

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 06-01-2009		2. REPORT TYPE Journal Article		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Accurate Methods for Large Molecular Systems (Preprint)				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Mark S. Gordon, Jonathan M. Mullin, Spencer R. Pruitt, & Luke B. Roskop (Iowa State University); Lyudmila V. Slipchenko (Purdue University); Jerry A. Boatz (AFRL/RZSP)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 23030421	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Research Laboratory (AFMC) AFRL/RZSP 10 E. Saturn Blvd. Edwards AFB CA 93524-7680				8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-RZ-ED-JA-2009-021	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory (AFMC) AFRL/RZS 5 Pollux Drive Edwards AFB CA 93524-70448				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S NUMBER(S) AFRL-RZ-ED-JA-2009-021	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited (PA #09071).					
13. SUPPLEMENTARY NOTES For publication in the Journal of Physical Chemistry.					
14. ABSTRACT Three exciting new methods that address the accurate prediction of processes and properties of large molecular systems are discussed. The systematic fragmentation method (SFM) and the fragment molecular orbital (FMO) method both decompose a large molecular system (e.g., protein, liquid, zeolite) into small subunits (fragments) in very different ways that are both designed to retain the high accuracy of the chosen quantum mechanical level of theory while greatly reducing the demands on computational time and resources. Both of these methods are inherently scalable and are therefore eminently capable of taking advantage of massively parallel computer hardware. The effective fragment potential (EFP) method is a very sophisticated approach for the prediction on non-binded and intermolecular interactions. Therefore, the EFP method provides a way to further reduce the computational effort while retaining accuracy, by treating the far field interactions in place of the full electronic structure method. The performance of the methods is demonstrated using applications to several systems, including benzene dimer, small organic species, pieces of the alpha helix, water, and ionic liquids.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 75	19a. NAME OF RESPONSIBLE PERSON Dr. Jerry A. Boatz
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code) N/A

ACCURATE METHODS FOR LARGE MOLECULAR SYSTEMS

(PREPRINT)

Mark S. Gordon^{*,‡}, Jonathan M. Mullin[‡], Spencer R. Pruitt[‡], Luke B. Roskop[‡], Lyudmila V. Slipchenko[§], and Jerry A. Boatz[&]

[‡] Department of Chemistry, Iowa State University, Ames, IA 50011

[§] Department of Chemistry, Purdue University, West Lafayette, IN 47907

[&] Air Force Research Laboratory, Space and Missile Propulsion Division, AFRL/RZS, 10 E. Saturn Blvd., Edward AFB, CA 93524

Abstract

Three exciting new methods that address the accurate prediction of processes and properties of large molecular systems are discussed. The systematic fragmentation method (SFM) and the fragment molecular orbital (FMO) method both decompose a large molecular system (e.g., protein, liquid, zeolite) into small subunits (fragments) in very different ways that are both designed to retain the high accuracy of the chosen quantum mechanical level of theory while greatly reducing the demands on computational time and resources. Both of these methods are inherently scalable and are therefore eminently capable of taking advantage of massively parallel computer hardware. The effective fragment potential (EFP) method is a very sophisticated approach for the prediction on non-binded and intermolecular interactions. Therefore, the EFP method provides a way to further reduce the computational effort while retaining accuracy, by treating the far field interactions in place of the full electronic structure method. The performance of the methods is demonstrated using applications to several systems, including benzene dimer, small organic species, pieces of the alpha helix, water, and ionic liquids.

Distribution A: approved for public release; distribution unlimited.

1. Introduction

The development of quantum chemistry methods in the 1980s and 1990s primarily focused on performing very accurate calculations on relatively small molecular systems. The desire for accurate calculations on larger molecular species led to several formulations employing more efficient scaling, as well as additivity of basis set improvement and higher levels of electron correlation. With regard to the latter, the Gaussian $G(n)$ ¹ methods and the Weizmann $W(n)$ ² methods are well known, along with several variants.³

Simultaneous progress in the development of systematically improving atomic basis sets also provided a path toward systematic increases in accuracy. It was recognized⁴ that basis functions optimized for atomic correlation are also capable of describing molecular correlation effects. Dunning and co-workers, for example, introduced a series of correlation consistent basis set sets⁵ based upon these conclusions, capable of accurately treating electron correlation with a compact set of primitive Gaussian functions. These basis sets can be used in a systematic way to obtain results approaching the complete basis set (CBS) limit. However, increasingly large basis sets must be used, and the convergence tends to be slow. Werner has recently introduced a series of F12 basis sets⁶ with improved convergence to the CBS limit. The high accuracy of these basis sets still comes at a significant computational cost, only feasible on relatively small systems.

Chemical phenomena occur in condensed phases as well as in the gas phase, and many methods have been developed to treat the chemical environment⁷ and condensed phase phenomena.⁸ The desire to study ever larger systems led to combining quantum

mechanics (QM) with molecular mechanics (MM). Several such combinations, known as QM/MM methods,⁹ have been developed since the initial work of Warshel,^{9a} including multi-layer methods such as ONIOM,¹⁰ the Truhlar MCMM methods,¹¹ and the effective fragment potential method (EFP)¹²⁻²⁷ developed by Gordon and co-workers. The EFP method will be discussed in detail as a means to investigate non-bonded and intermolecular interactions via the automatic generation of a model potential that is derived from first principles. While hybrid methods have expanded the size of systems that are accessible to computations, the use of classical model potentials for the description of the environment can be a limiting factor, given that the electron density of the MM region and its impact on the QM region is not usually properly accounted for.

Alternative approaches to QM/MM methods are fragmentation methods, in which the system is broken (“fragmented”) into smaller pieces, each of which is considered essentially independently by a specified level of electronic structure theory.

Fragmentation methods have the advantage that they are fully quantum mechanical in nature. Several general fragmentation methods have been proposed, including molecular fragmentation with conjugated caps (MFCC),²⁸ the elongation method,²⁹ the molecular tailoring approach (MTA),³⁰ the fast electron correlation method for molecular clusters developed by Hirata,³¹ Truhlar's electrostatically embedded many-body (EE-MB) expansion,³² multi-centered QM/QM methods,³³ the systematic fragmentation method (SFM),³⁴⁻³⁸ and the fragment molecular orbital (FMO) method.³⁹⁻⁴⁵ The latter two methods, the SFM and the FMO methods, will be discussed in detail in this work.

Instead of separating a system into two regions that are described by two very different levels of theory (QM and MM), fragmentation methods that divide a system into

many smaller pieces, all of which are described by the same level of QM theory, have been proposed since the 1970s.⁴⁶ By approaching a large system in this way, each smaller fragment can be treated using high levels of theory, providing the desired accuracy and an improvement in speed. The earliest attempts⁴⁶ constructed a set of fragments from common chemical groups (methyl, amino etc.) and used a selection of these fragments to build larger molecules. More recent fragmentation methods²⁸⁻⁴⁵ begin with the larger molecule of interest and break the system into smaller fragments.

Fragmentation methods must also treat the environment (e.g., the remainder of the entire molecular system, or a solvent) around each fragment in some approximate, but realistic manner. When fragmenting a molecule or a molecular system, each fragment no longer electronically “feels” the remainder of the initial system, unless one devises some way to retain the lost interactions. This issue is addressed in the FMO method³⁹ by performing each individual fragment calculation in a Coulomb “bath” represented by the electrostatic potential (ESP) of the entire system. Further corrections to the FMO method are achieved by performing fully quantum mechanical two-fragment (dimer) and three-fragment (trimer) calculations. In the SFM method³⁴ the effects of other fragments are incorporated by including overlapping fragments in such a manner that the double counting of atoms is accounted for, and non-bonded interactions are captured by employing classical potentials.¹² Accurately capturing non-bonded effects is essential to maintaining kcal/mol accuracy compared to full *ab initio* studies.

Traditional electronic structure methods, such as Hartree-Fock (HF), second order perturbation theory (MP2), and coupled cluster theory (e.g. CCSD(T)) have rapidly increasing resource requirements (e.g. time, memory, mass storage). For example, the

HF, MP2 and CCSD(T) computer time requirements scale as $O(n^4)$, $O(n^5)$ and $O(n^7)$ respectively, where n measures the size of the system, e.g., in terms of the basis set size. Further, CCSD(T) memory requirements scale as $O(n^4)$, while disk requirements are difficult to uniquely define. One approach to addressing the computational scaling issue is to develop highly parallel algorithms. The development of parallel algorithms for electronic structure theory has been an active research area for ~20 years, and considerable progress has been achieved for increasingly complex QM methods.⁴⁷ Such efforts may be referred to as fine-grained parallelism, in the sense that each energy or derivative evaluation itself takes advantage of many cores, usually in a distributed manner.⁴⁸ In many fragmentation methods each fragment calculation can be performed essentially independently of all the others. This leads to a multi-level parallelism, since the energy of each fragment can be obtained on a separate node (coarse-grained parallelism), while the fine-grained parallelism can be exploited within each node.⁴⁹ If a fragmentation method is implemented to take advantage of this ability, large reductions in required computational resources can be achieved, facilitating calculations on, for example, condensed phases, proteins and surfaces. Fragmentation approaches with multi-level parallelism also expand the capabilities of modest (e.g., single group or departmental) computer systems.

The present work focuses on three methods that have been designed to accurately treat large systems: EFP, SFM, and FMO. Benzene dimer is chosen as a representative example to illustrate the accuracy and efficiency of the EFP method, although several such studies have been carried out,⁵⁰ as have EFP molecular dynamics simulations.⁵¹ The EFP approach has also been included as a means to capture accurate non-bonded

interactions within the SFM framework,³⁹ as a replacement for simple electrostatics. It will be illustrated that the FMO method can be used to accurately describe a series of water clusters and ionic liquid systems.

2. Effective Fragment Potential (EFP) Method

The generalized effective fragment potential (EFP2) method¹² is a first-principles based model potential for the evaluation of intermolecular forces. This is a modification and extension of the original EFP1 water model¹³⁻¹⁷ to general systems. There are five EFP-EFP interaction terms in the EFP2 model potential, along with damping functions, each of which may be thought of as a truncated expansion:

Coulombic (electrostatic), induction (polarization), exchange repulsion, dispersion (Van der Waals), and charge transfer.

$$E = E_{\text{coul}} + E_{\text{ind}} + E_{\text{exrep}} + E_{\text{disp}} + E_{\text{ct}} \quad (1)$$

In EFP1, the exchange repulsion, E_{exrep} , and charge transfer, E_{ct} , components are folded into one term that contains fitted parameters, and there is no dispersion contribution. EFP1 has been integrated with HF,¹³ DFT,¹⁴ MCSCF,¹⁵ singly excited configuration interaction (CIS),¹⁶ and time-dependent density functional theory (TDDFT).¹⁷ The EFP2-QM interface is still under development.¹⁸

The five terms in the EFP potential may be grouped into long range, $(1/R)^n$ distance dependent, or short range interactions, which decay exponentially. The Coulombic, induction, and dispersion are long-range interactions, whereas the exchange repulsion and charge transfer are short range. EFP has been described in detail in several papers,¹²⁻¹⁷ therefore only a brief overview of the terms will be presented below.

The Coulomb portion of the electrostatic interaction, E_{Coul} , is obtained using the

Stone distributed multipolar analysis.¹⁹ This expansion is truncated at the octopole term. Atom centers and the bond midpoints are used as expansion points.

Induction (polarization), E_{ind} , arises from the interaction of an induced dipole on one fragment with the permanent dipole on another fragment, expressed in terms of the dipole polarizability. Truncating at the first (dipole) term in the polarizability expansion is viable, since the molecular polarizability tensor is expressed as a tensor sum of localized molecular orbital²⁰ (LMO) polarizabilities. Therefore, the number of bonds and lone pairs in the system is equal to the number of polarizability points. This induction term is iterated to self-consistency, so it is able to capture some many-body effects.

Because the Coulomb and induction terms discussed above, as well as the dispersion interaction, are treated by classical approximations, the shorter range interactions that occur when quantum mechanical charge densities begin to overlap are not correctly captured. Therefore, each term is multiplied by a damping (screening) expression. The relative merits of several approaches to damping have recently been analyzed and discussed extensively.²² Classical Coulombic interactions become too repulsive at short range, and must be moderated by a screening term, as discussed in several previous papers.^{21,22} Conversely, the induction interaction becomes too attractive in the short-range regime, so a damping term is needed here as well. The unphysical behavior is avoided by augmenting the electrostatic multipoles with exponential damping functions of the form:

$$f_{\text{damp}} = 1 - \exp(-\alpha R) \quad (2)$$

where parameters α are determined at each multipole expansion point by fitting the multipole damped potential to reproduce the Hartree–Fock potential. Damping terms in

the electrostatic energy are derived explicitly from the damped potential and the charge density. The damping procedure can be extended to higher-order electrostatic terms; that is, the charge–dipole, dipole–dipole, etc., interactions, and this is recommended.²¹

Damping is also applied to the induction and dispersion energies.^{21,22} For induction, both exponential damping, as in Eq. (2) and Gaussian damping are effective, but the Gaussian damping seems to be more generally applicable and is therefore recommended.

The exchange repulsion interaction between two fragments is derived as an expansion in the intermolecular overlap.²³ When this overlap expansion is expressed in terms of frozen LMOs on each fragment, the expansion can reliably be truncated at the quadratic term. This term does require each EFP to carry a basis set. Since the same basis set is used to generate the multipoles and the molecular polarizability tensor, EFP calculations are basis set dependent. The smallest recommended basis set is 6-31++G(d,p)⁵² The dependence of the computational cost of an EFP calculation on the basis set appears primarily in the initial generation of the EFP. Therefore, one can employ much larger basis sets with minimal cost. The tests presented below on the SFM method use the 6-311++G(3df,2p)⁵³ basis set. Since the basis set is used only to calculate overlap integrals, the computation is very fast and quite large basis sets are realistic.

Dispersion interactions are often expressed by an inverse R expansion,

$$E_{disp} = \sum_n C_n R^{-n} \quad (3)$$

where the coefficients C_n may be derived from the (imaginary) frequency dependent polarizabilities integrated over the entire frequency range.²⁴ The first term in the expansion, $n=6$, corresponds to the induced dipole-induced dipole (Van der Waals) interactions. In the EFP2 method, this term is evaluated using the time-dependent HF

method. In addition the contribution of the $n=8$ term is estimated empirically. The C_6 coefficients are derived in terms of interactions between pairs of LMOs, one each on the two interacting fragments. Because the dispersion interaction should decrease to zero at short range, each dispersion term is multiplied by a damping function. Tang-Toennies damping²⁵ is frequently used to damp dispersion. However, a new approach that is based on the overlap integrals between interacting fragments²² is free of fitted parameters and appears to be generally applicable. In future EFP applications, the overlap-based dispersion damping is recommended.

The charge transfer interaction is derived using a supermolecule approach, in which the occupied valence molecular orbitals on one fragment are allowed to interact with the virtual orbitals on another fragment. This interaction term leads to significant energy lowering in *ab initio* calculations on ionic or highly polar species when incomplete basis sets are employed. An approximate formula²⁶ for the charge transfer interaction in the EFP2 method was derived and implemented using a second order perturbative treatment of the intermolecular interactions for a pair of molecules at the Hartree–Fock level of theory. This approximate formula is expressed in terms of the canonical orbitals from a Hartree–Fock calculation of the isolated molecules and uses a multipolar expansion (through quadrupoles) of the molecular electrostatic potentials. Orthonormality is enforced between the virtual orbitals of the other molecule and all of the orbitals of the considered molecule, so that the charge transfer is not contaminated with induction. This approximate formula has been implemented in the EFP method and gives charge transfer energies comparable to those obtained directly from Hartree Fock calculations.²⁶ The analytic gradients of the charge transfer energy were also derived and implemented,

enabling efficient geometry optimization.²⁷

It is useful to consider the relative costs of the five EFP interaction terms. Based on relatively small molecules and taking the cost of the Coulomb and dispersion terms to be one unit, the induction interaction would cost approximately two units, exchange repulsion would cost about five units, and charge transfer would cost ~10 units. For larger molecules, the relative costs of exchange repulsion and charge transfer will decrease since they will scale linearly in the large molecule limit. As always in computational chemistry there is a trade off between computational cost and accuracy.

While the EFP model is currently a rigid body model potential, analytic gradients for all terms have been derived and implemented, so full intermolecular geometry optimizations, Monte Carlo and molecular dynamics simulations^{50,51} can be performed. Because the method involves no empirically fitted parameters, an EFP for any system can be generated by a “makefp” run in the GAMESS⁵⁴ suite of programs. This automatic generation makes possible the use of the EFP method for treating intermolecular and non-bonded interactions in fragmentation methods such as the SFM.

2.1 The EFP Method as a Model for Non-Bonded Interactions

Benzene dimer is used here to illustrate the accuracy of EFP-EFP non-bonded interactions, with a focus on the $\pi-\pi$ interactions between two benzene rings. These $\pi-\pi$ interactions are largely driven by dispersion and are therefore difficult to account for accurately by most *ab initio* electronic structure methods. Previous theoretical and experimental studies suggest that there are two minima on the benzene dimer potential energy surface⁵⁵⁻⁶² the perpendicular T-shaped and parallel-slipped configurations, as shown in Figure 1. A sandwich structure with two parallel benzene rings, also shown in

Figure 1, is a saddle point that connects two equivalent parallel slipped structures. Sherrill and co-workers calculated potential energy curves for these three structures^{56,57} using second order Møller-Plesset perturbation theory (MP2) and coupled-cluster theory with single, double and perturbative triple excitations [CCSD(T)]⁶³ A variety of augmented correlation consistent basis sets⁸ were used with both the MP2 and CCSD(T) levels of theory. Additionally, they employed symmetry-adapted perturbation theory (SAPT)⁶⁴ to decompose the benzene $\pi-\pi$ interaction energy into electrostatic, dispersion, induction and exchange-repulsion components of the total interaction energy. The binding energies, equilibrium separations and SAPT energy decomposition results from their work compare well with similar results obtained using the EFP method, as illustrated below.^{21a}

For this work, the EFP for benzene was constructed with the 6-311++G(3df,2p) basis set,⁵³ using the MP2/aug-cc-pVTZ⁶⁴ benzene monomer geometry taken from Ref. 57. The multipoles for benzene were generated using a numerical distributed multipolar analysis (DMA).¹⁹ The numerical DMA scheme is employed due to the instability and basis set dependence of the standard analytic DMA scheme, as well as the need for diffuse functions to properly describe the exchange-repulsion interactions within the EFP framework. Higher order (up to quadrupoles) damping terms were also used to provide an accurate description of charge penetration through screening of the potentials.²¹

The EFP binding energies and corresponding inter-ring distances for the three benzene dimer structures are in good agreement with the analogous *ab initio* values obtained by Sherrill and co-workers (See Table 1). Relative to the full CCSD(T)/aug-cc-pVQZ binding energies, the EFP method over-binds the sandwich dimer by 0.4 kcal/mol

and under-binds the T-shaped structure by 0.1 kcal/mol, while the equilibrium intermolecular separations are overestimated by approximately 0.1–0.2 Å. In comparison, MP2 with the same basis set overestimates the binding energies by 1.7 kcal/mol and 0.9 kcal/mol for the sandwich and T-shaped dimers, respectively, and underestimates the equilibrium distances by approximately 0.1–0.2 Å. In fact, the MP2 binding energies become successively worse compared with those predicted by CCSD(T) as the basis set is improved. The EFP and CCSD(T) predicted binding energies and structures are in reasonable agreement with each other, whereas the agreement between MP2 and CCSD(T) is not as good. Table 1 summarizes the MP2, CCSD(T) and EFP total interaction energies of all three benzene dimer structures. A comparison of the total interaction energy decompositions obtained using both SAPT and the EFP method shows good agreement for the sandwich and T-shaped isomers (See Figures 2 and 3). Specifically, the error in the EFP method compared to SAPT for the dispersion, exchange-repulsion and polarization interactions is in the range 0.2-0.5 kcal/mol for these two isomers.^{21a}

Highly accurate methods involve very demanding scaling of computational resources, such as time, memory and disk. For instance, a single point energy calculation in the 6-311++G(3df, 2p) basis set (660 basis functions) by MP2 requires 142 minutes of CPU time on one IBM Power5 processor, whereas the analogous EFP calculation requires only 0.4 seconds. The corresponding CCSD(T) calculation would be much more resource demanding than MP2. Taking into account the relatively good agreement of the EFP method results with the CCSD(T)/aug-cc-pVQZ results described above, this significant reduction in total computation time comes with a minimal loss of accuracy.

3. The Systematic Fragmentation Method (SFM)

The systematic fragmentation method (SFM) is designed to permit a large molecular system, such as a protein, to be fragmented into smaller pieces in such a way that retains the accuracy of a full *ab initio* calculation at the same level of theory, while significantly decreasing the computational expense. By treating the smaller sub-systems with accurate levels of theory, the total energy and properties of the full system are obtained through addition and subtraction of the contributions from the overlapping sub-systems or “groups.” Many body effects are accounted for including the nearest neighbor of each group. Non-bonded interactions between groups are also accounted for. In the original formulation³⁴ these non-bonded interactions were obtained using a classical electrostatic potential. Recently, this non-bonded description has been improved through the use of the EFP method, providing a more accurate representation of the non-bonded interactions.³⁸

Within the context of the SFM, a molecule can be thought of as a collection of functional groups. For example, ethanol contains three functional groups (CH_3 , CH_2 , and OH) according to the SFM prescription. To fragment the system into functional groups, single bonds are broken. This process splits a pair of bonding electrons; each of these electrons is assigned to one of the two resulting fragments. In order to avoid the resulting radical species, a hydrogen atom is used to “cap” the dangling bonds that are created by the fragmentation. The capping hydrogen points in the direction of the broken bond at a chemically reasonable distance. By design, double or triple bonds are not broken, keeping the relevant atoms as a part of one functional group. For example, ethanal would contain two functional groups (CH_3 and CHO), keeping the carbon and oxygen atoms of

the carbonyl together in one group. After the addition of the hydrogen caps, the ethanol groups would be CH₄, CH₄, and H₂O, and the ethanal groups would be CH₄ and CH₂O.

To gain a more quantitative understanding of the SFM, consider the general example of a linear molecule M containing K functional groups G_i :

$$M = G_1 G_2 G_3 \dots G_k. \quad (4)$$

Each group G_i is connected by single bonds to adjacent groups G_{i-1} and G_{i+1} . In order to separate the functional groups of M into smaller fragments, one can imagine breaking the $G_{i-1} - G_i$ single bond, then capping each new terminal atom with a hydrogen atom. This produces two new, smaller species,

$$M_1 = G_1 G_2 G_3 \dots G_{i-1} H_{i-1} \quad (5)$$

$$M_2 = H_i G_i G_{i+1} \dots G_k. \quad (6)$$

The internal geometries of M_1 and M_2 are preserved, except for the hydrogen atoms that have been used to cap the missing bond vector. The total energy can then be written, without approximation, as:

$$E(M) = E(M_1) + E(M_2) + dE_1, \quad (7)$$

where dE_1 is the correction for the differential change in the energy caused by breaking a bond and adding two hydrogen caps. This process can be repeated, since bonds can be broken at any point along the chain, decomposing the full system into many smaller fragments. As the separation between the bond breaks is increased, the accuracy of the SFM will increase, since the larger fragments will give a more accurate description of the full system. The separation between broken bonds can be described as different “levels” of the SFM.

The SFM levels are defined as follows.³⁴

Consider the molecule M :

$$M = G_1G_2G_3G_4G_5G_6G_7G_8 \quad (8)$$

In the Level 1 SFM, two bonds separated by just one functional group are sequentially broken. The fragments initially created would, for example, be as follows:

$$M \approx G_1G_2 + G_2G_3G_4G_5G_6G_7G_8 - G_2 \quad (9)$$

The G_2 fragment is subtracted off to conserve the number of atoms. Subsequently, this process is repeated exhaustively on the $G_2G_3G_4G_5G_6G_7G_8$ fragment until no fragment larger than 2 functional groups remains. In the end, the energy of M can be approximately decomposed into the simple sum of fragment energies for level 1 as follows:

$$E_{\text{level 1}}^{\text{bonded}}(M) = E(G_1G_2) + E(G_2G_3) + E(G_3G_4) + E(G_4G_5) + E(G_5G_6) + E(G_6G_7) + E(G_7G_8) \\ - E(G_2) - E(G_3) - E(G_4) - E(G_5) - E(G_6) - E(G_7) \quad (10)$$

Similarly in the Level 2 SFM, bonds separated by two functional groups are sequentially broken with the energy of M being decomposed into the following expression:

$$E_{\text{level 2}}^{\text{bonded}}(M) = E(G_1G_2G_3) + E(G_2G_3G_4) + E(G_3G_4G_5) + E(G_4G_5G_6) + E(G_5G_6G_7) + E(G_6G_7G_8) \\ - E(G_2G_3) - E(G_3G_4) - E(G_4G_5) - E(G_5G_6) - E(G_6G_7) \quad (11)$$

In the level 3 SFM, bonds separated by three functional groups are sequentially broken with the energy of M being decomposed into the following expression:

$$E_{\text{level 3}}^{\text{bonded}}(M) = E(G_1G_2G_3G_4) + E(G_2G_3G_4G_5) + E(G_3G_4G_5G_6) + E(G_4G_5G_6G_7) + E(G_5G_6G_7G_8) \\ - E(G_2G_3G_4) - E(G_3G_4G_5) - E(G_4G_5G_6) - E(G_5G_6G_7) \quad (12)$$

It is important to note that in the limit of SFM, that is for level n , where n is the number of groups in the system, one would be left with the un-fragmented system. So the higher the SFM level employed, the larger the fragments and the closer one should get to the

energy of the exact un-fragmented system.

There are some limitations of the SFM. First, as noted earlier, the SFM is unable to fragment conjugation in delocalized molecular systems. The second, less obvious, limitation is that the SFM is unable to fragment six member rings using level 3 since the capping hydrogens would approach each other too closely and would therefore cause unphysical repulsive interactions. To avoid this, the ring must remain intact and is considered to be a functional group itself. Similarly, five member rings can only be fragmented at level 1; four and three member rings cannot be fragmented at all. These exceptions are referred to as the ring repair rule.

3.1 Non-bonded interactions

The simplest approach to obtain the energy of the system of interest would be to calculate the energies of the individual hydrogen-capped fragments and sum them accordingly. The result obtained from this procedure would differ greatly from the analogous calculation on the full molecular system. This is because the (non-bonded) interactions among the separated fragments are unaccounted for. These non-bonded interactions are naturally incorporated into the full *ab initio* calculation. The non-bonded interactions are modeled within the SFM framework by using a modified many body expansion;³⁷ this expansion relies on the assumption that bonded interactions are much stronger than non-bonded ones.

3.2 Two-body interactions

The interaction energy between two functional groups G_1 and G_2 is given by

$$E_{\text{nb}}^{(1,1)}[G_1; G_2] = E(G_1 G_2) - E(G_1) - E(G_2), \quad (13)$$

where $E(G_1 G_2)$ is the super-molecular energy of the two separated functional groups

(placed in their positions in the original full molecule M) and $E(G_1), E(G_2)$ are the corresponding (one-body) fragment energies. The total two-body non-bonded energy of the system contains the energies of all possible pairs of functional groups that are not described by the fragmentation of the bonded system in the definition of M . That is, all pairs of groups G_1, G_2 that are not contained in any one fragment.

3.3 Three-body interactions

The mutual interaction of three functional groups G_1, G_2 and G_3 is assumed to be negligible unless any two of the groups are bonded to each other. For example, if G_3 is bonded directly to G_2 then the three-body interaction energy would be:

$$E_{\text{nb}}^{(1,2)}[G_1; G_2, G_3] = E(G_1 G_2 G_3) - E(G_1) - E(G_2 G_3) - E_{\text{nb}}^{(1,1)}[G_1; G_2] - E_{\text{nb}}^{(1,1)}[G_1; G_3] \quad (14)$$

In other words, the three-body energy is simply the super-molecular energy, $E(G_1 G_2 G_3)$, minus the one-body energies $E(G_1), E(G_2 G_3)$ and minus the two-body energies,

$E_{\text{nb}}^{(1,1)}[G_1; G_2], E_{\text{nb}}^{(1,1)}[G_1; G_3]$. The total three-body energy consists of all combinations containing any group (G_1) with any two bonded functional groups (G_2 or G_3), so long as G_1 is itself not present in any bonded fragment with G_2 or G_3 . This general trend can be extended to four body interactions and beyond; however for the purposes of this work only three-body terms will be treated.

The total SFM energy of a system is simply the addition of the bonded and non-bonded energies,

$$E_{\text{SFM}}^{\text{total}} = E^{\text{bonded}} + E^{\text{non-bonded}}, \quad (15)$$

where $E^{\text{non-bonded}}$ includes all terms up to n^{th} order from the modified many-body

approximation. For example, calculations employing 3rd order many body non-bonded energies would include the 2nd order many body non-bonded energies as well.

3.4 SFM and EFP

SFM molecular energy calculations corresponding to bonded level 3 including many body non-bonded interactions apparently provide, on average, the best balance between accuracy and computational effort.³⁵ Although the non-bonded approximation is important for reliability, it also hinders computational performance by significantly increasing the number of *ab initio* terms. For example, moderately sized proteins (~3500 residues) have on the order of 10⁶ non-bonded interactions. Because there are so many non-bonded terms, these terms can dominate the calculation. It is therefore advantageous to employ approximate methods for those non-bonded interactions that are sufficiently distant that classical approximations might be valid. The simplest approach, using just electrostatic interactions, were used in the original SFM implementation.³⁵ A more sophisticated approach, using effective fragment potentials (EFP), is described here. Compared to electrostatics, intermediate range (2.7-4.5Å) EFP interaction energies agree better with *ab initio* methods. This increases the number of non-bonded terms that can be calculated with model potentials.

The determination of whether a non-bonded term is treated with EFP or *ab initio* methods is based on a user-defined cutoff related to the nearest atom-atom distance between fragments. The short-range (< 2.7Å) non-bonded terms use *ab initio* methods, while long-range (≥2.7Å) ones use EFP. The original electrostatic approach³⁸ used a cutoff of 4.5Å. This shortened EFP cutoff comparatively reduces the number of *ab initio* non-bonded terms, thereby decreasing the computational expense.

Previously, Collins and Deev tested the SFM by calculating the isomerization energies of a set of organic molecules (12-44 heavy atoms) obtained from the Cambridge Structural Database.⁶⁵ A subset of this set of isomerization energies is examined here. The energies that are obtained by employing the EFP/*ab initio* non-bonded approach are compared with both the fully *ab initio* energies (no SFM) and to the SFM in which all non-bonded terms are calculated with the same *ab initio* method that is used for the bonded terms. The *ab initio* calculations here employ both the Hartree-Fock (HF) and second order perturbation theory (MP2) levels with the 6-31G(d,p) basis set. Additional SFM tests are presented for a small set of alpha helices using the 6-31++G(d,p) basis set. The larger 6-311++G(3df,2p) basis set is employed for creating all EFPs used for non-bonded interactions, since this basis set has been shown to produce reliable results and since the EFP basis set dependence does not significantly affect the computational cost relative to *ab initio* calculations. All of the SFM calculations presented here correspond to bonding level 3, including up to 3rd order many body non-bonded interactions. All calculations are performed with the GAMESS⁵⁴ electronic structure code.

Given in Table 2 are the errors in the isomerization energies. The corresponding structures are depicted in Scheme 1. It is evident that the two methods for treating the SFM non-bonded energy (EFP/*ab initio* and *ab initio* only) are in reasonable agreement with the fully *ab initio* (non SFM) energies, as the mean absolute error (MAE) in all cases is no more than 2.5 kcal/mol. Addition of the 3rd order non-bonded terms does not result in any improvement to the MAE. Interestingly, the MAE for the combined EFP/*ab initio* approach for the non-bonded terms are slightly smaller (~0.1-0.5 kcal/mol) than those obtained when the non-bonded terms are evaluated with the *ab initio* method (HF

or MP2). For the 21 molecules of interest here, as also noted by Collins and Deev³⁵, no improvement in the net CPU time is observed since the molecules themselves are small. Improvements in CPU timings are observed for larger molecular systems (>100 atoms), as discussed below.

SFM isomer energies for the larger model alpha helices (ranging from 125-170 atoms) are shown in Table 3, with the corresponding structures presented in Scheme 2. For these systems, adding the higher order non-bonded terms does improve the SFM performance. The MAE improves by ~1 kcal/mol when the 3rd order non-bonded terms are included. Here again, the SFM errors obtained when using the EFP/*ab initio* non-bonded energies are similar (~1 kcal/mol *smaller*) to those obtained using only *ab initio* non-bonded terms. Table 4 compares the CPU times for using the EFP method for non-bonded terms with those required for the *ab initio*-only SFM. The time needed to generate the EFP terms is also listed. This time becomes significant when the 3rd order many body terms are included. Further, since the EFP generation requires only calculations at the Hartree-Fock level of theory, the contribution of the EFP generation to the overall computation time will greatly decrease in importance when more accurate electronic structure methods are used.

Nonetheless, as shown in Table 4, employing EFP to treat a portion of the SFM 3rd order non-bonded terms results in an overall decrease in CPU time by roughly a factor of two. Including only the 2nd order non-bonded EFP/*ab initio* terms gives energies in good agreement with the full un-fragmented energies (Table 3: MAE = 2.6 kcal/mol), but the gain in computational efficiency is small, ~5-15% less CPU time. The advantage of using the EFP/*ab initio* approach is clearly seen in Table 5, where the number of non-

bonded terms that must be computed *ab initio* is compared for the EFP/*ab initio* and the electrostatic/*ab initio* methods. Since the EFP method is more effective at capturing interaction energies than electrostatics at close range, the EFP non-bonded cutoff can be set to the shorter distance of 2.7Å, instead of 4.5Å. This shorter cutoff value reduces the number of expensive *ab initio* non-bonded terms by up to 85-90%, while still retaining good accuracy. This increase in efficiency will be especially important when high levels of theory, such as MP2 or coupled cluster methods, are employed to treat large molecular systems. A major advantage of the SFM (and other fragment-like methods) is that it enables very accurate calculations on large molecular systems that would otherwise be impossible. As noted above, since the EFP generation requires only Hartree-Fock level calculations, the contribution of the EFP generation to the overall computation time will greatly decrease in importance when more accurate electronic structure methods are used.

4. The Fragment Molecular Orbital (FMO) Method

The FMO method³⁹⁻⁴⁵ relies upon the assumption that electron exchange and charge transfer are largely local phenomena in chemical systems. By breaking a system into fragments and treating the long-range interactions in a system using only a Coulomb operator, there are significant reductions in computational expense. In addition to this initial reduction in computational cost, the FMO method is further enhanced with the generalized distributed data interface (GDDI).⁴⁹ The GDDI uses a two-level parallelization scheme, assigning individual fragment calculations to different groups, each group performing its fragment calculation in parallel. The FMO method has also been interfaced with the polarizable continuum model (PCM)⁶⁶ and the effective fragment potential (EFP)⁸⁴ for the inclusion of solvent effects. There is also a multi-layer

FMO (MFMO) implementation⁶⁷ that allows for the use of different wavefunction types for different fragments. The combination of the long-range approximations to the system and the GDDI parallelization helps to facilitate the treatment of large molecular systems.^{45,68}

Creating fragments in the FMO method involves breaking bonds electrostatically, assigning two electrons of a covalent bond to one fragment and none to the other, with the fragment choice relying on the chemical intuition of the user. To avoid the charged species created by such a fragmentation scheme, a proton from the electron deficient fragment is reassigned to the electron rich species, creating two neutral fragments (indicated by the “1” and “5” in Figure 4). The “1” and “5” in the figure *both* carry sp^3 hybrid orbitals, to maintain the carbon character. The individual fragment (monomer) calculations are performed in the presence of a Coulomb “bath” that represents the electrostatic potential (ESP) of the system (Figure 5). Significant improvements^{69,70} to this description of the FMO method have been obtained by including many-body effects. Two-body (dimers; called FMO2) and three-body (trimers; called FMO3) interactions of the monomers are fully quantum, therefore all interactions are included in each calculation.

To calculate the energy of a system within the FMO method, first the initial electron density distribution is calculated for each monomer in the Coulomb bath of the system. The monomer Fock operators are constructed and the energy of each monomer is calculated. Each of the monomer energies is iterated to self-consistency in this manner, leading to the convergence of the ESP.

The total energy of a chemical system, within the FMO approximation, can be written as;

$$\begin{aligned}
E = & \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J) \\
& + \sum_{I>J>K}^N \{ (E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) \\
& - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I) \} + \dots
\end{aligned} \tag{10}$$

where monomer (I), dimer (IJ) and trimer (IJK) energies are obtained using the standard SCF method. Despite the seeming simplicity of Eq. (10), the FMO method encapsulates the deeper ideas of properly handling many-body effects, as clarified in the diagrammatic treatment⁶⁹ and the Green's function formalism.⁷³ This is a very important distinction between the FMO and other methods. The Fock equation;

$$\tilde{\mathbf{F}}^x \mathbf{C}^x = \mathbf{S}^x \mathbf{C}^x \tilde{\mathcal{E}}^x \quad x = I, IJ, IJK \tag{11}$$

$$\tilde{\mathbf{F}}^x = \tilde{\mathbf{H}}^x + \mathbf{G}^x \tag{12}$$

is modified from the standard form with the addition of a term, $V_{\mu\nu}^x$, that represents the ESP to the one-electron Hamiltonian $\tilde{\mathbf{H}}^x$.

$$\tilde{H}_{\mu\nu}^x = H_{\mu\nu}^x + V_{\mu\nu}^x + B \sum_i \langle \mu | \phi_i^h \rangle \langle \phi_i^h | \nu \rangle \tag{13}$$

The modified Hamiltonian also contains the projection operator, $B \sum_i \langle \mu | \phi_i^h \rangle \langle \phi_i^h | \nu \rangle$, needed for division of basis functions along the fractioned bonds, where B is a constant chosen to be sufficiently large to remove the corresponding orbitals out of variational space (normally $B=10^6$ a.u.).

The ESP of the system takes the form;

$$V_{\mu\nu}^x = \sum_{K(\neq x)} (u_{\mu\nu}^K + v_{\mu\nu}^K) \tag{14}$$

$$u_{\mu\nu}^K = \sum_{A \in K} \langle \mu | (-Z_A / |\mathbf{r} - \mathbf{r}_A|) | \nu \rangle \quad (15)$$

$$v_{\mu\nu}^K = \sum_{\lambda\sigma \in K} D_{\lambda\sigma}^K(\mu\nu | \lambda\sigma) \quad (16)$$

where the first term $u_{\mu\nu}^K$ is the nuclear attraction contribution and the second term $v_{\mu\nu}^K$ is the two-electron contribution, both of which are calculated for each of the surrounding monomers K with electron density D^K .

4.1 FMO Approximations

The formulation of the energy described above has limitations^{44,70}, such as the increasing cost of the two-electron term in the ESP. To reduce this cost, different approximations can be used to treat the ESP by creating a cut-off value R_{app} . Outside this cut-off the two-electron terms of the ESP can be treated in a more approximate way. However, the foregoing energy formulation loses some accuracy with such approximations, because the balance among the approximations in different FMO terms may be lost. For example, if there are three monomers I , J and L with some distance based approximation (R_{app}) applied, and the relative distances are as illustrated in Figure 6, then the electrostatic interaction of monomers I and L would be treated using the approximation, while the interaction of monomers J and L would be treated with the full ESP. However, there would be an interaction of dimer IJ with monomer L without any approximations, (because L is close to IJ and J , but far from I). This causes a loss of balance among some of the dimer energy terms in the expression

$$\sum_{I>J}^N (E_{IJ} - E_I - E_J) \quad (17)$$

for those dimers IJ in which some ESP contributions (i.e., those for fragment L) included in E_I are treated using the approximation, but in others they are not. There can be a great many dimer contributions to the energy in a single calculation, causing significant loss of accuracy in the energy of the full system.

The issue described above requires a reformulation of the energy that is equivalent to Eq. (10), but more accurate if approximations to the ESP are used⁴⁴. For FMO2,

$$\begin{aligned} E_{mol} &= \sum_I E'_I + \sum_{I>J} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J} \{Tr(\mathbf{D}^{IJ}\mathbf{V}^{IJ}) - Tr(\mathbf{D}^I\mathbf{V}^I) - (\mathbf{D}^J\mathbf{V}^J)\} \\ &= \sum_I E'_I + \sum_{I>J} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J} Tr(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ}) \end{aligned} \quad (18)$$

A similar expression has been derived for FMO3.⁷⁸

The new energy terms E'_x are defined as the internal energies of the monomers and dimers with the ESP contributions subtracted out;

$$E'_x = E_x - Tr(\mathbf{D}^x\mathbf{V}^x) \quad x = I, J, IJ \quad (19)$$

This is accomplished by contracting $\mathbf{V}^x$ with the electron density $\mathbf{D}^x$. $\Delta\mathbf{D}^x$ is the difference density matrix, defined as

$$\Delta\mathbf{D}^{IJ} = \mathbf{D}^{IJ} - \mathbf{D}^I \oplus \mathbf{D}^J = \begin{pmatrix} \mathbf{d}^{II} & \mathbf{d}^{IJ} \\ \mathbf{d}^{JI} & \mathbf{d}^{JJ} \end{pmatrix} - \begin{pmatrix} \mathbf{d}^I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}^J \end{pmatrix} \quad (20)$$

where $\mathbf{d}^{II}$, $\mathbf{d}^{IJ}$, $\mathbf{d}^{JI}$ and $\mathbf{d}^{JJ}$ are blocks of $\mathbf{D}^{IJ}$, and $\mathbf{d}^I$ ($\mathbf{d}^J$) is simply equal to $\mathbf{D}^I$ ($\mathbf{D}^J$). This formulation makes it possible to calculate the total energy explicitly from only the dimer ESP $\mathbf{V}^{IJ}$. By subtracting the monomer and dimer ESPs in the energy expression, approximations can be applied to the monomers and dimers separately. The dimer ESP

then directly contributes to the total energy, while the monomer ESP determines the monomer electron densities, only contributing to the total energy indirectly.

Two different levels of approximation are currently used in the FMO method, enabled by equation (18). For intermediate distances the Mulliken approximation⁷¹ to the two-electron integrals is used. Equation (16) can then be rewritten as;

$$v_{\mu\nu}^K \cong \sum_{\lambda \in K} (\mathbf{D}^K \mathbf{S}^K)_{\lambda\lambda} (\mu\nu | \lambda\lambda) \quad (21)$$

This approximation reduces the computational cost of the two-electron integrals by a factor of N_B (number of basis functions).

The fractional point charge approximation, using the Mulliken atomic populations of the monomers, is used for long distances. The two-electron term of equation (16) is then simplified as;

$$v_{\mu\nu}^K \cong \sum_{A \in K} \langle \mu | (Q_A / |\mathbf{r} - \mathbf{r}_A|) | \nu \rangle \quad (22)$$

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

Inter-fragment interactions have a similar approximation that evaluates the electrostatic contribution to the energy using the monomer densities for far separated dimers, instead of calculating the dimer density itself. This contribution is added to the dimer energy as

$$E'_{IJ} \cong E'_I + E'_J + \text{Tr}(D^I u^{1,I(J)}) + \text{Tr}(D^J u^{1,J(I)}) + \sum_{\mu\nu \in I} \sum_{\rho\sigma \in J} D_{\mu\nu}^I D_{\rho\sigma}^J (\mu\nu | \rho\sigma) \quad (23)$$

where $u^{1,I(J)}$ and $u^{1,J(I)}$ are one-electron Coulomb potentials of the force exerted by fragment J on fragment I, and fragment I on fragment J, respectively.

Other approximations of the same nature are implemented for correlated dimers and trimers, where the corresponding corrections for far separated pairs and triples of

fragments are neglected.^{72,74} Formal definitions and descriptions of the trimer interactions and cut-offs used in FMO3 have been described previously^{69,70} and will not be discussed here. All of these approximations are based on a distance definition R_{app} , defined as the minimum distance between atoms in n -mer I and monomer J divided by the sum of their van der Waals radii.

There have been several new developments in the FMO theory that cannot be discussed in detail here. Nonetheless, it is useful to mention a few of them briefly. As an alternative to the original bond fragmentation scheme, in which the electron density describing the detached bonds is variationally optimized, a new scheme has been suggested in which this density is obtained for a model system and is kept frozen in fragment calculations⁷⁵. This new scheme has been shown to work well for covalent crystals such as zeolites. The FMO method has also recently been implemented for the study of excited states⁷⁶, using multi-configurational self-consistent field (MCSCF) theory, configuration interaction (CI), and time-dependent density functional theory (TDDFT).

4.2 FMO2 and FMO3 Calculations on (H₂O)₃₂ Clusters

The unusual characteristics of liquid water make it both very important to chemical processes and particularly difficult to model accurately. The structure of small (H₂O)_n ($n=6$ through 20) clusters have recently been determined⁸³ using coupled cluster theory; however, the ability to model water clusters larger than this at the same level of theory is nearly impossible. The FMO method provides a way to model much larger water clusters at high levels of theory such as CCSD(T), while keeping the computational cost manageable.

In the present work, calculations of the energies of $(\text{H}_2\text{O})_{32}$ water clusters are reported, using fully *ab initio* Møller-Plesset second order perturbation theory (MP2) as well as the MP2 implementation of the FMO method^{72,77}. For these clusters a fragment (monomer) is defined as one water molecule for both FMO2 and FMO3 calculations. Initial structures were obtained from EFP Monte Carlo/Simulated Annealing (MC/SA) simulations followed by EFP optimizations of a representative set of structures. The MC/SA method with local minimization was used to sample the configuration space. For each global minimum found, the number of structures sampled was on the order of 500,000 – 1,100,000. The number of steps taken for each temperature was varied (100, 500, 1000, 10,000), along with changing the number of steps between local minimizations (10, 100, 1,000). The number of fragments moved per step was also varied between one and five. The starting temperature for the simulated annealing varied from 500 to 20,000 K and the final temperature was kept at 300 K. This selection of isomers (the lowest energy structure is shown in Figure 7) was used to investigate the accuracy of the FMO method by comparing both absolute and relative energies.

Average errors for the FMO2-MP2 calculations (Table 6) using the 6-31++G(d,p) basis set are very consistent, around 12 kcal/mol. The FMO3-MP2 results illustrate the importance of three-body interactions in water clusters,^{78,79} again with very consistent errors of ~2-3 kcal/mol (Table 6). Comparing results between basis sets in Table 6, when the basis set size is increased to 6-311++G(3df,2p), the FMO2 errors double to ~24-28 kcal/mol, while the FMO3 errors are cut in half to ~1 kcal/mol. This increase in errors with an increased basis set size for the two-body FMO method could be due to an increased importance of three-body contributions when the better basis set is used. The

larger basis set also provides a better description of three-body interactions, making the lack of these interactions in FMO2 even more detrimental.

Despite the large absolute errors present in the FMO2 description of water clusters, the relative energetics of the different isomers is captured quite well. On average, the FMO2 relative energies are in agreement with full *ab initio* results to within ~1-2 kcal/mol with both basis sets, shown in Table 7. The error increases for FMO2 as the relative energy of the isomers increases, showing an increased importance of three body contributions with higher energy isomers. For both basis sets, the FMO3 results are within ~0.5 kcal/mol or less for all isomers as shown in Table 7, effectively removing the error from the two body description used in FMO2.

4.3 The FMO Method Applied to Ionic Liquids

Previous studies of ionic liquids^{80,81,82} have focused on the decomposition of ion pairs (Figure 8), providing insight into the chemistry of their ignition as high energy fuels. The focus of this paper, however, will be to accurately describe larger systems beyond single anion-cation pairs. Recent work by Li et. al.⁸³ has provided an accurate structure of two ion pairs (two cations and two anions), providing a greater understanding of the molecular structure and intermolecular interactions. The same system will be modeled here, along with systems of three ion pairs to illustrate the effectiveness of the FMO method in accurately describing complex molecular clusters, with the goal of modeling much larger systems in the future.

Two ionic liquid systems, 1-H,4-H-1,2,4-triazolium dinitramide and 1-amino, 4-H-1,2,4-triazolium dinitramide (Figure 8), were studied using both *ab initio* Møller-Plesset second order perturbation theory (MP2) and the MP2 implementation of the FMO

method^{72,77} with one ion chosen as an FMO fragment or monomer. Structures composed of two cations and two anions (tetramers), shown in Figures 9 and 11, and larger clusters of three cations and three anions (hexamers), shown in Figures 10 and 12, were optimized at the MP2/6-31+G(d) level of theory. FMO2-MP2 and FMO3-MP2 single point energy calculations were then performed for comparison with the fully *ab initio* results. Mulliken charges on each cation and anion were also compared to ensure that the pronounced charge separation present in ionic liquids^{80,81,82} is captured using the FMO method.

Comparing the energies from FMO2 and FMO3, it can be seen immediately that the FMO method captures the total energy very well, within 2 kcal/mol in the worst case (Table 8). For the tetramers, both FMO2 and FMO3 are in good agreement; the FMO2 errors are less than 1 kcal/mol relative to the fully *ab initio* results. For the hexamers, the FMO2 errors are less than 2 kcal/mol, illustrating that FMO3 is not required to achieve the desired level of accuracy for these particular ionic liquid systems. Whether this trend persists as system size grows beyond three ion pairs, or for other ion pairs, must be tested further.

As shown in previous studies^{80,81,82}, ionic liquid ion pairs have a definite separation of charge (approximately +1 on the cations, -1 on the anions) at equilibrium geometries. This charge separation is also observed for tetramers, as shown in Table 9, and the charge separation between cations and anions is still present up to hexamer structures. FMO2 captures the qualitative charge separation quite well, however, the magnitude of charge present on both cations and anions is slightly overestimated by FMO2 for both tetramer structures (Table 9). However as the system size increases to three ion pairs, the difference between FMO2, FMO3 and *ab initio* results becomes

minimal. Future work using larger basis sets will help determine if FMO2 is accurate enough to describe larger ionic liquid clusters or if FMO3 will be required.

Another consideration for larger molecular systems is the computer time required. To illustrate the overall effectiveness of the FMO method in both providing accurate results and reducing computational requirements, timings were performed for the ionic liquid systems described above. Due to the fact that FMO2 is in good agreement with the *ab initio* results, only timings for FMO2 will be shown. However, it is noted here that because the tetramers and hexamers examined here are small, the FMO3 timings for these systems do not exhibit any time savings relative to the full MP2 calculations. The benefit of using FMO3 is only seen with larger systems⁷³.

Timings were performed on a Cray XT4 supercomputer using AMD Opteron64 processors running at 2.1 GHz, located at the U.S. Army Engineer Research and Development Center (ERDC) in Vicksburg, Mississippi. Single point Møller-Plesset second order perturbation theory (MP2) energy calculations were performed using 8, 16 and 32 processors with both FMO2 and MP2 using the 6-31+G(d) basis set. As shown in Table 10, FMO2 requires approximately half the computer time of a full MP2 calculation on the tetramers. With the increase in available processors, the overall time requirements are cut in half for both FMO2 and MP2, showing good scalability for both methods. With an increase in system size from ionic liquid tetramers to hexamers, the computer time required for a fully *ab initio* calculation increases more than 6 fold, while the FMO2 requirement only doubles. So, the FMO2 cost savings relative to full MP2 is much greater than that observed for the tetramers. Again, scalability for both methods is very

good for the hexamers, cutting the computational time in half when doubling the number of available CPUs.

It is apparent that as the system size increases to larger ionic liquid clusters, or as the basis set size increases (or both) the computational requirements for a fully *ab initio* calculation will rapidly and increasingly exceed those for FMO2. It may be that as the system size increases, the importance of three-body contributions to the interaction energy will also increase, requiring the use of FMO3. Future work will determine the importance of three-body terms in ionic liquid systems, as well as the ability of the FMO method to describe larger molecular clusters.

5. Summary and Conclusions

Obtaining chemical accuracy (1 kcal/mol) using model chemistries has been a major focus of quantum chemistry research for the last quarter of a century. The desire to study larger systems in order to capture novel chemical phenomena (e.g. solvent effects, surface science, enzyme and heterogeneous catalysis and polymerization phenomena), including the kinetics and dynamics of such processes, often requires very accurate predictions of potential energy surfaces for subsequent predictions to be even qualitatively correct. The computational effort of traditional methods such as Hartree-Fock (HF), density functional theory (DFT), 2nd order perturbation theory (MP2), and coupled cluster theory with perturbative triples (CCSD(T)) scale as $O(n^4)$, $O(n^4)$, $O(n^5)$, and $O(n^7)$, respectively, where n represents the size of the system, e.g., the size of the basis set. In practice, this limits the sizes of systems that can be studied with HF/DFT, MP2 and CCSD(T) to approximately a few hundred, one hundred, and twenty non-hydrogen atoms, respectively. By developing highly parallel algorithms, the goal of using

sophisticated electronic structure methods to investigate large molecular problems becomes more feasible, especially if one has access to massively parallel computing hardware. However, scalability beyond hundreds to a few thousand processors is generally a serious bottleneck for correlated electronic structure methods. Consequently, parallel computing is not the sole solution to enabling accurate calculations on extended molecular systems; other approaches are needed. If one is interested in performing long-time simulations at reliable levels of theory, the situation is only exacerbated.

Pioneering work by Warshel^{9a} and others⁹ introduced hybrid methods that employ both quantum mechanics (QM) and molecular mechanics (MM), leading to the now ubiquitous QM/MM approach. Importantly, the QM/MM approach is quite general, so it can be employed with any level of QM, including the fragmentation methods that have been the primary focus of the present work. Modern fragmentation methods have their roots in ideas from Murrell (1955)^{46a} and Christoffersen (1971).^{46b} More recently developed fragmentation methods, such as the fragment molecular orbital (FMO) method and the systematic fragmentation method (SFM), are now becoming capable of achieving chemical accuracy for extended molecular systems.

The effective fragment potential (EFP)¹² method has been developed to model non-bonded, intermolecular interactions. There are two related implementations of the EFP method: The original method, called EFP1, was developed specifically to study aqueous solvent effects on chemical processes. The more recently developed EFP2 method is completely general, in the sense that an EFP2 is generated automatically by a simple GAMESS run, so it contains no empirically fitted parameters. The Coulomb and induction terms are common to EFP1 and EFP2, but the remaining terms in EFP2 are

derived from first principles. Once an EFP2 has been built for a specific system, the evaluation of EFP-EFP interactions requires a small fraction of the computational cost compared to the fully QM calculation. The EFP computational effort scales in the range of quadratic to linear with an increasing number of fragments. EFP1/MP2 achieves an accuracy of ~ 1 kcal/mol for the relative energies of six-water clusters compared to CCSD(T)/aug-cc-pVTZ.^{47b} For benzene dimer binding energies, EFP2 achieves an accuracy of ~ 1 kcal/mol relative to CCSD(T)/aug-cc-pVTZ results. The EFP1 method has been interfaced with the QM methods HF,¹³ DFT,¹⁴ MCSCF,¹⁵ singly excited configuration interaction (CIS),¹⁶ and time-dependent density functional theory (TDDFT)¹⁷ within the GAMESS suite, so EFP1 is a fully QM/MM method. The EFP2-QM integration is currently under development.¹⁸

The SFM fragments a molecule based on the number of single bonds in each fragment, while considering the environmental effects of distant parts of the system via a many body expansion of the interactions not captured by the internal energies of the fragments. This framework allows the SFM to be widely applicable with a simple user interface, which has been integrated into the GAMESS suite. The SFM has been used to study small and medium sized organic molecules³⁵, as well as crystals.³⁷ In this paper, it was demonstrated that the SFM, when using EFPs for the non-bonded interactions, has a mean averaged error of 1.8 kcal/mol for several α -helical isomers at the HF/6-31++G(d,p) level of theory. The SFM is independent of the *ab initio* methods used in calculations of the fragments, thereby facilitating highly accurate energies and relative energies with nearly linear scaling as the size of the system is increased. The time requirements for the EFP part of a SFM calculation, when EFPs are used for the non-

bonded interactions, are determined by the cost of an initial HF single point calculation. Therefore, the EFP fraction of the overall computer time requirements decreases rapidly as the level of *ab initio* theory increases (e.g., from HF to MP2 to CCSD(T)).

The FMO method treats each fragment (monomer, dimer, etc.) in a Coulomb bath that represents the remainder of the full system. The energy of each monomer is iterated to self-consistency within this Coulomb bath. The FMO method is very flexible with regard to the definition of fragments (i.e., monomers), the assignments of distance cutoff parameters, and the level of many-body effects (i.e., dimer, trimer, etc.) to be included in the calculation. Combined with the avoidance of capping procedures, this facilitates the study of a wide variety of systems including clusters, zeolites, and proteins, and the ability to balance accuracy and computational efficiency. Within the GAMESS suite, the FMO method has been interfaced with the polarizable continuum method and the EFP method for studies of solvent effects on chemical processes. Each monomer in a molecular system of interest can be treated by most traditional electronic structure methods. In the present work, the FMO method has been shown to achieve accuracy within 1 kcal/mol for both ionic liquid systems and water clusters.

The EFP method provides a systematically improvable description of non-bonded interactions, while the FMO method and the SFM facilitate the description of large molecular systems with high levels of accuracy. The interface of the two fragmentation methods for internal and near-field *ab initio* calculations with the EFP method for non-bonded moderate and far-field interactions and for solvent effects provides a powerful and computationally effective combination. Additionally, the ability of these methods to take advantage of the standard theoretical electronic structure framework allows their

capabilities to move forward with new advances in electron correlation, wavefunction description and basis set development.

An important advantage of the FMO and SFM approaches described here is their ability to take great advantage of massively parallel computers. Because the energy of each fragment can be computed essentially independently, each fragment calculation can be determined on a separate compute node. Further, because most of the algorithms used in GAMESS for electronic structure functionalities are themselves highly scalable, the fragment-based calculations can take advantage of multi-level parallelism. This capability, which is enhanced by middleware developments like the generalized distributed data interface (GDDI), bodes well for the implementation of algorithms for “petascale” computers that are expected to come on line within the next 2-3 years. Simultaneous advancements in new approaches like the fragmentation methods discussed here, novel parallel algorithms, *ab initio* theory, and novel approaches in hardware development are all required if one is to successfully address the grand challenge problems in the chemical sciences, biological sciences and materials science and engineering.

Acknowledgements. The research described in this work has been supported by grants from the Air Force Office of Scientific Research and the National Science Foundation. A grant of computer time from the Department of Defense High Performance Computing Modernization Program is gratefully acknowledged. The authors have benefited from informative discussions with Dr. Dmitri Fedorov (AIST) and Professor Michael Collins (Australian National University).

References

1. (a) G4: Curtiss L.A.; Redfern, P.C.; Raghavachari, K. *J.Chem. Phys.* **2007**, *127*, 124105. (b) G4: Curtiss L.A.; Redfern, P.C.; Raghavachari, K. *J.Chem. Phys.* **2007**, *126*, 084108. (c) G3SX: Curtiss L.A.; Raghavachari, K.; Redfern, P.C.; Pople, J. A. *J.Chem. Phys.* **2001**, *114*, 108. (d) G3: Curtiss L.A.; Raghavachari, K.; Redfern, P.C. *J.Chem. Phys.* **1998**, *109*, 7764. (e) G3(MP2): Curtiss L.A.; Redfern, P.C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703. (f) G2(MP2): Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293. (g) G2: Curtiss L.A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J.Chem. Phys.* **1991**, *94*, 7221. (h) G1: Curtiss L.A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J.Chem. Phys.* **1990**, *93*, 2537. (i) G1: Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss L.A. *J.Chem. Phys.* **1989**, *90*, 5622.
2. (a) W4: Karton, A.; Rabinovich, E.; Martin, J.M. L.; Ruscic, B. *J.Chem. Phys.* **2006**, *125*, 144108. (b) W3: Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J.M. L.; Kállay, M.; Gauss, J. *J.Chem. Phys.* **2004**, *120*, 4129. (c) W1&W2: Martin, J.M. L.; Oliveira, G. *J.Chem. Phys.* **1999**, *111*, 1843.
3. (a) ccCA: DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *124*, 114103. (b) ccCA: DeYonker, N. J.; Grimes, T.; Yockel, S.; Dinescu, A.; Mintz, B.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *125*, 104011. (c) ROCBS-QB3: Wood, G. P. F. ; Random, L. ; Petersson, G. A. ; Barnes, E. C.; Frisch, M. J.; Montgomery, J. A. *J. Chem. Phys.* **2006**, *125*, 094106. (d) CBS-4M: Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532. (e) CBS-QB3: Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822. (f) SAC: Gordon, M.; Truhlar, D. G. *J. Am. Chem. Soc.* **1986**, *108*, 5412. (g) SAC: Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *234*, 64. (h) SAC: Fast, P. L.; Corchado, J. C.; Sanchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 5129. (i) SEC: Corchado, J. C.; Truhlar, D. G. *ACS Symp. Ser.* **712**, SEC: Brown, F. B.; Truhlar, D. G. *Chem. Phys. Lett.* **1985**, *117*, 307. (j) HEAT: Szalay, P. G.; Tajti, A.; Stanton, J. F. *Mol. Phys.* **2005**, *103*, 2159. (k) HEAT: Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquez, J.; Stanton, J. F. *J. Chem. Phys.* **2004**, *121*, 11599. (l) Kedziora, G. S.; Pople, J. A.; Rassolov, V. A.; Ratner, M. A.; Redfern, P. C.; Curtiss, L.A. *J.Chem.Phys.* **1999**, *110*, 7123. (m) Cioslowski, J.; Schimeczek, M.; Liu, G; Stoyanov, V. *J. Chem. Phys.* **2000**, *113*, 9377. (o) Grimme, S. *J Chem. Acc.* **1992**, *81*, 405. (p) Ricca, A.; Bauschlicher, C. W. *J. Phys. Chem. A* **1998**, *102*, 876. (q) Chase, J. M. W.; NIST-JANAF Tables (4th ed.); J. Phys. Chem. Ref. Data Monogr. 9, 1 1998
4. (a) Almlöf, J.; Taylor, P.R. *J. Chem. Phys.* **1987**, *86*, 4070. (b) Almlöf, J.; Helgaker, T.; Taylor, P.R.; *ibid*, **1988**, *92*, 3029.
5. (a) Dunning, T. H; Jr., *J. Chem. Phys.* **1989**, *90*, 1007. (b) S. S. Xantheas; T. H. Dunning, Jr. *J. Phys. Chem.* **1993**, *97*, 18. (c) T. H. Dunning, Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244. (d) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; T. H. Dunning, Jr. *J. Chem. Phys.* **1999**, *110*, 7667. (e)

- Wilson, A. K.; Woon, D. E.; Peterson, K. A.; T. H. Dunning, Jr. *J. Abstr. Pap.-Am. Chem. Soc.* **1997**, 213, 60. (f) Kendall, R. A.; T. H. Dunning, Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, 96, 6796. (g) Woon, D. E.; T. H. Dunning, Jr. *J. Chem. Phys.* **1994**, 100, 2975. (h) Woon, D. E.; T. H. Dunning, Jr. *J. Chem. Phys.* **1995**, 103, 4572. (i) Peterson, K. A.; T. H. Dunning, Jr. *J. Chem. Phys.* **2002**, 117, 10548. (j) G3 Basis: Curtiss, L. A.; Redfern, P. C.; Rassolov, V.; Kedziora, G.; Pople, J. A. *J. Chem. Phys.* **2001**, 114, 9287.
6. Peterson, K.A.; Adler, T.B.; Werner, H.-J. *J. Chem. Phys.* **2008**, 128, 084102.
 7. (a) Silagen: Redondo, A.; Goddard, W. A. III. *J. Vac. Sci. Technol.* **1982**, 21, 344. (b) SM8: Cramer, C. J.; Truhlar, D. G. *Acct. Chem. Res.* **2008**, 41, 760. (c) COSMO: Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, 11, 799. (d) PCM: Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, 94, 2027. (e) GCOSMO: Truong, T.N.; Stefanovich, E.V. *Chem. Phys. Lett.* **1995**, 240, 253.
 8. (a) Centroid MD: Jang, S.; Voth, G. A. *J. Chem. Phys.* **1999**, 111, 2371. (b) Force Matching: Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, 120, 10896. (c) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, 123, 044505. (d) Kuwajima S.; Warshel, A. *J. Chem. Phys.* **1988**, 89, 3751. (e) Warshel; A.; Hwang, J-K. *J. Chem. Phys.* **1986**, 84, 4938.
 9. (a) 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.Revi. 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.* **263**, 393, 1996. (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*, Vol. 7; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, **1996**, 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.
 10. (a) Svensson, M.; Humbel, S.; Froese, R. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, 100, 19357. (b) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, 24, 760. (c) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. *J. Chem. Phys.* **2006**, 124, 024301.
 11. (a) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, 107, 3898. (b) Fast, P. L.; Corchado, J. C.; Sanchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, 103, 5129. (c) Fast, P. L.; Sanchez, M. L.; Corchado, J. C.; Truhlar, D. G. *J. Chem. Phys.* **1999**, 110, 11679. (d) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, 306, 407. (e) Tratz, C. M.; Fast, P. L.; Truhlar, D. G. *PhysChemComm* **1999**, 2, 14. (f) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, 112, 1125. (g) Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, 104, 6111. (h) Curtiss, L. A.; Redfern, P. C.; Rassolov, V.; Kedziora, G.; Pople, J. A. *J. Chem. Phys.* **2001**, 114, 9287. (i) Lee, T.-H.; Chen, H.-R.; Hu, W.-P. *Chem. Phys. Lett.* **2005**, 412, 430. (j) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, 7, 43. (k) Sun, Y.-L.; Lee, T.-H.;

- Chen, J.-L.; Wu, K.-J.; Hu, W.-P. *Chem. Phys. Lett.* **2007**, *442*, 220. (l) Chen, J.L.; Sun, Y.L.; Wu, K.J.; Hu, W.P. *J. Phys. Chem. A* **2008**, *112*, 1064.
12. (a) Gordon, M.S.; Slipchenko, L.; Hui, L.; Jensen, J.H. *Ann. Rep. Comp. Chem.* **2007**, *3*, 177. (b) Gordon, M.S.; Freitag, M.A.; Bandyopadhyay, P.; Kairys, V.; Jensen, J.H.; Stevens, W.J. *J. Phys. Chem.* **2001**, *105*, 293.
 13. 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.
 14. Adamovic, I.; Gordon, M.S. *J. Phys. Chem.* **2005**, *109*, 1629.
 15. Webb, S.P.; Gordon, M.S. *J. Phys. Chem. A* **1999**, *103*, 1265.
 16. Pooja, A.; Gordon, M. S.; in preparation
 17. Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S.; *J. Chem. Phys.* **2008**, *129*, 14112.
 18. (a) Kemp, D. D.; Gordon, M. S. manuscript in preparation. (b) Smith, T.; Slipchenko, L.; Ruedenberg, K.; Gordon, M. S. manuscript in preparation.
 19. (a) Stone, J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047. (b) Stone, A.J., *The Theory of Intermolecular Forces*; Oxford Press; 1996. (c) Stone, A. J. *J. Chem. Theory Comp.* **2005**, *1*, 1128.
 20. (a) Edmiston C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457. (b) Raffenetti, R. C.; Ruedenberg, K.; Janssen, C. L.; Schaefer, H. F. III. *Theor. Chim. Acta* **1993**, *86*, 149.
 21. (a) Slipchenko, L.; Gordon, M.S. *J. Comp. Chem.* **2006**, *28*, 276. (b) Freitag, M.A.; Gordon, M.S.; Jensen, J.H.; Stevens W.J. *J. Chem. Phys.* **2000**, *112*, 7300.
 22. (a) Slipchenko, L.; Gordon, M. S.; in preparation. (b) Murrell, J. N.; Teixeira-Dias, J. J. C. *Mol. Phys.* **1970**, *19*, 521.
 23. (a) Jensen, J.H.; Gordon, M.S. *Mol. Phys.* **1996**, *89*, 1313. (b) Murrell, J. N. *Proc. Roy. Soc. (Lond.)* **1965**, *A284*, 566. (c) Murrell J. N. *J. chem. Phys.* **1967**, *47*, 4916.
 24. (a) Adamovic, I.; Gordon, M.S. *Mol. Phys.* **2005**, *103*, 379. (b) Amos, R.D.; Handy, N.C.; Knowles, P.J.; Rice, J.E.; Stone, A.J. *J. Phys. Chem.* **1985**, *89*, 2186. (c) Piecuch, P. *Mol. Phys.* **1986**, *59*, 1085.
 25. Tang, K.T.; Toennies, J.P. *J. Chem. Phys.* **1984**, *80*, 3726.
 26. Li, H.; Gordon, M.S.; Jensen, J.H. *J. Chem. Phys.* **2006**, *124*, 214107.
 27. Li, H.; Gordon, M.S., *Theor. Chem. Accts.* **2006**, *115*, 385.
 28. (a) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599. (b) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039. (c) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293. (d) Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 839. (e) Zhang, D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 1145. (f) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 11386 (g) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 044903. (h) Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Comput. Chem.* **2004**, *25*, 1431. (i) Chen, X.; Zhang, Y.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 184105. (j) Zhang, D. W.; Zhang, J. Z. H. *Int. J. Quantum Chem.* **2005**, *103*, 246. (k) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Phys. Chem. A* **2005**, *109*, 2. (l) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7251. (m) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112. (n) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 031103. (o) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *124*, 184703. (p) Mei,

- Y.; Ji, C.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 094906. (q) Mei, Y.; Wu, E. L.; Han, K. L.; Zhang, J. Z. H. *Int. J. Quantum Chem.* **2006**, *106*, 1267. (r) Mei, Y.; He, X.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *Proteins- Struct. Funct. Bioinform.* **2005**, *59*, 489. (s) He, X.; Mei, Y.; Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *Proteins- Struct. Funct. Bioinform.* **2005**, *61*, 423.
29. (a) Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419. (b) Aoki, Y.; Suhai, S.; Imamura, A. *Int. J. Quantum Chem.* **1994**, *52*, 267. (c) Aoki, Y.; Suhai, S.; Imamura, A. *J. Chem. Phys.* **1994**, *101*, 10808. (d) Mitani, M.; Aoki, Y.; Imamura, A. *Int. J. Quantum Chem.* **1995**, *54*, 167; (e) Mitani, M.; Aoki, Y.; Imamura, A. *Int. J. Quantum Chem.* **1997**, *64*, 301. (f) Rather, G.; Aoki, Y.; Imamura, A. *Int. J. Quantum Chem.* **1999**, *74*, 35. (g) Gu, F. L.; Aoki, Y.; Imamura, A.; Bishop, D. M.; Kirtman, B. *Mol. Phys.* **2003**, *101*, 1487. (h) Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. *J. Chem. Phys.* **2004**, *121*, 10385. (i) Korchowiec, J.; Gu, F. L.; Imamura, A.; Kirtman, B.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *102*, 785. (j) Korchowiec, J.; Gu, F. L.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *105*, 875. (k) Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. *J. Comput. Chem.* **2006**, *27*, 1603. (l) Orimoto, Y.; Aoki, Y. *J. Polym. Sci. B: Polym. Phys.* **2006**, *44*, 119.
30. (a) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C.; *J. Phys. Chem.* **1994**, *98*, 9165. (b) Babu, K.; Gadre, S. R.; *J. Comput. Chem.* **2003**, *24*, 484. (c) Babu, K.; Ganesh, V.; Gadre, S. R.; Ghermani, N. E.; *Theor. Chem. Acc.* **2004**, *111*, 255. (d) Gadre, S. R.; Ganesh, V.; *J. Theor. Comp. Chem.* **2006**, *5*, 835.
31. (a) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255. (b) Kamiya, M.; Hirata, S.; Valiev, M. *J. Chem. Phys.* **2008**, *128*, 074103. (c) Windus, T.L.; Bylaska, E.J.; Dupuis, M.; Hirata, S.; Pollack, L.; Smith, D.M.; Straatsma, T.P.; Aprà, E. In *Lecture Notes in Computer Science*, vol. 2660, P.M.A. Sloot, D. Abramson, A. Bogdanov, J.J. Dongarra, A. Zomaya, Y. Gorbachev (Eds), Springer, Berlin (2003). (d) Hirata, S. *J. phys. Chem. A* **2003**, *107*, 9887. (e) Hirata, S. *J. Chem. Phys.* **2004**, *121*, 51. (f) Hirata, S.; Zhan, C.-G.; Aprà, E.; Windus, T.L.; Dixon, D.A. *J. Phys. Chem. A* **2003**, *107*, 10154.
32. (a) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46. (b) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *4*, 1.
33. (a) Tschumper, G. S.; *Chem. Phys. Lett.* **2006**, *427*, 185. (b) Hopkins, B. W.; Tschumper, G. S. *Mol. Phys.* **2005**, *103*, 309. (c) Hopkins, B. W.; Tschumper, G. S. *Chem. Phys. Lett.* **2005**, *407*, 362.
34. Deev, V.; Collins, M.A. *J. Chem. Phys.* **2005**, *122*, 154102.
35. Collins, M A; Deev V A. *J. Chem. Phys.* **2006**, *125*, 104104.
36. Collins, M.A. *J. Chem. Phys.* **2007**, *127*, 024104.
37. Netzloff, HM; Collins, M.A. *J. Chem. Phys.* **2007**, *127*, 134113.
38. Mullin, J.M.; Collins, M.A.; Gordon, M.S. In preparation.
39. Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319.
40. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

41. Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614
42. Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
43. Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. *Chem. Phys. Lett.* **2002**, *364*, 139.
44. Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
45. Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.
46. (a) Christoffersen, R.; Genson, D. W.; Maggiora, G. M.; *J. Chem. Phys.* **1971**, *54*, 239. (b) Longuet-Higgins, H. C.; Murrell, J. N.; *Proc Phys. Soc.* **1955**, *A68*, 601.
47. (a) Curtis, L.; Janssen, I.; Nielsen, M. B. *Parallel Computing in Quantum Chemistry*, CRC press, 2008. (b) Olson, R.M.; Bentz, J. L.; Kendall, R. A.; Schmidt, M. W.; Gordon, M.S. *J. Chem. Theory Comput.* **2007**, *3*, 1312. (c) Bentz, J. L.; Olson, R. M.; Gordon, M. S.; Schmidt, M. W.; Kendall, R. A. *Comp. Phys. Com.* **2007**, *176*, 589. (d) Dudley, T.J.; Olson, R.M.; Schmidt, M.W.; Gordon, M.S. *J. Comp. Chem.* **2006**, *27*, 353. (e) Gordon, M.S.; Ruedenberg, K.; Schmidt, M.W.; Bytautas, L.; Dudley, T.J.; Nagata, T.; Olson, R.; Varganov, S. *J. Phys. Conf. Ser.* **2006**, *46*, 229. (f) Janowski, T.; Ford, A. R.; Pulay, P. *J. Chem. Theory Comput.* **2007**, *3*, 1368.
48. (a) Fletcher, G.D.; Schmidt, M.W.; Gordon, M.S. *Adv. Chem. Phys.* **1999**, *110*, 267. (b) Schmidt, M.W.; Fletcher, G.D.; Bode, B.M.; Gordon, M.S. *Comp. Phys. Comm.* **2000**, *128*, 190. (c) Olson, R.M.; Schmidt, M.W.; Gordon, M.S.; Rendell, A.P.; *Proc. Supercomputing 2004*. (d) Nieplocha, J.; Harrison, R. J.; Littlefield, R. *Proc. Supercomputing 1994*; IEEE Computer Society Press: Washington, D.C., **1994**, pp 340–346. (e) Nieplocha, J.; Palmer, B.; Tipparaju, V.; Manojkumar, K.; Trease, H.; Aprà, E. *Int. J. High Perform. Comput. Appl.* **2006**, *20*, 203.
49. Fedorov, D.G.; Olson, R.M.; Kitaura, K.; Gordon, M.S.; Koseki, S. *J. Comp. Chem.* **2004**, *25*, 872.
50. (a) Kemp, D.D. Gordon, M.S. *J. Phys. Chem A* **2005**, *109*, 7688. (b) Aikens, C.M.; Gordon, M.S. *J. Am. Chem. Soc.*, **2006**, *128*, 12835. (c) Adamovic, I.; Gordon, M.S. *J. Phys. Chem, A* **2006**, *110*, 10267. (d) Adamovic, I.; Gordon, M.S. *J. Phys. Chem.* **2005**, *109*, 1629. (e) Freitag, M.A.; Hillman, B.; Gordon, M.S. *J. Chem. Phys.* **2004**, *120*, 1197. (f) Bandyopadhyay, P.; Mennucci, B.; Tomasi, J.; Gordon, M.S. *J. Chem. Phys.* **2002**, *116*, 5023. (g) Webb, S.P.; Gordon, M.S. *J. Phys. Chem. A* **1999**, *103*, 1265. (h) Mullin, J; Gordon, M.S. (in preparation)
51. (a) Netzloff, H.M.; Gordon, M.S. *J. Chem. Phys.* **2004**, *121*, 2711. (b) Li, H.; Netzloff, H.M.; Gordon, M.S. *J. Chem. Phys.* **2006**, *125*, 194103.
52. (a) Hehre, W.J.; Ditchfield, R.; Pople, J.A. *J. Chem. Phys.* **1972**, *56*, 2257 (b) Francl, M.M.; Pietro, W.J.; Hehre, W.J.; Binkley, J.S.; Gordon, M.S.; DeFrees, D.J.; Pople, J.A. *J. Chem. Phys.* **1982**, *77*, 3654. (c) Hariharan, P.C.; Pople, J.A. *Theoret. Chim. Acta* **1973**, *28*, 213.
53. (a) Hariharan, P. C.; Pople, J. A. *Theoret. Chim. Acta* **1973**, *28*, 213. (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. (c)

- Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comp. Chem.* **1983**, *4*, 294.
54. (a) Schmidt, M. W.; Baldrige, 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, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347. (b) Gordon, M.S.; Schmidt, M.W., *Theory and Applications of Computational Chemistry, the first forty years*; Elsevier; Amsterdam; 2005
55. (a) Hobza, P., Selzle, H.L., Schlag, E.W. *J. Am. Chem. Soc.* **1994**, *116*, 3500. (b) Hobza, P., Selzle, H.L., Schlag, E.W. *J. Phys. Chem.* **1996**, *100*, 18790. (c) Jaffe, R.L., Smith, G.D. *J. Chem. Phys.* **1996**, *105*, 2780. (d) Park, Y.C., Lee, J.S. *J. Phys. Chem. A* **2006**, *110*, 5091. (e) Spirko, V., Engvist, O., Soldan, P., Selzle, H.L., Schlag, E.W., Hobza, P. *J. Chem. Phys.* **1999**, *111*, 572. (f) Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 10. (g) Tsuzuki, S., Uchimaru, T., Sugawara, K., Mikami, M. *J. Chem. Phys.* **2002**, *117*, 11216. (h) Bornsen, K.O., Selzle, H.L., Schlag, E.W. *J. Chem. Phys.* **1986**, *85*, 1726. Felker, P.M., Maxton, P.M., Schaeffer, M.W. *Chem. Rev.* **1994**, *94*, 1787. (i) Janda, K.C., Hemminger, J.C., Winn, J.S., Novick, S.E., Harris, S.J., Klemperer, W. *J. Chem. Phys.* **1975**, *63*, 1419. (j) Law, K.S., Schauer, M., Bernstein, E.R. *J. Chem. Phys.* **1984**, *81*, 4871. (k) Scherzer, W., Kratzschmar, O., Selzle, H.L., Schlag, E.W. *Zeitschrift Fur Naturforschung Section Journal of Physical Sciences* **1992**, *47*, 1248. (l) Steed, J.M., Dixon, T.A., Klemperer, W. *J. Chem. Phys.* **1979**, *70*, 4940. (m) Puzder, A., Dion, M., Langreth, D.C. *J. Chem. Phys.* **2006**, *124*, 1674105. (n) Sato, T., Tsuneda, T., Hirao, K. *J. of Chem. Phys.* **2005**, *123*, 104307.
56. Sinnokrot, M.O.; Sherrill, C.D. *J. Phys. Chem. A* **2004**, *108*, 10200.
57. Sinnokrot, M.O.; Valeev, E.F.; Sherrill, C.D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.
58. Arunan, E.; Gutowsky, H.S. *J. Chem. Phys.* **1993**, *98*, 4294.
59. Grover, J.R.; Walters, E.A.; Hui, E.T. *J. Phys. Chem.* **1987**, *91*, 3233.
60. Krause, H.; Ernstberger, B.; Neusser, H.J. *Chem. Phys. Lett.* **1991**, *184*, 411.
61. Jung, Y.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2831.
62. Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.
63. Paldus, J., Wilson, S., editor. *Handbook of Molecular Physics and Quantum Chemistry*, vol. 2. Chichester: Wiley; 2000, p. 272–313.
64. Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
65. Allen, F.H. *Acta Cryst.* **2002**, *B58*, 380.
66. Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. *J. Comp. Chem.* **2006**, *27*, 976.
67. Fedorov, D.G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
68. Mochizuki, Y.; Yamashita, K.; Murase, T.; Nakano, T.; Fukuzawa, K.; Takematsu, K.; Watanabe, H.; Tanaka, S. *Chem. Phys. Lett.* **2008**, *457*, 396.
69. Fedorov, D. G., Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
70. Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.
71. a) Mulliken, R.S. *J. Chim. Phys.* **1949**, *46*, 500. b) Mulliken, R.S. *J. Chim. Phys.* **1949**, *46*, 521.
72. Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.

73. Yasuda, Y.; Yamaki, K. *J. Chem. Phys.* **2006**, *125*, 154101.
74. Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
75. D. G. Fedorov, D.G.; Jensen, J.H.; Deka, R.C.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 11808.
76. (a) Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2005**, *122*, 054108; (b) Mochizuki, Y; Koikegami, S; Amari, S; Segawa, K; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.*, **2005**, *406*, 283; (c) Mochizuki, Y; Tanaka, K; Yamashita, K; Ishikawa, T; Nakano, T; Amari, S; Segawa, K; Murase, K; Tokiwa, H; Sakurai, M. *Theor. Chem. Acc.*, **2007**, *117*, 541; (d) Chiba, M; Fedorov, D.G.; Kitaura, K. *Chem. Phys. Lett.* **2007**, *444*, 346; (e) Chiba, M.; Fedorov, D.G.; Kitaura, K. *J. Chem. Phys.* **2007**, *127*, 104108; (f) Chiba, M.; Fedorov, D.G.; Kitaura, K. *J. Comp. Chem.* **2008**, *29*, 2667.
77. Fedorov, D. G.; Kitaura, K. *J. Comp. Chem.* **2007**, *28*, 222.
78. Fedorov, D.G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, *433*, 182.
79. Xantheas, S. *Struct. Bond.* **2005**, *116*, 119
80. Schmidt, M.W., Gordon, M.S., Boatz, J.A. *J. Phys. Chem. A* **2005**, *109*, 7285.
81. Zorn, D.D., Boatz, J.A., Gordon, M.S. *J. Phys. Chem.* **2006**, *B110*, 11110.
82. Pimienta, I.S.O., Elzey, S., Boatz, J.A., Gordon, M.S. *J. Phys. Chem. A* **2007**, *111*, 691.
83. Li, H., Boatz, J.A., Gordon, M.S. *J. Am. Chem. Soc.* in press.
84. T. Nagata and M.S. Gordon, manuscript in preparation.

Table 1. Binding energies (kcal/mol) and equilibrium separations R (Å) of benzene dimer structures.

Method	Basis set	Sandwich		T-Shaped		Parallel-displaced		
		R	Energy	R	Energy	R ₁	R ₂	Energy
MP2 ^a	aug-cc-pVDZ ^b	3.8	-2.83	5.0	-3.00	3.4	1.6	-4.12
	aug-cc-pVTZ	3.7	-3.25	4.9	-3.44	3.4	1.6	-4.65
	aug-cc-pVQZ ^b	3.7	-3.35	4.9	-3.48	3.4	1.6	-4.73
CCSD(T) ^a	aug-cc-pVDZ ^b	4.0	-1.33	5.1	-2.24	3.6	1.8	-2.22
	aug-cc-pVQZ ^b	3.9	-1.70	5.0	-2.61	3.6	1.6	-2.63
EFP	6-311++G(3df,2p)	4.0	-2.11	5.2	-2.50	3.8	1.2	-2.34

^a Reference 56

^b Basis sets as described in Ref. 56

Table 2. Mean absolute errors of isomerization energies (kcal/mol) calculated by SFM, relative to fully *ab initio* (no SFM) energies. The non-bonded terms use the combined EFP/*ab initio* approximation (cutoff $\geq 2.7\text{\AA}$). Given in parentheses is SFM with the non-bonded term fully *ab initio* (no approximation).

Isomer	2 nd Order Many Body 6-31G(d,p)		3rd Order Many Body 6-31G(d,p)	
	HF kcal/mol	MP2 kcal/mol	HF kcal/mol	MP2 kcal/mol
ODETAS-AHALUQ	0.0 (0.5)	0.6 (0.1)	0.6 (0.2)	0.4 (0.3)
ODETAS01-AHALUQ	0.5 (1.2)	0.0 (0.7)	0.9 (0.4)	0.0 (0.1)
BAZGEP-BAZGIT	0.4 (0.6)	0.4 (0.2)	0.2 (0.5)	0.3 (0.3)
BELDIF-NOTGAE	2.6 (2.6)	4.6 (5.1)	2.6 (2.4)	4.6 (4.9)
FDOURD01-BOFWIC	0.5 (0.7)	0.2 (0.9)	0.1 (0.2)	0.5 (1.6)
CONBAI-FDMUPD10	0.7 (0.3)	1.1 (2.4)	3.5 (4.2)	2.9 (5.4)
IDUFES-IDUFAO	1.6 (2.1)	3.1 (5.3)	0.7 (0.2)	1.8 (1.1)
LEDRAI-LEDRI	1.0 (0.8)	1.0 (2.6)	1.7 (1.9)	1.6 (2.8)
LEDRI-LEDRI	1.9 (1.5)	0.3 (0.0)	0.5 (0.4)	1.4 (1.0)
TAXYIA-MOGQOO	1.3 (0.4)	11.1 (8.4)	0.0 (1.2)	9.7 (6.8)
TAXYOG-MOGQOO	1.6 (2.1)	1.1 (3.7)	2.0 (2.6)	1.5 (4.2)
WINXIA-XEXXIH	0.3 (0.3)	0.1 (0.1)	0.3 (0.3)	0.1 (0.1)
MAE	1.0 (1.1)	2.0 (2.5)	1.1 (1.2)	2.1 (2.4)

Table 3. Absolute errors in isomerization energies (kcal/mol) at HF/6-31++G(d,p) for alpha helixes, relative to fully *ab initio* (no SFM). The non-bonded terms use the combined EFP/*ab initio* approximation (cutoff $\geq 2.7\text{\AA}$) or *ab initio* (given in parentheses).

Isomer	2nd Order	3rd Order
	HF/6-31++G(d,p)	HF/6-31++G(d,p)
	HF kcal/mol	HF kcal/mol
MAQWUW_1-MAQWUW_2	2.7 (1.1)	1.7 (0.2)
WUYCUO-WUYDAV	5.7 (5.8)	3.9 (6.1)
WUYCUO-WUYDEX	0.9 (2.5)	0.1 (3.1)
YETPES_1-YETPES_2	1.1 (5.3)	0.2 (1.0)
MAE	2.6 (3.7)	1.5 (2.6)

Table 4. Net CPU times (minutes) for the SFM HF/6-31++G(d,p) on a single core of a Xeon 2.66Ghz quad core Cloverton node, with 16 GB RAM. Net times include the time needed for EFP generation. The EFP generation time is given in parentheses. The total number of non-bonded terms is also listed. The heading EFP indicates the use of EFP for the non-bonded terms.

Isomer	2 nd Order Non-bonded			3 rd Order Non-bonded		
	# non-bonded terms	EFP	No EFP	# non-bonded terms	EFP	No EFP
MAQWUW_1	1113	128 (25)	144	3159	329 (191)	559
MAQWUW_2	1113	128 (26)	140	3155	333 (194)	562
WUYCUO	1752	155 (33)	182	5049	413 (214)	853
WUYDAV	1754	162 (33)	188	5059	439 (227)	909
WUYDEX	1754	150 (32)	180	5052	419 (217)	880
YETPES_1	932	115 (24)	122	2623	305 (170)	490
YETPES_2	929	117 (25)	126	2618	299 (166)	497

Table 5. Comparison of the number of *ab initio* non-bonded terms needed for non-bonded EFP/*ab initio* cutoffs set to 2.7 and 4.5Å at the 2nd and 3rd order many body approximation.

	2 nd Order Many Body		3 rd Order Many Body	
	2.7Å (terms)	4.5Å (terms)	2.7Å (terms)	4.5Å (terms)
MAQWUW_1	34	233	113	729
MAQWUW_2	34	224	106	693
WUYCUO	35	318	108	1054
WUYDAV	40	327	130	1075
WUYDEX	36	321	118	1055
YETPES_1	29	225	79	709
YETPES_2	25	225	70	708

Table 6. Absolute errors in the FMO2-MP2 and FMO3-MP2 total energy of the 32 water clusters selected from EFP Monte Carlo/Simulated Annealing simulations. Isomer names are only used to distinguish isomers from one another.

Isomer ^a	Absolute Error (kcal/mol)			
	6-31++G(d,p)		6-311++G(3df,2p)	
	FMO2-MP2	FMO3-MP2	FMO2-MP2	FMO3-MP2
32_1	11.8	2.2	26.8	1.0
32_2	12.4	2.5	28.0	1.2
32_3	11.4	1.9	27.3	1.3
32ab	12.5	2.5	27.4	1.3
32ad	11.8	2.5	27.3	1.2
32h	12.3	2.5	27.3	1.3
32o	12.5	2.5	25.8	1.0
32z	12.4	2.3	24.6	1.2

^aOne water molecule chosen as a monomer

Table 7. Relative FMO2-MP2 and FMO3-MP2 energies of the 32 water clusters selected from EFP Monte Carlo/Simulated Annealing simulations. Isomer names are only used to distinguish isomers from one another.

Isomer ^a	Relative Energies (kcal/mol)					
	6-31++G(d,p)			6-311++G(3df,2p)		
	FMO2-MP2	FMO3-MP2	<i>ab initio</i>	FMO2-MP2	FMO3-MP2	<i>ab initio</i>
32_1	0.0	0.0	0.0	0.0	0.0	0.0
32z	1.3	0.5	0.2	0.7	0.2	0.1
32_2	1.4	1.2	0.9	1.1	0.8	0.5
32ab	1.5	1.2	0.9	1.1	0.9	0.6
32h	1.4	1.2	0.9	1.4	1.0	0.7
32o	1.7	1.5	1.2	1.7	1.3	1.0
32ad	4.9	6.0	6.0	5.7	6.1	5.8
32_3	11.8	14.3	14.1	14.0	14.1	14.4

^aOne water molecule chosen as a monomer

Table 8. FMO2 errors (kcal/mol) for tetramer and hexamer ionic liquid clusters.

	Absolute Error (kcal/mol)	
	6-31+G(d)	
Tetramers	FMO2-MP2	FMO3-MP2
1-H,4-H-1,2,4-triazolium dinitramide	0.06	0.02
1-amino,4-H-1,2,4-triazolium dinitramide	0.69	0.03
Hexamers		
1-H,4-H-1,2,4-triazolium dinitramide	0.32	0.07
1-amino,4-H-1,2,4-triazolium dinitramide	1.35	0.27

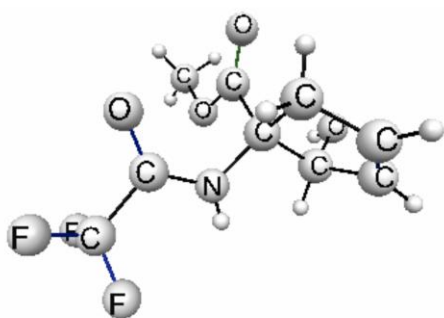
Table 9. Comparison of Mulliken charges for all ionic liquid systems investigated.

		Mulliken Charges 6-31+G(d)		
Tetramers		FMO2-MP2	FMO3-MP2	MP2
1-H,4-H-1,2,4-triazolium dinitramide	Cation 1	0.82	0.77	0.74
	Cation 2	0.82	0.77	0.74
	Anion 1	-0.82	-0.77	-0.74
	Anion 2	-0.82	-0.77	-0.74
1-amino,4-H-1,2,4-triazolium dinitramide	Cation 1	0.87	0.84	0.82
	Cation 2	0.82	0.82	0.82
	Anion 1	-0.89	-0.94	-0.93
	Anion 2	-0.80	-0.74	-0.71
Hexamers				
1-H,4-H-1,2,4-triazolium dinitramide	Cation 1	0.86	0.82	0.79
	Cation 2	0.88	0.85	0.80
	Cation 3	0.94	0.91	0.87
	Anion 1	-0.83	-0.79	-0.77
	Anion 2	-0.95	-0.92	-0.88
	Anion 3	-0.90	-0.86	-0.81
1-amino,4-H-1,2,4-triazolium dinitramide	Cation 1	0.84	0.84	0.88
	Cation 2	0.79	0.78	0.76
	Cation 3	0.90	0.89	0.89
	Anion 1	-0.81	-0.92	-0.95
	Anion 2	-0.83	-0.85	-0.83
	Anion 3	-0.88	-0.74	-0.75

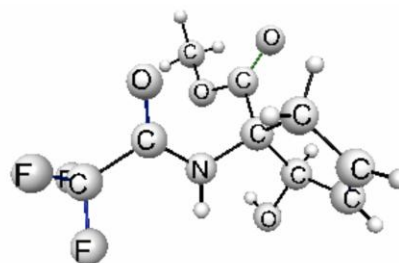
Table 10. Timings for ionic liquid clusters performed on a Cray XT4 with 2.1GHz AMD Opteron64 processors. Each node contains a 4 core CPU and 8 GB of RAM.

Tetramer	# CPUs	Timing (minutes)	
		6-31+G(d)	
		FMO2-MP2	MP2
1-amino,4-H-1,2,4-triazolium dinitramide	8	12.2	28.4
	16	6.4	14.7
	32	3.5	7.3
Hexamer			
1-amino,4-H-1,2,4-triazolium dinitramide	8	24.0	172.1
	16	12.5	83.9
	32	6.8	42.8

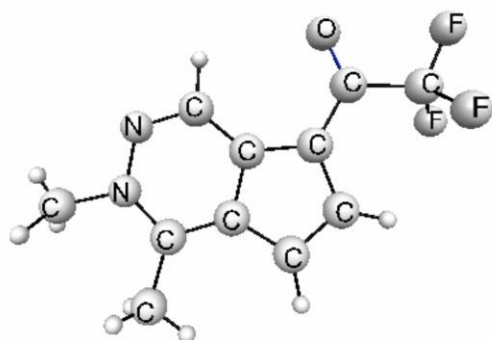
Scheme 1. Depiction of isomers used in Table 2. Structures are from the Cambridge Structural Database (CSD). Non-hydrogen atoms have been labeled.



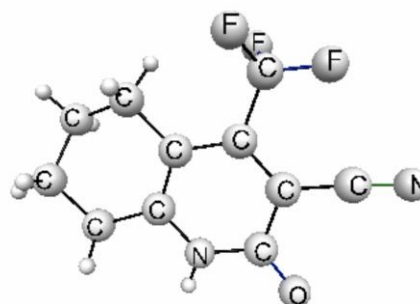
BAZGEP



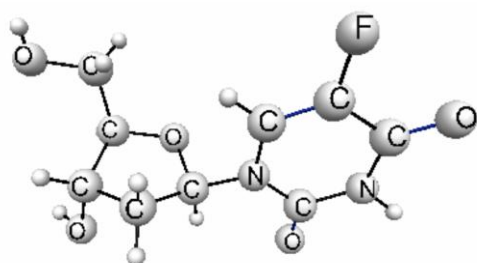
BAZGIT



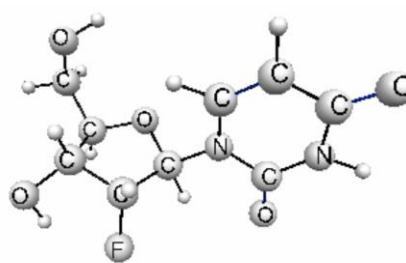
BELDIF



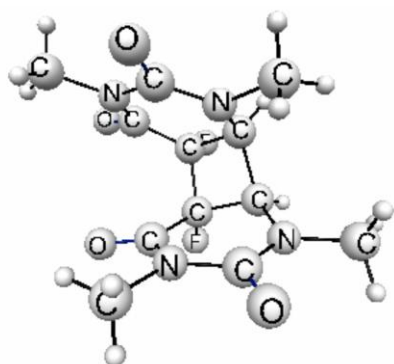
NOTGAE



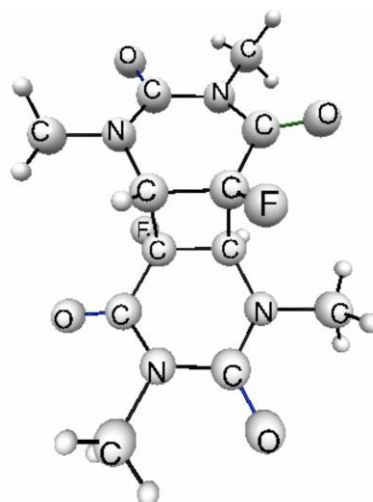
FDOURD01



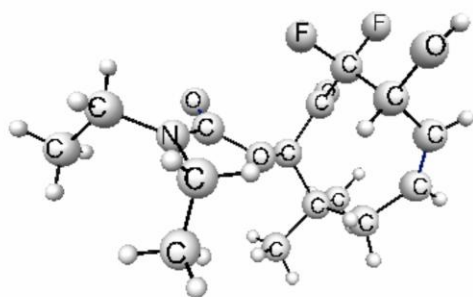
BOFWIC



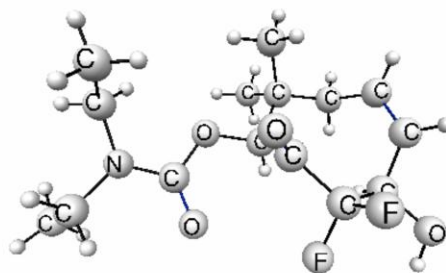
CONBAI



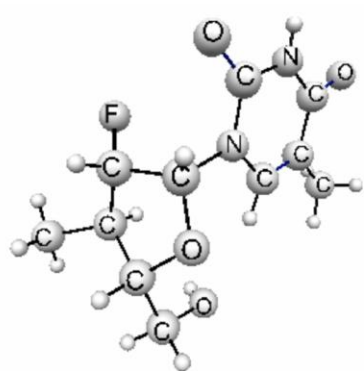
FDMUPDI0



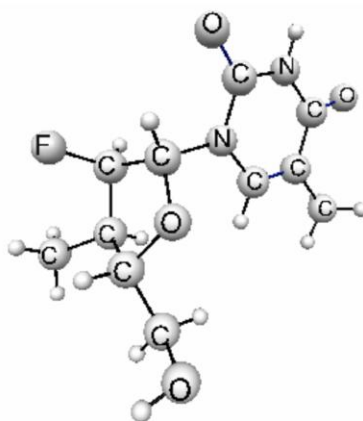
IDUFES



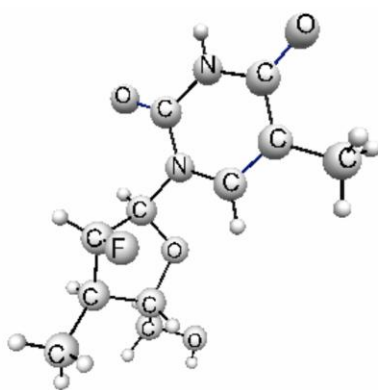
IDUFAO



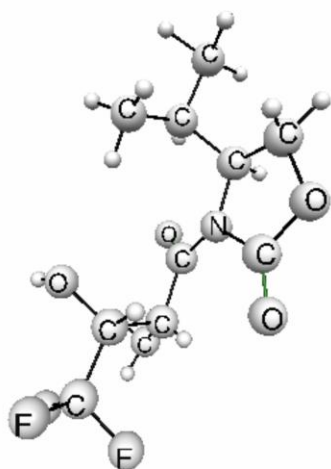
LEDRAN



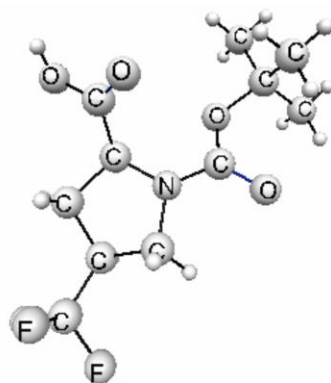
LEDRER



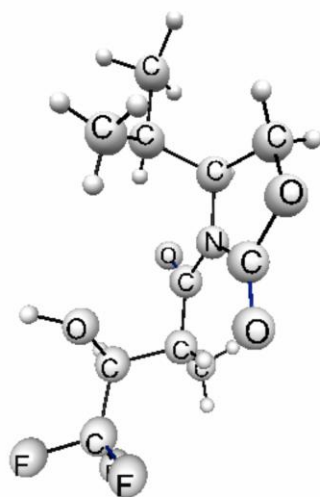
LEDRIV



TAXYIA

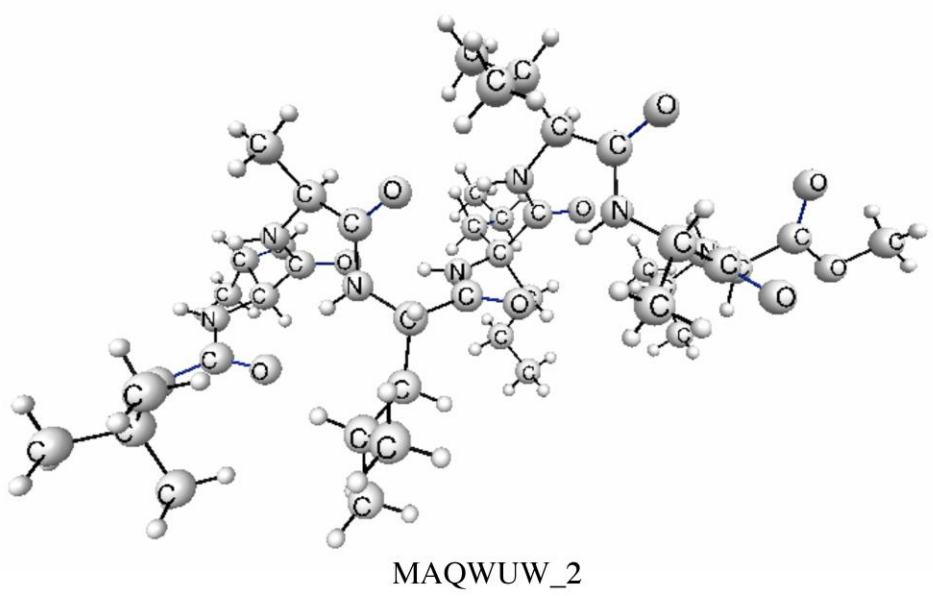
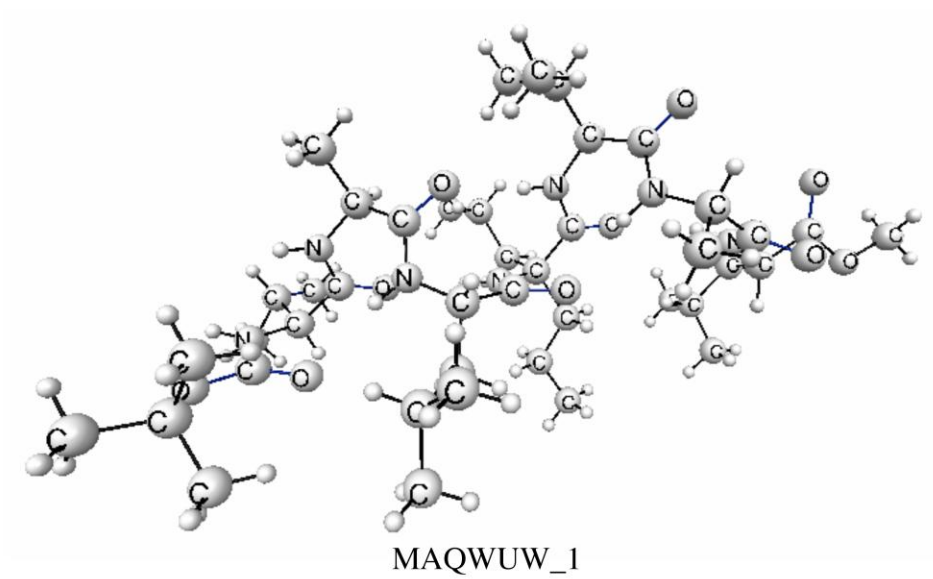


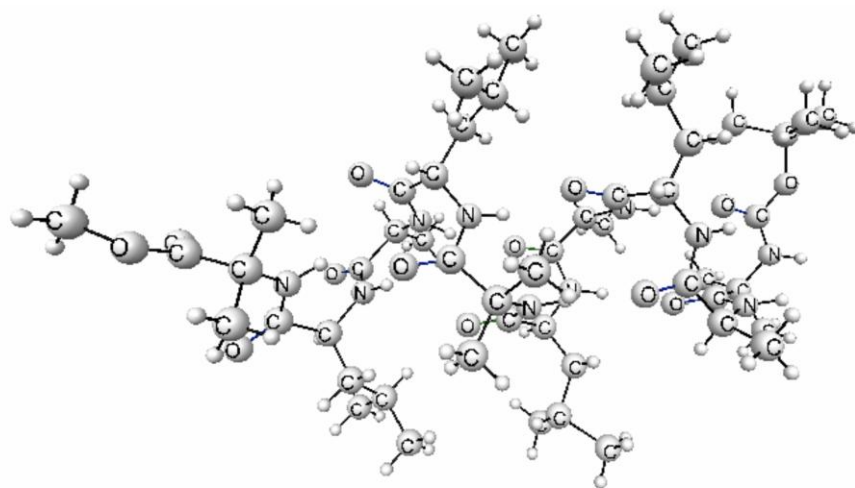
MOGQOO



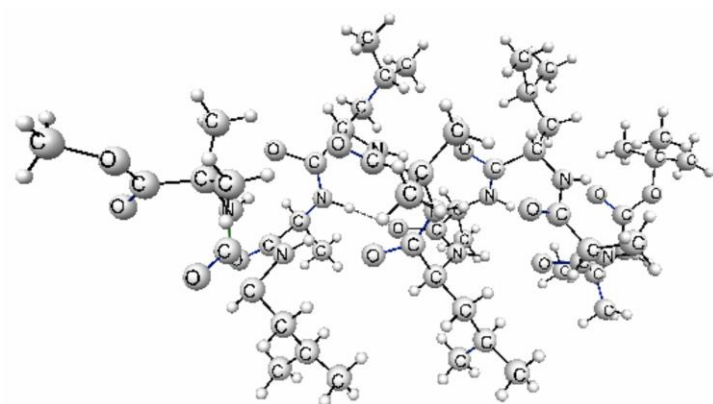
TAXYOG

Scheme 2. Depiction of alpha helix isomers used in Table 4. Structures are from the Cambridge Structural Database (CSD). Non-hydrogen atoms have been labeled.

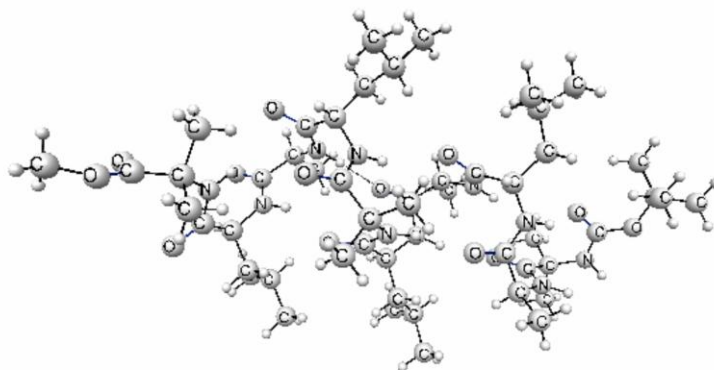




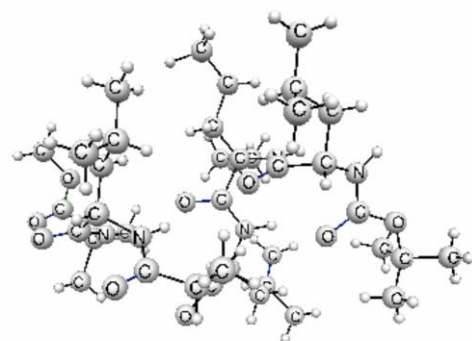
WUYCUO



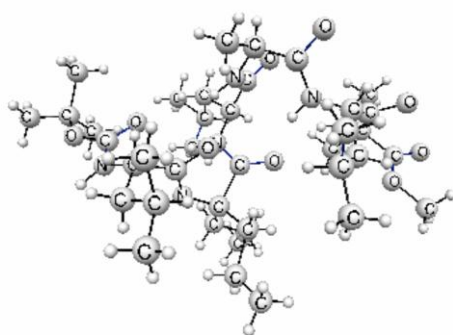
WUYDAV



WUYDEX



YETPES_1



YETPES_2

FIGURE CAPTIONS

Figure 1. Sandwich, T-shaped, and parallel-displaced configurations of the benzene dimer.

Figure 2. Comparison of SAPT and EFP interaction energy (kcal/mol) decomposition as a function of the separation ($\text{\AA}$) of benzene dimers in the sandwich configuration.

Figure 3. Comparison of SAPT and EFP interaction energy (kcal/mol) decomposition as a function of the separation R ($\text{\AA}$) of benzene dimers in the T-shaped configuration.

Figure 4. Electrostatic fractioning of bonds.

Figure 5. Monomer calculation performed in the ESP of the full system.

Figure 6. Illustration of FMO approximations applied to three monomers I, J, L (left) and as applied to dimer IJ and monomer L (right).

Figure 7. Lowest energy cluster of 32 water molecules obtained from EFP Monte Carlo/Simulated Annealing simulations.

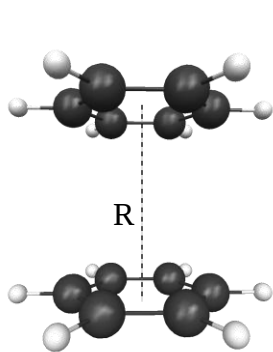
Figure 8. Ion pairs of 1-amino,4-H-1,2,4-triazolium dinitramide (top) and 1-H,4-H-1,2,4-triazolium dinitramide (bottom).

Figure 9. Lowest energy structure of 1-amino,4-H-1,2,4-triazolium dinitramide tetramer obtained from an *ab initio* MP2/6-31+G(d) optimization.

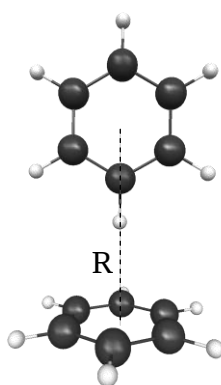
Figure 10. Lowest energy structure of 1-amino,4-H-1,2,4-triazolium dinitramide hexamer obtained from an *ab initio* MP2/6-31+G(d) optimization.

Figure 11. Lowest energy structure of 1-H,4-H-1,2,4-triazolium dinitramide tetramer obtained from an *ab initio* MP2/6-31+G(d) optimization.

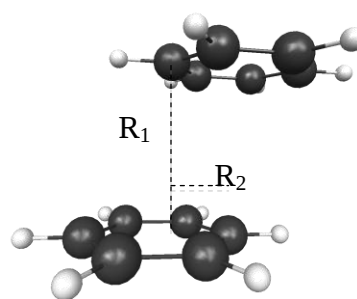
Figure 12. Lowest energy structure of 1-H,4-H-1,2,4-triazolium dinitramide hexamer obtained from an *ab initio* MP2/6-31+G(d) optimization.



sandwich



T-shaped



parallel-displaced

