ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

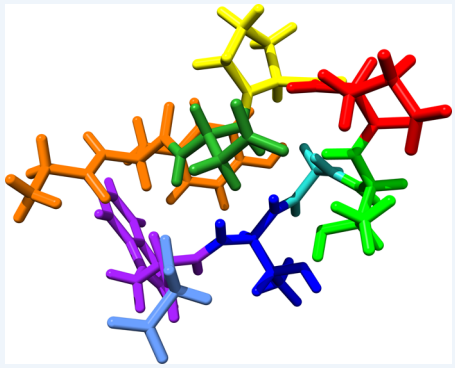
Efficient and Accurate Fragmentation Methods

Spencer R. Pruitt,[§,‡] Colleen Bertoni,[§] Kurt R. Brorsen,[§] and Mark S. Gordon*[,§]

[§]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

[‡]Argonne Leadership Computing Facility, Argonne, Illinois 60439, United States

CONSPECTUS: Three novel fragmentation methods that are available in the electronic structure program GAMESS (general atomic and molecular electronic structure system) are discussed in this Account. The fragment molecular orbital (FMO) method can be combined with any electronic structure method to perform accurate calculations on large molecular species with no reliance on capping atoms or empirical parameters. The FMO method is highly scalable and can take advantage of massively parallel computer systems. For example, the method has been shown to scale nearly linearly on up to 131 000 processor cores for calculations on large water clusters. There have been many applications of the FMO method to large molecular clusters, to biomolecules (e.g., proteins), and to materials that are used as heterogeneous catalysts.



The effective fragment potential (EFP) method is a model potential approach that is fully derived from first principles and has no empirically fitted parameters. Consequently, an EFP can be generated for any molecule by a simple preparatory GAMESS calculation. The EFP method provides accurate descriptions of all types of intermolecular interactions, including Coulombic interactions, polarization/induction, exchange repulsion, dispersion, and charge transfer. The EFP method has been applied successfully to the study of liquid water, π-stacking in substituted benzenes and in DNA base pairs, solvent effects on positive and negative ions, electronic spectra and dynamics, non-adiabatic phenomena in electronic excited states, and nonlinear excited state properties.

The effective fragment molecular orbital (EFMO) method is a merger of the FMO and EFP methods, in which interfragment interactions are described by the EFP potential, rather than the less accurate electrostatic potential. The use of EFP in this manner facilitates the use of a smaller value for the distance cut-off (R_{cut}). R_{cut} determines the distance at which EFP interactions replace fully quantum mechanical calculations on fragment−fragment (dimer) interactions. The EFMO method is both more accurate and more computationally efficient than the most commonly used FMO implementation (FMO2), in which all dimers are explicitly included in the calculation. While the FMO2 method itself does not incorporate three-body interactions, such interactions are included in the EFMO method via the EFP self-consistent induction term. Several applications (ranging from clusters to proteins) of the three methods are discussed to demonstrate their efficacy. The EFMO method will be especially exciting once the analytic gradients have been completed, because this will allow geometry optimizations, the prediction of vibrational spectra, reaction path following, and molecular dynamics simulations using the method.

1. INTRODUCTION

One of the goals of theoretical chemistry is to provide as accurate a description of chemical systems as possible within the limits of the currently available computational hardware and software. Over time, hardware limitations have receded and software has advanced, enabling accurate calculations on larger chemical systems than was heretofore possible. The most recent improvements have been multilevel parallel algorithms and accelerator hardware, for example, graphical processing units (GPUs) and many integrated core (MIC) coprocessors.

Developing codes that can make efficient use of many thousands of cores is challenging. On the one hand, one can modify an existing algorithm to take advantage of parallel architectures. Alternatively, one can develop new theoretical models that possess inherent parallelization that is straightforwardly implemented on parallel machines. This Account focuses on the latter path; in particular, the development of theoretical methods within the GAMESS computational package[1,2] that have been devised to take advantage of novel parallel architectures.

2. THEORETICAL METHODS

2.1. Effective Fragment Potential Method

The effective fragment potential (EFP) method was originally developed to describe effects that aqueous solvation can have on chemical reactions. That method (EFP1), contains three contributions to the interaction energy of the system: polarization, Coulomb, and repulsion terms. The details of the EFP methodology have been presented elsewhere,[3] so only a brief description of the method is presented here.

Special Issue: Beyond QM/MM: Fragment Quantum Mechanical Methods

Received: March 3, 2014
Published: May 8, 2014

   dx.doi.org/10.1021/ar500097m | *Acc. Chem. Res.* 2014, 47, 2786−2794

The total EFP1 interaction energy can be written as

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

In the presence of a quantum mechanics (QM) solute, the EFP1−QM contributions to the energy of the system are added as one-electron terms to the QM Hamiltonian.

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

The dipole−induced dipole polarization energy is iterated to self-consistency for both EFP−EFP and QM−EFP interactions, so this term captures many-body effects. Each induced dipole is located at the centroid of a localized molecular orbital (LMO), where the anisotropic distributed polarizability tensors are placed.

The EFP Coulomb term is calculated using a distributed multipolar analysis through octopoles. The distributed multipole moments, located at each atom center and bond midpoint, are obtained using the procedure outlined by Stone.[4,5] Short-range charge penetration effects are accounted for using a Gaussian like damping function of the form $1 - \beta \exp(-\alpha R^2)$, where the α and β parameters were obtained from a fit of the damped multipole potential to the QM Hartree−Fock (HF) or density functional theory (DFT) potential, and R is the distance between two multipole points or between a multipole point and an electron.

All remaining contributions to the EFP−QM interactions (exchange repulsion and charge transfer) are contained in the V^{rep} term, which is fitted to a functional form.

A more general formulation of the EFP method (termed EFP2, but herein referred to as simply EFP) was developed to allow for solvation effects using solvents other than water. In general, an EFP can be created for any molecule. The EFP energy expression can be written as

$$E^{\mathrm{EFP}} = E^{\mathrm{pol}} + E^{\mathrm{Coul}} + E^{\mathrm{rep}} + E^{\mathrm{disp}} + E^{\mathrm{ct}} \tag{3}$$

Equation 3 contains two additional terms, E^{disp} and E^{ct}. All of the terms in eq 3 are derived from first-principles, so there are no empirically fitted parameters.

The E^{disp} term between two fragments is evaluated as the leading dipole−induced dipole $1/R^6$ term. The higher order terms are estimated to be 1/3 of the $1/R^6$ term. The resulting formula is

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

The indices k and l refer to localized molecular orbitals (LMOs) on fragments A and B (A ≠ B). R_{kl} is the distance between the LMO centroids. C_6 is obtained by integrating over the imaginary frequency dependent polarizability tensors at points k and l.[6,7]

Both E^{rep} and E^{ct} depend on the intermolecular overlap of the fragment MOs. The expression for E^{rep} is truncated at the quadratic term, while the energy expression for E^{ct} is based on a second-order perturbative treatment of intermolecular interactions.[8−12] The E^{Coul}, E^{pol}, and E^{disp} terms are multiplied by a damping term that is based on the intermolecular overlap.[13]

2.2. Fragment Molecular Orbital Method

The fragment molecular orbital (FMO) method was developed to study biologically relevant systems such as proteins using QM methods.[14] By fragmentation of covalent bonds and division of a system into many separate fragments (monomers), the overall computational cost is reduced to that required for the largest fragment calculation. The general energy expression for the FMO2 method, involving explicit calculations on monomers and fragment pairs (dimers), is

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) \tag{5}$$

Monomer (I) and dimer (IJ) energies are obtained using standard QM methods. Three-body "trimer" calculations (IJK) can also be included for increased accuracy (FMO3), albeit at an increased computational cost.[15] The presence of the remaining fragments is accounted for during each monomer calculation through the use of an electrostatic potential (ESP), the Coulomb bath.[14] The overall procedure for the calculation of the total FMO energy is as follows:

(1) A fragmentation scheme is chosen, and the initial electron density distribution is calculated for each monomer.
(2) The fragment Fock operators are constructed using the electron densities from step 1, and the energy of each monomer is calculated in the presence of the Coulomb bath of the remaining fragments.
(3) Step 2 is repeated using the improved electron densities, resulting in a self-consistently converged ESP.
(4) The converged ESP is then incorporated into each dimer (and trimer) calculation, with each many-body energy being calculated only once (no self-consistency).

The ESP is incorporated into the one-electron Hamiltonian, $\tilde{H}^x$, as an additional term $V^x_{\mu\nu}$.

$$\tilde{H}^x_{\mu\nu} = H^x_{\mu\nu} + V^x_{\mu\nu} + B \sum_i \langle \mu || \varphi_i^h \rangle \langle \varphi_i^h || \nu \rangle \tag{6}$$

The third term in eq 6 is a projection operator used to divide basis functions across fractioned covalent bonds.

An important aspect of the FMO method is the inclusion of the ESP during individual fragment calculations. The specific form of $V^x_{\mu\nu}$ is

$$V^x_{\mu\nu} = \sum_{K(\neq x)} (u^K_{\mu\nu} + v^K_{\mu\nu}) \tag{7}$$

$$u^K_{\mu\nu} = \sum_{A \in K} \langle \mu | (-Z_A / |r - r_A|) | \nu \rangle \tag{8}$$

$$v^K_{\mu\nu} = \sum_{\lambda\sigma \in K} D^K_{\lambda\sigma} (\mu\nu | \lambda\sigma) \tag{9}$$

The $u^K_{\mu\nu}$ and $v^K_{\mu\nu}$ are the nuclear attraction and two-electron contributions, respectively. Both of these contributions are expressed in terms of the atomic orbitals (AOs) μ and ν and are calculated for all of the surrounding monomers K with electron density D^K. It is the self-consistent nature and subsequent inclusion of the ESP into each fragment calculation that allows the FMO method to achieve high accuracy, despite the approximations in the method. For chemical systems, inclusion of the ESP during individual fragment calculations also facilitates bond fractioning that does not require arbitrary hydrogen caps.

The FMO method may be considered to be an *ansatz* that can be combined with essentially any electronic structure method that is available in GAMESS. In recent years, the FMO

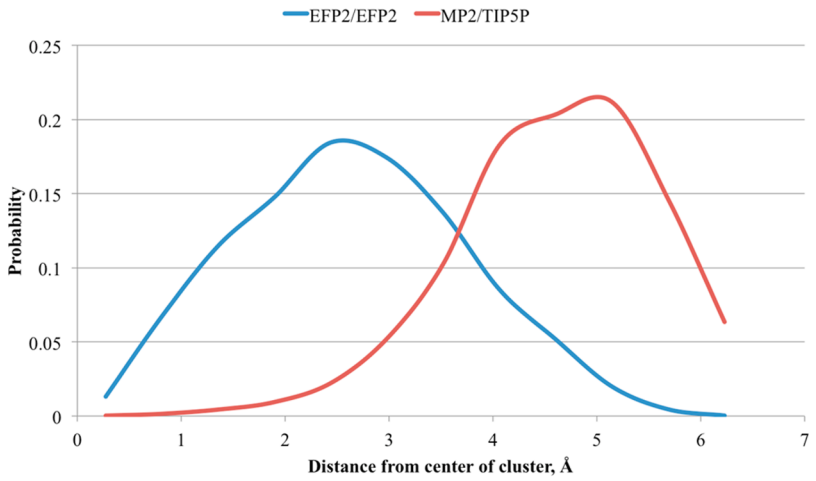


Figure 1. Probability distribution functions of the hydronium ion in a 32 water cluster using EFP2 for the entire system and using MP2 for the hydronium ion and TIP5P for the solvating water.

capability has been expanded to include multiconfigurational self-consistent field theory (MCSCF),[16] configuration interaction with single excitations (CIS),[17] restricted open-shell HF (ROHF),[18,19] density functional theory (DFT),[20] time-dependent DFT (TDDFT),[21] second order Møller−Plesset perturbation theory (MP2),[22] and coupled cluster theory (CC).[23] The FMO method can be combined with either explicit (EFP) or implicit methods to incorporate solvent effects.[24−26] Fully analytic gradients have been derived and implemented for FMO/HF,[27] FMO/DFT,[28] and FMO/MP2.[29] Fully analytic gradients facilitate the use of molecular dynamics (MD) simulations.[30−33] Each advancement of the FMO method has been derived separately, and all major aspects and improvements of the FMO method have been discussed and cited in a recent review.[3,34]

2.3. Effective Fragment Molecular Orbital Method

An important recent improvement of the FMO method is the development of the effective fragment molecular orbital (EFMO) method.[35] The EFMO method may be thought of as an integration of the FMO and EFP methods. The importance of this new development is best appreciated in the context of the FMO "separated dimers" approximation. Separated dimers are defined as pairs of fragments that are separated by a distance greater than a user defined cutoff value, R_{cut}. The FMO energy expression (eq 5) can be rewritten as

$$E^{FMO2} = \sum_{I}^{N} E_I + \sum_{I>J}^{R_{I,J}\le R_{cut}} \Delta E_{IJ} + \sum_{I>J}^{R_{I,J}>R_{cut}} \Delta E_{IJ}^{sep} \tag{10}$$

where

$$\Delta E_{IJ} = (E_{IJ} - E_I - E_J) \tag{11}$$

The unitless distance between two fragments I and J is defined as

$$R_{I,J} = \min_{i\in I, j\in J}\left\{\frac{|\vec{r_i} - \vec{r_j}|}{r_i^{vdw} + r_j^{vdw}}\right\} \tag{12}$$

So, R_{IJ} is the relative minimum distance between fragments I and J based on the van der Waals radii of the atoms contained in each fragment. For example, $R_{IJ} = 1$ means the fragments I and J are separated by the van der Waals distance. R_{cut} is the value of R_{IJ} beyond which two fragments are taken to be "separated". The separated dimers described above are represented by the third term in eq 10. Using this new formulation of the total FMO energy, the number of fully QM dimers evaluated can be reduced based on the value of R_{cut}.

The division of dimer calculations into QM and separated provides added flexibility to how the separated dimer calculations are evaluated. In the FMO method, the separated dimer calculations are approximated by a combination of one-electron Coulomb potentials and the nuclear repulsion energy. However, this approximate description of dimer interactions restricts the minimum distance at which R_{cut} can be set. If R_{cut} is too small, the result will be a decrease in accuracy, causing the FMO method to become unreliable. By performing the separated dimer calculations using a more accurate method, one can maintain, or even improve, the overall accuracy of the FMO method, while still reducing the computational cost of a given calculation. In the EFMO method, the EFP method is used to evaluate the separated dimer energies.

Equation 10 can be combined with eq 3 and rewritten as

$$E^{EFMO} = \sum_{I}^{N} E_I^0 + \sum_{I>J}^{R_{I,J}\le R_{cut}} (\Delta E_{IJ}^0 - E_{IJ}^{pol}) + \sum_{I>J}^{R_{I,J}>R_{cut}} (E_{IJ}^{Coul} + E_{IJ}^{disp} + E_{IJ}^{rep} + E_{IJ}^{ct}) + E_{tot}^{pol} \tag{13}$$

As shown in eq 13, all five intermolecular interactions from the EFP method are used to describe the separated dimers.[36] Importantly, the polarization term in eq 13 incorporates many-body effects into EFMO in a natural way. So, EFMO incorporates many-body effects within the FMO2 level. Through the use of a QM based model potential such as EFP, the minimum value of R_{cut} can be lowered, resulting in a reduction in computational cost compared with the FMO method. During a geometry optimization, an MD simulation, or a Monte Carlo simulation, new EFP potentials are generated in each step. The constraint of internally frozen EFP geometries is thereby removed. This will be a key improvement for phenomena like phase changes and the study of infrared spectra. As an additional benefit, an EFP energy decomposition analysis (EDA) is obtained for all separated dimer interactions.

dx.doi.org/10.1021/ar500097m | Acc. Chem. Res. 2014, 47, 2786−2794

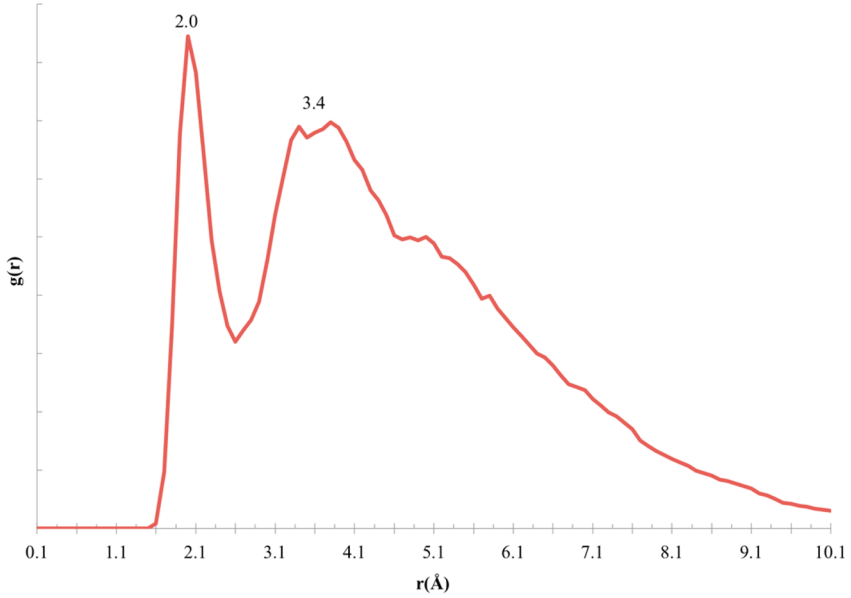


Figure 2. Radial distribution function of water H atoms around nitrate O atoms in $NO_3^-(H_2O)_{32}$ simulation. FMO-MD/HF/6-31+G(d) simulation run for 20 ps with a step size of 0.25 fs.

3. RECENT ADVANCES

3.1. Effective Fragment Potential Method Umbrella Sampling

Umbrella sampling[37] has been combined with the EFP method[38] to compute the surface affinity of the hydronium ion in water. Figure 1 shows probability distribution functions for a hydronium ion solvated by 32 water molecules. For the umbrella sampling calculations, a spherical harmonic boundary potential centered at the origin was applied to prevent evaporation. The edge of the spherical harmonic boundary potential was set to 6.2 Å so that the density of the cluster would be equal to the density of water at 300 K if all molecules were inside the spherical boundary. For this cluster, when the EFP method is used to describe both the hydronium ion and the solvating waters, umbrella sampling calculations predict that the proton most probably lies about 1/3 of the way from the center of the water cluster to the surface of the cluster. For calculations using the MP2/aug-cc-pVDZ level of theory to describe the hydronium ion and the TIP5P potential for the solvating waters, umbrella sampling predicts that the proton most probably lies about 3/4 of the way from the center of the cluster toward the surface of the cluster.

The EFP results predict that the H_3O^+ lies further from the surface of the water cluster than has been suggested by much of the recent computational and experimental results on the surface affinity of the solvated hydronium ion.[39−43] Since most of the previous computational studies have used classical potentials or a QM/MM level of theory and the EFP method has been shown to accurately reproduce geometries and energetics of water at the MP2 level theory,[44,45] it is possible that the EFP method is correctly mimicking *ab initio* molecular dynamics (MD) with MP2. To further investigate this possibility, future studies will examine the solvated hydronium using FMO-MD or EFMO-MD.

3.2. Fragment Molecular Orbital Molecular Dynamics

FMO-MD simulations were recently implemented[32,33] in the GAMESS program package after the addition of fully analytic energy gradients[27] to the FMO2 method. While improvements are still being made to the initial implementation, preliminary studies of the $(NO_3)^-$ ion in solution were performed to further investigate the accuracy of the FMO-MD method. Simulations were performed at the FMO2/HF/6-31+G(d,p) level of theory. The nitrate ion was solvated with 32 water molecules, and a fragmentation scheme of one molecule (water or nitrate) per fragment was used. The molecular cluster was equilibrated for 500 fs using the NVT ensemble, followed by 20 ps of simulation using the NVE ensemble. A step size of 0.25 fs was used with default values for the FMO2 approximations. The fully analytic energy gradients require that the electrostatic point charge approximation to the ESP be turned off.

Figure 2 shows the radial distribution function of water H atoms around the nitrate O atoms obtained from the 20 ps simulation. Two solvation shells are observed, the first at 2.0 Å and the second at 3.4 Å. Compared with previous studies,[46] the FMO-MD results agree well with EFP based results that predict two solvation shells at 2.1 and 3.4 Å. Amber 8 MD simulations from the same work give similar results, with peaks at 1.8 and 3.1 Å.

The FMO2-MD method does not include nonclassical three-body interactions that are essential for the accurate description of water clusters.[44,45,47−51] Nonetheless, the FMO-MD simulation reproduces the hydrogen bonding interactions within the $(NO_3)^-(H_2O)_{32}$ cluster, providing good agreement with previous work.[46] Future calculations will employ the EFMO/MP2 method, in order to incorporate both many-body effects and electron correlation.

3.3. Effective Fragment Molecular Orbital Method Applied to a Small Protein

Previously,[36] EFMO calculations were presented on molecular clusters. Now, consider the ability of the EFMO method to accurately describe a covalently bonded system. The small protein chignolin (PDB 1UAO) was used as a test system. The protein was divided into nine fragments, with the fragmentation scheme shown in Figure 3. MP2, FMO2-MP2, and EFMO-MP2 single point energy calculations using three different basis sets (6-31G(d), 6-31G(d,p), and 6-311G(d,p)) were performed on the structure obtained from the PDB database. Default

dx.doi.org/10.1021/ar500097m | *Acc. Chem. Res.* 2014, 47, 2786−2794

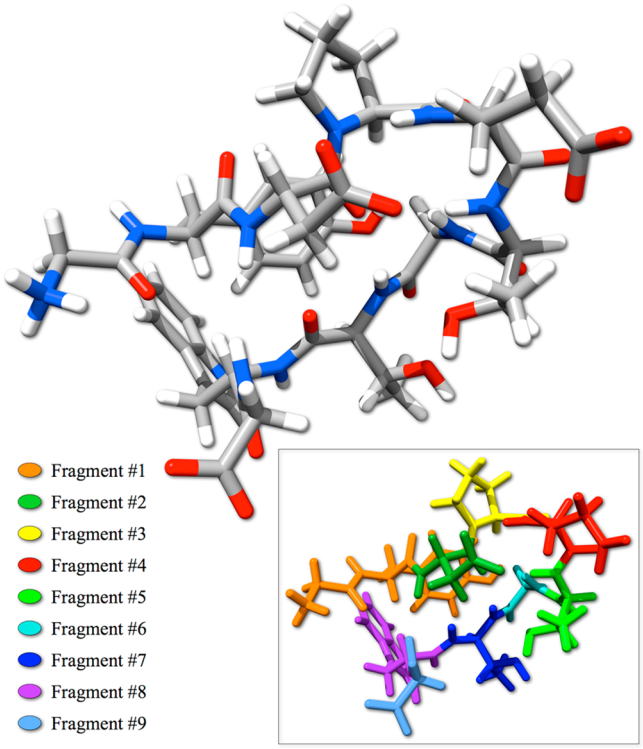


Figure 3. Fragmentation scheme used for FMO2 and EFMO calculations on chignolin.

values for FMO2 were used, while the R_{cut} value for the electrostatic dimer approximation was reduced by half (from 2.0 to 1.0) for the EFMO calculations, resulting in fewer QM dimer calculations (30 FMO2 QM dimers versus 14 EFMO QM dimers).

Table 1 shows the signed errors in the total energy of the protein for both FMO2 and EFMO compared with fully *ab initio* MP2 results. For the two smaller basis sets, FMO2 and EFMO perform similarly, with errors of 11.4 and −8.3 kcal/mol. The difference between the two methods decreases with an increase to the 6-31G(d,p) basis sets, providing errors of 10.3 and −10.8 kcal/mol for FMO2 and EFMO, respectively. The 6-311G(d,p) basis set proved troublesome for the FMO2 method, requiring many additional iterations to converge the ESP, producing an error of −54.4 kcal/mol. However, it is well-known that the FMO2 method performs poorly using extended basis sets.[52−54] On the other hand, the performance of the EFMO method improves when using the larger basis set, with the error dropping to only −4.0 kcal/mol compared with fully *ab initio* MP2 results.

Table 1. Errors (kcal/mol) for FMO2 and EFMO Using Three Different Basis Sets Compared with Fully *ab Initio* MP2 Energies

	FMO2	EFMO
6-31G(d)	11.4	−8.3
6-31G(d,p)	10.3	−10.8
6-311G(d,p)	−54.4	−4.0

Figure 4 shows the EDA among all separated dimers in the EFMO calculation. Since the EFMO method uses the EFP method to calculate the separated dimer interaction energies, the EDA is an added benefit of using the EFMO method. Two dimer interactions of particular interest, shown in Figure 5 (6.1 and 8.1), are investigated in greater detail. The interaction in dimer 6.1 (glycine/tyrosine) is characterized by a large repulsive contribution to the interaction energy (Figure 4). Figure 5 shows that the closest interaction is between the hydrogen atoms on the two amino acids. The dimer interaction between the tryptophan (fragment 8) and tyrosine (fragment 1) is dominated by a T-shaped dispersive interaction between two aromatic rings. This type of $\pi-\pi$ interaction has been studied previously using the EFP method. It was shown in that work[55] that EFP is capable of providing dispersive interaction energies that agree well with high level CC and symmetry adapted perturbation theory (SAPT) calculations.

3.4. Toward Fully Analytic EFMO Gradients

In the initial version of the EFMO method,[35] approximate gradients for the polarization energy term and the Coulomb energy term were derived and implemented. However, in order to perform accurate MD simulations and geometry optimizations using the EFMO method, the fully analytic gradient for all terms is required. Thus, the fully analytic gradients for all terms in the EFMO method are currently being derived and implemented.

The general EFMO gradient equation is represented as

$$\frac{\partial E^{\mathrm{EFMO}}}{\partial x_c} = \sum_{I}^{N} \frac{\partial E_I^0}{\partial x_c} + \sum_{I>J}^{R_{I,J}\le R_{\mathrm{cut}}} \left(\frac{\partial \Delta E_{IJ}^0}{\partial x_c} - \frac{\partial E_{IJ}^{\mathrm{pol}}}{\partial x_c} \right) + \sum_{I>J}^{R_{I,J}>R_{\mathrm{cut}}} \left(\frac{\partial E_{IJ}^{\mathrm{Coul}}}{\partial x_c} + \frac{\partial E_{IJ}^{\mathrm{disp}}}{\partial x_c} + \frac{\partial E_{IJ}^{\mathrm{rep}}}{\partial x_c} + \frac{\partial E_{IJ}^{\mathrm{ct}}}{\partial x_c} \right) + \frac{\partial E_{\mathrm{tot}}^{\mathrm{pol}}}{\partial x_c} \tag{14}$$

where the derivative of each term in eq 13 is taken with respect to the x-coordinate of atom c. The EFMO energy expression contains gas phase *ab initio* energy terms and EFP interaction energy terms. The *ab initio* terms are the monomer energy and the QM nonseparated dimer energy terms $(\partial E_I^0/(\partial x_c), \Delta E_{IJ}^0/(\partial x_c))$, and the EFP interaction energy terms are the separated dimer energy terms and the polarization energy terms $(\partial E_{IJ}^{\mathrm{pol}}/(\partial x_c), \partial E_{IJ}^{\mathrm{coul}}/(\partial x_c), \partial E_{IJ}^{\mathrm{disp}}/(\partial x_c), \partial E_{IJ}^{\mathrm{rep}}/(\partial x_c), \partial E_{IJ}^{\mathrm{ct}}/(\partial x_c), \partial E_{\mathrm{tot}}^{\mathrm{pol}}/(\partial x_c))$. The previously derived and implemented gradient for any *ab initio* term can be used directly. The previously derived EFP gradients cannot be directly used in the EFMO gradient, because in EFP calculations the fragments are rigid. Certain terms, such as the multipole moments in the EFP Coulomb term and polarizability tensors in the EFP polarization term, depend on the fragment internal geometry. If the fragments are rigid, as in the EFP method, the internal geometry does not change. Since the fragments are flexible in the EFMO method, the fragment internal geometry does change, and the derivative of these terms needs to be included in the gradient. Although gradients for the polarization energy term and the Coulomb energy term currently exist,[35] the gradients are not fully analytic, since the terms that arise due to the geometry dependence of the multipole moments and polarizability tensors are neglected.

Recently, the fully analytic gradient of the Coulomb term $(\partial E_{IJ}^{\mathrm{coul}}/(\partial x_c))$ has been derived and implemented. This includes the gradient of the previously neglected terms, which requires solving the coupled perturbed HF equations (in practice, a Z-vector equation is solved).[56,57] By inclusion of the

dx.doi.org/10.1021/ar500097m | Acc. Chem. Res. 2014, 47, 2786−2794

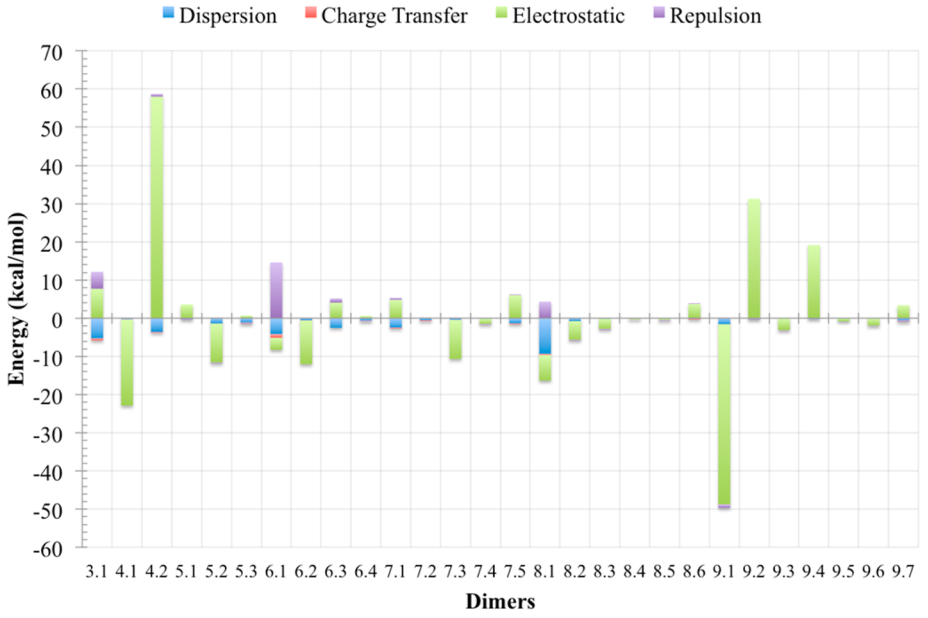


Figure 4. EFMO energy decomposition analysis between all separated dimers.

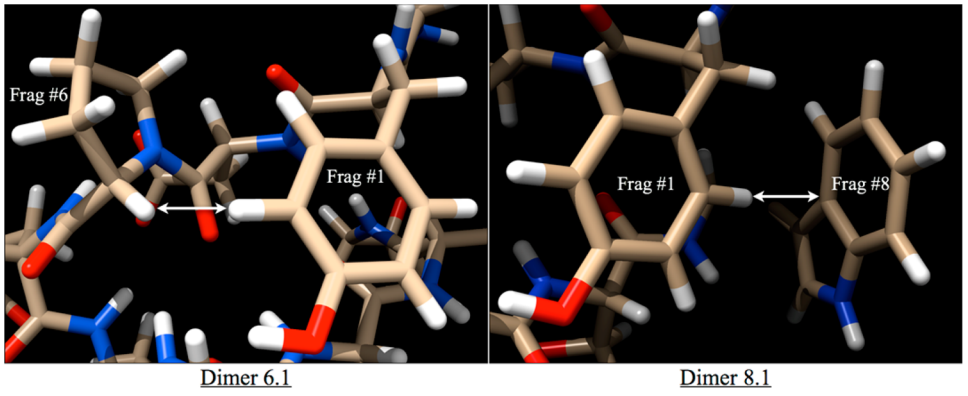


Figure 5. Local views of dimers 6.1 and 8.1.

terms missing from the original EFMO gradient, the Coulomb energy term gradient is now fully analytic. The implementation was tested by comparing the numeric gradient to the analytic gradient for a system containing 64 water fragments using a 6-31++G(d,p) basis set.

The numeric gradient was calculated using a step size of 0.005 Å and an R_{cut} value of 0.3. The small R_{cut} value forces all dimer energies to be considered separated and therefore calculated using the EFP gradient code. The polarization energy was removed to test only the Coulomb energy term. Therefore, the energy equation tested was $E^{EFMO} = \sum_I^N E_I^0 + \sum_{I>J}^{R_{IJ}>R_{cut}} E_{IJ}^{Coul}$. The results are shown in Table 2. The new implementation shows a significant improvement over the previous Coulomb energy term gradient implementation. The maximum gradient values resulting from the new implementation agree well with the numeric maximum gradient value, and the root-mean-square (RMS) error is small. The previous implementation has significantly larger RMS errors. Figure 6 shows this difference by plotting the error of the original approximate EFMO gradient and the fully analytic EFMO gradient relative to the numeric gradient.

Table 2. Comparison of EFMO Gradients and Numeric Gradients for a Cluster of 64 Water Molecules Using the 6-31++G(d,p) Basis Set[a]

	maximum gradient value	RMS error	maximum error
numeric	0.022 894		
original EFMO	0.023 659	0.002 904	0.013 334
analytic EFMO	0.022 883	0.000 013	0.000 037

[a]All values are in hartree/bohr.

The fully analytic gradients for the remaining separated dimer EFP interaction energy terms (polarization, exchange repulsion, dispersion, and charge transfer) are in the process of being derived and implemented. Once all of the gradient terms are fully analytic, accurate MD and geometry optimizations can be performed using EFMO.

4. CONCLUDING REMARKS

Three fragmentation methods have been discussed in this Account. The fragment molecular orbital (FMO) method can be combined with any electronic structure method to perform accurate calculations on large molecular species. The FMO method is highly scalable and can take advantage of massively parallel computer systems. Analytic gradients are available for the FMO implementations of Hartree−Fock, density functional

dx.doi.org/10.1021/ar500097m | Acc. Chem. Res. 2014, 47, 2786−2794

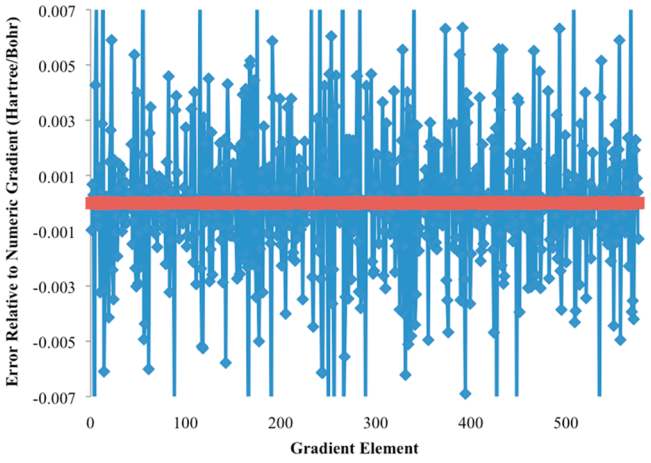


Figure 6. Error of the analytic gradients relative to the numeric gradient for each gradient element in a cluster of 64 water molecules. Blue, original implementation; red, new, fully analytic implementation.

theory, and second order perturbation theory, thereby facilitating geometry optimizations and molecular dynamics (MD) simulations for these methods. In general FMO simulations are expected to provide more accuracy than MD simulations with highly parametrized model potentials, providing *ab initio* insight into bulk properties.

The effective fragment potential (EFP) method is a model potential approach that is fully derived from first-principles and has no empirically fitted parameters. The EFP method provides accurate descriptions of all types of intermolecular interactions, ranging from Coulombic interactions to dispersion-controlled interactions. Fully analytic EFP gradients are also available, so that EFP MD simulations are possible when only intermolecular (not covalent) interactions are of interest.

The effective fragment molecular orbital (EFMO) method is a merger of the FMO and EFP methods, in which interfragment interactions are described by the EFP potential rather than the less accurate electrostatic potential. Using the EFP method in this manner provides three-body interactions without the need for more expensive higher orders of FMO theory. The EFMO method also provides a route to internally flexible fragments, thereby removing a limitation of the standalone EFP method. The EFMO method is both more accurate and more computationally efficient that the most commonly used FMO implementation, FMO2. The derivation and implementation of fully analytic EFMO energy gradients are in progress.

Several applications of the three methods demonstrate their efficacy.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

Biographies

Spencer R. Pruitt received his B.S. degree in chemistry from the University of Minnesota, Duluth, in 2006. Afterwards he entered the graduate program at Iowa State University, joining the research group of Prof. Mark S. Gordon. He received his Ph.D. in 2012 and continued his work in the Gordon group as a postdoctoral associate. He is currently a postdoctoral appointee in the Leadership Computing Facility at Argonne National Laboratory. His research interests include high performance computing, the study of intermolecular interactions, the development of the Fragment Molecular Orbital method, and *ab initio* molecular dynamics.

Colleen Bertoni received her B.S. degree in chemistry from the University of Texas at Austin in 2010. She is currently a graduate student at Iowa State University in the research group of Mark S. Gordon. She is interested in fragmentation methods.

Kurt R. Brorsen received his B.S. from the University of Oklahoma in 2008. He is currently a graduate student in Mark Gordon's group at Iowa State University. His research interests are focused on fragmentation methods and *ab initio* molecular dynamics.

Mark S. Gordon, Frances M. Craig Distinguished Professor of Chemistry at Iowa State University and Director of the Ames Laboratory Applied Mathematical and Computational Sciences program, was born and raised in New York City. After completing his B.S. in Chemistry in 1963, Professor Gordon entered the graduate program at Carnegie Institute of Technology, where he received his Ph.D. in 1967 under the guidance of Professor John Pople, 1998 Chemistry Nobel Laureate. Following a postdoctoral research appointment with Professor Klaus Ruedenberg at Iowa State University, Professor Gordon accepted a faculty appointment at North Dakota State University in 1970, where he rose through the ranks, eventually becoming distinguished professor and department chair. He moved to Iowa State University and Ames Laboratory in 1992. Professor Gordon's research interests are very broadly based in electronic structure theory and related fields, including solvent effects, the theory of liquids, surface science, the design of new materials, and chemical reaction mechanisms. He has authored more than 560 research papers and is a member of the International Academy of Quantum Molecular Science and of the World Association of Theoretical and Computational Chemistry (WATOC). He received the 2009 American Chemical Society Award for Computers in Chemical and Pharmaceutical Research and the 2014 WATOC Schrodinger medal. He is a Fellow of the American Chemical Society, the American Physical Society, and the American Association for the Advancement of Science.

ACKNOWLEDGMENTS

The development of the EFP, FMO, and EFMO methods has been supported by a National Science Foundation SI2 grant and by the Air Force Office of Scientific Research under AFOSR Award No. FA9550-11-1-0099. K.R.B. acknowledges the support of a Department of Energy Computational Science Graduate Fellowship. Generous allocations of computer time were provided by a Grand Challenge Grant from the Department of Defense High Performance Computing Modernization Office and by the Iowa State University NSF MRI cluster Cyence.

REFERENCES

(1) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347.

(2) Gordon, M. S.; Schmidt, M. W. Advances in Electronic Structure Theory: GAMESS a Decade Later. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.;Elsevier: Amsterdam, The Netherlands, 2005; pp 1167−1189.

(3) Gordon, M. S.; Fedorov, S. R.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632.

(4) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.
(5) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1996.
(6) Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. Ab-initio Prediction of Properties of Carbon Dioxide, Ammonia, and Carbon Dioxide···Ammonia. *J. Phys. Chem.* **1985**, *89*, 2186−2192.
(7) Adamovic, I.; Gordon, M. S. Dynamic Polarizability, Dispersion Coefficient C_6 and Dispersion Energy in the Effective Fragment Potential Method. *Mol. Phys.* **2005**, *103*, 379−387.
(8) Jensen, J. H.; Gordon, M. S. An Approximate Formula for the Intermolecular Pauli Repulsion Between Closed Shell Molecules. *Mol. Phys.* **1996**, *89*, 1313−1325.
(9) Jensen, J. H. Modeling Intermolecular Exchange Integrals between Non-orthogonal Molecular Orbitals. *J. Chem. Phys.* **1996**, *104*, 7795−7796.
(10) Jensen, J. H.; Gordon, M. S. An Approximate Formula for the Intermolecular Pauli Repulsion between Closed Shell Molecules. II. Application to the Effective Fragment Potential Method. *J. Chem. Phys.* **1998**, *108*, 4772−4782.
(11) Jensen, J. H. Intermolecular Exchange-Induction and Charge Transfer: Derivation of Approximate Formulas Using Non-orthogonal Localized Molecular Orbitals. *J. Chem. Phys.* **2001**, *114*, 8775−8783.
(12) Li, H.; Gordon, M. S.; Jensen, J. H. Charge Transfer Interaction in the Effective Fragment Potential Method. *J. Chem. Phys.* **2006**, *124*, No. 214108.
(13) Slipchenko, L. V.; Gordon, M. S. Damping Functions in the Effective Fragment Potential Method. *Mol. Phys.* **2009**, *107*, 999−1016.
(14) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(15) Fedorov, D. G.; Kitaura, K. The Three-Body Fragment Molecular Orbital Method for Accurate Calculations of Large Systems. *Chem. Phys. Lett.* **2006**, *433*, 182−187.
(16) Fedorov, D. G.; Kitaura, K. Multiconfiguration self-consistent-field theory based upon the fragment molecular orbital method. *J. Chem. Phys.* **2005**, *122*, No. 054108.
(17) Mochizuki, Y.; Koikegami, S.; Amari, S.; Segawa, K.; Kitaura, K.; Nakano, T. Configuration Interaction Singles Method with Multilayer Fragment Molecular Orbital Scheme. *Chem. Phys. Lett.* **2005**, *406*, 283−288.
(18) Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Open-Shell Formulation of the Fragment Molecular Orbital Method. *J. Chem. Theor. Comput.* **2010**, *6*, 1−5.
(19) Pruitt, S. R.; Fedorov, D. G.; Gordon, M. S. Geometry Optimizations of Open-Shell Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2012**, *116*, 4965−4974.
(20) Sugiki, S.; Kurita, N.; Sengoku, Y.; Sekino, H. Fragment Molecular Orbital Method with Density Functional Theory and DIIS Convergence Acceleration. *Chem. Phys. Lett.* **2003**, *382*, 611−617.
(21) Chiba, M.; Fedorov, D.; Kitaura, K. Time-Dependent Density Functional Theory with the Multilayer Fragment Molecular Orbital Method. *Chem. Phys. Lett.* **2007**, *444*, 346−350.
(22) Fedorov, D. G.; Kitaura, K. Second Order Møller-Plesset Perturbation Theory Based upon the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *121*, 2483−2490.
(23) Fedorov, D. G.; Kitaura, K. Coupled-Cluster Theory Based upon the Fragment Molecular-Orbital Method. *J. Chem. Phys.* **2005**, *123*, No. 134103.
(24) Nagata, T.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. (2009). A combined effective fragment potential−fragment molecular orbital method. I. The energy expression and initial applications. *J. Chem. Phys.* **2009**, *131*, 024101−12.
(25) Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K. Analysis of Solute-Solvent Interactions in the Fragment Molecular Orbital Method Interfaced with Effective Fragment Potentials: Theory and Application to a Solvated Griffithsin-Carbohydrate Complex. *J. Phys. Chem. A* **2012**, *116*, 9088−9099.
(26) Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J.; Gordon, M. S. The Polarizable Continuum Model (PCM) Interfaced with the Fragment Molecular Orbital Method (FMO). *J. Comput. Chem.* **2006**, *27*, 976−985.
(27) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully Analytic Energy Gradient in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2011**, *134*, No. 124115.
(28) Brorsen, K. R.; Zahariev, F.; Nakata, H.; Gordon, M. S. Analytic gradient for fragment molecular orbital density functional theory. Manuscript in preparation.
(29) Nagata, T.; Fedorov, D. G.; Ishimura, K.; Kitaura, K. Analytic Energy Gradient for Second-Order Møller-Plesset Perturbation Theory Based on the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2011**, *135*, No. 044110.
(30) Komeiji, Y.; Nakano, T.; Fukuzawa, K.; Ueno, Y.; Inadomi, Y.; Nemoto, T.; Uebayasi, M.; Fedorov, D. G.; Kitaura, K. Fragment Molecular Orbital Method: Application to Molecular Dynamics Simulation, '*ab Initio* FMO-MD'. *Chem. Phys. Lett.* **2003**, *372*, 342−347.
(31) Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Fedorov, D. G. Fragment Molecular Orbital-Based Molecular Dynamics (FMO-MD), a Quantum Simulation Tool for Large Molecular Systems. *J. Mol. Struct. THEOCHEM* **2009**, *898*, 2−7.
(32) Nagata, T.; Fedorov, D. G.; Kitaura, K. Analytic Gradient and Molecular Dynamics Simulations Using the Fragment Molecular Orbital Method Combined with Effective Potentials. *Theor. Chem. Acc.* **2012**, *131*, 1−15.
(33) Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment Molecular Orbital Molecular Dynamics with the Fully Analytic Energy Gradient. *J. Chem. Theor. Comput.* **2012**, *8*, 5008−5012.
(34) Fedorov, D. G., Kitaura, K., Eds.; *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009; pp 1−36.
(35) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **2010**, *114*, 8705−8712.
(36) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully Integrated Effective Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2013**, *9*, 2235−2249.
(37) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16*, 1339−1350.
(38) Choi, C. H.; Re, S.; Feig, M.; Sugita, Y. Quantum Mechanical/Effective Fragment Potential Molecular Dynamics (QM/EFP-MD) Study on Intra-molecular Proton Transfer of Glycine in Water. *Chem. Phys. Lett.* **2012**, *539*, 218−221.
(39) Petersen, P. B.; Saykally, R. J. On the Nature of Ions at the Liquid Water Surface. *Annu. Rev. Phys. Chem.* **2006**, *57*, 333−364.
(40) Buch, V.; Milet, A.; Vácha, R.; Jungwirth, P.; Devlin, J. P. Water Surface Is Acidic. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7342−7347.
(41) Tarbuck, T. L.; Ota, S. T.; Richmond, G. L. Spectroscopic Studies of Solvated Hydrogen and Hydroxide Ions at Aqueous Surfaces. *J. Am. Chem. Soc.* **2006**, *128*, 14519−14527.
(42) Kumar, R.; Knight, C.; Voth, G. A. Exploring the Behaviour of the Hydrated Excess Proton at Hydrophobic Interfaces. *Faraday Discuss.* **2013**, *167*, 263−278.
(43) Wick, C. D. Hydronium Behavior at the Air−Water Interface with a Polarizable Multistate Empirical Valence Bond Model. *J. Phys. Chem. C* **2012**, *116*, 4026−4038.
(44) Kemp, D. D.; Gordon, M. S. An Interpretation of the Enhancement of the Water Dipole Moment Due to the Presence of Other Water Molecules. *J. Chem. Phys. A* **2008**, *112*, 4885−4894.
(45) Pruitt, S. R.; Leang, S. S.; Xu, P.; Fedorov, D. G.; Gordon, M. S. Hexamers and Witchamers: Which Hex Do You Choose? *Comput. Theor. Chem.* **2013**, *1021*, 70−83.

(46) Miller, Y.; Thomas, J. L.; Kemp, D. D.; Finlayson-Pitts, B. J.; Gordon, M. S.; Tobias, D. J.; Gerber, R. B. Structure of Large Nitrate−Water Clusters at Ambient Temperatures: Simulations with Effective Fragment Potentials and Force Fields with Implications for Atmospheric Chemistry. *J. Phys. Chem. A* **2009**, *113*, 12805−12814.
(47) Xantheas, S. S.; Dunning, T. H. The Structure of the Water Trimer from ab Initio Calculations. *J. Chem. Phys.* **1993**, *98*, 8037−8040.
(48) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters (H2O)n n=1−6. II. Analysis of Many-Body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523−7534.
(49) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. Contribution of Many-Body Terms to the Energy for Small Water Clusters: A Comparison of ab Initio Calculations and Accurate Model Potentials. *J. Phys. Chem. A* **1997**, *101*, 9163−9168.
(50) Xantheas, S. S. Cooperativity and Hydrogen Bonding Network in Water Clusters. *Chem. Phys.* **2000**, *258*, 225−231.
(51) Xantheas, S. S. Interaction Potentials for Water from Accurate Cluster Calculations. *Struct. Bonding (Berlin)* **2005**, *116*, 119−148.
(52) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. Accuracy of the Three-Body Fragment Molecular Orbital Method Applied to Møller-Plesset Perturbation Theory. *J. Comput. Chem.* **2007**, *28*, 1476−1484.
(53) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic Study of the Embedding Potential Description in the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2010**, *114*, 8742−8753.
(54) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2014**, *597*, 99−105.
(55) Smith, T.; Slipchenko, L. V.; Gordon, M. S. Modeling $\pi-\pi$ Interactions with the Effective Fragment Potential Method: The Benzene Dimer and Substituents. *J. Phys. Chem. A* **2008**, *112*, 5286−5294.
(56) Handy, N. C.; Schaefer, H. F. On the Evaluation of Analytic Energy Derivatives for Correlated Wave Functions. *J. Chem. Phys.* **1984**, *81*, 5031−5033.
(57) Yamaguchi, Y.; Osamura, O.; Goddard, J. D.; Schaefer, H. F. *A New Dimension to Quantum Mechanics. Analytical Derivative Methods in ab Initio Molecular Electronic Structure Theory*. Oxford University Press: Oxford, 1994.