Development of Fragment-based Quantum Mechanical Methods and Combined Quantum Mechanical and Molecular Mechanical Methods

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Yingjie Wang

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

Prof. Jiali Gao, Advisor

August, 2014
Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Professor Jiali Gao, for his excellent guidance, caring, patience, enthusiasm, and providing me with an excellent atomosphere for doing research and financially supporting my study. His profound knowledge and unique insight on multiscale modeling, has guided me on different projects.

Besides, I would like to thank Professor Donald G. Truhlar, who is also one of my defense committee, for his supervision and guidance in the collaboration of the X-Pol project. His systematic strategy in analyzing and tackling difficult problems enlightens me of the professional and productive research philosophy. I would also like to thank another committee member, Professor Carlos P. Sosa, for his help on my learning of the Gaussian program and collaboration on the X-Pol project.

Special thanks are also dedicated to Mike Frisch and his colleagues in Gaussian, Inc. They provided me precious intern opportunities in developing high-quality computational chemistry software package, and invited me attending their productive workshops.

Many thanks go to my group members, both current and past, especially Dr. Peng Bao, Peng Zhang, Dr. Alessandro Cembran, and Dr. Michael M. Mazack. They offered generous help on my research and pleasant collaboration.

Finally, and most importantly, I offer my regards and blessings to my beloved parents. Their encouragement and love has always been my energy to face various challenges during the completion of my Ph.D. study.

Minneapolis, August 2014
Dedication

This dissertation is dedicated to my beloved parents, Chengjun Wang and Qinglian Wang.
Abstract

This thesis is dedicated to develop fragment-based QM methods and QM/MM schemes:

For the first part, the fragment-based explicit polarization (X-Pol) method has been extended in three aspects: I. the inclusion of exchange repulsion terms in the X-Pol model is examined by antisymmetrizing the X-Pol Hartree-product wave function; this yields X-Pol with full exchange, called X-Pol-X; II. the original X-Pol method, where all fragments are treated using the same electronic structure theory, is extended to a multilevel representations, called multilevel X-Pol, in which different electronic structure methods are used to describe different fragments; III. a fragment-based variational many-body (VMB) expansion method is described to directly account for exchange repulsion, charge delocalization (charge transfer) and dispersion interactions in the X-Pol method.

For the second part, a universal QM/MM scheme, the projected hybrid orbital (PHO) method, is proposed to handle the covalent boundary at QM/MM interface at ab initio level with arbitrary basis sets. As an extension to the generalized hybrid orbital (GHO) method in which hybrid orbitals are constructed using the valence orbitals on the boundary atom, the PHO method further represents the core and valence electrons with a secondary, minimal basis set by projecting the original (primary) basis set used in the QM system. The PHO method is then validated on several aspects: geometry optimization, charge population and proton-affinity calculation. Comparison with
standard QM results shows that PHO is a robust and balanced QM/MM scheme that yields satisfactory structural, electronic, and energetic properties.
Contents

Acknowledgements .................................................................................................................. i
Dedication .............................................................................................................................. ii
Abstract .................................................................................................................................. iii
List of Tables ........................................................................................................................... vii
List of Figures .......................................................................................................................... x
Preface ....................................................................................................................................... xii

Chapter 1. Introduction .......................................................................................................... 1
  1.1 General Introduction ......................................................................................................... 1
  1.2 Overview of QM/MM Schemes ...................................................................................... 3
  1.3 Overview of Linear-scaling and/or Fragment-based QM Methods ......................... 5
  1.4 Organization of the Thesis ............................................................................................. 8

Chapter 2. On the Interfragment Exchange in the X-Pol Method ...................................... 12
  2.1 Introduction .................................................................................................................... 12
  2.2 Method .......................................................................................................................... 15
  2.3 Computational Details .................................................................................................. 22
  2.4 Results and Discussion ................................................................................................. 24
  2.5 Conclusions .................................................................................................................. 29

Chapter 3. Multilevel X-Pol: A Fragment-based Method with Mixed Quantum Mechanical Representations of Different Fragments ....................................................... 36
  3.1 Introduction .................................................................................................................... 36
## List of Tables

Table 2.1. Computed relative energies for the cyclic water trimer minimum structure (e-W₃) and for a symmetric trimer geometry (s-W₃) (kcal/mol). All energies are determined using the 6-31+G(d) basis set. .......................................................... 30

Table 2.2. Additivity of dimer contributions to the Coulomb and exchange-repulsion interaction energies in the cyclic water trimer minimum structure (c-W₃) and in a symmetric trimer geometry (s-W₃) (kcal/mol). All energies are determined using the 6-31+G(d) basis set .................................................................31

Table 3.1. Computed binding energies and energy components (kcal/mol) between acetic acid (A) and water (B) using iterative charge-updating optimization in multilevel X-Pol. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries ........................................................................................................57

Table 3.2. Computed binding energies and energy components (kcal/mol) between acetic acid (A) and water (B) using the variational multilevel X-Pol. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries. ........................................59

Table 3.3. Computed binding energies and energy components (kcal/mol) between H₅O₂⁺ and (H₂O)₄ using the iterative charge-updating optimization and variational multilevel X-Pol methods. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries..........................................................60

Table 4.1. Binding energies (−ΔEₚ) in kcal/mol from standard methods, X-Pol, and VMB2 calculations, mean unsigned deviations (MUD), and root-mean-square deviations (RMSD). The 6-311G(dp) basis set is used in all calculations........................................73
Table 4.2. Computed two-body variational energy correction ($\Delta E_2$), pairwise charge transfer ($\Delta E_C^2$), exchange ($\Delta E_X^2$) and dispersion ($\Delta E_D^2$), and monomeric exchange ($\Delta E_{1\text{X}}^X$) and total dispersion energy ($\Delta E_{\text{MP2}}^D$) along with delocalization effects on pair additivity in exchange ($\Delta \Delta E_X^{\text{nadd}}$) and dispersion ($\Delta \Delta E_D^{\text{nadd}}$). All energies are given in kcal/mol. The 6-311G(dp) basis set is used in all calculations.

Table 5.1. Mean unsigned errors (MUE) of bond lengths (Å) and bond angles (deg) of ethane between results from the hybrid QM/MM optimizations and the corresponding full QM calculations with (PHO) and without inclusion (w/o) of the tetrahedral restraining potential, $E_B^{\text{tet}}$. For each basis set listed, the MUE is averaged over HF, B3LYP and M062X calculations. The subscript B indicates that the atom is associated with the boundary atom in the MM region.

Table 5.2. Mean unsigned errors (MUE) in bond length (Å) and bond angle (deg) between hybrid PHO-QM/MM and full Hartree-Fock optimizations for a set of 20 small compounds. Q indicates an atom present in the QM region, M is an atom treated classically by an MM model, and $C_A$ and $C_B$ are the frontier atoms in the QM and MM regions.

Table 5.3. Comparison of bond length $C_A$-$C_B$ of selected molecules between PHO and standard HF results at different basis sets. $C_b$ is the boundary atom, and atoms to the right of $C_b$ belong to the MM region.

Table 5.4. Proton affinities for carboxylate anions, alcoholic anions and amines at HF/6-31G(d) level. All energies are in kcal/mol. The symbol $C_b$ indicates the boundary atom.
for GHO calculations. Atoms on the right of $C_b$ are treated quantum mechanically. PrB denotes projected basis, and GLO denotes global lowdin orthogonalization.

Table 5.5. Proton affinities for carboxylate anions, alcoholic anions and amines at HF/6-311+G(d,p) level. All energies are in kcal/mol. The symbol $C_b$ indicates the boundary atom for GHO calculations. Atoms on the right of $C_b$ are treated quantum mechanically. PrB denotes projected basis, and GLO denotes global lowdin orthogonalization.
List of Figures

Figure 2.1. Illustration of the minimum water trimer structure (c-W₃) in (a) and a symmetric configuration (s-W₃) in (b) ................................................................. 33

Figure 2.2. Electron density difference contour between the polarized and unpolarized systems, \( \Delta \rho = \rho(\hat{A} \psi_a \psi_b \psi_c) - \rho^o(\psi_a^o \psi_b^o \psi_c^o) \), for (a) the minimum-energy structure (c-W₃) and (b) the symmetric configurations (s-W₃). The structure orientations are identical to that shown in Figure 1 and the contour levels are at 0.003 au with red contours representing gain in electron density and blue for charge depletion ................................. 34

Figure 2.3. Computed Coulomb (X-Pol) and exchange repulsion (X-Pol-X) interaction energy as a function of the hydrogen bond distance between the donor hydrogen and the acceptor oxygen atoms in the water dimer complex with Cs symmetry. Energies are computed using the 6-31+G(d) basis at the fully optimized geometries under the Cs symmetry constraint at the HF/6-31+G(d) level. The interaction energy profile is obtained from full geometry optimized with the hydrogen-bond distance specified above held fixed at various values shown in the figure. ................................................................. 35

Figure 3.1. Schematic illustration of the optimized configuration of acetic acid and water using M06/MG3S................................................................. 61

Figure 3.2. Fragment partition of the H₂O₂⁺(H₂O)₄ cluster optimized at the M06/MG3S .... ................................................................. 62

Figure 4.1. Geometries for the four configurations of the water hexamer optimized with B3LYP/6-311++G(2d,2p)........................................................................ 75

Figure 4.2. Geometry of the water decamer structure optimized with the TIP5P potential..
Figure 5.1. Classification of methods for treating the boundary between QM and MM fragments in a combined QM/MM approach...............................110

Figure 5.2. Illustration depicting the difference between the LSCF model and the GHO method, both of which utilize hybrid orbitals.................................................111

Figure 5.3. Illustration of difference on Mulliken population analysis charges of $C_A$, $C_B$ and the “QM” methyl group between PHO and standard calculations across a wide range of methods and basis sets.................................................................112

Figure 5.4. Illustration of difference on CM5 charges of $C_A$, $C_B$ and the “QM” methyl group between PHO and standard calculations across a wide range of methods and basis sets.................................................................113

Figure 5.5. Torsional barriers about the C2-C3 bond in n-butane. Complete energy minimization was performed at each constrained value of the dihedral angle. PHO calculations are done at HF/3-21G level, with the boundary atom placed at the C2, C3 and C4, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.................................................................114

Figure 5.6. Torsional barriers about the C2-C3 bond in n-butane. Complete energy minimization was performed at each constrained value of the dihedral angle. PHO calculations are done at HF/6-31G(d) level, with the boundary atom placed at the C2, C3 and C4 carbons, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.................................115
Preface
This dissertation contains previously published work and is reproduced with permission.

Chapter 2:

Chapter 3:

Chapter 4:
Chapter 1. Introduction

1.1 General introduction

Computer modeling has emerged as an indispensable tool for biochemical research by aiding in quantitative interpretation of experimental observations and providing predictions when experiments are unavailable. The research interests cover a wide range of spatial and temporal resolution: the system size varies from a single atom, small molecules to membrane protein and virus capsule comprising hundreds of millions atoms; the time scale spans from femtoseconds within which light absorption takes place, to milliseconds even seconds during which protein folding occurs. Various simulation techniques have been developed for biomolecular modeling, and the cornerstone is the molecular mechanical force field (MMFF). It was first proposed in the 1940s to study steric effects of organic molecules, and was extended to model biomolecular systems by Lifson and coworkers in the 1960s. Since that time, significant progress has been made, and a number of force fields that can be applied to model various biomolecular systems have been established.

Although the widely used force fields differ in details (for example, some of them include coupling between internal coordinates), the functional forms used in MMFFs have remained essentially unchanged over the past half century, and the functional form depicted in eq 1 captures the essence of a typical MMFF potential energy function:
\[ V = \sum_{b}^{\text{bonds}} \frac{1}{2} K_b (R_{b} - R_{b}^0)^2 + \sum_{a}^{\text{angles}} \frac{1}{2} K_a (\theta_{a} - \Theta_{a}^0)^2 + \sum_{t}^{\text{torsion}} \sum_{n}^{\infty} \frac{V_n}{2} [1 + \cos(n \varphi_{t} - \varphi_{t}^0)] \]

\[ + \sum_{i < j} E_{ij} \left[ \left( \frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] + q_i q_j \frac{1}{R_{ij}} \]

where the first three terms account for bond stretching, bond bending, and torsional strain, respectively; the last two terms account for the nonbonded van der Waals interaction and Coulomb interaction, respectively. The importance of polarization has long been recognized and major current efforts in improving MMFFs are being devoted to the explicit inclusion of polarization terms of various forms to account for inductive forces.\(^{30-50}\)

Despite the successes of molecular mechanics,\(^{1,29,51}\) there are also a number of limitations: there is no general approach to treat the coupling of internal degrees freedom, the treatment of electronic polarization is difficult, intermolecular charge transfer is neglected, excited electronic states cannot be treated, and in the form usually employed the methods are inapplicable to chemical reactions.\(^{29}\) Because polarization, excitation and reactivity are intrinsically quantum mechanical phenomena, quantum mechanics (QM) emerges as a natural approach to remedy these problems.

Quantum mechanical electronic structure calculations can provide both reactive and nonreactive potential energy surfaces, including not only electrostatics and van der Waals forces but also polarization and charge transfer effects. A wide range of approximate quantum chemical model chemistries have been developed, including both wave function theory (WFT)\(^{52}\) and density functional theory (DFT)\(^{53}\), which provide accurate energetic and structural properties for molecular system < 100 atoms. However,
it is a daunting task (essentially impossible) to solve the Schrödinger equation for condensed-phase systems. Therefore, different approaches have been actively proposed to reduce the computational costs of applying electronic structural methods to large systems: a natural solution to tackle this challenge is to use a QM method for the chemically active region and combine it with MMFFs to represent the environment, resulting in a combined quantum mechanical and molecular mechanical (QM/MM) scheme; a second strategy is to develop linear-scaling and/or fragment-based QM methods to accelerate the most time-consuming steps, e.g., evaluation of full electron repulsion integrals and diagonalization of the Fock matrix for the entire system. These two different strategies will be briefly reviewed in the following two sections.

1.2 Overview of QM/MM schemes

The combined quantum mechanical and molecular mechanical (QM/MM) method was pioneered by the seminal works of Karplus\textsuperscript{54}, Warshel and Levitt\textsuperscript{55} in 1970s, and has been established as a valuable tool for studying large biomolecular systems nowadays.\textsuperscript{56-61} The whole system is partitioned into a QM region that is treated quantum mechanically and an MM region that is represented by MMFFs. Therefore, the total Hamiltonian for the whole system is written as

\[ \hat{H}_{\text{tot}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \]  

(2)

where \( \hat{H}_{\text{QM}} \), \( \hat{H}_{\text{MM}} \), and \( \hat{H}_{\text{QM/MM}} \) represent the Hamiltonian for the QM region, Hamiltonian for the MM region, and the interactive Hamiltonian between the two regions, respectively. The former two terms are trivial, whereas the last term, the interactive Hamiltonian, represents the main challenge of coupling QM and MM methods. Two
primary concerns arise for the description of accurate QM/MM interaction at the interface: treatment of the electrostatic contribution and treatment of the covalent boundary.\textsuperscript{59-62}

A hierarchy of schemes is available for handling the electrostatic coupling between the electron density of the QM region and the point charges of the MM region. The simplest model is the mechanical embedding scheme, where both QM and MM regions are represented by fixed MM charges. Both regions are not polarized in this case, leading to inaccurate treatment for very polar environments. A better solution is the electrostatic embedding scheme that allows polarization in the QM region, because MM point charges are incorporated in the one-electron integral in the QM Hamiltonian. Most of current QM/MM methods use the electrostatic embedding scheme. Nevertheless, two major limitations are prominent: the MM charges placed in the immediate proximity to the QM electron density will cause overpolarization especially when the boundary passes a covalent bond; MM charges are fixed, and thus the MM region is not polarized. Therefore, the most sophisticated scheme is the polarized embedding scheme that allows the explicit mutual polarization between QM and MM regions. This can be achieved by employing general-purpose polarizable MMFF\textsuperscript{40,63} in the QM/MM scheme. Alternatively, the quantum mechanical force field that includes polarization effect naturally in the self-consistent-field (SCF) calculation,\textsuperscript{64-67} represents a new paradigm of polarization embedding model.

Another issue central to the accuracy of QM/MM interaction is proper treatment of the covalent boundary that separates a QM region from the remainder MM region.
This issue is mainly tackled by two categories of approaches: The first category is the link atom scheme and its several variants, in which a hydrogen atom, a fluorine atom, or a pseudobond, is added to the daggling bond of the QM region to saturate the valence.\textsuperscript{56,57,68-72} However, the link atom introduces additional degrees of freedom to the system, and the total energy expression needs to be revised. Besides, the charges on the link atom and the MM atoms close to the boundary are adjusted accordingly to avoid overpolarization at the interface. The second category is the localized-orbital scheme that places hybrid orbitals at the boundary to saturate the daggling bond. Examples of this category are local self-consistent field (LSCF)\textsuperscript{73-76} and the generalized hybrid orbital (GHO) method\textsuperscript{77-80}, and they differ in the number of frozen orbitals that are excluded from the SCF optimization. These methods do not introduce additional degrees of freedom, and they do not change the partial charges assigned to the MM region, thereby best maintaining the electrostatic interactions in the original system.

\subsection{1.3 Overview of Linear-scaling and/or Fragment-based QM methods}

Beyond QM/MM methods, linear-scaling and/or fragment-based QM methods are extensively developed to resolve the computational bottleneck of conventional electronic structural methods. Linear-scaling algorithms linearize the computational cost of a given method with the system size, such as the divide and conquer method,\textsuperscript{81} the linear scaling density functional and post-Hartree-Fock calculations.\textsuperscript{82,83} More efforts have been devoted to develop fragment-based methods based on electronic structure theories,
mainly at HF or DFT level,\(^6\)\(^4\)\(^{-6}\)\(^6\)\(^,\)\(^8\)\(^4\)\(^{-10}\) in which the whole system is partitioned into smaller parts that are called fragments, monomers or subgroups. These schemes can be further divided into two categories: density-based approaches and energy-based approaches. In the density-based schemes, the approximate electron density of the whole system is built from the density matrix of individual fragments, followed by the calculation of the energy. Examples of this category include the molecular tailoring approach (MTA),\(^8\)\(^4\) the adjustable density matrix assembler approach (ADMA),\(^8\)\(^8\) and the molecular fractionation with conjugate caps approach (MFCC).\(^9\)\(^0\) In the energy-based schemes, the total energy of the system is directly approximated as a linear combination of fragment energies. Examples of this category are the fragment molecular orbital (FMO) method,\(^8\)\(^6\) the explicit polarization (X-Pol) method,\(^6\)\(^4\)\(^{-6}\) the electrostatically embedded many-body (EE-MB)\(^9\)\(^2\) and electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE) method,\(^9\)\(^3\) and the multilevel fragment-based approach (MFBA).\(^9\)\(^6\) Compared with the density-based schemes, the energy-based schemes are more straightforward and easier for implementation in the existing quantum chemistry software package, and avoid the expensive step of calculating the energy from the density of the whole system.

Another general consideration for all the fragment-based methods is the treatment of fragment interactions. The inclusion of long-range electrostatic interaction is critical to account for the effect of other fragments in the system. The prototypical model of self-consistent fragment calculations was proposed by Otto and Ladik,\(^1\)\(^0\)\(^3\) in which every fragment is embedded in the Coulomb field of other fragments without the corresponding
exchange repulsion interaction. The idea was developed and integrated as a part of the modern method that treats mutual polarization effect explicitly, such as FMO\textsuperscript{86} and X-Pol.\textsuperscript{64-66} Besides electrostatic interaction and polarization, the remaining energy components, such as exchange repulsion and dispersion, can be compensated via two approaches: the perturbative treatment in the symmetry-adapted perturbation theory\textsuperscript{104,105} in which various fragment interaction terms are expressed as different orders of correction; the many-body expansion scheme in which pairs, triples or large unions of fragments are computed to account for the full fragment interaction.\textsuperscript{92,93,99,106}

The explicit polarization (X-Pol) method, an energy-based scheme and the focus of this thesis, was initially called molecular orbital derived potentials for liquids (MODEL),\textsuperscript{64} and was introduced in 1997 for the treatment of macromolecular systems in statistical mechanical Monte Carlo and molecular dynamics simulations using electronic structure theories.\textsuperscript{65,66,107-109} It is based on the partition of a whole system into subsystems, which are called fragments or blocks, each of which is represented by an antisymmetrized wave function. The system wave function is then a Hartree product of fragment wave functions, which are optimized by the self-consistent field (SCF) method in the presence of the external field of all other blocks until the energy or electron density of the entire system is converged.\textsuperscript{64-66} The short-range exchange repulsion interactions between fragments, the long-range dispersion interactions between different fragments, and the interfragment correlation energy are neglected in the Hartree product approximation, but are modeled empirically as in molecular mechanics.\textsuperscript{64,109,110} Compared to various QM/MM schemes, the mutual polarization effect among different fragments is treated
explicitly and consistently in the X-Pol method. Furthermore, the treatment of molecular polarization, as well as other energy component, can be systematically improved by using ab initio Hartree-Fock (HF), Kohn-Sham (KS) density functional theory (DFT), or correlation methods, such as perturbation, multiconfiguration, and couple cluster theories.

1.4 Organization of the thesis

This thesis consists of two parts of work in developing fragment-based QM methods and QM/MM schemes: from Chapter 2 to Chapter 4, the X-Pol model is further extended in several aspects; in Chapter 5, a universal QM/MM scheme, the projected hybrid orbital, is proposed to treat the covalent boundary at the QM/MM interface.

In Chapter 2, the inclusion of exchange repulsion terms in the X-Pol model is examined by antisymmetrizing the X-Pol Hartree-product wave function; this yields X-Pol with full eXchange, called X-Pol-X. When the monomers are treated by HF theory, this calculation can be accomplished by using the formalism of block-localized wave functions that has been used in a variety of applications. In this case the block-localized structure in the X-Pol-X wave function allows for decomposition of the full Fock matrix of a dimension of $M$ blocks into $M$ smaller Fock matrices. The method is illustrated by considering two trimer structures of water clusters, and it is found that the total exchange repulsion energies in these hydrogen-bonding test cases are adequately treated and—to a good approximation—are pairwise additive. We also present a formalism to yield a simplified Fock matrix by making use of the neglect of interfragment differential overlap
(NIDO) approximation, which is less severe than the neglect of diatomic differential overlap (NDDO) approximation.

In Chapter 3, the original X-Pol method, where all fragments are treated using the same electronic structure theory, is extended to multilevel representations, called multilevel X-Pol, in which different electronic structure methods are used to describe different fragments. The multilevel X-Pol method has been implemented into Gaussian 09. A key ingredient that is used to couple interfragment electrostatic interactions at different levels of theory is the use of the response density for post-self-consistent-field energy (The response density is also called the generalized density). The method is useful for treating fragments in a small region of the system such as the solute molecules or the substrate and amino acids in the active site of an enzyme with a high-level theory, and the fragments in the rest of the system by a lower-level and computationally more efficient method. The method is illustrated here by applications to hydrogen bonding complexes in which one fragment is treated with the M06 hybrid density functional, Mølter-Plesset perturbation theory, or coupled cluster theory, and the other fragments are treated by HF theory or the B3LYP or M06 hybrid density functionals.

In Chapter 4, a fragment-based variational many-body (VMB) expansion method is described to directly account for exchange repulsion, charge delocalization (charge transfer) and dispersion interactions in the X-Pol method. The present VMB/X-Pol approach differs from other fragment molecular orbital (FMO) techniques in two major aspects. First, the wave function for the monomeric system is variationally optimized using the standard X-Pol method, as opposed to the iterative update procedure adopted in
FMO. Second, the mutual polarization in the dimeric terms is also variationally determined, whereas single-point energy calculation of the individual dimers embedded in a static monomer field is used in FMO. The second-order (two-body) VMB (VMB2) expansion method is illustrated on a series of water hexamer complexes and one decamer cluster, making use of HF theory, MP2, and the PBE1 and M06 density functionals to represent the monomer and dimer fragments. The computed binding energies are within 2 kcal/mol of the corresponding results from fully delocalized calculations. Energy decomposition analyses reveal specific dimeric contributions to exchange repulsion, charge delocalization and dispersion. Since the wave functions for one-body and all two-body terms are variationally optimized in VMB2 and X-Pol, it is straightforward to obtain analytic gradient without the additional coupled-perturbed HF step. Thus, the method can be useful for molecular dynamics simulations.

In Chapter 5, the projected hybrid orbital (PHO) method is developed to treat the covalent boundary at the QM/MM interface. Both the energy and analytic gradient of PHO have been implemented in Gaussian Developmental Version (GDV-H35). As an extension to the GHO method in which hybrid orbitals are constructed using the valence orbitals on the boundary atom, the PHO method represents the core and valence electrons with a secondary, minimal basis set by projecting the original (primary) basis set used in the QM system. Then, the projected valence orbitals are transformed into a set of hybrid orbitals, defined in exactly the same way as that used in the GHO method. Since the minimal basis set used in the present projected hybrid orbitals (PHO) is a closest representation of the original basis set, it retains the essential properties to have a
balanced interaction with the QM fragment, and it is possible to obtain good results without system-dependent parameters in the present approach. The projected hybrid orbital method is implemented with two optimization procedures depending on the way the total Fock matrix is partitioned. The PHO method is then validated on various aspects: geometry optimization on ethane and a test set of 20 small organic compounds yields satisfactory structural and electronic properties as compared with the reference QM results; calculation of proton affinity on a set of acid anions, alcohol anions, and amine molecules shows good performance of PHO on energetics. In conclusion, the PHO method represents as a universal and robust approach to treat QM/MM covalent boundary.
Chapter 2. On the Interfragment Exchange in the X-Pol Method

2.1 Introduction

The explicit polarization (X-Pol) method, initially called molecular orbital derived potentials for liquids (MODEL), was introduced in 1997 for the treatment of macromolecular systems in statistical mechanical Monte Carlo and molecular dynamics simulations using electronic structure theory. It is based on the partition of a whole system into subsystems, which are called fragments or blocks, each of which is represented by an antisymmetrized wave function. The system wave function is then a Hartree product of fragment wave functions, which are optimized by the self-consistent field (SCF) method in the presence of the external field of all other blocks until the energy or electron density of the entire system is converged. The individual wave functions of the subsystems can be obtained at any level of theory—ab initio Hartree-Fock (HF), semiempirical molecular orbital theory, correlated wave function theory, or Kohn-Sham (KS) density functional theory (DFT). In general, the X-Pol method can be applied to post-SCF calculations such MP2 and multiconfigurational methods, and a post-SCF treatment can be used to estimate dispersion interaction energies between different fragments. In contrast to the methods of Stoll and Preuss (whose emphasis was on the localization of orbitals on atomic centers with the neighboring bonding partners and lone-pairs) and Yang (whose emphasis was on efficient full calculations for large systems by a divide-and-conquer approach), X-Pol is not designed to reproduce the full HF or KS
energy; rather it is intended to establish a theoretical framework for economical and accurate applications to macromolecular systems.\(^3,5\)

The Hartree-product approximation of the X-Pol wave function implies that interfragment exchange and correlation interactions are neglected.\(^1,3,6,8\) The most noticeable contributions to the interfragment exchange term are from the repulsion interactions at geometries where orbitals on different blocks have significant overlap (short-range), whereas the dominant contributions to interfragment correlation energy are dispersion interactions at geometries where they have negligible overlap (long-range). To achieve computational efficiency, we have proposed to use empirical Lennard-Jones terms to account for these interactions in the spirit of developing a quantal force field for macromolecular simulations.\(^1,3,5-6\) Alternatively, interfragment exchange can be included in the X-Pol method by fully antisymmetrizing the block-localized molecular orbitals (BLMOs) to form a determinant wave function, and this is equivalent to the block-localized wave function (BLW) method described by Mo,\(^13\) which has been applied to a variety of problems.\(^9,14-25\) In the present context, this approach will be called X-Pol with full eXchange or X-Pol-X. In such a wave function the orbital coefficient matrix is block-diagonal, which has the property that orbitals within each block are orthogonal, and orbitals between different subsystems are non-orthogonal.\(^11,13-14,26\) The latter is characteristic of valence bond theory.\(^9,16,23,27\) Methods for treating non-orthogonal determinantal wave functions have been described a long time ago and used in various applications, including the work of King et al.,\(^28\) Gallup,\(^29\) Stoll,\(^11,30\) Gianinetti and Raimondi,\(^31-32\) Mo et al.,\(^13-16,27\) Head-Gordon and coworkers,\(^33\) and Cembran et al.\(^9\) The
BLW method has been used to define valence bond-like states in the mixed molecular orbital and valence bond (MOVB) theory\textsuperscript{16,22-23,27} and in the multistate density functional theory (MSDFT) based on valence bond or VB-DFT.\textsuperscript{9} Here, we use the X-Pol-X (\textit{i.e.}, BLW) method to elucidate the magnitude of the exchange repulsion term that is required in the X-Pol method.\textsuperscript{15} In the Appendix we derive a simplified X-Pol-X method by invoking neglect of interfragment differential overlap (NIDO).

It should be noted that in 1999, Kitaura et al.\textsuperscript{34-35} used the wave functions of dimeric fragments in a post SCF correction to estimate the pair-wise exchange repulsion. The expansion of the monomeric subspace into two fragments also includes pair-wise charge transfer effects which are important for describing short-range interactions.\textsuperscript{10} This approach was called the fragment molecular orbital (FMO) method, and it has been used in numerous applications with various electronic structural models.\textsuperscript{36} However, aside from the dimer correction term, first described by Stoll and Preuss,\textsuperscript{10} the construction of block-localized (fragment) orbitals and the SCF computational algorithm (see page 321 of ref. 34) of the FMO method\textsuperscript{34-35} are essentially the same as that (see page 660 of ref 2) in X-Pol theory.\textsuperscript{2-3} Nevertheless, these studies illustrate the usefulness of block localization of molecular orbitals for macromolecular systems.

Section 2.2 presents the X-Pol method in a general way and then describes the X-Pol-X extension of the X-Pol method to include the exchange repulsion energy in SCF calculations. Section 2.3 describes an illustrative application at the ab initio Hartree-Fock level. Section 2.4 gives the results and discusses them, with special emphasis on comparison of X-Pol to X-Pol-X. Section 2.5 summarizes the conclusions. The NIDO
simplification procedure, which may be very useful in future work, is presented in Appendix A.

2.2 Method

We first describe the X-Pol method. Then, we present the SCF procedure for optimization of an X-Pol-X wave function. It is especially interesting to consider the case where the fragment wave functions are Hartree-Fock wave functions, obtained either by ab initio theory or by semiempirical molecular orbital theory. Then the fragment wave functions are optimized by the self-consistent field (SCF) method in the presence of the external field of all other blocks. This is sometimes called the double SCF (DSCF) procedure because one must achieve self-consistency both within each fragment and among all blocks in the system. As a consequence of the interfragment self-consistency, the fragment wave functions are polarized.

2.2.1 Block localization and the X-Pol wave function

We partition a large molecular system into $M$ blocks (which have been equivalently and interchangeably called subgroups, subsystems, molecules, residues, monomers, fragments, or bodies). The term block is more general than some other terms, such as fragment and body, in that orbitals on the same atom can be divided into different subsystems without the physical constraint of atomic structure fragmentation. Block $a$ contains $k_a$ basis functions and $n_a$ electrons. Thus, the total number of primitive basis functions, $K$, and the total number of electrons, $N$, in the system are:
The molecular orbitals $\varphi^a_j$ in each block $a$ are written as linear combinations of the primitive basis functions $\left\{ \chi^a_{\mu}, \mu = 1, \cdots, k_a \right\}$ in that specific subspace:

$$
\varphi^a_j = \sum_{\mu=1}^{k_a} c^a_{j\mu} \chi^a_{\mu}
$$

(2)

Thus, by construction, the molecular orbitals are strictly localized within the subspace of each block.

The X-Pol wave function is approximated as a product of the individual, antisymmetric fragment wave functions of all $M$ blocks:

$$
\Psi^{\text{X-Pol}} = \prod_{a=1}^{M} \hat{A} \Phi_a
$$

(3)

where $\hat{A}$ is an antisymmetrization operator, and $\Phi_a$ is a product of the occupied spin-orbitals in block $a$ (eq 2):

$$
\Phi_a = \varphi^a_1 \varphi^a_2 \cdots \varphi^a_n_a
$$

(4)

The Hartree-product wave function in eq 3 includes Coulomb interactions, but the exchange interactions are ignored between different blocks. If we make a further approximation by representing the Coulomb potentials by a classical multipole expansion or by the potential due to distributed monopoles calculated by Mulliken population analysis, the computation of the Fock matrix for block $a$, in the presence of the external field of the rest of the system, is simplified to
\[ F^a = H^a + \sum_i (J^a_i - \frac{1}{2} K^a_i) + \sum_{b \neq a} V^a_b \]  \hfill (5)

where \( H^a \) is the one-electron Hamiltonian matrix, \( J^a_i \) and \( K^a_i \) are the Coulomb and exchange integral matrices associate with molecular orbital \( i \) in block \( a \), and the last summation term is the Coulomb interaction matrix due to the nuclei and wave functions of all other blocks in the system. This is a level comparable to a force field for macromolecular systems.\(^{1-2,6,8}\) The Fock matrix \( F^a \) depends on the instantaneous polarization of the wave functions of other blocks through \( V^a_b \), which, in turn, depends on \( F^a \). The electronic energy of the full system is obtained by the DSCF method. We have presented a nonvariational version of the Fock operator to illustrate the main ideas as simply as possible, which is sufficiently accurate for use in statistical mechanical Monte Carlo simulations without the need for computing analytical gradients,\(^{2-4}\) and this approach has been adopted in other applications.\(^{34-36}\) A variational version that can be used to calculate analytic gradients, suitable for geometry optimization and molecular dynamics simulations,\(^{5}\) is presented elsewhere.\(^{7-8}\) The final SCF energy is called the BL term or \( E^{SCF} \).

We note that the electrostatic potential \( V^a_b \), which is derived from the wave function of block \( b \), can be calculated in more than one way. For example, it can be determined by integration over the BLMOs in that block, it can be represented by fitted atomic densities, or it can be approximated by electrostatic-potential-fitted charges or by
a multipole expansion. These choices may be dictated by the context in which X-Pol is used. In previous applications, we have used charges obtained by Mulliken population analysis to approximate the Coulomb potentials of each of the monomers. In the present work, we use the explicit evaluation of all two-electron Coulomb integrals to account for inter-block electrostatic interactions. Therefore, the calculation of Coulomb interactions between different monomer blocks does not introduce approximations here, and this makes it easier to evaluate the new theoretical approach to exchange repulsion.

The short-range exchange repulsion energies between different blocks can be approximated empirically by Lennard-Jones terms, $\varepsilon_{ab}^{LJ}$, which also account for long-range dispersion interactions. Thus, the total X-Pol energy of the system is

$$E^{X-Pol} = E^{SCF} + \frac{1}{2} \sum_a \sum_{b\neq a} \varepsilon_{ab}^{LJ}$$  

(6)

2.2.2 X-Pol-X: Block-localized wave function that accounts for exchange interactions

The use of the $R^{-2}$ repulsive terms in the Lennard-Jones potential (where $R$ is a distance between atoms in different monomers) to approximate the short-range exchange repulsion significantly reduces the computational cost needed to solve the Hartree-Fock-Roothaan equations for the entire block-localized (i.e., fragmental) system. Alternatively, the exchange energies can be determined without the use of an empirical potential by antisymmetrizing the X-Pol wave function, keeping the block-localized structure of the molecular orbitals.
\[ \Psi^{\text{X-Pol-X}} = \prod_{a=1}^{M} \Phi_a \]  

(7)

The coefficient matrix of the corresponding orbitals has the following block-diagonal form:

\[
C = \begin{pmatrix}
C^1 & 0 & \ldots & 0 \\
0 & C^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & C^k
\end{pmatrix}
\]  

(8)

where \( C^a \) is an \( k_a \times k_a \) matrix, with the orbital coefficients of eq 2 arranged as column vectors. As mentioned in the introduction, this is equivalent to constructing a block-localized wave function (BLW) using the block-localized molecular orbitals (BLMO), and the computational procedure has been described by several authors in various contexts, dating back to the work of Stoll.

An alternative to the expression of eq 7 is to use the antisymmetrized function of antisymmetrized block-localized wave functions of eq 3. However, in the present case, the molecular orbitals are strictly block-localized within each fragment space, and the difference from eq 7 is likely to be minimal (antisymmetrying determinant states is useful if each of the block-localized fragment is used as a reference to form multiconfigurational states). In addition, the present test results show that the exchange repulsion energy is adequately treated, suggesting that eq 7 in the current form is a good choice for the purpose of this study.
The molecular orbitals in eq 7 are subject to the constraint that orbitals within the same subgroup are orthogonal, whereas orbitals in different subgroups are non-orthogonal:

$$\langle \varphi_i^a \mid \varphi_j^b \rangle = \begin{cases} S_{ij}^{ab}, a = b \\ S_{ij}^{ab}, a \neq b \end{cases}$$  \hspace{1cm} (9)$$

where $S_{ij}^{ab}$ is the overlap integral between molecular orbitals $i$ and $j$ from blocks $a$ and $b$, respectively, and is given by

$$S_{ij}^{ab} = \left( c_i^a \right)^T R^{ab} c_j^b$$  \hspace{1cm} (10)$$

where $\{ c_i^a, i = 1, \ldots, n_a, a = 1, \ldots, k_a \}$ are the columns of $C^a$ in eq 8 (the component of these vectors are $\{ c_{i\mu}, \mu = 1, \ldots, k_a \}$), where $T$ denotes a transpose, and where $R^{ab}$ is the block of the overlap matrix $R$ corresponding to basis functions in fragments $a$ and $b$:

$$R_{\mu\nu}^{ab} = \left\langle \chi_\mu^a \mid \chi_\nu^b \right\rangle$$  \hspace{1cm} (11)$$

The orthogonality constraint on orbitals in the same block does not affect the total energy of the system because the energy is invariant to a unitary transformation of the orbitals within a block.

The density matrix is defined, using the occupied orbitals, which include the column vectors $\{ c_i^a \}$ for block $C^a$ in eq 8, by

$$D = C (C^T R C)^{-1} C^T$$  \hspace{1cm} (12)$$
The Fock matrix, $F$, has an identical form to that in Hartree-Fock theory, in which an element, in the atomic basis $\{ \chi_\mu \}$, is given for a closed-shell system as\textsuperscript{9,16,27}

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma}^K D_{\lambda\sigma} \left\{ (\chi_\mu \chi_\nu | \chi_\lambda \chi_\sigma) - \frac{1}{2} (\chi_\mu \chi_\lambda | \chi_\nu \chi_\sigma) \right\}$$  \hspace{1cm} (13)

where $H_{\mu\nu}$ is an element of the one-electron matrix representing kinetic energy and nuclear attraction, and the brackets are the Coulomb and exchange integrals in the Mulliken notation. The total X-Pol-X energy is

$$E^{X\text{-Pol}X} = \sum_{\mu\nu}^K D_{\mu\nu} \left\{ H_{\mu\nu} + F_{\mu\nu} \right\}$$  \hspace{1cm} (14)

In comparison with the X-Pol SCF energy $E^{\text{SCF}}$ (eq 6), the exchange energy included in the X-Pol-X expression (eq 14) originates from the orthogonalization constraint $(C^T R C)^{-1}$ in eq 12.\textsuperscript{15,40}

Since the orbital coefficient matrix is block diagonal (eq 8), the $K \times K$ Fock matrix can be reduced to smaller sizes, one for each block ($k_a \times k_a$) by transforming the original Fock matrix into a block-diagonal form using a projection operator.\textsuperscript{9,11} Thus, for block $a$,

$$\mathcal{F}^{aa} = F^{aa} + V_2 + V_3 + V_4$$  \hspace{1cm} (15)

where

$$V_2 = -\sum_{b \neq a}^k F^{ab} \left( \sum_{c \neq a}^k D^{bc} R^{ca} \right)$$  \hspace{1cm} (16)
\[
V_3 = - \sum_{b \neq a}^k \left( \sum_{c \neq a}^k D^{bc} R^{ca} \right)^T F^{ba} \tag{17}
\]

and

\[
V_4 = \sum_{d \neq a}^k \sum_{b \neq a}^k \left( \sum_{c \neq a}^k D^{bc} R^{ca} \right)^T F^{bd} \left( \sum_{c \neq a}^k D^{dc} R^{ca} \right) \tag{18}
\]

The upper indices in eqs 15–18 are used to indicate that the dimension of each matrix with such indices is determined by the number of basis functions in that block (see eq 1). Note that eq 15 is general in that it can be used both in Hartree-Fock\textsuperscript{13-16,27} and DFT calculations.\textsuperscript{9} In the latter case, the exchange potential in Hartree-Fock theory is replaced by the exchange-correlation potential.\textsuperscript{9}

### 2.3 Computational Details

All computations have been performed using a locally modified GAMESS program\textsuperscript{41} and the Xiamen University Valence Bond (XMVB) program.\textsuperscript{42} The minimally augmented polarized valence double-zeta basis set 6-31\textsuperscript{+}G(d) was used for all calculations. All X-Pol-X calculations are fully self-consistent, that is, the orbitals are optimized in the presence of interfragment exchange.

We present illustrative calculations for two examples, both of which are structures of water trimer. First, we chose the minimum energy configuration of a cyclic water trimer structure, \textbf{c-W}_3, in which each water molecule donates and accepts one hydrogen
bond from one of the other two water monomers. All three pairwise interactions are stabilizing. For the second structure, we constructed a trimer complex structure that has a significantly Coulomb repulsive interaction as follows. We first optimized the dimer complex structure, and then we placed the third water molecule at the $C_2$ image position of the donor water molecule about the bisection axis of the acceptor water molecule. This complex is denoted as the symmetric trimer, $s$-$W_3$. Both structures are depicted in Figure 2.1.

In the following discussion, each water monomer in the trimer complexes is treated as a single block or fragment, and the internal geometries of the monomers are kept the same as in the fully optimized configuration at the HF/6-31+G(d) level for $c$-$W_3$ and the same as constructed above for $s$-$W_3$. The X-Pol-X wave function is the three-block localized system $\hat{A} \psi_d \psi_b \psi_c$. The energy and energy components of the X-Pol-X wave function will be compared to those for the unrelaxed wave function $\hat{A} \psi_d^0 \psi_b^0 \psi_c^0$, where $\psi_w^0$ is the wave function of a monomer in the gas phase (unpolarized wave function) at the geometry in the trimer complex.

The energy components we consider are the Coulomb energy, the exchange energy, the polarization energy, and the charge transfer energy. The Coulomb and exchange components are the same in all energy decomposition analysis (EDA) methods, including the approach of Kitaura and Morokuma, whereas the polarization and charge transfer terms are defined in the block-localized wave-function energy decomposition (BLW-ED) method developed by Mo et al.
2.4 Results and Discussion

The main goal of the present article is to show that the theory presented in Section 2 can provide a reasonable estimate of the exchange repulsion energy in the X-Pol method.

Table A1 (see Appendix A) lists the total energies and energy components determined for the fully separated monomers and for the trimer complexes as calculated from \( \hat{A}\psi_a\psi_b\psi_c \) and from \( \hat{A}\psi^0_a\psi^0_b\psi^0_c \) and also by full Hartree-Fock calculations. In this discussion we shall focus on interaction energies, defined as the energy of the trimer minus the sum of the energies of the three separated monomers, each with the same geometry that it has in the trimer. We shall also discuss the components of the interactions. The interaction energies and their components are given in Table 2.1; all values in Table 2.1 were computed from the absolute energies in Table A1.

Row 1 of Table 2.1 gives the X-Pol-X interaction energy, and the following rows of the table report quantities that help us dissect this interaction energy. Row 2 is calculated using \( \psi^0_a\psi^0_b\psi^0_c \), and row 3 is calculated using \( \hat{A}\psi^0_a\psi^0_b\psi^0_c \). Row 2 gives the unpolarized Coulomb contribution to the X-Pol or X-Pol-X binding energy. The results in Table 2.2 show that the Coulomb interaction to the total binding energy is quite large in both water trimer complexes. In particular, with the monomers unpolarized, the total Coulomb interaction energies are -25.6 and -16.0 kcal/mol for \( \text{c-W}_3 \) and \( \text{s-W}_3 \) complexes, whereas row 3 shows that the exchange-repulsion energies are 16.3 and 10.8 kcal/mol, respectively. The sum of Coulomb and exchange energies is often called the total...
unpolarized electrostatic energy due to the charge density of each monomer in the gas phase alone; this is given in row 4, and it amounts to -9.3 and -5.2 kcal/mol, respectively, for c-W$_3$ and s-W$_3$. The water trimer has been extensively studied, but they are not listed here since our goal is not to investigate the water clusters. In one energy decomposition calculation using the reduced variational space (RVS) method, which is analogous to the BLW-ED process, Chen and Gordon found that the total electrostatic interaction energy is -9.0 kcal/mol using the 6-31++G(d,p) basis set at the minimum configuration.

Next we consider the analogous breakdown of the X-Pol-X interaction energies and interaction energy components. For these calculations, the wave function is $\hat{A}\psi_d\psi_b\psi_c$ where the antisymmetrization is performed before the orbitals are optimized, so that they are optimized in the presence of interfragment exchange. Now, as compared to rows 2–4, there is an additional component of the total interaction energy, namely the increase in the internal energy of the monomers due to the polarization of the orbitals. This is given in row 5, the Coulomb energy is in row 6, and the exchange energy is in row 7. The sum of rows 5–7 gives the X-Pol-X interaction energy in row 1. Note that row 4 is the total energy of the antisymmetrized fragment product wave function where the orbitals are the unpolarized monomer orbitals, whereas row 1 is the total energy for the X-Pol-X wave function, in which the mutual charge and exchange polarizations of the monomer charge density are included.

Mutual polarization of the block-localized monomer wave functions in the presence of other monomers enhances Coulomb interactions by -6.2 and -2.3 kcal/mol, but reduces the exchange repulsion slightly. Taking into account of the energy costs for
polarizing the monomer wave functions in the trimer complex and the small decrease in exchange repulsion due to polarization effects, we see that the overall polarization, calculated at this level, lowers the e-W₃ interaction energy from −9.3 to −12.5 kcal/mol and the s-W₃ interaction energy from −5.2 to −6.7 kcal/mol. The differences of these quantities are listed in row 8, which shows a net X-Pol polarization energy of -3.2 and -1.5 kcal/mol, or 26% and 22%, for the two complexes, respectively. The electronic polarization effects in these two water trimer complexes are illustrated in Figure 2.2, which depict electron density shifting away for the hydrogen atoms that donate hydrogen bonds to the acceptor oxygen atoms. Concomitantly, electron density is attracted towards regions where oxygen atoms bind the proton donor. It is of interest to notice that in the s-W₃ structure (Figure 2.2b), electron density is reduced from both hydrogen atoms in the W1 water molecule thanks to block-localization that disallows charge transfer between different water molecules, although they are not directly donating any hydrogen bonds. In the study of Chen and Gordon, polarization effects were found to contribute -2.7 kcal/mol with the slightly larger 6-31++G(d,p) basis set for e-W₃.⁴⁶

Row 10 depicts the interaction energy from a full (delocalized) Hartree-Fock calculation. The energy difference between the Hartree-Fock SCF result and the X-Pol-X energy for each trimer complex is defined as the charge-transfer energy, which is not included in the X-Pol-X method and will be addressed in a separate study. This charge transfer energy (row 9) is -3.0 and -2.1 kcal/mol using the 6-31+G(d) basis functions in the cyclic and symmetric complexes, respectively. For comparison, Chen and Gordon found a charge transfer energy of -1.7 kcal/mol using the RVS approach, smaller than the
present value.\textsuperscript{46} The BLW energy decomposition analysis (BLW-ED) method has been shown to be relatively stable in computed polarization and charge transfer energies with increased size of basis set.\textsuperscript{15,18,47}

It is interesting to further analyze the exchange repulsion energy of the X-Pol-X method. The numerical results in Table 2.1 show that the exchange repulsion is significant, reducing the Coulomb interaction energies by more than 50% in both cases. Table 2.2 shows the pair-wise contributions to the total Coulomb and exchange energies. The variation of Coulomb energies for the dimer interactions depicted in the first three rows in Table 2.2 suggests that stronger Coulomb attractions are accompanied by greater charge overlap and exchange repulsions. The Coulomb interaction energy is additive by definition in the energy decomposition analysis,\textsuperscript{15,40,44} but the exchange repulsion term is not necessarily additive due to orthogonalization of the product of nonorthogonal block localized orbitals. Nevertheless, Table 2.2 reveals that the exchange repulsion energies are additive to within 0.03 kcal/mol in both trimer complexes. This result holds even though one complex (\textit{e-W}_3) has three significant pairwise exchange repulsion interactions, while the other (\textit{s-W}_3) has only two. The finding of high additivity for exchange repulsion interactions suggests that it is sufficient to construct an X-Pol-X wave function that involves only pair-wise antisymmetrized blocks, which can greatly reduce the computational cost of the fully antisymmetrized system.

The dependence of exchange repulsion on intermolecular separation is depicted in Figure 2.3 for a water dimer as a function of the hydrogen-bond distance between the donor hydrogen and acceptor oxygen atoms. As is well-known, the exchange repulsion
decreases exponentially to a diminishing value as the hydrogen bond distance is elongated by more than 1 Å from the optimal geometry, well within the first solvation layer in liquid water. Figure 2.3 shows that for all practical purposes, one only needs to consider the explicit exchange repulsion for monomers within the first solvation shell, and there is no need to construct a fully antisymmetric X-Pol determinant wave function for the entire system. Further simplifications of the X-Pol-X method can be made by imposing the NIDO approximation presented in the Appendix.

Several topics suggest themselves for further work. First, it would be useful to examine more systems. Second, the method can be used to account for charge transfer effects. Third, the method may be examined by treating the monomers with density functional theory. Fourth, it would be interesting to examine the effect of computing the interfragment exchange terms of X-Pol-X in a post-SCF fashion, that is, optimizing the orbitals by X-Pol using a constant interfragment exchange based on the unpolarized fragmental wave functions and then using the optimized BLMOs to compute the exchange repulsion. Fifth, the NIDO approximation presented in the Appendix can be used with either full X-Pol-X calculations, including the use of full NDDO approximation in effective-Hamiltonian MOV method, or with post-SCF interfragment exchange. Finally, it would be interesting to use the interfragment exchange terms of X-Pol-X calculations as a basis for parameterizing more accurate analytical pairwise exchange repulsion terms.
2.5 Conclusions

In this article, we have examined the inclusion of exchange repulsion energies directly into the explicit polarization (X-Pol) model by antisymmetrizing the Hartree-product wave function in a way that is equivalent to the block-localized wave function (BLW) approach. In general, this may be called X-POL with full exchange or X-Pol-X. The block-localized structure in the X-Pol-X wave function allows for decomposition of the full Fock matrix of $M$ blocks into $M$ smaller Fock matrices, reducing the computational costs from $Q[(KM)^3]$ scaling to $Q[M(K)^3]$. Furthermore, the complexity of the Fock matrix can be reduced by making use of the neglect fragmental diatomic overlap approximation, a modification of the traditional neglect diatomic differential overlap (NDDO) method. Another way to reduce the cost would be to compute the exchange repulsion by a set of pairwise X-Pol-X dimer calculations, including the contributions of the nearby pairs. The computation of the exchange repulsion energy is illustrated by considering two trimer structures of water, and it was found that the total exchange repulsion energies in these simple, but general and strongly hydrogen-bonding cases are fully pair-wise additive.
Table 2.1. Computed relative energies for the cyclic water trimer minimum structure (c-W₃) and for a symmetric trimer geometry (s-W₃) (kcal/mol). All energies are determined using the 6-31+G(d) basis set.

<table>
<thead>
<tr>
<th>Row</th>
<th>Quantity</th>
<th>c-W₃</th>
<th>s-W₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\Delta E^{X-Pol-X}$</td>
<td>-12.5</td>
<td>-6.7</td>
</tr>
<tr>
<td>2</td>
<td>Coulomb energy with unpolarized monomer orbitals</td>
<td>-25.6</td>
<td>-16.0</td>
</tr>
<tr>
<td>3</td>
<td>Exchange energy with monomer orbitals</td>
<td>16.3</td>
<td>10.8</td>
</tr>
<tr>
<td>4</td>
<td>Total energy for monomer orbitals</td>
<td>-9.3</td>
<td>-5.2</td>
</tr>
<tr>
<td>5</td>
<td>X-Pol-X distortion energy</td>
<td>3.6</td>
<td>1.6</td>
</tr>
<tr>
<td>6</td>
<td>X-Pol-X Coulomb energy</td>
<td>-31.8</td>
<td>-18.3</td>
</tr>
<tr>
<td>7</td>
<td>X-Pol-X exchange energy</td>
<td>15.7</td>
<td>10.1</td>
</tr>
<tr>
<td>8</td>
<td>X-Pol-X polarization energy</td>
<td>-3.2</td>
<td>-1.5</td>
</tr>
<tr>
<td>9</td>
<td>Charge transfer energy</td>
<td>-3.0</td>
<td>-2.1</td>
</tr>
<tr>
<td>10</td>
<td>Full Hartree-Fock</td>
<td>-15.5</td>
<td>-8.8</td>
</tr>
</tbody>
</table>

*a* All energies are relative to the sum of the unpolarized monomer energies.
Table 2.2. Additivity of dimer contributions to the Coulomb and exchange-repulsion interaction energies in the cyclic water trimer minimum structure (c-W₃) and in a symmetric trimer geometry (s-W₃) (kcal/mol). All energies are determined using the 6-31+G(d) basis set.

<table>
<thead>
<tr>
<th></th>
<th>c-W₃</th>
<th>s-W₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coulomb</td>
<td>Exchange</td>
</tr>
<tr>
<td>Dimer 12</td>
<td>-10.92</td>
<td>5.43</td>
</tr>
<tr>
<td>Dimer 23</td>
<td>-9.86</td>
<td>4.75</td>
</tr>
<tr>
<td>Dimer 13</td>
<td>-11.02</td>
<td>5.56</td>
</tr>
<tr>
<td>Sum of dimers</td>
<td>-31.80</td>
<td>15.74</td>
</tr>
<tr>
<td>X-Pol-X</td>
<td>-31.80</td>
<td>15.71</td>
</tr>
<tr>
<td>Non-additivity</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 2.1. Illustration of the minimum water trimer structure (c-W_3) in (a) and a symmetric configuration (s-W_3) in (b).

Figure 2.2. Electron density difference contour between the polarized and unpolarized systems, \( \Delta \rho = \rho(\hat{A}\psi_a\psi_b\psi_c) - \rho^0(\psi_a^0\psi_b^0\psi_c^0) \), for (a) the minimum-energy structure (c-W_3) and (b) the symmetric configurations (s-W_3). The structure orientations are identical to that shown in Figure 1 and the contour levels are at 0.003 au with red contours representing gain in electron density and blue for charge depletion.

Figure 2.3. Computed Coulomb (X-Pol) and exchange repulsion (X-Pol-X) interaction energy as a function of the hydrogen bond distance between the donor hydrogen and the acceptor oxygen atoms in the water dimer complex with C_s symmetry. Energies are computed using the 6-31+G(d) basis at the fully optimized geometries under the C_s symmetry constraint at the HF/6-31+G(d) level. The interaction energy profile is obtained from full geometry optimized with the hydrogen-bond distance specified above held fixed at various values shown in the figure.
Figure 2.1.
Figure 2.2.

(a)

(b)
Figure 2.3.
Chapter 3. Multilevel X-Pol: A Fragment-based Method with Mixed Quantum Mechanical Representations of Different Fragments

3.1 Introduction

The early development of atomistic potential energy functions for polypeptides and the current coarse-grained models by Scheraga and coworkers have profoundly influenced the field of computational biology. In recent years, a number of fragment-based quantum mechanical methods have been explored. In these methods, a large system is partitioned into monomer blocks also called fragments, which may be separate individual molecules or covalently connected species such as amino acid residues in a protein. Fragment-based methods are computationally efficient, which enables electronic structure calculations to be applied to condensed-phase and biomolecular systems to gain a deeper understanding of intermolecular interactions such as polarization and charge transfer. Although linear scaling quantum mechanical calculations of proteins have been carried out, further approximations are needed to treat intermolecular electrostatic interactions in order to overcome the sampling computational bottleneck for statistical mechanical properties. To this end, the explicit polarization (X-Pol) method, making use of block localization of molecular orbitals within individual fragments, was developed for statistical mechanical Monte Carlo and molecular dynamics simulations of condensed phase and biomolecular systems. The block-localization scheme in the X-Pol method can also be applied to density functional theory.

In many applications, one is particularly interested in the properties of a small region of the system, which could be the solute molecule in solution or the active site of
an enzyme, and, a high level of theory is needed to yield accurate results for this region of the system. Yet, it is also important to incorporate explicitly the instantaneous polarization of the rest of the system. One approach is to use the method of combined quantum mechanics and molecular mechanics (QM/MM) with the former treating the small region of interest and the latter representing the solvent and protein environment; however, most molecular mechanics treatments do not include the mutual polarization effects. Beyond traditional QM/MM approaches, a mixed multilevel fragment-based quantum mechanical method will allow the environmental region also to be modeled by an electronic structure method. The present paper describes such a multilevel method to represent different fragments with different quantum mechanical methods within the X-Pol formalism. The present multilevel X-Pol procedure has been implemented into the Gaussian 09 program and it is sufficiently general that any theoretical method available in that program can be combined to represent any of the different fragments. Thus, the present approach differs from the strategy in other fragment-based molecular orbital methods to treat different level of theory separately, such as FMO-MP2, FMO-DFT, FMO-coupled cluster, FMO-multiconfiguration, or multilayer with different basis sets; they are all treated in the same footing here. Our method also represents a general strategy for a multilayer QM/QM coupling to study chemical reactions and intermolecular interactions.

In the following, we first briefly summarize the X-Pol method, which is followed by a discussion of the multilevel X-Pol strategy. The computational details are given in Section 3.3, and Section 3.4 presents applications of the multilevel X-Pol method to two
hydrogen bonding systems, one involving acetic acid and water and the other being a Zundel ion-water cluster. Section 3.5 summarizes the main findings of the present study.

3.2 Theoretical Background

X-Pol theory has a hierarchy of three elements: (1) the construction of the total molecular wave function, (2) the formulation of an effective Hamiltonian, and (3) the reduction of computational costs in electronic integral evaluation. We first briefly summarize the X-Pol method for the case in which all fragments are treated at the same self-consistent field (SCF) level, and discuss some methods that can be used to include exchange, dispersion and charge transfer contributions. Then, we describe the procedure for a multilevel X-Pol approach with mixed theoretical levels, focusing our attention on post-SCF methods. Throughout this paper, we consider systems that do not have covalent bond connections between different fragments, but the generalization for treating covalently connected fragments can be achieved using methods described previously, in particular by making use of the generalized hybrid orbital (GHO) scheme developed for combined QM/MM simulations at various levels of theory.

3.2.1 The X-Pol Method

In X-Pol, a macromolecular system is partitioned into $N$ monomer blocks, also called fragments, and the total wave function $\Psi$ of the system is written as a Hartree product of the antisymmetric wave functions of individual fragments:

$$\Psi = \prod_{A=1}^{N} \Psi_A$$  \hspace{1cm} (1)
where $\Psi_A$ is the wave function of fragment $A$, which may be approximated by a single determinant or by a multi-configurational wave function. The wave functions for different monomers do not have to be approximated using the same method or represented at the same level of theory.

The effective X-Pol Hamiltonian for the system is

$$\hat{H} = \sum_A^{N} \hat{H}_A^0 + \frac{1}{2} \sum_A^{N} \sum_{B \neq A}^{N} \left( \hat{H}_{AB}^{\text{int}} + E_{AB}^{\text{XD}} \right)$$

(2)

where $\hat{H}_A^0$ is the electronic Hamiltonian for an isolated fragment $A$, and $\hat{H}_{AB}^{\text{int}}$ and $E_{AB}^{\text{XD}}$ represent electrostatic and exchange-dispersion (XD) interactions between fragments $A$ and $B$.\textsuperscript{8,10,38} The interaction Hamiltonian $\hat{H}_{AB}^{\text{int}}$ depends on both electronic and nuclear degrees of freedom, and it can be viewed as an electrostatic embedding of the QM fragment $A$ in the external field of fragment $B$:

$$\hat{H}_{AB}^{\text{int}} = -\sum_{i=1}^{M_A} e \Phi_E^B(r_i^A) + \sum_{\alpha=1}^{N_A} Z_A^A \Phi_E^B(R_\alpha^A)$$

(3)

where $M_A$ and $N_A$ are, respectively, the number of electrons and nuclei of fragment $A$, $r_i^A$ and $R_\alpha^A$ are the corresponding positions for electron $i$ and nucleus $\alpha$, $Z_A^A$ is a nuclear charge, and $\Phi_E^B(r_x^A)$ is the electrostatic potential at $r_x^A$ due to the external charge density of fragment $B$:

$$\Phi_E^B(r_x^A) = \int \frac{\rho_x^B(r')}{|r_x^A - r'|} dr'$$

(4)
where \( \rho^B(r') = -\rho_{ele}^B(r') + \sum_b Z_b^B \delta(r'-R_b^B) \) is the total charge density of fragment B, including both the smooth electron density \( \rho_{ele}^B(r') \) and the nuclear charges \( \{ Z_b^B \} \) at \( \{ R_b^B \} \). In the present multilevel X-Pol, the embedding potential \( \Phi_E^B(r_A^A) \) is modeled by partial atomic charges \( \{ q_b^B \} \) derived from the corresponding charge density, \( \rho_{ele}^B \), for example by population analysis (Mulliken or Löwdin charges) or by electrostatic potential fitting (CHELPG or Merz-Kollman schemes), and this simplifies it to

\[
\Phi_E^B(r_A^A) = \sum_b \frac{q_b^B}{|r_A^A - R_b^B|} \quad (5)
\]

The theory can be extended to use higher multipole moments \( ^60 \) or even the full charge distribution of the fragments (eq 4) to compute the electrostatic potential, \( ^61 \) but that will not be considered here.

The total X-Pol energy is given as follows:

\[
E[\{ \rho \}] = \langle \Psi | \hat{H} | \Psi \rangle = \sum_A E_A + \frac{1}{2} \sum_A N \sum_{B \neq A} E_{AB}^{int}[\rho^A,\{ q_b^B \}] + E_{AB}^{XD} \quad (6)
\]

where \( E_A \) is the energy of fragment \( A \) (note that \( E_A \) is different from the gas-phase energy \( E_A^0 \) because the wave function \( \Psi_A \) has been polarized by the rest of the system in X-Pol), and \( E_{AB}^{int}[\rho^A,\{ q_b^B \}] \) is the electrostatic interaction energy between fragments \( A \) and \( B \), akin to that used in a QM/MM method.\(^{43-44}\)

The energy term \( E_{AB}^{XD} \) in eq 4 accounts for the effects of the approximation used in eq 1, which by construction neglects short-range exchange repulsion and long-range
dispersion interactions as well as charge transfer contributions. In the original X-Pol method, exchange repulsion is represented by a pairwise $R_{ij}^{12}$ dependence and the attractive non-covalent interaction by a pairwise $R_{ij}^{-6}$ term as in the Lennard-Jones potential, where $R_{ij}$ is an interatomic distance. In the present study, an exponential function for the repulsion, as in the Buckingham potential is used:

$$E_{AB}^{XD} = \sum_{I}^{N_A} \sum_{J}^{N_B} A_{ij} e^{-B_{ij} R_{ij}} - \frac{C_{ij}}{R_{ij}}$$  \hspace{1cm} (7)

where the parameters $A_{ij}$, $A_{ij}^{J}$, and $C_{ij}$ are determined using standard combining rules from atomic parameters such that $A_{ij} = (A_i A_j)^{1/2}$, $B_{ij} = (B_i + B_j)/2$, and $C_{ij} = (C_i C_j)^{1/2}$.

The effect of charge transfer is modeled indirectly. The strict block localization of molecular orbitals within individual monomers in X-Pol does not allow charge delocalization between different fragments (unless one uses a grand canonical formulation, which is not employed here). At distances longer than hydrogen bonding range, it is often a good approximation to neglect charge transfer, and interfragment electrostatic interactions can then be adequately described by the electrostatic embedding scheme using the Coulomb potential (eq 5). However, at short interfragment distances where there is significant orbital overlap, one needs to take into account the energy component due to charge delocalization (sometimes also called charge transfer). In the present work, we account for charge transfer only empirically, in particular (in the spirit often used in molecular mechanics) by modeling the charge delocalization
energy with enhanced electrostatic polarization. Consequently, the electrostatic potential \( \Phi_E^B(\mathbf{r}_X^A) \) in eq 5 is recognized as an effective potential that mimics both long-range Coulomb (electrostatic) interactions and short-range charge delocalization contributions, and this can be achieved to some extent by optimizing the parameters in the \( E_{AB}^{XD} \) term (eq 7) and possibly the charge model \( \left\{ q_b^B \left[ \rho_{ele}^B \right] \right\} \) (eq 5) to best reproduce hydrogen bonding interactions for a set of bimolecular complexes (however such optimization is beyond the scope of the present article).

Beyond the empirical approximations, a variational many-body expansion approach in X-Pol has been described, which includes exchange repulsion, charge delocalization and dispersion terms explicitly. Individually, one way to improve on the repulsive potential is to antisymmetrize the X-Pol Hartree-product wave function; this yields X-Pol with full eXchange, called X-Pol-X. When the monomers are treated by Hartree-Fock theory, this calculation can be accomplished by using the formalisms of block-localized wave function (BLW) or the SCF-MI method, and this procedure has been extended to density functional theory. To treat dispersion interactions, multiconfigurational methods and perturbation theories can be used; for example, one can adopt symmetry adapted perturbation theory (SAPT) as a post-SCF correction to the X-Pol energy, as has been done recently by Jacobson and Herbert. Both exchange-repulsion and dispersion interactions are of short range on the length scale of solutions and biopolymer systems, and only the close neighbors need to be explicitly considered.

Charge delocalization effects can also be estimated using a grand canonical ensemble, or by using the method of interaction energy expansion introduced by Stoll
and Preuss,\textsuperscript{71} which has been adopted by Kitaura and coworkers in a fragment molecular orbital implementation.\textsuperscript{11} Of course, a straightforward way of including charge delocalization effects is to use larger fragments that include charge transfer partners.\textsuperscript{28} Another approach, which has been recast in several ways, is the molecular fractionation with conjugated caps (MFCC) approach by Zhang and coworkers.\textsuperscript{13,72} In MFCC, the individual fragments are capped with a structure representative of the local functional group of the original system, and the total energy is obtained by subtracting the energies that account for the common fragments used in the “caps”.\textsuperscript{73-74} In both cases, the total energy can be conveniently determined using this addition-subtraction scheme; however, the total molecular wave function is no longer available, making energy gradient calculations more challenging. In this regard, we have developed a generalized explicit polarization (GX-Pol) method on the basis of a multiconfiguration self-consistent field (MCSCF) wave function that makes use of dimeric, charge-delocalized fragments.\textsuperscript{30,75} In the present study, we do not include the explicit treatment of charge delocalization energy, but this can be addressed in a separate study.

3.2.2 \textbf{Multilevel X-Pol}

The method outlined in the previous section has been implemented with all fragments treated at the same theoretical level (semiempirical,\textsuperscript{8,10,37} Hartree-Fock (HF), or density functional theory (DFT)\textsuperscript{38}). Here, we consider a system partitioned into $N$ fragments, of which $N'$ fragments are treated by a method denoted as high level (HL) and the remainder $N - N'$ fragments are modeled with a different approach specified as low level (LL). The former fragments are called HL fragments, and the latter are called
LL fragments. Generalization to any number of levels is straightforward, but for convenience, we restrict the following discussions to two levels. This division highlights the need for high accuracy in a small (HL) region of interest, such as the solute molecule in a solution or the active site of an enzyme, while retaining the need for a computationally efficient way to include polarization effects in the remainder of the system. In the present illustration of the method, the LL method is restricted to either HF or DFT.

A variety of methods can be used for fragments in the HL region, and they are divided into two categories: SCF and post-SCF. For methods such as DFT and multiconfiguration SCF (MCSCF), the treatment is the same as described previously for single-level X-Pol based on ab initio Hartree-Fock calculations or DFT. The second category includes post-SCF methods such as configuration interaction (CI), coupled cluster (CC) and Møller-Plesset (MP) perturbation theory; when such methods are employed for HL fragments, the total X-Pol energy is written as

$$E[\{\rho\}] = \sum_A^N \left( E_A^{\text{SCF}} + \frac{1}{2} \sum_{B \neq A}^N E_{AB}^{\text{int}}[\rho^A, \{q_b^B\}] + E_{AB}^{\text{XD}} \right) + \sum_A^{N'} E_A^{\text{corr}}$$

$$= E_{\text{tot}}^{\text{SCF}} + E_{\text{tot}}^{\text{corr}}$$

(8)

where $E_A^{\text{SCF}}$ is the SCF energy of the reference wave function, $E_A^{\text{corr}}$ is the post-SCF correction for fragment $A$, and $E_{\text{tot}}^{\text{SCF}}$ and $E_{\text{tot}}^{\text{corr}}$ are the total SCF energy and the total post-SCF correlation energy. Note that the SCF energy $E_A^{\text{SCF}}$ can be obtained either from a single determinant reference wave function in a CI, CC, or MP2 calculation or a multiconfiguration wave function in multireference CI, or CASPT2, etc. calculations. The
The main difference of this energy expression from that of eq 6 is that the total multilevel X-Pol energy is no longer written as the expectation value of an X-Pol wave function.

In using eq 8, the computation involves an initial optimization of the X-Pol SCF wave function, followed by determining the post-SCF energy corrections for fragments in the HL region. In the SCF procedure, the charge densities of the HL fragments that polarize other fragments are the response densities corresponding to the post-SCF calculation.\textsuperscript{76-78} The response density (which is also called the generalized density and relaxed density) is the sum of the SCF density and the relaxation density due to the post-SCF procedure, which is obtained using the Z-vector method,\textsuperscript{79} including a single coupled perturbed HF calculation for the occupied and virtual molecular orbital (MO) block, independent of the specific post-SCF method.\textsuperscript{77} The response density procedure allows the use of methods (such as MP perturbation theory) for which the energy does not correspond to a wave function expectation value; it is also more accurate for computing one-particle properties using CC and other post-SCF methods. The response density is obtained by adding the relaxation density to the SCF density and transforming into the atomic basis for population analysis and computation of one-particle properties including the electrostatic potential:

$$P_{\mu \nu}^A,_{\text{HL}} = P_{\mu \nu}^A,_{\text{SCF}} + P_{\mu \nu}^A,_{\text{rel}}$$

(9)

where $P_{\mu \nu}^A,_{\text{SCF}}$ and $P_{\mu \nu}^A,_{\text{rel}}$ are the SCF and relaxation densities for fragment $A$ in the HL region. If Mulliken population analysis (MPA) is used,\textsuperscript{80} the partial atomic charges in eq 8 for HL fragments can be written as
\[ q_{a,\text{HL}}^{A} = Z_{a}^{A} - \sum_{\mu \in A} P_{\mu \nu}^{A,\text{HL}} S_{\mu \nu}^{A} = q_{a,\text{SCF}}^{A} + q_{a,\text{rel}}^{A} \]  \hspace{1cm} (10)

The elements of the effective Hamiltonians (Fock or Kohn-Sham matrices), both for the HL and LL fragments, in a multilevel X-Pol method can be written similarly as\textsuperscript{22,81}

\[
F_{\mu \nu}^{A,\text{Xpol}} = \frac{\partial E^{\text{SCF}}[\{\rho\}]}{\partial P_{\mu \nu}^{A,\text{SCF}}} = F_{\mu \nu}^{A,\alpha} - \frac{1}{2} \sum_{b \in A} \sum_{b \in B} q_{b}^{B} (I_{b}^{B})_{\mu \nu}^T + \frac{1}{2} \sum_{a \in A} X_{a}^{A} (\Lambda_{a}^{A})_{\mu \nu} \]  \hspace{1cm} (11)

where \( F_{\mu \nu}^{A,\alpha} \) is the Fock matrix element for an isolated fragment \( A \), \( q_{b}^{B} \) is the partial charge on atom \( b \) in fragment \( B \) and it is understood that \( q_{b}^{B} \equiv q_{b,\text{HL}}^{B} \) for fragments in the HL region, \( I_{b}^{B} \) is the matrix of the pair potential in atomic basis as defined by

\[
(I_{b}^{B})_{\mu \nu}^T = \langle \mu | \frac{1}{|r - R_{b}^{B}|} | \nu \rangle \]  \hspace{1cm} (12)

and \( X_{a}^{A} \) is a vector arising from the derivative of the interaction energy with respect to partial atomic charge of atom \( a \):

\[
X_{a}^{A} = \sum_{b \in A} \left( \sum_{\lambda \in A} P_{\lambda \sigma}^{B} (I_{a}^{A})_{\sigma \lambda}^T + \sum_{b \in B} \frac{Z_{b}^{B}}{|R_{b}^{B} - R_{a}^{A}|} \right) \]  \hspace{1cm} (13)

Note that the notation \( P_{\lambda \sigma}^{B} \) is defined as \( P_{\lambda \sigma}^{B} \equiv P_{\lambda \sigma}^{B,\text{HL}} \) if \( B \leq N' \), and as \( P_{\lambda \sigma}^{B} \equiv P_{\lambda \sigma}^{B,\text{SCF}} \) if \( B > N' \). The elements of the response density matrix, \( \Lambda_{a}^{A} \), are given by

\[
(\Lambda_{a}^{A})_{\mu \nu} = \frac{\partial q_{a}^{A}}{\partial P_{\mu \nu}^{A,\text{SCF}}} = \frac{\partial q_{a}^{A,\text{SCF}}}{\partial P_{\mu \nu}^{A,\text{SCF}}} \]  \hspace{1cm} (14)
where \( q^A_a \) is the atomic charge on atom \( a \), and \( P^{A,SCF}_{\mu\nu} \) is an element of the density matrix of fragment \( A \) in the SCF optimization of the X-Pol fragment wave function. The charge derivatives have been given for a number of charge models.\(^{22,81}\) The interpretation of eqs 11–14 is that the wave functions for all fragments are optimized at the SCF level, but their polarization, by virtue of setting \( q^B_b \) to \( q^B_{b,SCF} + q^B_{b,rel} \) for \( B \leq N' \), includes contributions from the relaxation density corresponding to the post-SCF energy in the HL region.

### 3.2.3 Iterative Updating (IU) Method

In the standard X-Pol method, the SCF wave function of eq 1 is optimized variationally by using eq 11.\(^ {81}\) An alternative way of optimizing the SCF wave function, which is non-variational, is to consider each fragment as an isolated molecule embedded in the electrostatic field of the rest of the system. Then, the Fock matrix for each fragment can be written separately as follows:\(^ {8-10}\)

\[
F^{A,IU} = F^{A,o} - \sum_{B \neq A} \sum_b q^B_b \left( I^B_b \right)^A
\]

where \( F^{A,o} \) is the Fock matrix of fragment \( A \), \( q^B_b \) is a column vector of atomic charges of fragment \( B \) stretched to the dimension of the orbital basis and \( \left( I^B_b \right)^A \) is the matrix of pair potential (eq 12). The total electronic energy of the system can then be determined iteratively by a double self-consistent-field (DSCF) procedure.\(^ {8-10,82}\) Starting with an initial guess of the one-electron density matrix for each fragment, one loops over all fragments in the system and performs SCF optimization of the wave function for each
fragment in the presence of the instantaneous external charges of all other fragments (through \( \left( \mathbf{1}_{\mathcal{B}} \right)^{\mathcal{A}} \)). This is iterated (the SCF for the system) with an updated external potential until the total electronic energy and the charge density are converged. This iterative updating (IU) procedure is straightforward and was the approach proposed for fragment calculations in Ref. 8 and adopted in the subsequent FMO implementation.\(^{11,83}\) Such an iterative updating procedure can be found in many applications both in electronic structure theory\(^{82}\) and combined QM/MM approaches that include MM polarization.\(^{45}\) A main short coming of the above approach is that the Fock matrix in eq 15 is not variationally optimized,\(^{18,81}\) and it is not suitable for efficiently computing energy gradients. In this study, we use the superscript IU for the non-variational iterative updating procedure in eq 15, and simply Xpol for the variational method employing eq 11.

Note that both the sequential and variational optimization of the X-Pol wave function can be carried out by DSCF iterations, although they can also be done, if desired, as a single large SCF problem. In the illustrations of the multilevel X-Pol method presented below, we will we compare the energy difference between the two optimization procedures.

### 3.3 Computational Details

The goal of this study is to illustrate that the multilevel X-Pol fragment-based quantum mechanical model can be implemented with an arbitrary combination of different electronic structural methods for different fragments. The X-Pol method has
been implemented into a locally modified version of the Gaussian-09 program. In this program, the response density can be computed for a range of post-SCF methods, including MPn, QCISD, CCD, CCSD, CID, CISD, BD, and SAC-CI, thus, any of these—as well as SCF methods—can be used to represent a given fragment in multilevel X-Pol.

We choose two hydrogen bonding complexes, (1) acetic acid (fragment $A$) and water (fragment $B$), and (2) $\text{H}_2\text{O}_2^+$ (fragment $A$) and four water molecules (four water fragments as $B$ for a total of five fragments). The complexes and monomer structures are optimized using the hybrid M06/MG3S DFT, which are then used in all subsequent single-point energy calculations with various multilevel X-Pol methods. In the present study, we have used the hybrid density functional theory M06, second order Møller-Plesset perturbation theory (MP2), and coupled-cluster with singles and doubles (CCSD) method for acetic acid and the $\text{H}_2\text{O}_2^+$ ion, and we employed Hartree-Fock (HF), B3LYP and M06 density functional method for water. In all X-Pol calculations, the 6-31G(d) basis set was used.

The binding energy $\Delta E_b$ for the complex is defined as

$$\Delta E_b = E_{AB} - E_A^0 - E_B^0$$

(16)

In X-Pol, the binding energy can be decomposed into an intramolecular distortion term $\Delta E_{\text{dist}}$, including both the energy change due to geometric variation and the energy cost needed to polarize the electron density, and an intermolecular interaction contribution. The latter can be further separated into an electrostatic component
\( \Delta E_{\text{int}} \) and an exchange-dispersion energy term \( \Delta E_{\text{XD}} \). In this energy decomposition scheme, we rewrite eq 16 as

\[
\Delta E_b = \Delta E_{\text{dist}} + \Delta E_{\text{int}} + \Delta E_{\text{XD}}
\]  

These terms are defined as follows:

\[
\Delta E_{\text{dist}} = \sum_{A}^{N} (E_A - E_A^o)
\]  

where \( E_A \) and \( E_A^o \) are the intra-monomer components of the energies of monomer \( A \) in the complex or in isolation,

\[
\Delta E_{\text{XD}} = \sum_{A>B}^{N} E_{AB}^{XD}
\]

and

\[
\Delta E_{\text{int}} = \sum_{A>B}^{N} \Delta E_{AB}^{\text{int}} = \sum_{A>B}^{N} \frac{1}{2} [\Delta E_A^{\text{int}}(B) + \Delta E_B^{\text{int}}(A)]
\]

where \( \Delta E_{AB}^{\text{int}} \) is the interaction energy between monomers \( A \) and \( B \), and \( \Delta E_{X}^{\text{int}}(Y) \) is the electrostatic interaction energy of the “QM” fragment \( X \) polarized by the external potential of monomer \( Y \). Although \( \Delta E_{X}^{\text{int}}(Y) \) and \( \Delta E_{Y}^{\text{int}}(X) \) describe the same interaction between monomers \( X \) and \( Y \), they are not symmetric unless the same theoretical model is used for both monomers and the Coulomb integrals are explicitly computed over all basis functions. For convenience of discussion, we also define the total electrostatic component of binding energy as
∆E_{ele} = ∆E_{dist} + ∆E_{int}  \hspace{1cm} (21)

### 3.4 Results and discussion

Table 3.1 shows the computed electrostatic interaction energies between acetic acid and water for the optimized configuration shown in Figure 3.1 using the sequential optimization approach in multilevel X-Pol. Two charge models are used in this work, those from Mulliken population analysis (MPA) and those from Merz-Kollman electrostatic potential fitting (MK). The corresponding results obtained using variational optimization in multilevel X-Pol are given in Table 3.2. In this case, only the MPA charges are used.

The total interaction energy between an acetic acid and a water molecule at the configuration shown in Figure 3.1 is -6.9 kcal/mol from M06/MG3S optimization, which is reduced slightly to -6.6 kcal/mol using CCSD(T)/MG3S//M06/MG3S. The electrostatic interaction energy from X-Pol by iterative updating method using M06/6-31G(d) for both fragments is -7.7 kcal/mol when the MPA charges are used, and it is reduced to -7.0 kcal/mol when the MK charges are used. With a different combination in the multilevel X-Pol method in which acetic acid is treated by CCSD(T) and water by M06, the computed electrostatic interaction energies are -7.6 and -7.2 kcal/mol with the MPA and MK charges, respectively, similar to the single level results. Switching to the variational X-Pol method, we obtained an electrostatic interaction energy of -9.0 kcal/mol using the M06 representation of both monomers and the MPA charges. In this case, the variational optimization of the Kohn-Sham orbitals lowers the energy by 1.8 kcal/mol for this
bimolecular complex. Similar trends are found in other multilevel X-Pol combinations in Table 3.2. The results in Table 3.1 do not include the exchange repulsion energy, charge transfer contributions, or correlation effects that result from wave function delocalization in a full KS-DFT calculation. The latter two effects are not fully separable in energy decomposition analyses, but both make stabilizing contributions to the bimolecular complex, which tend to partially compensate the strong exchange repulsion energies. If we optimize the empirical parameters in eq 7 for the M06 and B3LYP combination in multilevel X-Pol, we obtained a energy of of 2.1 kcal/mol for the $\Delta E_{XD}$ term, which is applied to all multilevel X-Pol methods in Table 3.2 to yield the total binding energies $\Delta E_b$.

Table 3.1 shows that the MPA charges tend to provide stronger electrostatic polarization effect than do the MK charges. This results in overall interaction energies that are greater in magnitude. Both $\Delta E^\text{int}_A(B)$ and $\Delta E^\text{int}_B(A)$ describe the electrostatic interaction energy between fragments $A$ and $B$, but they differ numerically because the former specifies embedding of fragment $A$ in the classical field of fragment $B$ whereas $\Delta E^\text{int}_B(A)$ gives the embedding energy of fragment $B$ in the electrostatic field of fragment $A$. Across the series of five different combinations shown in Table 3.1, the computed $\Delta E^\text{int}_A(B)$ values are greater than the $\Delta E^\text{int}_B(A)$ terms by using IU optimization of the wave function within an electrostatic embedding picture; however, the ordering is reversed in the variational X-Pol method. Furthermore, the electrostatic polarization is significantly stronger by variational optimization than by IU optimization of the wave
function, both in single level and in multilevel X-Pol. On one hand, the difference between the $\Delta E_A^{\text{int}}(B)$ and $\Delta E_B^{\text{int}}(A)$ terms highlights the asymmetry in the representation of two fragments in a QM/MM type of treatment, and the average of the two terms is defined as the X-Pol dimer interaction energy (eq 20)\textsuperscript{8}\textsuperscript{9}. On the other hand, the difference between the IU and variational optimization procedures for the X-Pol wave function shows the importance of correctly accounting for the mutual polarization effects among different fragments that minimize the adiabatic ground state energy. Note that few existing fragment-based methods optimize the fragment wave functions variationally.

Table 3.3 gives the interaction energies between the Zundel ion $\text{H}_5\text{O}_2^+$ and four water molecules computed with various theoretical models using the optimized structure with M06/MG3S\textsuperscript{86} (Figure 3.2). The optimized structure for the complex is very similar to that optimized using B3LYP/6-311+G(dp) from ref 87.

It is interesting to first compare various methods of estimating the exchange repulsion–dispersion contributions to the energy of binding in this case. The exchange repulsion energy can be obtained as the difference between the energy from the antisymmetrized X-Pol wave function and that from the X-Pol at the Hartree-Fock level. We found that the results depend noticeably on the basis set and the charge model used for electrostatic coupling between different fragments (eq 11). The estimated exchange energies are 30.0 and 28.5 kcal/mol using iterative updating optimization in X-Pol with the MK and MPA charges, respectively. This increases to 35.8 kcal/mol using the variational X-Pol wave function and the MPA charges. In these cases, the 6-31G(d) basis is used.
To gain more insights into the magnitude of the contributions from intermonomer exchange, dispersion and charge transfer on the hydrogen bonding interactions in the Zundel ion complex, we have carried out an interaction energy decomposition analysis using the block-localized wave function method (BLW-ED) using a larger basis set.\textsuperscript{30,61} At the HF/aug-cc-pVDZ level, the exchange repulsion and charge transfer contributions to the energy of binding are estimated to be 38.8 and -13.3 kcal/mol, respectively, for a net contribution of 25.5 kcal/mol, and the total binding energy is -62.4 kcal/mol. If one uses the difference between the CCSD(T) binding energy (-69.7 kcal/mol) and that at the HF level (-62.4 kcal/mol) as a rough estimate of the dispersion contribution, a value of 7.3 kcal/mol is obtained. Then, the overall $E_{\text{AB}}^{\text{XD}}$ term including the effect of charge transfer may be estimated as 18.2 kcal/mol (that is, 25.5 minus 7.3 kcal/mol).

We optimized the Lennard-Jones parameters separately for the oxonium ion system with the M06 density functional for H$_3$O$^+$ and the B3LYP functional for (H$_2$O)$_4$, and we obtained $A_{\text{OO}} = 1.5221 \times 10^5$ kcal/mol, $B_{\text{OO}} = 3.754$ Å, and $C_{\text{OO}} = 756.3$ Å$^6$ kcal/mol for the Buckingham potential (eq 7). Then, eq 7 yields a value of 18.4 kcal/mol for $\Delta E_{\text{XD}}$, in good agreement with the above analysis. Although the $\Delta E_{\text{XD}}$ term ought be reoptimized for each multilevel X-Pol model, we have used to same Buckingham energy for all combinations listed in Table 3.3, and the total binding energies in the last column of Table 3.3 are reasonable in comparison with the CCSD(T) value (at the M06/MG3S geometry) of 69.7 kcal/mol. For comparison, the corresponding multilevel
X-Pol values without inclusion of the $\Delta E_{\text{XD}}$ term are significant greater than the full QM result, ranging from -83 to -92 kcal/mol.

For multilevel X-Pol in which both HL and LL energies are obtained at the SCF level, the energy of binding from the variational approach will be more negative than that obtained using the non-variational (iterative updating) procedure, which is also used in the fragment molecular orbital model.\textsuperscript{11-12} The results using the M06 density functional for the HL fragment in Table 3.3 are indeed consistent with this expectation. However, if the HL energy is determined by a post-SCF theory as in MP2 and CCSD calculations in Table 3.3, there is no guarantee that the “variational” multilevel X-Pol energy is lower than that of the IU optimization result because only the reference wave function used in the post-SCF calculation is optimized. This is seen in the CCSD and M06 combination, which yields a binding energy smaller than that from the IU optimization method (both using the MPA charges). The reference wave functions for the individual fragments are more strongly distorted than in other cases.

3.5 Concluding Remarks

The explicit polarization (X-Pol) method is a fragment-based quantum mechanical method, in which a macromolecular system is partitioned into monomer fragments and the total molecular wave function is written as a Hartree product of the antisymmetric wave functions for individual fragments. In the present study, a general formulation is presented to treat different fragments with different electronic structure methods. The current implementation of the multilevel X-Pol method in Gaussian-09
allows any method available in that program to be used to describe a given fragment. The key to the implementation is using the response density to compute the electrostatic coupling (and mutual polarization), in particular by using the response density in population analyses or in an electrostatic potential charge fitting procedure.

The computational method is illustrated by calculations on two hydrogen bonding complexes involving acetic acid and water, and the H$_5$O$_2^+$ ion and four water molecules. Acetic acid and H$_5$O$_2^+$ are treated using M06, MP2 and CCSD as the high-level theory, and these methods are paired with one of HF, M06 and B3LYP as the lower-level method. The present multilevel X-Pol method can be used to treat a small region of the system, such as the solute molecule in solution or the active site of an enzyme, with a high-level theory, and the remainder of the system with a more computationally efficient method.
Table 3.1. Computed binding energies and energy components (kcal/mol) between acetic acid (A) and water (B) using iterative charge-updating optimization in multilevel X-Pol. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>A dist</th>
<th>int $E_B$ (B)</th>
<th>int $E_A$ (A)</th>
<th>int $E$</th>
<th>$E_B$ $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Pol optimization by iteratively updating MK charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>1.4</td>
<td>-9.1</td>
<td>-7.8</td>
<td>-8.4</td>
<td>-7.0</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>1.4</td>
<td>-8.9</td>
<td>-7.6</td>
<td>-8.2</td>
<td>-6.8</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>1.6</td>
<td>-9.4</td>
<td>-8.2</td>
<td>-8.8</td>
<td>-7.2</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>1.2</td>
<td>-8.8</td>
<td>-7.8</td>
<td>-8.3</td>
<td>-7.1</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>1.0</td>
<td>-8.7</td>
<td>-7.7</td>
<td>-8.2</td>
<td>-7.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>A dist</th>
<th>int $E_B$ (B)</th>
<th>int $E_A$ (A)</th>
<th>int $E$</th>
<th>$E_B$ $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Pol optimization by iteratively updating MPA charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>1.5</td>
<td>-9.6</td>
<td>-8.7</td>
<td>-9.2</td>
<td>-7.7</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>1.5</td>
<td>-9.2</td>
<td>-8.5</td>
<td>-8.8</td>
<td>-7.3</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>1.8</td>
<td>-10.2</td>
<td>-9.2</td>
<td>-9.7</td>
<td>-7.9</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>1.3</td>
<td>-9.5</td>
<td>-8.4</td>
<td>-9.0</td>
<td>-7.7</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>1.1</td>
<td>-9.1</td>
<td>-8.3</td>
<td>-8.7</td>
<td>-7.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>A dist</th>
<th>int $E_B$ (B)</th>
<th>int $E_A$ (A)</th>
<th>int $E$</th>
<th>$E_B$ $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full wave function calculation without fragmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06 $^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCSD(T) $^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ For the purposes of this table, the multilevel X-Pol calculations of $E_b$ are set equal to $E_{ele}$, that is, they include the change in intramonomer energy and the electrostatic interaction energy but not the intermonomer exchange repulsion and dispersion terms.
b. Computed using the MG3S basis set at the M062X/MG3S geometries.
Table 3.2. Computed binding energies and energy components (kcal/mol) between acetic acid (A) and water (B) using the variational multilevel X-Pol. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>$\Delta E_{\text{dist}}$</th>
<th>$\Delta E_{\text{int}}^{A}$ (B)</th>
<th>$\Delta E_{\text{int}}^{B}$ (A)</th>
<th>$\Delta E_{\text{int}}$</th>
<th>$\Delta E_{\text{ele}}$</th>
<th>$\Delta E_{\text{XD}}$ a</th>
<th>$\Delta E_{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Pol by variational optimization with MPA charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>2.8</td>
<td>-10.1</td>
<td>-13.5</td>
<td>-11.8</td>
<td>-9.0</td>
<td>2.0</td>
<td>-7.0</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>2.6</td>
<td>-9.7</td>
<td>-13.0</td>
<td>-11.3</td>
<td>-8.7</td>
<td>2.0</td>
<td>-6.7</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>3.1</td>
<td>-10.7</td>
<td>-14.2</td>
<td>-12.5</td>
<td>-9.4</td>
<td>2.0</td>
<td>-7.4</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>4.3</td>
<td>-11.5</td>
<td>-13.1</td>
<td>-12.3</td>
<td>-8.0</td>
<td>2.0</td>
<td>-6.0</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>3.8</td>
<td>-11.0</td>
<td>-12.6</td>
<td>-11.8</td>
<td>-8.0</td>
<td>2.0</td>
<td>-6.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full wave function calculation without fragmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06 a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6.9</td>
</tr>
<tr>
<td>CCSD(T) b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6.6</td>
</tr>
</tbody>
</table>

a. The exchange-dispersion energy is estimated using the Buckingham terms with the parameters $A_{\text{OO}} = 1.5221 \times 10^5$ kcal/mol, $B_{\text{OO}} = 3.754$ Å$^{-1}$, and $C_{\text{OO}} = 756.3$ Å$^6$ kcal/mol for oxygen, and $A_{\text{CC}} = 2.50178 \times 10^6$ kcal/mol, $B_{\text{CC}} = 4.384$ Å$^{-1}$, and $C_{\text{CC}} = 1533.1$ Å$^6$ kcal/mol for carbon.

b. Computed at CCSD(T)/MG3S//M062X/MG3S.
Table 3.3. Computed binding energies and energy components (kcal/mol) between H$_5$O$_2^+$ and (H$_2$O)$_4$ using the iterative charge-updating optimization and variational multilevel X-Pol methods. The 6-31G(d) basis set is used in all calculations at the M06/MG3S optimized geometries.

<table>
<thead>
<tr>
<th></th>
<th>(H$_2$O)$_4$</th>
<th>$\Delta E_{\text{dist}}$</th>
<th>$\Delta E_{\text{int}}$</th>
<th>$\Delta E_{\text{ele}}$</th>
<th>$\Delta E_{\text{XD}}$</th>
<th>$\Delta E_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Pol optimization by iteratively updating MK charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>11.4</td>
<td>-100.5</td>
<td>-89.1</td>
<td>18.4</td>
<td>-70.7</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>11.4</td>
<td>-99.0</td>
<td>-87.7</td>
<td>18.4</td>
<td>-69.3</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>11.3</td>
<td>-103.3</td>
<td>-92.0</td>
<td>18.4</td>
<td>-73.6</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>10.9</td>
<td>-103.8</td>
<td>-92.9</td>
<td>18.4</td>
<td>-74.5</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>11.1</td>
<td>-100.6</td>
<td>-89.5</td>
<td>18.4</td>
<td>-71.1</td>
</tr>
<tr>
<td>X-Pol optimization by iteratively updating MPA charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>11.5</td>
<td>-99.0</td>
<td>-87.5</td>
<td>18.4</td>
<td>-69.1</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>11.4</td>
<td>-96.6</td>
<td>-85.2</td>
<td>18.4</td>
<td>-66.8</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>11.4</td>
<td>-103.1</td>
<td>-91.7</td>
<td>18.4</td>
<td>-73.3</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>11.1</td>
<td>-103.7</td>
<td>-92.7</td>
<td>18.4</td>
<td>-74.3</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>11.2</td>
<td>-99.2</td>
<td>-88.0</td>
<td>18.4</td>
<td>-69.6</td>
</tr>
<tr>
<td>X-Pol by variational optimization with MPA charges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>10.7</td>
<td>-101.7</td>
<td>-91.0</td>
<td>18.4</td>
<td>-72.6</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>11.3</td>
<td>-99.4</td>
<td>-88.1</td>
<td>18.4</td>
<td>-69.7</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>11.3</td>
<td>-105.8</td>
<td>-94.5</td>
<td>18.4</td>
<td>-76.1</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>12.3</td>
<td>-106.7</td>
<td>-94.4</td>
<td>18.4</td>
<td>-76.0</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>18.5</td>
<td>-102.4</td>
<td>-83.9</td>
<td>18.4</td>
<td>-65.5</td>
</tr>
</tbody>
</table>

Full wave function calculation without fragmentation

CCSD(T)$^a$ -69.7

---

a. Computed at CCSD(T)/MG3S//}}M062X/MG3S.
Figure 3.1. Schematic illustration of the optimized configuration of acetic acid and water using M06/MG3S.
Figure 3.2. Fragment partition of the $\text{H}_2\text{O}_2^+(\text{H}_2\text{O})_4$ cluster optimized at the M06/MG3S.
Chapter 4. Variational Many-Body Expansion: Accounting for Exchange Repulsion, Charge Delocalization, and Dispersion in the Fragment-based X-Pol Method

In 1977, Stoll and Preuss described a fragment-based many-body correction procedure for treating molecular clusters, layers and solids. In this approach, a system is partitioned into $N$ subsystems, represented by localized nonorthogonal orbitals. The spatial separation of the subsystems allows intra and inter-fragment interactions to be treated differently; the orbitals involved in the strong intra-group interactions are optimized using Hartree-Fock theory, and the relatively weak inter-fragment interactions are approximated by a Hartree-type potential. This gives the first-order energy $E_1$ for the monomer system. Then, a pair of subsystems, $I$ and $J$, are united to form a larger fragment, embedded in the remaining $N-2$ monomers to yield the two-body correction energy $\Delta E_{IJ}$ for subsystems $I$ and $J$. This is repeated for all relevant pairs, and if necessary, three-body terms, etc. to increase accuracy. The total energy of the system in this procedure can be written as

$$E_{\text{tot}} = E_1 + \Delta E_2 + \Delta E_3 + \cdots + \Delta E_N$$

where

$$\Delta E_2 = \sum_{I<J}^{N} \Delta E_{IJ} = \sum_{I<J}^{N} (E_{IJ} - E_1)$$

$$\Delta E_3 = \sum_{I<J<K}^{N} \Delta E_{IJK} = \sum_{I<J<K}^{N} (E_{IJK} - E_1 - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK})$$
with higher-order terms defined analogously. In eqs 2 and 3, $E_{IJ}$ is the energy of the two-body subsystem consisting of the $I$-$J$ superfragment and $N$-2 other monomers, and $E_{ILK}$ is the energy of the three-body subsystem involving the $I$-$J$-$K$ superfragment and $N$-3 other monomers. This expression is analogous to that used in local correlation energy expansion methods, but the aim is different here.

Although eq 1 is exact in the limit of order $N$, the method is only practical if the series converges quickly with negligible high-order contributions beyond two-body corrections. Therefore, it is important to use a method that can yield a first-order term as close to the ground-state energy as possible. The main advantage of the procedure of Stoll and Preuss is that each energy term is determined in the field of all other fragments. This greatly enhances the convergence of eq 1. Another critical element is the use of local orbitals, which increase the weights from lower-order terms, and a simple localization approach is to construct nonorthogonal orbitals with basis functions partitioned to a given region. Methods for optimizing these nonorthogonal orbitals have been described in various contexts.

One approach to reduce the computational effort for determining $E_1$ is the explicit polarization (X-Pol) method. In X-Pol, the first-order molecular wave function $\Phi_1$ is written as a Hartree product of the wave functions of the individual fragments $\{\Psi_I; I = 1, \cdots, N\}$. Consequently, $\Psi_I$ can be optimized independently including the Coulomb potentials from other fragments. In the original X-Pol method, dispersion, exchange repulsion and charge delocalization contributions between different fragments are approximated in spirit of molecular mechanics using
empirical Lennard-Jones terms. In this chapter, we present a second-order variational many-body X-Pol (VMB2/X-Pol) method to directly account for these energy components.

A number of fragmental quantum mechanical methods have been reported based on many-body expansion,\textsuperscript{1,19-22} notably the fragment molecular orbital (FMO) method of Kitaura and coworkers,\textsuperscript{19, 23-24} the method of electrostatic field adapted molecular fractionation with conjugated caps\textsuperscript{25} (EFA-MFCC),\textsuperscript{26} and the electrostatically embedded many-body expansion (EE-MB).\textsuperscript{27} Although the specific implementations differ, in FMO models, the two-body and higher order terms are approximated by single-point energy calculations in the field of the monomer-optimized densities,\textsuperscript{23-24} and in EFA-MFCC and EE-MB, the monomer terms are also determined with fixed point charges.\textsuperscript{26-27} Therefore, the mutual polarization effects in the many-body energy terms are excluded. In FMO,\textsuperscript{23-24} apparently unaware of the early work, a double self-consistent field (DSCF) procedure\textsuperscript{11-12, 28} was used to optimize the electron densities of monomer fragments and the total energy of the system. In two-body, and sometimes three-body, calculations, the charge densities of the monomers that polarize the higher-order fragments are fixed,\textsuperscript{23-24} and this makes gradient calculation rather challenging. Until recently,\textsuperscript{29} only approximate gradient methods were available for the FMO model.\textsuperscript{30-31}

There are two major differences between the present VMB2/X-Pol method and the FMO approach. First, in VMB2/X-Pol, the effective Hamiltonian (Fock or Kohn-Sham matrices) for each fragment includes the contributions from the response of all the other fragments.\textsuperscript{14}
where $F^{I, \text{XPol}}_{\mu \nu}$ is an element of the X-Pol Fock matrix for fragment $I$, $F^{I,o}_{\mu \nu}$ is the Fock matrix element for fragment $I$ in vacuum, $I^I_{\mu \nu}$ is the one-electron integral due to the atomic charges of all other fragments, $X^I_i$ is a vector arising from the derivative of the interaction energy, and $\Lambda^I_i$ is a response density matrix. In FMO, however, each fragment is embedded in a static electrostatic field of the rest of the system. Mutual polarization is included through a charge-update procedure, and the Fock matrix for each fragment is written separately as

$$
F^{I,\text{NV}}_{\mu \nu} = F^{I,o}_{\mu \nu} - I^I_{\mu \nu} 
$$

Note that the total energy obtained using this non-variational (NV) procedure is not the lowest energy of the monomer system. Second, the mutual polarizations between each dimer fragment and the surrounding monomer fragments in eq 2 are also included in VMB2, and the two-body wave function $\Phi_{IJ} = \Psi_{IJ} \prod_{K \neq I,J}^N \Psi_K$ is variationally optimized as in eq 4, where $\Psi_{IJ}$ is a determinant for the $I$-$J$ dimer treated as a single fragment. Consequently, standard analytic gradient techniques can be directly used for the VMB2 energy. Although the additional optimization step increases computational costs, the convergence is generally fast requiring two to three iterations since this only represents a small perturbation to the optimized monomer densities.
VMB2/X-Pol is illustrated on the binding energy of four water hexamer clusters (Figure 4.1) and one water decamer complex (Figure 4.2). The hexamer coordinates are taken from Ref. [33], which were optimized using B3LYP/6-311++G(2d,2p), while the \((\text{H}_2\text{O})_{10}\) configuration was optimized with the TIP5P potential\textsuperscript{34} from the Cambridge cluster database.\textsuperscript{35} For each system, we used Hartree-Fock (HF) theory, second-order Möller-Plesset perturbation theory (MP2), the PBE1 density functional theory (DFT), and the M06 hybrid density functional in VMB2/X-Pol calculations. The 6-311G(d,p) basis set was used; larger basis sets are needed to yield more accurate results for water clusters,\textsuperscript{33, 36} but the main objective here is to match the energies from the fully delocalized calculations at the corresponding level. We used Mulliken population charge to represent the embedding potential for interfragment interactions;\textsuperscript{11, 14} other alternatives may also be used, including the exact Coulomb potential. All calculations were performed using a developmental version of the *Gaussian 09* program.\textsuperscript{37}

Table 4.1 lists the computed binding energies using the conventional approach for the entire system, and the X-Pol and VMB2 fragment methods at different theoretical levels. In VMB2/X-Pol, each water is considered as a monomer fragment. The binding energy for a water cluster is defined by

\[
\Delta E_b = E_N - N \cdot E_w^o
\]

where \(E_N\) is the total energy of the \(N\)-mer water cluster \((\text{H}_2\text{O})_N\), and \(E_w^o\) is the energy of an isolated water molecule. Table 4.1 shows that the binding energies for the \((\text{H}_2\text{O})_6\) complexes from the monomeric X-Pol (without empirical exchange and dispersion terms) have significant errors in comparison with results from calculations on the full system;
the errors range from 7 to 18 kcal/mol. Inclusion of VMB2 corrections greatly improve the accuracy in binding energy, reducing MUD and RMSD errors to about 2 kcal/mol or lower, similar to that found in the FMO approach.\textsuperscript{38} Importantly, at the VMB2 level, the errors in total binding energy converge to a similar level for all methods despite the large differences in the monomeric X-Pol (Table 4.1), suggesting that the VMB2 method is capable of capturing the key energy components that are missing in the monomer partitioning. For the decamer configuration, both X-Pol and VMB2 results are surprisingly good in comparison with the full electronic structure calculations, except using HF with which the X-Pol method overestimates binding energy by 48 kcal/mol (see below).

We have further examined the energy components that contribute to the VMB2 correction energy, \( \Delta E_2 \), using an energy decomposition analysis.\textsuperscript{8, 39} First, at the HF level, \( \Delta E_2 \) includes both contributions from exchange (X) and charge delocalization (C), also called charge transfer. If we antisymmetrize both the monomer X-Pol wave function \( \Phi_1 \) and the \( I-J \) dimer wave function \( \Phi_{IJ} \), exchange repulsion is fully included in both systems. Then, the energy difference between these two states dominantly arise from electronic delocalization, and we define it as the \( I-J \) pairwise charge transfer energy:

\[
E_{IJ}^C = E_{IJ}[\hat{A}\{\Phi_{IJ}^{\text{HF}}\}] - E_1[\hat{A}\{\Phi_1^{\text{HF}}\}]
\]  

(7)

where the superscript HF emphasizes that this term is only clearly defined at the HF level, \( E_{IJ}[\hat{A}\{\Phi_{IJ}^{\text{HF}}\}] \) and \( E_1[\hat{A}\{\Phi_1^{\text{HF}}\}] \) are, respectively, the total energies, including exchange, for the \( I-J \) dimer and the monomer system, \( \hat{A} \) is the antisymmetrizer, and
\( \Phi^\text{HF}_I \) and \( \Phi^\text{HF}_{IJ} \) are the monomer and \( I-J \) dimer HF X-Pol wave functions.\(^{11, 15}\) The difference between the VMB2 correction energy, \( \Delta E_{IJ} \) (eq 2), and the charge transfer term gives the exchange repulsion between fragments \( I \) and \( J \):

\[
E_{IJ}^X = \Delta E_{IJ} - E_{IJ}^C
\]  
(8)

The charge delocalization energy can also be evaluated using the generalized X-Pol (GX-Pol) method,\(^{41}\) but it is not considered here.

We note that \( \Delta E_2 \) also includes dispersion/correlation (D) effects using MP2 to represent the individual fragments. In this case, by definition, the VMB2 correction energies determined at the MP2 and HF levels provide an estimate of the dispersion energy between two fragments:

\[
E_{IJ}^D = \Delta E_{IJ}^{\text{MP2}} - \Delta E_{IJ}^{\text{HF}}
\]  
(9)

The total two-body (pairwise) charge delocalization, exchange and dispersion energies are obtained by summing up all pair interactions:

\[
E_2^C = \sum_{I > J}^N E_{IJ}^C
\]  
(10)

\[
E_2^X = \sum_{I > J}^N E_{IJ}^X
\]  
(11)

\[
E_2^D = \sum_{I > J}^N E_{IJ}^D
\]  
(12)

For VMB2/X-Pol calculations using DFT, the energy decomposition can no longer be clearly defined, especially when hybrid functionals are used. One could use the
corresponding Kohn-Sham function to estimate charge transfer and exchange energies, and then, attribute the remaining energy to correlation effects. Validation of such a decomposition scheme in DFT calculations is beyond the scope of the present study.

Listed in Table 4.2 are the energy components contributing to the VMB2 correction energy, which is the sum of the exchange and charge delocalization terms: \( \Delta E_2 = E_2^C + E_2^X \). Table 4.2 reveals that there is significant compensation effect between these two opposing factors. For structures with few hydrogen bonds restrained by the geometric arrangement as in the prism and cage configurations in the hexamer complex, the exchange and charge delocalization components are roughly balanced with relatively small \( \Delta E_2 \) energies. However, for structures with restrained hydrogen bonds as in the book and cyclic configuration (Figure 4.1), the exchange repulsion increases exponentially with a small shortening of the hydrogen bond distances, but the resonance delocalization varies more slowly. Then, we see a large increase in exchange repulsion energy. This is even more vividly reflected in the decamer structure, which was optimized using the empirical TIP5P potential. In this case, to compensate for polarization effect implicitly, hydrogen bond distances are typically about 0.2 Å shorter than high-level ab initio values. In the \((\text{H}_2\text{O})_{10}\) structure (Figure 4.2), all O-O distances are less than 2.80 Å with several pairs as short as 2.63 Å. On the other hand, in the \((\text{H}_2\text{O})_6\) prism configuration from B3LYP/6-311++G(2d,2p) optimization, half O-O distances are greater than 2.90 Å and only one is below 2.76 Å. The closer interfragment contacts in the TIP5P structure result in significant orbital overlap and a seemingly disproportionally large exchange repulsion energy relative to the hexamer complexes.
We have computed the exchange energy for the monomer system as the difference between the energy of an antisymmetrized X-Pol wave function,\(^8\) also known as block-localized wave function (BLW),\(^7\) and that of the energy of the X-Pol wave function. In the latter case, explicit two-electron integrals are used to account for the Coulomb energy with fixed X-Pol wave function; this definition is identical to the original Morokuma analysis.\(^{42}\) An FMO-based energy decomposition has been reported.\(^{43-44}\) It is interesting to compare the monomeric exchange energy with that obtained from pairwise additions (eq 11). The exchange energy for the monomer system is defined. The non-additivity is about 1 kcal/mol per water for the prism and cage structures of the water hexamer, and increases to 1.5 kcal/mol per water for the cyclic configuration, and to 2.3 kcal/mol per water in the decamer geometry. Factors contributing to the non-additivity include electronic delocalization and the consequential mutual polarization with the surrounding monomers. Total dispersion energies are estimated using MP2 for the entire system, and are compared with those obtained from pairwise additions. Table 4.2 shows that dispersion energies are remarkably additive, in contrast to the great sensitivity of exchange repulsion to the polarization of the wave function. This is also consistent with the success of a simple \(R^{-6}\) dependence used in molecular mechanics.

The results for the \((\text{H}_2\text{O})_6\) and \((\text{H}_2\text{O})_{10}\) clusters indicate that reasonably accurate results (with errors in the order of millihartree) can be obtained using the VMB2/X-Pol method. A number of steps may be taken to further improve the accuracy, especially in the treatment of electrostatic interactions for close neighboring fragments. It is always
possible to include higher-order terms, but the computational costs will soon become intractable for macromolecular systems. One possibility is to use a small buffer region in which the two electron integrals between different fragments are explicitly computed.\textsuperscript{45} Other possibilities include the use of different charge models such as electrostatic potential fitted charges or a multipole representation of the electrostatic potential, or the use of screened charges for close interactions.\textsuperscript{46-47} Importantly, the wave functions for one-body and all two-body terms are variationally optimized in VMB2 and X-Pol, making it straightforward to obtain analytic gradient without the additional coupled-perturbed Hartree-Fock step. Thus, the method can be useful for molecular dynamics simulations.
Table 4.1. Binding energies ($-\Delta E_b$) in kcal/mol from standard methods, X-Pol, and VMB2 calculations, mean unsigned deviations (MUD), and root-mean-square deviations (RMSD). The 6-311G(dp) basis set is used in all calculations.

<table>
<thead>
<tr>
<th>structure</th>
<th>HF $\Delta E_b$</th>
<th>MP2 $\Delta E_b$</th>
<th>PBE1 $\Delta E_b$</th>
<th>M06 $\Delta E_b$</th>
<th>X-Pol</th>
<th>VMB2</th>
<th>X-Pol</th>
<th>VMB2</th>
<th>X-Pol</th>
<th>VMB2</th>
<th>X-Pol</th>
<th>VMB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>prism</td>
<td>47.1</td>
<td>46.2</td>
<td>66.4</td>
<td>47.0</td>
<td>67.0</td>
<td>71.9</td>
<td>48.5</td>
<td>73.6</td>
<td>72.0</td>
<td>52.4</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>cage</td>
<td>45.9</td>
<td>44.9</td>
<td>64.9</td>
<td>47.4</td>
<td>65.3</td>
<td>70.4</td>
<td>49.3</td>
<td>71.9</td>
<td>70.7</td>
<td>53.3</td>
<td>70.8</td>
<td></td>
</tr>
<tr>
<td>book</td>
<td>45.1</td>
<td>42.5</td>
<td>63.2</td>
<td>51.0</td>
<td>61.4</td>
<td>68.7</td>
<td>53.0</td>
<td>67.4</td>
<td>67.3</td>
<td>57.4</td>
<td>66.0</td>
<td></td>
</tr>
<tr>
<td>cyclic</td>
<td>44.3</td>
<td>39.8</td>
<td>60.6</td>
<td>54.7</td>
<td>57.0</td>
<td>66.0</td>
<td>56.8</td>
<td>62.1</td>
<td>64.3</td>
<td>61.5</td>
<td>60.4</td>
<td></td>
</tr>
<tr>
<td>MUD</td>
<td>0.0</td>
<td>2.2</td>
<td>13.8</td>
<td>1.6</td>
<td>0.0</td>
<td>17.3</td>
<td>2.1</td>
<td>0.0</td>
<td>12.4</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD</td>
<td>0.0</td>
<td>2.7</td>
<td>14.7</td>
<td>2.0</td>
<td>0.0</td>
<td>18.2</td>
<td>2.3</td>
<td>0.0</td>
<td>14.1</td>
<td>2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\left(\text{H}<em>2\text{O}\right)</em>{10}$</td>
<td>-88.1</td>
<td>-125.6</td>
<td>-124.8</td>
<td>-121.7</td>
<td>-134.9</td>
<td>-128.8</td>
<td>-133.7</td>
<td>-131.7</td>
<td>-138.5</td>
<td>-130.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>0.0</td>
<td>-48.2</td>
<td>6.5</td>
<td>0.0</td>
<td>3.9</td>
<td>6.1</td>
<td>1.2</td>
<td>0.0</td>
<td>-6.8</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.2. Computed two-body variational energy correction ($\Delta E_2^2$), pairwise charge transfer ($\Delta E_2^C$), exchange ($\Delta E_2^X$) and dispersion ($\Delta E_2^D$), and monomeric exchange ($\Delta E_1^X$) and total dispersion energy ($\Delta E_{MP2}^D$) along with delocalization effects on pair additivity in exchange ($\Delta \Delta E_{nadd}^X$) and dispersion ($\Delta \Delta E_{nadd}^D$). All energies are given in kcal/mol. The 6-311G(dp) basis set is used in all calculations.

<table>
<thead>
<tr>
<th>energy</th>
<th>prism</th>
<th>cage</th>
<th>book</th>
<th>cyclic</th>
<th>(H$<em>2$O)$</em>{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta E_2^2$</td>
<td>2.9</td>
<td>5.0</td>
<td>11.4</td>
<td>18.2</td>
<td>54.7</td>
</tr>
<tr>
<td>$\Delta E_2^C$</td>
<td>-39.6</td>
<td>-39.9</td>
<td>-41.2</td>
<td>-42.2</td>
<td>-102.0</td>
</tr>
<tr>
<td>$\Delta E_2^X$</td>
<td>42.5</td>
<td>44.9</td>
<td>52.6</td>
<td>60.3</td>
<td>156.7</td>
</tr>
<tr>
<td>$\Delta E_2^D$</td>
<td>-18.1</td>
<td>-18.0</td>
<td>-16.5</td>
<td>-14.7</td>
<td>-40.1</td>
</tr>
<tr>
<td>$\Delta E_1^X$</td>
<td>36.6</td>
<td>39.1</td>
<td>45.3</td>
<td>50.8</td>
<td>134.3</td>
</tr>
<tr>
<td>$\Delta E_{MP2}$</td>
<td>-16.6</td>
<td>-16.6</td>
<td>-15.7</td>
<td>-13.8</td>
<td>-37.5</td>
</tr>
<tr>
<td>$\Delta \Delta E_{nadd}^X$</td>
<td>5.9</td>
<td>5.8</td>
<td>7.3</td>
<td>9.5</td>
<td>22.7</td>
</tr>
<tr>
<td>$\Delta \Delta E_{nadd}^D$</td>
<td>-1.5</td>
<td>-1.4</td>
<td>-0.8</td>
<td>-0.9</td>
<td>-2.6</td>
</tr>
</tbody>
</table>
Figure 4.1. Geometries for the four configurations of the water hexamer optimized with B3LYP/6-311++G(2d,2p).
Figure 4.2. Geometry of the water decamer structure optimized with the TIP5P potential.\textsuperscript{35}
Chapter 5. Projected Hybrid Orbitals: A Universal QM/MM Method

5.1. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) methods provide a convenient and practical procedure to study chemical processes in condensed-phases and biological systems.\textsuperscript{1-5} In application to macromolecular systems, critical to success is the treatment of the covalent boundary that separates a QM region from the remaining MM region.\textsuperscript{6-12} One may wonder that this question has been well resolved since combined QM/MM methods are being widely used in a diverse range of fields for a long time,\textsuperscript{3-5} and indeed, satisfactory approaches are available in special cases, especially when semiempirical quantum mechanical models are used.\textsuperscript{13,14} However, there is still a lack of general approaches for ab initio molecular orbital and density functional theory with the use of any arbitrary basis sets without introducing system-dependent parameters. Consequently, the treatment of QM and MM boundary remains an active subject of current research.\textsuperscript{12-42} In this article, we describe a systematic approach that employs a projected hybrid orbital (PHO) technique for treating covalent boundaries between QM and MM molecular fragments. The PHO approach is system-independent and can be applied to any basis sets in ab initio wave function theory and density functional theory.

There are three main criteria that may be used to validate a method for treating QM-MM boundaries.\textsuperscript{9} First, the electron-withdrawing power between the QM region and MM region is properly balanced such that the electronegativity of the boundary atom of the MM region closely mimics that of the full QM system treated by the same quantum mechanical model. This allows a smooth transition from the QM region into the MM region without altering the reactivity of the central part (i.e., the QM region) of interest.
Second, a well-defined QM/MM system should preserve the integrity of the system without introducing or eliminating any degrees of freedom of the original system. This also includes the need that electrostatic interactions from atoms in close proximity of the QM region be preserved. Finally, the boundary method is capable of yielding consistent molecular geometry in comparison with that when the system is either fully treated by the QM method or by the corresponding MM approximation.

Numerous methods have been developed for treating the QM-MM boundary. Generally, these techniques may be grouped into two main categories (Figure 5.1). The first category includes methods that introduce additional degrees of freedom into the system or alter the local electrostatic environment of the original system. The boundary that separates QM and MM regions is defined by a chemical bond, typically between two sp$^3$ carbon atoms; these two atoms are called frontier atoms. The so-called link-atom approach is a prototypical example of this category, in which the valency of the molecular fragment treated quantum mechanically is saturated by a hydrogen atom, i.e., link-atom. The hydrogen link-atom is typically placed along the chemical bond between the two frontier atoms connecting the QM and MM region and a standard bond length of 1.09 Å for a typical carbon-hydrogen bond is adopted. Aside from hydrogen, halogen-like atoms parameterized to mimic the covalent bonds of the carbon atom in the MM region have also been used. The addition of the link-atom into the system has a number of consequences, including force redistribution, kinetic energy and temperature adjustment in molecular dynamics simulations (or simply ignored), and removal of electrostatic over-polarization from near contacts. The latter issue is a major short-coming of the link-atom type approach since the atom that the link-atom...
replaces is too close to the QM fragment and its atomic partial charges from the MM force field must be removed and redistributed. Furthermore, the atomic charges on atoms directly connected to the MM frontier atom also need to be adjusted to maintain charge neutrality of the MM fragment. Clearly, the local alteration of partial atomic charges would affect the local electrostatic environment and polarization of the QM region, although the significance of this effect in real systems remains to be carefully examined.

Methods belonging to the second category do not introduce nor eliminate any degrees of freedom of the system, and they do not change the partial charges of any atoms assigned to the MM region. Thus, these methods can best maintain the electrostatic interactions in the original system. Here, the boundary between QM and classical regions is anchored on an atomic site, often, but not restricted to, an sp\(^3\) hybridized carbon atom. Thus, literally, such a boundary atom is both a “QM” atom, having the same basis functions and electrons as the rest of carbon atoms in the QM region, and an “MM” atom, keeping the partial charge assigned to this atom in the force field. The local self-consistent field (LSCF) method developed by Rivail and coworkers, in which three hybrid orbitals on the boundary atom are included in the SCF optimization with the remaining hybrid orbital fixed at a parameterized density, is an early example in this category. The method has been extended to a broad range of situations by Friesner and coworkers. An alternative is the generalization hybrid orbital (GHO) approach, in which the hybrid orbitals are consistently dependent on the local, instantaneous geometry during molecular dynamics simulation and the parameters are no longer system-dependent. A main difference between the LSCF and GHO method, however, is
the way the boundary atom is recognized in SCF optimizations (Figure 5.2); three bonds from the boundary atoms are optimized in the LSCF method, whereas just one covalent bond is part of the SCF procedure. Thus, the error caused by the boundary approximation on the QM region is minimized in the GHO method.

Parameterization of the GHO method is straightforward using semiempirical QM models, requiring only two \( (U_{ss} \text{ and } U_{pp}) \) of the 14 parameters to be adjusted in the standard AM1 and PM3 set for carbon.\(^{13,14}\) A straightforward extension of the GHO approach to ab initio methods becomes more cumbersome if different basis sets were used in the QM fragment and the boundary atom because numerous electronic integrals would need to be adjusted to maintain a similar electronegativity of the boundary atom as that in a full QM system.\(^{37,38}\) In this article, we present a new strategy to circumvent the need for parameterization of the electronic integrals of the boundary atom in the QM region. The main idea is to represent the core and valence electrons with a secondary, minimal basis set by projecting the original (primary) basis set used in the QM system.\(^{43}\) Then, the projected valence orbitals are transformed into a set of hybrid orbitals, defined in exactly the same way as that used in the GHO method.\(^{13,14}\) The hybrid orbital pointing towards the QM fragment from the boundary atom is included in the standard SCF optimization. Since the minimal basis set used in the present projected hybrid orbitals (PHO) is a closest representation of the original basis set, it retains the essential properties to have a balanced interaction with the QM fragment, and we found that it is possible to obtain good results without system-dependent parameters in the present approach.
In the following, we first describe the projected hybrid orbital method, in which we present two optimization procedures depending on the way the total Fock matrix is partitioned. Then, we present test cases to validate the present PHO method as a simple and general approach to model QM and MM covalent bond separation. Finally, we summarize the major findings of this work and highlight several perspectives in future applications.

5.2. Method

The goal of this study is to develop a general approach to treat the QM and MM covalent boundary with any basis set without introducing system-dependent parameters in combined QM/MM calculations using ab initio wave function theory (WFT) and density functional theory (DFT). The generalized hybrid orbital (GHO) method introduced previously provides a good starting point, but two major issues that were absent in semiempirical methods must be addressed in ab initio calculations. First, the hybrid orbital representation of the boundary atom is chemically intuitive and conceptually simple for QM/MM applications. However, the hybrid orbitals are not conveniently defined using a large basis set that includes split valence and diffuse functions. Thus, a coarse-graining-like approach to reduce the basis set size on the boundary atom would be desired. Here, we choose to use a secondary minimal basis set to best represent the original (primary), typically larger, basis set used in the full QM calculation. This is accomplished by projecting the large basis set of the boundary atom onto the minimal basis set, which defines a representation most closely resembling the original basis set, and thus, best preserving the electronegativity of the boundary atom in the original basis. Second, orbitals in semiempirical methods are assumed to be orthogonal, but in ab initio
calculations, the auxiliary orbitals need to be specifically orthogonalized with respect to the active orbitals in the QM region. This has been carefully studied in the development of the LSCF method and the GHO model. Here, we employ two procedures to enforce the orthogonality constraints.

In this section, we briefly review the definition of hybrid orbitals on a QM-MM boundary atom that separates the two regions. For clarity, we use only one boundary atom in the discussion, while generalization to any number of boundary atoms between the two regions is straightforward. Then, we describe a strategy for a minimal basis representation in the sense of least square resemblance of the original, larger basis set on the boundary atom. Next, the energy formulation is presented, along with the SCF procedure and the associated density and Fock matrix formation. Finally, we present the expression for the first analytic energy derivatives of the present PHO method.

5.2.1. Generalized Hybrid Orbitals

We consider a system that is partitioned into two molecular fragments across a boundary atom treated by, respectively, an electronic structure method and an MM force field. The partition of QM and MM fragments takes place at the boundary atom $C_B$, which is assumed to be an $sp^3$ carbon (Figure 5.2). However, the boundary atom is not necessarily restricted to carbon and $sp^3$ hybridization, and the method can be extended to other situations. The $C_B$ atom is bonded with three other MM atoms, denoted by the symbols $M_1$, $M_2$ and $M_3$, respectively (Figure 5.2).

The atoms in the QM region other than $C_B$ are defined as “full QM atoms”, as compared to the boundary atom $C_B$ whose representation is both quantum mechanical and classical. In the original GHO method presented at both semiempirical and ab initio
levels,13,14 there are four atomic valence orbitals on the C_B atom. Here, we also include the 1s core orbital (χ_c) in the present study (the exact nature of the five atomic orbitals on the boundary atom will be discussed in the next section).

The atomic orbitals, \{χ_c, s, p_x, p_y, p_z\}, on the boundary atom are transformed into a set of core χ_c, and valence hybrid orbitals η:

\[
\begin{pmatrix}
χ_c \\
η_B \\
η_1 \\
η_2 \\
η_3 \\
\end{pmatrix} = T^\dagger_b
\begin{pmatrix}
χ_c \\
2s \\
p_x \\
p_y \\
p_z \\
\end{pmatrix}
\]

(1)

where \(T^\dagger_b\) is the basis transformation matrix, which depends on the local geometry about the C_B atom and has been explicitly defined previously,14 with the addition of a unity in the diagonal element corresponding to the core orbital. Hybrid orbitals are split into two categories: two active orbitals, denoted by χ_c and η_B, with the latter pointing towards the QM frontier atom C_Q, and three auxiliary orbitals that are not variationally optimized in the SCF of the QM system, denoted by η_x, η_y and η_z, pointing roughly to the three MM neighbors \{M_1, M_2, M_3\}.

The five core and valence hybrid orbitals plus the N atomic basis functions \{χ_μ\} on the fully QM atoms form an \((N+5)\)-dimensional hybrid-orbital (HO) space. The transformation \(T_H\) that relates the AO space with the HO space is written as

\[
T_H = \begin{pmatrix}
I & 0 \\
0 & T_b
\end{pmatrix},
\]

(2)

where the transformation on the basis functions of fully QM atoms is just an identity matrix of dimension N.
5.2.2. Projected Hybrid Orbitals

When we use an arbitrarily large basis set to represent the QM region, the number of AOs on the boundary atom $C_B$ is generally greater than that of the minimal basis functions needed to define the GHO orbitals above. When polarization and diffuse functions are included, it is even more difficult to divide these atomic orbitals directly into active and auxiliary bases. Previously, we explored an approximate approach, called GHO-AIHF,\(^{37}\) in which the boundary atom was represented by the valence-only minimal basis, STO-3G(v), different from the basis set used for the rest of the system.\(^{37}\) The imbalance due to mixing different basis sets on the QM fragment and on the boundary atom is compensated by scaling the electronic integrals involving the STO-3G(v) basis. To eliminate the need to scale electronic integrals in such a mixed basis approach, we present an orbital projection technique to construct hybrid orbitals from the same basis on the boundary atom as that for the fully QM atoms for QM-MM boundaries.

Let $\{\chi\}$ be the primary basis set used for the atoms in the QM region, including the boundary atom $C_B$ that is treated equally as the rest of the “full QM atoms”, and $\{\zeta^b\}$ be the secondary minimal basis functions located only on the boundary atom. The essential step is to use basis projection to transform the primary basis set of the boundary carbon onto the secondary, minimal basis representation. One straightforward strategy is to use Mulliken’s modified atomic orbital (MAO) projection scheme,\(^{47,48}\) which has been applied to population analysis on a minimal basis set.\(^{43}\) In particular, the projection from the primary basis set to the minimal basis on the boundary atom is given by,

$$P_b = S^{-\frac{1}{2}}S(\tilde{S}S^{-1}\tilde{S})^{-\frac{1}{2}}$$  \hspace{1cm} (3)
where $S$ is the atomic orbital overlap of the primary basis on the boundary atom $\{ \chi^b \}$, and $\tilde{S}_{\mu\nu} = \langle \chi^b_{\mu} | \zeta^b_{\nu} \rangle$ is the rectangular overlap matrix between the primary basis $\{ \chi^b \}$ and the secondary basis $\{ \zeta^b \}$. In the present study, we report results obtained with the STO-3G basis as the secondary orbitals on the boundary atom, although any STO-nG could be used. Löwdin symmetric orthogonalization $S^{-\frac{1}{2}}$ is required to preserve the maximum resemblance between the original basis and the projected basis. 49 Further, normalization is applied to retain the unitary property of the projection transformation. To this end, the $N^b$ atomic orbitals in the primary basis is reduced to five STO-3G orbitals on the boundary atom $C_B$, and the $(N+N_b)$ total primary atomic orbitals of the QM fragment is reduced to $(N+5)$ mixed (primary and secondary) atomic orbitals. The overall transformation is a $(N+N_b) \times (N+5)$ rectangular matrix that consists of an identity transformation $(N \times N)$ on the fully QM atoms and the minimal basis projection of the boundary atom B:

$$T_P = \begin{pmatrix} I & 0 \\ 0 & P_B \end{pmatrix}$$  \hspace{1cm} (4)

The five STO-3G orbitals on the boundary carbon atom include one 1s core orbital, one 2s orbital and three 2p orbitals. Hybridization transformation defined in eq 1 is then performed on the 2s and three 2p orbitals. Thus, the overall basis set projection and hybridization transformation is given by

$$T_{PH} = T_PT_H$$  \hspace{1cm} (5)

Furthermore, the 1s core orbital can be assigned either as an active orbital along with the active hybrid orbital to participate in self-consistent field (SCF) optimization or as a frozen orbital that contributes to the external potential from the three auxiliary hybrid orbitals. In the present study, the 1s core orbital, the one active hybrid orbital $\eta_B$ and the
$N$ basis functions on the fully QM atoms are grouped together to form an $(N+2)$-dimensional active HO space for the SCF iteration.

### 5.2.3. Orthogonality Constraint

The overlap matrix in the $(N+5)$-dimensional HO space, $S_{N+5}^{HO}$, is related to the overlap matrix of the primary atomic orbital basis, $S$, by the successive projection and hybridization transformations:

$$S_{N+5}^{HO} = T_H^* T_p^* S T_p T_H.$$  \hspace{1cm} (6)

The active orbitals are generally not orthogonal to the auxiliary orbitals that do not participate in the SCF optimization:

$$\langle \phi^H_a | \eta_b \rangle \neq 0 \ (a = 1, \ldots, N + 2; b = 1, 2, 3),$$  \hspace{1cm} (7)

where $\phi^H_a$ and $\eta_b$ are the active and the three non-optimized, hybrid auxiliary MOs, respectively. The superscript $H$ is to emphasize that the molecular orbitals in the QM region are linear combinations of the primary atomic orbitals on the fully QM atoms and the core and active hybrid orbitals on the boundary atom.

Orthogonality constraint between auxiliary and active MOs is a general condition for both GHO\textsuperscript{37} and LSCF-type methods,\textsuperscript{31,34} which must be imposed in the optimization of the active MOs. Given the orthogonality, by construction, among the four hybrid orbitals on the boundary atom, the essential constraints of the projected hybrid orbital (PHO) method are

$$S_{ub} = \langle \chi_u | \eta_b \rangle = 0 \ (u = 1, 2, \ldots, N + 1; b = 1, 2, 3),$$  \hspace{1cm} (8)

where $\chi_u$ includes the atomic basis on the full QM atoms and the projected core orbital ($\chi_c$) on the boundary atom $C_B$, and $\eta_b$ are the three auxiliary hybrid orbitals.
Several types of orthogonalization techniques have been proposed in the GHO approach, and two of them are adopted in the present PHO method: (a) the symmetric global Löwdin orthogonalization (GLO) method in which all the orbitals in HO space are diagonalized, and (b) the projected basis method in which the auxiliary basis is projected out of the active space and the Fock matrix is diagonalized in this projected active basis.

(a). Global Löwdin orthogonalization (GLO)

Global Löwdin transformation produces a set of orthogonal hybrid orbitals (OHO) of the entire mixed QM and boundary subsystems, which is denoted by

\[
\begin{pmatrix}
\phi_1^L \\
\phi_2^L \\
\vdots \\
\phi_{N+2}^L \\
\eta_1^L \\
\eta_2^L \\
\eta_3^L
\end{pmatrix}
= T_{LO}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_{N+2} \\
\eta_1 \\
\eta_2 \\
\eta_3
\end{pmatrix},
\]

(9)

where \( \phi_i \) \((i = 1, \cdots, N + 2)\) specifies the nonorthogonal active hybrid basis, and \( \eta_j \) \((j = 1, 2, 3)\) denotes the three auxiliary hybrid basis functions on the boundary atom, and \( T_{LO} \) is the symmetric Löwdin orthogonalization matrix, that is

\[
T_{LO} = (S_{N+5}^{HO})^{-\frac{1}{2}}.
\]

(10)

The transformation makes all the OHOs resemble the original HOs in a least-square sense. The OHOs form an orthonormal set, thereby, the orthonormality between the three auxiliary orbitals with other active basis functions is satisfied in OHOs.

The total transformation matrix \( T \) that relates the original AO basis to the OHO basis is the matrix product of consecutive transformations that include projection, hybridization and orthogonalization:
The transformation relates the Fock matrix in the OHO basis to the Fock matrix expressed in the primary AO basis by

$$F_{N+5}^{OHO} = T_{GLO}^\dagger F_{AO}^{} T_{GLO}.$$  

(12)

Due to the well-known orthogonalization tail in the Löwdin transformation, the elements in the rows and columns corresponding to the auxiliary orbitals in $F_{N+5}^{OHO}$ are generally not zero. In principle, a further projection step is required to uncouple the two block matrices; however, these elements are simply dropped as an approximation for simplicity to reduce the Fock matrix to $(N+2)$-dimension in the active OHO space for subsequent diagonalization.

(b). Projected basis method

The second approach to enforce the orthogonality constraint is the projected basis method, in which successive Schmidt orthogonalization is performed on the $(N+2)$ active basis $|\chi_u\rangle$ to remove its linear dependency with the three auxiliary hybrid orbitals. This results in a new set of projected hybrid basis, denoted by $|\bar{\chi}_u\rangle$:

$$|\bar{\chi}_u\rangle = [1 - \Sigma_{b=1}^3 (S_{ub}^{HO})^2]^{-\frac{1}{2}}(|\chi_u\rangle - \Sigma_{b=1}^3 |\eta_b\rangle \langle \eta_b |\chi_u\rangle),$$  

(13)

$$u = 1, \cdots, N + 1$$

where $u$ is the index of active orbitals on the fully QM atoms and the 1s core orbital on the boundary atom C$_B$. $S_{ub}^{HO}$ is the overlap matrix between mixed active atomic and hybrid orbitals $\chi_u$ and auxiliary orbital $\eta_b$. The transformation matrix of projected basis is explicitly given by,
The projected basis operation makes the active basis orthogonal to the auxiliary orbitals, but still retains the property of strict localization in the active basis:

\[
\tilde{\mathcal{S}}_{ub} = \langle \tilde{x}_u | \eta_b \rangle = 0 \quad (u = 1, \ldots, N + 2; b = 1, 2, 3).
\]

The total transformation matrix of the projected basis method that relates the original AOs to the projected active hybrid basis is written as

\[
T_{PHO} = T_p T_H M_0,
\]

The Fock matrix defined in the projected hybrid space is obtained from the Fock matrix in primary atomic orbital basis via the transformation:

\[
F_{N+5}^{PH} = T_{PHO}^\dagger F_{PH}^{AO} T_{PHO}.
\]

By projecting the auxiliary orbitals out of the active space, all the elements in the rows and columns of auxiliary orbitals in the \((N+5)\)-dimensional Fock matrix \(F_{N+5}^{PH}\) are strictly zero, resulting in the reduced Fock matrix \(F_{N+2}^{PH}\) in the projected active hybrid basis fully uncoupled to the auxiliary block. In contrast to the global Löwdin orthogonalization scheme where the \((N+2)\)-dimensional Fock matrix is constructed approximately (without the additional projection step), the reduced Fock matrix in the
projected basis method is variationally determined and the analytic gradient can be computed in a more straightforward fashion.

5.2.4. Self-Consistent Field Procedure for the PHO Energy

To summarize the successive transformations and subsequent SCF steps, we provide a general procedure in the PHO calculation.

(a). Construct the minimal basis projection matrix $T_p$ and the hybridization matrix $T_H$ according to eqs 4 and 2, respectively.

(b). Construct the orthogonalization matrix, either using the global Löwdin transformation (eq 10), or the projected basis scheme (eq 14). The total transformation matrix $T_t$ is the matrix product of the consecutive operations that include projection, hybridization, and orthogonalization. Specifically, $T_t = T_{GLO}$ in the global Löwdin transformation (eq 11), and $T_t = T_{PHO}$ in the projected basis procedure (eq 17).

(c). Form the $(N+N_b)$-dimensional Fock matrix in the primary AO basis as in standard electronic structure calculations, and transform it into the $(N+5)$-dimensional Fock matrix in the HO basis in the PHO method using either eq 12 in the global Löwdin orthogonalization scheme or eq 18 in the projected basis method.

\[ F_{N+5}^{HO} = T_t^\dagger F^{AO} T_t, \]

(19)

(d). Diagonalize the $(N+2)$-dimensional block Fock matrix of the active orbitals in the QM region by dropping the columns and rows that correspond to the auxiliary hybrid orbitals from $F_{N+5}^{HO}$, and solve the Roothaan equation in the reduced active HO space to obtain the orbital coefficients $C_{N+2}^{HO}$:

\[ F_{N+2}^{HO} C_{N+2}^{HO} = \epsilon S_{N+2}^{HO} C_{N+2}^{HO}, \]

(20)

Note that in the global löwdin orthogonalization, $S_{N+2}^{HO}$ is the identity matrix already.
(e). Form the new density matrix $p_{N+2}^{HO}$ using $C_{N+2}^{HO}$, and append the electron densities of auxiliary orbitals to the diagonal terms to form the total density matrix in the $(N+5)$-dimensional HO space.

$$p_{N+5}^{HO} = \begin{pmatrix} p_{N+2}^{HO} & 1 - q_m/3 \\ 1 - q_m/3 & \end{pmatrix}, \quad (21)$$

where the charge density $q_m$ has been defined previously in the GHO method and is adopted in the present PHO approach.

(f). Transform the density matrix in the HO space back to the density matrix in the primary AO space:

$$p^{AO} = T_t p_{N+5}^{HO} T_t^\dagger \quad (22)$$

(g). Check if the density and total electronic energy of the combined QM/MM system are converged. If convergence is not yet achieved, the procedure returns to step (c) and the SCF iteration is repeated by another increment.

The total energy of QM/MM systems in the PHO method is determined as follows,

$$E = \sum_{\mu\nu} p^{AO}_{\mu\nu} H_{\mu\nu}^{eff} + \frac{1}{2} \sum_{\mu\nu} \sum_{\lambda\sigma} p^{AO}_{\mu\nu} p^{AO}_{\lambda\sigma} \left[ (\mu\nu, \sigma\lambda) - \frac{1}{2} (\mu\sigma, \lambda\nu) \right] \quad (23)$$

$$+ E_{nuc}^{QM} + E_{nuc}^{QM/MM} + E_{MM} + E_{et}^{et},$$

where the first two summations account for the QM electronic energy, in which all matrix elements are expressed in terms of the primary atomic orbitals, including the boundary atom, $E_{nuc}^{QM}$ and $E_{nuc}^{QM/MM}$ are the nuclear repulsion energy in the QM region and between QM and MM atoms, respectively, and $E_{MM}$ is the energy of the MM region. In eq 23, the
term $E_{b}^{tet}$ is a tetrahedral restoring potential of the boundary atom, which is introduced to model the effect of reduced electron-pair repulsions from the auxiliary orbitals that only have the electron density from the boundary atom rather than fully localized “bond” orbitals for the MM fragment. $E_{b}^{tet}$ is expressed by standard bond stretch terms:

$$E_{b}^{tet} = \sum_{b}^{6} K_{b}^{tet} (\theta_{b} - \theta_{tet}^{o})^{2}$$  \hspace{1cm} (24)

where $b$ specifies the 6 possible bond angles about the boundary atom, $\theta_{tet}^{o} = 109.47^{o}$, and $K_{b}^{tet} = 300 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. $K_{b}^{tet}$ may be considered as the only parameter of the PHO method, which is basis set and system-independent.

To accelerate the SCF convergence, we used an energy-based DIIS procedure to minimize a function that is based on the total energy. Note that in the PHO method, the density matrices are optimized in the active HO space, whereas the Fock matrix constructed in the AO space has an implicit dependence both on the optimized density and on the frozen density of the auxiliary orbitals. Because of the contribution from frozen densities, the popular DIIS scheme based on the commutator of the density and Fock matrices is not applicable to the PHO procedure.

5.2.5. Analytic First Gradient of the PHO Energy

The reduced Fock matrix in the projected active hybrid orbital basis is variationally determined, resulting in a straightforward derivation of the analytic first gradient. The same expression may be used as an approximation to the first derivatives of the energy computed using the global Lowdin orthogonalization method. The gradient of the PHO energy (eq. 23) with respect to the nuclear coordinates $\{R_{a}\}$ is defined by
The gradient includes contribution from the QM electronic energy, nuclear Coloumbic energy in the QM fragment and between QM and MM atoms, and the MM energy. All terms but the first are trivial to derive.

We use Hartree-Fock theory to illustrate the gradient calculation, arising from the QM electronic energy term:

\[
\frac{\partial E_{\text{QM}}}{\partial R_a} = 2 \sum_{\mu\nu} p^{AO}_{\mu\nu} \frac{\partial H_{\mu\nu}}{\partial R_a} + \sum_{\mu\nu\rho\sigma} \left(2 p^{AO}_{\mu\nu} p^{AO}_{\rho\sigma} \right) \times \frac{\partial G_{\mu\nu}}{\partial R_a}
\]

\[
+ 2 \sum_{\mu\nu} \frac{\partial P^{AO}_{\mu\nu}}{\partial R_a} \left(H_{\mu\nu} + G_{\mu\nu}\right)
\]

where the first two summations are the derivatives on the one-electron and two-electron integrals, respectively, which are readily obtained from standard HF gradient calculations. The main difference between the PHO gradient and the standard HF gradient is the “density force” terms, denoted by \( \frac{\partial P^{AO}_{\mu\nu}}{\partial R_a} \). Note that in standard HF, the “density force” term is transformed to energy weighted density matrix and computed indirectly, but in PHO, \( \frac{\partial P^{AO}_{\mu\nu}}{\partial R_a} \) has explicit dependency on the derivatives of the transformation matrix as follows:

\[
\frac{\partial P^{AO}_{\mu\nu}}{\partial R_a} = \frac{\partial(T_{PHO} P^{H}_{N+5} T_{PHO}^\dagger)}{\partial R_a}
\]

\[
= \frac{\partial T_{PHO}}{\partial R_a} P^{H}_{N+5} T_{PHO}^\dagger + T_{PHO} P^{H}_{N+5} \frac{\partial T_{PHO}^\dagger}{\partial R_a} + T_{PHO} \frac{\partial P^{H}_{N+5}}{\partial R_a} T_{PHO}^\dagger
\]
where the derivatives of the transformation matrix \( \partial T_{PHO}/\partial R_a \) is computed by the chain rule:

\[
\frac{\partial T_{PHO}}{\partial R_a} = \frac{\partial T_P}{\partial R_a} T_b M + T_p \frac{\partial T_p}{\partial R_a} M + T_p T_p \frac{\partial M_O}{\partial R_a}
\]

(28)

Note that \((\partial T_P/\partial R_a) = 0\), because the projection operation is on the boundary atom only and is invariant to coordinate changes. The explicit expressions on the specific terms \((\partial T_b/\partial q)\) and \((\partial M_O/\partial q)\) have been detailed previously in the GHO method.\(^{14,37}\)

The last term in eq 27 includes the “density term” in the \((N+5)\)-dimensional hybrid basis which is written as,

\[
\frac{\partial P_{N+5}^H}{\partial R_a} = \frac{\partial p_{N+2}^H}{\partial R_a} + \frac{\partial p_{bb}^H}{\partial R_a} = \frac{\partial P_{N+2}^H}{\partial R_a}
\]

(29)

where the fixed density terms \(P_{bb}^H\) do not contribute to the gradient, and for \((\partial P_{N+2}^H/\partial R_a)\), energy weighted density matrix is constructed variationally in the \((N+2)\)-dimensional projected active hybrid basis and is transformed back to AO representation.

5.3. Computational details

We have implemented the projected hybrid orbital (PHO) method into the Gaussian program, developmental version (gdv-H35).\(^{50}\) Single point energy calculations using PHO are available both for the global Löwdin orthogonalization and the projected basis method, and the exact first analytic gradient of PHO is available for the projected basis method. In PHO calculations, the QM part can be represented either by WFT or by DFT, whereas the MM region adopts the AMBER force field\(^{51}\) that has been implemented in the Gaussian program. We follow the same strategy in the original GHO scheme\(^{13}\) to determine the MM energy terms: the general rule is that any MM energy terms containing at least a single MM atom will be retained. Therefore, for bonded
interactions between QM and MM atoms, i.e., bond stretching, angle bending and torsional terms, we discard covalent terms among all QM atoms, but preserve MM energy terms that contain one or more MM atoms. The non-bonded van der Waals interaction between QM and MM atoms is fully counted in MM energy expression, whereas the electrostatic interaction between QM and MM atoms is determined quantum mechanically by including the partial charges of all MM atoms in the QM/MM interaction Hamiltonian.

The performance of the PHO method is examined in various aspects as follows.

(a) *Geometry and charge distribution of ethane.* We investigated the optimized structure of ethane using PHO with different QM methods and basis sets, making use of HF, B3LYP and M06-2X and the STO-3G, 6-31G(d), 6-311+G(d,p) and aug-cc-pVTZ basis sets and an example. This simple test in fact is the most crucial to validate the balance of electronegativity between the two methyl groups, one of which is treated as the MM boundary. In addition, the force constant in the tetrahedral restraining term of eq 24 was obtained to yield agreement in the optimized geometries and partial charges with the corresponding full QM calculations.

(b) *Equilibrium geometries.* To validate the performance of the PHO method and the tetrahedral restraining term, we performed geometry optimizations on a test set that consists of various alkane and molecules with different functional groups. The optimized geometries from PHO calculations using HF with different basis sets are compared with the corresponding standard full HF results.
(c) **Torsion energy profile.** We examined the potential energy profile of about rotation around the central C2-C3 bond of n-butane both at HF/3-21G and at HF/6-31G(d) level. The PHO results are compared with standard QM calculations and MM results. Since the purpose here is to compare the difference between PHO and full ab initio results, the use of the relatively small basis is reasonable.

(d) **Energetics.** We performed PHO single-point energy calculations at the HF/6-311+G(d,p) optimized geometries to determine proton affinities for a set of organic molecules using HF at both 6-31G(d) and 6-311+G(d,p) basis sets. Both orthogonalization schemes, global Löwdin orthogonalization and projected basis, were examined.

### 5.4. Result and discussion

**A. Ethane and the optimization of the PHO method**

As in the original development of the GHO method, ethane was selected as the prototypical target for developing the present PHO algorithm. Although ethane might be considered to be too small for this purpose, it, in fact, provides the most direct test of the electron-withdrawing ability of a boundary model simply by inspecting the net partial charge in each methyl group since a balanced QM and MM division would have a minimal amount of charge transfer across the nonpolar bond. The system was partitioned into two methyl groups, one into the QM region and the other into the MM region. The carbon atom of the latter was treated as the boundary atom (C_B). To test the PHO method, we used Hartree-Fock (HF) theory and two popular hybrid density functionals, B3LYP and M06-2X, combined with a few representative basis sets, including STO-3G, 6-
31G(d), 6-311+G(d,p) and aug-cc-pVTZ to illustrate its performance. In each QM model and basis set combination, the optimized structure and electronic properties obtained from the hybrid calculations were compared with the results from the corresponding full QM results.

Table 5.1 lists the mean unsigned errors (MUEs) in bond lengths and angles of ethane between the hybrid QM/MM-PHO method and full QM calculations over HF, B3LYP and M06-2X optimizations using four different basis sets. The range of these basis functions are deemed to be sufficient to illustrate the generality of the PHO method here. Both results obtained with and without the inclusion of the $E_B^{tet}$ term in eq 24 are shown. First, the optimized bond lengths both in the QM region and the MM region are in good accord with the full QM results, whereas the C-C_B bond across the two regions exhibits greater deviations, ranging from 0.054 to 0.155 Å. The errors in Table 5.1 do not show systematic dependence with respect to the size of the basis set, although it is important to distinguish that the C-H bond reports the perturbation to the QM calculation due to the boundary atom and that results of the C_B-H_B bond are dictated by the MM force field. We attribute the relatively greater errors in the C-C_B bond length to the restriction that the projected orbitals are not relaxed in the SCF optimization. The error trends in Table 5.1 have been found previously in the semiempirical and ab initio GHO-AIHF models, and we consider that they are acceptable in a combined QM/MM method.

Without introducing any correction terms (Table 5.1, w/o column), we found that the bond angles about the boundary atom C_B can have significant errors in QM/MM geometry optimization. The origin of this discrepancy is due to the use of the system-
independent auxiliary hybrid orbitals that are not optimized for the specific bonding environment. In PHO, as well as the original GHO, the three auxiliary orbitals are represented by half bond-orbitals with an effective charge density of \((1-q_B/3)\), where \(q_B\) is the partial atomic charge on the boundary atom defined in the MM force field. The electron-pair repulsions are much weaker among the three auxiliary orbitals than that with the C-C\(_B\) bond orbital in the full QM system, resulting in a locally distorted tetrahedral geometry. To correct this local structural distortion, a tetrahedral restraining potential, \(E_{tet}^t\), is introduced (eq 24). As it turns out, a single parameter (the force constant) is sufficient for all basis sets examined. With the inclusion of this term, which is an intrinsic term of the PHO method, the mean unsigned errors in the optimized bond angles are within the same error range as that in all other bond types (Table 5.1, PHO column).

The electronic polarization between the two methyl groups of ethane represented, respectively, by the QM and MM models was characterized by the estimated partial atomic charges obtained from the hybrid method and the corresponding full QM calculation. Because of symmetry, the net atomic charge of each methyl fragment should be zero, irrespective of the specific charge model used to derive these charges, and this provides a simple, but most critical, test of the relative electron-withdrawing power for a balanced QM and MM partition. Two charge models, namely Mulliken population analysis (MPA) and Charge Model 5 (CM5), were used; the former is known to be a poor charge model especially for large basis sets, whereas the latter is a parametrized approach that shows remarkable stability across different basis sets and theoretical models.
Figures 5.3 and 5.4 illustrate deviations of the PHO model in partial charge from the full QM results, i.e., $q = q(\text{PHO}) - q(\text{QM})$, based on the MPA and CM5 analysis. Not surprisingly, the partial atomic charges on the individual atoms, the frontier atoms $C_A(\text{QM})$ and $C_B(\text{QM/MM})$ from MPA are very sensitive to the basis set used and have large differences between the PHO and full QM results (Figure 6.3), but the total net charges of the methyl groups are quite stable. Importantly, the total charge on the $C_BH_3$ group in the hybrid method exhibits only small deviations from the ideal value, indicating that charge transfer between QM and MM fragments is relatively small. The largest errors came from HF using 6-31G(d) and aug-cc-pVTZ basis sets, which showed a net charge accumulation of about 0.07 e. The partial charges determined using the CM5 method have small differences between the PHO calculation and the full QM method not only for individual atoms, but also for the fragments. For all methods examined, the PHO deviation is less than 0.05 e from the full QM results. As in the $C_A$-$C_B$ bond, the small deviation may be attributed to the projected secondary orbitals that are not further optimized in the molecular environment. This restriction can be relaxed, which will be addressed in a forthcoming study, but the results displayed in Figures 5.3 and 5.4 suggest that the small difference is acceptable for QM/MM applications.

**B. Molecular geometries and further validation of the PHO method.**

To validate the performance of the PHO method, we constructed a set of twenty small compounds that include alkanes and molecules with different functional groups (Supporting Information). The optimized geometries from the hybrid PHO-QM/MM method were compared against results from HF treatment of the full system with three basis sets, STO-3G, 6-31G(d), and 6-311+G(d,p). We focus on bond lengths and bond
angles associated with the frontier atom \( C_A \) in the QM fragment and the boundary atom \( C_B \) of the MM fragment (Table 5.2). The MUEs in bond length are about 0.005 Å for covalent bonds between \( C_A \) and atoms in the QM region (\( C_A-Q \)), 0.03 Å between \( C_B \) and atoms in the MM region (\( C_B-M \)), and 0.068 to 0.158 Å in the QM/MM frontier bond (\( C_A-C_B \)). For bond angles in the QM region, the MUEs are less than 1.5°, whereas the errors are slightly larger for bond angles about \( C_B \). Overall, the geometrical results in Table 5.2 have mean unsigned errors of similar magnitude as that in ethane used to define the PHO method.

To illustrate the variation of the frontier bond between QM and MM regions as it is influenced by the neighboring functional groups, the optimized bond lengths, \( C_A-C_B \), for some key functional groups are compared between PHO and standard HF methods using different basis sets (Table 5.3). Although the optimized frontier bonds from the PHO method are uniformly longer than the corresponding HF results, the trends due to functional group substitutions are in reasonable accord.

**C. Torsional potential energy profile of \( n \)-butane**

To test the ability of the PHO method to reproduce the shape of the potential energy surface as well as the equilibrium geometries, we investigated the internal rotation of \( n \)-butane with PHO at both HF/3-21G and HF/6-31G(d) level. In addition, we examined different QM/MM partition schemes by placing the \( C_B \) atom at C2, C3 and C4 carbons, corresponding to the enlarged QM region from a methyl to an ethyl and to a propyl group. Pure HF results at both 3-21G and 6-31G(d) level and MM results using the OPLS force field are also included for comparison. As shown in Figure 5.5 and Figure 5.6, pure QM calculations quite similar results pure MM calculations. All PHO
calculations predict reasonable energy for the Gauche conformation, however, they report different results for the potential energy barrier. For PHO calculation with boundary atom placed at C4, the predicted potential energy barrier is about 1.0 kcal/mol larger than the QM results. When the boundary atom is placed at C2 or C3 position, the predicted energy barrier is lower than that in pure QM or MM calculations. In these cases, the QM fragment does not contain the torsional terms around the C2-C3 bond, and special dihedral MM parameters for torsions that involve frontier and boundary atoms are desirable.

**D. Proton affinities**

Finally, we study the energetic properties of the PHO method by conducting the calculation of proton affinities (PA) over a wide range of organic molecules, including carboxylate anions, alcoholic anions and amines. In present work, proton affinity is defined as the zero-point-exclusive energy difference between the protonated state (denoted X or XH+) and the deprotonated state (denoted X− or X). The structures of both states for each species are optimized at HF/6-311+G(d,p) level. Subsequent single point energy calculations are performed based on the optimized structure to yield the reference values of PA at HF/6-31G(d) and HF/6-311+G(d,p) level. Single-point PHO energy calculations using both orthogonalization schemes, global löwdin orthogonalization (GLO) and projected basis (PrB), are also carried out at each level. Table 5.4 and 5.5 list reference values and PHO results of PA at HF/6-31G(d) and HF/6-311+G(d,p) level, respectively.

Table 5.4 lists the PA values for standard HF and PHO calculations over all selected species at HF/6-31G(d) level. The deprotonated state of each species is specified
in the first column, where the boundary atom is denoted by $C_b$, and atoms to the right of the $C_b$ are defined as the QM region. As the results show, the mean unsigned error (MUE) for all the species is 3.8 kcal/mol using PrB compared to the HF reference values, and is 3.6 kcal/mol using GLO. When the boundary atom is moving away from position of the protonated site to -position, predicted proton affinity improves significantly with MUE decreasing from 5.5 kcal/mol to 2.5 kcal/mol in PrB, and the MUE decreasing from 5.3 kcal/mol to 2.9 kcal/mol in GLO. When the boundary atom is placed at -position or even further, the MUE of PA is less than 1.0 kcal/mol. These improvements are consistent with the common insight that the partition of QM/MM boundary should be at least two bonds away from the QM active site.

The sensitivity of PHO energy on basis set is further tested by the calculations on whole set of PA species at a larger basis set of 6-311+G(d,p) (Table 5.5). With the larger basis set, MUE of PA is 2.6 kcal/mol for PrB, and is 4.8 kcal/mol for GLO, showing similar MUE compared with that at HF/6-31G(d) level. Overall, MUE of all PHO calculations, with different basis sets and QM method, falls in the range of 3 to 5 kcal/mol, and this is comparable to the typical error on PA in other QM/MM boundary treatments: Amara et al. used the link atom model and got the error $\sim$ 3 kcal/mol on PA;\textsuperscript{17} Zhang et al. proposed the pseudobond approach with various basis sets and yielded MUE in PA of 2.9 to 7.7 kcal/mol.\textsuperscript{23} In the GHO-AIHF method with parameterized STO-3G(v) basis on the boundary carbon, the error was reported to be 2.6 kcal/mol with the MIDI!

5.5. Conclusion and Perspective

In this article, we describe a systematic approach that employs a projected hybrid orbital (PHO) technique for treating covalent boundaries between QM and MM
molecular fragments. The PHO approach can be used in ab initio wave function theory with any basis set without introducing empirical parameters. Three important issues are investigated in the PHO approach: (a) the representation of core and valence electrons on the boundary carbon atom using a secondary minimal basis by projecting the regular, primary basis set, (b) the constraint of orthogonality between the auxiliary orbitals with the active orbitals via the global Löwdin orthogonalization and the projected basis method, and (c) the analytic first gradient of the PHO energy based on the projected basis method.

The PHO method is then applied to geometry optimization on ethane using a variety of electronic structure methods and basis sets. A single parameter of force constant $K_B^{tet} = 300 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, which is basis set and system independent, is introduced to represent the tetrahedral restraining potential, $E_B^{tet}$. Comparison with standard results shows that PHO is a robust and balanced QM/MM scheme that preserves the structural and electronic properties of ethane when the covalent boundary is split. The introduction of $E_B^{tet}$ is further validated via geometry optimization on a set of small organic compounds that contain alkanes and molecules with different functional group. The mean unsigned errors on bond lengths and bond angles of the test set are similar in magnitude as those in ethane. The performance of PHO on energetics is investigated by a systematic study on proton affinity over a wide range of carboxylates, alcohols and amines with different electronic structure theories and basis sets. Results show that the PHO scheme yields MUEs between 3 to 5 kcal/mol for both the projected basis method, and the global Löwdin orthogonalization in all cases. This is comparable to the 3-7
kcal/mol MUE on the calculation of proton affinity in other QM/MM schemes with boundary treatment.

The series of transformation in the PHO method are currently performed on a sp$^3$ carbon atom. We are devising a general extension to other types of boundary atom with a more universal hybridization scheme. Note that in the current PHO method, the projection and hybridization is based on modified Mulliken atomic orbitals (MAO).\textsuperscript{47,48} There is another approach to form hybrid basis called the polarized atomic orbitals (PAO) scheme.\textsuperscript{55} The idea was proposed by Martin Head-Gordon and coworkers to transform a large “secondary” basis set variationally to a small “primary” basis set. The key difference between the PAO and MAO schemes is that in the former scheme, the primitive expansion coefficients of small basis sets are not fixed but variationally optimized. In principle, the projection and subsequent hybridization in MAO can be combined into the variational optimization of PAO, allowing flexible representation on different types of boundary atoms. This issue will be explored in the subsequent work.

Beside, the PHO method can also be applied to fragment-based QM methods, such as X-Pol.\textsuperscript{56-59} Hybrid orbitals on the boundary atoms can be divided evenly between the two connected fragments. In a typical X-Pol calculation with boundary Carbon atom, the two active hybrid orbitals are optimized in the SCF iterations of current "QM" fragment, whereas the two auxiliary orbitals serve as effective core potential along with other "MM" fragments. The self-consistent calculation on every fragment ensures converged electron density distribution on the boundary atom as well as the whole system. The derivation and implementation of PHO in the X-Pol model is undergoing.
Table 5.1. Mean unsigned errors (MUE) of bond lengths (Å) and bond angles (deg) of ethane between results from the hybrid QM/MM optimizations and the corresponding full QM calculations with (PHO) and without inclusion (w/o) of the tetrahedral restraining potential, $E_B^{tet}$. For each basis set listed, the MUE is averaged over HF, B3LYP and M062X calculations. The subscript B indicates that the atom is associated with the boundary atom in the MM region.

<table>
<thead>
<tr>
<th></th>
<th>STO-3G</th>
<th>6-31G(d)</th>
<th>6-311+G(d,p)</th>
<th>aug-cc-pVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w/o</td>
<td>PHO</td>
<td>w/o</td>
<td>PHO</td>
</tr>
<tr>
<td>C-H</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>C-C_B</td>
<td>0.054</td>
<td>0.072</td>
<td>0.140</td>
<td>0.155</td>
</tr>
<tr>
<td>C_B-H_B</td>
<td>0.012</td>
<td>0.013</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>H-C-H</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>H-C-C_B</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>C_A-C_B-H_B</td>
<td>12.8</td>
<td>1.9</td>
<td>8.6</td>
<td>0.4</td>
</tr>
<tr>
<td>H_B-C_B-H_B</td>
<td>15.8</td>
<td>2.0</td>
<td>10.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Table 5.2. Mean unsigned errors (MUE) in bond length (Å) and bond angle (deg) between hybrid PHO-QM/MM and full Hartree-Fock optimizations for a set of 20 small compounds. Q indicates an atom present in the QM region, M is an atom treated classically by an MM model, and C_A and C_B are the frontier atoms in the QM and MM regions.

<table>
<thead>
<tr>
<th></th>
<th>STO-3G</th>
<th>6-31G(d)</th>
<th>6-311+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_A-Q</td>
<td>0.005</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>C_A-C_B</td>
<td>0.068</td>
<td>0.158</td>
<td>0.108</td>
</tr>
<tr>
<td>C_B-M</td>
<td>0.022</td>
<td>0.026</td>
<td>0.029</td>
</tr>
<tr>
<td>Q-C_A-Q</td>
<td>1.3</td>
<td>1.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Q-C_A-C_B</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>C_A-C_B-M</td>
<td>2.6</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>M-C_B-M</td>
<td>2.4</td>
<td>1.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Table 5.3. Comparison of bond length $C_A$-$C_B$ of selected molecules between PHO and standard HF results at different basis sets. $C_b$ is the boundary atom, and atoms to the right of $C_b$ belong to the MM region.

<table>
<thead>
<tr>
<th></th>
<th>STO-3G</th>
<th></th>
<th>6-31G(d)</th>
<th></th>
<th>6-311+G(d,p)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PHO</td>
<td>HF</td>
<td>PHO</td>
<td>HF</td>
<td>PHO</td>
<td>HF</td>
</tr>
<tr>
<td>HOOC-C$_3$H$_3$</td>
<td>1.589</td>
<td>1.537</td>
<td>1.619</td>
<td>1.502</td>
<td>1.569</td>
<td>1.500</td>
</tr>
<tr>
<td>HCOCH$_2$-C$_3$H$_3$</td>
<td>1.616</td>
<td>1.542</td>
<td>1.629</td>
<td>1.504</td>
<td>1.578</td>
<td>1.502</td>
</tr>
<tr>
<td>C$_6$H$_5$-C$_3$H$_3$</td>
<td>1.596</td>
<td>1.527</td>
<td>1.659</td>
<td>1.512</td>
<td>1.613</td>
<td>1.511</td>
</tr>
<tr>
<td>HOCH$_2$C$_3$H$_3$</td>
<td>1.613</td>
<td>1.547</td>
<td>1.675</td>
<td>1.522</td>
<td>1.622</td>
<td>1.520</td>
</tr>
<tr>
<td>HOOCCCH$_2$-C$_3$H$_3$</td>
<td>1.611</td>
<td>1.539</td>
<td>1.681</td>
<td>1.524</td>
<td>1.631</td>
<td>1.523</td>
</tr>
<tr>
<td>HOCH$_2$CH$_2$-C$_3$H$_3$</td>
<td>1.613</td>
<td>1.540</td>
<td>1.692</td>
<td>1.528</td>
<td>1.642</td>
<td>1.528</td>
</tr>
<tr>
<td>H$_2$NCH$_2$-C$_5$H$_3$</td>
<td>1.615</td>
<td>1.546</td>
<td>1.690</td>
<td>1.529</td>
<td>1.639</td>
<td>1.528</td>
</tr>
<tr>
<td>H$_2$NCH$_2$CH$_2$-C$_5$H$_3$</td>
<td>1.614</td>
<td>1.540</td>
<td>1.693</td>
<td>1.528</td>
<td>1.643</td>
<td>1.528</td>
</tr>
<tr>
<td>HCO-C$_4$H$_5$</td>
<td>1.594</td>
<td>1.536</td>
<td>1.696</td>
<td>1.533</td>
<td>1.642</td>
<td>1.532</td>
</tr>
<tr>
<td></td>
<td>HF</td>
<td>PrB</td>
<td>Diff</td>
<td>GLO</td>
<td>Diff</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>-----------</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CO}_2^-$</td>
<td>366.7</td>
<td>370.7</td>
<td>4.0</td>
<td>354.1</td>
<td>-12.6</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CO}_2^-$</td>
<td>367.5</td>
<td>372.8</td>
<td>5.3</td>
<td>357.8</td>
<td>-9.7</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-CO}_2^-$</td>
<td>367.5</td>
<td>371.9</td>
<td>4.4</td>
<td>369.2</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CH}_2\text{-CO}_2^-$</td>
<td>366.3</td>
<td>373.1</td>
<td>6.8</td>
<td>370.7</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-CH}_2\text{-CO}_2^-$</td>
<td>366.3</td>
<td>369.1</td>
<td>2.8</td>
<td>367.8</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>$(\text{CH}_3)_2\text{H}_2\text{-CO}_2^-$</td>
<td>365.4</td>
<td>375.5</td>
<td>10.1</td>
<td>361.2</td>
<td>-4.2</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}(\text{CH}_3)\text{-CO}_2^-$</td>
<td>365.4</td>
<td>371.9</td>
<td>6.5</td>
<td>369.6</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-H}_2\text{C}_b\text{-CO}_2^-$</td>
<td>405.4</td>
<td>410.8</td>
<td>5.4</td>
<td>403.4</td>
<td>-2.0</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CH}_2\text{-CO}_2^-$</td>
<td>404.4</td>
<td>410.9</td>
<td>6.5</td>
<td>401.9</td>
<td>-2.5</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-CH}_2\text{-CO}_2^-$</td>
<td>404.4</td>
<td>406.9</td>
<td>2.5</td>
<td>408.0</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}(\text{CH}_3)\text{-CO}_2^-$</td>
<td>403.3</td>
<td>409.9</td>
<td>6.6</td>
<td>402.3</td>
<td>-1.0</td>
<td></td>
</tr>
<tr>
<td>$(\text{H}_3\text{C}_b)_2\text{C}(\text{CH}_3)_2\text{-CO}_2^-$</td>
<td>401.8</td>
<td>398.6</td>
<td>-3.2</td>
<td>390.5</td>
<td>-11.3</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}(\text{CH}_3)_2\text{-CH}_2\text{-CO}_2^-$</td>
<td>399.0</td>
<td>402.9</td>
<td>3.9</td>
<td>401.9</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CH}_2\text{-CH}_2\text{-CO}_2^-$</td>
<td>404.1</td>
<td>408.1</td>
<td>4.0</td>
<td>406.9</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-CO}_2^-$</td>
<td>404.1</td>
<td>405.1</td>
<td>1.0</td>
<td>403.8</td>
<td>-0.3</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-NH}_2$</td>
<td>228.2</td>
<td>231.7</td>
<td>3.5</td>
<td>230.1</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-NH}_2$</td>
<td>231.5</td>
<td>236.1</td>
<td>4.6</td>
<td>235.2</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-NH}_2$</td>
<td>231.5</td>
<td>232.1</td>
<td>0.6</td>
<td>228.2</td>
<td>-3.3</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CH}_2\text{-NH}_2$</td>
<td>232.7</td>
<td>233.7</td>
<td>1.0</td>
<td>230.3</td>
<td>-2.4</td>
<td></td>
</tr>
<tr>
<td>$\text{H}_3\text{C}_b\text{-CH}_2\text{-CH}_2\text{-NH}_2$</td>
<td>232.7</td>
<td>233.4</td>
<td>0.7</td>
<td>232.5</td>
<td>-0.2</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-CH}_2\text{-H}_2\text{C}_b\text{-CH}_2\text{-NH}_2$</td>
<td>233.5</td>
<td>233.4</td>
<td>-0.1</td>
<td>230.1</td>
<td>-3.4</td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_3\text{-H}_2\text{C}_b\text{-CH}_2\text{-CH}_2\text{-NH}_2$</td>
<td>233.5</td>
<td>234.2</td>
<td>0.7</td>
<td>233.7</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td><strong>MUE</strong></td>
<td><strong>3.8</strong></td>
<td></td>
<td></td>
<td><strong>3.6</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.5. Proton affinities for carboxylate anions, alcoholic anions and amines at HF/6-311+G(d,p) level. All energies are in Kcal/mol. The symbol C_b indicates the boundary atom for GHO calculations. Atoms on the right of C_b are treated quantum mechanically.

PrB denotes projected basis, and GLO denotes global lowdin orthogonalization.

<table>
<thead>
<tr>
<th>Species</th>
<th>HF</th>
<th>PrB</th>
<th>Diff</th>
<th>GLO</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₃C₆-CO₂⁻</td>
<td>361.3</td>
<td>360.8</td>
<td>-0.5</td>
<td>346.7</td>
<td>-14.6</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CO₂⁻</td>
<td>361.3</td>
<td>361.7</td>
<td>0.4</td>
<td>349.4</td>
<td>-11.9</td>
</tr>
<tr>
<td>H₃C₆-CH₂-CO₂⁻</td>
<td>361.3</td>
<td>363.4</td>
<td>2.1</td>
<td>364.0</td>
<td>2.7</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CH₂-CO₂⁻</td>
<td>361.0</td>
<td>364.9</td>
<td>3.9</td>
<td>366.3</td>
<td>5.3</td>
</tr>
<tr>
<td>H₃C₆-CH₂-CO₂⁻</td>
<td>361.0</td>
<td>361.7</td>
<td>0.7</td>
<td>363.6</td>
<td>2.6</td>
</tr>
<tr>
<td>(CH₃)₂-H₃C₆-CO₂⁻</td>
<td>360.6</td>
<td>363.1</td>
<td>2.5</td>
<td>350.8</td>
<td>-9.8</td>
</tr>
<tr>
<td>H₃C₆-CH(CH₃)-CO₂⁻</td>
<td>360.6</td>
<td>364.6</td>
<td>4.0</td>
<td>365.8</td>
<td>5.2</td>
</tr>
<tr>
<td>H₃C₆-H₂C-O⁻</td>
<td>397.2</td>
<td>399.1</td>
<td>1.9</td>
<td>393.3</td>
<td>-3.9</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CH₂-O⁻</td>
<td>396.7</td>
<td>404.2</td>
<td>7.3</td>
<td>396.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>H₃C₆-CH₂-CH₂-O⁻</td>
<td>396.7</td>
<td>398.7</td>
<td>2.0</td>
<td>398.3</td>
<td>1.6</td>
</tr>
<tr>
<td>H₃C₆-CH(CH₃)-O⁻</td>
<td>395.7</td>
<td>398.3</td>
<td>2.6</td>
<td>391.3</td>
<td>-4.4</td>
</tr>
<tr>
<td>(H₃C₆)-C(CH₃)₂-O⁻</td>
<td>394.5</td>
<td>393.6</td>
<td>-0.9</td>
<td>380.3</td>
<td>-14.2</td>
</tr>
<tr>
<td>H₃C₆-C(CH₃)₂-CH₂-O⁻</td>
<td>393.4</td>
<td>397.5</td>
<td>4.1</td>
<td>386.9</td>
<td>-6.5</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CH₂-CH₂-O⁻</td>
<td>396.4</td>
<td>398.6</td>
<td>2.2</td>
<td>400.8</td>
<td>4.4</td>
</tr>
<tr>
<td>H₃C₆-CH₂-CH₂-CH₂-O⁻</td>
<td>396.4</td>
<td>396.4</td>
<td>0.0</td>
<td>397.6</td>
<td>1.2</td>
</tr>
<tr>
<td>H₃C₆-NH₂</td>
<td>227.7</td>
<td>222.4</td>
<td>-5.3</td>
<td>229.3</td>
<td>1.6</td>
</tr>
<tr>
<td>CH₃-H₂C₆-NH₂</td>
<td>231.0</td>
<td>225.9</td>
<td>-5.1</td>
<td>235.5</td>
<td>4.5</td>
</tr>
<tr>
<td>H₃C₆-CH₂-NH₂</td>
<td>231.0</td>
<td>228.5</td>
<td>-2.5</td>
<td>228.4</td>
<td>-2.6</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CH₂-NH₂</td>
<td>232.2</td>
<td>229.8</td>
<td>-2.4</td>
<td>231.0</td>
<td>-1.2</td>
</tr>
<tr>
<td>H₃C₆-CH₂-CH₂-NH₂</td>
<td>232.2</td>
<td>230.5</td>
<td>-1.7</td>
<td>233.5</td>
<td>1.3</td>
</tr>
<tr>
<td>CH₃-CH₂-H₂C₆-CH₂-NH₂</td>
<td>233.0</td>
<td>229.6</td>
<td>-3.4</td>
<td>230.6</td>
<td>-2.4</td>
</tr>
<tr>
<td>CH₃-H₂C₆-CH₂-CH₂-NH₂</td>
<td>233.0</td>
<td>231.2</td>
<td>-1.8</td>
<td>235.1</td>
<td>2.1</td>
</tr>
</tbody>
</table>

MUE 2.6 4.8
Figure 5.1. Classification of methods for treating the boundary between QM and MM fragments in a combined QM/MM approach. A. Category one models are depicted, which separate a QM fragment from an MM subsystem across a covalent bond, typically, but not required, between two sp\(^3\) carbons. The valency of the QM fragment is satisfied by placing a link-atom, X, which can be a hydrogen or a pseudo-halogen atom to mimic the electronic properties of the C\(_M\) atom. X can be parameterized to mimic the C\(_b\)-C\(_M\) bond, namely a pseudobond model. Constraints are needed to enforce the link-atom aligned along the C\(_b\)-C\(_M\) bond; thus, it belongs both to the QM and MM region as indicated by inclusion in the two circles. B. In type two models, the QM and MM partition is made over a boundary atom, C\(_b\), typically an sp\(^3\) carbon atom, which is both a “QM” atom and an “MM” atom. Typically hybrid orbitals are used. Unlike category one models, no degrees of freedom are added (Cartesian coordinates of the link-atom) or eliminated (neighboring atomic charges) here.
Figure 5.2. Illustration depicting the difference between the LSCF model and the GHO method, both of which utilize hybrid orbitals. Orbitals colored in red are used as part of the basis functions in the self-consistent field optimization, whereas orbitals in green are auxiliary orbitals that are frozen.
Figure 5.3. Illustration of difference on Mulliken population analysis charges of $C_A$, $C_B$ and the “QM” methyl group between PHO and standard calculations across a wide range of methods and basis sets. BS1-4 denote STO-3G, 6-31G(d), 6-311+G(d,p) and aug-cc-pVTZ, respectively.
Figure 5.4. Illustration of difference on CM5 charges of $C_A$, $C_B$ and the “QM” methyl group between PHO and standard calculations across a wide range of methods and basis sets. BS1-4 denote STO-3G, 6-31G(d), 6-311+G(d,p) and aug-cc-pVTZ, respectively.
Figure 5.5. Torsional barriers about the C2-C3 bond in n-butane. Complete energy minimization was performed at each constrained value of the dihedral angle. PHO calculations are done at HF/3-21G level, with the boundary atom placed at the C2, C3 and C4 carbons, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.
Figure 5.6. Torsional barriers about the C2-C3 bond in n-butane. Complete energy minimization was performed at each constrained value of the dihedral angle. PHO calculations are done at HF/6-31G(d) level, with the boundary atom placed at the C2, C3 and C4 carbons, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.
Chapter 1


(14) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard, W. A.; Skiff, W. J.


---

**Chapter 2**


Chapter 3


Chapter 4

(15) L. Song, J. Han, Y. L. Lin, W. Xie, and J. Gao, J. Phys. Chem. A 113, 11656 (2009).
(37) M. J. Frisch et al., (Gaussian, Inc., Wallingford, CT, 2009).

Chapter 5

(2) Gao, J.; Thompson, M. A. 1998.
(50) Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. *GAUSSIAN, Development Version 2013.*
Appendix A

Supplementary Material for Chapter 2:

X-Pol-X with the neglect of interfragment differential overlap approximation

The neglect of diatomic differential overlap (NDDO)\textsuperscript{1,2} approximation is adopted in most modern semiempirical quantum mechanical methods such as MNDO,\textsuperscript{3} AM1,\textsuperscript{4} PM3,\textsuperscript{5} and RM1.\textsuperscript{6} In this method, electronic integrals involving the charge density $\chi^A \chi^B$ are assumed to be zero when the basis functions $\chi^A_\mu$ and $\chi^B_\nu$ are located on different atoms, $A \neq B$, with the exception that the two-center one-electron nuclear attraction integrals are retained due to their importance for chemical bonding. Here we introduce a less severe approximation called the neglect of interfragment differential overlap (NIDO). In this approximation, we retain all electronic integrals within the same monomer, but electronic integrals involving differential overlap between basis functions belonging to different fragments are neglected. Applying the NIDO approximation greatly simplifies eq 15, as explained in the rest of this appendix. Below, we present the explicit expression for each term in eqs 16-18 by applying the NIDO approximation.

First, the block diagonal terms of the un-projected matrix are reduced to

$$F^{aa}_{\mu\nu} = \left[ H^{aa}_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{aa} \left\{ (\chi^a_\mu \chi^a_\nu \mid \chi^a_\lambda \chi^a_\sigma) - \frac{1}{2} (\chi^a_\mu \chi^a_\nu \mid \chi^a_\lambda \chi^a_\sigma) \right\} \right]$$

(A1)

$$+ \left[ \sum_{b \neq a} \sum_{\delta\sigma} D_{\delta\sigma}^{bb} (\chi^b_\mu \chi^b_\nu \mid \chi^b_\delta \chi^b_\sigma) \right]$$

Note that the “exchange” terms from orbitals on other fragments, $D_{\lambda\sigma}^{ab}(\chi^a_\mu \chi^a_\lambda \mid \chi^b_\nu \chi^b_\sigma)$, are all zero because of the NIDO approximation. Thus, the un-projected Fock matrix element for orbitals on fragment $a$ can be written as:
\[ F^{aa}_{\mu v} = (F^{aa}_{\mu v})^\dagger + <\chi^a_\mu | \sum_{b \neq a} V^a_b | \chi^a_v > \]  

(A2)

Eq A2 includes contributions only from the Coulomb integrals between different fragments, which is similar to a combined quantum mechanical and molecular mechanical (QM/MM) method. In the original X-Pol potential,\textsuperscript{7,8} in which the Hartree-product wave function is used, the rest of the density projection terms in eq 15 are not needed.

The \( V_2 \) term of eq 16 simplifies under the NIDO approximation to

\[
(V_2)^a_{\mu v} = \sum_{b \neq a} \sum_{\beta} \left[ H^{ab}_{\mu \beta} + \sum_{c,d} \sum_{\lambda,\sigma} D^{cd}_{\lambda \sigma} \{ (\chi^a_\mu \chi^b_\beta | \chi^c_\lambda \chi^d_\sigma) - \frac{1}{2} (\chi^a_\mu \chi^c_\lambda | \chi^b_\beta \chi^d_\sigma) \} \right] P^{ba}_{\beta \nu} \\
= -\frac{1}{2} \sum_{b \neq a} \sum_{\beta} \left[ \sum_{\lambda} \sum_{\sigma} D^{ab}_{\lambda \sigma} (\chi^a_\mu \chi^a_\beta | \chi^b_\lambda \chi^b_\sigma) \right] P^{ba}_{\beta \nu} 
\]

(A3)
in which, only the exchange terms having \( a = c \) and \( b = d \) remain to be non-zero.

The \( V_3 \) term becomes

\[
(V_3)^a_{\mu v} = \sum_{b \neq a} \sum_{\beta} \left[ (P^{ba}_{\beta \nu})^\dagger \left[ H^{ba}_{\beta \nu} + \sum_{c,d} \sum_{\lambda,\sigma} D^{cd}_{\lambda \sigma} \{ (\chi^b_\beta \chi^b_\lambda | \chi^a_\nu \chi^d_\sigma) - \frac{1}{2} (\chi^b_\beta \chi^d_\lambda | \chi^b_\nu \chi^a_\sigma) \} \right] \right] \\
= \sum_{b \neq a} \left[ -\frac{1}{2} \sum_{\beta} \left( P^{ba}_{\beta \nu} \right)^\dagger \sum_{\lambda} \sum_{\sigma} D^{ba}_{\lambda \sigma} (\chi^b_\beta \chi^b_\lambda | \chi^a_\nu \chi^a_\sigma) \right] 
\]

(A4)

Similarly, the non-zero terms are in the exchange part for orbitals that are on fragments such that \( b = c \) and \( a = d \) in the double summation over fragments in parentheses.

Finally, the last term of eq 15 becomes
We presented the NIDO approximation here for a general case of X-Pol-X, that is the monomers can be treated by ab initio Hartree-Fock, by post-Hartree-Fock correlation methods, or by density functional theory. However, if desired, the monomers could also be treated by semiempirical molecular theory employing the NDDO approximation within the monomers. This procedure has been implemented into the program CHARMM for defining diabatic states in the effective Hamiltonian mixed molecular orbital and valence bond (EH-MOVB) method.\(^9\)

References:

Table A1. Computed total energies and energy components (in hartrees) for the cyclic water trimer minimum structure and for a symmetric trimer geometry. All calculations were performed using HF/6-31+G(d).

<table>
<thead>
<tr>
<th>Method</th>
<th>c-W₃</th>
<th>s-W₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(ψₐ⁰) + E(ψₐ⁰) + E(ψₖ)</td>
<td>-228.05297</td>
<td>-228.05318</td>
</tr>
<tr>
<td>E_Coulomb(ψₐ⁰,ψₐ⁰,ψₖ)</td>
<td>-228.09376</td>
<td>-228.07863</td>
</tr>
<tr>
<td>E_BLW(A ψₐ⁰ψₐ⁰ψₖ)</td>
<td>-228.06774</td>
<td>-228.06150</td>
</tr>
<tr>
<td>E(ψₐ) + E(ψₖ) + E(ψₖ)</td>
<td>-228.04723</td>
<td>-228.05067</td>
</tr>
<tr>
<td>E_Coulomb(ψₐψₖψₖ)</td>
<td>-228.09790</td>
<td>-228.07990</td>
</tr>
<tr>
<td>E_BLW(A ψₐψₖψₖ)</td>
<td>-228.07286</td>
<td>-228.06382</td>
</tr>
<tr>
<td>E_HF(ψₐψₖψₖ)</td>
<td>-228.07775</td>
<td>-228.06724</td>
</tr>
</tbody>
</table>
Appendix B

Supplementary Material for Chapter 5:

Table B1. Complete set of bond length (Å) in ethane from standard calculation and PHO (w/o tetrahedral restraining potential $E_B^{tet}$) for different methods and basis sets

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard</th>
<th>PHO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A$-$C_B$</td>
<td>$C_A$-$H$</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>1.538</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>1.527</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/6-31G+G(d,p)</td>
<td>1.527</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>1.525</td>
<td>1.084</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>1.552</td>
<td>1.101</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>1.531</td>
<td>1.096</td>
</tr>
<tr>
<td>B3LYP/6-31G+G(d,p)</td>
<td>1.531</td>
<td>1.094</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>1.527</td>
<td>1.091</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>1.542</td>
<td>1.095</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>1.526</td>
<td>1.094</td>
</tr>
<tr>
<td>M062X/6-31G+G(d,p)</td>
<td>1.527</td>
<td>1.092</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>1.524</td>
<td>1.090</td>
</tr>
<tr>
<td>Method</td>
<td>H-C&lt;sub&gt;A&lt;/sub&gt;-H</td>
<td>H-C&lt;sub&gt;A&lt;/sub&gt;-C&lt;sub&gt;B&lt;/sub&gt;</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Standard Method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>107.6</td>
<td>111.2</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>107.6</td>
<td>111.2</td>
</tr>
</tbody>
</table>
Table B3. Complete set of bond length (Å) in ethane from standard calculation and PHO (with tetrahedral restraining potential $E_B^{\text{tet}}$) for different methods and basis sets

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>PHO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A-C_B$</td>
<td>$C_A-H$</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>1.538</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>1.527</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>1.527</td>
<td>1.086</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>1.525</td>
<td>1.084</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>1.552</td>
<td>1.101</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>1.531</td>
<td>1.096</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>1.531</td>
<td>1.094</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>1.527</td>
<td>1.091</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>1.542</td>
<td>1.095</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>1.526</td>
<td>1.094</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>1.527</td>
<td>1.092</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>1.524</td>
<td>1.090</td>
</tr>
</tbody>
</table>
Table B4. Complete set of bond angles (Degree) in ethane from standard calculation and PHO (with tetrahedral restraining potential $E_{B}^{tet}$) for different methods and basis sets

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard Method</th>
<th>PHO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H-C&lt;sub&gt;A&lt;/sub&gt;-H</td>
<td>H-C&lt;sub&gt;A&lt;/sub&gt;-C&lt;sub&gt;B&lt;/sub&gt;</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>107.5</td>
<td>111.4</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>108.2</td>
<td>110.7</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>107.6</td>
<td>111.2</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>107.7</td>
<td>111.2</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>107.6</td>
<td>111.2</td>
</tr>
</tbody>
</table>
Table B5. Comparison of Mulliken population charges in ethane between standard calculation and PHO (w/o tetrahedral restraining potential $E_B^{tet}$) for different methods and basis sets

<table>
<thead>
<tr>
<th>Method</th>
<th>C_A</th>
<th>H_qm</th>
<th>C_B</th>
<th>C_A</th>
<th>H_qm</th>
<th>C_B</th>
<th>(C_AH_3)q</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/STO-3G</td>
<td>-0.175</td>
<td>0.058</td>
<td>-0.175</td>
<td>-0.173</td>
<td>0.037</td>
<td>-0.209</td>
<td>-0.061</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>-0.476</td>
<td>0.159</td>
<td>-0.476</td>
<td>-0.463</td>
<td>0.122</td>
<td>-0.173</td>
<td>-0.097</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>-0.329</td>
<td>0.110</td>
<td>-0.329</td>
<td>-0.072</td>
<td>0.016</td>
<td>-0.245</td>
<td>-0.025</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>-0.856</td>
<td>0.285</td>
<td>-0.856</td>
<td>-0.624</td>
<td>0.176</td>
<td>-0.174</td>
<td>-0.096</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>-0.216</td>
<td>0.072</td>
<td>-0.216</td>
<td>-0.222</td>
<td>0.056</td>
<td>-0.215</td>
<td>-0.055</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>-0.433</td>
<td>0.144</td>
<td>-0.433</td>
<td>-0.420</td>
<td>0.121</td>
<td>-0.212</td>
<td>-0.058</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>-0.389</td>
<td>0.130</td>
<td>-0.389</td>
<td>-0.146</td>
<td>0.054</td>
<td>-0.285</td>
<td>0.015</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>-0.600</td>
<td>0.200</td>
<td>-0.600</td>
<td>-0.363</td>
<td>0.116</td>
<td>-0.256</td>
<td>-0.015</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>-0.209</td>
<td>0.070</td>
<td>-0.209</td>
<td>-0.216</td>
<td>0.052</td>
<td>-0.209</td>
<td>-0.061</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>-0.479</td>
<td>0.160</td>
<td>-0.479</td>
<td>-0.473</td>
<td>0.130</td>
<td>-0.188</td>
<td>-0.082</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>-0.470</td>
<td>0.157</td>
<td>-0.470</td>
<td>-0.228</td>
<td>0.071</td>
<td>-0.254</td>
<td>-0.016</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>-0.735</td>
<td>0.245</td>
<td>-0.735</td>
<td>-0.603</td>
<td>0.184</td>
<td>-0.219</td>
<td>-0.051</td>
</tr>
</tbody>
</table>
Table B6. Comparison of Mulliken population charges in ethane between standard calculation and PHO (with tetrahedral restraining potential $E_{\text{tet}}^B$) for different methods and basis sets.

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th></th>
<th>PHO</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A$</td>
<td>$H_{\text{qm}}$</td>
<td>$C_B$</td>
<td>$C_A$</td>
<td>$H_{\text{qm}}$</td>
<td>$C_B$</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>-0.175</td>
<td>0.058</td>
<td>-0.175</td>
<td>-0.194</td>
<td>0.035</td>
<td>-0.182</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>-0.476</td>
<td>0.159</td>
<td>-0.476</td>
<td>-0.475</td>
<td>0.123</td>
<td>-0.163</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>-0.329</td>
<td>0.110</td>
<td>-0.329</td>
<td>-0.120</td>
<td>0.020</td>
<td>-0.209</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>-0.856</td>
<td>0.285</td>
<td>-0.856</td>
<td>-0.676</td>
<td>0.189</td>
<td>-0.161</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>-0.216</td>
<td>0.072</td>
<td>-0.216</td>
<td>-0.237</td>
<td>0.054</td>
<td>-0.196</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>-0.433</td>
<td>0.144</td>
<td>-0.433</td>
<td>-0.424</td>
<td>0.121</td>
<td>-0.211</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>-0.389</td>
<td>0.130</td>
<td>-0.389</td>
<td>-0.187</td>
<td>0.058</td>
<td>-0.257</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>-0.600</td>
<td>0.200</td>
<td>-0.600</td>
<td>-0.404</td>
<td>0.127</td>
<td>-0.246</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>-0.209</td>
<td>0.070</td>
<td>-0.209</td>
<td>-0.231</td>
<td>0.050</td>
<td>-0.189</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>-0.479</td>
<td>0.160</td>
<td>-0.479</td>
<td>-0.477</td>
<td>0.131</td>
<td>-0.186</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>-0.470</td>
<td>0.157</td>
<td>-0.470</td>
<td>-0.268</td>
<td>0.075</td>
<td>-0.227</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>-0.735</td>
<td>0.245</td>
<td>-0.735</td>
<td>-0.650</td>
<td>0.197</td>
<td>-0.210</td>
</tr>
</tbody>
</table>
Table B7. Comparison of CM5 charges in ethane between standard calculation and PHO (w/o tetrahedral restraining potential $E_B^{tet}$) for different methods and basis sets

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>PHO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A$</td>
<td>$H_{qm}$</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>-0.251</td>
<td>0.084</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>-0.244</td>
<td>0.081</td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>-0.244</td>
<td>0.081</td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>-0.245</td>
<td>0.082</td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>-0.272</td>
<td>0.091</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>-0.265</td>
<td>0.088</td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>-0.261</td>
<td>0.087</td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>-0.261</td>
<td>0.087</td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>-0.269</td>
<td>0.090</td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>-0.265</td>
<td>0.088</td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>-0.264</td>
<td>0.088</td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>-0.265</td>
<td>0.088</td>
</tr>
</tbody>
</table>
Table B8. Comparison of CM5 charges in ethane between standard calculation and PHO (with tetrahedral restraining potential $E_B^{tet}$) for different methods and basis sets

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard C_A</th>
<th>H_qm</th>
<th>C_B</th>
<th>PHO C_A</th>
<th>H_qm</th>
<th>C_B</th>
<th>(C_AH_3)q</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF/STO-3G</td>
<td>-0.251</td>
<td>0.084</td>
<td>-0.251</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>-0.244</td>
<td>0.081</td>
<td>-0.244</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF/6-311+G(d,p)</td>
<td>-0.244</td>
<td>0.081</td>
<td>-0.244</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF/Aug-cc-pvtz</td>
<td>-0.245</td>
<td>0.082</td>
<td>-0.245</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3LYP/STO-3G</td>
<td>-0.272</td>
<td>0.091</td>
<td>-0.272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>-0.265</td>
<td>0.088</td>
<td>-0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3LYP/6-311+G(d,p)</td>
<td>-0.261</td>
<td>0.087</td>
<td>-0.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3LYP/Aug-cc-pvtz</td>
<td>-0.261</td>
<td>0.087</td>
<td>-0.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M062X/STO-3G</td>
<td>-0.269</td>
<td>0.090</td>
<td>-0.269</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M062X/6-31G(d)</td>
<td>-0.265</td>
<td>0.088</td>
<td>-0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M062X/6-311+G(d,p)</td>
<td>-0.264</td>
<td>0.088</td>
<td>-0.264</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M062X/Aug-cc-pvtz</td>
<td>-0.265</td>
<td>0.088</td>
<td>-0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B9. Optimized Cartesian Coordinates and Total Energy Using PHO@HF/6-311+G(d,p) for all the Molecules in the Test Set

<table>
<thead>
<tr>
<th>Compound</th>
<th>PHO @ HF/6-311+G(d,p) Total Energy</th>
<th>Coordinates (A.U.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₂H₅C₆H₃</td>
<td>-115.055784692 A.U.</td>
<td>C-CT(Fragment=1) 2.012761 0.135269 -0.000032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 2.785180 -0.629825 -0.000120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 2.173417 0.759536 0.876669</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CT(Fragment=1) 0.622519 -0.502498 -0.000027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 0.570313 -1.156319 0.862917</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 0.570247 -1.156186 0.863068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CG--0.18(Fragment=1) -0.651161 0.539516 0.000107 GHO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -0.662012 1.213315 0.891136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CT--0.27(Fragment=2) -1.988410 -0.169606 -0.000013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -2.068067 -0.790349 0.883158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -2.786878 0.561088 -0.000045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -2.067941 -0.790287 -0.883240</td>
</tr>
<tr>
<td>C₂H₅C₆H₃</td>
<td>-115.060969233 A.U.</td>
<td>C-CT(Fragment=1) 1.319919 -0.284703 0.000016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 2.210963 0.338126 0.000035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 1.371130 -0.927047 0.876620</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CT(Fragment=1) 0.059176 0.580339 -0.000004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 0.116658 1.233446 -0.863401</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC(Fragment=1) 0.116625 1.233440 0.863399</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CG--0.27(Fragment=1) -1.370212 -0.231718 -0.000037 GHO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -1.493198 -0.892259 -0.890636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -1.493217 -0.892142 0.890646</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -2.253362 0.449985 0.000080</td>
</tr>
<tr>
<td>C₆H₅C₆H₃</td>
<td>-266.568260381 A.U.</td>
<td>C-CA(Fragment=1) -0.169828 1.191224 0.000325</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CA(Fragment=1) 1.221279 1.195816 -0.000112</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CA(Fragment=1) 1.927937 -0.000070 -0.000334</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CA(Fragment=1) 1.221306 -1.195821 0.000013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CA(Fragment=1) -0.170029 -1.191145 0.000216</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CA(Fragment=1) -0.899595 0.000012 0.000015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HA(Fragment=1) -0.684968 2.136004 0.001376</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HA(Fragment=1) 1.747401 2.134353 0.000145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HA(Fragment=1) 3.003275 -0.000122 -0.000614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HA(Fragment=1) 1.747325 -2.134444 0.000243</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HA(Fragment=1) -0.685184 -2.135998 0.001234</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-CG--0.27(Fragment=1) -2.512711 0.000013 -0.000316 GHO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-HC-0.09(Fragment=2) -2.945920 -0.894947 -0.508097</td>
</tr>
</tbody>
</table>
H-HC-0.09 (Fragment=2)  -2.946138  0.007915  1.028628
H-HC-0.09 (Fragment=2)  -2.945950  0.887068 -0.521760

H₂NCH₂-C₃H₃ PHO @ HF/6-311+G(d,p) Total Energy -131.060133412 A.U.
N-N(3) (Fragment=1)  1.326036 -0.215291  0.000022
H-H(1) (Fragment=1)  1.398264 -0.800941  0.807420
H-H(1) (Fragment=1)  1.398183 -0.801343 -0.807092
C-CT(1) (Fragment=1)  0.103555  0.570802 -0.000105
H-HC(1) (Fragment=1)  0.169576  1.219086  0.862553
H-HC(1) (Fragment=1)  0.169555  1.218789 -0.862987
C-CG--0.27 (Fragment=1)  -1.323127 -0.235104  0.000049 GHO
H-HC-0.09 (Fragment=2)  -1.447594 -0.895236 -0.890576
H-HC-0.09 (Fragment=2)  -2.205231  0.448049 -0.000041
H-HC-0.09 (Fragment=2)  -1.447574 -0.895553 -0.890246

H₂NCH₂CH₂-C₃H₃ PHO @ HF/6-311+G(d,p) Total Energy -170.103564555 A.U.
N-N(3) (Fragment=1)  -1.981227 -0.033770  0.000047
H-H(1) (Fragment=1)  -2.140331 -0.605789 -0.805209
H-H(1) (Fragment=1)  -2.140576 -0.604573  0.806133
C-C(1) (Fragment=1)  -0.635873  0.520217 -0.000145
H-H(1) (Fragment=1)  -0.549827  1.167850 -0.869319
H-H(1) (Fragment=1)  -0.549654  1.168285  0.806133
C-CT(1) (Fragment=1)  0.496036 -0.512555 -0.000081
H-HC(1) (Fragment=1)  0.345921 -1.151797 -0.864717
H-HC(1) (Fragment=1)  0.345804 -1.151973 -0.864423
C-CG--0.27 (Fragment=1)  2.021159  0.098688  0.000084 GHO
H-HC-0.09 (Fragment=2)  2.803797 -0.696715 -0.000198
H-HC-0.09 (Fragment=2)  2.232705  0.736917 -0.890474
H-HC-0.09 (Fragment=2)  2.232820  0.736087  0.891211

HCO-C₃H₃ PHO @ HF/6-311+G(d,p) Total Energy -149.740970777 A.U.
C-CT(1) (Fragment=1)  0.286438  0.392940 -0.003991
O-OH(1) (Fragment=1)  1.243518 -0.293508  0.001023
H-HC(1) (Fragment=1)  0.419499  1.478576  0.008300
C-CG--0.27 (Fragment=1)  -1.206115 -0.117845  0.000417 GHO
H-HC-0.09 (Fragment=2)  -1.309820 -1.150080 -0.414961
H-HC-0.09 (Fragment=2)  -1.649689 -0.155926  1.024940
H-HC-0.09 (Fragment=2)  -1.890073  0.524925 -0.605016

HCOCH₂-C₃H₃ PHO @ HF/6-311+G(d,p) Total Energy -188.771268117 A.U.
C-CT(1) (Fragment=1)  -0.826776 -0.271789  0.255115
O-OH(1) (Fragment=1)  -1.855029 -0.031779 -0.280951
H-HC(1) (Fragment=1)  -0.748751 -1.158316  0.902478
C-CT(1) (Fragment=1)  0.417697  0.577027  0.158342
H-HC(1) (Fragment=1)  0.224422  1.344514 -0.576529
H-HC(1) (Fragment=1)  0.483710  1.073306  1.121927
C-CG--0.27(Fragment=1)          1.804041 -0.243577 -0.158710 GHO
H-HC-0.09(Fragment=2)          2.699121  0.422842 -0.193281
H-HC-0.09(Fragment=2)          2.033043 -1.024366  0.605616
H-HC-0.09(Fragment=2)          1.778918  0.773719 -1.141081

HOCH$_2$-C$_b$H$_3$ PHO @ HF/6-311+G(d,p) Total Energy -150.899041492 A.U.
O-OH(Fragment=1)          1.255119 -0.292531  0.116528
H-HO(Fragment=1)          1.361165  0.839430  0.640851
C-CT(Fragment=1)          0.142761  0.550550  0.041155
H-HC(Fragment=1)          0.257373  1.100572  0.968193
C-CG--0.27(Fragment=1)          1.289530  0.208908  0.013014 GHO
H-HC-0.09(Fragment=2)          2.781298  0.674598  0.033912
H-HC-0.09(Fragment=2)          2.185430  0.491345  0.068832
H-HC-0.09(Fragment=2)          1.442649  0.774712 -0.962942

HOCH$_2$CH$_2$-C$_b$H$_3$ PHO @ HF/6-311+G(d,p) Total Energy -189.939752977 A.U.
O-OH(Fragment=1)          1.924717  0.092476  0.081071
H-HO(Fragment=1)          2.054799  0.693689  0.630596
C-CT(Fragment=1)          0.658694  0.506317  0.021748
H-HC(Fragment=1)          0.596427  1.091281  0.939921
H-HC(Fragment=1)          0.598842  1.204077  0.805996
C-CT(Fragment=1)          0.473180  0.514464 -0.045509
H-HC(Fragment=1)          0.321624  1.215770  0.770560
H-HC(Fragment=1)          0.328500  1.083236  0.956930
H-HC(Fragment=1)          0.328500  1.083236  0.956930
C-CG--0.27(Fragment=1)          1.990000  0.110565  0.018038 GHO
H-HC-0.09(Fragment=2)          2.781298  0.674598 -0.033912
H-HC-0.09(Fragment=2)          2.185430  0.491345  0.068832
H-HC-0.09(Fragment=2)          1.442649 -0.774712 -0.962942

HOOC-C$_b$H$_3$ PHO @ HF/6-311+G(d,p) Total Energy -224.663987976 A.U.
C-C(Fragment=1)          0.113387  0.118593  0.000022
O-O(Fragment=1)          0.651155  1.169328  0.000014
O-OH(Fragment=1)          0.824796  1.005053  0.000023
H-HO(Fragment=1)          1.741389 -0.772829 -0.000083
C-CG--0.27(Fragment=1)          1.437624 -0.118084  0.000100 GHO
H-HC-0.09(Fragment=2)          1.792896 -0.689098  0.892905
H-HC-0.09(Fragment=2)          1.792349 -0.692610 -0.890668
H-HC-0.09(Fragment=2)          2.018332  0.837287 -0.001880

HOOCCH$_2$-C$_b$H$_3$ PHO @ HF/6-311+G(d,p) Total Energy -263.693749301 A.U.
C-C(Fragment=1)          0.599888  0.112929  0.000004
O-O(Fragment=1)          0.691523  1.292646  0.000014
O-OH(Fragment=1)          1.682057 -0.668184 -0.000008
H-HO(Fragment=1)          2.449823 -0.116206  0.000000
C-CT(Fragment=1)         -0.666434 -0.714716 -0.000020
H-HC(Fragment=1)  -0.577325 -1.367080 -0.859847
H-HC(Fragment=1)  -0.577315 -1.367138  0.859761
C-CG-0.27(Fragment=1) -2.083774  0.091738  0.000015 GHO
H-HC-0.09(Fragment=2) -2.207524  0.753164 -0.891230
H-HC-0.09(Fragment=2) -2.966884 -0.591225  0.000010
H-HC-0.09(Fragment=2) -2.207500  0.753089  0.891318

(CH$_3$)$_2$CHC$(CH_3)_3$    PHO @ HF/6-311+G(d,p) Total Energy -154.155699084 A.U.
C-CT(Fragment=1)        -0.942532  0.000002 -0.453159
H-HC(Fragment=1)        -1.013245  0.000008 -1.536436
C-CT(Fragment=1)        -1.700900 -1.253080  0.014256
H-HC(Fragment=1)        -2.728431 -1.250129 -0.342590
H-HC(Fragment=1)        -1.236874 -2.163758 -0.355879
H-HC(Fragment=1)        -1.733592 -1.323819  1.100614
C-CT(Fragment=1)        -1.700905  1.253075  0.014269
H-HC(Fragment=1)        -1.733598  1.323803  1.100628
H-HC(Fragment=1)        -1.236883  2.163759 -0.355857
H-HC(Fragment=1)        -2.728437  1.250124 -0.342576
C-CG(Fragment=1)        0.660791  0.000003 -0.032554 GHO
C-CT-0.27(Fragment=2)   1.406683  1.195761 -0.551228
H-HC-0.09(Fragment=2)   0.967436  2.097288 -0.145711
H-HC-0.09(Fragment=2)   1.348780  1.215044 -1.631196
H-HC-0.09(Fragment=2)   2.441534  1.128478 -0.245339
C-CT-0.27(Fragment=2)   0.883546  0.000017  1.452549
H-HC-0.09(Fragment=2)   0.434526 -0.884705  1.883723
H-HC-0.09(Fragment=2)   1.946219 -0.000020  1.651860
H-HC-0.09(Fragment=2)   0.434526  0.884658  1.883747
C-CT-0.27(Fragment=2)   1.406693 -1.195742 -0.551245
H-HC-0.09(Fragment=2)   2.441545 -1.128452 -0.245360
H-HC-0.09(Fragment=2)   0.967456 -2.097278 -0.145738
H-HC-0.09(Fragment=2)   1.348786 -1.215012 -1.631214

(CH$_3$)$_2$CHCbH$_2$CH$_2$CH$_3$    PHO @ HF/6-311+G(d,p) Total Energy -154.109780467 A.U.
C-CT(Fragment=1)        -1.104511 -0.028377 -0.325708
H-HC(Fragment=1)        -1.127097  0.043673 -1.408110
C-CT(Fragment=1)        -2.235546  0.998146  0.054696
H-HC(Fragment=1)        -3.194429 -0.653251 -0.324733
H-HC(Fragment=1)        -2.067564 -1.991180 -0.354632
H-HC(Fragment=1)        -2.330951 -1.104610  1.134411
C-CT(Fragment=1)        -1.441229  1.377157  0.196503
H-HC(Fragment=1)        -1.485418  1.404550  1.284251
H-HC(Fragment=1)        -0.700288  2.108251 -0.116227
H-HC(Fragment=1)        -2.404338  1.716493 -0.177641
C-CG-0.18(Fragment=1)   0.396468 -0.556674  0.127751 GHO
H-HC-0.09(Fragment=2)   0.654718 -1.552942 -0.305380
<table>
<thead>
<tr>
<th>Bond</th>
<th>Distance 1</th>
<th>Distance 2</th>
<th>Distance 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>0.502690</td>
<td>-0.669517</td>
<td>1.233597</td>
</tr>
<tr>
<td>C- CT-0.18(Fragment=2)</td>
<td>1.506470</td>
<td>0.382850</td>
<td>-0.301920</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>1.358118</td>
<td>0.480443</td>
<td>-1.382300</td>
</tr>
<tr>
<td>C- CT-0.27(Fragment=2)</td>
<td>2.846520</td>
<td>-0.198789</td>
<td>0.159052</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>3.648645</td>
<td>0.462097</td>
<td>-0.140919</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.842072</td>
<td>-0.297983</td>
<td>1.236000</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.989318</td>
<td>-1.170819</td>
<td>-0.293075</td>
</tr>
</tbody>
</table>

(CH$_3$)$_2$CHCbH$_2$CH$_3$ PHO @ HF/6-311+G(d,p) Total Energy -154.099201088 A.U.

<table>
<thead>
<tr>
<th>Bond</th>
<th>Distance 1</th>
<th>Distance 2</th>
<th>Distance 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-0.514820</td>
<td>-0.019874</td>
<td>-0.320129</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.453175</td>
<td>0.011856</td>
<td>-1.402898</td>
</tr>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-1.791616</td>
<td>-0.812870</td>
<td>0.005123</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-2.665880</td>
<td>-0.357222</td>
<td>-0.453609</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-1.733932</td>
<td>-1.834886</td>
<td>-0.361139</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-1.974700</td>
<td>-0.862932</td>
<td>1.077645</td>
</tr>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-0.690735</td>
<td>1.437694</td>
<td>0.133549</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.805301</td>
<td>1.514022</td>
<td>1.213872</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>0.163451</td>
<td>2.048677</td>
<td>-0.145960</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-1.570352</td>
<td>1.888614</td>
<td>-0.320075</td>
</tr>
<tr>
<td>C- CG-0.18(Fragment=1)</td>
<td>0.864428</td>
<td>-0.726604</td>
<td>0.257490</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>1.013723</td>
<td>-1.764677</td>
<td>-0.127314</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>0.877147</td>
<td>-0.810468</td>
<td>-1.371439</td>
</tr>
<tr>
<td>C- CT-0.27(Fragment=2)</td>
<td>2.116405</td>
<td>0.038602</td>
<td>0.098702</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.065280</td>
<td>1.038662</td>
<td>0.295563</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.984615</td>
<td>-0.469969</td>
<td>0.283298</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.197179</td>
<td>0.096633</td>
<td>-1.193458</td>
</tr>
</tbody>
</table>

(CH$_3$)$_2$CHCbH$_2$CH$_3$ PHO @ HF/6-311+G(d,p) Total Energy -154.104426442 A.U.

<table>
<thead>
<tr>
<th>Bond</th>
<th>Distance 1</th>
<th>Distance 2</th>
<th>Distance 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-0.518598</td>
<td>0.000000</td>
<td>-0.358356</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.688199</td>
<td>0.000000</td>
<td>-1.442361</td>
</tr>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-0.779390</td>
<td>-1.254747</td>
<td>0.098702</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-1.801017</td>
<td>-1.252818</td>
<td>-0.273729</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.307143</td>
<td>-2.164145</td>
<td>-0.263948</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.827265</td>
<td>-1.324822</td>
<td>1.184358</td>
</tr>
<tr>
<td>C- CT(Fragment=1)</td>
<td>-0.779394</td>
<td>1.254745</td>
<td>0.098702</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.827269</td>
<td>1.324820</td>
<td>1.184358</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-0.307148</td>
<td>2.164145</td>
<td>-0.263947</td>
</tr>
<tr>
<td>H- HC(Fragment=1)</td>
<td>-1.801020</td>
<td>1.252815</td>
<td>-0.273730</td>
</tr>
<tr>
<td>C- CG-0.27(Fragment=1)</td>
<td>1.572676</td>
<td>0.000002</td>
<td>0.085309</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.126857</td>
<td>-0.890593</td>
<td>-0.294530</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>2.126863</td>
<td>0.890594</td>
<td>-0.294528</td>
</tr>
<tr>
<td>H- HC-0.09(Fragment=2)</td>
<td>1.714198</td>
<td>0.000000</td>
<td>1.191919</td>
</tr>
</tbody>
</table>

(CH$_3$)$_2$CHCbH(CH$_3$)$_2$ PHO @ HF/6-311+G(d,p) Total Energy -154.116280798 A.U.
C-CT(Fragment=1) -0.775331 0.263025 -0.418055
H-HC(Fragment=1) -0.929181 0.812007 -1.342172
C-CT(Fragment=1) -1.855749 -0.830843 -0.396449
H-HC(Fragment=1) -2.852313 -0.403040 -0.477799
H-HC(Fragment=1) -1.739021 -1.528117 -1.222481
H-HC(Fragment=1) -1.824066 -1.411092 0.524606
C-CT(Fragment=1) -1.046078 1.265844 0.714163
H-HC(Fragment=1) -0.964572 0.798938 1.694535
H-HC(Fragment=1) -0.344163 2.095296 0.690285
H-HC(Fragment=1) -2.045127 1.689081 0.636924
C-CG--0.09(Fragment=1) 0.769111 -0.333536 -0.416057
H-HC-0.09(Fragment=2) 0.975172 -1.014369 -0.879432
C-CT--0.27(Fragment=2) 1.816964 0.748206 -0.486308
H-HC-0.09(Fragment=2) 1.717198 1.399240 0.372113
H-HC-0.09(Fragment=2) 1.686499 1.322643 -1.394140
H-HC-0.09(Fragment=2) 2.799700 0.295968 -0.480904
C-CT-0.27(Fragment=2) 1.092255 -1.129654 0.822817
H-HC-0.09(Fragment=2) 0.417741 -1.972489 0.900103
H-HC-0.09(Fragment=2) 2.112164 -1.485505 0.764088
H-HC-0.09(Fragment=2) 0.982928 -0.496812 1.693611

(CH₃)₃CCbH₂CH₂CH³ PHO @ HF/6-311+G(d,p) Total Energy -193.150077556 A.U.
C-CT(Fragment=1) 0.950735 0.050443 -0.000040
C-CT(Fragment=1) 1.211445 -0.823857 1.249250
H-HC(Fragment=1) 2.225602 -1.217623 1.257689
H-HC(Fragment=1) 1.077664 -0.260344 2.169752
H-HC(Fragment=1) 0.536019 -1.675079 1.291167
C-CT(Fragment=1) 1.211636 -0.823870 -1.249281
H-HC(Fragment=1) 0.536219 -1.675094 -1.291292
H-HC(Fragment=1) 1.077995 -0.260368 -2.169811
H-HC(Fragment=1) 2.225795 -1.217634 -1.257561
C-CT(Fragment=1) 1.994591 1.193397 0.000034
H-HC(Fragment=1) 1.893276 1.830737 -0.875563
H-HC(Fragment=1) 1.893144 1.830744 0.875610
H-HC(Fragment=1) 3.011435 0.807040 0.000112
C-CG--0.18(Fragment=1) -0.604395 0.653652 -0.000164
H-HC-0.09(Fragment=2) -0.822822 1.290822 0.890391
H-HC-0.09(Fragment=2) -0.822780 1.291000 -0.890601
C-CT--0.18(Fragment=2) -1.659657 -0.435393 0.000027
H-HC-0.09(Fragment=2) -1.550807 -1.050513 0.886923
H-HC-0.09(Fragment=2) -1.550979 -1.050658 -0.886790
C-CT-0.27(Fragment=2) -3.042807 0.222895 0.000107
H-HC-0.09(Fragment=2) -3.146278 0.838131 -0.883307
H-HC-0.09(Fragment=2) -3.806662 -0.543040 0.000244
H-HC-0.09(Fragment=2) -3.146106 0.838275 0.883441
(CH$_3$)$_3$CCbH$_3$CH$_3$  PHO @ HF/6-311+G(d,p) Total Energy -193.139562467 A.U.

| C-CT(Fragment=1) | -0.445750 | 0.030481 | 0.000001 |
| C-CT(Fragment=1) | -0.567265 | -0.873705 | -1.249207 |
| H-HC(Fragment=1) | -1.507350 | -1.421362 | -1.257295 |
| H-HC(Fragment=1) | -0.524030 | -0.296138 | -2.169770 |
| H-HC(Fragment=1) | 0.233282 | -1.608326 | -1.291159 |
| C-CT(Fragment=1) | -0.567263 | -0.873603 | 1.249283 |
| H-HC(Fragment=1) | 0.233284 | -1.608221 | 1.291294 |
| H-HC(Fragment=1) | -0.524027 | -0.295961 | 2.169799 |
| H-HC(Fragment=1) | -1.507348 | -1.421260 | 1.257418 |
| C-CT(Fragment=1) | -1.654695 | 0.997289 | -0.000038 |
| H-HC(Fragment=1) | -1.653909 | 1.642666 | 0.875589 |
| H-HC(Fragment=1) | -1.653907 | 1.642596 | -0.875716 |
| H-HC(Fragment=1) | -2.599037 | 0.457413 | -0.000017 |
| C-CG--0.18(Fragment=1) | 0.996014 | 0.867944 | -0.000033 |
| H-HC-0.09(Fragment=2) | 1.112458 | 1.531666 | -0.891007 |
| H-HC-0.09(Fragment=2) | 1.112462 | 1.531749 | 0.890878 |
| C-CT-0.27(Fragment=2) | 2.204910 | -0.043522 | -0.000008 |
| H-HC-0.09(Fragment=2) | 2.187118 | -0.668534 | -0.883637 |
| H-HC-0.09(Fragment=2) | 3.108164 | 0.552871 | -0.000038 |
| H-HC-0.09(Fragment=2) | 2.187132 | -0.668467 | 0.883668 |
Table B10. Potential energy for internal rotation around the C2-C3 bond in n-butane, using PHO, pure QM (HF/3-21G), and MM (OPLS). PHO calculations are done at HF/3-21G level, with the boundary atom placed at the C2, C3 and C4 carbons, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.

<table>
<thead>
<tr>
<th>Dihedral Angle</th>
<th>PHO @C2_B</th>
<th>PHO @C3_B</th>
<th>PHO @C4_B</th>
<th>QM HF/3-21G</th>
<th>OPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>4.38</td>
<td>4.76</td>
<td>6.86</td>
<td>6.04</td>
<td>6.13</td>
</tr>
<tr>
<td>10.0</td>
<td>4.17</td>
<td>4.51</td>
<td>6.54</td>
<td>5.71</td>
<td>5.80</td>
</tr>
<tr>
<td>20.0</td>
<td>3.63</td>
<td>3.85</td>
<td>5.66</td>
<td>4.82</td>
<td>4.92</td>
</tr>
<tr>
<td>30.0</td>
<td>2.92</td>
<td>2.99</td>
<td>4.42</td>
<td>3.62</td>
<td>3.77</td>
</tr>
<tr>
<td>40.0</td>
<td>2.20</td>
<td>2.16</td>
<td>3.10</td>
<td>2.40</td>
<td>2.65</td>
</tr>
<tr>
<td>50.0</td>
<td>1.64</td>
<td>1.54</td>
<td>2.00</td>
<td>1.44</td>
<td>1.75</td>
</tr>
<tr>
<td>60.0</td>
<td>1.34</td>
<td>1.22</td>
<td>1.34</td>
<td>0.88</td>
<td>1.20</td>
</tr>
<tr>
<td>70.0</td>
<td>1.33</td>
<td>1.24</td>
<td>1.28</td>
<td>0.78</td>
<td>1.06</td>
</tr>
<tr>
<td>80.0</td>
<td>1.59</td>
<td>1.58</td>
<td>1.79</td>
<td>1.12</td>
<td>1.37</td>
</tr>
<tr>
<td>90.0</td>
<td>2.01</td>
<td>2.14</td>
<td>2.69</td>
<td>1.82</td>
<td>2.04</td>
</tr>
<tr>
<td>100.0</td>
<td>2.47</td>
<td>2.75</td>
<td>3.65</td>
<td>2.67</td>
<td>2.85</td>
</tr>
<tr>
<td>110.0</td>
<td>2.78</td>
<td>3.23</td>
<td>4.35</td>
<td>3.34</td>
<td>3.50</td>
</tr>
<tr>
<td>120.0</td>
<td>2.80</td>
<td>3.34</td>
<td>4.57</td>
<td>3.60</td>
<td>3.72</td>
</tr>
<tr>
<td>130.0</td>
<td>2.50</td>
<td>3.01</td>
<td>4.20</td>
<td>3.36</td>
<td>3.41</td>
</tr>
<tr>
<td>140.0</td>
<td>1.93</td>
<td>2.31</td>
<td>3.34</td>
<td>2.69</td>
<td>2.68</td>
</tr>
<tr>
<td>150.0</td>
<td>1.24</td>
<td>1.47</td>
<td>2.21</td>
<td>1.78</td>
<td>1.74</td>
</tr>
<tr>
<td>160.0</td>
<td>0.60</td>
<td>0.71</td>
<td>1.10</td>
<td>0.88</td>
<td>0.85</td>
</tr>
<tr>
<td>170.0</td>
<td>0.16</td>
<td>0.19</td>
<td>0.29</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>180.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>190.0</td>
<td>0.16</td>
<td>0.19</td>
<td>0.29</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>200.0</td>
<td>0.60</td>
<td>0.71</td>
<td>1.10</td>
<td>0.88</td>
<td>0.85</td>
</tr>
<tr>
<td>210.0</td>
<td>1.24</td>
<td>1.47</td>
<td>2.21</td>
<td>1.78</td>
<td>1.74</td>
</tr>
<tr>
<td>220.0</td>
<td>1.93</td>
<td>2.31</td>
<td>3.34</td>
<td>2.69</td>
<td>2.68</td>
</tr>
<tr>
<td>230.0</td>
<td>2.50</td>
<td>3.01</td>
<td>4.20</td>
<td>3.36</td>
<td>3.41</td>
</tr>
<tr>
<td>240.0</td>
<td>2.80</td>
<td>3.34</td>
<td>4.57</td>
<td>3.60</td>
<td>3.72</td>
</tr>
<tr>
<td>250.0</td>
<td>2.78</td>
<td>3.23</td>
<td>4.35</td>
<td>3.34</td>
<td>3.50</td>
</tr>
<tr>
<td>260.0</td>
<td>2.47</td>
<td>2.75</td>
<td>3.65</td>
<td>2.67</td>
<td>2.85</td>
</tr>
<tr>
<td>270.0</td>
<td>2.01</td>
<td>2.14</td>
<td>2.69</td>
<td>1.82</td>
<td>2.04</td>
</tr>
<tr>
<td>280.0</td>
<td>1.59</td>
<td>1.58</td>
<td>1.79</td>
<td>1.12</td>
<td>1.37</td>
</tr>
<tr>
<td>290.0</td>
<td>1.33</td>
<td>1.24</td>
<td>1.28</td>
<td>0.78</td>
<td>1.06</td>
</tr>
<tr>
<td>300.0</td>
<td>1.34</td>
<td>1.22</td>
<td>1.34</td>
<td>0.88</td>
<td>1.20</td>
</tr>
<tr>
<td>310.0</td>
<td>1.64</td>
<td>1.54</td>
<td>2.00</td>
<td>1.44</td>
<td>1.75</td>
</tr>
<tr>
<td>320.0</td>
<td>2.20</td>
<td>2.16</td>
<td>3.10</td>
<td>2.40</td>
<td>2.65</td>
</tr>
<tr>
<td>330.0</td>
<td>2.92</td>
<td>2.99</td>
<td>4.42</td>
<td>3.62</td>
<td>3.77</td>
</tr>
<tr>
<td>Temperature</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>340.0</td>
<td>3.63</td>
<td>3.85</td>
<td>5.66</td>
<td>4.82</td>
<td>4.92</td>
</tr>
<tr>
<td>350.0</td>
<td>4.17</td>
<td>4.51</td>
<td>6.54</td>
<td>5.71</td>
<td>5.80</td>
</tr>
<tr>
<td>360.0</td>
<td>4.38</td>
<td>4.76</td>
<td>6.86</td>
<td>6.04</td>
<td>6.13</td>
</tr>
</tbody>
</table>
Table B11. Potential energy for internal rotation around the C2-C3 bond in n-butane, using PHO, pure QM (HF/6-31G(d)), and MM (OPLS). PHO calculations are done at HF/6-31G(d) level, with the boundary atom placed at the C2, C3 and C4 carbons, denoted by PHO@C2_B, PHO@C3_B and PHO@C4_B, respectively.

<table>
<thead>
<tr>
<th>Dihedral Angle</th>
<th>PHO @C2_B</th>
<th>PHO @C3_B</th>
<th>PHO @C4_B</th>
<th>QM HF/3-21G</th>
<th>OPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>4.45</td>
<td>4.69</td>
<td>7.02</td>
<td>6.19</td>
<td>6.13</td>
</tr>
<tr>
<td>10.0</td>
<td>4.24</td>
<td>4.44</td>
<td>6.68</td>
<td>5.83</td>
<td>5.80</td>
</tr>
<tr>
<td>20.0</td>
<td>3.68</td>
<td>3.78</td>
<td>5.72</td>
<td>4.88</td>
<td>4.92</td>
</tr>
<tr>
<td>30.0</td>
<td>2.93</td>
<td>2.93</td>
<td>4.40</td>
<td>3.63</td>
<td>3.77</td>
</tr>
<tr>
<td>40.0</td>
<td>2.20</td>
<td>2.11</td>
<td>3.00</td>
<td>2.42</td>
<td>2.65</td>
</tr>
<tr>
<td>50.0</td>
<td>1.63</td>
<td>1.49</td>
<td>1.84</td>
<td>1.49</td>
<td>1.75</td>
</tr>
<tr>
<td>60.0</td>
<td>1.32</td>
<td>1.17</td>
<td>1.16</td>
<td>1.02</td>
<td>1.20</td>
</tr>
<tr>
<td>70.0</td>
<td>1.32</td>
<td>1.20</td>
<td>1.09</td>
<td>0.99</td>
<td>1.06</td>
</tr>
<tr>
<td>80.0</td>
<td>1.59</td>
<td>1.54</td>
<td>1.61</td>
<td>1.37</td>
<td>1.37</td>
</tr>
<tr>
<td>90.0</td>
<td>2.05</td>
<td>2.10</td>
<td>2.55</td>
<td>2.05</td>
<td>2.04</td>
</tr>
<tr>
<td>100.0</td>
<td>2.54</td>
<td>2.73</td>
<td>3.59</td>
<td>2.83</td>
<td>2.85</td>
</tr>
<tr>
<td>110.0</td>
<td>2.88</td>
<td>3.21</td>
<td>4.37</td>
<td>3.45</td>
<td>3.50</td>
</tr>
<tr>
<td>120.0</td>
<td>2.92</td>
<td>3.33</td>
<td>4.64</td>
<td>3.65</td>
<td>3.72</td>
</tr>
<tr>
<td>130.0</td>
<td>2.60</td>
<td>2.99</td>
<td>4.29</td>
<td>3.36</td>
<td>3.41</td>
</tr>
<tr>
<td>140.0</td>
<td>2.01</td>
<td>2.30</td>
<td>3.41</td>
<td>2.67</td>
<td>2.68</td>
</tr>
<tr>
<td>150.0</td>
<td>1.29</td>
<td>1.47</td>
<td>2.25</td>
<td>1.75</td>
<td>1.74</td>
</tr>
<tr>
<td>160.0</td>
<td>0.63</td>
<td>0.71</td>
<td>1.11</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>170.0</td>
<td>0.16</td>
<td>0.19</td>
<td>0.29</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>180.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>190.0</td>
<td>0.16</td>
<td>0.19</td>
<td>0.29</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>200.0</td>
<td>0.63</td>
<td>0.71</td>
<td>1.11</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>210.0</td>
<td>1.29</td>
<td>1.47</td>
<td>2.25</td>
<td>1.75</td>
<td>1.74</td>
</tr>
<tr>
<td>220.0</td>
<td>2.01</td>
<td>2.30</td>
<td>3.41</td>
<td>2.67</td>
<td>2.68</td>
</tr>
<tr>
<td>230.0</td>
<td>2.60</td>
<td>2.99</td>
<td>4.29</td>
<td>3.36</td>
<td>3.41</td>
</tr>
<tr>
<td>240.0</td>
<td>2.92</td>
<td>3.33</td>
<td>4.64</td>
<td>3.65</td>
<td>3.72</td>
</tr>
<tr>
<td>250.0</td>
<td>2.88</td>
<td>3.21</td>
<td>4.37</td>
<td>3.45</td>
<td>3.50</td>
</tr>
<tr>
<td>260.0</td>
<td>2.54</td>
<td>2.73</td>
<td>3.59</td>
<td>2.83</td>
<td>2.85</td>
</tr>
<tr>
<td>270.0</td>
<td>2.05</td>
<td>2.10</td>
<td>2.55</td>
<td>2.05</td>
<td>2.04</td>
</tr>
<tr>
<td>280.0</td>
<td>1.59</td>
<td>1.54</td>
<td>1.61</td>
<td>1.37</td>
<td>1.37</td>
</tr>
<tr>
<td>290.0</td>
<td>1.32</td>
<td>1.20</td>
<td>1.09</td>
<td>0.99</td>
<td>1.06</td>
</tr>
<tr>
<td>300.0</td>
<td>1.32</td>
<td>1.17</td>
<td>1.16</td>
<td>1.02</td>
<td>1.20</td>
</tr>
<tr>
<td>310.0</td>
<td>1.63</td>
<td>1.49</td>
<td>1.84</td>
<td>1.49</td>
<td>1.75</td>
</tr>
<tr>
<td>320.0</td>
<td>2.20</td>
<td>2.11</td>
<td>3.00</td>
<td>2.42</td>
<td>2.65</td>
</tr>
<tr>
<td>330.0</td>
<td>2.93</td>
<td>2.93</td>
<td>4.40</td>
<td>3.63</td>
<td>3.77</td>
</tr>
<tr>
<td>Value</td>
<td>Column 1</td>
<td>Column 2</td>
<td>Column 3</td>
<td>Column 4</td>
<td>Column 5</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>340.0</td>
<td>3.68</td>
<td>3.78</td>
<td>5.72</td>
<td>4.88</td>
<td>4.92</td>
</tr>
<tr>
<td>350.0</td>
<td>4.24</td>
<td>4.44</td>
<td>6.68</td>
<td>5.83</td>
<td>5.80</td>
</tr>
<tr>
<td>360.0</td>
<td>4.45</td>
<td>4.69</td>
<td>7.02</td>
<td>6.19</td>
<td>6.13</td>
</tr>
</tbody>
</table>
Table B12. OPLS-AA Bond Stretching and Angle Bending Parameters of Butane

<table>
<thead>
<tr>
<th>Type</th>
<th>$R_{eq}$ or $\theta_{eq}$</th>
<th>$K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT-CT</td>
<td>1.529</td>
<td>268.0</td>
</tr>
<tr>
<td>HC-CT</td>
<td>1.090</td>
<td>340.0</td>
</tr>
<tr>
<td>HC-CT-HC</td>
<td>107.8</td>
<td>33.00</td>
</tr>
<tr>
<td>HC-CT-CT</td>
<td>110.7</td>
<td>37.50</td>
</tr>
<tr>
<td>CT-CT-CT</td>
<td>112.7</td>
<td>58.35</td>
</tr>
</tbody>
</table>
Table B13. OPLS-AA Fourier Coefficients (kcal/mol) for Torsional Energy Functions of Butane

<table>
<thead>
<tr>
<th>Dihedral</th>
<th>V₁</th>
<th>V₂</th>
<th>V₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-C-C-H</td>
<td>0.000</td>
<td>0.000</td>
<td>0.318</td>
</tr>
<tr>
<td>H-C-C-C</td>
<td>0.000</td>
<td>0.000</td>
<td>0.366</td>
</tr>
<tr>
<td>C-C-C-C</td>
<td>1.740</td>
<td>-0.157</td>
<td>0.279</td>
</tr>
</tbody>
</table>
Table B14. OPLS-AA Non-Bonded parameters of Butane

<table>
<thead>
<tr>
<th>Atom</th>
<th>$q, e^-$</th>
<th>$A$</th>
<th>kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>C,RCH$_3$</td>
<td>-0.180</td>
<td>3.500</td>
<td>0.066</td>
</tr>
<tr>
<td>C,RCH$_2$</td>
<td>-0.120</td>
<td>3.500</td>
<td>0.066</td>
</tr>
<tr>
<td>H,RH,alkanes</td>
<td>0.060</td>
<td>2.500</td>
<td>0.030</td>
</tr>
</tbody>
</table>
Appendix C

The following pages include my other publications that I have contributed as a part of collaboration during doctoral studies.
Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields

Jiali Gao,*1,2 Donald G. Truhlar,2 Yingjie Wang,2 Michael J. M. Mazack,2 Patrick Löffler,2 Makenzie R. Provorse,3 and Pavel Rehak2

1. Theoretical Chemistry Institute
State Key Laboratory of Theoretical and Computational Chemistry
Jilin University, Changchun, Jilin Province 130028, P. R. China

and

2. Department of Chemistry and Supercomputing Institute
University of Minnesota, Minneapolis, MN 55455, U.S.A.

Email: gao@jialigao.org

Conspectus

Molecular mechanical force fields have been successfully used to model condensed-phase and biological systems for a half century. Thanks to careful parametrization, such classical force fields can be used to provide useful interpretations of experimental findings. Yet, there is a need to further improve computational accuracy for quantitative prediction of biomolecular interactions. A new strategy is presented to construct the potential energy surface for macromolecular simulations on the basis of quantum mechanical formalisms rather than merely using quantum chemical results to fit parameters. In this spirit, this approach is called a quantum mechanical force field (QMFF).

QMFF is based on partition of a macromolecular system into individual molecular fragments to achieve computational efficiency, whereas the mutual electronic polarization of interfragment interactions is treated explicitly using quantum mechanics.
Such an explicit polarization (X-Pol) theory not only can be used as a general, multilevel electronic structure model for macromolecular systems, but also it provides a framework for developing next-generation force fields. As a quantum chemical model, a variational many-body (VMB) expansion approach is used to approximate and systematically improve interfragment interactions, including exchange repulsion, charge delocalization, and dispersion/correlation energies, $\Delta E_{\text{XCD}}$. Here, it is critical to define a good monomer energy such as that from variational X-Pol to minimize many-body correction contributions. This is important because the number of high-order terms increases rapidly, severely hampering computational efficiency along with large error propagation.

As a quantum mechanical force field, $\Delta E_{\text{XCD}}$ is approximated by using empirical functions in the spirit of traditional molecular mechanics to strive for accuracy. In this work, the feasibility of carrying out molecular dynamics simulations of solvated proteins in aqueous solution using an X-Pol QMFF and the accuracy of the X-Pol potential in condensed phase simulations are illustrated. The method can also provide insights on physical properties that cannot be directly obtained using classical force fields. For example, the average charges of the same functional group along the peptide chain depend strongly on the secondary structure interactions, whereas they are fixed in most classical force fields. VMB2 calculations of the ionic liquid of ethyl-methyl-imidazolium acetate reveal that the amount of charge transfer between anion and cation is in fact rather small in this system, in contrast to classical simulations employing charge-scaling as large as 70% of the formal value, Importantly, the development of QMFFs offers an opportunity to extend the scope of applications of current models to problems involving
the change of potential energy surfaces, such as molecular spectroscopies, chemical reactions, and electron and energy transfer processes.
1. Introduction

Molecular mechanical (MM) force fields are traditionally used in computer simulations of macromolecular systems, including proteins and nucleic acids in aqueous solution.\textsuperscript{1,2} Although these models are computationally efficient for applications to large systems, there are also a number of well-known difficulties that are not easily resolved within the current model,\textsuperscript{3} including the choice of specific energy terms, harmonic approximation for bond stretch and angle bending and the coupling of internal degrees of freedom, polarization effects and charge transfer, and the direct modeling of system-wide chemical reactions such as proton transport. There is a great need to further improve the accuracy to achieve quantitative prediction of biomolecular interactions such as ligand binding, electron and energy transfer as well as enzymatic reactions. It is timely to ask the question what type of potential energy functions will be used in the future for biomolecular simulations. One possibility is to continue improving the current MM approach, including the incorporation of classical polarization terms in the force fields, and the second is to develop a new theoretical framework with the capability of including quantum mechanical effects explicitly to model intermolecular interactions.\textsuperscript{4-8}

Quantum mechanics, in principle, can provide both reactive and nonreactive potential energy surfaces, including not only electrostatics and van der Waals forces, but also polarization and charge transfer effects. Yet, it is limited by the computational cost that increases rapidly with the size of the system. To this end, a variety of fragment-based electronic structure methods have been developed to increase the computational speed for large molecular systems,\textsuperscript{9} many of which are illustrated in this Special Issue. In particular, York and coworkers pioneered a version of the divide-and-conquer strategy
for linear-scaling quantum-mechanical (QM) representation of macromolecular systems as a force field, and Herbert discussed the issue of error propagation in many-body expansion theory, highlighting the necessity for an excellent approximation of the monomer reference state to reduce energy correction in many-body terms. Recently, a good summary of fragment-based methods can be found in Ref., and thus, for brevity, we shall not make an extensive list of citations here on these methods. This Account presents one alternative, namely, the explicit polarization theory, for the construction of a next-generation force field on the basis of quantum chemical formalisms.

The explicit polarization (X-Pol) model is a fragment-based QM method, in which the entire system is divided into molecular subunits, which can be individual molecules, ions, ligands or cofactors, and amino acid residues or a group of these entities. The key assumption in the X-Pol method is that the wave function of the entire system is approximated as a Hartree product of the antisymmetric wave functions of individual fragments. Consequently, the optimization of the total wave function can be reduced to the optimization of each fragment embedded in and polarized by the rest of the system, thereby, reducing the computational cost to nearly linear scaling with respect to the number of fragments. Clearly, variational optimization of the mutual dependence of the fragmental wave functions is critical to the success of this method. As a force field, the energy of each fragment is determined by the electronic structure method used, whereas intermolecular interactions are modeled through electrostatic embedding. The short-range exchange-repulsion interactions, charge transfer, and dispersion interactions and correlation energy between different fragments are modeled empirically as in
Alternatively, these energy contributions can be modeled by density-dependent functionals, by Hartree-Fock (HF) exchange, or by making use of many-body expansion corrections.\textsuperscript{13}

In the following, we describe the theoretical approach of X-Pol and illustrate computational results that demonstrate the feasibility and accuracy of macromolecular simulations employing the X-Pol quantum mechanics force field (QMFF).

2. Theoretical Background

In X-Pol, a macromolecular system is partitioned into molecular fragments,\textsuperscript{4,5,15} which may be called monomers. For solutions with small solute molecules, a fragment can be a single solute or solvent molecule.\textsuperscript{15,18,21} For a polypeptide chain, the peptide unit as defined by the IUPAC nomenclature (rather than the conventional residue)\textsuperscript{22} is assigned to be the smallest fragment, which contains the atoms \(-C^l_{\alpha}R^lCO - NH&C^l_{\alpha}^{+1}H\) – in the conventional residues \(I\) and \(I+1\), where each \(C^l_{\alpha}\) atom is shared by two adjacent peptide units (Figure 1).\textsuperscript{5,23} The connection approach is an extension of the generalized hybrid orbital (GHO) method,\textsuperscript{24} originally developed for combined QM/MM applications.\textsuperscript{25} Unlike other capping schemes by hydrogen atom or small functional groups,\textsuperscript{26} the GHO method does not alter the number of degrees of freedom, nor electrostatic interactions between neighboring fragments. A buffering scheme was used to accelerate the convergence.\textsuperscript{27} Thus, the hybrid orbital approach provides a seamless transition from one fragment to the next across chemical bonds. Several peptide units can be combined into the same fragment, if desired, which can be useful for modeling systems containing disulfide bonds.
The X-Pol model is derived from a standard electronic structure method by a hierarchy of three approximations. First, the molecular wave function of the entire system $\Phi$ is approximated as a Hartree product of the antisymmetric wave functions of the individual fragments $\{\Psi_A; A = 1, \cdots, N\}$, where $N$ is the number of fragments.

$$\Phi = \prod_A^{N} \Psi_A \quad (1)$$

The effective Hamiltonian of the system is given by
\[ \hat{H} = \sum_{A} \hat{H}_{A}^{0} + \frac{1}{2} \sum_{A} \sum_{B \neq A} (\hat{H}_{A}^{\text{int}}[V_{E}^{B}] + \Delta E_{AB}^{\text{XCD}}), \]  

where the first term sums over the Hamiltonians of all isolated fragments \( \{\hat{H}_{A}^{0}\} \), and the second, double summation accounts for pairwise interactions among all the fragments. In eq 2, \( \hat{H}_{A}^{\text{int}}[V_{E}^{B}] \) represents electrostatic interactions between fragments \( A \) and \( B \) with \( V_{E}^{B} \) being the potential due to \( B \), and the final term \( \Delta E_{AB}^{\text{XCD}} \) specifies exchange-repulsion (X), charge delocalization (C), dispersion and other interfragment correlation (D) energy contributions.

A general approach to represent the external potential \( V_{E}^{B} \) due to the charge density of fragment \( B \), is to use a multicenter multipole expansion,\(^4,5,17\) of which the simplest form is to limit the expansion to the monopole terms, so the result only depends on the partial atomic charges.\(^15\) The use of partial atomic charges to approximate \( V_{E}^{B} \) is particularly convenient for constructing the effective Hamiltonian of eq 2, and this is the strategy that has been adopted in the X-Pol method.\(^4,5,15-18,21\) This strategy has been used in other studies.\(^8,12,28\)

The total energy of the system in the X-Pol method is simply given by

\[ E_{\text{X-Pol}} = \langle \Phi | \hat{H} | \Phi \rangle. \]  

The Hartree product wave function in eq 1 implies that interfragment charge delocalization and exchange-repulsion interactions arising from the Pauli exclusion principle are neglected.\(^29,30\) However, these interactions as well as interfragment dispersion interactions make critical contributions to intermolecular interactions. Together, they are called \( \Delta E_{\text{XCD}} \), and they must be properly accounted for in the X-Pol
energy expression. As a QM model, a brute force approach is to employ variational many-body expansion (VMB) theory to make two-body, three-body, and higher order corrections. Of course, there are also numerous non-variational, charge-embedding many-body expansion approaches, and the review by Gordon et al. extensively covered these methods. When two- and three-body terms are included, we have, respectively, the corresponding VMB2 and VMB3 corrections,

\[
\Delta E_{XCD}^2 \approx \Delta E_{VMB}^2 = \sum_{I<J}^N (E_{IJ} - E_{X-Pol}), \quad (4)
\]

\[
\Delta E_{XCD}^3 \approx \Delta E_{VMB}^3 = \sum_{I<J<K}^N (E_{IK} - E_{X-Pol} - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK}). \quad (5)
\]

In eqs 4 and 5, \( E_{IJ} \) and \( E_{IK} \) are, respectively, the X-Pol energies when two and three monomers are grouped into a single super-fragment. Thus, the difference in parentheses in eq 4 gives the two-body correction energy, \( \Delta E_{IJ} \), for exchange and charge delocalization effects at the Hartree-Fock level plus dispersion-correlation contributions when correlated methods are used. Typically, the size of the monomer and dimer fragments are significantly smaller than the full system, making the computation costs negligibly cheaper than a single calculation of the entire system. However, the number of terms increases rapidly with the number of fragments and the order of correction included, rendering the many-body expansion approach impractical beyond the two-body correction terms. Another major issue is error propagation accompanying SCF convergence as the number of many-body correction terms increases. Herbert showed that the expansion energy diverges even at the three-body level for systems as small as a cluster of 64 water molecules. Thus in using this approach, it is critical to define a
reference state for the monomer energy such that higher-order (>2) terms are negligible.\textsuperscript{13,34}

When the X-Pol method is used as a theoretical framework to develop force fields for condensed-phase and macromolecular systems, a simpler, empirical approach can be adopted, such as Lennard-Jones or Buckingham potentials (as used in molecular mechanics)\textsuperscript{7,15,17,20} or perturbation theory\textsuperscript{35} to estimate the dimer $\Delta E^{XCD}_{AB}$ terms.

There are two ways of constructing the Fock matrix for optimizing the fragmental wave functions, and they distinguish the variational X-Pol from other fragment-based methods.

(a) \textbf{Variational X-Pol}. With the use of partial atomic charges to approximate $V_E^B$, the Fock operator for a fragment, $A$, is derived variationally to yield\textsuperscript{16,17,27}

$$
F^{A,X\text{pol}}_{\mu\nu} = F^{A,o}_{\mu\nu} - \frac{1}{2} \sum_{B \neq A} \sum_{b \in B} q_b^B \left( I_b^B \right)_\mu^A + \frac{1}{2} \sum_{a \in A} X_a^A \left( \Lambda_a^A \right)_\mu^\nu,$$

where $F^{A,o}_{\mu\nu}$ is the Fock matrix element for the Hamiltonian of the isolated fragment $A$, $q_b^B$ is the point charge on atom $b$ of fragment $B$, $I_b^B$ is the matrix of the one-electron integrals of the embedding potential due to fragment $B$, and the last term represents the response of the charge density of fragment $B$ due to variational optimization of the wave function of fragment $A$ ($X_a^A$ is the derivative of the energy with respect to the atomic charge on $a$, and $\Lambda_a^A$ is the charge derivative with respect to the density, see Ref. \textsuperscript{36}).

(b) \textbf{Embedding X-Pol}. If each fragment is considered to be embedded in the static field of the rest of the system, one can construct a Fock operator for fragment $A$ heuristically as follows,\textsuperscript{4,9,15,33} akin to a combined QM/MM approach:\textsuperscript{25}
In the charge-embedding approach, the mutual polarization among all fragments in the system is achieved by iteratively updating the partial atomic charges \( \{ q_b^B \} \). However, it should be emphasized that the use of eq 7 is not a variational procedure and it does not include the response of the charge density due to the changes of other charges. To emphasize this difference, in the rest of this article and in general, method (a) will be specifically called variational X-Pol, and the subsequent many-body expansion corrections are called VMB2 and VMB3. The embedding approach (b) is simply called X-Pol, and the many-body corrections by MB2 and MB3.

In comparing methods (a) and (b), the variational X-Pol method has the advantage of allowing the computation of analytic gradients for efficient geometry optimization and dynamics simulations.\(^{16,17,27}\) Furthermore, the total energy obtained from the variational procedure is necessarily lower than that from the charge-embedding scheme. Consequently, it is expected that the use of the variational X-Pol energy as the monomer energy reference state in many-body energy expansion is more efficient than other alternatives.\(^{13}\) Although it is possible to obtain analytic gradients for the non-variational, charge-embedding approaches,\(^{37}\) it generally involves solution of coupled-perturbed self-consistent field (CPSCF) equations. Often in fragment-based methods, those response terms have simply been ignored\(^{38}\) without their numerical consequences fully investigated.
3. Illustrative Examples

3.1. Multilevel X-Pol as a quantum chemical model for macromolecules

The X-Pol method can be used in terms of different electronic structure representations for different fragments. This provides a general, multilevel QM/QM treatment for large systems, where the region of interest could be modeled by a high-level electronic structure theory, embedded in the environment modeled by a lower level representation. To illustrate this possibility, a number of arbitrarily selected combinations of different electronic models have been used to describe the bimolecular complexes between CH$_3$CO$_2$H and H$_2$O and between a Zundel ion H$_5$O$_2^+$ and four water molecules. In a hybrid X-Pol calculation with different QM models, the interaction between different fragments is coupled by the electrostatic potential $V_E^B$ in eq 2. Two different charge models, namely Mulliken population analysis (MPA) and electrostatic potential (ESP) charge-fitting, were used to construct the charge-embedding Fock matrix (eq 7), whereas only the MPA charges were used in the variational X-Pol (eq 6) model.

The X-Pol interaction energies between H$_5$O$_2^+$ and (H$_2$O)$_4$ are listed in Table 1, and they tend to overestimate the binding interactions since interfragment exchange-repulsion interactions are neglected. With the inclusion of the $\Delta E_{XCD}$ energy, derived from an energy decomposition analysis using CCSD/MS3G and HF/MS3G results, the computed binding energies are significantly improved. The CCSD/M06 combination yielded reasonable agreement (-68.0 kcal/mol) with the binding energy estimated using full CCSD(T)/MS3G (-69.7 kcal/mol). The errors from other combinations range from 2.3 kcal/mol to 8.9 kcal/mol. Clearly, it is important to develop a systematic procedure to improve the accuracy of X-Pol (both variational and embedding models) beyond the
monomer approximation. Two alternatives can be followed, which distinguish the use of X-Pol as a multi-level quantum chemical method presented in this section, or as a quantum mechanical force field that introduces empirical energy terms in the next section.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Charge-embedding</th>
<th>Variational</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ESP</td>
<td>MPA</td>
</tr>
<tr>
<td>M06</td>
<td>M06</td>
<td>-89.1</td>
<td>-87.5</td>
</tr>
<tr>
<td>M06</td>
<td>B3LYP</td>
<td>-87.7</td>
<td>-85.2</td>
</tr>
<tr>
<td>M06</td>
<td>HF</td>
<td>-92.0</td>
<td>-91.7</td>
</tr>
<tr>
<td>MP2</td>
<td>HF</td>
<td>-92.9</td>
<td>-92.7</td>
</tr>
<tr>
<td>CCSD</td>
<td>M06</td>
<td>-89.5</td>
<td>-88.0</td>
</tr>
</tbody>
</table>

The effects due to the Hartree-product approximation to the molecular wave function in X-Pol can be systematically corrected using variational and embedding many-body expansions (VMBn and MBn). Table 2 displays the total interaction energies of four minimum energy structures for the water hexamer, determined up to the third order VMB correction. Although there are significant errors at the monomer level for neglecting the ΔE_{XCD} contributions, the binding energies are
significantly improved with the dimer correction (VMB2), and the computed results at the VMB3 level are in quantitative agreement with those from the corresponding full QM calculations. The VMB results are systematically in better agreement with the binding energies from full QM calculations than the corresponding charge-embedding approach, even the charges have been iteratively optimized at the X-Pol level. The difference is particularly striking at the two-body correction level, with average errors of 2.2 and 1.5 kcal/mol from VMB2/HF and VMB2/PMO, and 4.2 and 6.1 kcal/mol from MB2/HF and MB2/PMO, respectively. This emphasizes the critical role of defining a good reference, monomer energy in MB expansion theory, and the variational X-Pol energy leads to
better convergence in correction terms. Including three-body energies, the errors are less than 1 kcal/mol.

Table 3. Computed interaction energies for a (H₂O)₆₅ water cluster using variational and embedding X-Pol and two-body corrections. The 6-31G(d) basis set was used in ab initio and DFT calculations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Full QM</th>
<th>VX-Pol</th>
<th>VMB2</th>
<th>X-Pol</th>
<th>MB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>-645.5</td>
<td>-1008.7</td>
<td>-691.2</td>
<td>-942.4</td>
<td>-638.8</td>
</tr>
<tr>
<td>MP2</td>
<td>-975.6</td>
<td>-915.3</td>
<td>-917.3</td>
<td>-877.3</td>
<td>NC</td>
</tr>
<tr>
<td>B3LYP</td>
<td>-887.1</td>
<td>-861.5</td>
<td>-953.7</td>
<td>-797.6</td>
<td>-874.4</td>
</tr>
<tr>
<td>B3LYP-D</td>
<td>-1042.7</td>
<td>-861.5</td>
<td>-1109.4</td>
<td>-797.6</td>
<td>-1030.1</td>
</tr>
<tr>
<td>PMO</td>
<td>-735.5</td>
<td>-432.6</td>
<td>-720.9</td>
<td>-344.8</td>
<td>-602.6</td>
</tr>
</tbody>
</table>

a. VMB3 and MB3 values, respectively.

Although results for the small water clusters are encouraging, we found that significant errors exist for larger water clusters (Table 3). We examined a system containing 65 molecules, taken from a configuration in liquid simulations by keeping water molecules within 7.5 Å of a monomer center. We used HF, MP2, B3LYP and B3LYP-D, and the semiempirical PMO methods to describe each monomer and dimer pairs, all with the 6-31G(d) basis set. The absolute binding energies are certainly not converged from these methods, but the main purpose here is to test the performance of many-body expansion theory. The average unsigned error using the variational model at
the VMB2 level is about 60 kcal/mol, or about 1 kcal/mol per monomer among the four
ab initio QM models. Interestingly, the error is much smaller employing the charge-
embedding MB2 approach with fixed background charges, which is about 11 kcal/mol
for the complex, or 0.17 kcal/mol per monomer. Interestingly, the performance of the
semiempirical PMO model\textsuperscript{18} for water is exceptionally good. For comparison, in the
fragment molecular orbital approach, the error for a 64-water cluster was 0.003 a.u. (1.9
kcal/mol) per monomer fragment from a two-body correction (FMO2) calculation
employing MP2.\textsuperscript{43} The FMO2 procedure is very similar to the MB2 approach. The origin
of relatively large errors in VMB is due to charge penetration effects in X-Pol
optimization. While this is not significant with a minimal basis such as that employed in
semiempirical methods, it can become unphysically large when large basis sets are used.
In future studies, an explicit exchange potential may be needed to prevent unphysical
charge penetration effects between different fragments.

Charge penetration effects are much smaller in many-body corrections using the
PMO semiempirical method, as illustrated for a range of water clusters with and without
periodic boundary conditions. In this case, both VMB3 and MB3 binding energies are
nearly in quantitative agreement with the full QM results (Supporting Information);
however, MB3 is much more efficient than VMB3 in energy calculations since the SCF
is only carried out on dimer fragments with fixed background charges.
Figure 2. Computed dimeric charge transfer energy versus the amount of charge transferred from a donor water molecule into an acceptor water molecule for a water cluster system consisting of 65 water molecules.

Figure 2 displays the interaction energy due to charge delocalization with respect to charge transferred for the cluster of 65 water molecules at the two-body VMB2 level with HF/6-31G(d). Interestingly, there is a good correlation between charge transfer energy and the net amount of charge migration between two monomers. Similar correlations have been observed in cation-... complexes from energy decomposition analysis, and it can be rationalized based on perturbation theory.\(^{44}\)

3.2. X-Pol as a quantum mechanical force field

Although \textit{ab initio} molecular orbital theory and density functional theory can be used to improve the accuracy of X-Pol results for large systems, it is still impractical to use these methods to perform molecular dynamics simulations for an extended period of time. With increased computing power, this will become feasible in the future; however, at present, it is desirable to use semiempirical methods such as the popular approaches...
based on neglect of diatomic differential overlap (NDDO) \(^{45}\) or the more recent self-consistent-charge tight-binding density functional (SCC-DFTB) \(^{46}\) method to model condensed-phase systems and biomacromolecules.

The construction of a QMFF based on the variational X-Pol formalism has two components. First, a computationally efficient quantum chemical model is needed to describe the electronic structure of individual molecular fragments. Recently, a polarized molecular orbital (PMO) method was introduced by Truhlar and Gao and coworkers, \(^{47}\) which is based on the MNDO formalism with the addition of a set of \(p\)-orbitals on each hydrogen atom. It was found that the computed molecular polarizabilities for a range of compounds containing hydrogen, carbon, oxygen and fluorine are significantly improved. \(^{21,47-49}\) In addition to the enhancement in computed molecular polarizability, a damped dispersion function is included as a post-SCF correction to the electronic energy. \(^{50}\) Second, a practical and *parametrizable* procedure is desired to model interfragment electrostatic and exchange-dispersion interactions. For the electrostatic component, a dipole-preserving and polarization consistent (DPPC) method was used to derive partial atomic charges that can exactly reproduce the instantaneous molecular dipole moment from the polarized electron density of each fragment. \(^{51}\) Since the DPPC charges are optimized by a Lagrange multiplier technique, there are no empirical parameters. For the \(\Delta E_{\text{XCD}}\) term, pairwise Lennard-Jones potentials were adopted, which contain two parameters for each atomic number. \(^{15,17,20}\) Employing this strategy, X-Pol quantum chemical models for water \(^{18}\) and for hydrogen fluoride \(^{21}\) have been developed for fluid simulations.
Table 4. Computed liquid properties of the XP3P model for water along with those from experiments, and the TIP3P, AMOEBA, and SWM4-NDP models.a

<table>
<thead>
<tr>
<th></th>
<th>XP3P</th>
<th>TIP3P</th>
<th>AMOEBA</th>
<th>SWM4-NDP</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_v$, kcal/mol</td>
<td>10.42 ± 0.01</td>
<td>10.41</td>
<td>10.48</td>
<td>10.51</td>
<td>10.51</td>
</tr>
<tr>
<td>density, g/cm³</td>
<td>0.996 ± 0.001</td>
<td>1.002</td>
<td>1.000</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>$C_p$, cal mol⁻¹ K⁻¹</td>
<td>21.8 ± 1.0</td>
<td>20.0</td>
<td>20.9</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td>$10^6$, atm⁻¹</td>
<td>25 ± 2</td>
<td>60</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$, K⁻¹</td>
<td>37 ± 3</td>
<td>75</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gas, D</td>
<td>1.88</td>
<td>2.31</td>
<td>1.77</td>
<td>1.85</td>
<td>1.85</td>
</tr>
<tr>
<td>liq, D</td>
<td>2.524 ± 0.002</td>
<td>2.31</td>
<td>2.78</td>
<td>2.33</td>
<td>2.3-2.6</td>
</tr>
<tr>
<td>$10^3 D$, cm²/s</td>
<td>2.7</td>
<td>5.1</td>
<td>2.02</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>$\kappa$, (ps)</td>
<td>97 ± 8</td>
<td>92</td>
<td>82</td>
<td>79 ± 3</td>
<td>78</td>
</tr>
<tr>
<td>$D_r$, (ps)</td>
<td>8.8</td>
<td>11 ± 2</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$NMR_r$, (ps)</td>
<td>2.6</td>
<td>1.87 ± 0.03</td>
<td>2.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. $H_v$, heat of vaporization; $C_p$, heat capacity; $\kappa$, isothermal compressibility; $\alpha$, coefficient of thermal expansion; $\beta$, dipole moment; $D$, diffusion constant; $\varepsilon$, dielectric constant; $D_r$, Debye relaxation time; and $NMR_r$, NMR rotational relaxation time.

The results from simulations of liquid water are displayed in Table 4 at 25 °C and 1 atm, along with results from the TIP3P pairwise potential, and from two polarizable force fields, namely AMOEBA and SWM4-NDP. The data in Table 4 were obtained from an assemblage of over 100 million configurations in Monte Carlo and 500 ps of molecular dynamics simulations for a cubic system containing 267 water molecules; the latter took about 24 hours (1 fs integration step) on an eight-core workstation, while about 7 million configurations can be executed in Monte Carlo sampling. The standard
errors (± 1σ) were obtained from fluctuations of separate averages over blocks of 2-4×10^5 configurations. The average density and heat of vaporization of the XP3P water are within 1% of the experimental values and are similar to results obtained with other polarizable and non-polarizable force fields (see Table 4). Quantities involving intermolecular fluctuations such as isothermal compressibility, coefficient of thermal expansion and dielectric constant are more difficult to converge, but, overall, the agreement with experiment is good, and the performance of the XP3P model is as good as any other empirical force fields in dynamics simulations.

Monte Carlo simulations of liquid water were performed at temperatures ranging from -40 to 100 °C, and Figure 3 shows that water molecules in the liquid experience a wide spectrum of instantaneous electrostatic fields from the rest of the system, reflected

![Figure 3. Distribution of the molecular dipole moment of water in the liquid at temperatures ranging from -40 °C to 100 °C.](image)
by the shift in $\langle \mu_{\text{liq}} \rangle$ and the distribution of the instantaneous molecular dipole moments. It is interesting to note that the maximum value in the dipole probability density distribution is the same at different temperatures, despite the large change in the dipole fluctuations, suggesting that there is polarization saturation in liquid water. The average molecular dipole moment of water in the liquid at 25 °C, $\langle \mu_{\text{liq}} \rangle$, was determined to be 2.524 ± 0.002 D, which represents an increase of 35% relative to the gas-phase value (1.88 D from the PMOw Hamiltonian). There is no experimental data for direct comparison. Based on analysis of dielectric screening effects of water, Sprik pointed out that an average dipole moment of 2.5–2.6 D in liquid water would most likely yield the correct dielectric constant.\(^{52}\)
All other thermodynamic and dynamic properties determined using the XP3P model in Table 4 are in reasonable accord with experiments and are of similar accuracy in comparison with other empirical models. In contrast to the large number of polarizable force fields in the literature that are based on classical approximations, the electronic polarization from the present XP3P model is explicitly described based on a quantum chemical formalism.

Figure 4 shows the structure of liquid water characterized by radial distribution functions, $g_{xy}(r)$. In comparison with the neutron scattering data, the computational results are in excellent agreement with experiments. In particular, a well-resolved minimum following the first peak in the O-O distribution was obtained. For the XP3P potential, the location of the maximum of the first peak of the O-O RDF is $2.78 \pm 0.05$ Å.
with a peak height of 3.0. For comparison, the corresponding experimental values are 2.73 Å and 2.8 from neutron diffraction.\textsuperscript{53}

The X-Pol method was used in molecular dynamics simulations of an ionic liquid for a system consisting of 125 1-ethyl-3-methylimidazolium (EMIM) and acetate ion pairs in a periodic box using the AM1 model\textsuperscript{54} to represent each ion as a molecular fragment. Traditional molecular dynamics simulations often employ scaled-charge force fields to approximate charge delocalization (transfer) effects between anions and cations, and scaling factors ranging from 0.7 to 0.9 have been used.\textsuperscript{55-57} Employing the two-body VMB2 scheme, it was found that there is relatively a small amount of charge transfer from acetate to EMIM, with an average departure of 0.028 ± 0.015 e from unity.\textsuperscript{58} In contrast, the average charge transfer over five minimum energy configurations in the gas phase was 0.104 ± 0.026 e, similar to other studies.\textsuperscript{55} Thus, it appears that condensed phase polarization effects reduces significantly the amount of charge transfer between cation and anions in ionic liquids, casting doubt on the use of scaled charge models significantly deviating from unity in dynamics simulations.

To illustrate the importance of electronic polarization effects in macromolecular simulations, the small protein bovine pancreatic trypsin inhibitor (BPTI) in water with periodic boundary conditions was carried out using the standard AM1 model.\textsuperscript{23} The system contained a total of 14,281 atoms and 29,026 basis functions, for which 3.2 ps (1 fs time-step) could be performed per day on a single processor (1.66 GHz) computer. Figure 5 depicts the average partial atomic charge of the carbonyl group for each amino acid of the polypeptide chain. There is significant fluctuation of the charge distribution in the backbone carbonyl group depending on its sequence location that has different
electrostatic environment as a result of specific hydrogen-bonding interactions with the solvent and with other amino acids to form secondary structures. For comparison, the standard CHARMM22 force field makes use of neutral group convention where the total net atomic charge of the peptide carbonyl group is zero and invariant in all conformation substates. Since standard semiempirical methods are known to underestimate intermolecular polarization effects, the findings from this study likely represent a lower limit of the polarization effects.

4. Future Perspective

Molecular mechanical force fields have traditionally been used to model condensed-phase and biological systems. Thanks to careful parametrization, such classical force fields can be used to provide useful interpretation of experimental findings. In this Account, we presented a new strategy to construct the potential energy

![Figure 5. Average net partial charges (in atomic units) on the backbone carbonyl (C=O) groups of BPTI. The carbonyls are arranged in order of sequence number.](image)
surface for macromolecular systems on the basis of a quantum mechanical formalism. Here, electronic structure theory is directly used to model intermolecular interactions, rather than using quantum chemical results merely as the target for fitting empirical parameters. In this spirit, this class of approaches is called quantum mechanical force fields (QMFF).

Our strategy is based on the partition of condensed-phase and macromolecular systems into fragments to achieve computational efficiency. The mutual polarization between different fragments is treated explicitly using quantum mechanics. The present explicit polarization (X-Pol) theory can be used as a general, multilevel electronic structure model for macromolecular systems such that different fragments can be represented by different quantum chemical methods. The X-Pol wave function can be optimized either using a variational Fock operator or through a charge-embedding iterative approach. In both cases, interfragment electrostatic interactions are directly included, whereas exchange-repulsion, charge delocalization, and dispersion and correlation energies, \( \Delta E_{XCD} \), can be systematically incorporated using variational many-body (VMB) expansion. On the other hand, X-Pol provides a framework for developing next-generation force fields by approximating \( \Delta E_{XCD} \) with empirical functions. In the spirit of traditional classical force fields, X-Pol QMFF strives for accuracy by carefully parametrizing these empirical terms.

So far, we have shown the feasibility of X-Pol for molecular dynamics simulations of a solvated protein in aqueous solution, and the accuracy of the X-Pol QMFF for condensed phase simulations, including liquid water and hydrogen fluoride. The method can also provide insights on physical properties that cannot be directly
obtained using classical force fields. For example, in contrast to the widely used charge-scaling schemes with fractional charges as small as 70% of the formal value in dynamics simulations of ionic liquids, VMB2 calculations showed that the amount of charge transfer between anions and cations is rather small in 1-ethyl-3-methylimidazolium acetate ionic liquid. Yet, much needs to be improved in order for a QMFF such as X-Pol to yield accurate results comparable to the current MM force fields since the latter has been thoroughly parametrized for over half a century by numerous research groups around the globe.

Since an important goal is to carry out large temporal and spatial scale simulations of macromolecular systems, it is essential to further develop more accurate and efficient quantum chemical models to represent the electronic structure of individual fragments. It appears that both the DFTB\textsuperscript{46} and NDDO-based models that include explicitly orthogonalization constraints may provide the essential ingredients for more accurate conformational energies. Second, it would be desirable to explicitly model exchange and dispersion interactions in the effective Hamiltonian. Third, since molecular fragments can be considered as effective valence bond states, it is possible to construct multistate methods for studying solvent reorganization energies and non-adiabatic processes in condensed phases. Finally, in many-body expansion theory, it is critical to start with a good monomer reference state such as that defined by the variational X-Pol to minimize high-order correction energies. This is important because the rapid increase in high-order terms can severely hamper computational efficiency along with increased error accumulation. The development of QMFFs is still in its infancy,\textsuperscript{5-8} but it offers an opportunity to further improve computational accuracy on properties involving the
change of potential energy surfaces in condensed-phase and biomolecular systems, such as molecular spectroscopies, system-wide chemical reactions, and electron and energy transfer processes.

**Acknowledgments:** We thank the National Institutes of Health (GM46376 and GM091445) for partially supporting this research.

**Supporting Information Available:** A table containing the PMO energies from variational and charge-embedding X-Pol and many-body expansion theory up to three-body terms for water clusters and periodic systems is provided, along with the Cartesian coordinate for a cluster of 65 water molecules. This information is available free of charge via the Internet at http://pubs.acs.org/.

**Biography**

**Jiali Gao** studied at Beijing University, followed by graduate education at Purdue and postdoctoral research at Harvard. He was on the faculty of the State University of New York at Buffalo, and he is currently a Professor of Chemistry with partial appointments at the University of Minnesota and Jilin University.

**Donald G. Truhlar** is the Regent Professor of Chemistry. He has been on the faculty of the University of Minnesota since 1972.

**Yingjie Wang** earned a B.S. in chemistry (2009) from Nanjing University. He is currently a fifth-year graduate student of Chemical Physics at University of Minnesota.

**Michael J. M. Mazack** received a B.S. in mathematics with a physics minor from Western Washington University in 2007. He obtained an M.S. in mathematics from the same institution in 2009, studying numerical linear algebra under Professor Tjalling
Ypma. He completed a Ph.D. in scientific computation at the University of Minnesota in 2014 under Professor Jiali Gao, researching the explicit polarization theory and developing coarse-grained models for proteins.

Patrick Löffler earned a B.S. in biology (2009) from the University of Erlangen-Nuremberg and a M.S. in bioinformatics (2012) from the University of Hamburg. After having spent time as a Visiting Researcher at the University of Minnesota, he is currently a Ph.D. student at the University of Regensburg.

Makenzie Provorse received a B.S. in Chemistry from Kansas State University in 2009 and a M.S. and Ph.D. in Chemistry from the University of Minnesota under the guidance of Prof. Jiali Gao in 2010 and 2014, respectively. She is currently a postdoctoral associate with Assistant Prof. Christine Isborne at the University of California at Merced.

Pavel L. Rehak earned a BS from SUNY Stony Brook in 2010, double majoring in Chemistry and Mathematics. He is currently a graduate student under the supervision of Dr. Jiali Gao

References.


