gradient expression within a divide-and-conquer self-consistent field approach: Exact formula and its approximate treatment. *J. Chem. Phys.* **2011**, *134*, 034105.

398 Song G.; Li, Z; Liu Z.; Cao, X.; Wang, W.; Fan, K.; Xie, Y.; Schaefer, H. F. III Local Hybrid Divide-and-Conquer Method for the Computation of Medium and Large Molecules. *J. Chem. Theory Comput.* **2008**, *12*, 2049-2056.

399 Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 024501.

400 E. Fermi, "L'interpretazione del principio di causalit`a nella meccanica quantistica", *Rend. Lincei*, **1930**, *11*, 980-985

401 Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.*, **1993**, *115*, 12423

402 Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.*, **1994**, *116*, 12022

403 Michl, J.; Kaszynski, K.; Friedli, A. C.; McMurdi, N. D.; Kim, T. *NATO ASI Ser., Ser. C*, **1989**, *273*, 469.

404 Mathias, J. P.; Stoddart, J. F. *Chem. Soc. Rev.*, **1992**, 215

405 Mezey, P. G. *J. Math. Chem.*, **1995**, *18*, 141

406 Mezey, P. G., *Computational Chemistry: Reviews and Current Trends*, J. Leszczynski, Ed. (World Scientific, Singapore, 1996), Volume 1, 109

407 Exner, T. E.; Mezey, P. G. Ab initio-quality electrostatic potentials for proteins: An application of the ADMA approach. *J. Phys. Chem. A* **2002**, *106*, 11791-11800.

408 Exner, T. E.; Mezey, P. G. Evaluation of the field-adapted ADMA approach: absolute and relative energies of crambin and derivatives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061-4069.

409 Szekeres, Z.; Mezey, P. G. A one-step, diophantine solution to the density matrix purification problem. *Mol. Phys.* **2005**, *103*, 1013-1015.

www.manaraa.com

410 Eckard, S.; Exner, T. E. Generalized hybrid orbitals in the FA-ADMA method. *Zeitschrift Für Physikalische Chemie-Int. J. Res. Phys.Chem. Chem. Phys.* **2006**, *220*, 927-944.

411 Exner, T. E.; Mezey, P. G. Ab initio quality properties for macromolecules using the ADMA approach. *J. Comput. Chem.* **2003**, *24*, 1980-1986.

412 Exner, T. E.; Mezey, P. G. The field-adapted ADMA approach: Introducing point charges. *J. Phys. Chem. A* **2004**, *108*, 4301-4309.

413 Szekeres, Z.; Exner, T.; Mezey, P. G. Fuzzy fragment selection strategies, basis set dependence and HF-DFT comparisons in the applications of the ADMA method of macromolecular quantum chemistry. *Int. J. Quantum Chem.* **2005**, *104*, 847-860.

414 Szekeres, Z.; Mezey, P. G.; Surjan, P. R., Diagonalization-free initial guess to SCF calculations for large molecules. *Chem. Phys. Lett.* **2006**, *424*, 420-424.

415 Eckard, S.; Exner, T. E. Improvements in the Generalized Hybrid Orbital Method. *Int. J. Quantum Chem.* **2009**, *109*, 1451-1463.

416 Das, G. P.; Yeates, A. T.; Dudis, D. S. *Int. J. Quant. Chem.* **2003**, *92*, 22.

417 Sakai, S.; Morita, S. Ab Initio Integrated Multi-Center Molecular Orbitals Method for Large Cluster Systems: Total Energy and Normal Vibration. *J. Phys. Chem. A* **2005**, *109*, 8424.

418 Mata, R. A.; Stoll, H.; Cabral, B. J. C. *J. Chem. Theory Comput.* **2009**, *5*, 1829.

419 Bettens, R.P.A; Lee, A.M. A New Algorithm for Molecular Fragmentation in Quantum Chemical Calculations. *J. Phys. Chem. A* **2006**, *110*, 8777.

420 Lee, A. M.; Bettens, R. P. A. First Principles NMR Calculations by Fragmentation. *J. Phys. Chem. A* **2007**, *111*, 5111–5115.

421 Le, H.-A.; Lee, A. M.; Bettens, R. P. A. Accurately Reproducing Ab Initio Electrostatic Potentials with Multipoles and Fragmentation. *J. Phys. Chem. A*, **2009**, *113*, 10527.

422 Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796.

423 Wang, L.; Zhao, Z.; Meza, J. Linear-scaling three-dimensional fragment method for large-scale electronic structure calculations. *Phys. Rev. B* **2008**, *77*, 165113.

424 Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

425 Gordon, M. S.; Schmidt, M. W. Title? in *Theory and Applications of Computational Chemistry, the first forty years*, Elsevier, Amsterdam, 2005.

426 Komeiji, Y.; Inadomi, Y.; Nakano, T. PEACH 4 with ABINIT-MP: a general platform for classical and quantum simulations of biological molecules. *Comput. Biol. Chem.* **2004**, *28*, 155-161.

427 RSS21 Project. http://www.ciss.iis.u-tokyo.ac.jp/rss21/result/download/index.php#download_2 (in Japanese).

www.manaraa.com

428 PAICS. http://www.paics.net/index_e.html .

429 Ganesh, V.; Kavathekar, R.; Rahalkar, A.; Gadre, S. R. WebProp: Web interface for ab initio calculation of molecular one-electron properties. *J. Comput. Chem.* **2008**, *29*, 488–495.

430 Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. WebMTA: A Web-Interface for Ab Initio Geometry Optimization of Large Molecules Using Molecular Tailoring Approach. *J. Comput. Chem.* **2009**, *30*, 1167-1173.

431 Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Mod.* **1998**, *16*, 133-138.

http://www.scl.ameslab.gov/~brett/MacMolPlt/

432 Suenaga, M. *J. Comput. Chem. Jpn.* **2008**, *7*, 33 (in Japanese).

http://www1.bbiq.jp/zzzfelis/Facio.html .

433 Schmidt, M. W.; Fletcher, G. D.; Bode, B. M.; Gordon, M. S. The Distributed Data Interface in GAMESS. *Comput. Phys. Comm.* **2000**, *128*, 190.

434 Olson, R. M.; Schmidt, M. W.; Gordon, M. S.; Rendell, A. P. Enabling the Efficient Use of SMP Clusters: The GAMESS/DDI Model. *Proc. Supercomputing*, 2003.

435 G. D. Fletcher et al., *J. Chem. Theor. Comp.*, to be submitted.

436 Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.

437 Sato, M.; Yamataka, H.; Komeiji, Y.; Mochizuki, Y.; Ishikawa, T.; Nakano, T. How Does an SN2 Reaction Take Place in Solution? Full Ab Initio MD Simulations for the Hydrolysis of the Methyl Diazonium Ion. *J. Am. Chem. Soc.* **2008**, *130*, 2396-2397.

438 Sato, M.; Yamataka, H.; Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Does amination of formaldehyde proceed through a zwitterionic intermediate in water? Fragment molecular orbital molecular dynamics simulations by using constraint dynamics. *Chem. Eur. J.* **2010**, *16*, 6430-6433.

439 Pomogaev, V.; Pomogaeva, A.; Aoki, Y., Absorption Spectra of Estradiol and Tryptophan Constructed by the Statistical and Elongation Methods. *J. Phys. Chem. A* **2009**, *113*, 1429-1433.

440 Kistler, K. A.; Matsika, S. Solvatochromic Shifts of Uracil and Cytosine Using a Combined Multireference Configuration Interaction/Molecular Dynamics Approach and the Fragment Molecular Orbital Method. *J. Phys. Chem. A*, **2009**, *113*, 12396-12403.

441 Fujiwara, T.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Mori, H.; Nakano, T.; Miyoshi, E. Fragment molecular orbital-based molecular dynamics (FMO-MD) simulations on hydrated Zn(II) ion. *Chem. Phys. Lett.* **2010**, *490*, 41-45.

442 Bandyopadhyay, P. Assessment of Two Surface Monte Carlo (TSMC) method to find stationary points of (H2O)15 and (H2O)20 clusters. *Theo. Chem. Acc.* **2008**, *120*, 307-312.

443 Kemp, D.A.; Gordon, M.S. An interpretation of the enhancement of the water dipole moment due to the presence of other water molecules. *J. Phys. Chem. A*, **2008**, *112*, 4885-4894.

www.manaraa.com

444 Kemp, D.D.; Gordon, M.S. Theoretical study of the solvation of fluorine and chlorine anions by water. *J. Phys. Chem. A*, **2005**, *109*, 7688-7699.

445 Merrill, G.N.; Webb, S.P. Anion-water clusters A- (H2O)1-6, A = OH, F, SH, Cl, and Br. An effective fragment potential test case. *J. Phys. Chem. A*, **2003**, *107*, 7852-7860.

446 Merrill, G.N.; Webb, S.P. The application of the effective fragment potential method to molecular anion solvation: A study of ten oxyanion-water clusters, A-(H2O)1-4. *J. Phys. Chem. A*, **2004**, *108*, 833-839.

447 Merrill, G.N.; Webb, S.P.; Bivin, D.B. Formation of alkali metal/alkaline earth cation water clusters, M(H2O)1-6 M = Li+, Na+, K+, Mg2+, and Ca2+: An effective fragment potential (EFP) case study. *J. Phys. Chem. A*, **2003**, *107*, 386-396.

448 Chandrakumar, K.R.S.; Ghanty, T.K.; Ghosh, S.K.; Mukherjee, T. Hydration of uranyl cations: Effective fragment potential approach. *J. Mol. Str. (Theochem)*, **2007**, *807*, 93-99.

449 Merrill, G.N.; Fletcher, G.D. A microsolvation approach to the prediction of the relative enthalpies and free energies of hydration for ammonium ions. *Theo. Chem. Acc.* **2008**, *120*, 5-22.

450 Petersen, C.P.; Gordon, M.S. Solvation of sodium chloride: An effective fragment study of NaCl(H2O)n. *J. Phys. Chem. A*, **1999**, *103*, 4162-4166.

451 Yoshikawa, A.; Morales, J.A. The onset of dissociation in the aqueous LiOH clusters: a solvation study with the effective fragment potential model and quantum mechanics methods. *Journal of Molecular Structure-Theochem*, **2004**, *681*, 27-40.

452 Balawender, R.; Safi, B.; Geerlings, P. Solvent effect on the global and atomic DFT-based reactivity descriptors using the effective fragment potential model. Solvation of ammonia. *J. Phys. Chem. A*, **2001**, *105*, 6703-6710.

453 Safi, B.; Balawender, R.; Geerlings, P. Solvent effect on electronegativity, hardness, condensed Fukui functions, and softness, in a large series of diatomic and small polyatomic molecules: Use of the EFP model. *J. Phys. Chem. A*, **2001**, *105*, 11102-11109.

454 Day, P.N.; Pachter, R. A study of aqueous glutamic acid using the effective fragment potential method. *J. Chem. Phys.* **1997**, *107*, 2990-2999.

455 Mullin, J.M.; Gordon, M.S. Alanine: Then There Was Water. *J. Phys. Chem. B*, **2009**, *113*, 8657-8669.

456 Mullin, J.M.; Gordon, M.S. Water and Alanine: From Puddles(32) to Ponds(49). *J. Phys. Chem. B*, **2009**, *113*, 14413-14420.

457 Song, J.; Gordon, M.S.; Deakyne, C.A.; Zheng, W.C. Theoretical investigations of acetylcholine (ACh) and acetylthiocholine (ATCh) using ab initio and effective fragment potential methods. *J. Phys. Chem. A*, **2004**, *108*, 11419-11432.

458 Adamovic, I.; Gordon, M.S. Methanol-water mixtures: A microsolvation study using the effective fragment potential method. *J. Phys. Chem. A*, **2006**, *110*, 10267-10273.

459 Adamovic, I.; Li, H.; Lamm, M.H.; Gordon, M.S. Modeling styrene-styrene interactions. *J. Phys. Chem. A*, **2006**, *110*, 519-525.

www.manaraa.com

460 Slipchenko, L.V.; Gordon, M.S. Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer. *J. Comput. Chem.* **2007**, *28*, 276-291.

461 Smith, T.; Slipchenko, L.V.; Gordon, M.S. Modeling pi-pi interactions with the effective fragment potential method: The benzene dimer and substituents. *J. Phys. Chem. A*, **2008**, *112*, 5286-5294.

462 Smith, Q.A.; Gordon, M.S.; Slipchenko, L.V. Benzene-Pyridine Interactions Predicted by the Effective Fragment Potential Method. *J. Phys. Chem. A*, **2011**, *in press*.

463 Slipchenko, L.V.; Gordon, M.S. Water-Benzene Interactions: An Effective Fragment Potential and Correlated Quantum Chemistry Study. *J. Phys. Chem. A*, **2009**, *113*, 2092-2102.

464 Mohri, F.; Granovsky, A.A. A molecular orbital explanation for the B-N bond shortening in H3BNH3 on going from the gaseous to the solid state. *Int. J. Quant. Chem.* **2008**, *108*, 544-557.

465 Krauss, M. Effective Fragment Potentials and Spectroscopy at Enzyme Active-Sites. *Computers & Chemistry*, **1995**, *19*, 33-38.

466 Krauss, M. Effective Fragment Potentials and Spectroscopy at Enzyme Active-Sites. *Computers & Chemistry*, **1995**, *19*, 199-204.

467 Kina, D.; Arora, P.; Nakayama, A.; Noro, T.; Gordon, M.S.; Taketsugu, T. Ab Initio QM/MM Excited-State Molecular Dynamics Study of Coumarin 151 in Water Solution. *Int. J. Quant. Chem.* **2009**, *109*, 2308-2318.

468 Kina, D.; Nakayama, A.; Noro, T.; Taketsugu, T.; Gordon, M.S. Ab initio QM/MM molecular dynamics study on the excited-state hydrogen transfer of 7-azaindole in water solution. *J. Phys. Chem. A*, **2008**, *112*, 9675-9683.

469 Atadinc, F.; Gunaydin, H.; Ozen, A.S.; Aviyente, V. A quantum mechanical approach to the kinetics of the hydrogen abstraction reaction H2O2+(OH)-O-center dot -> HO2+H2O. *Int. J. Chem. Kinetics*, **2005**, *37*, 502-514.

470 Ferreira, D.E.C.; Florentino, B.P.D.; Rocha, W.R.; Nome, F. Quantum Mechanical/Effective Fragment Potential (QM/EFP) Study of Phosphate Monoester Aminolysis in Aqueous Solution. *J. Phys. Chem. B*, **2009**, *113*, 14831-14836.

471 Hush, N.S.; Schamberger, J.; Bacskay, G.B. A quantum chemical computational study of the relative stabilities of cis- and trans-platinum dichloride in aqueous solution. *Coordination Chemistry Reviews*, **2005**, *249*, 299-311.

472 Nemukhin, A.V.; Topol, I.A.; Grigorenko, B.L.; Burt, S.K. On the origin of potential barrier for the reaction OH-+CO2 -> HCO3- in water: Studies by using continuum and cluster solvation methods. *J. Phys. Chem. B*, **2002**, *106*, 1734-1740.

473 Nemukhin, A.V.; Grigorenko, B.L.; Topol, I.A.; Burt, S.K. QM/MM modeling of the glutathione-hydroxymethyl radical reaction in watery. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1031-1038.

474 Jose, K. V. J.; Gadre, S. R. Electrostatic guidelines and molecular tailoring for density functional investigation of structures and energetics of (Li)n clusters. *J. Chem. Phys.* **2008**, *129*, 164314.

475 Jose, K. V. J.; Gadre, S. R. Ab Initio Study on (CO2)n Clusters via Electrostatics- and Molecular Tailoring-Based Algorithm. *Int. J. Quantum Chem.* **2009**, *109*, 2238-2247.

www.manaraa.com

476 Mahadevi, A. S.; Rahalkar, A.P.; Gadre, S. R.; Sastry. G. N. Ab initio investigation of benzene clusters: Molecular tailoring approach. *J. Chem. Phys.* **2010**, *133*, 164308.

477 Yang, Z.; Hua, S. G.; Hua, W. J.; Li, S. H. Low-Lying Structures and Stabilities of Large Water Clusters: Investigation Based on the Combination of the AMOEBA Potential and Generalized Energy-Based Fragmentation Approach. *J. Phys. Chem. A* **2010**, *114*, 9253-9261.

478 Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical calculations of binding enthalpy for protein–ligand complexes. *Int. J. Quant. Chem.* **2004**, *97*, 747–763.

479 Stewart, J. J. P. Application of the PM6 method to modeling proteins. *J. Mol. Mod.* **2009**, *15*, 765–805.

480 *Combined Quantum Mechanical and Molecular Mechanical Methods*. Gao, J.; Thompson, M. A., Eds., ACS Symposium Series 712, Oxford University Press, 1998.

481 Scuseria, G. E. Linear Scaling Density Functional Calculations with Gaussian Orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.

482 Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, 5, 19-2628.

483 Li, H.; Hains, A.W.; Everts, J.E.; Robertson, A.D.; Jensen, J.H. The Prediction of Protein pKa's Using QM/MM: The pKa of Lysine 55 in Turkey Ovomucoid Third Domain. *J. Phys. Chem. B* **2002**, *106*, 3486-3494.

484 Jensen, J.H.; Li, H.; Robertson, A.D.; Molina, P.A. Prediction and rationalization of protein pK(a) values using QM and QM/MM methods. *J. Phys. Chem. A*, **2005**, *109*, 6634-6643.

485 Minikis, R.M.; Kairys, V.; Jensen, J.H. Accurate intraprotein electrostatics derived from first principles: An effective fragment potential method study of the proton affinities of lysine 55 and tyrosine 20 in turkey ovomucoid third domain. *J. Phys. Chem. A*, **2001**, *105*, 3829-3837.

486 Porter, M.A.; Hall, J.R.; Locke, J.C.; Jensen, J.H.; Molina, P.A. Hydrogen bonding is the prime determinant of carboxyl pK(a) values at the N-termini of alpha-helices. *Proteins-Str., Funct., Bioinf.* **2006**, *63*, 621-635.

487 Naor, M.M.; Jensen, J.H. Determinants of cysteine pK(a) values in creatine kinase and alpha 1-antitrypsin. *Proteins-Str., Funct., Bioinf.* **2004**, *57*, 799-803.

488 Wang, P.F.; Flynn, A.J.; Naor, M.M.; Jensen, J.H.; Cui, G.L.; Merz, K.M.; Kenyon, G.L.; McLeish, M.J. Exploring the role of the active site cysteine in human muscle creatine kinase. *Biochemistry*, **2006**, *45*, 11464-11472.

489 Li, H.; Robertson, A.D.; Jensen, J.H. The determinants of carboxyl pK(a) values in Turkey ovomucoid third domain. *Proteins-Str., Funct., Bioinf.* **2004**, *55*, 689-704.

490 Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *J. Chem. Theory Comp.* **2009**, *5*, 459-467.

491 Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. Change in a protein's electronic structure induced by an explicit solvent: An ab initio fragment molecular orbital study of ubiquitin, *J. Comput. Chem.* **2007, *28*,** 1750-1762.

www.manaraa.com

492 He, X.; Fusti-Molnar, L.; Cui, G.; Merz, K. M. Jr. Importance of dispersion and electron correlation in ab initio protein folding. *J. Phys. Chem. B* **2009,** *113***,** 5290-5300.

493 Sawada, T.; Fedorov, D. G.; Kitaura, K. Structural and interaction analysis of helical heparin oligosaccharides with the fragment molecular orbital method. *Int. J. Quant. Chem.* **2009,** *109***,** 2033-2045.

494 Huang, L.; Massa, L.; Karle, J. Kernel energy method applied to vesicular stomatitis virus nucleoprotein. *Proc. Nat. Acad. Sciences U.S.A.* **2009,** *106*, 1731-1736.

495 Duan, L. L.; Tong, Y.; Mei, Y.; Zhang, Q. G.; Zhang, J. Z. H. Quantum study of HIV-1 protease-bridge water interaction. *J. Chem. Phys.* **2007,** *127*, 145101.

496 Dong, H.; Hua, S. G.; Li, S. H. Understanding the Role of Intra- and Intermolecular Interactions in the Formation of Single- and Double-Helical Structures of Aromatic Oligoamides: A Computational Study. *J. Phys. Chem. A* **2009**, *113*, 1335-1342.

497 Deshmukh, M. M.; Bartolotti, L. J.; Gadre, S. R. Intramolecular hydrogen bonding and cooperative interactions in carbohydrates via the molecular tailoring approach. *J. Phys. Chem. A* **2008**, *112*, 312-321.

498 Deshmukh, M. M.; Gadre, S. R. Estimation of N-H•••O=C Intramolecular Hydrogen Bond Energy in Polypeptides. *J. Phys. Chem. A* **2009**, *113*, 7927-7932.

499 Fukuzawa, K.; Kitaura, K.; Nakata, K.; Kaminuma, T.; Nakano, T. Fragment molecular orbital study of the binding energy of ligands to the estrogen receptor, *Pure Appl. Chem.* **2003**, *75***,** 2405-2410.

500 Fukuzawa, K.; Kitaura, K.; Uebayasi, M.; Nakata, K.; Kaminuma, T.: Nakano, T. Ab initio quantum mechanical study of the binding energies of human estrogen receptor with its ligands: An application of fragment molecular orbital method, *J. Comput. Chem.* **2005**, *26*, 1-10.

501 Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. Molecular Interactions between Estrogen Receptor and Its Ligand Studied by the ab Initio Fragment Molecular Orbital Method. *J. Phys. Chem. B* **2006**, *110*, 16102-16110.

502 Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Ishida, H.; Kiso, M. Why does avian influenza A virus hemagglutinin bind to avian receptor stronger than to human receptor? Ab initio fragment molecular orbital studies, *Biochem. Biophys. Res. Comm.* **2006**, *351*, 40-43.

503 Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y.; Ishida, H.; Kiso, M. Influenza viral hemagglutinin complicated shape is advantageous to its binding affinity for sialosaccharide receptor, *Biochem. Biophys. Res. Comm.* **2007**, *355*, 6-9.

504 Sawada, T.; Hashimoto, T.; Tokiwa, H.; Suzuki, T.; Nakano, H.; Ishida, H.; Kiso, M.; Suzuki, Y. Ab initio fragment molecular orbital studies of influenza virus hemagglutinin-sialosaccharide complexes toward chemical clarification about the virus host range determination. *Glycoconj. J.* **2008**, 25, 805-815.

505 Sawada, T.; Hashimoto, T.; Tokiwa, H.; Suzuki, T.; Nakano, H.; Ishida, H.; Kiso, M.; Suzuki, Y. Ab initio base fragment molecular orbital studies of influenza viral hemagglutinin HA1 full-domains in complex with sialoside receptors. *J. Mol. Genet. Med.* **2009**, *3*, 133-142.

506 Sawada, T.; Fedorov, D. G.; Kitaura, K. Role of the key mutation in the selective binding of avian and human influenza hemagglutinin to sialosides revealed by quantum-mechanical calculations. *J. Am. Chem. Soc.* **2010**, *132*, 16862-16872.

www.manaraa.com

507 Sawada, T.; Fedorov, D. G.; Kitaura, K. Binding of influenza A virus hemagglutinin to the sialoside receptor is not controlled by the homotropic allosteric effect. *J. Phys. Chem. B* **2010**, *114*, 15700-15705.

508 Takematsu, K.; Fukuzawa, K.; Omagari, K.; Nakajima, S.; Nakajima, K.; Mochizuki, Y.; Nakano, T.; Watanabe, H.; Tanaka, S. Possibility of mutation prediction of influenza hemagglutinin by combination of hemadsorption experiment and quantum chemical calculation for antibody binding. *J. Phys. Chem. B* **2009**, *113*, 4991-4994.

509 Iwata, T.; Fukuzawa, K.; Nakajima, K.; Aida-Hyugaji, S.; Mochizuki, Y.; Watanabe, H.; Tanaka, S.; Theoretical analysis of binding specificity of influenza viral hemagglutinin to avian and human receptors based on the fragment molecular orbital method. *Comput. Biol. Chem.* **2008**, *32*, 198-211.

510 Yamagishi, K.; Yamamoto, K.; Yamada, S.; Tokiwa, H. Functions of key residues in the ligand-binding pocket of vitamin D receptor: Fragment molecular orbital-interfragment interaction energy analysis, *Chem. Phys. Lett.* **2006**, *420*, 465-468.

511 Motoyoshi, S.; Yamagishi, K.; Yamada, S.; Tokiwa, H. Ligand-dependent conformation change reflects steric structure and interactions of a vitamin D receptor/ligand complex: a fragment molecular orbital study. *J. Ster. Biochem. Mol. Biol.* **2010**, *121*, 56-59.

512 Yamagishi, K.; Tokiwa, H.; Makishima, M.; Yamada, S. Interactions between 1alpha,25(OH)2D3 and residues in the ligand-binding pocket of the vitamin D receptor: a correlated fragment molecular orbital study. *J. Ster. Biochem. Mol. Biol.* **2010**, *121*, 63-67.

513 Ito, M.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Ab Initio Fragment Molecular Orbital Study of Molecular Interactions between Liganded Retinoid X Receptor and Its Coactivator: Roles of Helix 12 in the Coactivator Binding Mechanism. *J. Phys. Chem. B* **2007**, *111*, 3525-3533.

514 Ito, M.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Ab Initio Fragment Molecular Orbital Study of Molecular Interactions between Liganded Retinoid X Receptor and Its Coactivator; Part II: Influence of Mutations in Transcriptional Activation Function 2 Activating Domain Core on the Molecular Interactions, *J. Phys. Chem. A* **2008**, *112*, 1986-1998.

515 Ito, M.; Fukuzawa, K.; Ishikawa, T.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Ab Initio Fragment Molecular Orbital Study of Molecular Interactions in Liganded Retinoid X Receptor: Specification of Residues Associated with Ligand Inducible Information Transmission. *J. Phys. Chem. B.* **2008**, *112*, 12081-12094.

516 Nakanishi, I.; Fedorov, D. G.; Kitaura, K. Molecular recognition mechanism of FK506 binding protein: An all-electron fragment molecular orbital study. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 145-158.

517 Sugiki, S.-I.; Matsuoka, M.; Usuki, R.; Sengoku, Y.; Kurita, N.; Sekino, H.; Tanaka, S. Density functional calculations on the interaction between catabolite activator protein and cyclic AMP using the fragment molecular orbital method, *J. Theor. Comput. Chem.* **2005**, *4*, 183-195.

518 Nemoto, T.; Fedorov, D. G.; Uebayasi, M.; Kanazawa, K.; Kitaura, K.; Komeiji, Y. Ab initio fragment molecular orbital (FMO) method applied to analysis of the ligand-protein interaction in a pheromone-binding protein, *Comput. Biol. Chem.* **2005**, *29*, 434.

519 Watanabe, H.; Enomoto, T.; Tanaka, S. Ab initio study of molecular interactions in higher plant and Galdieria partita Rubiscos with the fragment molecular orbital method, *Biochem. Biophys. Res. Comm.* **2007**, *361*, 367-372.

www.manaraa.com

520 Harada, T.; Yamagishi, K.; Nakano, T.; Kitaura, K.; Tokiwa, H.; Ab initio fragment molecular orbital study of ligand binding to human progesterone receptor ligand-binding domain. *Naunyn-Schmiedeberg's Arch. Pharmac.* **2008**, *377*, 607-615.

521 Tada, M.; Nagasima, T.; Udagawa, T.; Tachikawa, M.; Sugawara, H. Ab initio fragment molecular orbital (FMO) analysis of the structure of the phosphoinositide-binding peptide from gelsolin. *J. Mol. Str. (THEOCHEM)* **2009**, *897*, 149-153.

522 Dedachi, K.; Khan, M. T. H., Sylte, I.; Kurita, N. A combined simulation with ab initio MO and classical vibrational analysis on the specific interactions between thermolysin and dipeptide ligands. *Chem. Phys. Lett.* **2009**, *479*, 290-295.

523 Van Schouwen, B. M. B., Nakano, M.; Watanabe, H.; Tanaka, S.; Gordon, H. L.; Rothstein, S. M. Molecular mechanics and all-electron fragment molecular orbital calculations on mutated polyglutamine peptides. *J. Mol. Str. (THEOCHEM)* **2010**, *944*, 12-20.

524 Yamagishi, K.; Yamamoto, K.; Mochizuki, Y.; Nakano, T.; Yamada, S.; Tokiwa, H. Flexible ligand recognition of peroxisome proliferator-activated receptor-g (PPARg). *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3344-3347.

525 Yoshikawa, E.; Miyagi, S.; Dedachi, K.; Ishihara-Sugano, M.; Itoh, S.; Kurita, N. Specific interactions between aryl hydrocarbon receptor and dioxin congeners: Ab initio fragment molecular orbital calculations. *J. Mol. Graph. Mod.* **2010**, *29*, 197-205.

526 Ishikawa, T.; Kuwata, K. Interaction Analysis of the Native Structure of Prion Protein with Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 538-547.

527 Hasegawa, K.; Mohri, S.; Yokoyama, T. Fragment molecular orbital calculations reveal that the E200K mutation markedly alters local structural stability in the human prion protein. *Prion* **2010**, *4*, 38-44.

528 Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Kato, A.; Nakano, T.; Tanaka, S. Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: Ab initio fragment molecular orbital study. *J. Comput. Chem.* **2006**, *27*, 948-960.

529 Kurisaki, I.; Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Imada, J.; Chmielewski, A.; Rothstein, S. M.; Watanabe, H.; Tanaka, S. Visualization analysis of inter-fragment interaction energies of CRP-cAMP-DNA complex based on the fragment molecular orbital method. *Biophys. Chem.* **2007**, *130*, 1-9.

530 Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. DNA and Estrogen Receptor Interaction Revealed by Fragment Molecular Orbital Calculations, *J. Phys. Chem. B.* **2007**, *111*, 9621-9627.

531 Kurisaki, I.; Fukuzawa, K.; Nakano, T.; Mochizuki, Y.; Watanabe, H.; Tanaka, S. Fragment molecular orbital (FMO) study on stabilization mechanism of neuro-oncological ventral antigen (NOVA)-RNA complex system. *J. Mol. Str. (THEOCHEM)* **2010**, *962*, 45-55.

532 Ozawa, T.; Okazaki, K. CH/p hydrogen bonds determine the selectivity of the Src homology 2 domain to tyrosine phosphotyrosyl peptides: An ab initio fragment molecular orbital study. *J. Comput. Chem.* **2008**, *29*, 2656-2666.

533 Ozawa, T.; Tsuji, E.; Ozawa, M.; Handa, C.; Mukaiyama, H.; Nishimura, T.; Kobayashi, S.; Okazaki, K. The importance of CH/p hydrogen bonds in rational drug design: An ab initio fragment molecular orbital study to leukocyte-specific protein tyrosine (LCK) kinase, *Bioorg. Med. Chem.* **2008**, *16*, 10311-10318.

www.manaraa.com

534 Fujimura, K.; Sasabuchi, Y. The Role of Fluorine Atoms in a Fluorinated Prostaglandin Agonist. *Chem. Med. Chem.* **2010**, *5*, 1254-1257.

535 Ohno, K.; Mori, K.; Orita, M.; Takeuchi, M. Computational Insights into Binding of Bisphosphates to Farnesyl Pyrophosphate Synthase. *Curr. Med. Chem.* **2011**, *18*, 220-233.

536 Zhang, D. W.; Zhang, J. Z. H. Full quantum mechanical study of binding of HIV-1 protease drugs. *Int. J. Quantum Chem.* **2005**, *103*, 246-257.

537 Mei, Y.; He, X.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Quantum study of mutational effect in binding of efavirenz to HIV-1 RT. *Proteins-Str., Funct., Bioinf.* **2005**, *59*, 489-495.

538 He, X.; Mei, Y.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations. *Proteins-Str., Funct., Bioinf.* **2005**, *61*, 423-432.

539 Wu, E. L.; Han, K. L.; Zhang, J. Z. H. Computational study for binding of oscillarin to human a-thrombin. *J. Theor. Comput. Chem.* **2009**, *8*, 551-560.

540 Tong, Y.; Mei, Y.; Zhang, J. Z. H.; Duan, L. L.; Zhang, Q. G. Quantum calculation of protein solvation and protein-ligand binding free energy for HIV-1 protease/water complex. *J. Theor. Comput. Chem.* **2009**, *8*, 1265-1279.

541 Huang, L. L.; Massa, L.; Karle, J. Drug target interaction energies by the kernel energy method in aminoglycoside drugs and ribosomal A site RNA targets. *Proc. Nat. Acad. Sciences U.S.A.* **2007**, *104*, 4261-4266.

542 Orimoto, Y.; Gu, F. L.; Imamura, A.; Aoki, Y., Efficient and accurate calculations on the electronic structure of B-type poly(dG)· poly(dC) DNA by elongation method: First step toward the understanding of the biological properties of aperiodic DNA. *J. Chem. Phys.* **2007,** *126*, *215104.*

543 Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *J. Cheminf.* **2011**, *3*, 2.

544 Hayik, S. A.; Dunbrack, R., Jr.; Merz, K. M., Jr. Mixed Quantum Mechanics/Molecular Mechanics Scoring Function To Predict Protein-Ligand Binding Affinity. *J. Chem. Theory Comput.* **2010**, *6*, 3079-3091.

545 Whittaker, M.; Law, R. J.; Ichihara, O.; Hesterkamp, T.; Hallett, D. Fragments: past, present and future. *Drug Discovery Today: Technologies* **2010**, *7*, e163-e171, and references therein.

546 Yoshida, T.; Fujita, T.; Chuman, H.; Novel Quantitative Structure-Activity Studies of HIV-1 Protease Inhibitors of the Cyclic Urea Type Using Descriptors Derived from Molecular Dynamics and Molecular Orbital Calculations. *Curr. Comput.-Aided Drug Des.* **2009,** *5***,** 38-55.

547 Ishikawa, T.; Mochizuki, Y.; Amari, S.; Nakano, T.; Tanaka, S.; Tanaka, K. An application of fragment interaction analysis based on local MP2. *Chem. Phys. Lett.* **2008**, *463*, 189-194.

548 Yoshida, T.; Yamagishi, K.; Chuman, H. QSAR Study of Cyclic Urea Type HIV-1 PR Inhibitors Using Ab Initio MO Calculation of Their Complex Structures with HIV-1. *QSAR Comb. Sci.* **2008,** *27***,** 694-703.

549 Fischer, B.; Fukuzawa, K.; Wenzel, W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins: Struct., Funct., Bioinf.* **2008,** *70***,** 1264-1273.

www.manaraa.com

550 Yoshida, T.; Munei, Y.; Hitaoka, S.; Chuman, H. Correlation analyses on binding affinity of Substituted benzenesulfonamides with carbonic anhydrase using ab initio MO calculations on their Complex structures. *J. Chem. Inf. Model.* **2010**, *50*, 850-860.

551 Hitaoka, S.; Harada, M.; Yoshida, T.; Chuman, H. Correlation Analyses on Binding Affinity of Sialic Acid Analogues with Influenza Virus Neuraminidase-1 Using ab Initio MO Calculations on Their Complex Structures. *J. Chem. Inf. Model.* **2010**, *50*, 1796-1805.

552 Munei, Y.; Shimamoto, K.; Harada, M.; Yoshida, T.; Chuman, H. Correlation analyses on binding affinity of substituted benzenesulfonamides with carbonic anhydrase using ab initio MO calculations on their complex structures (II). *Bioorg. Med. Chem. Lett.* **2011**, *21*, 141-144.

553 Ishida, T.; Fedorov, D. G.; Kitaura, K. All Electron Quantum Chemical Calculation of the Entire Enzyme System Confirms a Collective Catalytic Device in the Chorismate Mutase Reaction, *J. Phys. Chem. B* **2006**, *110*, 1457-1463.

554 Nakamura, T.; Yamaguchi, A.; Kondo, H.; Watanabe, H.; Kurihara, T.; Esaki, N.; Hirono, S.; Tanaka, S. Roles of K151 and D180 in L-2-haloacid dehalogenase from Pseudomonas sp. YL: Analysis by molecular dynamics and ab initio fragment molecular orbital calculations. *J. Comput. Chem.* **2009**, *30*, 2625-2634.

555 Mochizuki, Y.; Nakano, T.; Amari, S.; Ishikawa, T.; Tanaka, K.; Sakurai, M.; Tanaka, S. Fragment molecular orbital calculations on red fluorescent protein (DsRed). *Chem. Phys. Lett.* **2007**, *433*, 360-367.

556 Tagami, A.; Ishibashi, N.; Kato, D.; Taguchi, N.; Mochizuki, Y.; Watanabe, H.; Ito, M.; Tanaka, S. Ab initio quantum-chemical study on emission spectra of bioluminescent luciferases by fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *472*, 118-123.

557 Taguchi, N.; Mochizuki, Y.; Nakano, T.; Amari, S.; Fukuzawa, K.; Ishikawa, T.; Sakurai, M.; Tanaka, S. Fragment Molecular Orbital Calculations on Red Fluorescent Proteins (DsRed and mFruits). *J. Phys. Chem. B* **2009**, *113*, 1153-1161.

558 Taguchi, N.; Mochizuki, Y.; Nakano, T. Fragment molecular orbital calculations for excitation energies of blue- and yellow-fluorescent proteins. *Chem. Phys. Lett.* **2011**, *504*, 76-82.

559 Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. Fragment molecular orbital study of the electronic excitations in the photosynthetic reaction center of Blastochloris viridis. *J. Comput. Chem.* **2010**, *31*, 447-454.

560 Milne, B. F.; Marques, M. A. L.; Nogueira, F. Fragment molecular orbital investigation of the role of AMP protonation in firefly luciferase pH-sensitivity. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14285-14293.

561 Huang, L.; Massa, L.; Karle, I.; Karle, J. Calculation of strong and weak interactions in TDA1 and RangDP52 by the kernel energy method. *Proc. Nat. Acad. Sciences U.S.A.* **2009**, *106*, 3664-3669.

562 Fukunaga, H.; Fedorov, D. G.; Chiba, M.; Nii, K.; Kitaura, K. Theoretical Analysis of the Intermolecular Interaction Effects on the Excitation Energy of Organic Pigments: Solid State Quinacridone. *J. Phys. Chem. A* **2008**, *112*, 10887-10894.

563 Zhang, R. J.; Tian, W. Q.; Gu, F. L.; Aoki, Y. Theoretical studies on the adsorption of Si and C chains onto unfaulted and faulted Si(111) surfaces. *J. Phys. Chem. C* **2007**, *111*, 6350-6356.

www.manaraa.com

564 Chen, W.; Yu, G. T.; Gu, F. L.; Aoki, Y. Investigation on the Electronic Structures and Nonlinear Optical Properties of Pristine Boron Nitride and Boron Nitride-Carbon Heterostructured Single-Wall Nanotubes by the Elongation Method. *J. Phys. Chem. C* **2009**, *113*, 8447-8454; ibid, 11424-11424.

565 Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. Analytic gradient for the adaptive frozen orbital bond detachment in the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *477*, 169-175.

566 Orimoto, Y.; Aoki, Y., Quantum-chemical approach to the solvatochromic transition in polysilane derivatives. *J. Polymer Science Part B-Polymer Phys.* **2006**, *44*, 119-133.

567 Ohnishi, S. I.; Orimoto, Y.; Gu, F. L.; Aoki, Y. Nonlinear optical properties of polydiacetylene with donor-acceptor substitution block. *J. Chem. Phys.* **2007**, *127*, 084702.

568 Yu, G. T.; Chen, W.; Gu, F. L.; Orimoto, Y.; Aoki, Y. Theoretical study on static (hyper)polarizabilities for polyimide by the elongation finite-field method. *Mol. Phys.* **2009**, *107*, 81-87.

569 Yu, G. T.; Chen, W.; Gu, F. L.; Aoki, Y. Theoretical Study on Nonlinear Optical Properties of the Li+calix 4 pyrrole Li(-)Dimer, Trimer and its Polymer with Diffuse Excess Electrons. *J. Comput. Chem.* **2010**, *31*, 863-870.

570 Pomogaeva, A.; Gu, F. L.; Imamura, A.; Aoki, Y. Electronic structures and nonlinear optical properties of supramolecular associations of benzo-2,1,3-chalcogendiazoles by the elongation method. *Theor. Chem. Acc.* **2010**, *125*, 453-460.

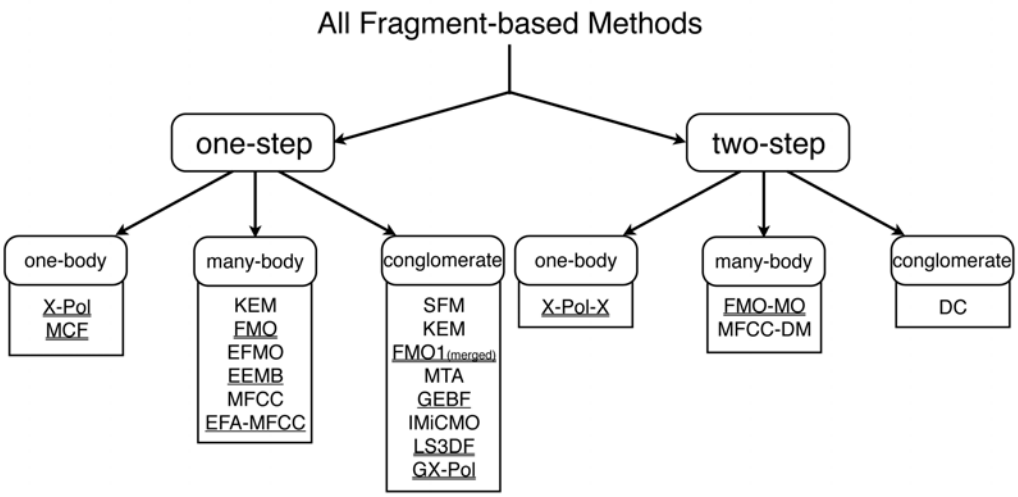
571 Orimoto, Y.; Gu, F. L.; Korchowiec, J.; Imamura, A.; Aoki, Y. Application of the elongation method to the electronic structure of spin-polarized molecular wire under electric field. *Theor. Chem. Acc.* **2010,** *125*, 493-501.

572 Yan, L. K.; Pomogaeva, A.; Gu, F. L.; Aoki, Y. Theoretical study on nonlinear optical properties of metalloporphyrin using elongation method. *Theor. Chem. Acc.* **2010**, *125*, 511-520.

573 Pomogaev, V.; Gu, F. L.; Pomogaeva, A.; Aoki, Y. Elongation Method for Calculating Excited States of Aromatic Molecules Embedded in Polymers. *Int. J. Quantum Chem.* **2009**, *109*, 1328-1340.

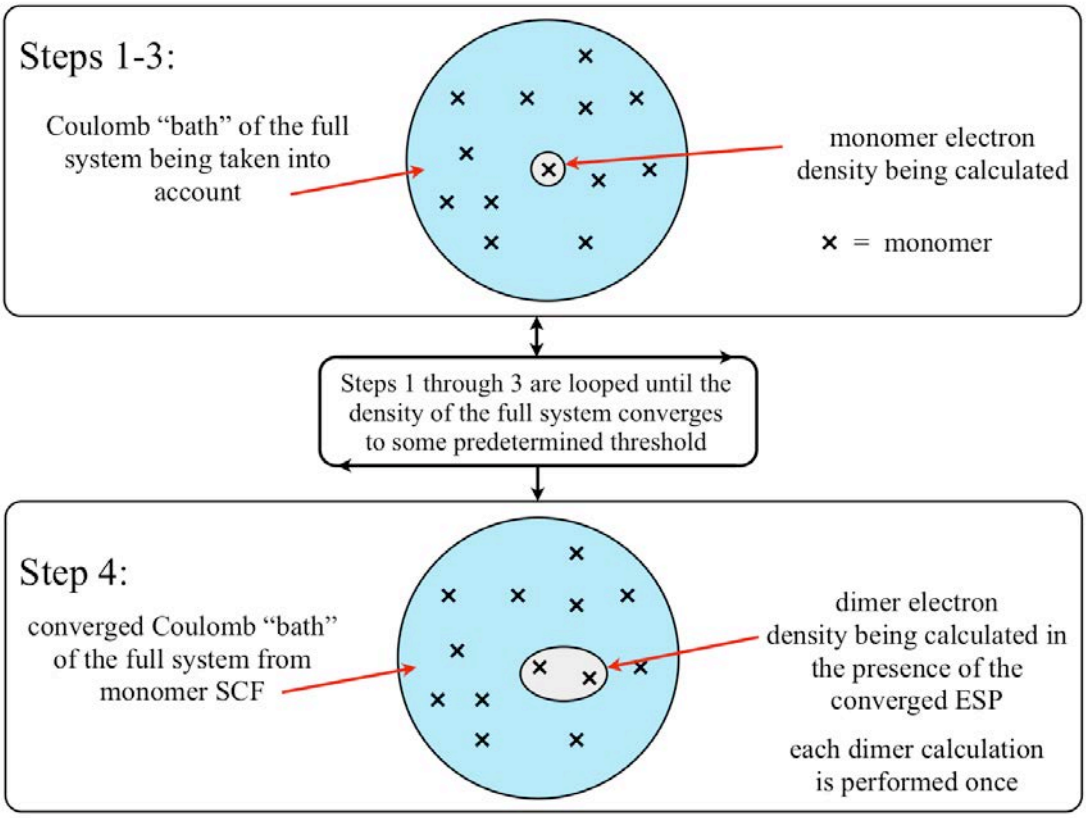
www.manaraa.com

Figure. 1. Classification of fragment-based methods. Singly underlined approaches include constant embedding potentials, while doubly underlined approaches include some form of self-consistent fragment embedding potential (SCC), other approaches have no embedding potential.



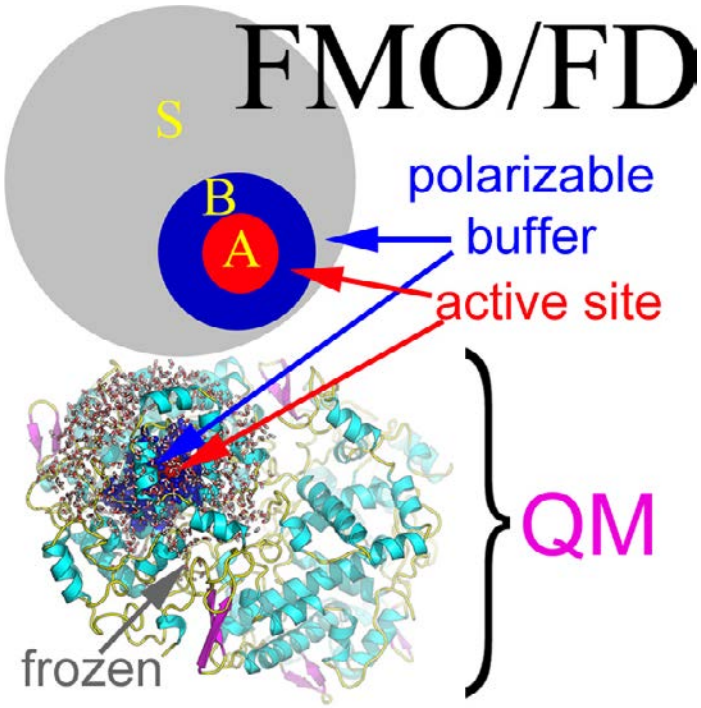
www.manaraa.com

Figure. 2. General outline of a standard FMO calculation including up to dimer interactions.



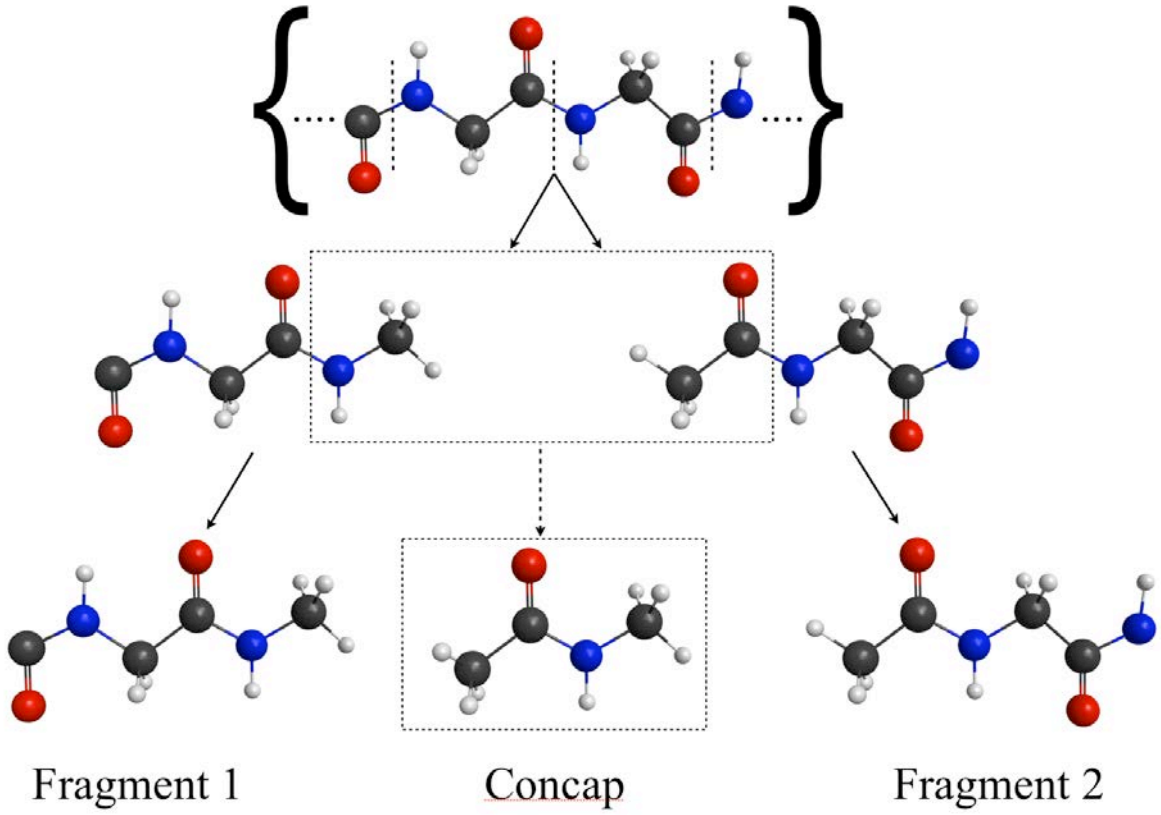
www.manaraa.com

Figure. 3. Schematic representation of FMO/FD and FMO/FDD. Fragments are divided into three domains, frozen (grey), polarizable (blue) and active (red). All three domains are computed with QM (using multilayer FMO). An actual geometry optimization of the partially solvated prostaglandin H(2) synthase-1 in complex with the reversible competitive inhibitor ibuprofen (PDB: 1EQG) is shown at the lower part [this Figure is reproduced from the TOC graphic of J. Phys. Chem. Lett. 2011, 2, 282-288].



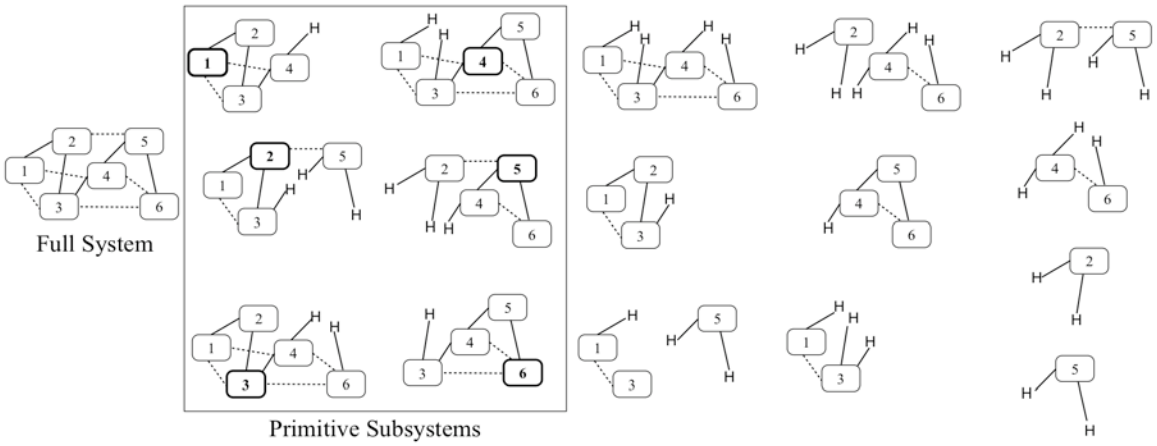
www.manaraa.com

Figure. 4. Example of a typical MFCC fractionation scheme. The parent system is divided at the peptide bond into two fragments that are then capped. The caps are then combined into a single concap to be subtracted from the sum of fragment contributions.



www.manaraa.com

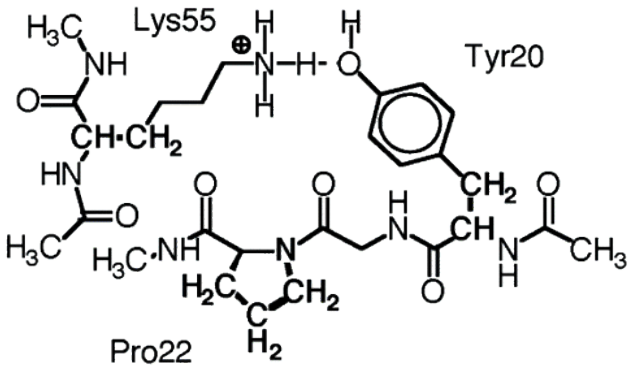
Figure. 5. Depiction of a fragmentation scheme created by the GEBF method containing primitive subsystems derived from each central fragment. Solid lines depict covalent bonds while dashed lines represent hydrogen bonds. All possible subsystems are shown, including those that would be removed (i.e. 1234) due to double counting.



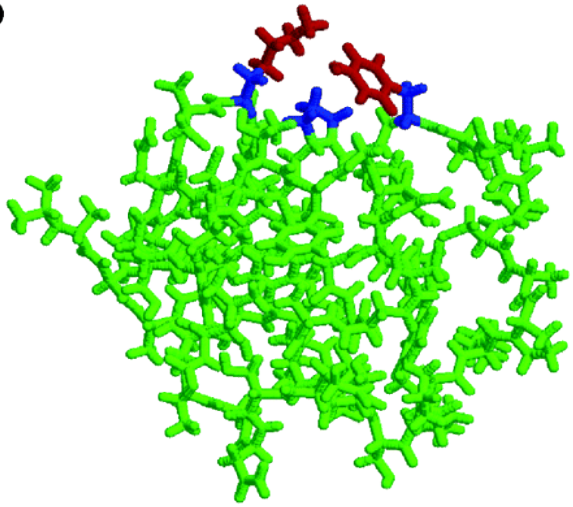
www.manaraa.com

Figure. 6. Subsystem of OMTKY3 (protein turkey ovomucoid third domain) used to obtain the buffer region (bold) used for (b) ab initio/buffer/EFP regions (red/blue/green) used for the computation of the pKa of Lys55. (This Figure is reproduced from Jensen, J.H.; Li, H.; Robertson, A.D.; Molina, P.A. J. Phys. Chem. A, 2005, 109, 6634-6643.)

a

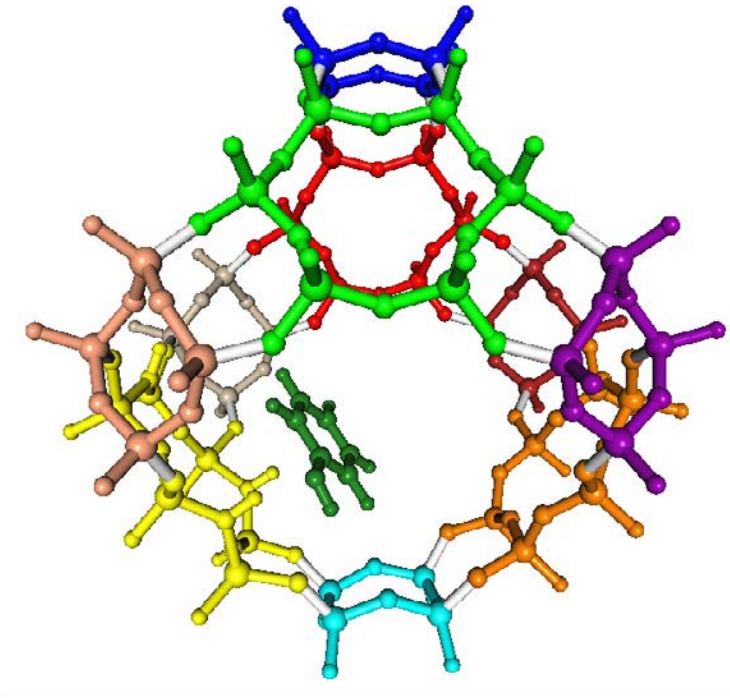


b



www.manaraa.com

Figure. 7. Fragmentation of the complex of phenol and faujasite zeolite, computed with FMO.[231] The phenol molecule is shown in dark green, and the active catalytic site containing three aluminum atoms and three acidic hydrogens is shown in yellow.



www.manaraa.com

Scheme I. Comparison of the Effective fragment molecular orbital method (EFMO) with EFP and FMO.

Scheme I.	
Differences between EFMO and EFP	**Differences between EFMO and FMO**
1. The monomer energies are explicitly included in EFMO to give an estimate of the total energy of the system and allow internal geometries of the fragments to change. 2. In EFMO, short-range interactions between the fragments are computed *ab initio* rather than with EFP expression: $E^{EFP} = E^{Coul} + E^{pol} + E^{disp} + E^{exrep} + E^{ct}$ 3. In EFMO, long-range interactions are modeled by electrostatic Coulomb and polarization terms only; dispersion, exchange-repulsion and charge-transfer are considered to be negligible at large separations.	1. In EFMO, monomer calculations are performed in the absence of external field and done once. 2. In EFMO, all long-range and many-body interactions are included through classical Coulomb and polarization terms.

www.manaraa.com

Chapter 3. The Fragment Molecular Orbital and Systematic Molecular Fragmentation Methods Applied to Water Clusters

A paper published in *The Journal of Physical Chemistry and Chemical Physics*

Spencer R. Pruitt, Matthew A. Addicoat, Michael A. Collins, and Mark S. Gordon

Abstract

Two electronic structure methods, the fragment molecular orbital (FMO) and systematic molecular fragmentation (SMF) methods, that are based on fragmenting a large molecular system into smaller, more computationally tractable components (fragments), are presented and copared with fully *ab initio* results for the predicted binding energies of water clusters. It is demonstrated that, even when explicit three-body effects are included (especially necessary for water clusters due to their complex hydrogen-bonded networks) both methods present viable, computationally efficient alternatives to fully *ab initio* quantum chemistry.

1. Introduction

Understanding the dynamics of molecules and ions in water has been a major aim of theoretical chemistry in order to fully describe how reactions take place in aqueous solution. Water is the most important solvent existing. The structure and function of all biological molecules rely on solvation by water. The local structure and role of water molecules around active sites in enzymes are essential components of substrate binding. The activation energy of all aqueous chemical processes that involve some charge rearrangement depends on both the static and dynamic polarization of the nearby water molecules. All of these chemical and biological processes depend on the "delicate molecular inter-linking" (hydrogen bonding) of water molecules.[1]

This unusual interlinking of water molecules presents substantial difficulties for accurate modelling of bulk water or large water clusters. The energetics of small water clusters can be treated *ab initio* using high levels of electronic structure theory such as coupled cluster theory with large basis sets. [2] High level methods have also been employed to model larger clusters.[3] However, the computational time for these very high level methods scale as the seventh power of the number of water molecules. In order to model bulk water with sufficient accuracy, it is useful to establish the accuracy of more computationally feasible methods that scale much more slowly with the number of water molecules considered.

The task of estimating the electronic energy of large systems at low computational cost has been the subject of much study, using multi-layer quantum mechanics (QM)/molecular mechanics (MM) methods,[4-6] electrostatic embedding methods,[7] and the effective fragment potential (EFP) method[8, 9] [see Ref. [10], and references therein]. A number of groups [10-27] have reported successful methods for estimating the *ab initio* electronic energy (and other properties) of large molecules by breaking the molecule into fragments. Here, two of these approaches are considered which provide systematic hierarchies for estimating the energies of

www.manaraa.com

molecules or clusters, the fragment molecular orbital (FMO) method and the systematic molecular fragmentation (SMF) method.

The FMO method was recently applied to the calculation of the energy of some clusters of 32 water molecules.[10] In the present paper, a range of water clusters (16, 20, 32 and 64 water molecules) is studied using both the FMO and SMF methods. This allows a comprehensive comparison of the accuracy of the two methods for the binding energy of water clusters. The sizes of the clusters, the sizes of the basis sets employed, and the use of second order Møller-Plesset perturbation theory (MP2) [28] is determined by the necessity to calculate the "exact" total energy of the clusters for comparison with the FMO and SMF estimates. Since the MP2 level of theory gives at least a semi-quantitatively reasonable description of the interaction of water molecules, it provides a useful "test-bed" to determine if the FMO and SMF methods are robust methods for estimating the energetics of large clusters of water molecules.[29]

The paper is arranged as follows: Section 2 describes new modifications to the SMF method to account for the charge distribution in the whole water cluster, and includes a brief description of the FMO method. Section 3 presents and discusses the results of tests of the accuracy of the FMO and SMF methods for the binding energies and relative energies of water molecules in clusters of 16, 20, 32 and 64 water molecules. Discussion and concluding remarks are contained in the final section.

2. Methodology

2.1 Systematic molecular fragmentation

The systematic molecular fragmentation (SMF) method decomposes the whole molecule into small molecular fragments by breaking single bonds between functional groups, according to an established algorithm.[22-24] Recently a simple variation on this approach, called systematic molecular fragmentation by annihilation (SMFA), has been reported [30] which is better suited for the estimation of cluster energies.

In order to treat water clusters as a molecule, each water monomer is defined as a functional group. In this application of the SMFA method to molecular clusters, a simple *ansatz* is adopted for bonding: A bond exists between two water molecules if the shortest atom-atom distance between the molecules is less than 2.3 Å (significantly larger than a typical intermolecular hydrogen bond length of about 1.8 Å). For any given cluster geometry, single bonds are assigned according to this criterion. The cluster is then fragmented, using the SMFA algorithm, except that hydrogen atom "caps"[22, 23] on broken bonds are not required. This fragmentation is represented by

$$Cluster \rightarrow \sum_{n=1}^{N_{frag}} f_n F_n , \qquad (1)$$

www.manaraa.com

where the $\{F_n\}$ are small sub-clusters of water molecules, and the $\{f_n\}$ are integer coefficients. The SMFA method is hierarchical, meaning that there are a series of approximations denoted as Level 1, Level 2, and so on, in which the water sub-clusters increase in size. In simple terms, at Level *n* any two water molecules separated by up to n bonds will occupy at least one sub-cluster in common, whereas pairs of molecules separated by n+1 bonds (or more) will not both be found in any sub-cluster.

The "bonding energy" of the whole cluster is given by

$$E_b = \sum_{n=1}^{N_{frag}} f_n E(F_n), \tag{2}$$

where $E(F_n)$ denotes the electronic energy of fragment F_n. For a Level 1 fragmentation, Eqs (1) and (2) give a simple description of the cluster and cluster energy in terms of water dimers (for those dimers in which the shortest atom-atom distance between the molecules is less than 2.3 Å). For higher levels of fragmentation, the energy is given in terms of larger sub-clusters of water molecules. For Level 2, the largest sub-clusters would generally contain five water molecules; one central molecule linked by hydrogen bonds to four nearby molecules. Level 3 sub-clusters often contain as many as nine water molecules.

Eq,(2) neglects the energy, E_{nb}, due to longer range (i.e., longer than bonded interactions) non-bonded interactions between water molecules. In previous applications of the SMF method, this longer range interaction has been included in terms of Coulomb, induction and dispersion energies

$$E = E_b + E_{nb}, \tag{3}$$

where

$$E_{nb} = E_{ele} + E_{ind} + E_{disp} \tag{4}$$

Because a water molecule has a significant dipole moment, it is expected that the Coulomb and induction interactions will be relatively significant in large water clusters. Due to these significant contributions, the Coulomb and induction interactions are included by "embedding" [7, 26] the fragments, F_n, in the charge distribution associated with all other molecules in the cluster not contained in F_n. Each water molecule in the cluster has an associated charge distribution. The electronic energy of each fragment is evaluated in the presence of the charge distributions for all water molecules that are not in that fragment. The energy of fragment F_n in the presence of the charge distribution associated with all other groups,[*i*], is written as $E\left[F_n; i \notin F_n\right]$. This energy replaces $E(F_n)$ in Eq. (2). It has previously been noted by Li et. al[26] that setting

www.manaraa.com

$$E = \sum_{n=1}^{N_{frag}} f_n E\left[F_n; i \notin F_n\right], \tag{5}$$

evaluates the Coulomb interaction of non-bonded groups twice, and must be corrected.

The total energy that accounts for both bonded and non-bonded interactions is then given by

$$E = \sum_{n=1}^{N_{frag}} f_n E\left[F_n; i \notin F_n\right] - \frac{1}{2} \sum_{i=1}^{N_{groups}} \sum_{\substack{j=1 \\ j,i \notin \{F_k\}}}^{N_{groups}} E_{ele}(i,j) + E_{disp}, \tag{6}$$

where the number of groups (water monomers) is denoted by N_{groups}. In Eq. (6), $E_{ele}(i,j)$ denotes the classical Coulomb interaction of the charge distributions associated with groups *i* and *j*. Note that by embedding each fragment in all other charges, the many-body nature of the induction energy is accounted for and is included in the *ab initio* calculation of the fragment energies. The double summation in Eq. (6) includes all pairs of groups that are never both contained in any fragment. The dispersion energy is approximated here by a pair-wise dispersion interaction between groups that are never both contained in any fragment:

$$E_{disp} = \sum_{i=1}^{N_{groups}} \sum_{\substack{j>i \\ j,i \notin \{F_k\}}}^{N_{groups}} E_{disp}(i,j). \tag{7}$$

This pair-wise dispersion energy, $E_{disp}(i,j)$, is evaluated as previously described,[24] using the anisotropic static polarizability and the imaginary-frequency-dependent polarizability of each group evaluated at the same level of *ab initio* theory used to evaluate the $E\left[F_n; i \notin F_n\right]$.

Charge distributions

The question arises as to what charge distribution is associated with each water molecule. The set of charges associated with each water molecule,$\{q(i)\}$, is evaluated self-consistently as follows. In the first iteration, an *ab initio* calculation at the given level of theory is carried out for each water molecule. The calculated electron density is analysed as a multipole expansion centered on each atom, using the Stone distributed multipole analysis (DMA) method.[31] This expansion includes a charge, dipole and quadrupole on each atom. In the next iteration, the calculation of the electronic wavefunction for each water molecule is repeated, but now embedded in the presence of the charges (but not the higher order multipoles) on all other atoms, as evaluated in the previous iteration. The newly calculated electron density is analysed using the DMA method to give updated values for the multipole moments centered on each atom (including the charges). This process is iterated to self-consistency, however it has been found that only one iteration is required. This

www.manaraa.com

process provides charges, dipoles and quadrupoles centered on each atom that reflect the polarization of the charge density on each water molecule by all other water molecules. In some *ab initio* program packages that do not have full multipole capability (e.g., Gaussian03), these multipole moments must be represented as closely spaced groups of charges. Table S1 in the supplementary material presents expressions for the locations and magnitudes of the charges in the vicinity of each atom, which represent the charge, dipole and quadrupole for each atom.

Bonding between water monomers has been defined in terms of a specific "cut-off" distance criterion described above. Such a definition must produce a discontinuity in the total energy at configurations where the distance between two monomers exactly equals this cut-off distance. A switching function that smoothly exchanges the bonded fragment energy for the corresponding non-bonded calculation might be used to overcome this problem. This task will be the focus of future work.

2.2 The Fragment Molecular Orbital (FMO) method

The FMO method [32] treats electron exchange and charge transfer as largely local phenomena. By making this approximation, a chemical system may be broken into many smaller, localized pieces, treating long range effects of the full system using only a Coulomb operator. Large reductions in computational cost required for calculations can be achieved using this approximation, allowing for calculations on much larger systems than would be possible using fully *ab initio* methods. As a means to further reduce the computational cost, the FMO method employs multi-level parallelization through the use of the Generalized Distributed Data Interface (GDDI). [33] This combination of theoretical approximations and efficient computational algorithms has allowed the FMO method to perform all electron calculations on over 20,000 atoms.[34]

Fragments are created by breaking bonds electrostatically,[35, 36] with the user choosing the most appropriate fragmentation scheme using their chemical intuition. Broken bonds are divided heterolytically, with two electrons assigned to one fragment and no electrons assigned to the other. To avoid the charged fragments that would be created by stopping here, a proton from the electron deficient fragment is reassigned to the electron rich fragment. This is analogous to the localized charge distribution method.[37, 38] This procedure creates two neutral fragments, both of which carry sp^3 hybridized orbitals at the broken bond. Once all of the fragments are created, individual (monomer) calculations are performed for each fragment in the Coulomb “bath” of the full system, representing the electrostatic potential (ESP) of the remaining N-1 fragments (N = total number of fragments in the system). This first FMO level, called FMO1, may be improved upon by including explicit two-body (dimer) and three-body (trimer) calculations. [39, 40] All of the many-body interactions are obtained through fully quantum calculations, including all relevant short range interactions (electron exchange, electron correlation and charge transfer), and are performed in the rigorous ESP of the full system.

The following outline is followed during any given FMO calculation:

www.manaraa.com

1. The initial electron density distribution is calculated for each monomer in the Coulomb bath of the rest of the system
2. The monomer Fock operators are then constructed using these densities and the energy of each monomer is calculated
3. Each of the monomer energies is iterated to self-consistency, leading to a converged ESP. This step is commonly called the "Monomer SCF"
4. If one chooses to use the FMO2 level, fragment dimer calculations are then performed in the converged ESP of the rest of the system. Each dimer calculation is only performed once (not iterated to self-consistency).
5. If one chooses to use the FMO3 level, fragment trimer calculations are performed next in the converged ESP of the rest of the system. Each trimer calculation is only performed once (not iterated to self-consistency).

The total energy of the system within the FMO approximation can be written as:

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$

$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) \tag{8}$$

$$-(E_{IJ} - E_I - E_J) - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\}$$

with monomer (*I*), dimer (*IJ*) and trimer (*IJK*) energies all obtained through the standard SCF method. The inclusion of higher order interactions has recently been reported[40]. The Fock equation used to calculate each of the energies in Eq. (8) is modified from the standard form by adding terms to the one electron Hamiltonian that represent the ESP:

$$\tilde{\mathbf{F}}^x \mathbf{C}^x = \mathbf{S}^x \mathbf{C}^x \tilde{\varepsilon}^x \quad x = I, IJ, IJK \tag{9}$$

$$\tilde{\mathbf{F}}^x = \tilde{\mathbf{H}}^x + \mathbf{G}^x \tag{10}$$

$$\tilde{H}^x_{\mu\nu} = H^x_{\mu\nu} + V^x_{\mu\nu} + B\sum_i \left\langle \mu \middle| \varphi_i^h \right\rangle \left\langle \varphi_i^h \middle| \nu \right\rangle \tag{11}$$

Proper division of basis functions across broken bonds is achieved by adding a projection operator, $B\sum_i \left\langle \mu \middle| \varphi_i^h \right\rangle \left\langle \varphi_i^h \middle| \nu \right\rangle$, to the Hamiltonian, where the superscript *h* indicates a hybrid orbital as mentioned above. The constant *B* is chosen to be sufficiently large so as to remove any corresponding orbitals out of the

www.manaraa.com

variational space (B=10^6 a.u.). The last term in Eq. (11) is not used if no covalent bonds are broken. This occurs, for example, if the fragmentation occurs between molecules in a liquid.

The exact form of the ESP, $V_{\mu\nu}^{x}$, is:

$$V_{\mu\nu}^{x} = \sum_{K(\neq x)} (\mu_{\mu\nu}^{K} + \nu_{\mu\nu}^{K}) \tag{12}$$

$$\mu_{\mu\nu}^{K} = \sum_{A \in K} \langle \mu | (-Z_A / |\mathbf{r} - \mathbf{r_A}|) | \nu \rangle \tag{13}$$

$$\nu_{\mu\nu}^{K} = \sum_{\lambda\sigma \in K} D_{\lambda\sigma}^{K} (\mu\nu | \lambda\sigma) \tag{14}$$

The first term in Eq. (12) $\mu_{\mu\nu}^{K}$ represents the electron-nuclear attraction contribution to the total energy and the second term $\nu_{\mu\nu}^{K}$ represents the two-electron contribution, both expressed in terms of AOs μ and ν. Both terms are calculated in the presence of the surrounding monomers K with electron density D^K.

Savings in computational cost can be achieved through the use of approximations to the two-electron term in the ESP, Eq. (14). A user-defined cut-off value R_{ESPPC} is created, defined as the interfragment distance between two fragments I and J, outside of which the two-electron terms are treated with more approximate methods. This approximation is applied by comparing R_{ESPPC} with the relative minimum distance between fragments R_{IJ},

$$R_{IJ} = \min_{A \in I, B \in J} \left\{ \frac{|\mathbf{R}_A - \mathbf{R}_B|}{W_A + W_B} \right\} \tag{15}$$

where indices A and B run over atoms in fragments I and J respectively, and W_A and W_B are the corresponding van der Waals radii.[41] Due to the cancellation of units (Å) in Eq. (15), all of the user-defined and calculated values are unitless. Analogous cut-off values, R_{ESDIM} and R_{CORSD}, can also be defined by the user to approximate the SCF procedure using a Coulomb interaction and neglect the inclusion of electron correlation during dimer calculations when using methods such as MP2 and coupled cluster theory (e.g., CCSD(T)) respectively. In addition to these cut-off values, an array of four values, $\boldsymbol{R_{ITRIM}}$, can be specified to completely neglect the inclusion of separated trimer interactions.

www.manaraa.com

Two types of approximations to the ESP are employed using the cut-off values based on Eq. (15). The first approximation, applied at intermediate distances, is the Mulliken atomic orbital population approximation to the two electron integrals. [42, 43] Eq. (14) can be rewritten as

$$v_{\mu\nu}^{K} \cong \sum_{\lambda \in K} (\mathbf{D}^{K}\mathbf{S}^{K})_{\lambda\lambda} (\mu\nu \mid \lambda\lambda) \tag{16}$$

This approximation can effectively reduce the computational cost of calculating the two-electron integrals by a factor of the number of basis functions (N_B).

The second type of approximation, called the point charge approximation, is applied at long distances and further simplifies Eq. (15) as

$$v_{\mu\nu}^{K} \cong \sum_{A \in K} \left\langle \mu \left| \left(\frac{Q_A}{|\mathbf{r} - \mathbf{r}_A|} \right) \right| \nu \right\rangle \tag{17}$$

where Q_A are fractional atomic charges derived from the Mulliken atomic populations of the monomers. This approximation further reduces the computational cost of calculating the two-electron integrals by another factor of N_B.

Further refinements to the formalism have been performed to increase the accuracy of the method,[44] as well as the addition of a long list of functionality.[14, 45-54] Detailed descriptions of the FMO method formalism can be found in a number of review articles and books.[10, 55, 56 57]

3. Results

A sample of water cluster configurations was obtained using the EFP1 [58] implementation of the Monte Carlo algorithm with simulated annealing (MC/SA)[59, 60]. Minimum energy structures of 16 and 20 waters, were chosen from previous work[60], while isomers of 32 and 64 water molecules were obtained using the MC/SA approach. The starting temperature for the SA was set at 20,000 K and lowered to 300 K. Local minimizations were performed every 100 steps; the number of molecules moved per step was varied from one to four. Of the ~16,000 geometries sampled for both 32 waters and 64 waters, several unique isomers were chosen for each and re-optimized using the EFP1 method[58]. It should be noted that all optimizations using the EFP method use the same frozen internal geometry for each water molecule. The final sample consists of seven clusters each of 16, 20 and 64 water molecules and eight clusters of 32 water molecules.

The total electronic energy of the 16, 20 and 32 clusters was calculated at the MP2 level of theory using the 6-31G(d), 6-31++G(d,p) and 6-311++G(3df,2p) Pople-type basis sets. [61-63] The total electronic energy of the 64 water clusters was evaluated with the first two basis sets only. Calculations of these total energies, as well as

www.manaraa.com

all FMO calculations, were carried out using the General Atomic and Molecular Electronic Structure System (GAMESS) program package.[35, 36] All of the SMFA energies were evaluated by MAC and MAA using the GAUSSIAN03 program package.[64] SRP and MSG did not use, nor would they consider using, the aforementioned program package.

Since each water monomer has the same internal geometry in every cluster, the total binding energy of each cluster was easily evaluated by subtracting the gas phase monomer energy (multiplied by the appropriate number of monomers) from the total energy. To allow direct comparison between different sized clusters, the binding energy per water molecule is reported. The total energies for all clusters for MP2 calculations compared to the FMO and SMFA estimates are presented in the supplementary material Table S2.

3.1 Approximations

The SMFA energies are obtained by embedding the fragments in a background charge distribution. Table 1 presents examples of the accuracy of the SMFA estimate of the cluster binding energy when no background charges are used, or when the multipole expansion of the charge distribution is truncated at various orders. In the table, "two body only" denotes a Level 1 fragmentation into hydrogen bonded pairs of water molecules with no account of non-bonded Coulomb, induction or dispersion effects. Results for "no field" include non-bonded dispersion but no Coulomb or induction effects. The remaining results include background charge distributions that mimic atom centered multipoles up to second order. Table 1 indicates that inclusion of dipole and quadrupole moments is necessary to achieve reliable estimates of the cluster binding energy. In the results presented below, the background charge distribution follows the findings from Table 1 and is represented by charges (see Table S1) that mimic the atom centered distributed charges, dipole and quadrupole moments.

The FMO method employs two approximations to reduce the total number of explicit dimer and trimer calculations performed. Table 2 shows the total number of n-mer calculations to be performed for one isomer of 16, 20, 32 and 64 water molecules. The total number of trimer calculations to be performed increases rapidly, nearly an order of magnitude, on increasing the cluster size from 32 to 64 water molecules. Many of these n-mer interactions occur between far separated water molecules and can be approximated using one-electron Coulomb potentials, or neglected completely, by specifying the cut-off values R_{ESDIM} and $\boldsymbol{R_{ITRIM}}$ for dimer and trimer interactions, respectively. The keyword $\boldsymbol{R_{ITRIM}}$ is an array of four values. The first three of these array elements are used to determine whether a trimer SCF calculation is performed, while the fourth element specifies the cut-off value for the calculation of correlation contributions for trimers. The omission of electron correlation for far separated dimers is specified with the cut-off value R_{CORSD}. Approximations to the two-electron term in the ESP are defined using the cut-off value $R_{ESPPC.}$ For this work only the point charge approximation to the ESP was used. Due to the local nature of electron correlation, a great number of MP2 calculations can also be neglected. Table 3 shows the reduction in dimer and trimer calculations, with

www.manaraa.com

approximated interactions labelled as “separated” and the total number of explicit HF and MP2 calculations performed labelled as “SCF” and “correlated” respectively.

It becomes apparent that for small clusters of fewer than 20 water molecules, the total number of n-mer calculations is not significantly reduced, even for the most aggressive cut-off approximation used. Figure 1 shows the average error by cut-off value employed for all three basis sets and all cluster sizes. It can be seen that reducing the number of n-mer calculations for these small clusters has little effect on the accuracy of the calculation for FMO2 and FMO3. Indeed, the level of error remains relatively low and unchanged for all three basis sets for these small clusters. The error begins to increase with the 32- and 64-water clusters, particularly for FMO2 where the average error for these cluster sizes jumps to over 10 kcal/mol. However, even though all dimer calculations are retained for the clusters of 32 water molecules, and nearly all dimer calculations are retained for the 64-water clusters, the level of error does not decrease significantly. This lack of error reduction implies that the error incurred from using the FMO method for these calculations is not due to a reduction in the number of dimer calculations that are included for large clusters.

The inclusion of trimer interactions via FMO3 provides a substantial reduction in the average error, shown in Figure 1d-f. As was the case for the FMO2 approximations being used for clusters of 20 waters or less, the inclusion of all three-body interactions has little effect on the average error. An improvement in the average error can be seen in Figure 1e for the clusters of 32 water molecules, but more dramatically in the average error of the 64 water clusters.

For consistency and to achieve the most accurate energies for the largest water clusters, all results shown hereinafter employ the most relaxed level of approximation (A3), with the fewest number of excluded calculations. This corresponds to an average reduction of approximately 200 correlated dimer and 2000 correlated trimer calculations for the clusters of 64 water molecules. For calculations with fewer than 32 water molecules, however, the authors recommend values of R_{ESDIM}=3.5 and R_{CORSD}=3 for FMO2 calculations and $\boldsymbol{R_{ITRIM}}$=3,-1,3.5,3 for FMO3 calculations, using the same R_{ESDIM} and R_{CORSD} values as FMO2. Due to the importance of many-body interactions in water clusters, any clusters of 64 water molecules or larger necessitate the inclusion of as many dimer and trimer interactions as possible.

3.2 Binding Energy per Molecule

As a first criterion for comparison between the FMO and SMFA methods with fully *ab initio* results, the binding energy per molecule was calculated for all clusters and basis sets for both methods and compared to *ab initio* results. The mean absolute errors in the binding energies for all clusters and basis sets are shown in Table 4. Average errors for each size cluster are shown in Tables 5 and 6 for the FMO method and the SMFA method, respectively. The binding energies per molecule for both the FMO and SMFA methods are also plotted in

www.manaraa.com

Figures 2-6. This data is characterized by distinct groups resulting from the substantial reduction in the magnitude of the binding energy with increasing basis set size.

For both methods, the error in average binding energy per water molecule for all clusters and basis sets falls between 0.4% and 4.3%, as shown in Table 4. While the errors for individual clusters are dependent on both the particular cluster size and the basis set chosen, Table 4 indicates that as a whole the average errors for FMO2 and SMFA Level 1 are comparable, with errors around 4.3%. Similar conclusions can be made for the agreement of FMO3 and SMFA Level 3 with errors between 0.4 and 0.5%. SMFA Level 2 appears to fall somewhere between FMO2 and FMO3 and between SMFA Levels 1 and 3, giving errors of approximately 1.3%. While errors of 4.3% are perhaps too large to be useful, the errors for FMO3 and SMFA Level 3 are very small and within chemical accuracy. Based upon the data in Table 4, SMFA Level 2 appears to be a reasonable compromise approach, providing an acceptable level of error. It is also apparent that the level of error incurred by including only pair interactions is unreliable. Water clusters require at a minimum some level of higher order many-body interactions in order to provide reliable binding energies.

The results for FMO2 (Figure 2) show a trend of increasing error in the binding energy compared to *ab initio* results as the size of the basis set increases. While there is an increase in the magnitude of the errors, the size of the error between different isomers of each cluster size corresponding to a given basis set is very consistent. This is illustrated by the relatively close grouping of each cluster of dots in Figure 2. SMFA Level 1 errors (Figure 4) appear to be much more random, with the errors also growing worse with an increase in basis set size. The consistent level of error in predicted FMO2 binding energies versus the random errors in SMFA Level 1 manifests in the ability of FMO2 to provide a better estimate of the relative energies of different cluster isomers than can be achieved with SFMA Level 1. However, both FMO2 and SMFA Level 1 reproduce the fully *ab initio* average binding energy per molecule to within 0.5 kcal/mol.

Further comparisons between FMO2 and SMFA Level 1 are shown in Tables 5 and 6. Table 5 shows the average FMO errors in the binding energy per water molecule as a function of basis set and cluster size, while Table 6 contains the corresponding SMFA results. Table 5 illustrates a slight increase in the FMO2 error as the cluster size increases. This effect becomes more pronounced with the larger basis sets. In contrast, SMFA Level 1 does not exhibit a consistent increase in error with either cluster size or basis set size. Compared to FMO2, the SMFA Level 1 errors are not nearly as consistent between cluster size and isomers, and are therefore more difficult to interpret. In general it can be said that, in addition to the inability of the pair interactions to properly describe water clusters, an increase in basis set size with the FMO2 method appears to place even more significance on the higher order many body interactions that are not explicitly included in FMO2. For the SMFA Level 1 method, there is also a trend of increasing error with increasing cluster size, especially for the largest basis set. This trend leads to the same conclusions about the importance of higher order many body interactions with increasing cluster size.

www.manaraa.com

FMO3 and SMFA Levels 2 and 3 contain higher order many body interactions. Figure 3 shows the FMO3 errors in binding energy per molecule. It is readily apparent that the use of FMO3 nearly eliminates the error incurred through the use of only pair interactions. As is the case for FMO2, FMO3 also provides extremely consistent errors for all clusters and basis sets. The binding energy is underestimated slightly when using the small 6-31G(d) basis set, however the errors are still quite small. SMFA Level 2 results, shown in Figure 5, also suffer from some of the random errors found for SMFA Level 1, but to a much smaller degree. The accuracy of SMFA Level 2, on average, is reasonable, as illustrated by the lack of any exceptionally large errors (see Figure 5). For SMFA Level 3 (Figure 6) the predicted binding energies/water molecule appear to be very close to those predicted by FMO3, including the consistent nature of the errors.

3.3 Relative Energies

A more important metric to gauge the quality of results provided by either the FMO or the SMFA method is how accurately the relative energies of the different cluster isomers are reproduced. Figures 7 and 8 show the FMO and SMFA relative energies, respectively, for all cluster sizes and basis sets. Each figure consists of eleven graphs, a through k, with the first four showing results for the 6-31G(d) basis set, the next four showing results for the 6-31++G(d,p) basis set and the last three showing results for the 6-311++G(3df,2p) basis set. Each set of graphs displays the results in order of increasing cluster size for 16, 20, 32 and 64 water molecules, with the exception of the graphs for the 6-311++G(3df,2p) basis set which only shows results for 16, 20 and 32 water molecules. In some instances, the errors for FMO2 and SMFA Level 1 are very large, requiring the use of two different ordinate axes displaying the FMO2 or SMFA Level 1 energies on the right ordinate while displaying the *ab initio*, FMO3 and SMFA Levels 2 and 3 energies on the left ordinate. All energies are in kcal/mol and all cluster geometries are shown in Figures S1 through S4.

The results for the FMO method in Figure 7 provide additional evidence that the use of only pair interactions, while insufficient for absolute energies, reproduces the relative energies for all cluster sizes and basis sets with remarkable accuracy. In all cases the FMO2 relative energies closely mirror the relative energies of the *ab initio* calculations, with only a few discrepancies appearing for the 6-31++G(d,p) basis set for the 64-water clusters and the 6-311++G(3df,2p) basis set for 16 and 20 water molecules. As expected, the addition of three body interactions dramatically improves the overall accuracy when compared to *ab initio* energies. In agreement with previous conclusions based on the binding energy per molecule, the absolute errors for FMO3 decrease with an increase in the size of the basis set, but the errors increase slightly with an increase in cluster size. These errors in the absolute energies have little effect on the relative energies other than a corresponding shift along the ordinate, with FMO3 closely mirroring the *ab initio* curve in nearly all cases. There are, however, a few notable discrepancies, with increases in error for certain isomers of 16 and 20 water molecules for the 6-31++G(d,p) and 6-311++G(3df,2p) basis sets. These discrepancies can be attributed to unique, highly symmetric structural arrangements of the smaller clusters that exhibit unusually large electron delocalization. Such species are not present in the larger, more globular clusters of 32 and 64 water molecules.

www.manaraa.com

Figure 8 follows the same layout conventions as Figure 7, with certain graphs containing two ordinate axes in order to better illustrate the relative energetics compared to *ab initio*. In contrast to the FMO method, the use of only pair interactions in SMFA Level 1 fails to reproduce accurate relative energies in most cases. Increasing to SMFA Level 2 provides a marked improvement, particularly for smaller clusters and basis sets. However, as the basis set is increased the agreement begins to decline. This is particularly true for the energy of the 16-water s43 isomer. In most other cases, however, the agreement is adequate. Further increasing the accuracy to SMFA Level 3 eliminates all spurious errors such as those found for the s43 isomer. Agreement with *ab initio* results for SMFA Level 3 is good, and in some cases better than the relative energies produced by FMO3.

Further investigation of the improvement in agreement upon going from SMFA Level 2 to Level 3 can be found in the supplementary information (Table S3). There it is shown that the contribution of non-bonded dispersion is more significant for the larger, 32- and 64-water clusters. For SMFA at Level 2, the non-bonded dispersion lowers the binding energy per molecule by about 0.3 kcal/mol. This contribution falls to about 0.07 kcal/mol, at Level 3. It appears, therefore, that the dominant non-bonded dispersion interactions are between those molecules that are not connected in sub-clusters at Level 2, but are so connected at Level 3. The results for SMFA Levels 2 and 3 are consistent with those obtained using a different fragmentation scheme by Li and co-workers.[26] Although that work did not include non-bonded dispersion effects, the sub-clusters considered are comparable in size to those for Level 3, so that the total contribution of long range dispersion is not large.

3.4 Timings

One of the advantages of the SMFA approach is the ease of trivial parallelism, achieved by calculating each non-interacting fragment on a separate CPU. It is illustrative to consider the number and size of fragments produced by fragmentation. Tables 7 and 8 show the distribution of fragment sizes for each SMFA Level 2 and SMFA Level 3 cluster, respectively. The total number of fragments produced by the SMFA method is shown in the final column. For SMFA Level 2, the total number of fragments is approximately constant up to a cluster size of 32. Doubling the size of the cluster from 32 to 64 water molecules increases the number of fragments from 125 to 356, roughly a factor of three. It is important to note, however, that the majority of the increase is from the smallest fragments. For example, while the number of 4-water fragments doubles, the number of 2- and 3-water fragments approximately triples, and the number of the latter is greater to begin with.

The FMO method has similar advantages when it comes to parallelism, including the ability to calculate individual fragments, as well as dimers and trimers independently of one another. This allows for two levels of parallelism through the use of the GDDI in GAMESS. FMO calculations that employ the GDDI assign individual fragment n-mer calculations to separate nodes or “groups”, with each group performing its assigned fragment calculation in parallel. Timings for energy calculations at the MP2 level of theory using the 6-31++G(d,p) basis set and clusters of up to 32 water molecules are shown in Table 9. The timing calculations

www.manaraa.com

were performed on six nodes containing two quad-core 2.66 GHz Intel Xeon processors and 16 GB of RAM per node. The FMO calculations used GDDI with one node per group for a total of 6 groups.

It is apparent that the wall time for the *ab initio* calculations increases by a factor of greater than 25 when doubling the system size from 16 to 32 water molecules. In contrast to this quickly increasing wall time, the FMO2 calculations provide a time savings between 97 and 99% of the *ab initio* calculations with errors between 2.65 and 9.30 kcal/mol. In order to obtain reasonable errors, however, FMO3 must be used with the current fragmentation scheme. The use of FMO3 reduces the errors to between 0.28 and 1.73 kcal/mol while still providing a time savings between 77 and 96% compared to the *ab initio* timings.

4. Conclusions

It has been shown that the FMO and SMFA methods both accurately estimate the MP2 energies of water clusters. In contrast, the pair-wise interaction model provides a very poor estimate of binding in water clusters. On average the mean errors of the FMO methods are remarkably close to the corresponding mean errors for SMFA methods. However, the FMO method tends to give systematic, rather than random, deviations from the exact cluster energies. This means that if isomerization energies are important for the physical measurement of interest, then the FMO method would be more reliable than the SMFA method.

The computational effort required for the SMFA method increases nearly linearly with the cluster size, since the number of sub-clusters increases approximately linearly with the size of the whole cluster. Level 2 fragmentation typically results in sub-clusters of five (and less) water monomers, while Level 3 typically results in sub-clusters of eight to nine (and less) water monomers. Moreover, Level 3 fragmentation typically produces a much greater number of fragments. Hence, Level 3 SMFA is significantly more computationally expensive than Level 2. Level 2 SMFA produces an average error of about 0.15 kcal/mol per water monomer (about 1.3% of the total), which would appear to be sufficiently accurate for many applications. It is important to note that non-bonded dispersion must be accounted for if Level 2 fragmentation is used.

The computational effort for FMO also increases nearly linearly for FMO2, while the computational effort of FMO3 increases by a factor of ~4.8. This deviation from nearly linear scaling is due to the quickly increasing number of three body interactions that are required with increasing cluster size. Despite this increase in three body calculations required, the time savings obtained from FMO3 when compared to *ab initio* calculations also increases with increasing system size (77 to 96% with a doubling of system size). The accuracy of FMO2 is more comparable to the SMFA Level 1 method, with an average error of about 0.46 kcal/mol per water monomer. This level of error is clearly unacceptable for many applications, despite the accurate reproduction of the relative energies between water cluster isomers. To produce an acceptable level of error FMO3 must be used, producing errors in agreement with SMFA Level 3 but significantly increasing the computational cost. It is important point out that the accuracy of FMO2 can be increased for single point

www.manaraa.com

energy calculations by placing more than one water molecule per fragment, however this type of fragmentation scheme can become problematic for optimizations and molecular dynamics (MD) simulations where individual water molecules may move apart.

MD simulations in aqueous solution are at present primarily carried out using classical force fields. Simulations of water (and solutes in water) require a large number of evaluations of the energy and energy gradient of the whole simulation system. As noted here and elsewhere, the water monomer-monomer interaction fails to account for a large percentage of the water binding energy per molecule, and so does not provide the basis for a quantitatively accurate study of water dynamics, even though pair models are frequently used.[65-67] FMO and SMFA appear to provide a quantitative approach to the energetics of water dynamics; the challenging task is to incorporate these approaches into a computationally feasible algorithm for dynamical simulations.

References

1. P. Ball, *Nature*, 2008, **452**, 291.
2. D. M. Bates and G. S. Tschumper, *J. Phys. Chem. A*, 2009, **113**, 3555.
3. S. Yoo, E. Apra, X. C. Zeng and S. Xantheas, *J. Phys. Chem. Lett.*, 2010, **1**, 3122.
4. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357.
5. T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel and H. B. Frisch, *J. Comput. Chem.*, 2003, **24**, 760.
6. P. Canfield, M. G. Dahlbom, N. S. Hush and J. R. Riemers, *J. Chem. Phys.*, 2006, **124**, 024301.
7. E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 46.
8. M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys and W. J. Stevens, *J. Phys. Chem. A*, 2001, **105**, 293.
9. M. S. Gordon, L. Slipchenko, H. Li and J. H. Jensen, *Ann. Rep. Comp. Chem.*, 2007, **3**, 177.
10. 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.
11. F. Cortés-Guzmán and R. F. W. Bader, *Chem. Phys. Lett.*, 2003, **379**, 183.
12. D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599.
13. W. Li and S. Li, *J. Chem. Phys.*, 2005, **122**, 194109.
14. D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **121**, 2483.
15. D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2005, **122**, 194109.
16. K. Babu and S. R. Gadre, *J. Comput. Chem.*, 2003, **24**, 484.
17. W. Yang and T. Lee, *J. Chem. Phys.*, 1995, **103**, 5674.

www.manaraa.com

18. Y. Mei, C. Ji and J. Z. H. Zhang, *J. Chem. Phys.*, 2006, **125**, 094906.

19. X. H. Chen, D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2004, **120**, 839.

20. X. H. Chen and J. Z. H. Zhang, *J. Chem. Phys.*, 2004, **120**, 11386.

21. X. H. Chen and J. Z. H. Zhang, *J. Theor. Comp. Chem.*, 2004, **3**, 277.

22. V. Deev and M. A. Collins, *J. Chem. Phys.*, 2005, **122**, 154102.

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

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

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

26. W. Li, S. Li and Y. Jiang, *J. Phys. Chem. A*, 2007, **111**, 2193.

27. J. M. Mullin, L. B. Roskop, S. R. Pruitt, M. A. Collins and M. S. Gordon, *J. Phys. Chem. A*, 2009, **113**, 10040.

28. C. Møller and S. Plesset, *Phys. Rev.*, 1934, **46**, 618.

29. E. Apra, A. P. Rendell, R. J. Harrison, V. Tipparaju, W. A. D. Jong and S. S. Xantheas, Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (SC '09), Portland, Oregon, 2009.

30. M. A. Collins, *to be published*, 2011.

31. A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128.

32. K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **313**, 701.

33. D. G. Fedorov, R. M. Olsen, K. Kitaura, M. S. Gordon and S. Koseki, *J. Comput. Chem.*, 2004, **25**, 872.

34. T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa and S. Sekiguchi, *Proc. of Supercomputing 2005, IEEE Computer Society*, 2005.

35. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.

36. M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry, the First Forty Years*, eds. C. E. Dykstra, G. Frenking, K. S. Kim and G. Scuseria, Elsevier, Amsterdam, 2005.

37. W. England and M. S. Gordon, *J. Am. Chem. Soc.*, 1971, **93**, 4649.

38. J. H. Jensen and M. S. Gordon, *J. Phys. Chem.*, 1995, **99**, 8091.

39. D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **120**, 6832.

40. (a) D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2004, **389**, 129; (b) T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, S. Tanaka, Chem. Phys. Lett., 2012, **523**, 128.

41. T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, U. M and K. Kitaura, *Chem. Phys. Lett.*,

www.manaraa.com

2002, **351**, 475.

42. K. Ruedenberg, *J. Chem. Phys.*, 1951, **19**, 1433.

43. A. L. Companion and R. G. Parr, *J. Chem. Phys.*, 1961, **35**, 2268.

44. D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2006, **433**, 182.

45. D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2005, **122**, 054108.

46. D. G. Fedorov, T. Ishida and K. Kitaura, *J. Phys. Chem. A*, 2005, **109**, 2638.

47. Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura and T. Nakano, *Chem. Phys. Lett.*, 2005, **406**, 283.

48. D. G. Fedorov, K. Kitaura, H. Li, J. H. Jensen and M. S. Gordon, *J. Comput. Chem.*, 2006, **27**, 976.

49. D. G. Fedorov and K. Kitaura, *J. Comput. Chem.*, 2007, **28**, 222.

50. M. Chiba, D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2007, **444**, 346.

51. M. Chiba, D. G. Fedorov and K. Kitaura, *J. Comput. Chem.*, 2008, **29**, 2667.

52. T. Nagata, D. G. Fedorov, K. Kitaura and M. S. Gordon, *J. Chem. Phys.*, 2009, **131**, 024101.

53. S. R. Pruitt, D. G. Fedorov, K. Kitaura and M. S. Gordon, *J. Chem. Theory and Comput.*, 2010, **6**, 1.

54. C. Steinmann, D. G. Fedorov and J. H. Jensen, *J. Phys. Chem. A*, 2010, **114**, 8705.

55. D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 6904.

56. D. G. Fedorov and K. Kitaura, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC Press, Boca Raton, 2009.

57. M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. Slipchenko, *Chem. Rev.*, in press.

58. P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch and D. Cohen, *J. Chem. Phys.*, 1996, **105**, 1968.

59. N. Metropolis, A. Reosenbluth and A. Teller, *J. Chem. Phys.*, 1953, **21**, 1087.

60. P. N. Day, R. Pacher, M. S. Gordon and G. N. Merrill, *J. Chem. Phys.*, 2000, **112**, 2063.

61. W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257.

62. M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654.

63. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213.

64. G. W. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. Montgomery, J. A., T. Vreven, K. N. Kudin, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill,

www.manaraa.com

B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2003.

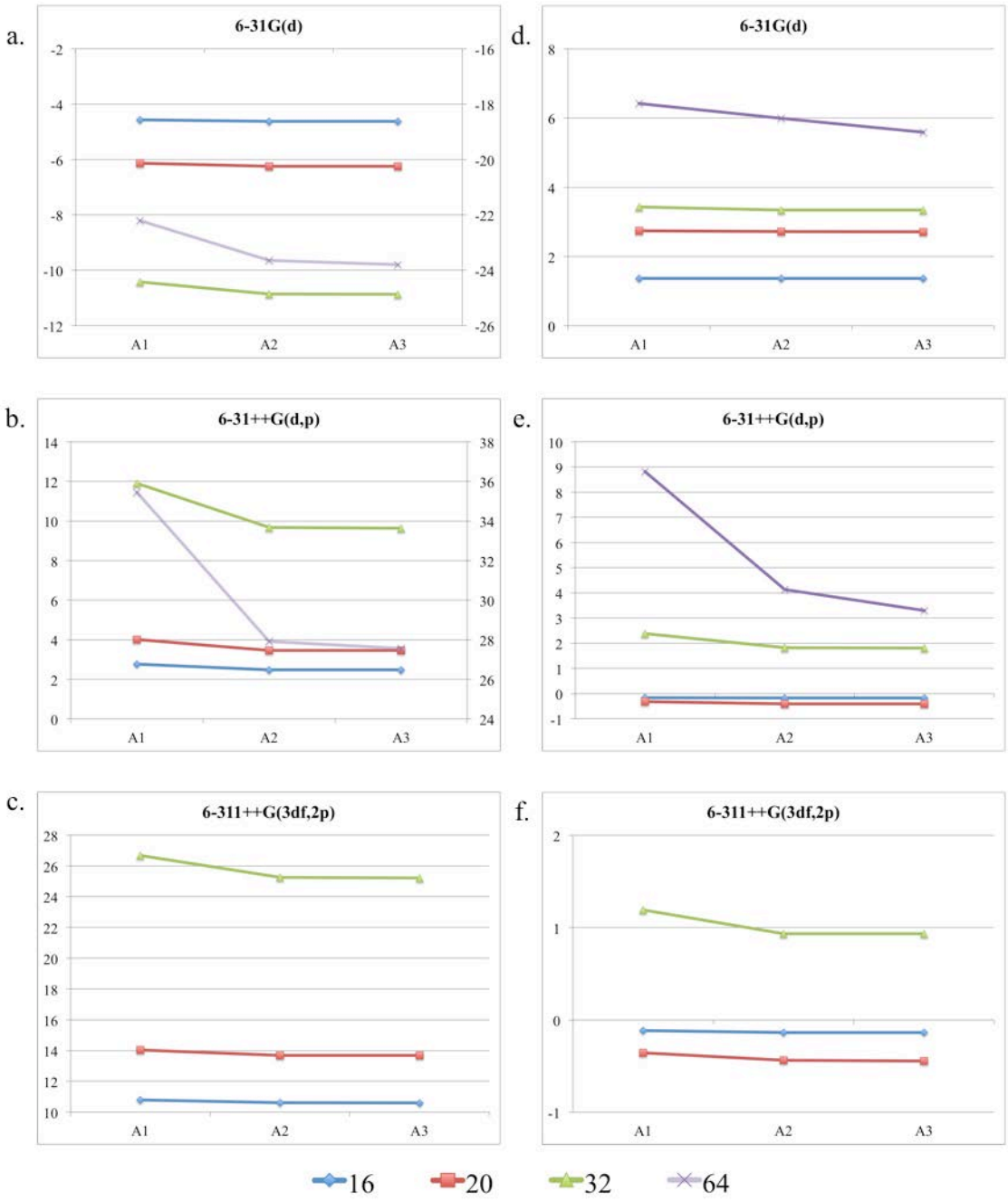
65. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Clarendon Press, Oxford, 1987.

66. A. Wallqvist and R. D. Mountain, *Rev. Comp. Chem.*, 1999, **13**, 183.

67. W. L. Jorgensen and J. Tirado-Rives, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 6665.

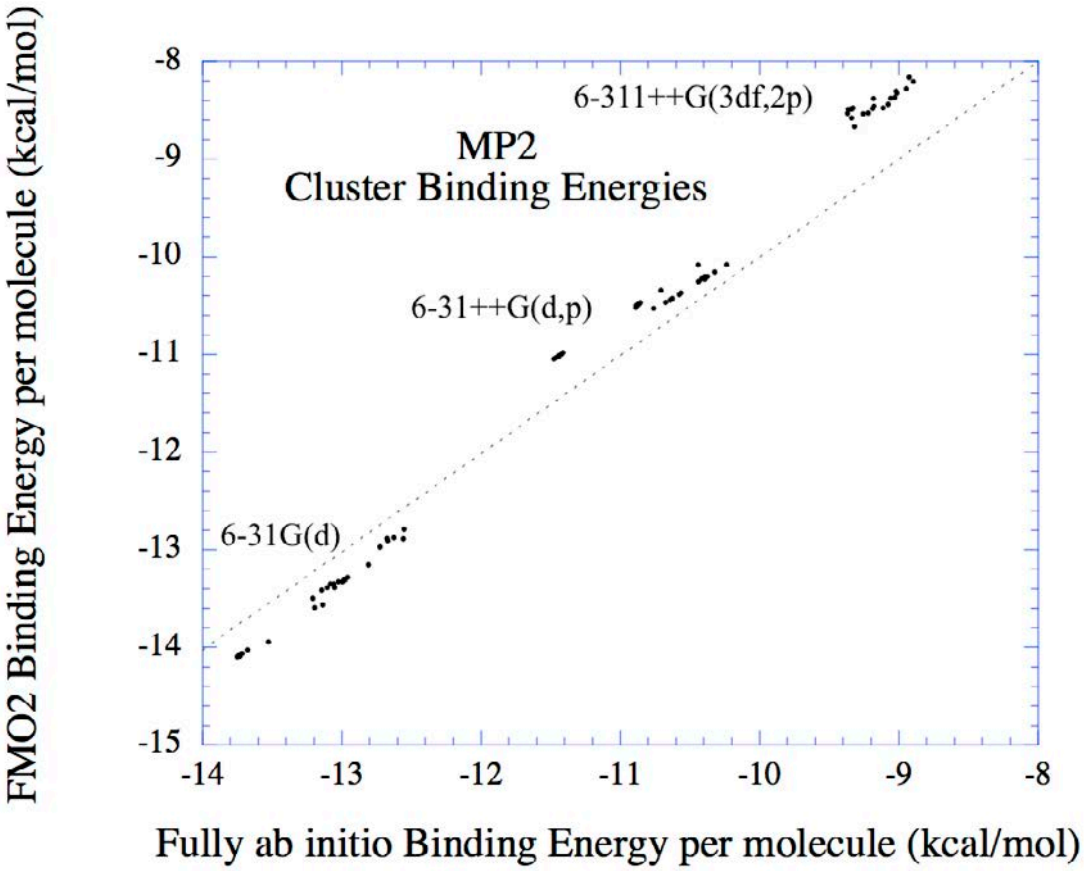
www.manaraa.com

Figure 1. Average error according to basis set and cut-off values described in Table 3. Approximation labels A1, A2, A3 correspond to the following cut-off values: (RESDIM/RCORSD) for FMO2 graphs a-c: A1 = 2.5/2 A2 = 3.5/3 A3 = 4.5/4. Cut-off values (RITRIM) corresponding to FMO3 graphs d-f: A1 = 2,-1,2.5,2 A2 = 3,-1,3.5,3 A3 = 4,-1,4.5,4. Two ordinate axes are used in cases of unusually high errors (graphs a and b) with the errors for clusters of 64 water molecules shown on the right ordinate.



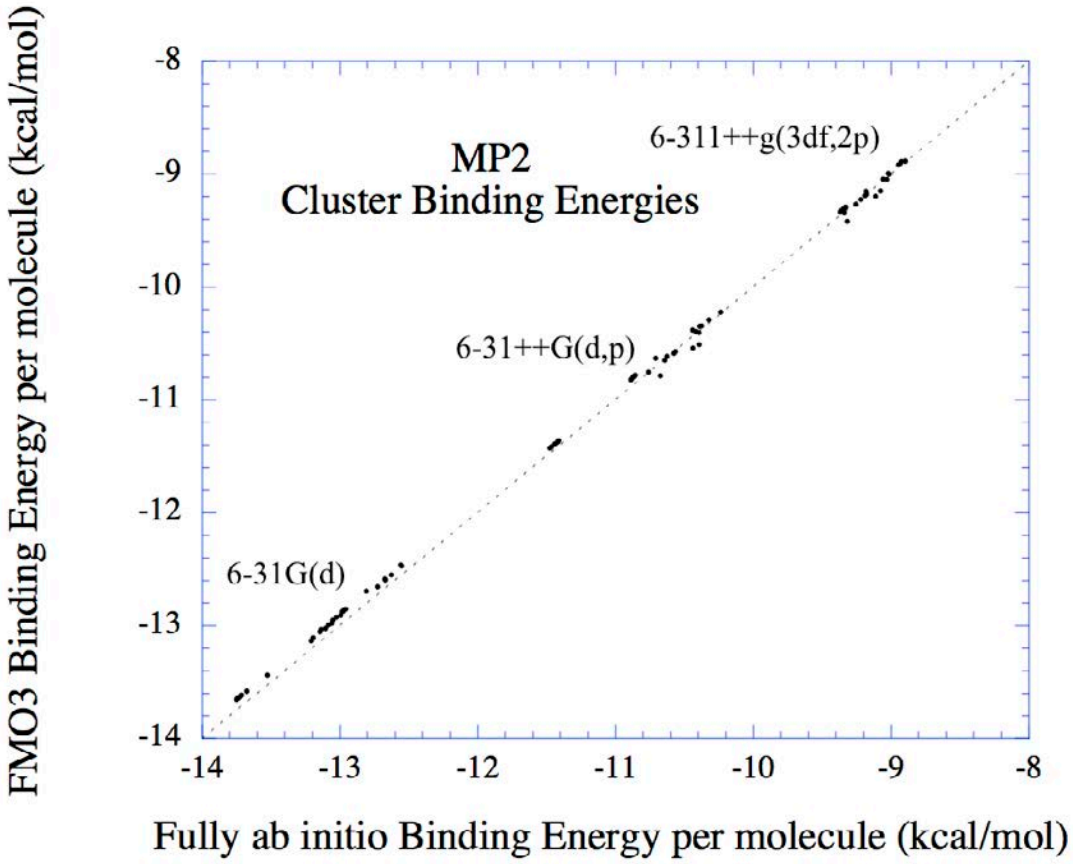
www.manaraa.com

Figure 2. The binding energy (per molecule) for each cluster and basis set using the FMO2 method is shown versus the corresponding "exact" value calculated for the whole cluster. The dotted line is merely a visual aid indicating exact agreement.



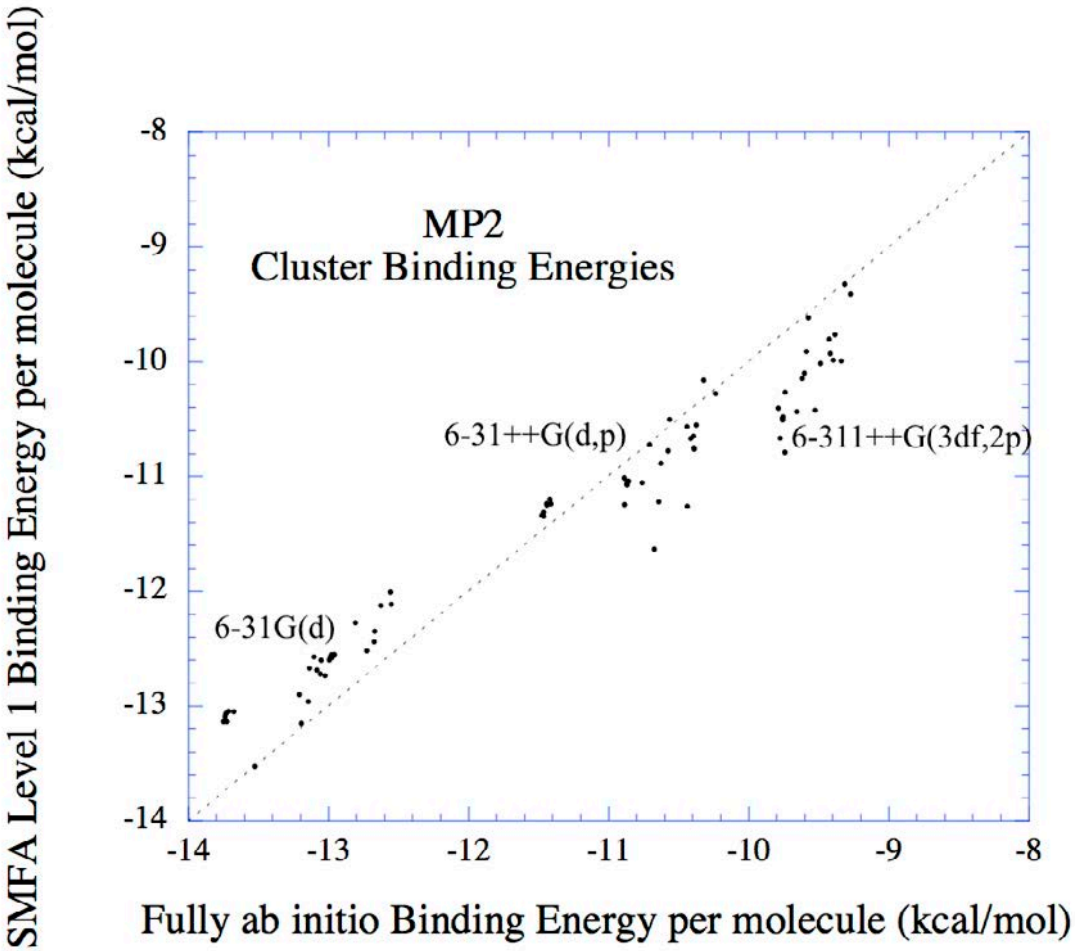
www.manaraa.com

Figure 3. The corresponding data to Figure 2 for the FMO3 method.



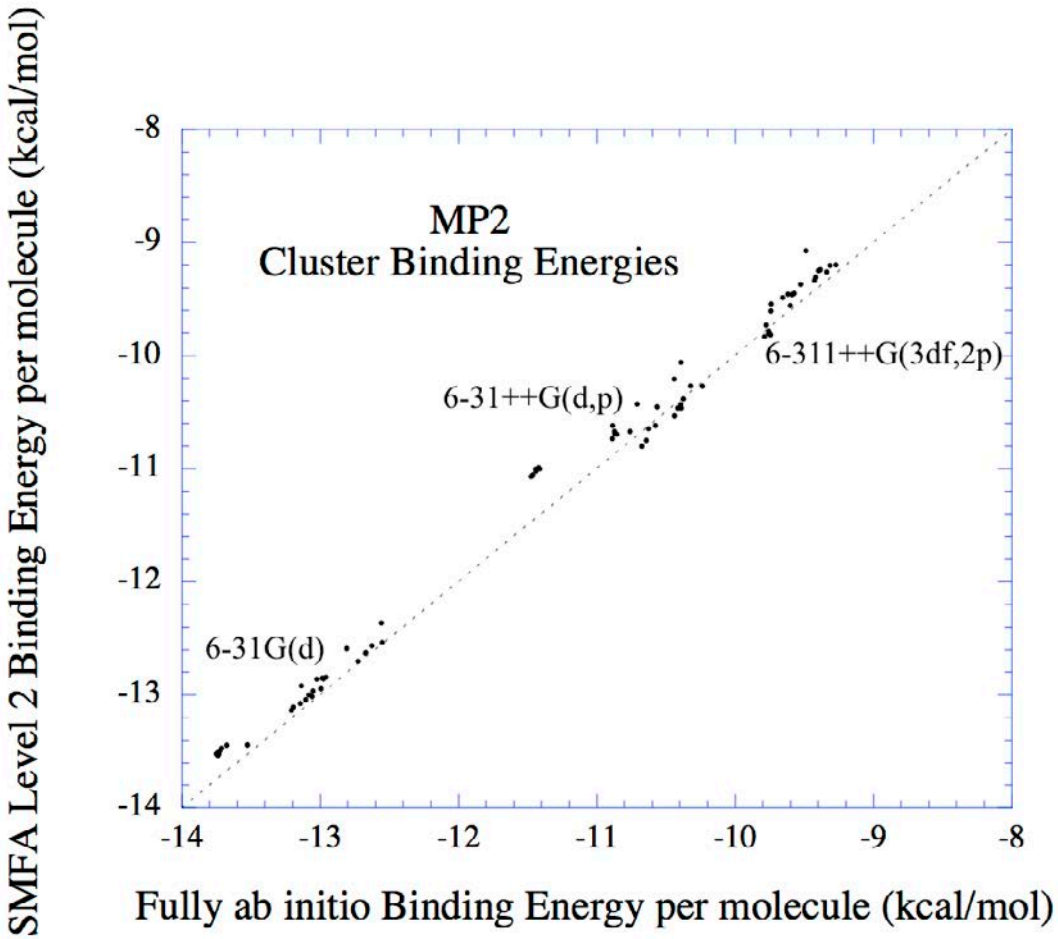
www.manaraa.com

Figure 4. The corresponding data to Figure 2 for the SMFA Level 1 method.



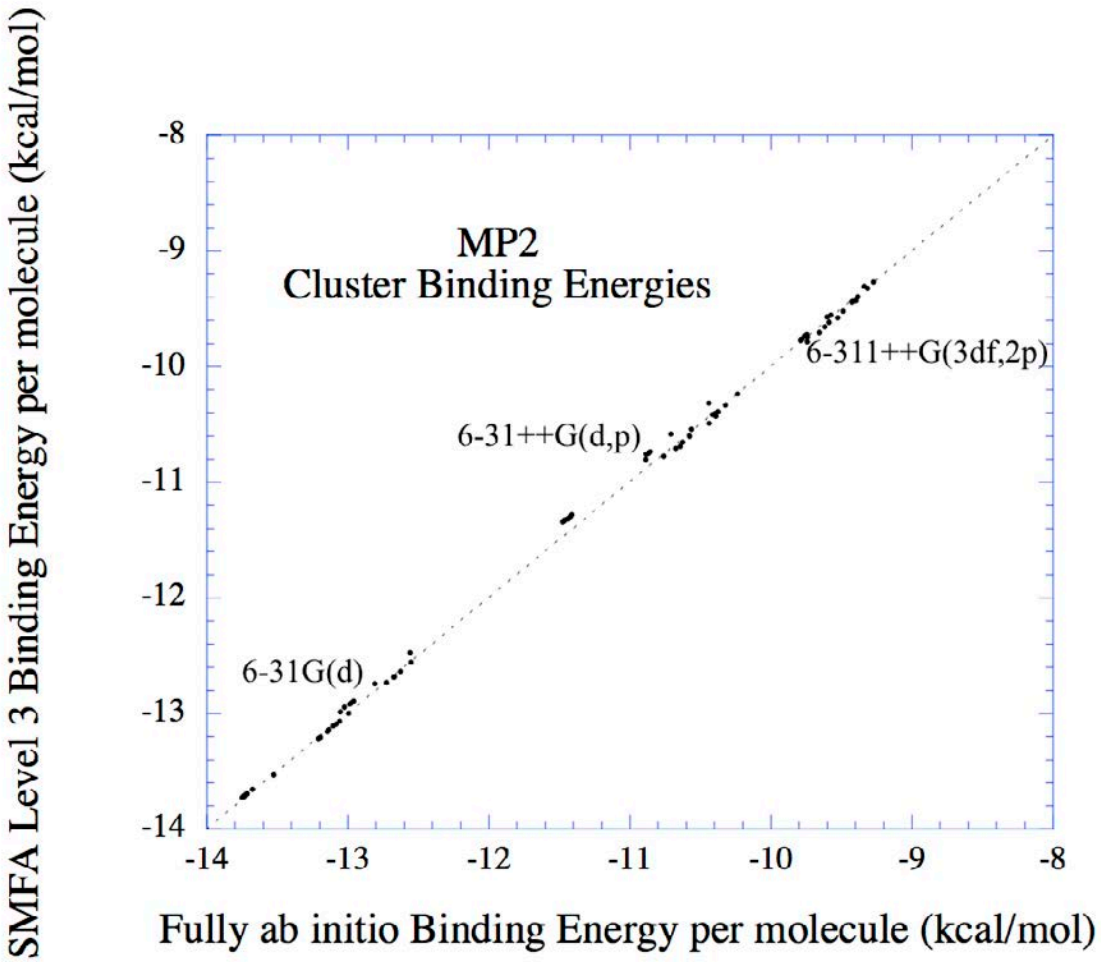
www.manaraa.com

Figure 5. The corresponding data to Figure 2 for the SMFA Level 2 method.



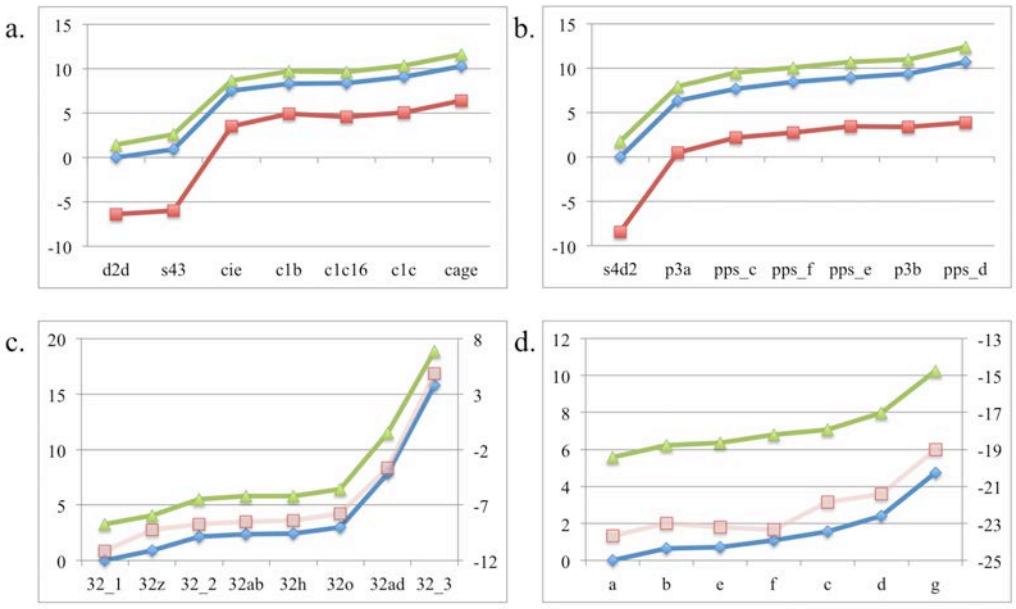
www.manaraa.com

Figure 6. The corresponding data to Figure 2 for the SMFA Level 3 method.



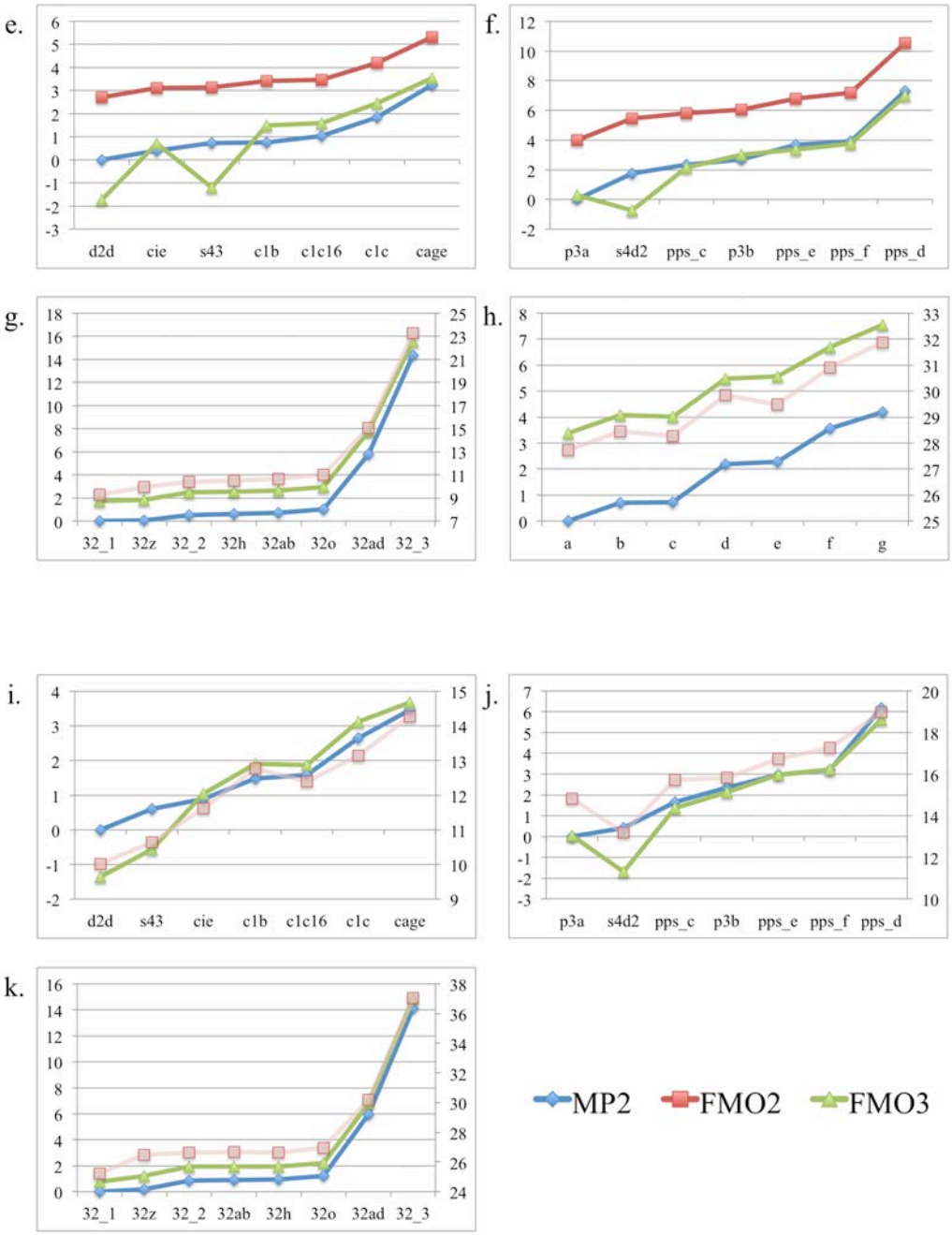
www.manaraa.com

Figure 7. Comparison of relative energies between fully ab initio MP2, FMO2-MP2 and FMO3-MP2 for all cluster sizes and basis sets. Graphs a-d show relative energies for 16, 20, 32 and 64 water clusters respectively for the 6-31G(d) basis set. Following the same trend in cluster size, graphs e-h correspond to the 6-31++G(d,p) basis set results and graphs i-k correspond to the results using the 6-311++G(3df,2p) basis set. In graphs c, d and g-k the FMO2 energy scale is shown on the right ordinate while the energy scale for ab initio and FMO3 results is shown on the left ordinate. Isomers are represented on the abscissa of each graph and all energies are in kcal/mol.



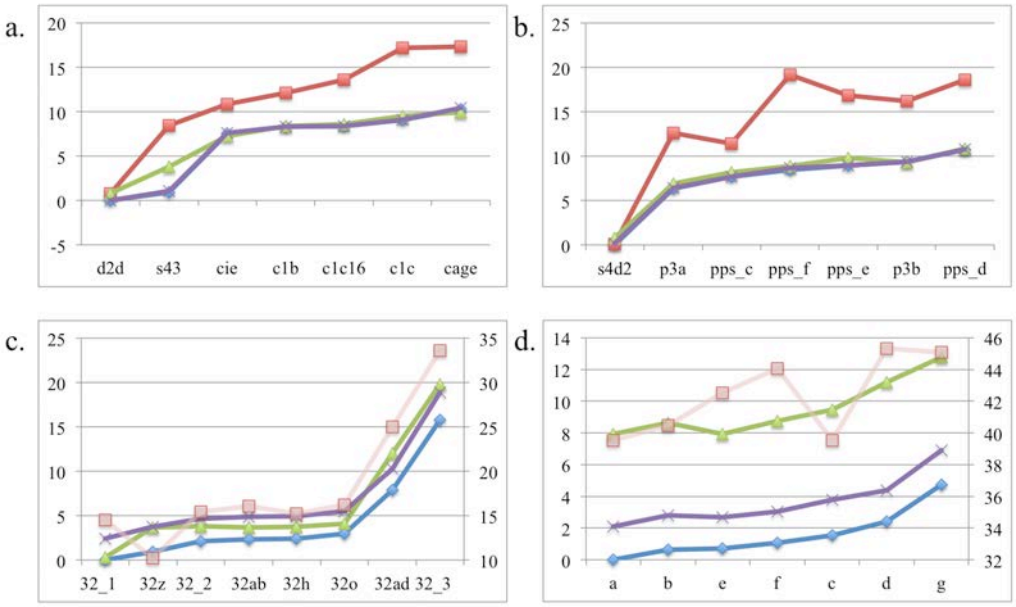
www.manaraa.com

Figure 7. (continued)



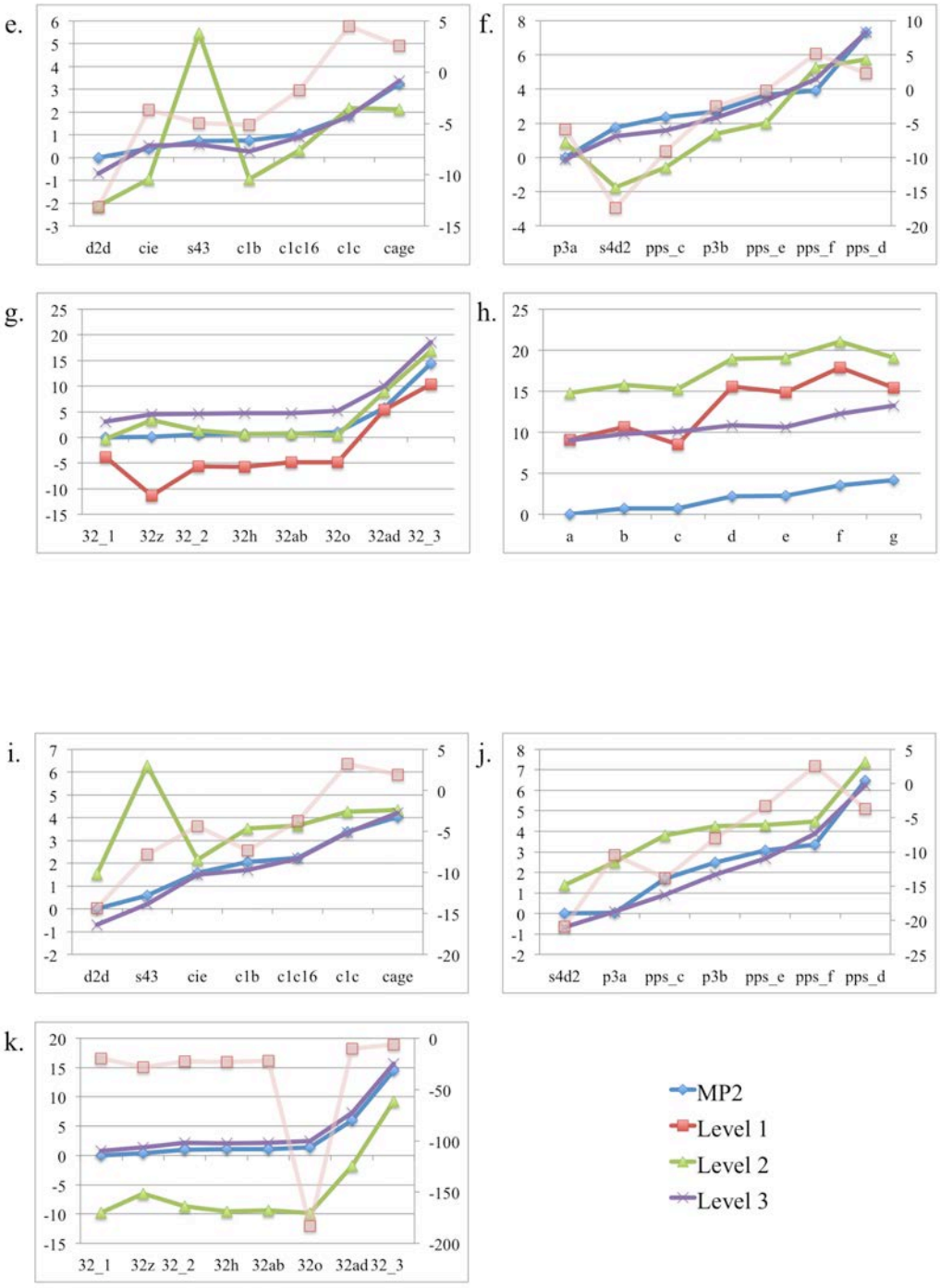
www.manaraa.com

Figure 8. Comparison of relative energies between fully ab initio MP2 and SMFA Levels 1, 2 and 3 for all cluster sizes and basis sets. Graphs a-d show relative energies for 16, 20, 32 and 64 water clusters respectively for the 6-31G(d) basis set. Following the same trend in cluster size, graphs e-h correspond to the 6-31++G(d,p) basis set results and graphs i-k correspond to the results using the 6-311++G(3df,2p) basis set. In graphs a-f and i-k the SMF Level 1 energy scale is shown on the right ordinate while the energy scale for ab initio and SMF Level 2 and Level 3 results is shown on the left ordinate. Isomers are represented on the abscissa of each graph and all energies are in kcal/mol.



www.manaraa.com

Figure 8. (continued)



www.manaraa.com

Table 1. The average percentage error in the binding energy for clusters of 64 water molecules calculated at SMF Level 1. Results are given for the two-body interaction with fragmentation plus long range dispersion and long range electrostatic interactions (see text for details).

Method	MP2/6-31G(d)	MP2/6-31++G(d,p)
Two body only	14.3	46.1
No Field	13.5	41.2
Charges	6.0	21.1
Charges + dipoles	6.1	5.4
Charges + dipoles + quadrupoles	6.4	1.5

www.manaraa.com

Table 2. Total number or dimers and trimers potentially evaluated without the use of approximations using FMO2 and FMO3, for examples of each cluster size.

Cluster size	Total number of n-mers	
	dimers	trimers
16	120	560
20	190	1140
32	496	4960
64	2016	41664

www.manaraa.com

Table 3. Number of dimer and trimer calculations actually performed based on the FMO approximation cut-offs (RESDIM, RCORSD for dimers and RITRIM for trimers).

	n-mer calculations based on cut-off thresholds					
	dimers			trimers		
	RESDIM=2.5 RCORSD=2	RESDIM=3.5 RCORSD=3	RESDIM=4.5 RCORSD=4	RITRIM 2,-1,2.5,2	RITRIM 3,-1,3.5,3	RITRIM 4,-1,4.5,4
W16						
separated	11	0	0	-	-	-
SCF	109	120	120	541	560	560
correlated	92	120	120	448	560	560
W20						
separated	29	0	0	-	-	-
SCF	161	190	190	1034	1140	1140
correlated	127	187	190	776	1140	1140
W32						
separated	128	7	0	-	-	-
SCF	368	489	496	3705	4949	4960
correlated	259	454	496	2369	4797	4960
W64						
separated	1047	417	71	-	-	-
SCF	969	1599	1945	14898	34377	41187
correlated	599	1299	1827	7363	26797	39700

www.manaraa.com

Table 4. Average errors in the binding energy per water molecule.

Method	Mean absolute error (kcal mol^{-1})	Mean absolute error (percentage)
FMO2	0.43	4.1
FMO3	0.06	0.5
SMFA Level 1	0.46	4.3
SMFA Level 2	0.15	1.3
SMFA Level 3	0.04	0.4

www.manaraa.com

Table 5. Average absolute errors in the MP2 binding energy per water molecule by basis set for FMO2 and FMO3 (kcal/mol).

	6-31G(d)		6-31++G(d,p)		6-311++G(3df,2p)	
Cluster Size	FMO2	FMO3	FMO2	FMO3	FMO2	FMO3
16	0.29	0.09	0.18	0.05	0.68	0.04
20	0.31	0.09	0.20	0.02	0.70	0.02
32	0.33	0.11	0.38	0.07	0.84	0.04
64	0.35	0.10	0.43	0.05		

www.manaraa.com

Table 6. Average absolute errors in the MP2 binding energy per water molecule for the SFMA method by fragmentation level and basis set (kcal/mol).

Cluster Size	6-31G(d)	6-31++G(d,p)	6-311++G(3df,2p)
Level 1			
16	0.32	0.31	0.41
20	0.31	0.37	0.53
32	0.44	0.17	1.32
64	0.63	0.17	
Level 2			
16	0.07	0.09	0.16
20	0.07	0.08	0.15
32	0.15	0.21	0.05
64	0.23	0.42	
Level 3			
16	0.007	0.02	0.02
20	0.006	0.03	0.03
32	0.07	0.12	0.03
64	0.02	0.13	

www.manaraa.com

Table 7. Fragment size distribution for SMFA Level 2.

Cluster Size	$(H_2O)_6$	$(H_2O)_5$	$(H_2O)_4$	$(H_2O)_3$	$(H_2O)_2$	N_{frag}
16	0	8	21	56	30	115
20	0	12	24	76	44	156
32	4	21	18	55	27	125
64	6	58	36	170	86	356

www.manaraa.com

Table 8. Fragment size distribution for SMFA Level 3.

Cluster Size	$(H_2O)_{11}$	$(H_2O)_{10}$	$(H_2O)_9$	$(H_2O)_8$	$(H_2O)_7$	$(H_2O)_6$	$(H_2O)_5$	$(H_2O)_4$	N_{frag}
16	0	2	12	20	12	8	0	1	55
20	0	2	17	37	37	20	0	2	115
32	0	6	23	30	38	27	41	37	220
64	7	30	79	99	108	94	100	83	626

www.manaraa.com

Table 9. Timing comparison[a,b] between fully ab initio MP2, FMO2-MP2 and FMO3-MP2 energy calculations. For the FMO calculations, the number of separated dimers (sd), SCF dimers (d), correlated dimers (cd), SCF trimers (t) and correlated trimers (ct) for each calculation is shown. For FMO2 calculations RESDIM=3.5 and RCORSD=3 while for FMO3 calculations RESDIM=4.5, RCORSD=4 and RITRIM=3,-1,3.5,3.

	6-31++G(d,p)						
	MP2	FMO2-MP2			FMO3-MP2		
Cluster Size	Wall Time	Wall Time	Error (kcal/mol)	sd/d/cd	Wall Time	Error (kcal/mol)	d/t/ct
16	394	12	2.65	0/120/120	89	0.72	120/560/560
20	1119	17	4.00	0/190/187	161	0.28	190/1140/1140
32	9989	36	9.30	7/489/454	429	1.73	496/4949/4797

[a]FMO timings performed using six GDDI groups with 8 CPUs/group; MP2 timings performed on 48 CPUs.
[b]Times in sec., errors in kcal/mol.

www.manaraa.com

Table S1. For an atom with Cartesian coordinates (0,0,0), with charge q, dipole moment $\boldsymbol{\mu}$, and traceless Cartesian quadrupole $\boldsymbol{\Theta}$, this table shows the coordinates and charges that mimic these moments. The small displacement, ε, is taken to be 0.01 Å.

Coordinates	Charge
0,0,0	q
ε,0,0	$\mu_x/2\varepsilon$
-ε,0,0	$-\mu_x/2\varepsilon$
0,ε,0	$\mu_y/2\varepsilon$
0,-ε,0	$-\mu_y/2\varepsilon$
0,0,ε	$\mu_z/2e$
0,0,-ε	$-\mu_z/2e$
ε,ε,0	$-\left(\Theta_{zz}-\Theta_{xy}\right)/6\varepsilon^2$
ε,0,ε	$-\left(\Theta_{yy}-\Theta_{xz}\right)/6\varepsilon^2$
0,ε,ε	$-\left(\Theta_{xx}-\Theta_{yz}\right)/6\varepsilon^2$
-ε,-ε,0	$-\left(\Theta_{zz}-\Theta_{xy}\right)/6\varepsilon^2$
-ε,0,-ε	$-\left(\Theta_{yy}-\Theta_{xz}\right)/6\varepsilon^2$
0,-ε,-ε	$-\left(\Theta_{xx}-\Theta_{yz}\right)/6\varepsilon^2$
ε,-ε,0	$-\left(\Theta_{zz}+\Theta_{xy}\right)/6\varepsilon^2$
ε,0,-ε	$-\left(\Theta_{yy}+\Theta_{xz}\right)/6\varepsilon^2$
0,ε,-ε	$-\left(\Theta_{xx}+\Theta_{yz}\right)/6\varepsilon^2$
-ε,ε,0	$-\left(\Theta_{zz}+\Theta_{xy}\right)/6\varepsilon^2$
-ε,0,-ε	$-\left(\Theta_{yy}+\Theta_{xz}\right)/6\varepsilon^2$
0,-ε,-ε	$-\left(\Theta_{xx}+\Theta_{yz}\right)/6\varepsilon^2$

www.manaraa.com

Table S2. Absolute errors in kcal/mol between *ab initio* results and the FMO and SMFA methods for each cluster. The geometry for each cluster is given in Table S4, identified by the labels shown here.

	MP2/6-31G(d)					MP2/6-31++G(d,p)					MP2/6-311++G(3df,2p)				
16	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3
c1b	3.79	0.08	-0.05	-3.35	1.41	-5.90	-1.72	-0.49	2.95	0.71	-9.41	1.48	-0.37	11.48	0.45
c1c16	5.23	0.22	0.02	-3.73	1.29	-2.77	-0.70	-0.11	2.80	0.49	-5.99	1.42	-0.04	11.03	0.26
c1c	8.08	0.42	0.01	-4.00	1.25	2.62	0.32	-0.02	2.68	0.56	-0.13	0.89	-0.03	10.70	0.51
cage	7.05	-0.33	0.10	-3.81	1.34	-0.65	-1.13	0.12	2.48	0.28	-2.16	0.31	0.17	11.07	0.21
cie	3.36	-0.22	0.07	-3.90	1.19	-4.05	-1.33	0.14	3.07	0.28	-5.97	0.58	-0.07	10.93	0.24
d2d	0.75	0.80	-0.01	-6.33	1.45	-13.14	-2.12	-0.69	2.97	-1.63	-14.35	1.51	-0.70	10.22	-1.32
s43	7.50	2.88	0.10	-6.84	1.67	-5.68	4.72	-0.15	2.65	-1.85	-8.38	5.68	-0.39	10.22	-1.14
20	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3
p3a	6.21	0.57	0.08	-5.73	1.58	-5.86	0.88	-0.12	4.66	0.17	-10.47	2.49	0.06	15.24	-0.06
p3b	6.81	-0.06	0.01	-5.85	1.61	-5.18	-1.33	-0.37	4.04	0.26	-10.49	1.79	-0.57	13.90	-0.02
pps_c	3.69	0.51	-0.04	-5.40	1.82	-11.49	-2.95	-0.78	4.03	-0.10	-15.51	2.09	-0.79	14.42	-0.20
pps_d	7.93	0.13	0.09	-6.68	1.73	-5.05	-1.57	-0.01	3.93	-0.14	-10.18	0.90	-0.23	13.23	-0.33
pps_e	7.94	0.85	0.01	-5.36	1.81	-3.92	-1.64	-0.34	3.83	-0.16	-6.35	1.24	-0.38	14.16	0.04
pps_f	10.70	0.43	0.22	-5.58	1.66	1.29	1.35	0.68	3.95	-0.06	-0.79	1.12	0.54	14.45	0.06
s4d2	0.09	0.79	0.18	-8.32	1.81	-19.11	-3.48	-0.48	4.16	-2.18	-20.91	1.37	-0.67	13.07	-2.00

www.manaraa.com

المنارة للاستشارات

Table S2. (continued)

	MP2/6-31G(d)					MP2/6-31++G(d,p)					MP2/6-311++G(3df,2p)				
32	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3
32_1	14.54	0.29	2.42	-10.68	3.36	-3.89	-0.30	3.08	11.80	2.21	-19.67	-9.76	0.75	26.83	0.98
32_2	13.33	1.66	2.53	-10.43	3.50	-6.13	0.82	4.15	12.37	2.53	-23.56	-9.69	1.13	27.34	1.30
32_3	17.80	4.00	3.12	-10.50	3.16	-4.00	2.53	4.21	11.45	1.92	-20.89	-5.21	1.19	24.57	1.18
32ab	13.69	1.30	2.48	-10.42	3.51	-5.51	0.10	4.08	12.48	2.53	-23.16	-10.45	1.11	27.38	1.30
32ad	17.14	4.17	2.40	-11.06	3.69	-0.36	3.16	4.29	11.77	2.55	-15.95	-7.74	1.30	25.75	1.04
32h	12.79	1.31	2.52	-10.32	3.48	-6.43	-0.01	4.13	12.30	2.48	-24.00	-10.68	1.00	27.27	1.23
32o	13.21	1.08	2.51	-10.33	3.51	-5.90	-0.56	4.13	12.46	2.49	-184.33	-11.19	1.07	27.32	1.26
32z	9.30	2.79	2.84	-9.64	3.26	-11.41	3.25	4.37	12.44	2.33	-28.43	-6.93	0.96	28.00	1.25

	MP2/6-31G(d)					MP2/6-31++G(d,p)				
64	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3	SMFA L1	SMFA L2	SMFA L3	FMO2	FMO3
a	39.52	7.93	2.10	-22.08	6.41	9.12	14.77	9.02	27.72	3.37
b	39.85	8.00	2.14	-22.05	6.40	9.95	15.05	9.06	27.75	3.36
c	37.98	7.93	2.25	-21.81	6.36	7.76	14.53	9.37	27.54	3.29
d	42.93	8.79	1.98	-22.23	6.43	13.37	16.78	8.65	27.66	3.30
e	41.81	7.20	1.95	-22.32	6.47	12.59	16.80	8.35	27.21	3.29
f	42.98	7.67	1.97	-22.79	6.55	14.32	17.47	8.71	27.36	3.15
g	40.37	8.04	2.16	-22.14	6.34	11.22	14.87	9.06	27.69	3.35

www.manaraa.com

Table S3. The SMFA error in the total energy is compared to the total dispersion energy for each cluster (in kcal/mol).

	MP2/6-31g(d)					
	Level 1		Level 2		Level 3	
16 Waters	E(disp)	L1 error	E(disp)	L2 error	E(disp)	L3 error
c1b	-4.49	6.05	-0.95	0.13	-0.03	-0.08
c1c16	-4.43	8.34	-0.95	0.35	-0.03	0.03
c1c	-4.29	12.88	-0.94	0.66	-0.03	0.02
cage	-4.35	11.24	-0.95	-0.52	-0.03	0.15
cie	-4.40	5.36	-0.94	-0.34	-0.03	0.11
d2d	-5.09	1.19	-1.03	1.28	-0.05	-0.02
s43	-4.90	11.96	-1.02	4.60	-0.05	0.16
20 Waters						
p3a	-6.38	9.90	-1.53	0.91	-0.07	0.12
p3b	-6.32	10.85	-1.52	-0.09	-0.07	0.02
pps_c	-6.27	5.89	-1.40	0.82	-0.07	-0.06
pps_d	-6.50	12.64	-1.36	0.20	-0.07	0.14
pps_e	-6.12	12.66	-1.40	1.35	-0.07	0.02
pps_f	-6.13	17.05	-1.43	0.69	-0.07	0.35
s4d2	-6.82	0.15	-1.45	1.27	-0.08	0.28
32 Waters						
32_1	-15.10	23.17	-4.08	0.46	-1.43	3.86
32_2	-15.54	21.24	-4.37	2.65	-1.54	4.03
32_3	-14.80	28.37	-3.83	6.37	-1.32	4.97
32ab_2b	-15.49	21.81	-4.36	2.07	-1.54	3.96
32ad_2b	-15.70	27.32	-4.51	6.65	-1.57	3.82
32h_2b	-15.53	20.38	-4.37	2.08	-1.54	4.02
32o_2b	-15.53	21.05	-4.36	1.72	-1.54	4.01
32z_2b	-15.33	14.82	-4.11	4.44	-1.48	4.53
64 Waters						
a	-35.03	62.98	-10.93	12.64	-2.93	3.34
b	-35.00	63.50	-10.91	12.75	-2.92	3.41
c	-35.06	60.53	-10.94	12.64	-2.92	3.59
d	-34.89	68.42	-10.90	14.01	-2.92	3.16
e	-33.62	66.63	-9.82	11.48	-2.86	3.10
f	-33.76	68.50	-9.90	12.23	-2.88	3.14
g	-35.04	64.33	-10.94	12.81	-2.92	3.44

www.manaraa.com

Table S3. (continued)

	MP2/6-31++G(d,p)					
	Level 1		Level 2		Level 3	
16 Waters	E(disp)	L1 error	E(disp)	L2 error	E(disp)	L3 error
c1b	-7.27	-9.40	-0.95	-2.73	-0.05	-0.79
c1c16	-7.24	-4.42	-0.95	-1.11	-0.06	-0.18
c1c	-7.11	4.17	-0.94	0.51	-0.06	-0.03
cage	-7.21	-1.04	-0.95	-1.81	-0.05	0.20
cie	-7.21	-6.45	-0.94	-2.12	-0.06	0.23
d2d	-8.37	-20.93	-1.03	-3.38	-0.07	-1.11
s43	-8.34	-9.05	-1.02	7.52	-0.07	-0.24
20 Waters						
p3a	-10.48	-9.34	-1.53	1.40	-0.12	-0.20
p3b	-10.40	-8.25	-1.52	-2.11	-0.12	-0.60
pps_c	-10.15	-18.31	-1.40	-4.70	-0.12	-1.24
pps_d	-10.82	-8.05	-1.36	-2.50	-0.12	-0.02
pps_e	-10.02	-6.25	-1.40	-2.62	-0.12	-0.54
pps_f	-10.11	2.05	-1.43	2.15	-0.12	1.09
s4d2	-11.14	-30.46	-1.45	-5.55	-0.13	-0.76
32 Waters						
32_1	-25.21	-6.20	-8.55	-0.47	-2.57	4.91
32_2	-25.89	-9.76	-9.16	1.31	-2.76	6.62
32_3	-24.68	-6.38	-7.94	4.02	-2.31	6.71
32ab_2b	-25.84	-8.79	-9.14	0.16	-2.76	6.50
32ad_2b	-26.02	-0.57	-9.41	5.04	-2.77	6.84
32h_2b	-25.85	-10.25	-9.16	-0.02	-2.77	6.58
32o_2b	-25.85	-9.40	-9.16	-0.88	-2.76	6.58
32z_2b	-25.35	-18.19	-8.63	5.18	-2.66	6.97
64 Waters						
a	-57.72	14.53	-18.27	23.54	-5.05	14.37
b	-57.68	15.85	-18.24	23.99	-5.05	14.43
c	-57.71	12.37	-18.28	23.16	-5.03	14.93
d	-57.68	21.31	-18.26	26.74	-5.05	13.79
e	-55.65	20.06	-16.47	26.78	-4.93	13.30
f	-55.87	22.82	-16.58	27.84	-4.95	13.88
g	-57.74	17.87	-18.28	23.70	-5.05	14.43

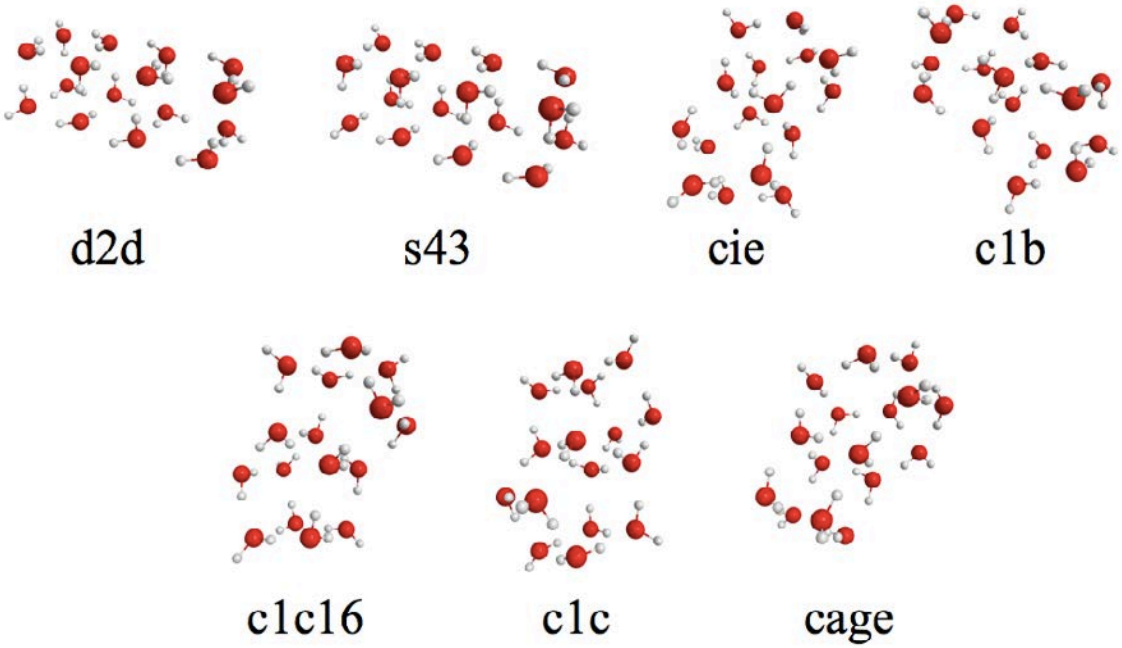
www.manaraa.com

Table S3. (continued)

	MP2/6-31++G(3df,2p)					
	Level 1		Level 2		Level 3	
16 Waters	E(disp)	L1 error	E(disp)	L2 error	E(disp)	L3 error
c1b	-11.33	-14.99	-1.46	2.35	-0.08	-0.58
c1c16	-11.28	-9.55	-1.46	2.27	-0.09	-0.06
c1c	-11.07	-0.21	-1.45	1.42	-0.09	-0.05
cage	-11.20	-3.45	-1.46	0.50	-0.08	0.27
cie	-11.23	-9.51	-1.45	0.93	-0.09	-0.12
d2d	-13.06	-22.87	-1.59	2.41	-0.12	-1.11
s43	-13.03	-13.36	-1.57	9.05	-0.11	-0.63
20 Waters						
p3a	-16.32	-16.69	-2.35	3.97	-0.19	0.09
p3b	-16.17	-16.72	-2.34	2.86	-0.19	-0.91
pps_c	-15.81	-24.71	-2.16	3.34	-0.18	-1.26
pps_d	-16.85	-16.22	-2.09	1.44	-0.18	-0.37
pps_e	-15.61	-10.12	-2.15	1.98	-0.18	-0.60
pps_f	-15.72	-1.26	-2.20	1.78	-0.19	0.85
s4d2	-17.38	-33.33	-2.23	2.18	-0.20	-1.07
32 Waters						
32_1	-39.21	-31.35	-13.29	-15.55	-4.03	1.19
32_2	-40.26	-37.55	-14.24	-15.44	-4.33	1.80
32_3	-38.38	-33.28	-12.30	-8.30	-3.60	1.90
32ab_2b	-40.19	-36.90	-14.22	-16.65	-4.33	1.77
32ad_2b	-40.42	-25.41	-14.62	-12.33	-4.34	2.07
32h_2b	-40.20	-38.25	-14.26	-17.01	-4.35	1.60
32o_2b	-40.21	-293.75	-14.25	-17.84	-4.34	1.71
32z_2b	-39.43	-45.31	-13.43	-11.04	-4.19	1.53

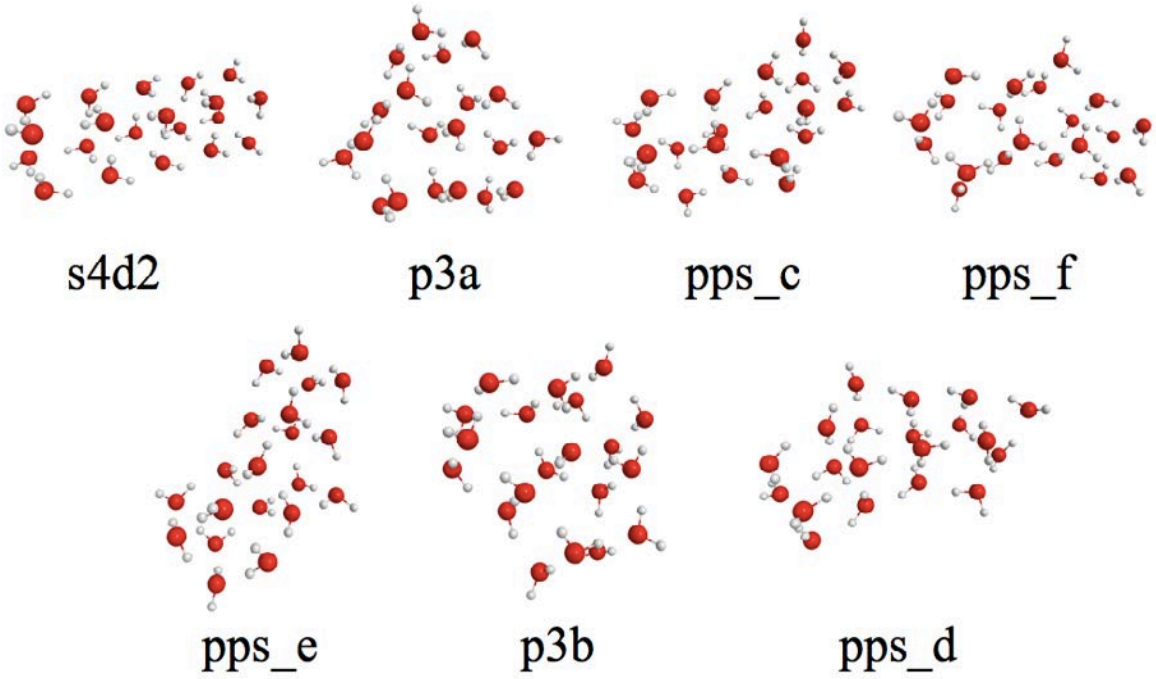
www.manaraa.com

Figure S1. Geometries of all clusters of 16 waters studied.



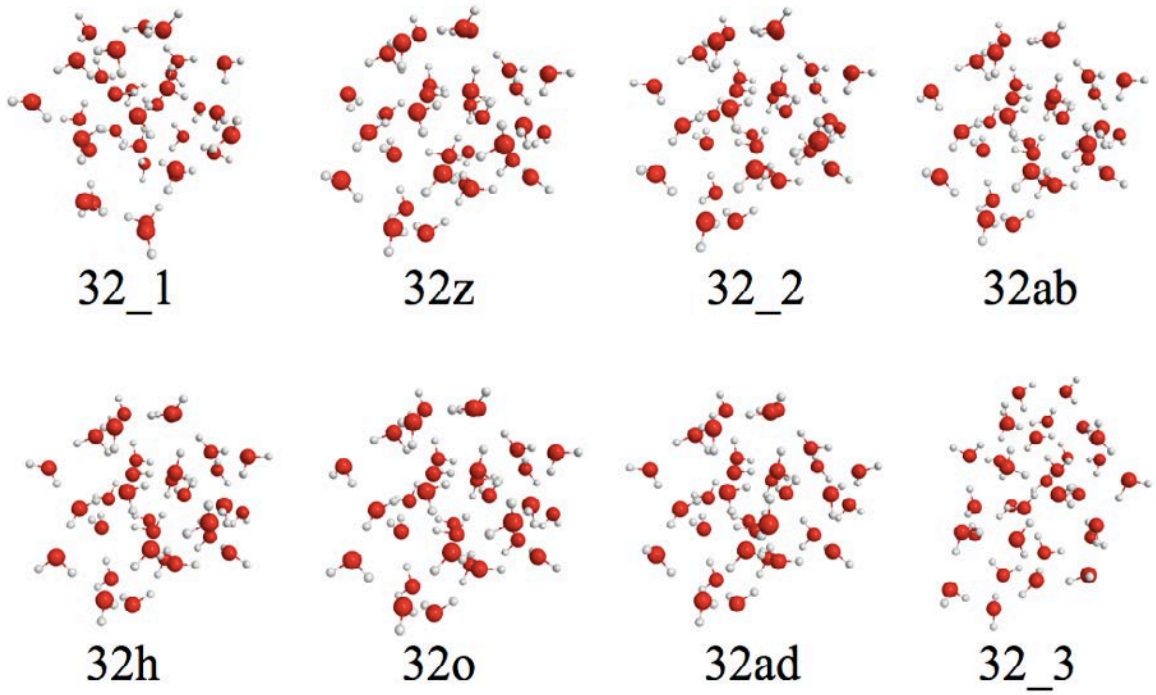
www.manaraa.com

Figure S2. Geometries of all clusters of 20 waters studied.



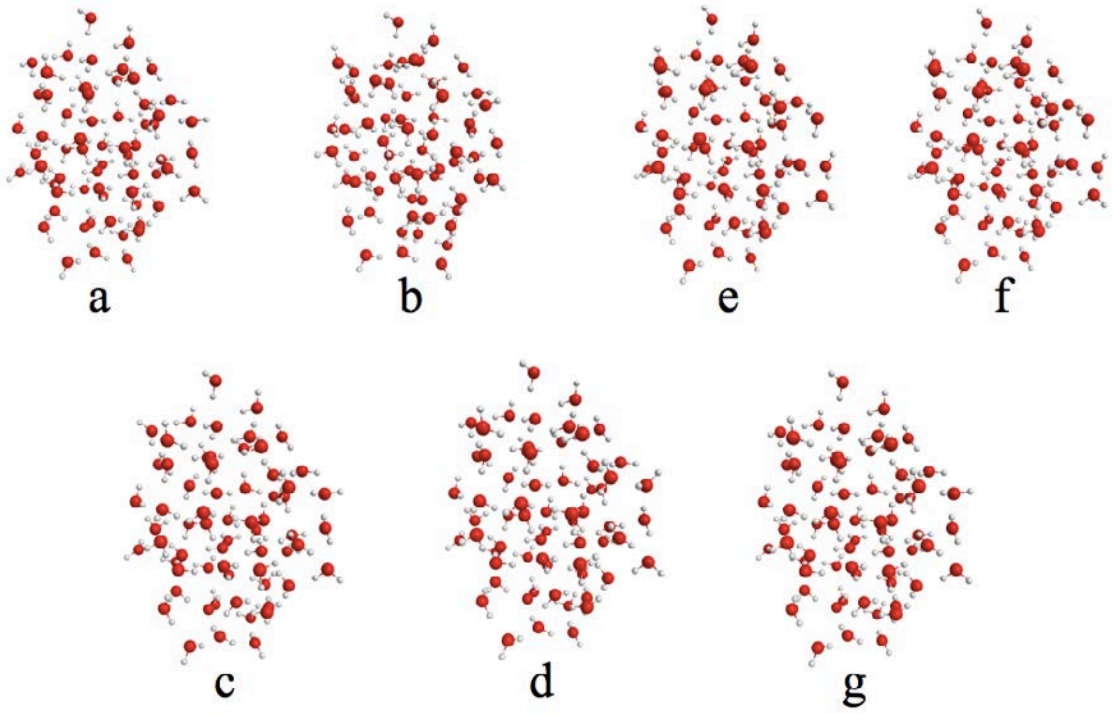
www.manaraa.com

Figure S3. Geometries of all clusters of 32 waters studied.



www.manaraa.com

Figure S4. Geometries of all clusters of 64 waters studied.



www.manaraa.com

Chapter 4. Open-Shell Formulation of the Fragment Molecular Orbital Method

A paper published in *The Journal of Chemical Theory and Computation*

Spencer R. Pruitt, Dmitri G. Fedorov, Kazuo Kitaura, Mark S. Gordon

Abstract

Performing accurate calculations on large molecular systems is desirable for closed and open-shell systems. In this work, the fragment molecular orbital (FMO) method is extended to open-shell systems and implemented in the GAMESS (General Atomic and Molecular Electronic Structure System) program package. The accuracy of the method is tested, and the ability to reproduce reaction enthalpies is demonstrated. These tests also demonstrate its utility in providing an efficient means to model large open-shell systems.

1. Introduction

Recently, a large number of fragment-based methods[1-7] have been developed, including the fragment molecular orbital (FMO) method.[8-10] The aim of these methods is to treat complex molecular species efficiently, while retaining *ab initio* accuracy, by dividing the system into many much smaller fragments. However, few of these methods are capable of treating open-shells.[11] Several wave function types have already been interfaced with FMO,[12-17] however most of them are for closed-shell methods. The only exception is the multi-configuration self-consistent field (MCSCF),[14] which can treat both closed and open-shell species. Although the MCSCF method is very useful in some systems, it is also desirable to have a single-reference open-shell method, which could be efficiently combined with restricted open-shell (RO) second order Møller-Plesset perturbation theory (MP2)[18] or coupled cluster theory (CC).[19]

Open-shell systems play a very important role in many processes, for instance, in radical chemistry,[20] electron transfer,[21] and many transition metal compounds.[22] In addition, transition states in chemical reactions involve breaking chemical bonds and thus possess considerable open-shell character. Although more difficult cases such as transition metal complexes may require a more sophisticated treatment (e.g. MCSCF), in many organic, inorganic and biochemical systems the open-shell character can be well described by an open-shell single-reference wave function. Therefore, it is useful to be able to perform large-scale calculations of open-shell systems with single-reference methods. The FMO method is particularly appealing in this regard, since it has been shown to accurately reproduce fully *ab initio* calculations with high accuracy.[6,9] Due to the inherently parallel nature of the method, it also scales ~ linearly with system size.

2. Methodology

The *n*-body FMO energy (FMO*n*) of *N* fragments is given by:[8]

www.manaraa.com

$$E^{\mathrm{FMO2}} = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \left(E_{IJ} - E_I - E_J \right) \tag{1}$$

$$E^{\mathrm{FMO3}} = E^{\mathrm{FMO2}} + \sum_{I>J>K}^{N} \left(E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK} \right) \tag{2}$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \tag{3}$$

and E_I, E_{IJ} and E_{IJK} are the monomer (single fragment), dimer (fragment pair) and trimer (fragment triple) energies, respectively, computed in the electrostatic field of other fragments.[6]

This expression is also used for the restricted open-shell FMO method. The computational scheme is methodologically similar to that of FMO-based MCSCF or time-dependent[23] density functional theory (DFT), with one fragment chosen to be the open-shell fragment. There are two types of dimers and trimers in an open-shell system: (a) open-shell, if they include the open shell fragment; (b) closed shell otherwise. The electrostatic field added to all monomer, dimer and trimer calculations is computed from the total density of either open- or closed-shell fragments. In this work all open-shell systems have doublet spin multiplicity; however, there is no limitation on the multiplicity of the open-shell FMO method. The FMO energy has been implemented for the RO-based Hartree-Fock method (ROHF), MP2 (ROMP2) and CC (ROCC).

In addition, a multilayer scheme[24] was also implemented in which several layers with varying levels of electronic structure theory (HF, MP2, CC) can be defined. The multilayer FMO method uses the notation of listing the wave functions in increasing order of layers, e.g., FMO2-ROMP2:ROCC means that the fragment densities are obtained self-consistently at the uncorrelated level (ROHF) and used in the correlated calculations at the ROMP2 level for the less important fragments (substituents) and ROCC for (for example) a reaction center. Dimer calculations are performed at the lower level of the two layers to which the two fragments belong. In the case of the reaction described below there was only one fragment in the higher level, so that all dimer calculations were done with ROMP2.[25] The same basis set is used in both layers.

Especially for CC, which is a very steeply scaling method (N^7) with the system size N, the use of FMO is beneficial even for very small systems, such as that with only three fragments discussed below. In addition to the computational cost scaling, CC requires very large memory; the huge memory demand prevented the full *ab initio* CC calculations while the FMO-CC computations are feasible.

The open-shell FMO code was parallelized with the generalized distributed data interface (GDDI),[26,27] using a two-level hierarchical scheme. Since the open-shell MP2 method in GAMESS (General Atomic and Molecular Electronic Structure System)[28] is fully parallelized, it can take advantage of both levels of GDDI.

www.manaraa.com

This is not the case for the open-shell CC code, which can only take advantage of the inherent coarse-grained level parallelism of the FMO method using GDDI. All methods discussed here have been implemented in GAMESS.

All calculations discussed here used the default values of thresholds, except for water clusters the point charge representation of the electrostatic potentials in FMO was used. Spherical basis functions were used throughout, and the core electrons (e.g., 1s on C and O) were not correlated in MP2 and CC.

3. Tests

3.1 Open-shell FMO2 and FMO3 Calculations on $OH(H_2O)_5$ Clusters.

A preliminary test of the open-shell FMO method employed clusters of six water molecules. The ability to accurately model the solvated OH radical has implications in biological applications and atmospheric processes.[29] The large charge transfer present in solvated OH clusters adds an additional degree of difficulty, providing an excellent test case for the open-shell FMO method.

The structures of the six clusters shown in Figure 1 were determined by optimizing previously determined minima[30] with MP2 and the aug-cc-pVTZ basis set.[31] To create the open-shell test systems, one hydrogen atom was arbitrarily removed from one of the water molecules in each cluster. Fragments were chosen by placing the open shell OH in a fragment with both nearest neighbor H_2O molecules, while placing the other H_2O molecules in fragments by themselves, with the exception of the prism and bag isomer. For these two isomers, there were two other H_2O molecules with a significant interaction between them. This required them to be placed into one fragment together, while the open shell OH was placed in a fragment with only one nearest neighbor. The final fragmentation scheme created four FMO monomers for all isomers.

Errors for FMO2-ROHF calculations (Table 1) relative to *ab initio* calculations are between 0.2-2.8 kcal/mol, while the errors are between 0.0-2.7 kcal/mol for FMO2-ROMP2. The addition of *ab initio* three body interactions with FMO3-ROHF significantly reduces the error to 0.0-0.1 kcal/mol, while the error for FMO3-ROMP2 falls to 0.1-0.3 kcal/mol. The improvement in accuracy for FMO3 is not a surprising result, as the importance of three-body effects in water has been shown previously.[12,32]

Relative energetics (Table 2) are of similar accuracy, with the ordering of isomers being captured correctly with both FMO2 and FMO3, illustrated in Figure 2. The choice of fragments is important for the accurate reproduction of the relative energies, especially for FMO2 and in systems in which fragments may have very strong interactions such as charge transfer. These strongly interacting fragments should be grouped together to improve accuracy.

Considering the distributed memory requirements of the *ab initio* ROMP2 calculations (~2 GB of RAM) versus that of the FMO2-ROMP2 calculation (~512 MB of RAM) or the FMO3-ROMP2 calculation

www.manaraa.com

(~1GB of RAM), the open-shell FMO method is capable of providing accurate energies at a much lower cost. One can imagine that as the size of the cluster *N* increases, the memory requirement of the *ab initio* calculation will also increase substantially (~N^4); however, for FMO it will remain the same, no matter how large the cluster is.

3.2 Multilayer FMO2 Calculation of the Reversible Addition Fragmentation Chain Transfer (RAFT) Reaction Enthalpy (Figure 3).

As a further test of the open-shell FMO method, the initiation step in the RAFT reaction[33,34] was chosen as a small test case. The initiation step consists of two reactants, one an open-shell radical, that combine to form an open-shell radical product. Initial structures were optimized using DFT with the B3LYP functional and the 6-31G(d) basis set.[35] FMO2-ROMP2 single point energy calculations using the 6-311G(d,p) basis set were performed with the fragmentation scheme shown in Figure 3. Higher level calculations were also performed, using the completely renormalized coupled cluster single and double excitations using left eigenstates for perturbative triple excitation corrections (CR-CC(2,3))[19,36] method with the 3-21G basis set to model the open-shell fragment, and MP2 with the 3-21G basis set to model the closed shell fragments (FMO2-MP2:CR-CC(2,3)), shown in Figure 4. The reason for using a smaller basis set is the huge memory requirements for the CC code, which even for 3-21G was 8 GB, while for the 6-31G(d) basis set the requirements are more than 32 GB.

Table 3 gives the absolute energy differences between *ab initio* ROMP2 and FMO2-ROMP2 for both reactants and product of the RAFT reaction. The FMO2-ROMP2 method gives accurate energies, with errors between 0.4-0.9 kcal/mol. This accuracy in absolute energies translates to equivalent accuracy when calculating the reaction enthalpy (Table 4). Comparing the enthalpy from FMO2-ROMP2 using the 6-311G(d,p) basis set with the ab initio ROMP2 enthalpy, the error is only 0.9 kcal/mol. Calculations performed with the 3-21G basis set also show very good agreement between ab initio ROMP2 and FMO2-ROMP2 calculations. With the addition of the CC correction in the FMO2-MP2:CR-CC(2,3) calculation the enthalpy changes by +2.2 kcal/mol. This suggests that the use of ROMP2 to calculate enthalpies is adequate in this case, however, a higher level of theory may be required in other instances to properly describe open-shell systems.

4. Conclusions

The open-shell FMO method has been implemented in the GAMESS program package and parallelized using GDDI for the HF, MP2, and CC levels of electronic structure theory. The accuracy of the method was tested by calculating the absolute and relative energetics of open-shell molecular clusters. The ability of the method to reproduce reaction enthalpies was also tested using the RAFT reaction. It was demonstrated that in both cases the open-shell FMO method provides energies and properties within 0.0-2.0 kcal/mol of *ab initio* calculations.

www.manaraa.com

The need for a single reference open-shell FMO method was fulfilled through this work, providing a scalable method for use on large chemical systems such as the RAFT reaction. The combination of accuracy and reduction in computational expense provides a means for accurate calculations on much larger open-shell radical chemical systems than was previously available.

References

1. Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.
2. Pomogaev, V.; Pomogaeva, A.; Aoki, Y. *J. Phys. Chem. A* **2009**, *113*, 1429.
3. Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comp.* **2009**, *5*, 459.
4. Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comp.* **2009**, *5*, 1573.
5. Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comp.* **2009**, *5*, 1667.
6. 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.
7. Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927.
8. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
9. Fedorov, D.G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.
10. Fedorov, D.G.; Kitaura, K.; *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009.
11. Korchowiec, J.; Gu, F. L.; Aoki, Y. *Int. J. Quant. Chem.* **2005**, *105*, 875.
12. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
13. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.
14. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108.
15. Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.* **2005**, *406*, 283.
16. Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
17. Chiba, M.; Fedorov, D.G.; Nagata, T.; Kitaura, K. *Chem. Phys. Lett.* **2009**, *474*, 227.
18. Lee, T.J.; Rendell, A.P.; Dyall, K.G.; Jayatilaka, D. *J.Chem.Phys.* **1994**, *100*, 7400.
19. Włoch, M.; Gour, J.R.; Piecuch, P. *J. Phys. Chem. A*, **2007**, *111*, 11359.
20. Pryor, W.A.; *Free Radicals*, McGraw-Hill: New York, NY, 1966.
21. Griese, B.; Wang, M.; Gao, J.; Stoltz, M.; Müller, P.; Graber, M. *J. Org. Chem.* **2009**, *74*, 3621.
22. Astruc, D.; *Electron Transfer and Radical Processes in Transition-Metal Chemistry*; Wiley-VCH: New York, NY, 1995.

www.manaraa.com

23. Chiba, M.; Fedorov, D.G. *J. Chem. Phys.* **2007**, *127*, 104108.

24. Fedorov, D.G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A.* **2005**, *109*, 2638.

25. Aikens, C.M.; Fletcher, G.D.; Schmidt, M.W.; Gordon, M.S. *J. Chem. Phys.* **2006**, *124*, 014107.

26. Fletcher, G.D.; Schmidt, M.W.; Bode, B.M.; Gordon, M.S. *Comp. Phys. Comm.* **2000**, *128*, 190.

27. Fedorov, D.G.; Olson, R.M.; Kitaura, K.; Gordon, M.S.; Koseki, S. *J. Comp. Chem.* **2004**, *25*, 872.

28. Schmidt, M.W.; Baldridge, K.K.; Boatz, J.A.; Elbert, S.T.; Gordon, M.S.; Jensen, J.H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.; Windus, T.L.; Dupuis, M.; Montgomery, Jr., J.A. *J. Comp. Chem.* **1993**, *14*, 1347.

29. Seinfeld, J.H.; Pandis, S.N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; Wiley, New York, NY, 2006.

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

31. (a) Dunning, T. H; Jr., *J. Chem. Phys*. **1989**, *90*, 1007. (b) Woon, D. E.; T. H. Dunning, Jr. *J. Chem. Phys*. **1995**, *103*, 4572.

32. Xantheas, S. *Struct. Bond.* **2005**, *116*, 119.

33. Coote, M.L.; Izgorodina, E.I.; Krenske, E.H.; Busch, M.; Barner-Kowollik, C. *Macromol. Rapid Comm.* **2006**, *27*, 1015.

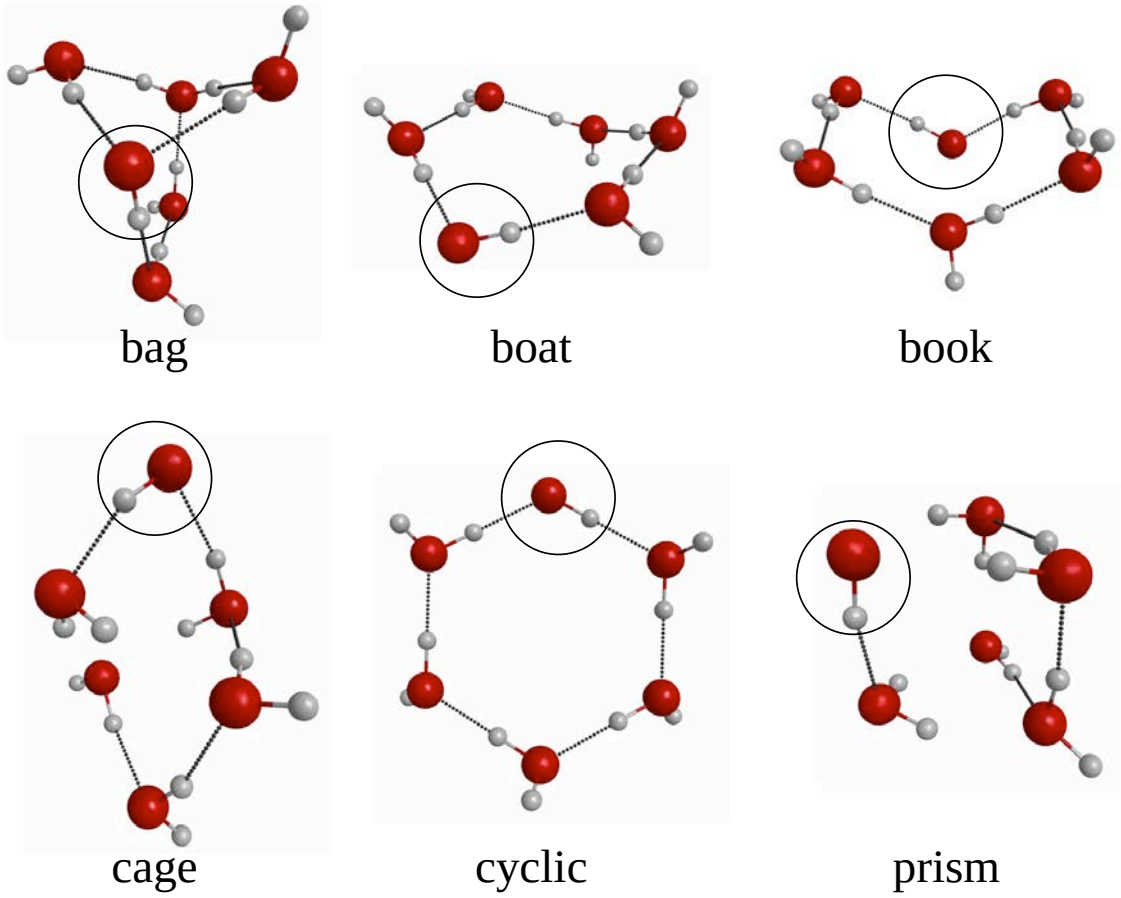
34. Chiefari, J.; Chong, Y.K.; Ercole, F.; Krstina, J.; Jeffery, J.; Le, T.P.T; Mayadunne, R.T.A.; Meijs, G.F.; Moad, C.L.; Moad, G.; Rizzardo, E.; Thang, S.H. *Macromol.* **1998**, *31*, 5559.

35. Structures used for the RAFT reaction calculations were obtained from Professor Michelle Coote at the Austrailian National University.

36. Piecuch, P.; Włoch, M. *J. Chem. Phys.* **2005**, *123*, 224105-1.

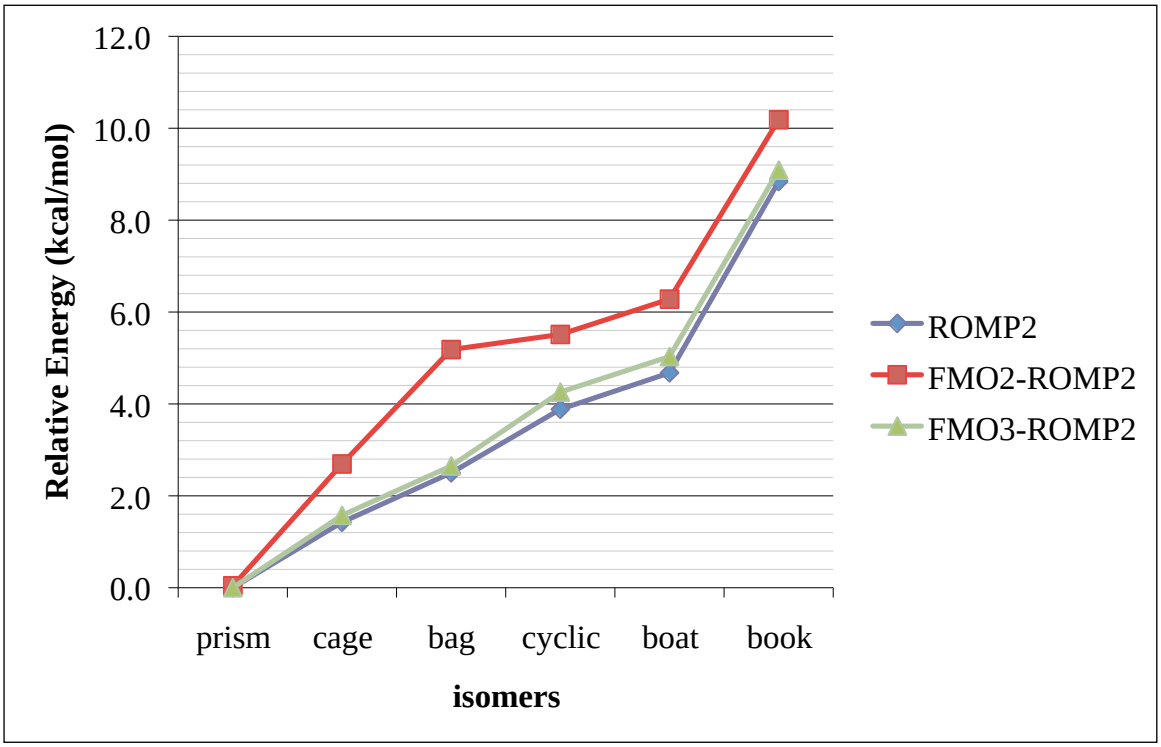
www.manaraa.com

Figure 1. The six isomers of $OH(H_2O)_5$ clusters used for testing. Open shell OH molecules are circled and the naming convention is from ref. 25.



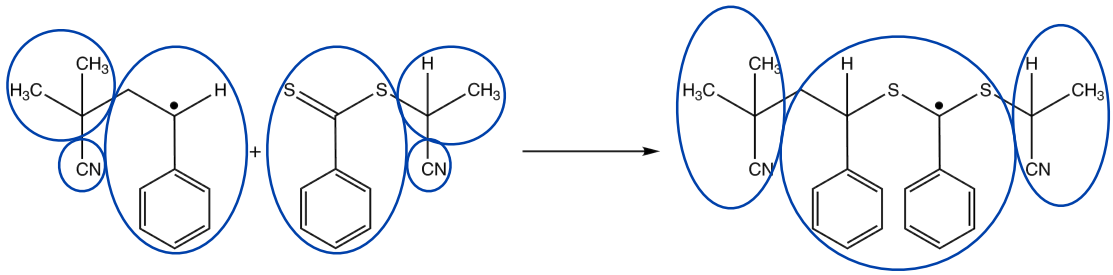
www.manaraa.com

Figure 2. Graph of the relative energies of the six $OH(H_2O)_5$ clusters computed using *ab initio* ROMP2, FMO2-ROMP2 and FMO3-ROMP2.



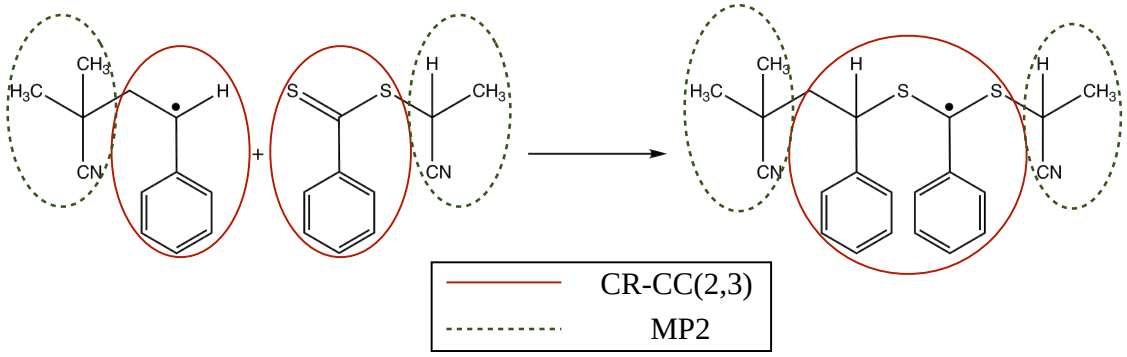
www.manaraa.com

Figure 3. Reaction scheme for the RAFT reaction with the choice of FMO fragments shown in blue.



www.manaraa.com

Figure 4. Reaction scheme for the RAFT reaction with the multilayer FMO details: higher layer fragments (CC) are circled in red and lower levels fragments (MP2) are in green dashed circles.



www.manaraa.com

Table 1. Absolute errors between *ab initio* ROMP2 and FMO-ROHF for six $OH(H_2O)_5$ clusters.

	absolute errors (kcal/mol)			
	aug-cc-pVTZ			
isomer	FMO2-ROHF	FMO3-ROHF	FMO2-ROMP2	FMO3-ROMP2
Prism	0.2	0.0	0.0	-0.1
Cage	1.4	0.1	1.3	0.1
Bag	2.8	0.0	2.7	0.1
cyclic	1.7	0.1	1.6	0.3
boat	1.7	0.1	1.6	0.3
book	1.4	0.1	1.3	0.2

www.manaraa.com

Table 2. Comparison of the ROMP2 relative energies[a] of six $OH(H_2O)_5$ clusters.

	relative energies (kcal/mol)		
	aug-cc-pVTZ		
isomer	ROMP2	FMO2-ROMP2	FMO3-ROMP2
prism	0.0	0.0	0.0
cage	1.4	2.7	1.6
bag	2.5	5.2	2.7
cyclic	3.9	5.5	4.3
boat	4.7	6.3	5.0
book	8.8	10.2	9.1

[a] Zero energy for all methods is set to be the prism isomer.

www.manaraa.com

Table 3. Absolute errors between *ab initio* ROMP2 calculations and FMO2-ROMP2 for the RAFT reaction.

	absolute errors (a.u.)		
	6-311G(d,p)		
	ROMP2	FMO2-ROMP2	error
reactant 1	-518.94377299	-518.94481925	-0.7
reactant 2	-1235.33260039	-1235.33407266	-0.9
product	-1754.31552043	-1754.31616825	-0.4

www.manaraa.com

Table 4. Calculated enthalpy of the RAFT reaction using *ab initio* ROMP2, FMO2-ROMP2 and FMO2-MP2:CR-CC(2,3).

		energy (a.u.)			enthalpy (kcal/mol)
basis set	method	reactant 1	reactant 2	product	
3-21G	FMO2-ROMP2	-515.27782025	-1228.15256231	-1743.46127536	-19.4
3-21G	FMO2-MP2:CR-CC(2,3)	-515.37480438	-1228.23664205	-1743.63878524	-17.2
6-311G(d,p)	FMO2-ROMP2	-518.94481925	-1235.33407266	-1754.31616825	-23.4
3-21G	ROMP2	-515.27782024	-1228.15256232	-1743.46133847	-19.4
6-311G(d,p)	ROMP2	-518.94377300	-1235.33260039	-1754.31552043	-24.6

www.manaraa.com

Chapter 5. Geometry Optimizations of Open-Shell Systems with the Fragment Molecular Orbital Method

A paper submitted to *The Journal of Physical Chemistry A*

Spencer R. Pruitt, Dmitri G. Fedorov, Mark S. Gordon

Abstract

The ability to perform geometry optimizations on large molecular systems is desirable for both closed and open-shell species. In this work, the restricted open-shell Hartree Fock (ROHF) gradients for the fragment molecular orbital (FMO) method are presented. The accuracy of the gradients is tested, and the ability of the method to reproduce adiabatic excitation energies is also investigated. Timing comparisons between the FMO method and full *ab initio* calculations are also performed, demonstrating the efficiency of the FMO method in modeling large open-shell systems.

1. Introduction

With recent increases in computational power the need for methods that are able to exploit the massively parallel nature of modern computers is becoming more important. A number of methods have been developed[1-7] that attempt to exploit the ability to divide large systems into more computationally tractable pieces.[8] One such method, the fragment molecular orbital (FMO) method[9-11], has been at the forefront of this effort, particularly since the development of the generalized distributed data interface (GDDI).[12] The GDDI allows the FMO method to take advantage of large, modern computer clusters and massively parallel computers, facilitating the solutions of much larger chemical systems than has previously been possible.[6]

With the broad range of interests in chemical research, including silicon nano pores,[13] proteins,[14] condensed phases[15] and radical chemistry,[16] there is also an increasing need for broadened functionality to be added to the FMO method. Since the original formulation[9-11,17] there have already been a number of extensions added to the FMO method to help treat a broad range of chemical problems, including second order Møller-Plesset perturbation theory (MP2),[18] coupled cluster theory (CC),[19] density functional theory (DFT),[20] solvation models such as the effective fragment potential (EFP)[21] and the multilayer formulation of the FMO method (MFMO).[22]

Until recently, the only way to treat open-shell systems with the FMO method was to use the multi-configuration self-consistent field implementation (FMO-MCSCF),[23] which can treat both closed and open-shell species. Although the MCSCF method can be very useful in treating multi-reference systems, it is also desirable to have a single-reference open-shell method that can be efficiently combined with restricted open-shell second order Møller-Plesset perturbation theory (ROMP2)[13] or coupled cluster theory (CC).[14] This ability to treat systems containing single-reference open-shell character was recently added to the FMO method with

www.manaraa.com

the implementation of the restricted open-shell wave function (FMO-ROHF)[16]. Since the original effort was limited to single point energy calculations, it is the goal of the present work to develop open-shell FMO energy gradients and to demonstrate the efficacy of the FMO-ROHF method in geometry optimizations[24] and excitation energy calculations.[25]

2. Methodology

Since the details of the FMO-ROHF method and relevant functionality have been discussed in detail previously[11,26] a brief overview of the general theory is presented here.

The FMO energy of N fragments for the n-body FMOn expansion is given by:[17]

$$E^{FMO2} = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \left(E_{IJ} - E_I - E_J \right) \tag{1}$$

$$E^{FMO3} = E^{FMO2} + \sum_{I>J>K}^{N} \left(E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK} \right) \tag{2}$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \tag{3}$$

and E_I, E_{IJ} and E_{IJK} are the monomer (single fragment), dimer (fragment pair) and trimer (fragment triple) energies, respectively, computed in the electrostatic field of all other fragments.[17] The first step of all FMO calculations is to converge the electronic state of each fragment (monomer) with respect to the electrostatic field of the whole system. This is followed by all dimer calculations for FMO2 and all trimer calculations for FMO3, performed in the field determined in the monomer step.

The expressions in eqs 1 and 2 that are used for the restricted open-shell FMO method resemble those of the closed-shell method.[17] The FMO-ROHF computational scheme is similar in nature to the FMO-based MCSCF or time-dependent DFT (FMO-TDDFT) methods.[15] In all of these schemes, one fragment is chosen to be the open-shell fragment, treated with the corresponding wave function or density, while all other fragments are taken to be closed shell species. Two types of dimers and trimers are present in an open-shell FMO calculation: (a) open-shell if they include the open shell fragment; (b) closed shell otherwise. The electrostatic field added to all n-mer (monomer, dimer, trimer) calculations is computed from the total density of either open- or closed-shell fragments. In covalently bonded systems, the boundaries are treated exactly as in the closed-shell FMO method,[9,17] that is, by assigning two electrons from the detached bond to one fragment and none to the other.

www.manaraa.com

There is an important difference between the perturbative-like treatment used in the configuration interaction (CI) and TDDFT approaches and the treatment used in the ROHF and MCSCF approaches. For CI and TDDFT calculations, the molecular orbitals and their energies from the ground electronic state are used in the subsequent excited state calculation without additional orbital relaxation. In contrast, for ROHF and MCSCF calculations the orbitals are optimized for the state of interest. In addition, within the FMO method, the embedding electrostatic field in CI and TDDFT calculations is computed for the ground state, whereas this field is fully relaxed for the state of interest in ROHF and MCSCF calculations. The importance of the electrostatic field relaxation has been discussed recently in conjunction with EFP-based excitation studies.[27]

Following the derivation for the closed shell FMO gradient,[28] the gradient for the open-shell FMO-ROHF method has been derived by taking the derivative with respect to a nuclear coordinate of the energy in eqs 1 and 2 for FMO2 and FMO3, respectively. This work, similarly to closed-shell TDDFT,[29] follows the derivation of the gradient in which the small contributions arising from the derivatives of the electrostatic potentials are neglected. The latter contributions will be addressed in a subsequent effort. The gradient was implemented in a development version of the GAMESS (General Atomic and Molecular Electronic Structure System) program package[30] and fully parallelized with the GDDI.[12]

For molecular clusters the Mulliken point charge representation of the electrostatic potential in the FMO method was used.[31] More specifically, all fragment calculations were performed in the field of point charges derived self-consistently from the fragment densities, and were repeated until their densities converged with respect to the field (dimers and trimers are computed in the converged monomer field only once). The core electrons (e.g., 1s on both C and O) were not correlated during the MP2 calculations. Otherwise, all calculations discussed below used the default values for all FMO method approximations.

3. Results

Test calculations are organized by first evaluating the accuracy of the open-shell FMO method for hydroxyl radical solvated by seven water molecules using the same geometry for both FMO and *ab initio* calculations. Next, the accuracy of the newly developed analytic gradient on solvated phenol and polypeptides is investigated, followed by full geometry optimizations of a solvated phenol system, two isomers of a polypeptide, as well as a small test system composed of the products of a reversible addition-fragmentation transfer (RAFT) reaction. Finally, the open-shell FMO method is applied to a small protein.[32]

3.1 Accuracy of Open-Shell FMO Energies for a Solvated Hydroxyl Radical

The use of water clusters provides a convenient test system for non-bonded molecular clusters. The ability to accurately model water clusters requires a proper treatment of three-body effects, as has been shown previously[17,18,33]. The implications of solvated OH radicals can be found in many biological processes and atmospheric reactions.[34] In addition to three-body effects, other effects such as electron correlation and strong

www.manaraa.com

charge transfer adds another degree of difficulty to the solvated OH clusters when compared to purely water clusters.

The structures of the eight $OH(H_2O)_7$ clusters considered in this work were determined by the following procedure using previously determined minima[35]:

1. The structures of $(H_2O)_8$ were optimized at the MP2 level of theory using the aug-cc-pVTZ basis set.[36]
2. One hydrogen atom was arbitrarily removed from a single water molecule in each cluster to create an open-shell system with doublet multiplicity.
3. The OH radical in each cluster plus the two nearest neighbor H_2O molecules comprise one fragment. The other fragments each contain one water molecule.
4. Test calculations were performed to determine if any additional merging into larger (and fewer) fragments was advisable based upon the level of charge transfer between fragments.
5. Single point MP2 energy calculations were performed with *ab initio* (fully MP2), FMO2 and FMO3.

The final fragmentation scheme can be found in Table 1, while Figure 1 shows the eight isomers used including the location of the open shell $OH^{\cdot}$ radical. The names assigned to the clusters are taken from previous work[35].

Errors for both the ROHF and ROMP2 levels of theory, calculated as

$$\text{Error} = E^{FMOn} - E^{ab\ initio} \qquad (4)$$

where FMO*n* represents the level of FMO used, are shown in Tables 2 and 3. All subsequent errors were calculated using eq 4. The FMO2 error for the aug-cc-pVDZ basis set is in the range 0.5-10 kcal/mol for ROHF (Table 2) and 0.1-9.3 kcal/mol for ROMP2 (Table 3), with mean absolute errors (MAE) of 5.1 kcal/ and 4.3 kcal/mol for ROHF and ROMP2, respectively. The FMO3 errors are much lower in both cases, 0.1-1.9 kcal/mol for both ROHF and ROMP2, with MAE values of 0.5 and 0.7 kcal/mol respectively. Improving the basis set to aug-cc-pVTZ reduces the FMO2 errors slightly; however the MAE is still 3.4 kcal/mol for ROHF and 3.1 kcal/mol for ROMP2. For both levels of theory the FMO2 errors are much too large to be considered reliable, although improvements can be made by placing more than one water molecule in each fragment. With the aug-cc-pVTZ basis set, the FMO3 errors also decrease (a similar trend was observed for closed shell systems[31]), producing a MAE of 0.24 kcal/mol for ROHF and 0.35 kcal/mol for ROMP2.

More important than errors in the total energies is how well the relative energies of the different isomers are captured. Table 4 shows the ROMP2 relative energies of the ten clusters, with Figures 2-5 showing both the ROHF and ROMP2 relative energies graphically. Both ROHF and ROMP2 relative energies reinforce the conclusions drawn based upon the total energy errors, namely the unreliability of FMO2 for this

www.manaraa.com

particular system, as well as the improved agreement with *ab initio* results with an increase in basis set size. The MAE for the FMO3 relative energies is less than 1 kcal/mol for both the aug-cc-pVDZ and aug-cc-pVTZ basis sets at 0.96 and 0.30 kcal/mol respectively.

Timing comparisons were also obtained to analyze the performance of the open-shell FMO method for energy calculations using the lowest energy isomer (S4) of the $OH(H_2O)_7$ clusters. MP2 energy calculations using ROMP2, FMO2-ROMP2 and FMO3-ROMP2 were performed on a Cray XE6 computer with two 2.4 GHz AMD Opteron 64-bit 8-core processors and 32 GB of RAM per node. All FMO calculations used the GDDI, assigning individual fragment n-mer calculations to separate nodes or "groups", each group performing its fragment calculation in parallel. Both *ab initio* and FMO timings were performed on 4, 6 and 8 nodes, for totals of 4, 6 and 8 groups when using the GDDI with FMO. The timings in Table 5 show the *ab initio*, FMO2 and FMO3 calculations scale quite well when doubling the number of nodes. In terms of time savings, the FMO2 calculations take an order of magnitude less time than the full *ab initio* calculations while the FMO3 calculations provide no time savings for a system this small.

While there is little time savings achieved for FMO3 for these small clusters, the real advantage of using FMO3 for a system of this size is the memory requirements, illustrated in Table 5. Compared to the computational time scaling of N^5, MP2 memory requirements generally scale as N^4, while the FMO memory requirement is determined by the largest *n*-mer size (dimer for FMO2 and trimer for FMO3) which is much smaller than the full system. This reduction in memory requirements through the use of the FMO method is very powerful for large system calculations since the memory requirements are independent of the total system size *N*.

3.2 Comparison of FMO-ROHF Analytic Gradient and ab initio Analytic Gradient

The accuracy of the FMO-ROHF gradient was investigated by explicit comparison with *ab initio* fully analytic gradient calculations using the solvated phenol and polypeptide test systems at both their equilibrium geometries and at selected non-equilibrium structures. Timing comparisons were also made between the fully analytic *ab initio* gradient and the FMO2 analytic gradient to show the efficacy of FMO2 for geometry optimizations. Due to the relatively small size of the test systems, a timing comparison between *ab initio* and FMO3 gradients was not made.

Results from single point gradient calculations at equilibrium geometries are shown in Table 6. The maximum gradient as well as the root mean squared (RMS) gradient and the mean absolute error (MAE) are compared for all three structures. For the two isomers of $(Ala)_2Phe(Ala)_2$, the error in the maximum gradient is between 0.01 and 0.03 mH/bohr for both FMO2 and FMO3; the error in the RMS gradient is less than 0.01 mH/bohr in all cases. The error in the maximum gradient for the solvated phenol system is comparable at 0.01 mH/bohr for FMO2 and FMO3, while the error in the RMS gradient is similar to that of the $(Ala)_2Phe(Ala)_2$

www.manaraa.com

system at less than 0.01 mH/bohr for both FMO2 and FMO3. In all cases, the MAE is less than $2.6x10^{-5}$ mH/bohr. Corresponding results from single point gradient calculations at non-equilibrium geometries are shown in Table 7. The error in the maximum gradient for the two isomers of $(Ala)_2Phe(Ala)_2$ is between 0.5 and 1.1 mH/bohr for both FMO2 and FMO3; the error in the RMS gradient is less than 0.1 mH/bohr in all cases. The error in the maximum gradient for the solvated phenol system is 0.2 mH/bohr for FMO2 and 0.1 mH/bohr for FMO3. The error in the RMS gradient is significantly smaller than those of the $(Ala)_2Phe(Ala)_2$ systems, with errors of $8.0x10^{-6}$ mH/bohr for FMO2 and $1.7x10^{-5}$ for FMO3. In all cases, the MAE is less than 1.5 mH/bohr.

Timing comparisons on the initial geometry of the α-$(Ala)_2Phe(Ala)_2$ system were performed on four nodes of a Cray XE6 computer with two 2.4 GHz AMD Opteron 64-bit 8-core processors and 32 GB of RAM per node. All FMO calculations employed the GDDI, with group division chosen to be one node per group. The *ab initio* ROHF gradient calculation took approximately 362 seconds while the FMO2-ROHF gradient calculation only took approximately 249 seconds. This corresponds to a time savings of approximately 30% compared to the full *ab initio* calculation.

3.3 Open-shell Optimizations and Adiabatic Excitation Energies of $C_6H_5OH(H_2O)_8$

To further investigate the accuracy of the open-shell FMO method compared to conventional *ab initio* methods, first consider a closed shell cluster consisting of one phenol molecule solvated by eight water molecules. The FMO fragmentation scheme used for this system, shown in Figure 6, places the phenol molecule and the two closest H-bonded water molecules in one fragment. Each of the remaining water molecules was chosen as a single fragment, creating seven fragments total. Both FMO2 and FMO3 optimizations were performed using the same fragmentation scheme. The starting structure for both *ab initio* and FMO optimizations was obtained by placing five of the eight water molecules approximately 5-6 Å away from the phenol molecule on the x, y and z coordinate axes, with the remaining three molecules placed around the OH moiety (Figure 7). This starting structure was then optimized using the EFP1 method[37] to model the water molecules and RHF with the STO-3G basis set to model the phenol molecule, with the resultant geometry used as the initial structure in all further optimizations. RHF and ROHF optimizations were performed. Closed and open shell MP2 energy calculations were then performed at the RHF and ROHF equilibrium geometries. For the *ab initio* ROHF calculation the multiplicity of the system was chosen as a triplet, with the singly occupied orbitals located on the phenol molecule. For both FMO2 and FMO3 calculations the triplet electronic configuration was placed on the fragment containing the phenol molecule, because the triplet state is mainly localized in this region.

Table 8 shows the errors in kcal/mol for FMO2 and FMO3 total energies relative to fully *ab initio* energies, using the 6-31G(d) basis set. For the closed shell (singlet) optimizations, the FMO2 method produces an error of -0.23 kcal/mol, while the FMO2 error for the open-shell (triplet) optimizations is -0.48 kcal/mol. As

www.manaraa.com

expected for a system containing more than two water molecules, the FMO3 method produces smaller errors of 0.07 kcal/mol for the singlet optimizations and 0.08 kcal/mol for the triplet optimizations. For the FMO3 optimizations it is apparent that the size of the error is consistent between singlet and triplet states, while the FMO2 errors approximately double upon going from the singlet to the triplet. Despite this doubling of the FMO2 error, the actual values are all less than 0.50 kcal/mol, which is well within chemical accuracy. The errors for the MP2 energy calculations at the RHF and ROHF optimized geometries are also all less than 1.00 kcal/mol. Adiabatic excitation energies were also calculated, giving FMO2 errors of -0.25 kcal/mol and -0.70 kcal/mol for the HF and MP2 levels of theory, respectively. FMO3 again outperforms FMO2 for the excitation energies, giving errors of less than 0.10 kcal/mol for both levels of theory.

The root mean squared deviation (RMSD) between the *ab initio* and FMO optimized structures was computed using the unit quaternion method[38a] implemented in the freely available program Jmol.[38b] Table 9 shows the RMSD values for both the singlet and triplet geometries optimized with FMO2 and FMO3. The RMSD value for FMO2 is lower for the singlet structure than for the triplet structure, with values of 0.038 Å and 0.254 Å respectively. The performance of FMO3 is significantly better, producing a RMSD value of 0.013 Å for the singlet structure and 0.006 Å for the triplet structure. As mentioned previously for the $OH(H_2O)_7$ system, the improved accuracy of the FMO3 method for water clusters is not surprising considering the importance of many-body polarization effects. Despite the decrease in accuracy for the FMO2 optimized structures, the overall performance of FMO2 is acceptable. The higher-order many-body effects could be captured using FMO2 by having two water molecules per fragment, thereby lowering the errors in energy as well as the RMSD values, but this would increase the computational expense.

3.4 Open-shell Optimizations and Adiabatic Excitation Energies of $(Ala)_2Phe(Ala)_2$ Chains

To test the ability of the FMO-ROHF method to treat larger, covalently bonded systems, two different isomers of the $(Ala)_2Phe(Ala)_2$ chain were chosen for geometry optimizations. Each geometry optimization was performed using HF, FMO2-HF and FMO3-HF with the 6-31G(d) basis set, followed by MP2, FMO2-MP2 and FMO3-MP2 single point energy calculations. Both singlet and triplet structures were optimized with the above methods and adiabatic excitation energies from the singlet to the triplet state were calculated. The default FMO method settings for the electrostatic field were used for all optimizations and energy calculations. A one amino acid residue per fragment partition was used, and the open shell fragment was chosen to contain only phenylalanine.

Table 10 shows the errors in the energy for FMO optimized structures relative to the corresponding *ab initio* optimized structures, as well as the errors in the MP2 energies calculated at the RHF and ROHF optimized geometries. For all FMO2 HF optimized structures the errors in the total energy are between -0.47 and -1.89 kcal/mol, with errors in the excitation energies of -0.85 kcal/mol for the alpha isomer and -0.49 kcal/mol for the beta isomer. The MP2 energies calculated at the HF geometries show a decrease (relative to

www.manaraa.com

HF) in the error of the FMO2 total energy to 0.17 and 0.64 kcal/mol for the singlet and triplet structures of the alpha structure, respectively, with an error of -0.47 kcal/mol for the excitation energy. However, the MP2 total energy errors increase to -1.48 and -2.59 kcal/mol for the singlet and triplet structures of the beta isomer, while the error in excitation energy only increases slightly to -1.11 kcal/mol. The FMO3 errors in the total HF energy for the optimized structures are 0.01 and -0.62 kcal/mol for the singlet and triplet alpha isomer structures respectively, with an error of 0.61 kcal/mol in the excitation energy. For the beta isomer, FMO3 produces total HF energy errors of 0.02 and -0.49 kcal/mol for the optimized singlet and triplet structures and an error of -0.47 kcal/mol in the excitation energy. MP2 single point energy calculations performed with FMO3 give total energy errors of -0.09 and -0.32 kcal/mol for the singlet and triplet states of the alpha isomer. The beta isomer again gives slightly worse results for FMO3, with total energy errors of -0.04 and -0.68 kcal/mol for the singlet and triplet structures. The errors in the FMO3 excitation energies are -0.23 kcal/mol for the alpha isomer and -0.64 kcal/mol for the beta isomer. For both FMO2 and FMO3, the triplet structures of the $(Ala)_2Phe(Ala)_2$ isomers give errors that are much worse than those obtained in the corresponding singlet optimizations. However, in all but one case, FMO3 reduces the error by greater than half of the FMO2 errors. The one exception to this trend is the FMO3-ROMP2 energy, which is roughly twice as large as the FMO2-ROMP2 error for the triplet state of the alpha isomer. All excitation energies produced by FMO3 are between 0.02 and 0.47 kcal/mol lower than the FMO2 excitation energies.

The total RMSD values for FMO2 and FMO3 compared to the *ab initio* optimized structures are shown in Table 11. The FMO2 geometries are within 0.563 Å of the *ab initio* geometries in the worst case, corresponding to the triplet state of the beta isomer. The singlet state of the beta isomer produces a similar error of 0.484 Å for FMO2. The FMO2 RMSD values are much lower for the alpha isomer at 0.177 and 0.210 Å for the singlet and triplet structure respectively. FMO3 performs significantly better than FMO2 in all cases except the triplet structure of the alpha isomer. The error in this one case is 0.254 Å, compared to an error of 0.210 Å for FMO2. This is indicative of a decreased importance of three-body effects in the triplet state of the alpha isomer. FMO3 produces errors of 0.026 and 0.019 Å for the singlet structures of the alpha and beta isomers, as well as an error of 0.176 Å for the triplet state of the beta isomer.

3.5 Optimization of $(CH_3)_2C(CN)$-CH_2-CH(Ph)-CH_2-C•H(Ph)

As an additional test of the ability of the open-shell FMO method to optimize a small system, the open-shell reactant from the initiation step in a RAFT reaction[39] was chosen. The initiation step consists of two reactants that combine to form an open-shell radical product, shown in Figure 8. The initial structure of the open-shell reactant of this reaction was optimized using DFT with the B3LYP functional and the 6-31G(d) basis set.[39c] This structure was then optimized with both HF and FMO3 at the ROHF/6-31G(d) level of theory, with the multiplicity of the system chosen as a doublet in both cases. Single point MP2 energy calculations were then performed at the respective ROHF optimized geometry for each method.

www.manaraa.com

The fragmentation scheme for the FMO calculations is illustrated in Figure 9, with the backbone chosen as one fragment, each phenyl group chosen as a fragment and the $(CH_3)_2C(CN)$ moiety chosen as a fragment, for a total of four fragments. The open-shell (doublet) fragment is shown in orange in Figure 9, with the open-shell carbon atom circled in black.

An FMO2 optimization using the same fragmentation scheme as the FMO3 optimization was unable to converge to an equilibrium geometry. The inability of FMO2 to converge to an equilibrium geometry, coupled with the results discussed for the $(Ala)_2Phe(Ala)_2$ chains in subsection 3.4 showing that the use of FMO3 provides marked reductions in the structural RMSD compared to *ab initio* structures, further illustrates the importance of using FMO3 for optimizations of smaller open-shell systems composed of fewer than ~40-50 heavy atoms. However, larger systems can benefit from larger fragment choices, thereby improving the accuracy of the calculation and potentially facilitating the use of FMO2.

The overall accuracy of the FMO3 optimization of $(CH_3)_2C(CN)$-CH_2-CH(Ph)-CH_2-C•H(Ph) is quite good, with an error in the total energy of the system of 0.05 kcal/mol compared to the full *ab initio* optimization. The MP2 single point energy calculation also shows good agreement with *ab initio* results, with a difference in energy of -0.53 kcal/mol. The structure obtained from the FMO3 optimization is compared to the *ab initio* structure in Figure 10. The RMSD value between these two structures is 0.172 Å.

3.6 Adiabatic Excitation Energy Calculation of a Small Protein

To demonstrate the ability of the FMO-ROHF method to perform open-shell calculations on a large system, the adiabatic excitation energy of a small protein, chignolin (PDB ID: 1UAO), was calculated. The initial structure, obtained from the PDB database, was optimized with the 6-31G(d) basis set using FMO2-RHF and FMO2-ROHF, followed by FMO2-MP2 single point calculations for both the closed shell (singlet) and open-shell (triplet) states. The fragmentation scheme chosen for the geometry optimizations, shown in Figure 11, is two amino acid residues per fragment creating a total of 5 fragments. The fragmentation scheme used for the open-shell calculations is the same, with the fragment having triplet multiplicity chosen to contain the tryptophan residue. The calculated adiabatic excitation energy is 68.4 kcal/mol and 91.9 kcal/mol for HF and MP2 respectively (Table 12).

In order to provide some assessment of the accuracy of the FMO-ROHF method on a large system such as chignolin, full *ab initio* MP2 energy calculations were performed at the FMO2 optimized structures. The other systems investigated with the open-shell FMO method up to this point have been relatively small and in all cases the corresponding *ab initio* calculations have not been prohibitively expensive to perform. This is not the case for chignolin, particularly for the MP2 single point calculation. While the full geometry optimizations on this system were not possible, full MP2 single point energy calculations were performed. The error between the FMO2 and *ab initio* energies is ~0.7 kcal/mol for the singlet and triplet HF optimized

www.manaraa.com

structures while the error between *ab initio* and FMO2 for the MP2 energy calculations is -0.07 and -0.60 kcal/mol for the singlet and triplet structures respectively.

Both the FMO2 and *ab initio* singlet state MP2 energy calculations were performed on 10 nodes of a Cray XE6 computer with two 2.4 GHz AMD Opteron 64-bit 8-core processors and 32 GB of RAM per node. The FMO2 calculation took approximately 30 minutes while the full *ab initio* calculation took 181.5 minutes. However, the triplet state *ab initio* MP2 energy calculation requires significantly more memory, using 33 nodes on the same Cray XE6, or 528 cores and 1 TB of RAM. The same FMO2-MP2 energy calculation on the triplet state was performed on only 10 nodes, requiring approximately 0.1 GB of RAM per node, or four orders of magnitude less memory. Even with more than three times the number of nodes, the *ab initio* calculation took 140.7 minutes on 33 nodes while the FMO2 calculation took only 47 minutes on 10 nodes. It is important to note that the FMO2 calculation does not require such a large computer and could easily be performed on a much smaller computer system.

4. Conclusions

The gradients for the open-shell FMO method have been derived and then implemented in the GAMESS program package and parallelized using the GDDI for the ROHF level of electronic structure theory. The ability of the FMO-ROHF method to reproduce accurate total energies and geometries for a variety of chemical systems with varying multiplicities was tested. The accuracy of adiabatic excitation energies was also investigated and it was demonstrated that the open-shell FMO method is capable of producing both accurate geometries as well as adiabatic excitation energies within 0.01 to 0.85 kcal/mol of *ab initio* calculations. Timings and memory requirements for the relatively small test systems also show the ability of the open-shell FMO method to provide a route to geometry optimizations on larger systems.

This work contributes a scalable method for geometry optimizations on large chemical systems through the implementation of the gradient for the single-reference open-shell FMO method. Through the combination of reduced computational cost as well as chemical accuracy shown, the open-shell FMO method provides a means for accurate geometry optimizations on open-shell radical systems much larger than previously possible.

References

1. Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419.
2. Gao, J. *J. Phys. Chem. B* **1997,** *101*, 657.
3. Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.
4. Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comp.* **2009**, *5*, 1573.
5. Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comp.* **2009**, *5*, 1667.

www.manaraa.com

6. 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.

7. Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927.

8. Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.*, in press.

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

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

11. Fedorov, D.G.; Kitaura, K.; *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009.

12. (a) Fletcher, G.D.; Schmidt, M.W.; Bode, B.M.; Gordon, M.S. *Comp. Phys. Comm.* **2000**, *128*, 190. (b) Fedorov, D.G.; Olson, R.M.; Kitaura, K.; Gordon, M.S.; Koseki, S. *J. Comp. Chem.* **2004**, *25*, 872. (c) Fletcher, G. D.; Fedorov, D. G.; Pruitt, S. R.; Windus, T. L.; Gordon, M. S. *J. Chem. Theory Comput.*, **2012**, *8*, 75.

13. (a) Lee, T.J.; Rendell, A.P.; Dyall, K.G.; Jayatilaka, D. *J.Chem.Phys.* **1994**, *100*, 7400. (b) Roskop, L.; Fedorov, D.G.; Gordon, M.S. *Chem. Phys. Lett.*, in preparation.

14. (a) Włoch, M.; Gour, J.R.; Piecuch, P. *J. Phys. Chem. A*, **2007**, *111*, 11359. (b) Sawada, T.; Fedorov, D. G.; Kitaura, K. *J. Am. Chem. Soc.* **2010**, *132*, 1686–16872.

15. Fukunaga, H.; Fedorov, D. G.; Chiba, M.; Nii, K.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 10887.

16. Pruitt, S.R.; Fedorov, D.G.; Kitaura, K.; Gordon, M.S. *J. Chem. Theory Comput.* **2010**, *6*, 1

17. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.

18. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.

19. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108.

20. Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.

21. Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K.; Gordon, M. S. *J. Chem. Phys.* **2011**, *134*, 034110.

22. Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.

23. Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.

24. Fedorov, D. G.; Alexeev, Y.; Kitaura, K. *J. Phys. Chem. Lett.* **2011**, *2*, 282.

25. Mochizuki, Y.; Yamashita, K.; Murase, T.; Nakano, T.; Fukuzawa, K.; Takematsu, K.; Watanabe, H.; Tanaka, S *Chem. Phys. Lett.* **2008**, *457*, 396.

26. T. Nagata, D. G. Fedorov, K. Kitaura. Mathematical Formulation of the fragment molecular orbital method. In *Linear-Scaling Techniques in Computational Chemistry and Physics*. R. Zalesny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski (Eds.), Springer, New York, 2011, pp. 17-64.

27. DeFusco, A.; Minezawa, N.; Slipchenko, L.V.; Zahariev, F.; Gordon, M.S. *J. Phys. Chem. Lett.* **2011**, *2*, 2184.

28. See Kitaura, K.; Sugiki, S.I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.*, **2001**, *336*, 163 for a description of the closed shell gradient method.

www.manaraa.com

29. Chiba, M.; Fedorov, D.G.; Nagata, T.; Kitaura, K. *Chem. Phys. Lett.* **2009**, *474*, 227.

30. (a) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. et al. *J. Comput. Chem.* **1993**, *14*, 1347. (b) Gordon, M. S.; Schmidt, M. W. *Advances in Electronic Structure Theory: GAMESS a Decade Later* in Theory and Applications of Computational Chemistry, the first forty years, Elsevier, Amsterdam, 2005.

31. Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. *J. Phys. Chem. A* **2010**, *114*, 8742.

32. Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. *Structure*, **2004**, *12*, 1507.

33. (a) Xantheas, S. S.; Dunning Jr., T. H. *J. Chem. Phys.* **1993**, *98*, 8037 (b) Xantheas, S. S. *J. Chem. Phys*. **1994**, *100*, 7523 (c) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163 (d) Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225 (e) Xantheas, S. S. *Struct. Bonding* **2005**, *116*, 119

34. Seinfeld, J.H.; Pandis, S.N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; Wiley, New York, NY, 2006.

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

36. (a) Dunning Jr., T. H *J. Chem. Phys*. **1989**, *90*, 1007. (b) Woon, D. E.;Dunning, Jr., T. H. *J. Chem. Phys*. **1995**, *103*, 4572.

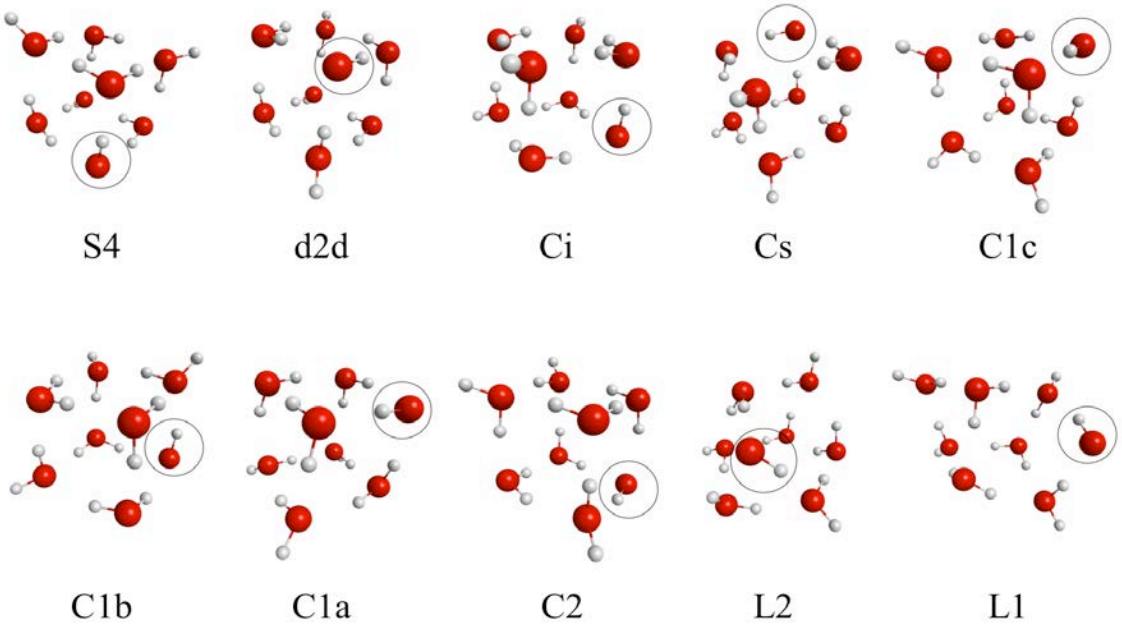
37. Day, P.N.; Jensen, J.H.; Gordon, M.S.; Webb, S.P.; Stevens, W.J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968-1986.

38. (a) Horn, B.K.P. *J. Opt. Soc. Am.* **1987**, *4*, 629. (b) Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

39. (a) Coote, M.L.; Izgorodina, E.I.; Krenske, E.H.; Busch, M.; Barner-Kowollik, C. *Macromol. Rapid Comm.* **2006**, *27*, 1015. (b) Chiefari, J.; Chong, Y.K.; Ercole, F.; Krstina, J.; Jeffery, J.; Le, T.P.T; Mayadunne, R.T.A.; Meijs, G.F.; Moad, C.L.; Moad, G.; Rizzardo, E.; Thang, S.H. *Macromol.* **1998**, *31*, 5559. (c) Structures used for the RAFT reaction calculations were obtained from Professor Michelle Coote at the Austrailian National University.

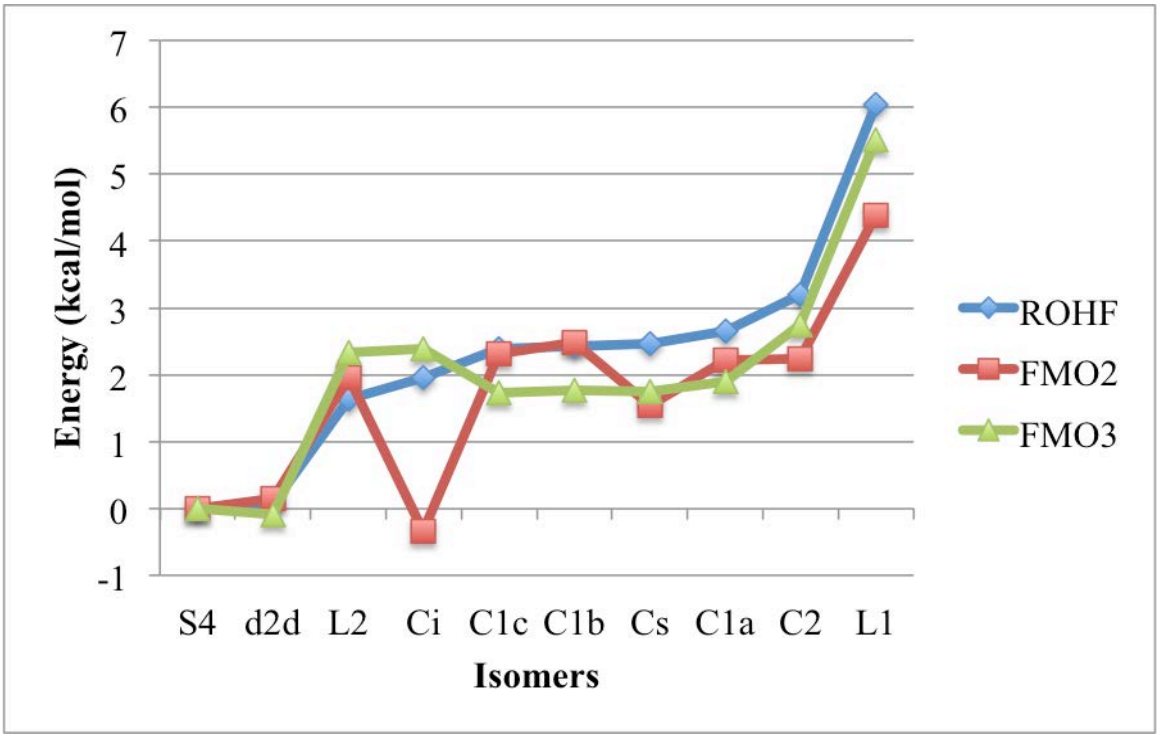
www.manaraa.com

Figure 1. Ten isomers of $OH(H_2O)_7$ clusters used for testing. Open shell OH molecules are circled. The naming convention for the isomers is taken from reference 28.



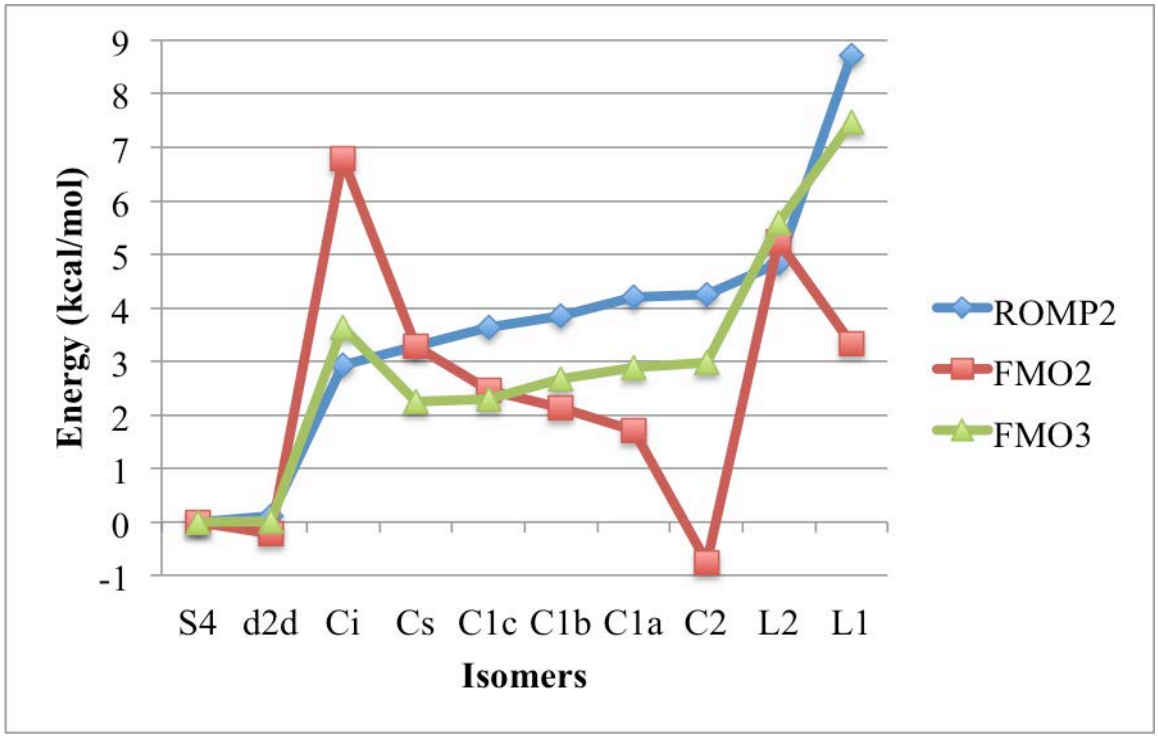
www.manaraa.com

Figure 2. Graph of the relative energies of the ten $OH(H_2O)_7$ clusters computed using *ab initio* ROHF, FMO2-ROHF and FMO3-ROHF with the aug-cc-pVDZ basis set.



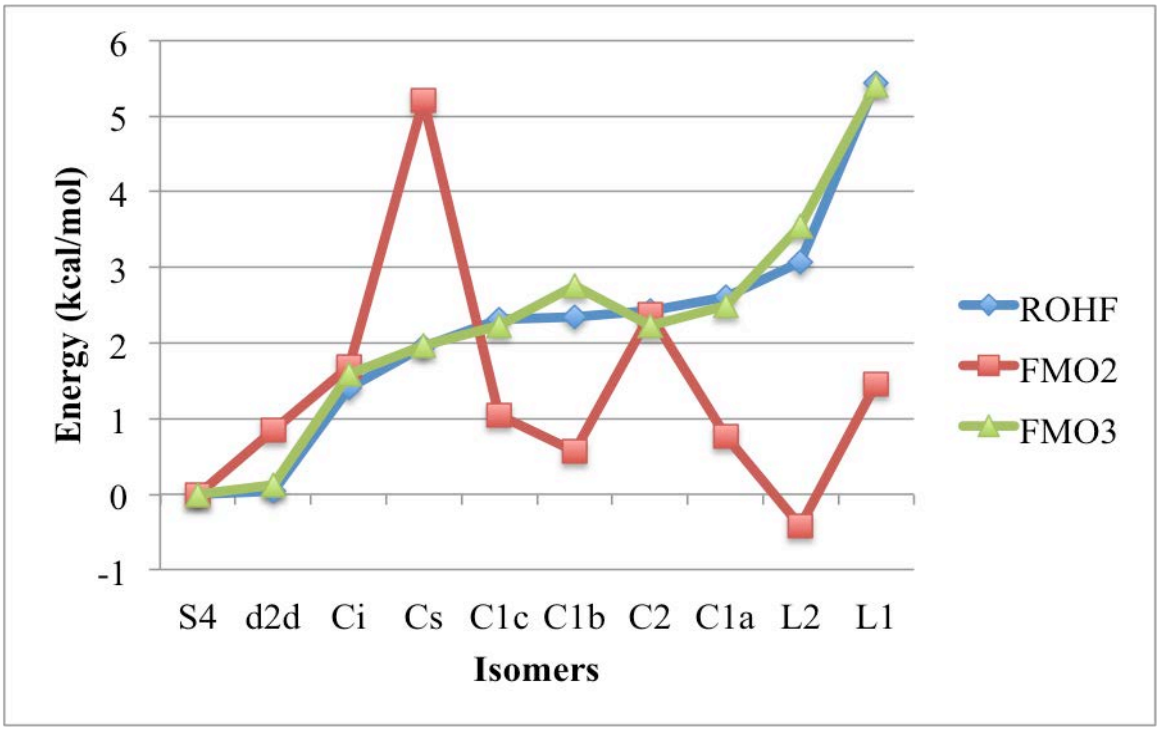
www.manaraa.com

Figure 3. Graph of the relative energies of the ten $OH(H_2O)_7$ clusters computed using *ab initio* ROMP2, FMO2-ROMP2 and FMO3-ROMP2 with the aug-cc-pVDZ basis set.



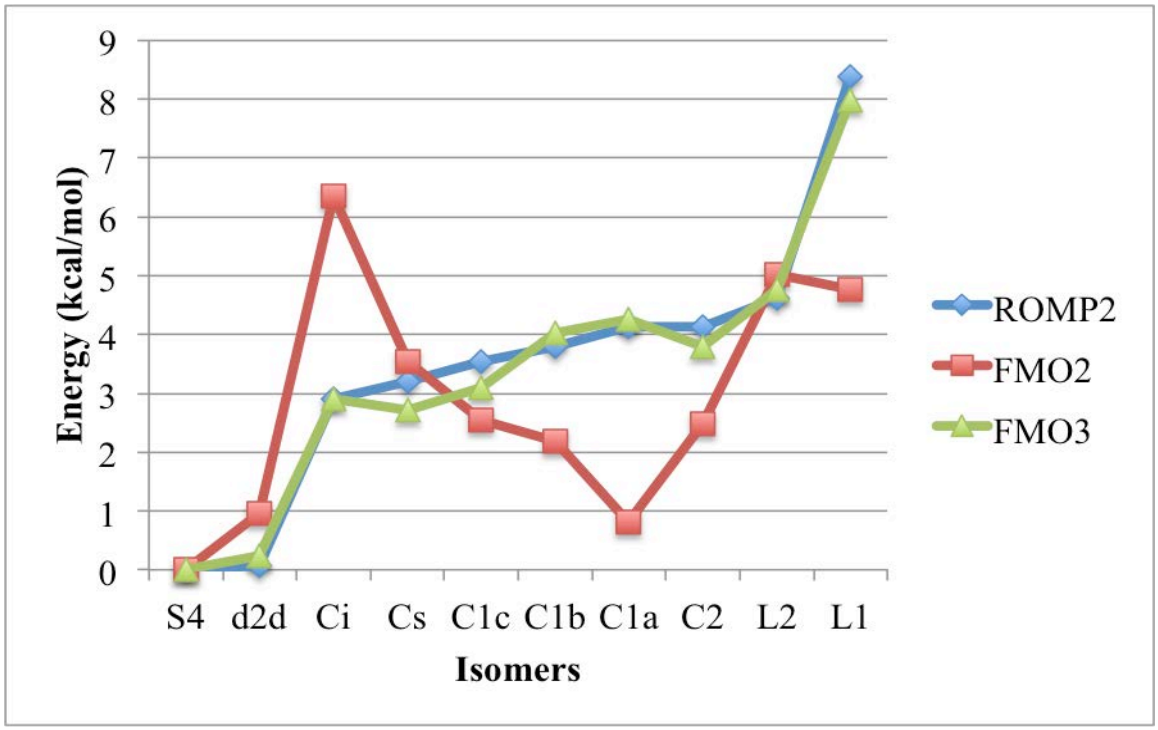
www.manaraa.com

Figure 4. Graph of the relative energies of the ten $OH(H_2O)_7$ clusters computed using *ab initio* ROHF, FMO2-ROHF and FMO3-ROHF with the aug-cc-pVTZ basis set.



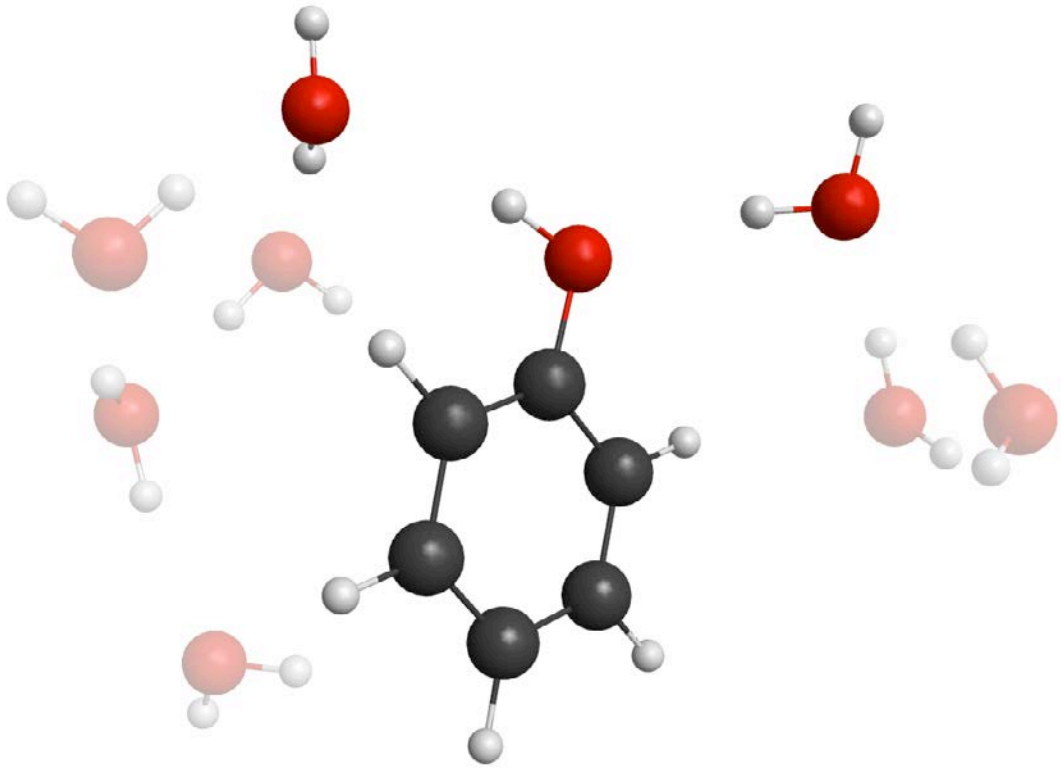
www.manaraa.com

Figure 5. Graph of the relative energies of the ten $OH(H_2O)_7$ clusters computed using *ab initio* ROMP2, FMO2-ROMP2 and FMO3-ROMP2 with the aug-cc-pVTZ basis set.



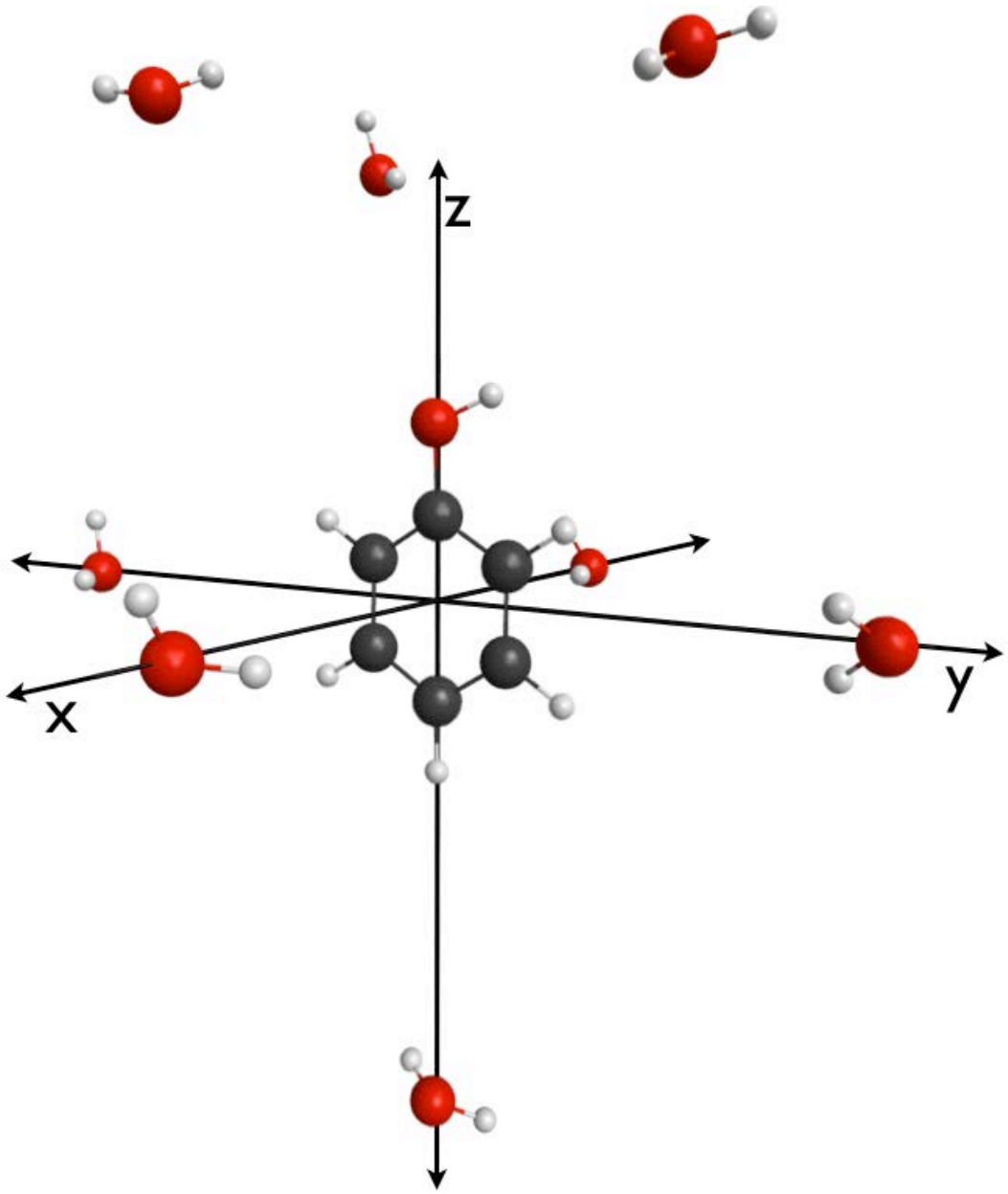
www.manaraa.com

Figure 6. Fragmentation scheme used for open-shell solvated phenol calculations. Highlighted molecules make up the triplet fragment during open-shell calculations, with each remaining water molecule assigned as a single fragment.



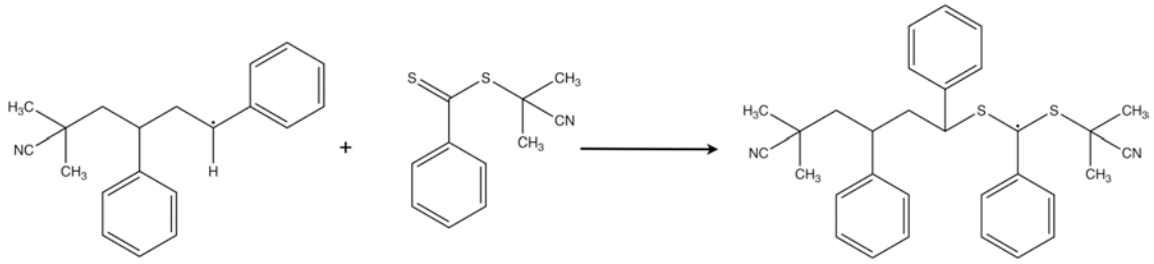
www.manaraa.com

Figure 7. Initial orientation of water molecules around phenol before EFP1/RHF optimization.



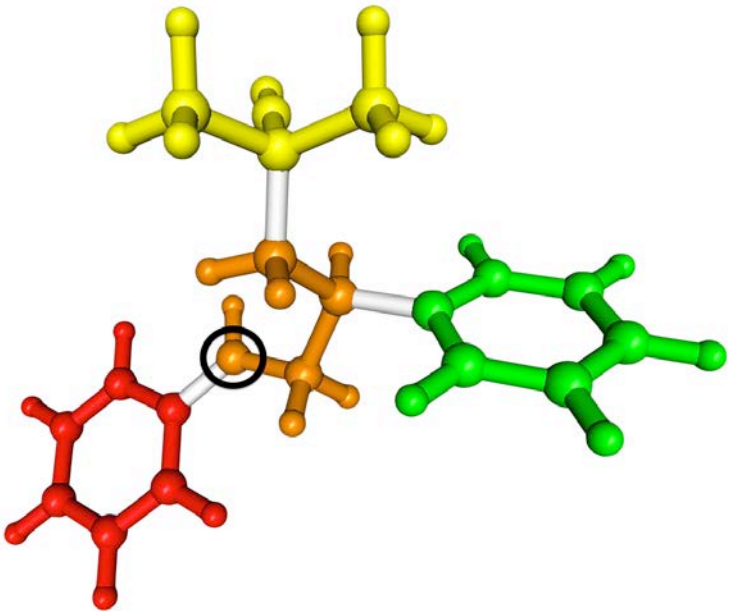
www.manaraa.com

Figure 8. Schematic representation of a RAFT reaction.



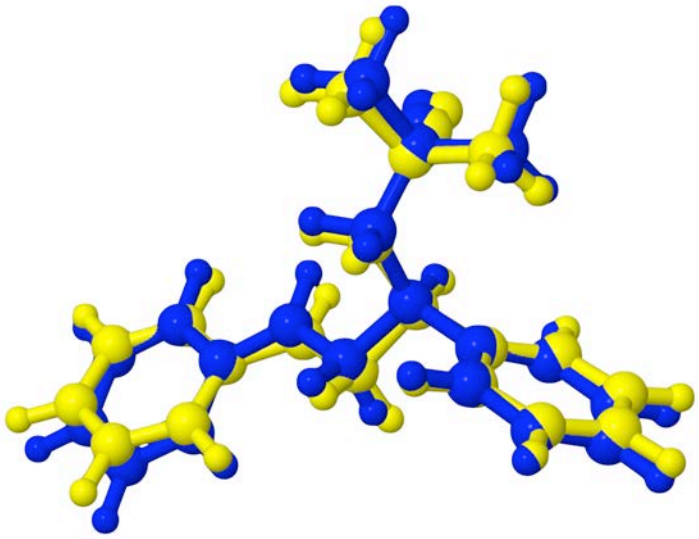
www.manaraa.com

Figure 9. Fragmentation scheme used for $(CH_3)_2C(CN)\text{-}CH_2\text{-}CH(Ph)\text{-}CH_2\text{-}C\bullet H(Ph)$. The open-shell fragment is shown in orange.



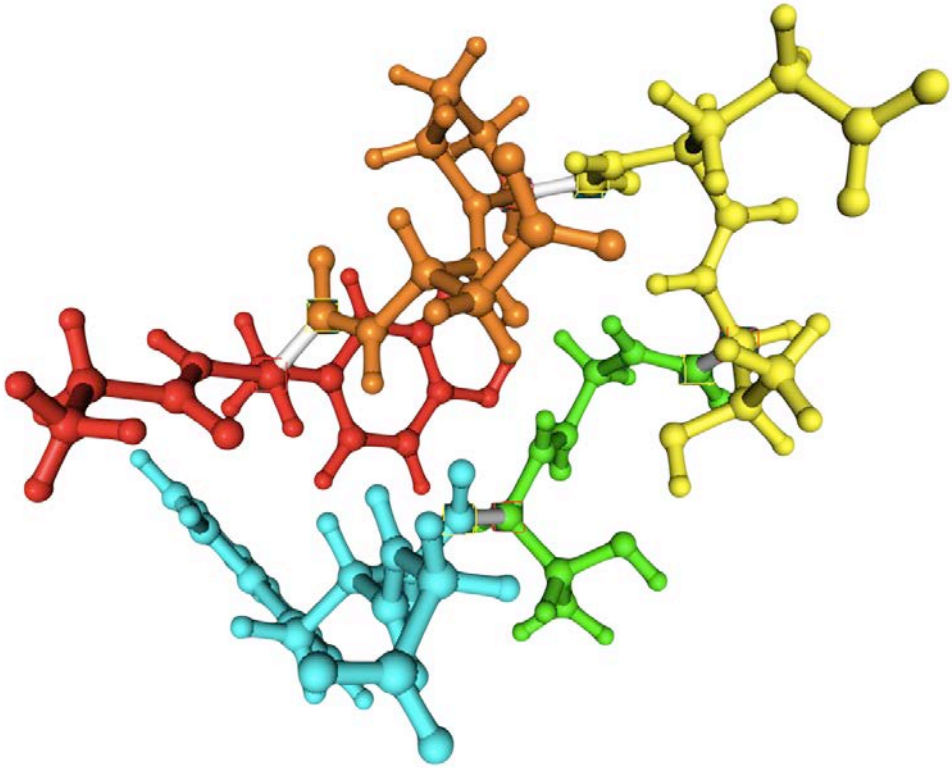
www.manaraa.com

Figure 10. Overlay of FMO3 optimized (yellow) and *ab initio* optimized (blue) structures of $(CH_3)_2C(CN)$-CH_2-CH(Ph)-CH_2-C•H(Ph).



www.manaraa.com

Figure 11. Fragmentation scheme used for chignolin (1UAO).



www.manaraa.com

Table 1. Fragmentation schemes used for ten $OH(H_2O)_7$ clusters.

	Number of fragments
S4	8
d2d	8
Ci	8
Cs	7
C1c	7
C1b	7
C1a	7
C2	5
L2	8
L1	5

www.manaraa.com

Table 2. Errors of FMO-ROHF(MP2) energies relative to *ab initio* ROHF(MP2) energies using the aug-cc-pVDZ basis set for ten $OH(H_2O)_7$ clusters.

	absolute errors (kcal/mol)			
	aug-cc-pVDZ			
Isomer	FMO2-ROHF	FMO3-ROHF	FMO2-ROMP2	FMO3-ROMP2
S4	6.54	0.43	5.47	1.15
d2d	6.35	0.28	5.16	1.07
Ci	10.22	0.88	9.33	1.87
Cs	6.10	-0.28	5.47	0.12
C1c	4.89	-1.08	4.30	-0.90
C1b	4.41	-0.23	3.74	-0.03
C1a	3.66	-0.33	2.97	-0.16
C2	1.16	-0.03	0.47	-0.11
L2	6.79	1.12	5.88	1.90
L1	0.47	-0.09	0.10	-0.09

www.manaraa.com

Table 3. Errors in the total energy between *ab initio* ROHF(MP2) and FMO-ROHF(MP2) using the aug-cc-pVTZ basis set for ten $OH(H_2O)_7$ clusters.

	errors in the total energy (kcal/mol)			
	aug-cc-pVTZ			
Isomer	FMO2-ROHF	FMO3-ROHF	FMO2-ROMP2	FMO3-ROMP2
S4	4.23	-0.06	3.71	0.30
d2d	5.03	0.02	4.61	0.48
Ci	7.49	-0.05	7.17	0.29
Cs	4.18	-0.25	4.04	-0.19
C1c	2.97	-0.89	2.71	-0.73
C1b	2.45	0.35	2.11	0.54
C1a	2.38	-0.17	2.06	-0.05
C2	0.73	0.41	0.39	0.44
L2	4.50	0.11	4.10	0.44
L1	0.26	-0.09	0.10	-0.09

www.manaraa.com

Table 4. Comparison of the ROMP2 and FMO-ROMP2 relative energies of ten $OH(H_2O)_7$ clusters using the aug-cc-pVDZ and aug-cc-pVTZ basis sets.

	relative energies (kcal/mol)					
	aug-cc-pVDZ			aug-cc-pVTZ		
isomer	FMO2-	FMO3-ROMP2	ROMP2	FMO2-ROMP2	FMO3-ROMP2	ROMP2
S4	0.00	0.00	0.00	0.00	0.00	0.00
d2d	-0.21	0.02	0.10	0.96	0.24	0.06
Ci	6.78	3.64	2.92	6.36	2.90	2.91
Cs	3.27	2.25	3.28	3.54	2.72	3.21
C1c	2.46	2.29	3.63	2.55	3.10	3.54
C1b	2.12	2.68	3.86	2.19	4.03	3.79
C1a	1.70	2.89	4.21	0.81	4.27	4.14
C2	-0.75	2.99	4.25	2.49	3.79	4.13
L2	5.24	5.59	4.84	5.01	4.77	4.63
L1	3.34	7.47	8.71	4.77	7.99	8.38

www.manaraa.com

Table 5. Timing comparison between *ab initio* and FMO single point energy calculations using the lowest energy $OH(H_2O)_7$ isomer (S4) at the MP2/aug-cc-pVTZ level of theory.

	Wall Clock Time (minutes)			Memory Requirements
	4 nodes	6 nodes	8 nodes	MB RAM/CPU
ROMP2	42.2	31.8	26.6	200
FMO2-ROMP2	3.7	3.1	2.3	9
FMO3-ROMP2	43.1	30.4	25.8	16

www.manaraa.com

Table 6. FMO-ROHF/6-31G(d) analytic gradient (in hartree/bohr) compared to ROHF/6-31G(d) fully analytic gradient at equilibrium geometries.

	Fully Analytic	FMO2	Error	FMO3	Error
			α-$(Ala)_2Phe(Ala)_2$		
Maximum Gradient	0.0000590	0.0000722	0.0000132	0.0000952	0.0000362
RMS Gradient	0.0000166	0.0000250	0.0000084	0.0000237	0.0000071
MAE			0.0000228		0.0000216
			β-$(Ala)_2Phe(Ala)_2$		
Maximum Gradient	0.0000860	0.0000957	0.0000097	0.0000481	-0.0000379
RMS Gradient	0.0000183	0.0000333	0.0000150	0.0000064	-0.0000119
MAE			0.0000298		0.0000138
			$C_6H_5OH(H_2O)_8$		
Maximum Gradient	0.0000727	0.0000538	-0.0000189	0.0000847	0.0000120
RMS Gradient	0.0000251	0.0000174	-0.0000077	0.0000278	0.0000027
MAE			0.0000218		0.0000259

www.manaraa.com

Table 7. FMO-ROHF/6-31G(d) analytic gradient (in hartree/bohr) compared to ROHF/6-31G(d) fully analytic gradient at non-equilibrium geometries.

	Fully Analytic	FMO2	Error	FMO3	Error
			α-$(Ala)_2Phe(Ala)_2$		
Maximum Gradient	0.0092148	0.0096807	0.0004659	0.0093471	0.0001323
RMS Gradient	0.0021575	0.0023306	0.0001731	0.0022209	0.0000634
MAE			0.0003201		0.0000640
			β-$(Ala)_2Phe(Ala)_2$		
Maximum Gradient	0.0079401	0.0068328	-0.0011073	0.0074678	-0.0004723
RMS Gradient	0.0020780	0.0020709	-0.0000071	0.0011626	-0.0009154
MAE			0.0007022		0.0013018
			$C_6H_5OH(H_2O)_8$		
Maximum Gradient	0.0070689	0.0072404	0.0001715	0.0069375	-0.0001314
RMS Gradient	0.0017368	0.0017288	-0.0000080	0.0017197	-0.0000171
MAE			0.0001502		0.0000337

www.manaraa.com

Table 8. Total energy deviations calculated between FMO-ROHF and *ab initio* ROHF optimized geometries for $C_6H_5OH(H_2O)_8$.

	absolute errors (kcal/mol)			
	HF		MP2	
	FMO2	FMO3	FMO2	FMO3
Singlet	-0.23	0.07	0.69	0.13
Triplet	-0.48	0.08	0.01	0.19
Excitation	-0.25	0.01	-0.70	0.06

www.manaraa.com

Table 9. Root mean squared deviations (RSMD) between fully *ab initio* optimized structures and FMO optimized structures of $C_6H_5OH(H_2O)_8$.

	RMSD (Å)	
	6-31G(d)	
	FMO2	FMO3
Singlet	0.038	0.013
Triplet	0.254	0.006

www.manaraa.com

Table 10. Errors in isomer and adiabatic excitation energies for FMO-ROHF geometry optimizations and MP2 single point energy calculations relative to full *ab initio* optimized alpha and beta isomers of $(Ala)_2Phe(Ala)_2$.

	absolute errors (kcal/mol)			
	HF		MP2	
	FMO2	FMO3	FMO2	FMO3
α-$(Ala)_2Phe(Ala)_2$				
Singlet	-1.04	0.01	0.64	-0.09
Triplet	-1.89	-0.62	0.17	-0.32
Excitation	-0.85	0.61	-0.47	-0.23
β-$(Ala)_2Phe(Ala)_2$				
Singlet	-0.47	0.02	-1.48	-0.04
Triplet	-0.97	-0.49	-2.59	-0.68
Excitation	-0.49	-0.47	-1.11	-0.64

www.manaraa.com

Table 11. Root mean squared deviations (RSMD) between *ab initio* optimized structures and FMO optimized structures of alpha and beta isomers of $(Ala)_2Phe(Ala)_2$.

	RMSD (Å)	
	FMO2	FMO3
α-$(Ala)_2Phe(Ala)_2$		
Singlet	0.177	0.026
Triplet	0.210	0.254
β-$(Ala)_2Phe(Ala)_2$		
Singlet	0.484	0.019
Triplet	0.563	0.176

www.manaraa.com

Table 12. Adiabatic FMO2 excitation energy of the small protein chignolin (1UAO) compared to *ab initio* energy values calculated at the FMO2 optimized structure.

	6-31G(d)			
	FMO2-HF	Error (kcal/mol)	FMO2-MP2	Error (kcal/mol)
Singlet	-3799.987012704	-0.71	-3811.124860473	-0.07
Triplet	-3799.878011966	-0.70	-3810.978345282	-0.60
Excitation	68.40	0.02	91.94	-0.53

www.manaraa.com

Chapter 6. Extension of the Effective Fragment Molecular Orbital Method

A paper to be submitted to *The Journal of Chemical Theory and Computation*

Spencer R. Pruitt, Casper Steinmann, Jan H. Jensen, Mark S. Gordon

Abstract

In this work, the effective fragment molecular orbital (EFMO) method is extended to include all relevant intermolecular interactions currently accounted for in the effective fragment potential (EFP) method. The accuracy of the method is tested and compared to both the standard fragment molecular orbital (FMO) method as well as fully *ab initio* methods. It is shown that the extended EFMO method provides significant reductions in error while at the same time reducing the computational cost associated with standard FMO calculations by up to 93%.

1. Introduction

Modern computational chemistry methods strive to accurately model chemical systems using efficient computational algorithms. Unfortunately, it is difficult to reconcile both of these goals, since most methods that are widely viewed as the most accurate[1] also require the most computational effort. The most effective compromise reached so far has been through the application of fragmentation approaches to these computationally intensive methods. Many such fragmentation methods have been introduced in recent years[2-8], with many showing the ability to accurately model large molecular systems. Methods such as the systematic molecular fragmentation (SMF) method[9-11], molecular fractionation with conjugate caps (MFCC)[12], the molecular tailoring approach (MTA)[13] and the explicit polarization potential (X-Pol)[14,15] have all shown a variety of successes in describing different chemical systems.

One such method, the fragment molecular orbital (FMO) method[16], has been extensively developed[17] since the original implementation by Kitaura et al. Based upon a many-body expansion of the energy, the FMO method takes the effects of the entire system into account during each step of a given calculation through the use of an electrostatic potential (ESP). The FMO method, as well as other fragmentation methods[18], also benefit from the relative ease with which calculations can be parallelized on modern computer hardware. This inherent parallelizability aids in lowering the computational cost of the most accurate *ab initio* methods.

While not a fragmentation method in the same vein as the FMO method, the effective fragment potential (EFP) method[19-21] was originally developed to accurately introduce solvent effects into chemical processes without the use of any fitted parameters. The importance of modeling chemistry in solution is apparent in many applications[22-32], making the EFP method an attractive solution to the study of solvent effects and intermolecular interactions. Recently, the generalized EFP method[33,34] (EFP2) has been developed as an *ab initio* based method for capturing all intermolecular interactions including, but not limited to, solvent effects. As

www.manaraa.com

implemented in the General Atomic and Molecular Electronic Structure System (GAMESS) program package,[35,36] the EFP2 method requires a preliminary calculation to generate the potential for each fragment. The generated potentials are then incorporated into a production calculation on the chemical system of interest. Since the EFP method is essentially a classical model, it can only provide intermolecular interactions. Bond breaking/bond making processes must be described by quantum mechanics. EFP fragments have internally frozen geometries.

In an effort to combine the accuracy of the FMO method with the speed of the EFP method, Steinmann et al. recently developed a combined method called the effective fragment molecular orbital (EFMO) method[37,38]. The EFMO method uses the fragmentation scheme from the FMO method and treats the separated fragment interactions using the Coulomb interaction of the EFP method. Additionally, the ESP used during standard FMO calculations is replaced with the many-body polarization of the EFP method. Other intermolecular interactions such as exchange repulsion, charge transfer and dispersion that are present in the EFP method are not included in the original EFMO method. The EFMO method also incorporates the intramolecular energy of each fragment, and it removes the restriction of frozen internal EFP geometries.

Although the EFMO method has many benefits, the inclusion of only the EFP polarization and Coulomb interactions essentially limits the applicability of the method to Hartree Fock (HF) or density functional theory (DFT) calculations. The present work extends the EFMO method to include all relevant intermolecular interactions (Coulomb, polarization, exchange repulsion, charge transfer and dispersion) to enable the use of correlated *ab initio* methods. By including all fundamental types of intermolecular interactions, this extension of the EFMO method can also reduce the total number of explicit quantum mechanics (QM) calculations required, providing a significant reduction in computational cost. An additional benefit of this extension is the detailed analysis of the individual contributions to the total interaction energy, or energy decomposition analysis[39,40] (EDA), between two fragments provided by the EFP method. A dimer EDA can provide insights into the most important intermolecular interactions in a given chemical system.

The first section of this paper introduces the theoretical background of the FMO, EFP and extended EFMO methods. Next, the newly extended EFMO method is tested on water clusters of varying size, comparing the energies to those from the FMO method and to fully *ab initio* energies. Timings are then presented that compare the extended EFMO, FMO and *ab initio* methods, followed by conclusions and future directions.

2. Methodology

2.1 The Fragment Molecular Orbital Method

The FMO method is based on a many-body expansion of the total energy, truncated at three-body interactions:

www.manaraa.com

$$E = \sum_I E_I + \sum_{I>J} (E_{IJ} - E_I - E_J) + \sum \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{IK} - E_I - E_K) - (E_{JK} - E_J - E_K)\} \quad (1)$$

Each energy in Eq. (1) is calculated using quantum mechanical (QM) methods. Eq. (1) includes energies for individual fragments *I* (monomers), fragment pairs *IJ* (dimers) and fragment triples *IJK* (trimers). The first two terms in Eq. (1) omit explicit trimer calculations and comprise the FMO2 method. The FMO3[41] method includes the third (trimer) term. The FMO2 and FMO3 energy expressions can be written as:

$$E^{FMO2} = \sum_I E_I + \sum_{I>J} (E_{IJ} - E_I - E_J) \quad (2)$$

$$E^{FMO3} = E^{FMO2} + \sum_{I>J>K} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{IK} - E_I - E_K) - (E_{JK} - E_J - E_K)\} \quad (3)$$

Each energy in Eqs. (2) and (3) can be calculated in one of two ways: as isolated *n*-mers with energy E_x^0, or in the presence of the rest of the fragments with energy E_x (*x* = *I*, *J*, *K*, *IJ*, *IK*, *JK*, *IJK*). The standard FMO method calculates the energies E_x by incorporating an electrostatic potential (ESP) derived from the densities of all other fragments[16]. The ESP has the form

$$V_{\mu\nu}^x = \sum_{K \neq x} (\mu_{\mu\nu}^K + \nu_{\mu\nu}^K) \quad (4)$$

$$\mu_{\mu\nu}^K = \sum_{A \in K} \langle \mu | (-Z_A / |\mathbf{r} - \mathbf{r}_A|) | \nu \rangle \quad (5)$$

$$\nu_{\mu\nu}^K = \sum_{\lambda\sigma \in K} D_{\lambda\sigma}^K (\mu\nu | \lambda\sigma) \quad (6)$$

where $\mu_{\mu\nu}^K$ represents the nuclear attraction contribution to the energy and $\nu_{\mu\nu}^K$ represents the two-electron contribution. Both of these terms are expressed in terms of one- (Eq. (5)) and two- (Eq. (6)) electron integrals over AOs μ and ν and are calculated for each of the surrounding monomers *K* with electron density D^K.

While the formalism up to this point provides a means to obtain accurate energies, efficient algorithms must be used in order to take advantage of modern computer hardware. The FMO implementation in GAMESS makes use of the generalized distributed data interface[42] (GDDI) to accomplish this goal. One performs each *n*-mer calculation on a separate computer node (coarse grained parallel); the availability of multiple cores within

www.manaraa.com

each node facilitates the fine-grained parallel calculation for each fragment. The two-level parallelism described here allows the FMO method to take advantage of massively parallel computers.

Despite the computational savings that can be achieved through the use of the GDDI, the rapid increase in the number of dimer and trimer calculations with system size for large molecular systems still requires an increasingly significant computational effort. The number of *n*-mer calculations performed can be partially reduced by employing approximations for the interaction energies of *n*-mers that are farther apart than a unitless, predefined cut-off value R_{cut}. The unitless distance between two fragments, $R_{I,J}$, is given by

$$R_{I,J} = \min_{i \in I, j \in J} \left\{ \frac{\left| \vec{r}_i - \vec{r}_j \right|}{r_i^{vdw} + r_j^{vdw}} \right\} \tag{7}$$

where $R_{I,J}$ is the relative minimum interatomic distance between fragment *I* and fragment *J* based on the van der Waals radii r^{vdw} of the atoms *i* and *j*. Through a reformulation of Eq. (2), E^{FMO2} can be expressed as

$$E^{FMO2} = \sum_I E_I + \sum_{I>J}^{R_{I,J} \le R_{cut}} \Delta E_{IJ} + \sum_{I>J}^{R_{I,J} > R_{cut}} \Delta E_{IJ}^{sep} \tag{8}$$

where

$$\Delta E_{IJ} = (E_{IJ} - E_I - E_J) \tag{9}$$

with a corresponding expression[41] for E^{FMO3}. The total number of QM dimers (term 2) in Eq. (8) can now be varied based on the user-defined value of R_{cut} (default value = 2.0). The remaining dimers (last term in Eq. (8)) are now approximated by

$$\Delta E_{IJ}^{sep} \cong E_I + E_J + \mathrm{Tr}(D^I u^{1,I(J)}) + \mathrm{Tr}(D^J u^{1,J(I)}) + \sum_{\mu\nu \in I} \sum_{\lambda\sigma \in J} D_{\mu\nu} D_{\lambda\sigma} (\mu\nu \mid \lambda\sigma) + \Delta E_{IJ}^{NR} \tag{10}$$

where $u^{1,I(J)}$ and $u^{1,J(I)}$ are the one-electron Coulomb potentials of the force exerted by fragment *I* (*J*) on fragment *J* (*I*). The electron-electron interaction (5th term in Eq. (10)) and the nuclear repulsion energy, ΔE_{IJ}^{NR}, are also included. The standard SCF procedure is not employed during separated dimer calculations, resulting in significant time savings compared to QM dimer calculations.

2.2 The Effective Fragment Potential Method

An extension of the original EFP1 method[19-21] to the generalized EFP2 method[33-34] facilitates the inclusion of solvent effects and the evaluation of intermolecular interactions for any molecular species. The EFP2 intermolecular energy is composed of five terms:

www.manaraa.com

$$E^{EFP2} = E^{Coul} + E^{pol} + E^{disp} + E^{exrep} + E^{ct} \quad (11)$$

The first two interactions, E^{Coul} and E^{pol}, are identical to the same terms in EFP1. E^{Coul} is based on a distributed multipole analysis at atom centers and bond midpoints[43,44], truncated at the octopole term. The polarization energy, E^{pol}, arises from the interaction of induced and permanent dipoles between fragments. The polarization energy is iterated to self-consistency, enabling the EFP method to capture some of the many-body effects present in chemical systems. The third term in Eq. (11), E^{disp}, is expressed using an inverse R expansion[45]:

$$E^{disp} = \sum_n C_n R^{-n} \quad (12)$$

In the EFP method, this expansion is truncated at the leading induced dipole-induced dipole R^{-6} term. The contribution of the R^{-8} term is estimated as one-third of the R^{-6} term[45]. The coefficients C_n are derived from imaginary frequency-dependent polarizabilities integrated over the entire frequency range. In particular, the C_6 coefficients are derived in terms of the interactions between pairs of localized molecular orbitals (LMOs), one on each molecular species.

The last two terms in Eq. (11), E^{exrep} and E^{ct}, are based on approximate energy expressions that depend on the intermolecular overlap of molecular orbitals. Since the EFP method uses frozen LMOs, the overlap expansion used for E^{exrep} can be reliably truncated at the quadratic term[46-47]. The calculation of E^{exrep} requires each generated potential to carry a basis set, making the EFP method basis set dependent. The charge transfer energy[48-49], E^{ct}, between two fragments is calculated by considering the interaction between the occupied orbitals of one fragment and the virtual orbitals of a second fragment. The approximate formula used to calculate E^{ct} is based on a second-order perturbative treatment of the intermolecular interactions. This formula is expressed in terms of canonical HF orbitals for a pair of fragments, using a truncated multipolar expansion (through quadrupoles) to represent the molecular electrostatic potential. The addition of E^{ct} to the total energy results in a significant lowering of the energy for ionic and highly polar species.

The first three terms in Eq. (11) can fail at short inter-molecular distances. For example, the Coulomb interaction becomes repulsive at short distances and the polarization interaction becomes too attractive. The failures that occur at small distances can be mitigated through the addition of damping functions.[50] For the Coulomb term, an exponential damping function is used[51,52]:

$$f_{damp} = 1 - \exp(-\alpha R) \quad (13)$$

The parameter α is obtained through the fitting of the damped multipole potential to the Hartree-Fock potential. An alternate approach to damping the short-range Coulomb interactions is through an approximation of the short-range charge penetration energy based on the intermolecular overlap[52,53]. The polarization term can be damped using either an exponential term as in Eq. (13) or Gaussian damping. The dispersion term can be

www.manaraa.com

damped using the Tang-Toennies[45,54] formula, or through the use of a damping formula based on the intermolecular overlap[52]. The latter is preferable since the intermolecular overlap is already calculated for the exchange repulsion and since no arbitrarily fitted parameter is required.

2.3 The Effective Fragment Molecular Orbital Method

The EFMO method was developed to integrate the FMO and EFP methods in an effort to provide a generally applicable, accurate and efficient approach to large molecular systems. The original EFMO energy[37] is given by

$$E^{EFMO} = \sum_{I} E_I^0 + \sum_{I>J}^{R_{I,J} \le R_{cut}} (\Delta E_{IJ}^0 - E_{IJ}^{pol}) + \sum_{I>J}^{R_{I,J} > R_{cut}} E_{IJ}^{Coul} + E_{tot}^{pol} \qquad (14)$$

where the standard fragment energies in the FMO method (which include the ESP of each other fragment) are replaced with the isolated energies described previously. The use of isolated fragment energies eliminates the need to calculate the ESP used in standard FMO method calculations. The many-body interaction energy formerly computed using the ESP is replaced by the total EFP polarization energy E_{tot}^{pol}. Each fragment pair polarization energy, E_{IJ}^{pol}, is subtracted from the corresponding dimer energy E_{IJ}, since each E_{IJ}^{pol} is already contained in E_{tot}^{pol}. The separated dimer energies (ΔE_{IJ}^{sep}) in Eq. (8) are replaced by the EFP Coulomb energies E_{IJ}^{Coul}.

The original formulation of the total EFMO energy is extended in this work to include all five components of the EFP energy as described in Section 2.2. The expression for the extended EFMO energy is:

$$E^{EFMO} = \sum_{I} E_I^0 + \sum_{I>J}^{R_{I,J} \le R_{cut}} (\Delta E_{IJ}^0 - E_{IJ}^{pol}) + \sum_{I>J}^{R_{I,J} > R_{cut}} (E_{IJ}^{Coul} + E_{IJ}^{disp} + E_{IJ}^{exrep} + E_{IJ}^{ct}) + E_{tot}^{pol} \qquad (15)$$

Including the dispersion energy in the separated dimer energies allows the EFMO method to be used with correlated *ab initio* methods such as second order Møller-Plesset perturbation theory (MP2) and coupled cluster theory (CC). Additionally, by including all intermolecular interactions, the user defined cut-off value R_{cut} can be reduced to neglect additional QM dimers. The reduction in QM dimers lowers the computational requirements of EFMO calculations relative to standard FMO method calculations.

The extended EFMO method has been implemented in a modified version of the GAMESS program package. All of the default screening parameters for the EFP method were used with the exception of the Coulomb energy, for which the overlap based damping[52,53] is used.

www.manaraa.com

3. Results

Clusters of water molecules, shown in Figures 1 through 4, were used as test systems to assess the accuracy of the extended EFMO method versus both FMO2 and fully *ab initio* methods. Seven minimum energy structures for clusters of both 8 and 16 water molecules were chosen from previous work[55]. Larger clusters of 32 and 64 water molecules were obtained using a Monte Carlo algorithm with simulated annealing (MC/SA)[55,56]. The starting temperature for the SA was set to 20,000 K and then lowered to 300 K. Every 100 steps a local minimization was performed, with the number of molecules moved per step varied from one to four. Approximately 16,000 geometries were sampled for both 32 and 64 water molecules. Seven unique isomers were chosen and re-optimized using the EFP1 method[20]. Fully *ab initio* energies for each of the 8, 16, 32 and 64 water clusters were calculated at the MP2 level of theory using two Pople-type basis sets[57] (6-31++G(d,p) and 6-311++G(3df,2p)). Due to computational limitations, the MP2 energies of the 64 water clusters were evaluated using only the 6-31++G(d,p) basis set.

Table 1 shows the various R_{cut} values tested and the average number of separated and QM dimer calculations performed for each cut-off value and cluster size. Recall that the smaller the value of R_{cut}, the fewer is the number of dimers that are calculated using a fully QM level of theory. For example, if R_{cut} = 0.6, no dimers are calculated with the QM level of theory for all of the water cluster sizes considered, so there are only separated dimer interactions for both the extended EFMO and FMO2 methods. The largest value of R_{cut} in Table 1 (2.0, the default value in GAMESS) results in the QM calculation of many more dimers, as many as 598 for the 64-water clusters.

3.1 Average signed errors

The total cluster energies produced by the extended EFMO method are compared with FMO2 and MP2 energies in Table 2. The average signed errors in Table 2 are calculated as

$$Error = \frac{\sum_{i=1}^{n} (E_i^X - E_i^{MP2})}{n} \qquad (15)$$

where n is the number of isomers and E_i^X is either the EFMO or FMO2 energy.

For the 6-31++G(d,p) basis set and R_{cut} = 0.6, the extended EFMO method produces consistently smaller errors than does FMO2. The EFMO errors average 18.6 kcal/mol for clusters of 8 water molecules and 262.2 kcal/mol for clusters of 64 water molecules. Average FMO2 errors are significantly larger, with an average error of -104.1 kcal/mol for clusters of 8 water molecules and -1348.2 kcal/mol for clusters of 64 water molecules. Increasing R_{cut} to 0.8 provides a sharp decrease in the average FMO2 errors for all cluster sizes, with errors between -7.2 and -72.2 kcal/mol. The corresponding errors for the extended EFMO method

www.manaraa.com

decrease as well, falling in between -5.7 and -44.6 kcal/mol for all cluster sizes. The inclusion of additional QM dimers by increasing R_{cut} to 1.4 and then 2.0 shows a steady increase in average error for both the extended EFMO and FMO2 methods. Despite this increase in average error, the extended EFMO method produces smaller errors than the FMO2 method for all cluster sizes and values of R_{cut}.

Errors obtained using the larger 6-311++G(3df,2p) basis set and R_{cut} = 0.6 follow the same trend of increasing with cluster size reported above for the smaller basis set. In general, average FMO2 errors increase for all cluster sizes compared to those obtained using the smaller basis set. However, the average EFMO errors drop substantially for all clusters sizes when the larger basis set is used. For example, the average EFMO error for the 8-water clusters with R_{cut} = 0.6 is 2.3 kcal/mol for the larger basis set vs. 8.6 kcal/mol for the smaller basis set. The analogous comparison for the 32-water clusters is 16.2 kcal/mol vs. 122.3 kcal/mol. This significant reduction in the extend EFMO method errors is due to the use of a better EFP potential generated with the larger basis set. Increasing the number of QM dimers causes an increase in error for both FMO2 and EFMO. However, the errors produced by the extended EFMO method converge more quickly for each cluster size. For example, with R_{cut} = 0.8, 1.4 and 2.0 the extended EFMO method produces errors of -8.5, -9.4 and -9.6 kcal/mol for the smallest clusters of 8 water molecules. The corresponding FMO2 errors are -8.5, -20.8 and -22.9 kcal/mol. This behavior is consistent for all cluster sizes, with the extended EFMO method errors differing by less than 8.6 kcal/mol when R_{cut} is increased from 0.8 to 2.0.

3.2 Relative Energies

Seven different isomers were chosen for each cluster size to test the ability of both the extended EFMO and FMO2 methods to reproduce relative energies compared to MP2 results. The same set of R_{cut} values (0.6, 0.8, 1.4 and 2.0) that were discussed in Section 3.1 are used.

Figure 5 shows the relative energies for isomers of eight water molecules for EFMO, FMO2 and MP2. Graphs a through d present results for R_{cut} values of 0.6, 0.8, 1.4 and 2.0, respectively, for the 6-31++G(d,p) basis set. Graphs e through f present the corresponding data for the 6-311++G(3df,2p) basis set. For clusters of 8 water molecules, the EFMO method provides more accurate relative energies than does the FMO2 method. As noted above for the average signed errors, the best EFMO results are produced using R_{cut} = 0.6. The quality of the EFMO relative energies does not change appreciably with an increase in basis set size.

Relative energies for the 16-water isomers are shown in Figure 6, following the same scheme as in Figure 5. Results for both FMO2 and EFMO using the smaller basis set are not as accurate as the results for the 8-water isomers. The most accurate results are now produced using R_{cut} = 1.4 for EFMO, while the most accurate FMO2 results are produced when using R_{cut} = 2.0. The accuracy for both EFMO and FMO2 when using the larger basis set are relatively poor. The inability of either method to accurately produce the relative energies of the 16 water molecule clusters may be due to the stacking of dipole moments in these highly

www.manaraa.com

structured clusters. The stacked dipoles can create much larger dipole moments than either the smaller, similarly structured 8-water clusters or the more globular arrangements of the larger 32 and 64 water clusters.

When the size of the clusters is increased to 32 water molecules, there are no highly structured isomers; globular droplets are more prevalent. Figure 7 shows that for both the smaller and larger basis sets, the extended EFMO method consistently produces extremely accurate relative energies, in some cases (Figures 7b-e) nearly overlaying the MP2 curve exactly. The relative energies produced by the FMO2 method are not dissimilar to the MP2 results, however the two lowest energy structures are found to be farther apart in energy than the relative energies obtained by the extended EFMO method or MP2. Consequently, the FMO2 relative energy curve is shifted up, in all cases predicting the highest energy isomer to be greater than 10 kcal/mol higher than predicted by either EFMO or MP2; the largest difference in Figure 7e is ~45 kcal/mol.

The relative energies of the 64-water clusters using only the smaller 6-31++G(d,p) basis set are shown in Figure 8. For R_{cut} = 0.6, the extended EFMO method does not reproduce the relative energies correctly. However, the EFMO relative energies are an improvement over the FMO2 relative energies. Adding nearest neighbor QM dimers improves both the EFMO and FMO2 results, however inclusion of additional QM dimers (R_{cut} = 1.4) provides the best EFMO results. For all R_{cut} values used, FMO2 fails to give even reasonably accurate relative energies.

3.3 Average Binding Energy per Water Molecule

As a final test of the extended EFMO method, the average binding energy per molecule was calculated for all clusters and basis sets for both the extended EFMO method and the FMO2 method and compared to the MP2 results in Table 3. Each water molecule has the same internal geometry. The total binding energy of each cluster can be easily evaluated by subtracting the gas phase monomer energy, multiplied by the appropriate number of monomers, from the total energy.

The extended EFMO method outperforms the FMO2 method for all clusters and basis sets investigated. The FMO2 error compared to MP2 for the smallest basis set are in the range 4-20 kcal/mol, while the corresponding EFMO binding energies are in error by only 1-3 kcal/mol. For the larger basis set, the extended EFMO method produces even more accurate binding energies, with errors between 0.5 and 2 kcal/mol. The FMO2 errors do not change appreciably: they are still in the range of 2 - 20 kcal/mol. The most accurate FMO2 binding energies are in error by 1.9-2.1 kcal/mol compared to MP2 for R_{cut} = 0.8. In comparison, the most accurate EFMO binding energies are in error by only 0.3 kcal/mol for R_{cut} = 0.6; the largest EFMO error of 2.5 kcal/mol is comparable in magnitude to the most accurate FMO2 binding energy. It is especially noteworthy that the largest difference in predicted binding energies between FMO2 and EFMO occurs consistently for R_{cut} = 0.6. This means that many fewer fully MP2 calculations are required if one uses the extended EFMO method that is presented here.

www.manaraa.com

3.4 Timings

Timing comparisons are presented in Table 4 for the extended EFMO, FMO2 and MP2 methods, using the 6-311++G(3df,2p) basis set on a single 2.66 GHz Intel Xeon processing core for one isomer each of 8 and 32 water molecules. Due to computational requirements, MP2 timings are only shown for the cluster of 8-water molecules. Timings are compared for R_{cut} values that produced the best average errors for both the extended EFMO method (R_{cut} = 0.6) and FMO2 method (R_{cut} = 0.8).

For the smallest water cluster and R_{cut} = 0.6, the FMO2 method produces a wall clock time of ~353 seconds with an error of -104.5 kcal/mol, compared to an MP2 wall clock time of ~4348 seconds. The corresponding extended EFMO method calculation takes only ~61 seconds, but with a much smaller error of 2.9 kcal/mol. With R_{cut} = 0.8 for the 8-water cluster, the errors for FMO2 and EFMO are similar at -7.6 and -7.9 kcal/mol respectively, however the extended EFMO method is approximately 170 seconds faster than the FMO2 method. The difference in wall clock times between the extended EFMO and FMO2 methods increases for the cluster of 32 water molecules. Using R_{cut} = 0.6, the EFMO energy calculation takes ~147 seconds compared to ~2390 seconds for the FMO2 calculation, roughly 16 times faster. Increasing R_{cut} to 0.8 reduces the EFMO speed-up to approximately 3 times faster compared to the FMO2 calculation, with wall clock times of 2203 and 6969 seconds respectively. For both cluster sizes, the extended EFMO method provides reductions in wall clock times between 25 and 93% compared to the FMO2 method while producing superior results.

4. Conclusions

Through this work, the EFMO method has been extended to include all five intermolecular interactions present in the EFP method. The inclusion of short-range intermolecular interactions allows the extended EFMO method to be used with correlated *ab initio* methods such as MP2 and CC. The accuracy of the extended EFMO method was tested versus the standard FMO2 method and MP2. The extended EFMO method was shown to provide superior average errors, relative energies and binding energies for all cluster sizes and basis sets compared to the FMO2 method. Significant reductions in wall clock times compared to the FMO2 method were also shown. Through the reduction of the number of explicit QM dimers performed during extended EFMO method calculations, time savings of up to 93% compared to the FMO2 method were achieved while at the same time providing a more accurate estimate of the MP2 energies.

Future work on the extended EFMO method will include the implementation of gradients for the newly added energy terms in order to enable geometry optimizations. With the addition of energy gradients, as well as the improved accuracy and reduction in computational requirements versus the FMO2 method, the extended EFMO method could provide a computationally feasible algorithm for dynamical simulations.

www.manaraa.com

References

1. T. J. Lee and G. E. Scuseria, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, edited by S. R. Langhoff (Kluwer, Dordrecht, The Netherlands, 1995), pp. 47–108.

2. Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.

3. Pomogaev, V.; Pomogaeva, A.; Aoki, Y. *J. Phys. Chem. A* **2009**, *113*, 1429.

4. Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comp.* **2009**, *5*, 459.

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

6. Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comp.* **2009**, *5*, 1667.

7. 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.

8. Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927.

9. V. Deev and M. A. Collins, *J. Chem. Phys.*, 2005, **122**, 154102.

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

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

12. Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039-12041.

13. Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484-495.

14. Gao, J. *J. Phys. Chem. B* **1997,** *101*, 657–663.

15. Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346–2354.

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

17. Fedorov, D. G.; Kitaura, K.; Eds., The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems, CRC Press, Boca Raton, FL, 2009.

18. Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.

19. Jensen, J.H.; Day, P.N.; Gordon, M.S.; Basch, H.; Cohen, D.; Garmer, D.R.; Kraus, M.; Stevens, W.J. *Modeling the Hydrogen Bond*, **1994**, *569*, 139–151.

20. Day, P.N.; Jensen, J.H.; Gordon, M.S.; Webb, S.P.; Stevens, W.J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968-1986.

21. Gordon, M.S.; Freitag, M.A.; Bandyopadhyay, P.; Jensen, J.H.; Kairys, V.; Stevens, W.J. *J. Phys. Chem. A* **2001**, *105*, 293-307.

22. Petersen, C.P.; Gordon, M.S. *J. Phys. Chem. A*, **1999**, *103*, 4162-4166.

23. Balawender, R.; Safi, B.; Geerlings, P. *J. Phys. Chem. A*, **2001**, *105*, 6703-6710.

24. Merrill, G.N.; Webb, S.P. *J. Phys. Chem. A*, **2003**, *107*, 7852-7860.

25. Merrill, G.N.; Webb, S.P.; Bivin, D.B.. *J. Phys. Chem. A*, **2003**, *107*, 386-396.

www.manaraa.com

26. Yoshikawa, A.; Morales, J.A. *Journal of Molecular Structure-Theochem*, **2004**, *681*, 27-40.

27. Merrill, G.N.; Webb, S.P. *J. Phys. Chem. A*, **2004**, *108*, 833-839.

28. Kemp, D.D.; Gordon, M.S. *J. Phys. Chem. A*, **2005**, *109*, 7688-7699.

29. Chandrakumar, K.R.S.; Ghanty, T.K.; Ghosh, S.K.; Mukherjee, T.. *J. Mol. Str. (Theochem)*, **2007**, *807*, 93-99.

30. Kemp, D.A.; Gordon, M.S. *J. Phys. Chem. A*, **2008**, *112*, 4885-4894.

31. Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C.F.; Herbert, J.M.; Gordon, M.S.; Schmidt, M.W.; Slipchenko, L.V.; Krylov, A.I. *J. Phys. Chem. A* **2010**, *114*, 12739-12754.

32. Slipchenko, L.V.; Gordon, M.S. *J. Comput. Chem.* **2007**, *28*, 276-291.

33. Gordon, M.S.; Slipchenko, L.V.; Li, H.; Jensen, J.H.. *Annual Reports in Computational Chemistry*, **2007**, *3*, 177-193.

34. Gordon, M.S.; Mullin, J.M.; Pruitt, S.R.; Roskop, L.B.; Slipchenko, L.V.; Boatz, J.A. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646-9663.

35. M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis and J.J.A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.

36. M.S. Gordon and M.W. Schmidt, in *Theory and Applications of Computational Chemistry, the First Forty Years*, eds. C. E. Dykstra, G. Frenking, K. S. Kim and G. Scuseria, Elsevier, Amsterdam, 2005.

37. Steinmann, C.; Fedorov, D. G.; Jensen, J. H. *J. Phys. Chem. A* **2010**, *114*, 8705-8712.

38. Steinmann, C; Fedorov, D.G.; Jensen, J.H. *PLoS One*, submitted

39. Kitaura, K.; Morokuma, K. *Int. J. Quant. Chem.* **1976**, ***10***, 325.

40. Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *100*, 14316.

41. Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, *433*, 182-187.

42. Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.* **2004**, *25*, 872.

43. Stone, A.J. *Chem. Phys. Lett.* **1981**, *83*, 233-239.

44. Stone, A.J., *The Theory of Intermolecular Forces*. 1996: Oxford University Press: Oxford.

45. Adamovic, I.; Gordon, M.S. *Mol. Phys.* **2005**, *103*, 379-387.

46. Jensen, J.H.; Gordon, M.S. *Mol. Phys.* **1996**, *89*, 1313-1325.

47. Jensen, J.H.; Gordon, M.S. *J. Chem. Phys.* **1998**, *108*, 4772-4782.

48. Jensen, J.H. *J. Chem. Phys.* **2001**, *114*, 8775-8783.

49. Li, H.; Gordon, M.S.; Jensen, J.H. *J. Chem. Phys.* **2006**, *124*, 214108.

50. Freitag, M.A.; Gordon, M.S.; Jensen, J.H.; Stevens, W.J. *J. Chem. Phys.* **2000**, *112*, 7300-7306.

51. Slipchenko, L.V.; Gordon, M.S. *J. Comput. Chem.* **2007**, *28*, 276-291.

www.manaraa.com

52. Slipchenko, L.V.; Gordon, M.S. *Mol. Phys.* **2009**, *107*, 999-1016.

53. Kairys, V.; Jensen, J.H. *Chem. Phys. Lett.* **1999**, *315*, 140-144.

54. Tang, K.T.; Toennies, J.P. *J. Chem. Phys.* **1984**, *80*, 3726-3741.

55. P. N. Day, R. Pacher, M. S. Gordon and G. N. Merrill, *J. Chem. Phys.*, 2000, **112**, 2063.

56. N. Metropolis, A. Reosenbluth and A. Teller, *J. Chem. Phys.*, 1953, **21**, 1087.

57. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213.

www.manaraa.com

Table 1. Average total number separated and QM dimers for all cluster sizes and values of R_{cut}.[a]

	8 waters		16 waters		32 waters		64 waters	
R_{cut}	separated	QM	separated	QM	separated	QM	separated	QM
0.6	28	0	120	0	496	0	2016	0
0.8	16	12	93	27	444	52	1904	112
1.4	5	23	65	55	363	133	1730	286
2.0	0	28	30	90	237	259	1418	598

[a] "separated" indicates the number of dimers calculated with approximations. QM indicates the number of dimers calculated using quantum mechanics.

www.manaraa.com

Table 2. Average signed errors (kcal/mol) relative to *ab initio* energies for the EFMO and FMO2 methods for each cluster size and R_{cut} value.

	6-31++G(d,p)		6-311++G(3df,2p)	
R_{cut}	FMO2	EFMO	FMO2	EFMO
	8 water molecules			
0.6	-104.1	18.6	-104.4	2.3
0.8	-7.2	-5.7	-8.5	-8.5
1.4	-12.5	-8.4	-20.8	-9.4
2.0	-14.1	-8.8	-22.9	-9.6
	16 water molecules			
0.6	-310.7	48.1	-312.7	7.9
0.8	-23.2	-17.2	-30.1	-27.3
1.4	-38.9	-24.4	-70.7	-30.4
2.0	-48.2	-27.1	-85.5	-31.3
	32 water molecules			
0.6	-642.2	122.3	-654.2	16.2
0.8	-40.3	-25.6	-61.0	-57.0
1.4	-79.9	-45.8	-153.3	-64.2
2.0	-111.9	-53.9	-204.5	-65.6
	64 water molecules			
0.6	-1348.2	262.2	-	-
0.8	-72.2	-44.6	-	-
1.4	-155.7	-88.6	-	-
2.0	-238.0	-109.8	-	-

www.manaraa.com

Table 3. Binding energy per water molecule for FMO2, EFMO and MP2 (kcal/mol).

	6-31++G(d,p)			6-311++G(3df,2p)		
Rcut	FMO2	EFMO	MP2	FMO2	EFMO	MP2
	8 water molecules					
0.6	-23.7	-8.4	-10.7	-22.3	-9.0	-9.3
0.8	-11.6	-11.4	-10.7	-10.3	-10.3	-9.3
1.4	-12.3	-11.7	-10.7	-11.9	-10.4	-9.3
2.0	-12.4	-11.8	-10.7	-12.1	-10.5	-9.3
	16 water molecules					
0.6	-29.8	-7.4	-10.4	-28.6	-8.5	-9.0
0.8	-11.8	-11.5	-10.4	-10.9	-10.7	-9.0
1.4	-12.8	-11.9	-10.4	-13.4	-10.9	-9.0
2.0	-13.4	-12.1	-10.4	-14.4	-11.0	-9.0
	32 water molecules					
0.6	-30.9	-7.0	-10.8	-29.7	-8.8	-9.3
0.8	-12.1	-11.6	-10.8	-11.2	-11.1	-9.3
1.4	-13.3	-12.2	-10.8	-14.1	-11.3	-9.3
2.0	-14.3	-12.5	-10.8	-15.7	-11.3	-9.3
	64 water molecules					
0.6	-32.5	-7.4	-11.5	-	-	-
0.8	-12.6	-12.2	-11.5	-	-	-
1.4	-13.9	-12.8	-11.5	-	-	-
2.0	-15.2	-13.2	-11.5	-	-	-

www.manaraa.com

Table 4. Timing comparison[a,b] for EFMO, FMO2 and fully MP2 energy calculations.

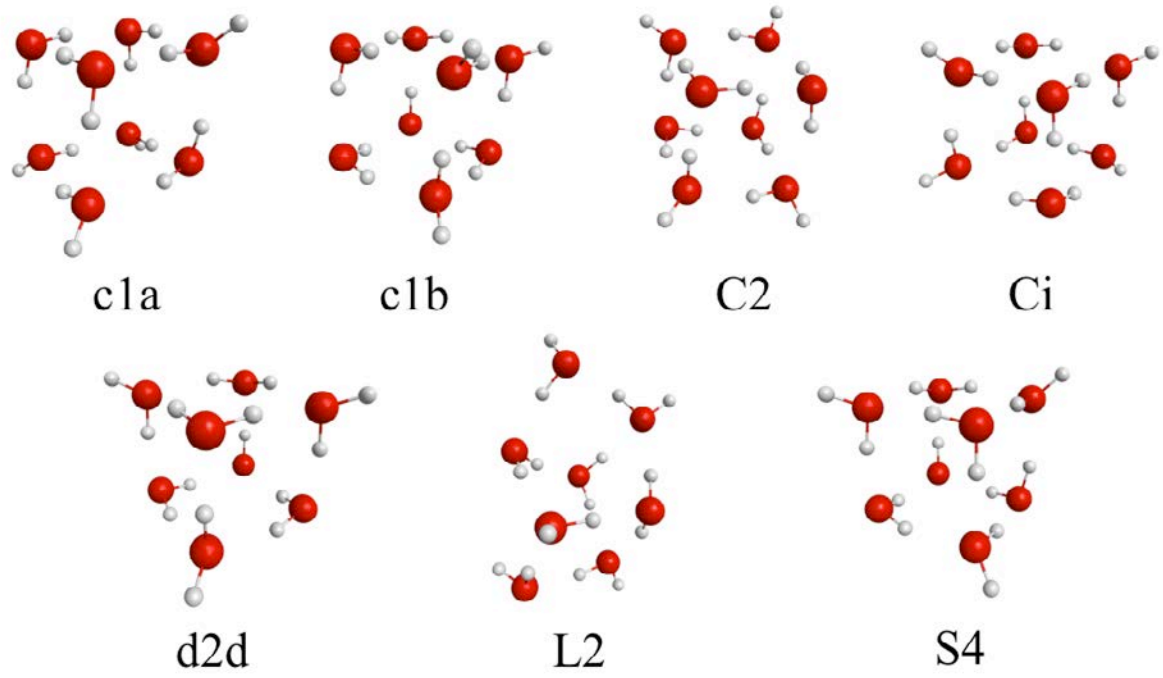
	6-311++(3df,2p)				
	EFMO		FMO2		MP2
Rcut	Wall Time	Error	Wall Time	Error	Wall Time
	8 waters				
0.6	61	2.9	353	-104.5	4348
0.8	515	-7.9	686	-7.6	-
	32 waters				
0.6	147	16.7	2390	-668.5	-
0.8	2203	-57.4	6969	-64.1	-

[a] All timings performed on a single CPU core.

[b] Times in are in seconds and errors are in kcal/mol.

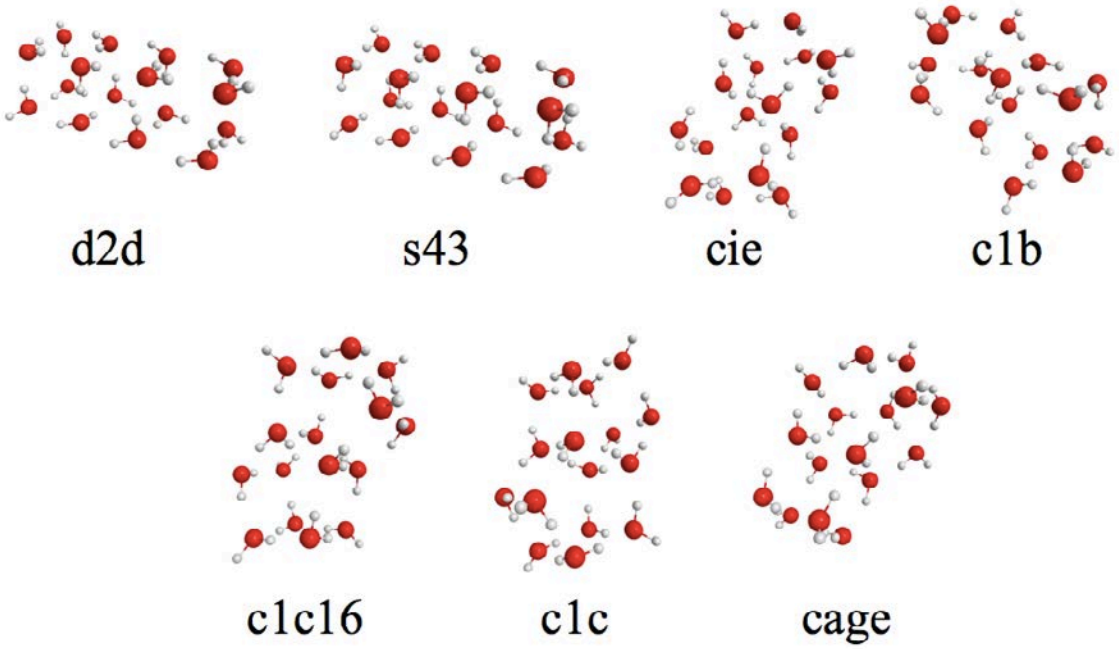
www.manaraa.com

Figure 1. Structures of all 8-water clusters.



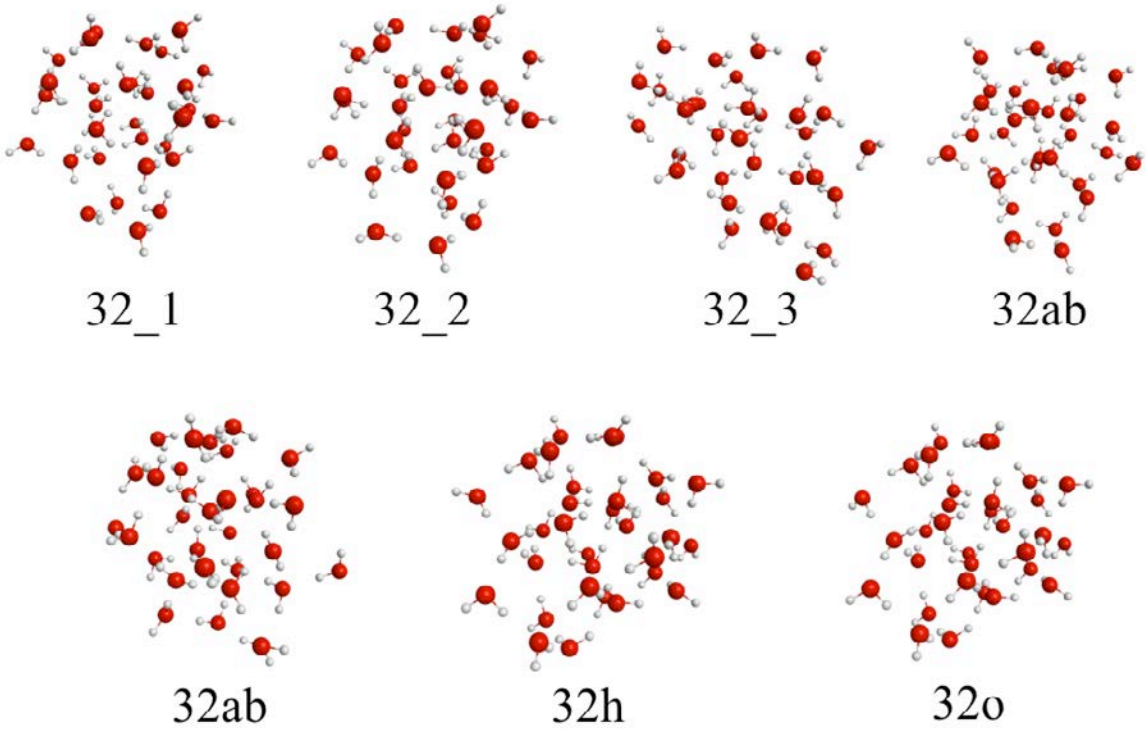
www.manaraa.com

Figure 2. Structures of all 16-water clusters.



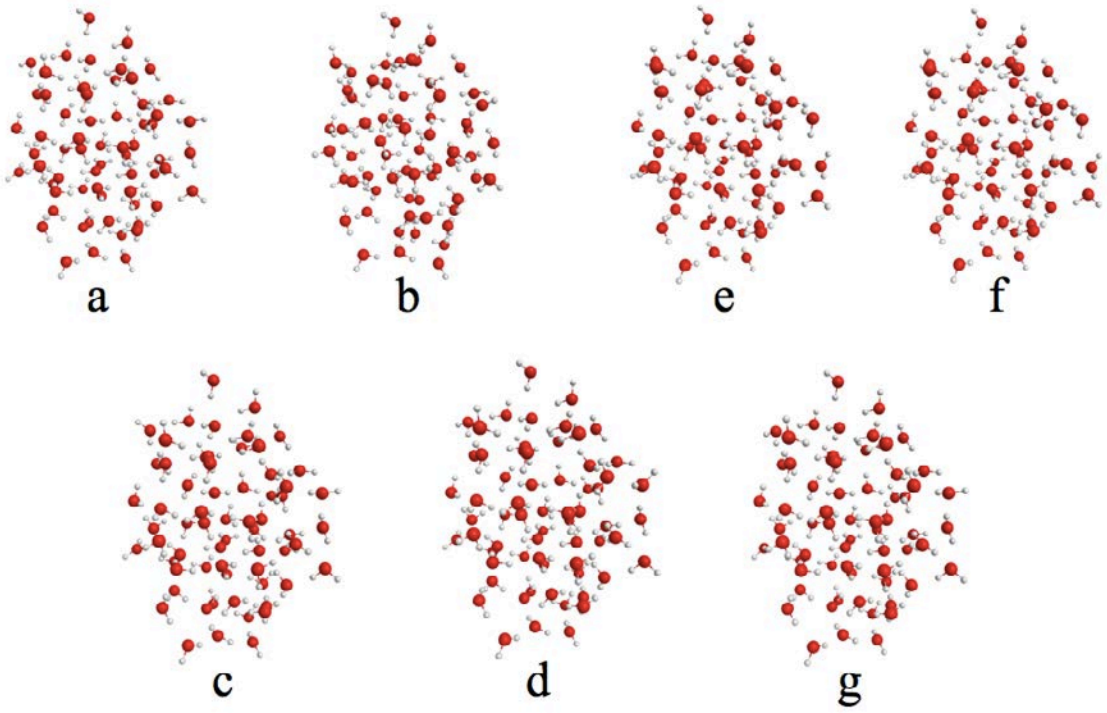
www.manaraa.com

Figure 3. Structures of all 32-water clusters.



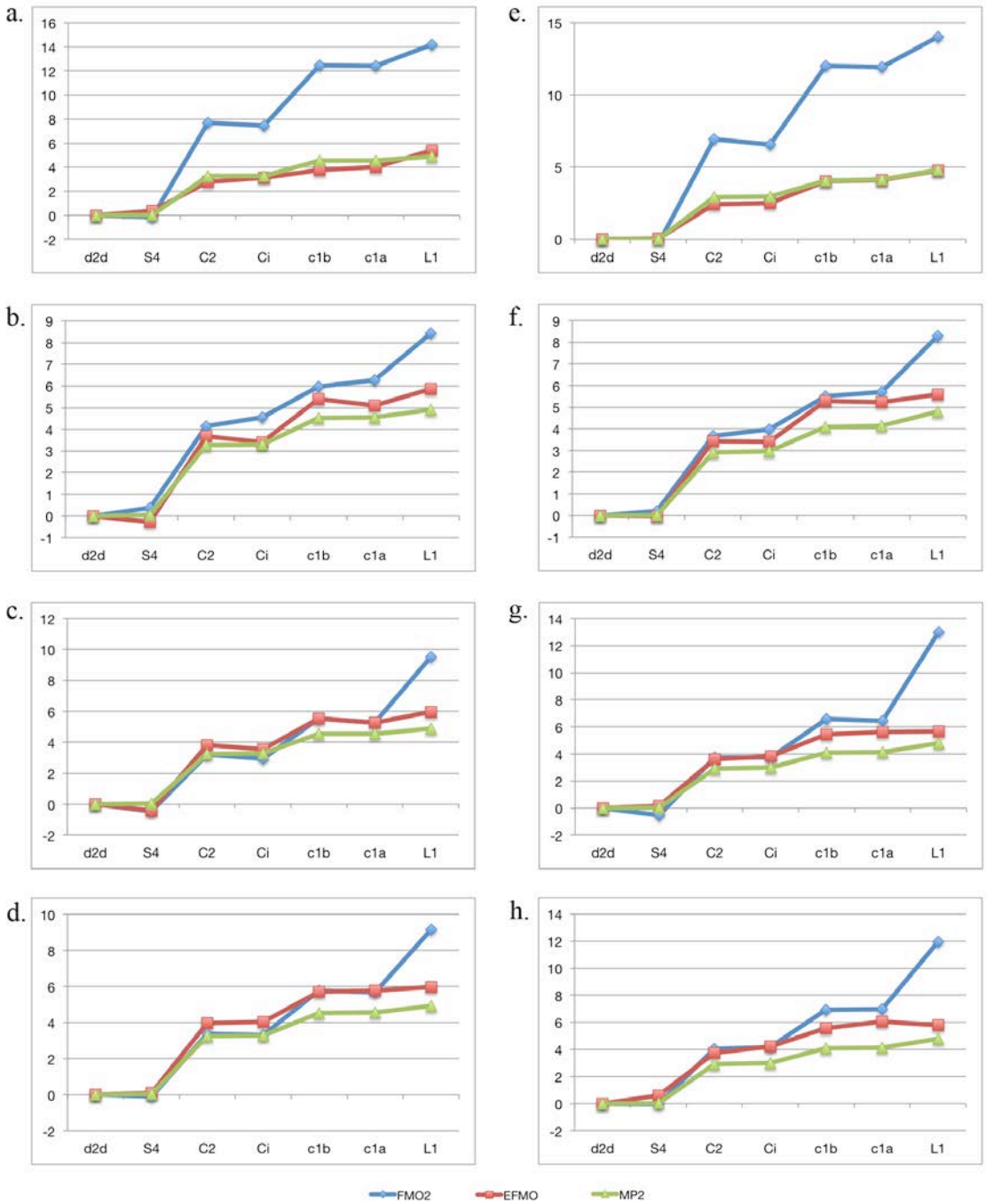
www.manaraa.com

Figure 4. Structures of all 64-water clusters.



www.manaraa.com

Figure 5. Comparison of relative energies for FMO2, EFMO and fully MP2 for clusters of 8 water molecules. Graphs a-d show relative energies for the 6-31++G(d,p) basis set using R_{cut} values of 0.6, 0.8, 1.4 and 2.0 respectively. Graphs e-h correspond to the 6-311++G(3df,2p) basis set results. Isomers are represented on the abscissa of each graph and all energies represented on the ordinate are in kcal/mol.

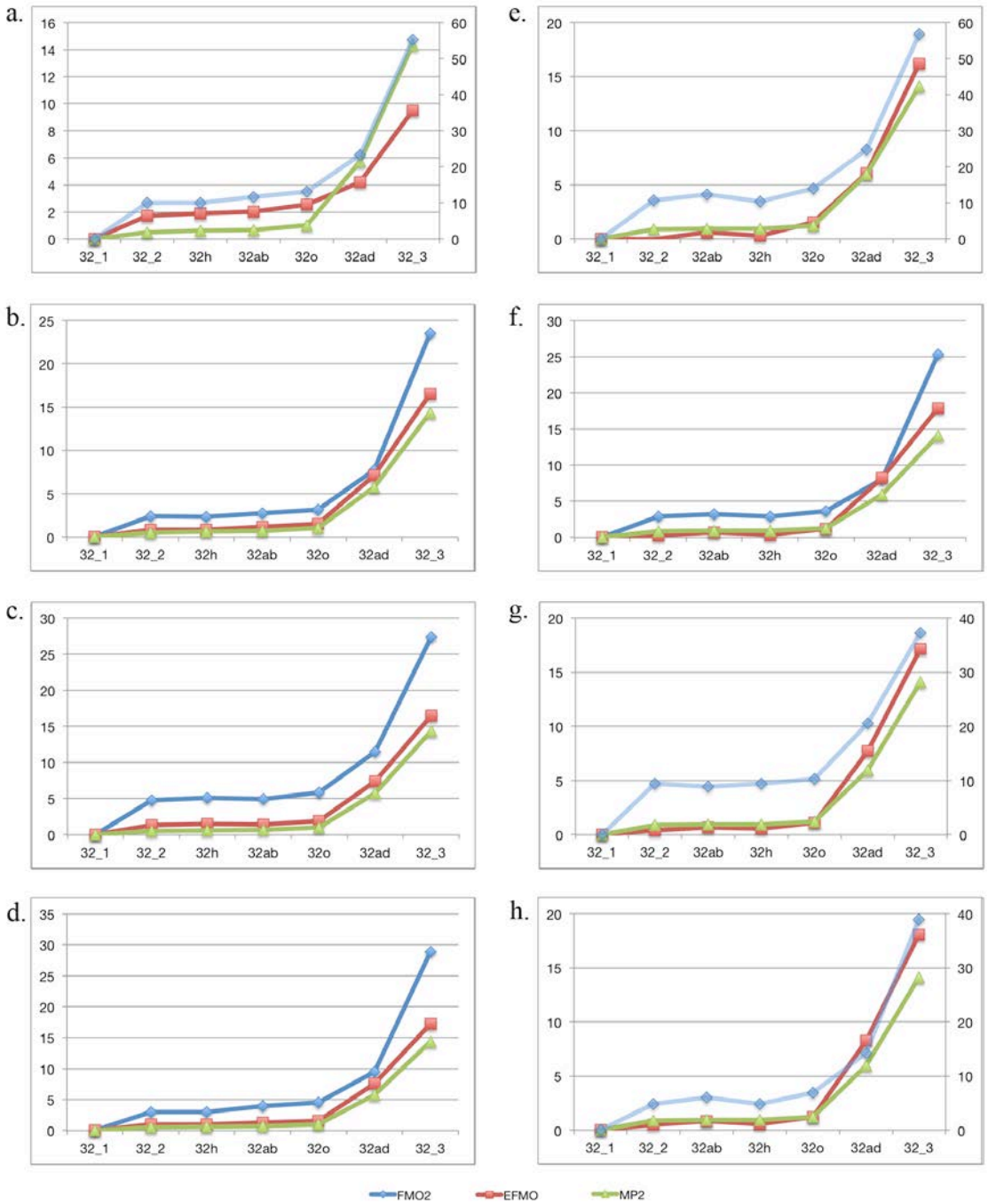


www.manaraa.com

Figure 6. Comparison of relative energies for FMO2, EFMO and fully MP2 for clusters of 16 water molecules. Graphs a-d show relative energies for the 6-31++G(d,p) basis set using R_{cut} values of 0.6, 0.8, 1.4 and 2.0 respectively. Graphs e-h correspond to the 6-311++G(3df,2p) basis set results. Isomers are represented on the abscissa of each graph and all energies represented on the ordinate are in kcal/mol.

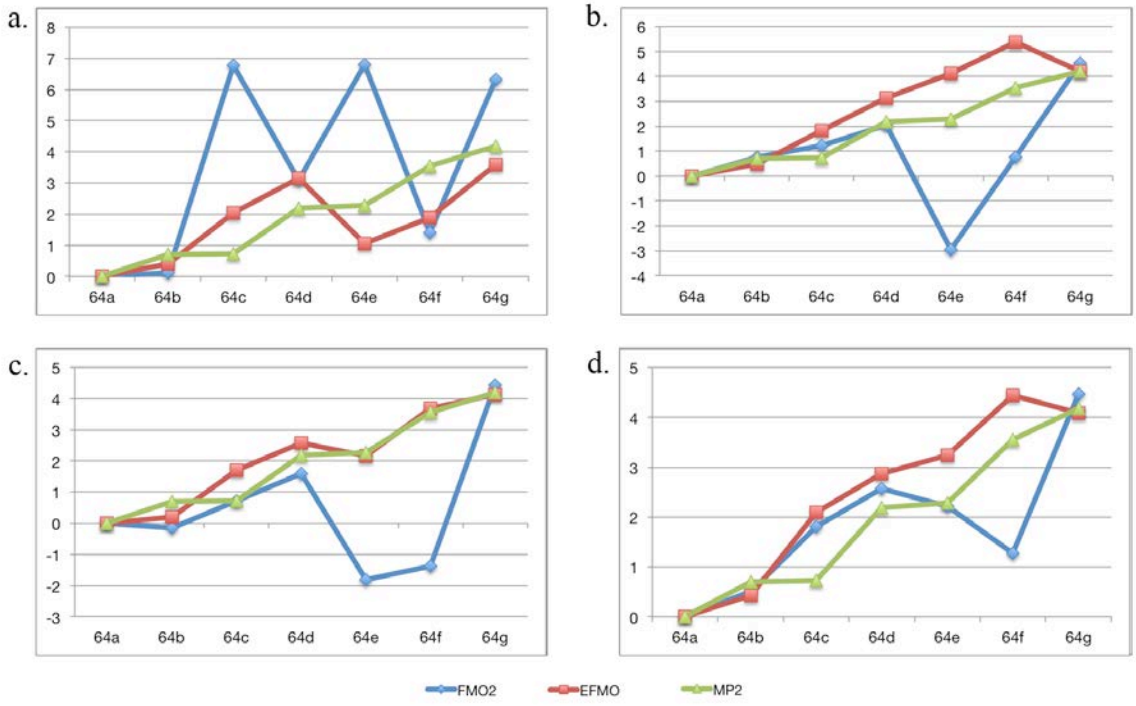
www.manaraa.com

Figure 7. Comparison of relative energies for FMO2, EFMO and fully MP2 for clusters of 32 water molecules. Graphs a-d show relative energies for the 6-31++G(d,p) basis set using R_{cut} values of 0.6, 0.8, 1.4 and 2.0 respectively. Graphs e-h correspond to the 6-311++G(3df,2p) basis set results. In graphs a, e, g and h the FMO2 energy scale is shown on the right ordinate while the energy scale for MP2 and EFMO results is shown on the left ordinate. Isomers are represented on the abscissa of each graph and all energies are in kcal/mol.



www.manaraa.com

Figure 8. Comparison of relative energies for FMO2, EFMO and fully MP2 for clusters of 64 water molecules. Graphs a-d show relative energies for the 6-31++G(d,p) basis set using R_{cut} values of 0.6, 0.8, 1.4 and 2.0 respectively. In graph a the FMO2 energy scale is shown on the right ordinate while the energy scale for MP2 and EFMO results is shown on the left ordinate. Isomers are represented on the abscissa of each graph and all energies are in kcal/mol.



www.manaraa.com

Chapter 7. Conclusions

It has been shown in this thesis that the FMO and SMFA methods both accurately estimate the MP2 energies of water clusters. In contrast, the pair-wise interaction model provides a very poor estimate of binding in water clusters. On average the mean errors of the FMO methods are remarkably close to the corresponding mean errors for SMFA methods. However, the FMO method tends to give systematic, rather than random, deviations from the exact cluster energies. This means that if isomerization energies are important for the physical measurement of interest, then the FMO method would be more reliable than the SMFA method.

The computational effort required for the SMFA method increases nearly linearly with the cluster size, since the number of sub-clusters increases approximately linearly with the size of the whole cluster. Level 2 fragmentation typically results in sub-clusters of five (and less) water monomers, while Level 3 typically results in sub-clusters of eight to nine (and less) water monomers. Moreover, Level 3 fragmentation typically produces a much greater number of fragments. Hence, Level 3 SMFA is significantly more computationally expensive than Level 2. Level 2 SMFA produces an average error of about 0.15 kcal/mol per water monomer (about 1.3% of the total), which would appear to be sufficiently accurate for many applications. It is important to note that non-bonded dispersion must be accounted for if Level 2 fragmentation is used.

The computational effort for FMO also increases nearly linearly for FMO2, while the computational effort of FMO3 increases by a factor of ~4.8. This deviation from nearly linear scaling is due to the quickly increasing number of three body interactions that are required with increasing cluster size. Despite this increase in three body calculations required, the time savings obtained from FMO3 when compared to *ab initio* calculations also increases with increasing system size (77 to 96% with a doubling of system size). The accuracy of FMO2 is more comparable to the SMFA Level 1 method, with an average error of about 0.46 kcal/mol per water monomer. This level of error is clearly unacceptable for many applications, despite the accurate reproduction of the relative energies between water cluster isomers. To produce an acceptable level of error FMO3 must be used, producing errors in agreement with SMFA Level 3 but significantly increasing the computational cost. It is important point out that the accuracy of FMO2 can be increased for single point energy calculations by placing more than one water molecule per fragment, however this type of fragmentation scheme can become problematic for optimizations and molecular dynamics (MD) simulations where individual water molecules may move apart.

MD simulations in aqueous solution are at present primarily carried out using classical force fields. Simulations of water (and solutes in water) require a large number of evaluations of the energy and energy gradient of the whole simulation system. As noted here and elsewhere, the water monomer-monomer interaction fails to account for a large percentage of the water binding energy per molecule, and so does not provide the basis for a quantitatively accurate study of water dynamics, even though pair models are frequently used.[65-67]

www.manaraa.com

FMO and SMFA appear to provide a quantitative approach to the energetics of water dynamics; the challenging task is to incorporate these approaches into a computationally feasible algorithm for dynamical simulations.

As part of this work, the open-shell FMO method has been implemented in the GAMESS program package and parallelized using GDDI for the HF, MP2, and CC levels of electronic structure theory. The accuracy of the method was tested by calculating the absolute and relative energetics of open-shell molecular clusters. The ability of the method to reproduce reaction enthalpies was also tested using the RAFT reaction. It was demonstrated that in both cases the open-shell FMO method provides energies and properties within 0.0-2.0 kcal/mol of *ab initio* calculations.

The need for a single reference open-shell FMO method was fulfilled through this work, providing a scalable method for use on large chemical systems such as the RAFT reaction. The combination of accuracy and reduction in computational expense provides a means for accurate calculations on much larger open-shell radical chemical systems than was previously available.

The gradients for the open-shell FMO method have also been derived and then implemented in the GAMESS program package and parallelized using the GDDI for the ROHF level of electronic structure theory. The ability of the FMO-ROHF method to reproduce accurate total energies and geometries for a variety of chemical systems with varying multiplicities was tested. The accuracy of adiabatic excitation energies was also investigated and it was demonstrated that the open-shell FMO method is capable of producing both accurate geometries as well as adiabatic excitation energies within 0.01 to 0.85 kcal/mol of *ab initio* calculations. Timings and memory requirements for the relatively small test systems also show the ability of the open-shell FMO method to provide a route to geometry optimizations on larger systems.

This work contributes a scalable method for geometry optimizations on large chemical systems through the implementation of the gradient for the single-reference open-shell FMO method. Through the combination of reduced computational cost as well as chemical accuracy shown, the open-shell FMO method provides a means for accurate geometry optimizations on open-shell radical systems much larger than previously possible.

Through this work, the EFMO method has been extended to include all five intermolecular interactions present in the EFP method. The inclusion of short-range intermolecular interactions allows the extended EFMO method to be used with correlated *ab initio* methods such as MP2 and CC. The accuracy of the extended EFMO method was tested versus the standard FMO2 method and MP2. The extended EFMO method was shown to provide superior average errors, relative energies and binding energies for all cluster sizes and basis sets compared to the FMO2 method. Significant reductions in wall clock times compared to the FMO2 method were also shown. Through the reduction of the number of explicit QM dimers performed during extended

www.manaraa.com

EFMO method calculations, time savings of up to 93% compared to the FMO2 method were achieved while at the same time providing a more accurate estimate of the MP2 energies.

Future work on the extended EFMO method will include the implementation of gradients for the newly added energy terms in order to enable geometry optimizations. With the addition of energy gradients, as well as the improved accuracy and reduction in computational requirements versus the FMO2 method, the extended EFMO method could provide a computationally feasible algorithm for dynamical simulations.

In closing, during the last several decades fragment-based methods have come a long way from the initial stage of method development to large scale applications, which span many types of systems: molecular clusters, proteins, DNA, oligosaccharides, zeolites, quantum dots, nanowires and others. Despite the very considerable progress, they remain underused; there may be several reasons for this. Some of the software developments are only locally implemented, making it difficult for most interested users to utilize the methods. Many, perhaps most methods are specific to one particular program; users who are unfamiliar with that program may have an inertial barrier to using it. Second, many applications so far have been performed in what should be considered demonstrative fashion, with low level wave functions and basis sets. Third, in some cases the applications are not conducted while properly considering all necessary effects and factors; the most conspicuous example is the need to incorporate solvent effects and entropy in biochemical applications.

Nevertheless, fragment-based methods also offer many advantages. One is the efficiency and the ability to compute realistic systems. Another is the additional information that they can deliver, such as the intrinsic details of the physical picture of the interactions in the system. The application field is very broad, encompassing most systems of finite size that many chemists and physicists are interested in. It is expected that in near future with the revolutionary progress in the computer technology and the advent of multicore CPUs and GPUs, increasing level of calculations and the ease of their performance, the fragment-based methods will grow more popular in computational community.

www.manaraa.com

Acknowledgements

I would like to thank my family and friends for all of their support and understanding during this 6-year undertaking. Specifically, I want to thank my parents, Mary and Walter Pruitt, for their continued support and encouragement throughout my entire life.

I would also like to thank my advisor Prof. Mark S. Gordon for giving me a chance, as well as Renee Harris for fighting for me at the beginning of my graduate work. Thank you for looking past transcripts and seeing potential.

A special thank you to my wife, Maggie. Without you, none of this would have been possible.

This doctoral thesis is dedicated to both the memory of my grandparents, Russell and Mary Barnard, and the life of my son, Benjamin Rian Pruitt.

www.manaraa.com