TOWARDS MORE EFFICIENT *AB INITIO* COMPUTATION OF PHYSICAL PROPERTIES

A Dissertation
Presented to
The Academic Faculty

By

Michael Zott

In Partial Fulfillment
of the Requirements for the Degree
Bachelors of Science with the Research Option in the
School of Chemistry and Biochemistry

Georgia Institute of Technology

May 2018

TOWARDS MORE EFFICIENT *AB INITIO* COMPUTATION OF PHYSICAL PROPERTIES

Approved by:

Dr. Charles David Sherrill, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. E. Kent Barefield
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: May 1, 2018

Nothing contributes so much to tranquilize the mind as a stedy purpose — a point on which the soul can focus its intellectual eye.

Mary Shelley

This thesis is dedicated to my parents Elizabeth and Richard Zott who have given me the confidence as well as the resources to help pursue my passion in chemistry.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Prof. Charles David Sherrill for trusting me enough to allow me to work with him starting my very first day at Georgia Tech. He paired me with excellent mentors in postdoctoral scholars Drs. Ryan Richard and Daniel Smith who helped me experience a breadth of computational research across both quantum chemistry and molecular mechanics. In addition to providing me with these mentors, Dr. Sherrill was also a great proponent in getting me to present my research. I was able to present at five conferences, including presenting a poster at regional ACS meeting as well as being an invited speaker at the Psi4 Conference. Without his support, I would not have been able to share my research in such diverse settings.

In addition to this mentorship in the Sherrill group, several other faculty at Georgia Tech helped enrich my academic and personal education. Extensive conversations with Prof. Gary Schuster helped me to become more academically diverse and consider new scientific avenues. Likewise Prof. E. Kent Barefield helped me learn to enjoy experimental chemistry and consider the electronic structure of inorganic systems, a path I hope to follow in my graduate career. Prof. Charles Liotta introduced me to spiroconjugation, an interaction owing to perpendicular yet phase aligned orbitals in π-systems, also serving to broaden my knowledge of electronic structure and the computational procedures for radical systems.

Outside of faculty members, several graduate students in the Sherrill lab as well as in teaching labs had great impacts on my education. In the Sherrill group, Brandon Bakr, Dominic Sirianni, and Trent Parker were immensely helpful in understanding the theoretical and practical aspects of computational chemistry. Teaching assistants for several of my courses such as Abraham Jordan, Brandon Yik, and Sam Evans gave me confidence in my chemical abilities in addition to imparting vast knowledge regarding practical lab skills.

Finally, my parents Richard and Elizabeth Zott as well as my brother Ricky were constantly supportive of my studies and critical in giving me the confidence to pursue my

academic goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The introduction of the modern computer has been a boon to myriad scientific communities. Scientific experiment can be categorized into the categories of physical experiment and thought experiment. In the chemical arena, these thought experiments are now able to be tested for validity through advanced semi-empirical and *ab initio* computational methods. Theoretical chemistry continues to increase in efficacy, and the spread of classical, wavefunction, and density functional methods into experimental communities is now undeniable. An aspiration of computational chemistry is to provide predictive power to lower the number of physical experiments that need to be performed. This is especially important when systems arise that are difficult to study experimentally. This has the possibility to lower financial and environmental costs, in addition to reducing the time needed to perform physical experiments.

Here, methods to computationally study solvent effects and crystal lattice energies are reported on. Both of these physical properties have substantial relevance to human-focused enterprises such as targeted drug design. For example, drugs are often delivered in solid, crystalline form and must dissolve into molecular form prior to being pharmaceutically active. Although the specific research reported on here does not use systems directly related to such applications, it is posited that fundamental advances in computational methods for computing physical properties for arbitrary systems will contribute to solving problems in drug design, material development, and biomolecule recognition.

CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Computational Theory

The fundamental tool in computational research is the *theory*. Herein, both classical and quantum mechanical theories are studied and developed. Although the rigorous theoretical frameworks for both classical and quantum mechanics have been established for centuries and decades, respectively, there is still active development in the application and implementation of these frameworks to many particle, chemically relevant systems. Brief descriptions of the frameworks of MM and QM follow. The overarching goal of both methods is to provide quantities that can be compared to experimental results for predictive or descriptive power. Examples of such quantities include the positions of atoms in a system, the dipole moment of a molecule, or the permeability of a membrane.

1.1.1 Molecular Mechanics

In order to calculate properties from the principles of molecular mechanics, Newtonian physics is used. In MM, a potential energy function is defined (Equation 1.1) whereby a potential energy surface can be generated; the terms, in order, correspond to stretching, bending, torsion, Van der Waals, electrostatic, and "cross" forces.

$$U = E_{str} + E_{bend} + E_{tors} + E_{VDW} + E_{el} + E_{cross} \tag{1.1}$$

Each of these terms is computed according to Newton's Laws. For example, E_{str} is computed using the Taylor series for displacement about an equilibrium point (*i.e.* Hooke's law when truncated to second order): $E_{str} = k_{ab}(r_{ab} - r_{0,ab})^2$, where k_{ab} is a constant and r_{ab} is the distance between particles at points a and b with r_0 corresponding to the equilib-

rium interparticle distance. The usage of these parameters k_{ab} and $r_{0,ab}$ is a quintessential aspect of MM. In MM jargon, the set of all parameters such as k_{ab} is known as a force field (FF).

Molecular mechanics operates under the assumption that an arbitrary chemical system can be modeled by analogy to a more general system for which these parameters are already known. Depending on the granularity of the FF, any particular chemical motif (*e.g.* functional group) may have one or several parameters. For example, there may be only one parameter for a C=O stretch (k_{CO}) which would mean that the C=O stretch force constant is equivalent in aldehydes, ketones, carboxylic acids, *etc.*, or there could be individual parameters for each of these different functional groups ($k_{CO,\ ald}$, $k_{CO,\ ket}$, $k_{CO,\ carb\ acid}$). This granularity is defined by the number of atom types available for any given element. In order to run a MM simulation, then, the atom types for each atom in the system must be known such that parameters can for constants such as k_{ab} can be assigned to calculate the force at each particle.

Using the equation for the potential energy for our MM system (Equation 1.1), the forces on every particle can be computed (note the introduction of bold variables to symbolize vectors):

$$\mathbf{F} = -\nabla U. \tag{1.2}$$

From Newton's second law, $\mathbf{F} = \mathbf{m}a$, the acceleration of each of these particles is given by the force impinging on this particle divided by that particle's mass. Once the acceleration is known, the trajectory of the particle can be computed. Additional corrections can be added to modify the trajectory to account for conditions such as temperature and pressure, but theoretical details on the implementation of such aspects of MM simulations is outside of this work's scope.

An aspect of MM critical to note is that all of the parameters relevant to the system encoded in the FF are known prior to the beginning of the simulation. Therefore, the com-

putational cost in performing MM simulations is in integrating the equations of motion. This allows very large systems to be studied by MM, where very large in the computational arena means over one million individual atoms. With systems of this size, dynamic properties are able to be visualized such as water molecules flowing through a membrane channel or a protein folding into its secondary structure from an unfolded state. Additionally, statistical behavior of a system can be detected due to the modeling of the effect of thermal energy on a system.

1.1.2 Quantum Mechanics

Calculating properties according to the laws of quantum mechanics is another important area of computational chemistry. Quantum chemistry includes many levels of theory, but here only wavefunction and density functional (DFT) methods are described.

As in molecular mechanics, a key focus of QM is the calculation of energies. However, the focus described here is the calculation of *electronic* energies, the energies owing to the particular configuration of electrons in a system. In MM, the electronic energy is accounted for through the force field's parameters, but in QM the energy is computed from first principles (*ab initio*). The equation governing energy in QM is Schrödinger's equation,

$$\mathcal{H}\psi = \mathcal{E}\psi, \tag{1.3}$$

where $\mathcal{H}$ is the Hamiltonian, ψ is the wavefunction, and $\mathcal{E}$ is the energy. The Hamiltonian is described in matrix mechanics by a matrix, thus this equation defines an eigenvalue problem. In QM, the computational difficulty lies in finding the solution to this equation. Depending on the exact method employed (wavefunction versus density functional), the form of the electronic Hamiltonian $\mathcal{H}$ will change. For wavefunction methods (*i.e.* Hartree-Fock, Möller-Plesset perturbation theory, coupled cluster theory, *etc.*), the Hamiltonian is based on a potential energy function dependent on two particle interactions through the $\frac{1}{r_{ab}}$ term. In density functional methods, this potential energy term is replaced by a term that

is a function of electron density. This allows larger systems to be studied by DFT than may be studied by wavefunction methods as all operators in the electronic Hamiltonian can be represented by one electron operators unlike the two electron electrostatic potential operator in Hartree-Fock methods.

Unlike MM where all the parameters governing a system are known before the calculation begins, the QM methods described here require all of the terms (operators) in the Hamiltonian to be computed in order to find the electronic energy. This means that QM calculations require more computations to be performed than MM calculations, and for identical systems, a lesser number of atoms can be included in QM calculations. Generally, the number of atoms in a QM simulation are on the tens to hundreds scale. Additionally, the trajectory of particles is more commonly computed for MM simulations than for QM simulations, and dynamic properties such as vibrational frequencies require more effort to compute.

1.2 Computational Chemistry Software

In computational chemistry, both commercial and open source software packages exist. Here, all computations are performed using free and open source software. The molecular mechanics program OpenMM[1] is used, and it is compatible with many force fields. Additionally, it has several benefits such as allowing arbitrary potential energy functions to be added to the potential energy function (Equation 1.1), offering several types of integrators for the equations of motion, and having well written, graphics processor unit (GPU) optimized code. For QM calculations, the Psi4 package[2] was used. Psi4 has a large number of density-fitted methods which speed up computation, and the frozen natural orbit coupled cluster with single, double, and perturbative triple excitations (FNO-CCSD(T)) is very fast as well.

From a software development perspective, both of these packages have efficient codebases written in C++ as well as easy to use interfaces written in Python. Being written

in the same programming languages enables easier passing of data between these softwares. Although not formally computational chemistry software, there are many useful and well-funded modules written for Python that are able to be used in harmony with Psi4 and OpenMM such as SciPy, NumPy, and NetworkX.

CHAPTER 2

PARALLEL MM AND QM SIMULATIONS

From even a cursory consideration of computational theory, it is evident that large systems are the domain of MM, and small systems can be treated through MM or QM simulations. When QM theory can be applied to a system, it almost invariably yields superior results to MM. For many years, the goal of applying QM and MM simultaneously (QM/MM) has been pursued, and this methodology is implemented in several software packages. Here, a new approach to tackling unified QM and MM simulations is studied in the context of modeling explicit solvation.

2.1 Theory and Methodology

In QM/MM, the principal difficulty lies in accommodating the boundary between the components of the system treated quantum mechanically versus classically. An example relevant to the research discussed here would be modeling a benzene molecule in an aqueous solution. Chemical intuition would indicate that for understanding solvation of the nonpolar benzene molecule in a polar aqueous solution, the interaction between water molecules in the solvation sphere and the benzene molecule should be treated quantum mechanically whereas water molecules outside of the solvation sphere could be treated classically through an atomistic simulation (explicit solvent) or perhaps even using a solvent model such as the polarizable continuum model (PCM).[3] However, description of where the boundary between quantum and classical mechanics should be applied in the system is not always clear.

Here, an alternative to the traditional QM/MM scheme is proposed and tested. The accuracy of single point energy calculations from quantum mechanics and the ability to perform rapid dynamics simulations of larger system from molecular mechanics are both

maintained. However, to avoid the issue of embedding the QM system into the MM system as in traditional QM/MM, the simulations are performed separately from one another. Dynamics simulations are performed in OpenMM in order to generate fluctuations in the structure, and single point energies (here, DFT, but any level of theory is applicable) are computed along this trajectory of structures using Psi4. By calculating large numbers of single point energies for different conformations along this trajectory, thermodynamic properties should be able to be computed using statistical mechanics. For example, the electronic enthalpy should be able to be computed by taking the Boltzmann averaged energy of the electronic energies of these conformations' single point QM energies.

In order to carry out this new approach, first the issue of sharing information between separate programs (Psi4 and OpenMM) needed to be dealt with. While passing information between these programs at the broadest level is facile due to both programs having wrappers written in Python, the issue that needs to be confronted is that the information required by a MM simulation is markedly different from that required by a QM simulation.

As introduced in Section 1.1, the potential energy function in MM simulations relies on knowing parameters for the individual energy terms in Equation 1.1. For known systems, molecular mechanics programs often can assign parameters based on the structure as read in from a Protein Data Bank (PDB) file. For example, this works well for proteins, nucleic acids, lipids, and some small molecules such as guanidine or water. However, quantum mechanics simulations only need the Cartesian coordinates and atomic numbers of involved atoms in order to run a simulation. Thus, when starting from a QM calculation, all of these parameters for the MM simulation need to be identified and assigned. In MM, there are several force fields which all define different rules for assigning atom types. Some of these force fields have software which enables automatic assignment of atom types to a system based on only the geometry of the system. However, at the time of initiation of this project, there was no unified software that could assign atom types for multiple different force fields (this software now exists independently of this project, thus work in this direction

has been stopped). Existing atom type assignment programs can be found for the General Amber Force Field (GAFF),[4] the CHARMM General Force Field (CGenFF),[5] and the Optimized Potentials for Liquid Simulations (OPLS) force fields.[6]

The goal of this project was to create a program to identify functional groups and larger chemical motifs (*e.g.* aminos acids) in order to be able to assign atom types to any arbitrary system for any arbitrary force field. Once the functional groups were identified for every atom in a system, then a mapping between functional group and atom type in any force field would allow assignment of atom types. The difficulty with implementing this algorithm is that each force field has a different granularity; for some force fields, there may be 3 carbon atom types for carbonyl carbons, and for others there may be 8 atom types for this type of carbon.

2.2 Implementation of Atom Typing Algorithm

In order to address this problem, molecular systems were represented as mathematical graphs where each atom in a molecule is represented by a node and each bond is represented by an edge. However, the notion of what constitutes a chemical bond is also variable depending on the force field. This is due to the difficulty of finding bond order from data that only includes interatomic distances. Various schemes exist to calculate bond order in this situation, and different force field atom type identification algorithms adopt different standards. Once bond orders are identified, properties such as the number of atoms connected to each atom, the existence of conjugation at this atom, or the presence of a cyclic arrangement of atoms including this atom can be computed. From these properties, functional groups can be identified. Upon identification, the nodes (atoms) involved in this functional group can be assigned to a new, single node corresponding to the functional group. This reduces the size of the graph. Iterations of this strategy, forming larger and larger functional groups, leads to assignment of the most specific functional group possible to each atom and enables atom types to be assigned to individual atoms.

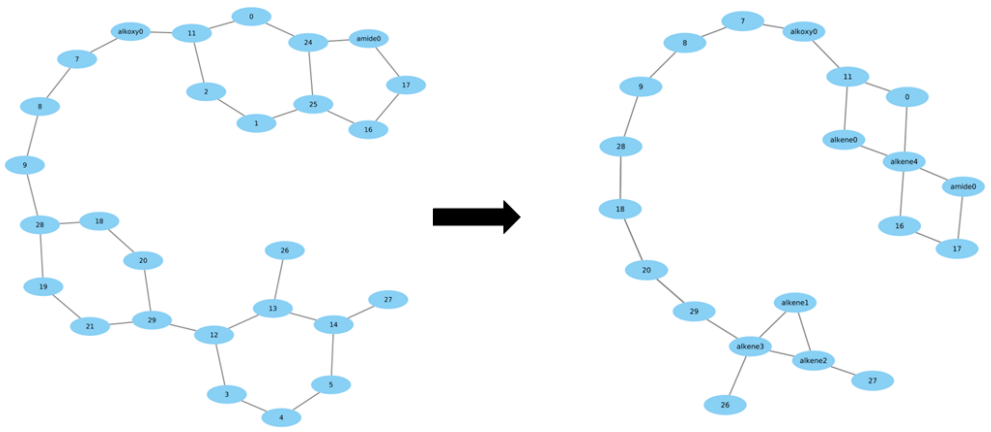


Figure 2.1: An example of the graph reduction strategy reducing phenyl, alkenyl, and amide groups into reduced notes. The reduction in size and complexity of the graphs is visually apparent.

Table 2.1: The percentage of atom types equal to those assigned by Antechamber is shown along with the percentage of molecules where every single atom type was assigned equivalently. The Rx200 database has 200 pharmaceutical molecules, and the IQmol database has approximately 250 general small molecules covering all of the standard functional groups.

	Rx200	**IQmol Database**
Atom Types Correct (%)	>90	>90
Molecules Correct (%)	≈25	≈40

To simplify the first version of this atom type assignment program, the GAFF was targeted. Results of this graph reduction strategy are seen in Figure 2.1. Upon visual inspection, it is clear that the graph has been reduced in size and that functional groups are now identified. In the translation to atom types from functional groups, accuracy is assessed by comparison to Antechamber,[7] the GAFF atom type assignment program, in Table 2.1.

Upon testing our new implementation with the reference Antechamber implementation, it is clear that there are discrepancies especially for the percentage of molecules that have all of their atom types assigned correctly (Table 2.1). The importance of assigning all atom types correctly is that occasionally if an incorrect atom type is assigned, multi-atom type properties will become grossly erroneous. For example, the harmonic angle bend term (E_{bend} in Equation 1.1) relies on three separate atom type parameters. The force field does

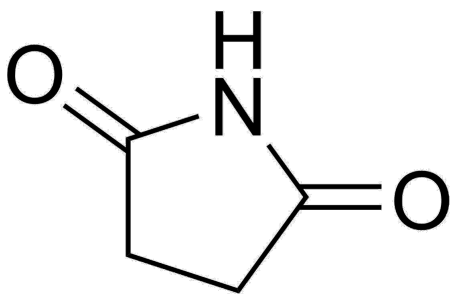


Figure 2.2: Pyrrolidine-1,4-diketone.

not typically provide parameters for every possible combination of three atom types because they are typically not all chemically relevant. However, in one instance an incorrect atom type in the chloroform molecule led to a non-tetrahedral structure because there was no restriction on the angles between chlorine atoms due to missing harmonic bend parameters. This leads to extreme errors in the subsequent calculations. The percentage of atom types correct for both Rx200 and IQmol is very promising, though. In some cases, achieving a perfect 100% is ostensibly not desirable. For example, the Antechamber program assigns the atom type of ha, or a hydrogen on an aromatic carbon, for the hydrogen atoms in the ethylene molecule. It is unlikely that any chemist would ever describe the hydrogen atoms in ethylene to be on aromatic carbons. Our implementation assigns the hc atom type which is hydrogen on a general carbon atom. In this case, choices made by the Antechamber program can occasionally seem arbitrary or even wrong. Due to the degeneracy of potential atom type choices, occasionally a choice does need to be made to decide between two atom types that both fit the chemistry of the molecule. In this case, it is very difficult to match with the Antechamber program because there seems to be no perfect metric for making decisions in instances such as this.

At present, the SMIRNOFF program[8] has been released by another group based on the same idea of assigning atom types for any arbitrary force field. This program has the advantage over my implementation of both funding and number — development of Smirnoff includes several developers and is supported by a multi-site NSF grant. Thus, for

atom type assignment in simulation workflows, programs other than that developed my me are currently used.

2.3 QM-MM Calculations Through the Psi4-OpenMM Interface

After solving the problem of assigning atom types to arbitrary molecules with only structural information available, the problem of transferring data from the MM and QM programs to one another was next in line. Apart from atom types, the other information required by an MM program to perform a simulation is the connectivity of the atoms (which atoms are bonded to each other — this is solved during the atom type assignment procedure) and the partial charges on each atom. The assignment of partial charges can be done through several charge partioning schemes, but each force field maintains its own recommendations on how charges are to be computed. Two common approaches are through the Restrained Electrostatic Potential (RESP)[9] scheme or the AM1-BCC scheme.[10] The RESP protocol is implemented in Psi4, and RESP charges are typically of higher accuracy than those generated by AM1-BCC.

After generation of all of these missing descriptors, data can be passed between MM and QM programs. Here, an oepn source interface for passing data between OpenMM and Psi4 was developed. The interface allows simple inputs from the user to access all of the functionality of both OpenMM and Psi4. A MM calculation to find the trajectory of a system can be performed in OpenMM, and the structures found can be directly transferred to Psi4 for calculation of quantum mechanical energies and properties.

2.4 Studying Solvent Effects in Shimizu Torsion Balances

To demonstrate an application of performing these parallel, independent QM-MM calculations and develop new methodology, a torsion balance system (Figure 2.3) synthesized by the Shimizu group to study interactions between C-H bonds and arene π electron density was targeted.[11] This system has previously been studied in the Sherrill group by Trent

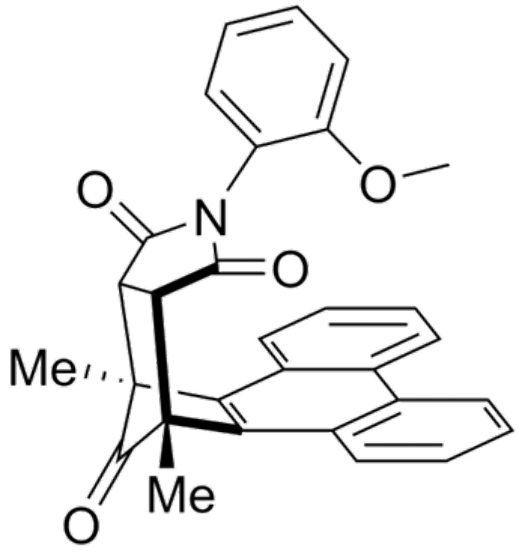


Figure 2.3: One of the torsion balance molecules studied due to its solvent dependent rotamer equilibrium.

Parker without explicit solvent molecules in the calculations.[12] Study of this system in relation to solvent effects is motivated by the fact that quantum mechanical calculations for the difference in electronic energy between the open and closed rotamer conformations (Figure 2.5). From NMR studies of the equilibrium population of each rotamer, the Gibbs Free Energy was able to be calculated as +0.31 kcal/mol favoring the open conformation. When solvent effects are ignored and gas phase single point B3LYP-D3 calculations with the aug-cc-pVDZ basis set are performed on the open and closed conformations, an enthalpic difference of -0.56 kcal/mol is calculated, favoring the closed conformation. Therefore, entropic effects must be at play in favoring the open conformation.

One contributor to entropy in the solvated system (Figure 2.4) is the increased freedom of motion of the methoxy substituent as well as the increased number of solvent binding sites in the open conformation. Through performing a large number of calculations on various torsion balance-solvent (chloroform) interactions, the entropy can be predicted based on statistical mechanics principles. However, not enough computations have been performed yet in order to estimate the entropy of this system. However, identification of important torsion balance-chloroform interactions did occur. In Figure 2.6, the conforma-

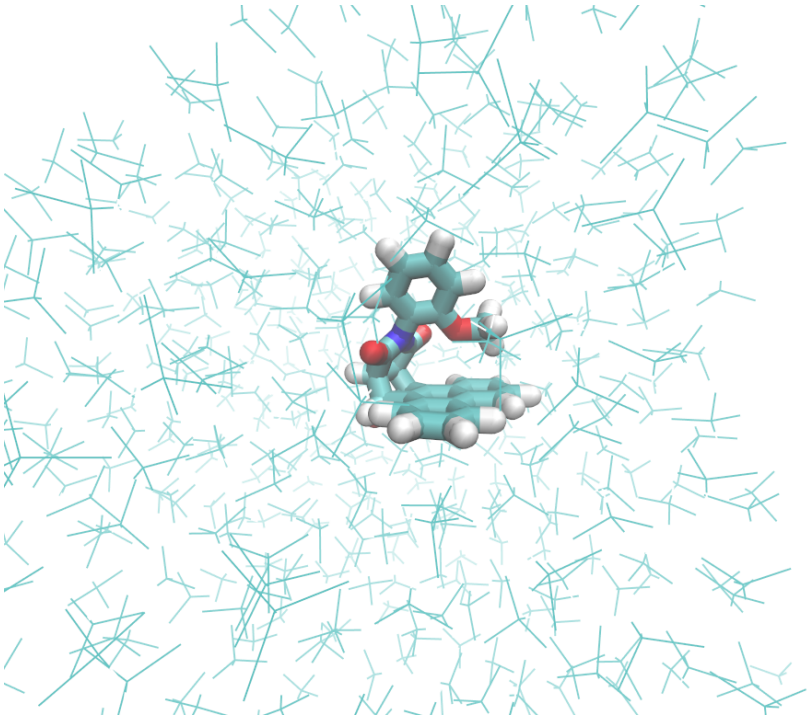


Figure 2.4: Example of the automatic addition of solvent (here, chloroform) to a simulation.

tions identified by molecular mechanics trajectories with the lowest quantum mechanical B3LYP-D3 electronic energy are visualized. In these systems, the Boltzmann averaged difference in electronic energy is -1.58 kcal/mol. This leads to prediction of preference for the closed rotamer, contrary to experiment. Thus, QM computations with a larger number of chloroform molecules are needed. QM computations with up to 5 chloroform molecules included were calculated, but no improvement in the predicted population of each rotamer state was seen. However, Boltzmann averaging of the electronic energy of each rotamer state for the torsion balance without any chloroforms present in the QM single point energies led to a small shift towards predicting the correct equilibrium populations, with the difference in electronic energy being -0.54 kcal/mol (Figure 2.7) versus -0.56 kcal/mol difference seen above for only the minimum energy conformers. Clearly, more study of solvent effects as well as further development of this methodology to predict entropy values is needed.

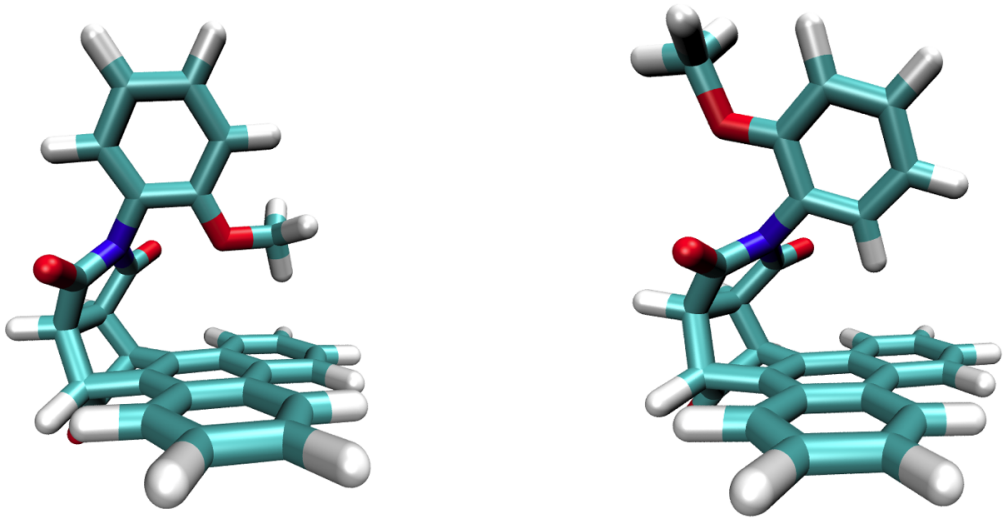


Figure 2.5: Three dimensional views of the closed (left) and open (right) conformations of a Shimizu torsion balance molecule.

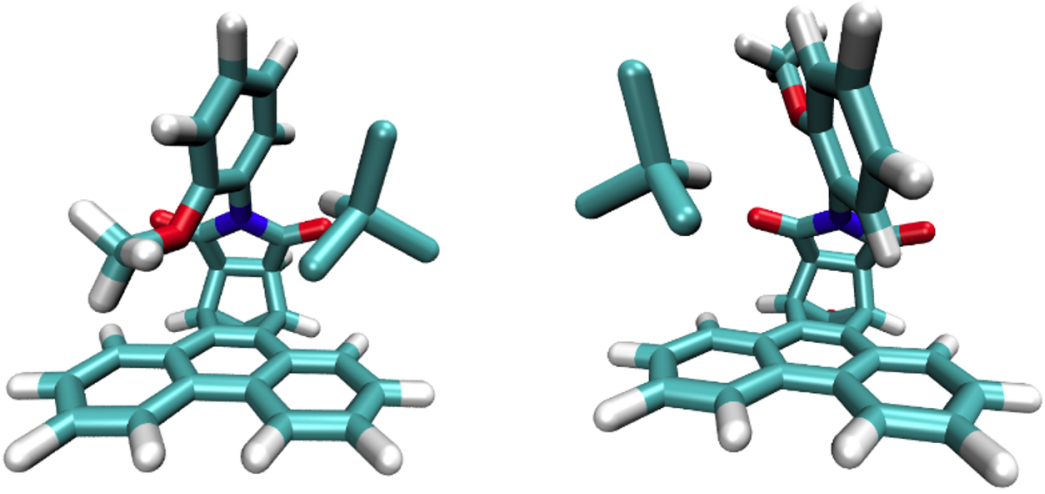


Figure 2.6: Important interactions between chloroform and the torsion balance identified by QM-MM methods. These interactions represent the torsion balance-chloroform interaction with the strongest binding energy as computed by B3LYP-D3 with the aug-cc-pVDZ basis. Both interactions represent a balance between chloroform C-H to arene π system and chloroform C-H to carbonyl oxygen bonding.

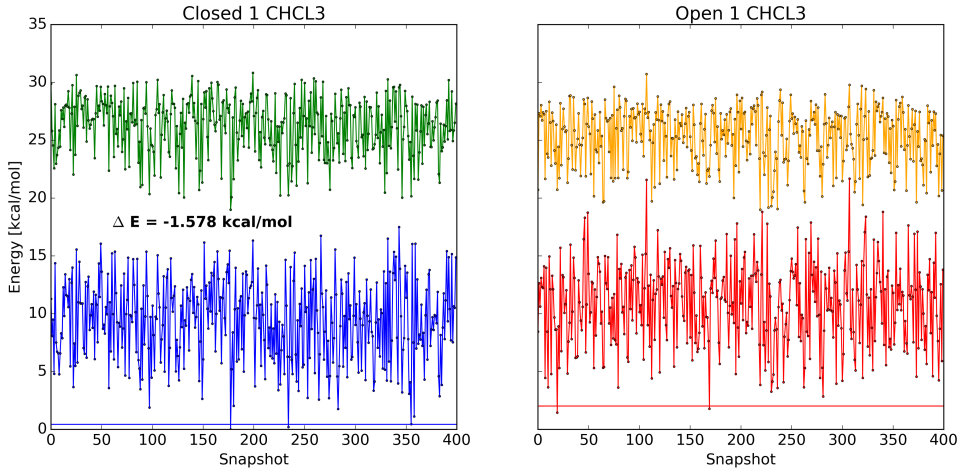


Figure 2.7: An example of data generated from MM-QM calculations. 400 MM and QM calculations are reported in each plot; the horizontal lines represent the Boltzmann averaged energy of each system. The systems under study here is the Shimizu torsion balance shown above. It is studied in explicit microsolvation using chloroform. The energies in these plots are for the torsion balance alone, although the exact configurations are influenced by interactions with solvent. The blue and red traces are QM energies, and the green and orange traces are MM energies (shifted arbitrarily).

2.5 Conclusion

A new interface between the open source software programs Psi4 and OpenMM has been made. As part of this development, the undertaking of developing a new atom typing was begun. However, a NSF funded project with similar atom typing goals has supplanted this effort.

Using this new interface, calculations on a molecular torsion balance have also been started in order to identify important chloroform-torsion balance interactions that may favor one rotamer over another.These preliminary calculations have succeeded in identifying several strongly favorable chloroform-torsion balance interactions. If this methodology allows the relative populations of rotamers to be predicted accurately with reference to experimental values, this new QM-MM approach will be useful for studying many other molecular systems with explicit solvent. However, better development of this method to compute entropy values is needed.

CHAPTER 3

BENCHMARKING OF THE MANY BODY EXPANSION FOR REDUCING CRYSTAL LATTICE ENERGY COMPUTATIONAL COST

Finding the crystal lattice energy of solid materials is an important application of computational chemistry. In the application of drug delivery, drug molecules are often delivered as solid, crystalline material. Inside the target organism, this crystalline material must dissolve prior to becoming pharmaceutically active. The lattice energy of the crystal is a key parameter in determining the kinetics of dissolution.

As unit cells in crystal structures often contain a large number of atoms relative to standard quantum chemistry system sizes, the method chosen to compute lattice energies must be highly efficient such that the computation is tractable. One method that may enable reduced computational cost is the many body expansion. The many body expansion represents the interaction energy of a system by analyzing the 1, 2, ..., $\mathcal{M}$-body contributions to the interaction energy of the total system. By finding the order m with $m < \mathcal{M}$ where the interaction energy of the system stabilizes to the interaction energy computed accounting for all $\mathcal{M}$-body interactions, calculations of the total interaction energy and thus lattice energy can be accelerated. This is especially true as m is typically much less than $\mathcal{M}$; for a system of 100 molecules, it is unlikely that concerted interactions between more than 3 individual molecules contribute strongly to the system's interaction energy based on chemical intuition. Here, a particular formalism of the many body expansion, the Valiron-Meyer Functional Counterpoise correction (VMFC) is tested in order to find the general order m where the interaction energy converges to the $\mathcal{M}$-body interaction energy. Previous studies have assumed an order m equal to 3, and some studies have investigated this quantitatively with trimer systems. Here, 14 pentamer geometries (Figure 3.1) are studied; further, these pentamer geometries cover the fundamental intermolecular forces: dispersion, electrostat-

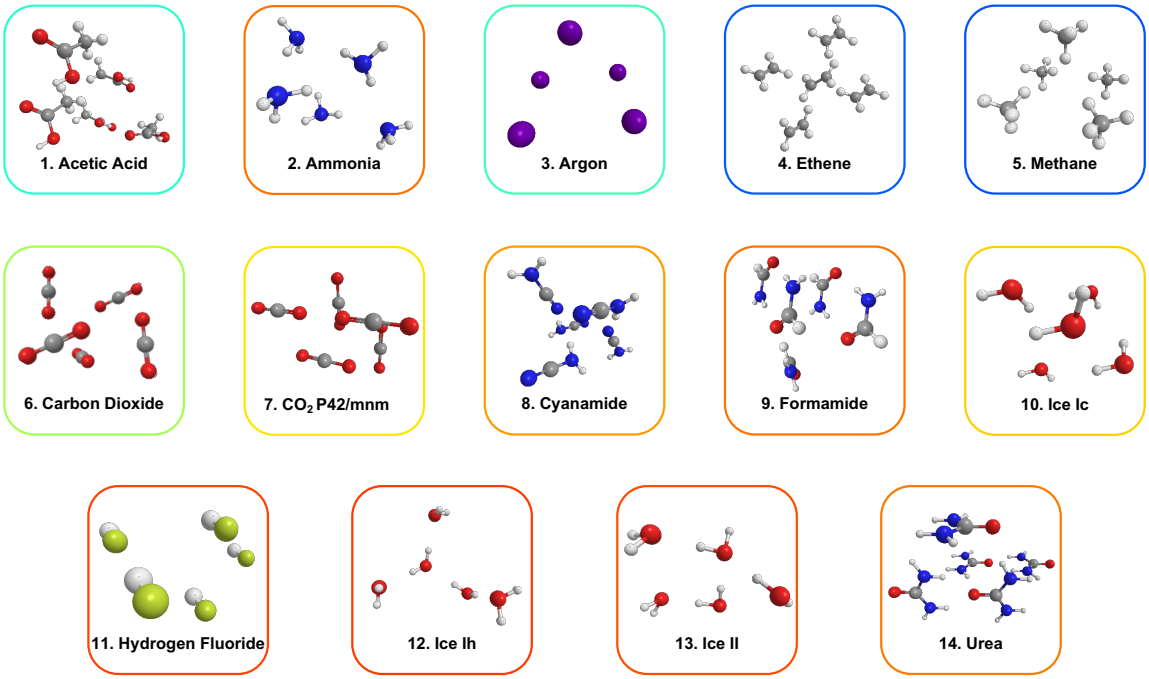


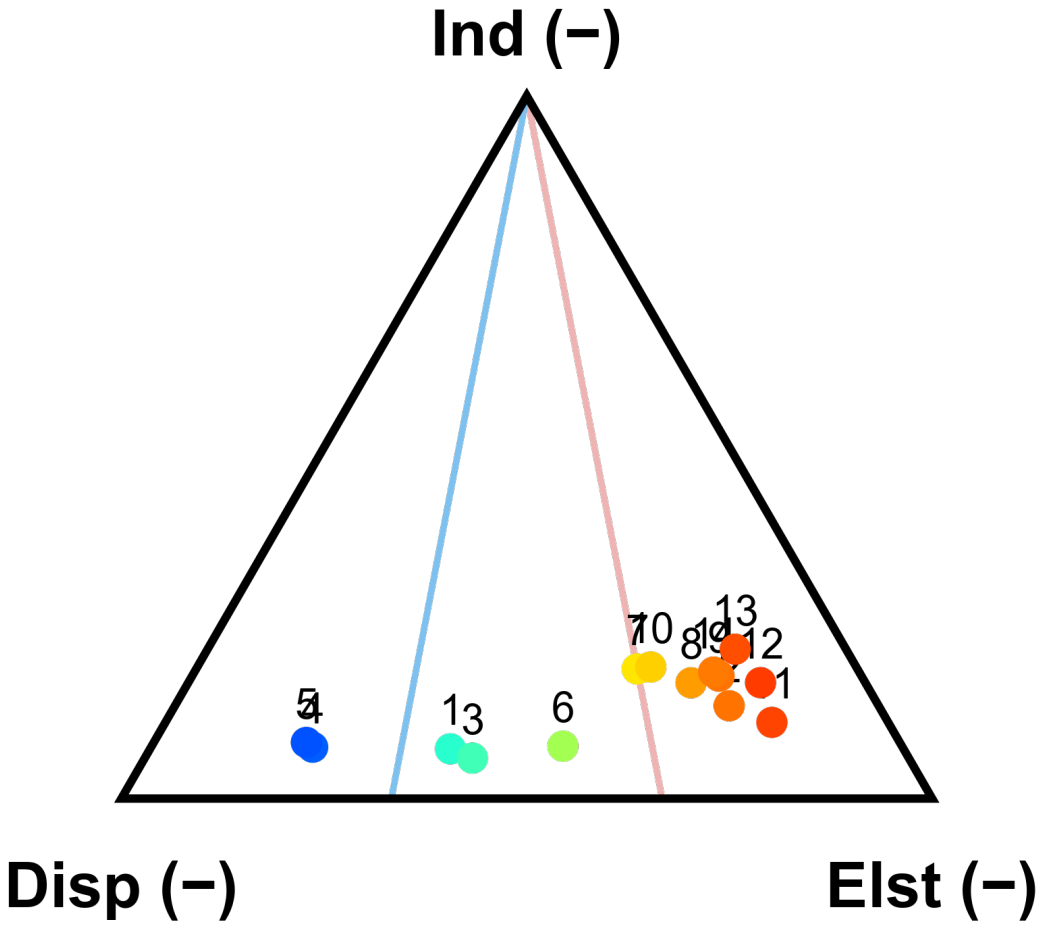
Figure 3.1: The 14 pentamer systems under study; the labeling of each system from 1–14 is propagated throughout the figures in this chapter. Additionally, the color scheme is maintained. The color scheme is based on the primary intermolecular interaction in the system according to SAPT0 (see Figure 3.2) — cooler colors are for systems dominated by dispersion forces, and warmer colors for those dominated by electrostatics.

ics, induction. Gold standard benchmark calculations using density fitted frozen natural orbital coupled cluster with single, double, and perturbative triple excitation (DF-FNO-CCSD(T)) calculations are presented in addition to density functional theory calculations (DFT) for comparison.

3.1 Theory

To introduce the idea of the many body expansion, first the notion of an "interaction energy" must be introduced. The simplest example of an intermolecular interaction is the interaction between two molecules — a dimeric interaction (note that the language of dimer, trimer, m-mer will be used here to refer to 2, 3, m interacting molecules with no constraint on these molecules being equivalent). To compute the energy of this intermolecular interaction, the energies of the dimer system as well as each monomer will be needed. Then, the interaction

Figure 3.2: SAPT0/jun-cc-pvdz breakdown of the induction (Ind), electrostatics (Elst), and dispersion (Disp) components of the intermolecular interactions in the pentamer systems. Numbering is from Figure 3.1. Blue and magenta lines demarcate dispersion, mixed, and electrostatics dominant systems. The "monomers" making up the "dimer" required for the SAPT0 are a single molecule from the pentamer as one "monomer" and the other four molecules as the other "monomer."



energy ΔE_{IJ} can be defined for the dimer IJ composed of monomers I and J,

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J, \tag{3.1}$$

where $E_{\mathcal{F}_k}$ is the energy of the fragments in the set $\mathcal{F}_k$ (here IJ, I, and J, respectively) where

$$E = \sum_{I=1}^{M} E_I + \sum_{I<J}^{{}_MC_2} \Delta E_{IJ} + \sum_{I<J<K}^{{}_MC_3} \Delta E_{IJK} + \cdots + \Delta E_{IJK\cdots M}, \tag{3.2a}$$

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K. \tag{3.2b}$$

For systems of size greater than 2, there exist different formalisms to account for the contribution to the interaction energy from $\mathcal{M}$-mer interactions where $m > 2$. One such formalism is the many body expansion (Equation 3.2a). The many body expansion is simply a generalization of the interaction energy for a dimer (Equation 3.1) where the trimer, tetramer, ..., $\mathcal{M}$-mer interaction energies are as represented in Equation 3.2b

$$\epsilon_{\mathcal{M}} = \sum_{k=1}^{M} \sum_{\mathcal{F}_k \in \mathcal{P}_k(\mathcal{M})}^{{}_MC_k} \Delta\mathcal{E}_{\mathcal{F}_k}. \tag{3.3a}$$

$$\Delta\mathcal{E}_{\mathcal{F}_k} = E_{\mathcal{F}_k}(\mathcal{F}_k) - \sum_{\ell=1}^{k-1} \sum_{\mathcal{F}_\ell \in \mathcal{P}_\ell(\mathcal{F}_k)}^{{}_kC_\ell} \Delta E_{\mathcal{F}_\ell}(\mathcal{F}_k) \tag{3.3b}$$

The shortcoming of the many body expansion in this form is that it does not correct for basis set superposition error (BSSE) when the m-body correction to the interaction energy is computed as in Equation 3.2b. For the extensive theoretical argument for why BSSE is not accounted for, refer to a paper preceding this project by Richard.[13] However, when the m-body correction is written as in Equation 3.3b, the interaction energy computed is BSSE free and referred to as the Valiron and Mayer functional counterpoise correction (VMFC).

Note the change in notation in Equation 3.3b — the ΔE terms in this equation now include a term in parentheses, $\mathcal{F}_k$, which corresponds to the set of k fragments contributing basis functions to the set of l fragments contributing nuclei and electrons. Note that the set k can be larger than the set l; the number of fragments k contributing basis functions is equal to the size of the m-body correction to the interaction energy being computed. This confers an advantage to the VMFC, however, as the maximum value of m necessary to compute the interaction energy of the system of $\mathcal{M}$ fragments is often much lower than $\mathcal{M}$, $m << \mathcal{M}$. Often computations on a system of M fragments require at least one computation to be performed where all $\mathcal{M}$ fragments are contributing basis functions; here, only m fragments contribute basis functions for the m-body correction to the interaction energy. This enables an enormous reduction to the size of the computation in exchange for an increase in the number of individual, smaller calculations. Therefore, for applications such as computing crystal lattice energies where there may be hundreds or more fragments, this approach is highly appealing. Providing evidence for what value of m, the truncation order, is suitable for replicating the total $\mathcal{M}$-body energy is thus crucial such that accurate usage of the truncated VMFC approach is feasible. These truncated VMFC energies will be referred to as VMFC(m), where m is the equal to the largest m-body correction to the interaction energy.

3.2 Methodology

All electronic structure computations have been performed using density-fitted (DF), frozen natural orbital (FNO) coupled-cluster with single, double, and perturbative triple excitations [DF-FNO-CCSD(T)][14, 15] as implemented in Psi4. The correlation consistent basis sets aug-cc-pVXZ with $X = D, T, Q, 5$ are used and abbreviated aXZ. Using Helgaker CBS extrapolation of the correlation energy[16] we expect our results to be of gold-standard quality. Here, the correlation energy often refers to either the MP2 interaction energy alone,

$$\delta^{MP2} = E_{total,MP2} - E_{total,SCF}, \tag{3.4}$$

or for DF-FNO-CCSD(T), *only* the coupled cluster contribution to the correlation energy,

$$\delta_{MP2}^{CCSD(T)} = E_{total,DF-FNO-CCSD(T)} - E_{total,MP2}. \tag{3.5}$$

The total DF-FNO-CCSD(T) correlation energy is the sum of the δ^{MP2} correlation energy (Equation 3.4) and the $\delta_{MP2}^{CCSD(T)}$ correlation energy (Equation 3.5).

Even with the cost savings of the DF and FNO approximations, DF-FNO-CCSD(T) is still too prohibitively expensive to be applied to systems much larger than the pentamers considered here. In an effort to test the convergence of the MBE for density functional methods as well, we will also consider the accuracy of several functionals, namely B3LYP[17, 18], B2PLYP[19], M05-2X[20], ωPBE, ωB97X-D[21]. B3LYP, B2PLYP, and ωPBE use Grimme's D3 dispersion correction[22]; B2PLYP additionally employees Becke-Johnson dampening[23]. These DFTs have been selected based on previous calibration studies by our group.[24] Geometries as well as details pertaining to how the geometries were obtained are available in the appendix.

In addition to these levels of theory chosen for benchmarking analysis, symmetry adapted perturbation theory (SAPT) was also employed without intramonomer electron correlation (SAPT0) with the *jun-cc-pVDZ* basis set. SAPT allows the the components of the interaction energy — electrostatics, dispersion, induction, and exchange repulsion — to be calculated. The usage here is to verify that the systems chosen for study cover a diverse range of interaction motifs. The truncation of the fluctuation potential to the zeroth order (*i.e.* the usage of SAPT0) is not expected to change the qualitative conclusions about which

Table 3.1: Benchmark interaction energies (kcal mol^{-1}) at the SCF/a5Z + δ^{MP2}/a[Q, 5]Z + $\delta_{MP2}^{CCSD(T)}$/aXZ level of theory. Non-additive m-body contributions to the interaction energy are computed using the VMFC procedure, and their sum is the overall VMFC(5) interaction energy.

ID	**X**	$m = 2$	$m = 3$	$m = 4$	$m = 5$	**VMFC(5) IE**
1	D	-14.039 $\pm$ 0.697	0.119 $\pm$ 0.024	-0.013 $\pm$ 0.018	0.0 $\pm$ 0.007	-13.933 $\pm$ 0.745
2	Q	-13.003 $\pm$ 0.145	-0.284 $\pm$ 0.019	0.015 $\pm$ 0.012	-0.001 $\pm$ 0.006	-13.272 $\pm$ 0.182
3	Q	-2.285 $\pm$ 0.216	0.056 $\pm$ 0.004	-0.005 $\pm$ 0.001	0.002 $\pm$ 0.001	-2.231 $\pm$ 0.222
4	D	-4.487 $\pm$ 0.092	0.091 $\pm$ 0.003	0.001 $\pm$ 0.001	-0.0 $\pm$ 0.003	-4.396 $\pm$ 0.1
5	Q	-2.484 $\pm$ 0.048	0.057 $\pm$ 0.015	-0.002 $\pm$ 0.005	-0.0 $\pm$ 0.002	-2.429 $\pm$ 0.07
6	T	-8.46 $\pm$ 0.238	0.01 $\pm$ 0.007	0.007 $\pm$ 0.005	-0.0 $\pm$ 0.0	-8.443 $\pm$ 0.25
7	T	27.218 $\pm$ 1.599	-2.149 $\pm$ 0.036	0.055 $\pm$ 0.017	0.003 $\pm$ 0.006	25.127 $\pm$ 1.659
8	T	-26.773 $\pm$ 0.341	0.625 $\pm$ 0.025	-0.219 $\pm$ 0.001	0.012 $\pm$ 0.003	-26.355 $\pm$ 0.369
9	D	-36.807 $\pm$ 0.835	0.902 $\pm$ 0.026	-0.082 $\pm$ 0.018	-0.001 $\pm$ 0.006	-35.987 $\pm$ 0.886
10	Q	-0.006 $\pm$ 0.113	0.192 $\pm$ 0.023	0.003 $\pm$ 0.001	0.002 $\pm$ 0.001	0.191 $\pm$ 0.139
11	Q	2.344 $\pm$ 0.115	0.19 $\pm$ 0.035	0.009 $\pm$ 0.019	-0.007 $\pm$ 0.001	2.536 $\pm$ 0.17
12	Q	-13.477 $\pm$ 0.206	0.132 $\pm$ 0.013	0.074 $\pm$ 0.018	-0.004 $\pm$ 0.008	-13.275 $\pm$ 0.246
13	Q	-17.234 $\pm$ 0.337	0.014 $\pm$ 0.009	0.328 $\pm$ 0.001	-0.001 $\pm$ 0.001	-16.892 $\pm$ 0.348
14	D	-40.341 $\pm$ 0.863	-2.339 $\pm$ 0.024	0.102 $\pm$ 0.018	0.001 $\pm$ 0.006	-42.577 $\pm$ 0.911

component of the IE is dominant.

3.3 Results and Discussion

In order to compute benchmark values for the pentamers listed in Figure 3.1, DF-FNO-CCSD(T) was used to compute the energies of the VMFC. Additionally, extrapolations to the complete basis set (CBS) limit were performed for the SCF, MP2, and coupled cluster components of the DF-FNO-CCSD(T) energy such that the total energy is SCF/a[Q,5]Z + δ^{MP2}/a[Q, 5]Z + $\delta_{MP2}^{CCSD(T)}$/aXZ where terms in square brackets indicate the basis sets used in the CBS extrapolation. Values for X are listed in Table 3.1. While the precise benchmark energies will be useful for future studies where highly accurate interaction energies for pentamers are required (to our knowledge, no such database exists), here the most interesting aspect is the magnitude of the m-body corrections. The relative size of m-body corrections to the interaction energy is studied in Figure 3.3. For convenience, a quantity called the truncation error $\mathcal{T}(m)$ is defined to quantify the difference in the energy computed using a

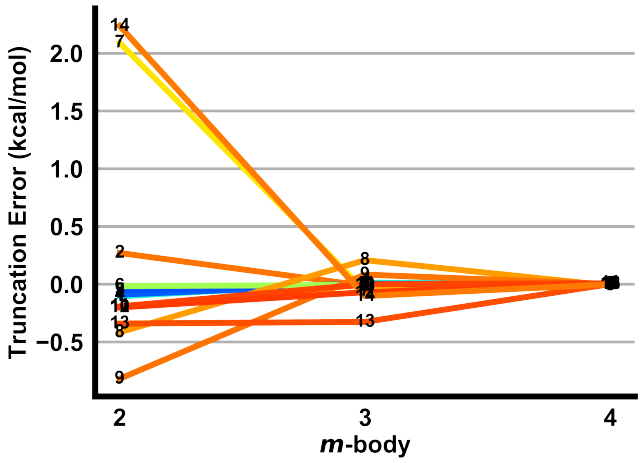


Figure 3.3: Truncation errors, $\mathcal{T}(m)$ (Equation 3.6), due to truncating the many-body expansion at m-bodies vs the full VMFC(5) result, for the benchmark method, VMFC-corrected SCF/a5Z + δ^{MP2}/a[Q,a5]Z + $\delta^{CCSD(T)}_{MP2}$/aXZ (values of X in Table 3.1). $\mathcal{T}(5) = 0$ by definition. Individual systems are labeled by their numbers in Figure 3.1.

truncated VMFC versus the entire VMFC series,

$$\mathcal{T}(m) = \mathcal{E}(m) - \mathcal{E}(\mathcal{M}). \tag{3.6}$$

Figure 3.3 displays the important finding that by $m = 3$, the truncation error is less than 0.5 kcal mol^{-1}, and by $m = 4$, the truncation error is below 0.1kcal mol^{-1}. This indicates that truncating the VMFC to $m = 3$ is a viable option for reducing computational cost while maintaining accuracy. In the full manuscript, it is also demonstrated that this convergence behavior is maintained for SCF, the δ^{MP2} correlation energy, and the $\delta^{CCSD(T)}_{MP2}$ correlation energy.[25] This should not be surprising as any errors manifested in these individual components would also be seen in the benchmark data. As calculations on solid state crystalline materials are often carried out using DFT, a study on the truncation errors of various functionals was also performed. In Figure 3.4, it is seen that there is similar convergence behavior to the benchmark method; by $m = 3$, the truncation error is below 0.5 kcal mol^{-1}. Another interesting feature is that all of the DFT functionals converge at

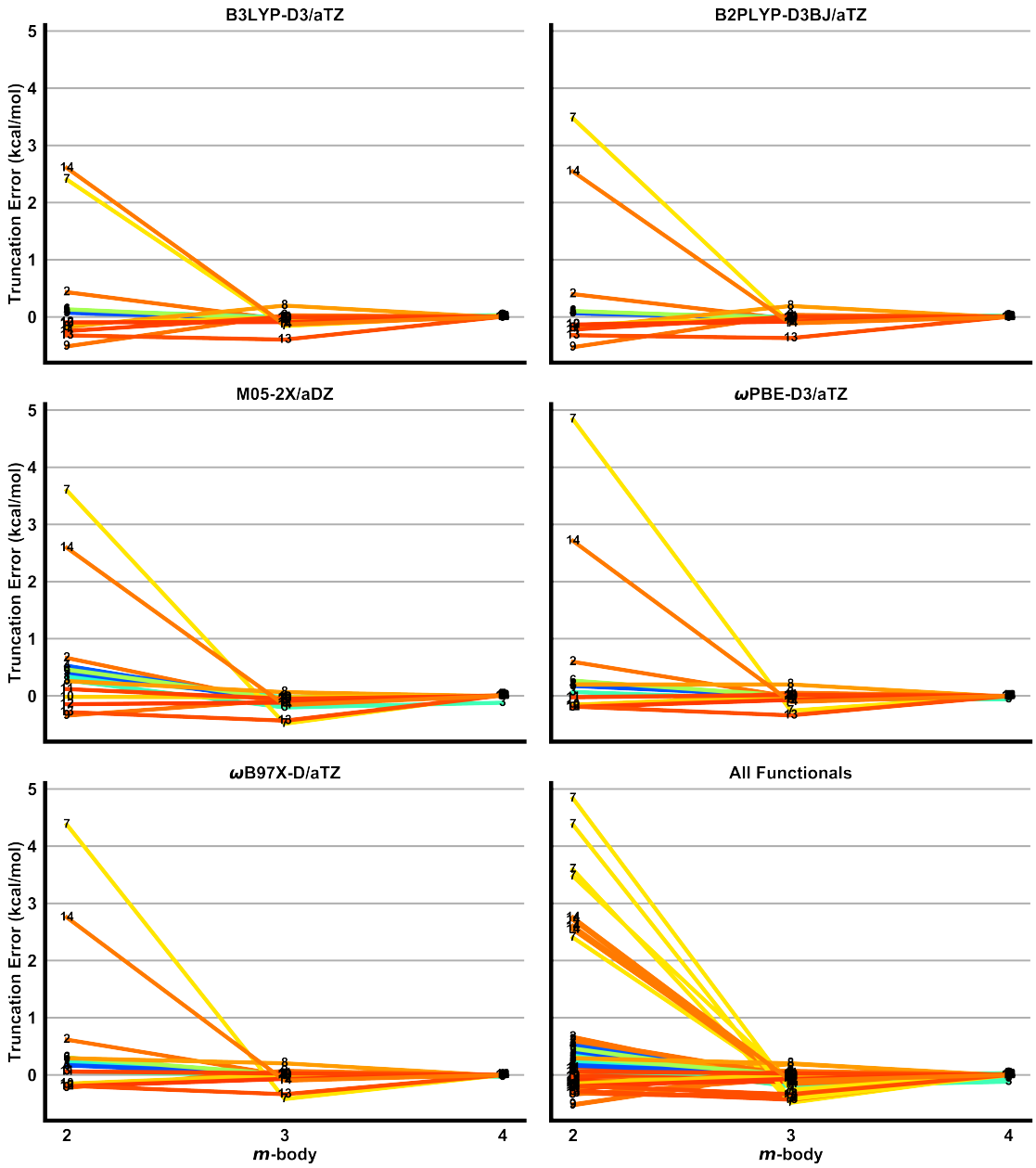


Figure 3.4: Truncation errors, $\mathcal{T}(m)$ (Equation 3.6), in VMFC-corrected DFT interaction energies due to truncating the many-body expansion at order m. DFT methods are the same as in previous panels. Individual systems are labeled by their numbers in Figure 3.1.

approximately the same rate with respect to the number of m-body interactions needed. This is not necessarily a given as the exact makeup of DFT functionals with respect to the weighting of exact exchange, MP2 correlation energy (B2PLYP), and other components is not constant between the functionals.

Identifying and proving that the truncation order of $m = 3$ or $m = 4$, depending on the reduction in truncation error desired, is one of the most important findings of this work. The breadth of intermolecular interactions covered by our test set (Figure 3.2) indicates that this truncation order should be general to any chemical system. The only concern for a loss of generality would be in large, extended systems where induction through multiple, separate, adjacent fragments is important. An example of such a system could be anthracene interacting with ethanol as a solute-solvent type interaction; the ethanol is much smaller than the anthracene, so several ethanol molecules would be interacting with the anthracene. However, the generalized many body expansion has shown that this should not be a problem.[26]

After considering the issue of minimizing truncation order, a second component to leveraging the many body expansion for facilitating larger calculations is identifying ways to maintain accuracy while lowering computational cost. One such method is to compute different terms of the many body expansion (*i.e.* the 2-body, 3-body, etc correction to the interaction energy) with different levels of theory. For example, the 2-body term could be computed with a high level of theory while higher order terms could be computed with DFT. In order to know when such approaches are efficacious, the accuracy of each method considered here compared to the benchmark method must be considered. For the full data set, see the manuscript for this project.[25] Here, the comparison between the accuracy of each DFT functional versus the benchmark, DF-FNO-CCSD(T) CBS extrapolated, values is presented for each m-body correction (Figure 3.5). From the comparison of DFT and benchmark VMFC values, it is seen that the 3-body correction on average (gray horizontal bars in the boxplots) contains less than 0.5 kcal mol^{-1} of error versus the benchmark. Note

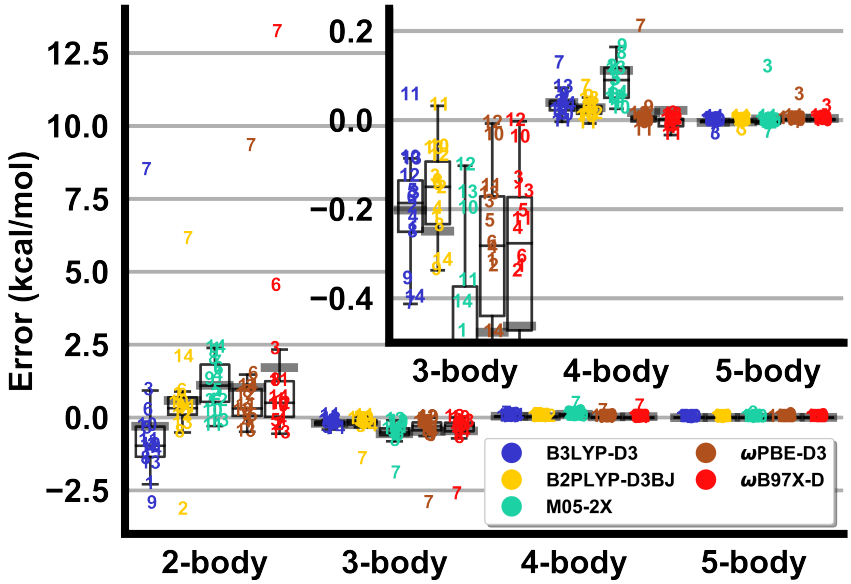
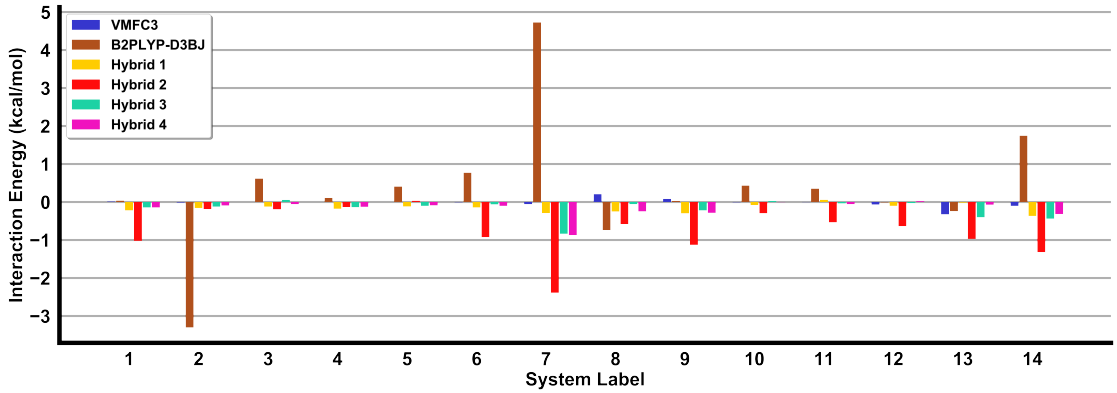


Figure 3.5: Errors in VMFC-corrected DFT values for non-additive m-body interaction energies vs benchmark values. Individual systems are labeled by their numbers in Figure 3.1. Boxplots show the first, second(median), and third quartiles as horizontal lines on the box, and the whiskers extend outside the box to 1.5 times the interquartile range (IQR). The mean is shown as a thick horizontal gray line.

that at $m = 3$, the truncation error $\mathcal{T}(m)$ is of this same magnitude for both DFT and the benchmark method. Therefore, the substitution of DFT for coupled cluster or MP2 3-body corrections to the interaction energy should introduce little if any more error than the truncation error already introduces. However, 2-body corrections have a larger magnitude of error and should be calculated with higher levels of theory as the dimer calculations are small and the largest contribution to the total interaction energy comes from the 2-body interaction.

Using this finding, alternative, multi-method VMFC schemes for computing the interaction energy of the pentamer systems in Figure 3.1 were proposed. In Figure 3.6), it is seen that some of these hybrid approaches yield excellent results in comparison to the benchmark method at drastically reduced computational cost. For example, hybrid 3 strikes an excellent balance of accuracy and diminished computational cost. Instead of using large basis sets like aTZ or aQZ for the coupled cluster component, the aDZ basis set is used. Additionally, the a[Q,5]Z CBS extrapolated MP2 components is reduced to a[T,Q]Z CBS extrapolated values for the 2-body correction to the interaction energy. Although MP2 energies are known to be slow to converge to a stable energy with respect to basis set size, this does not manifest itself in the form of large errors here. Finally, the VMFC is truncated to $m = 3$ for hybrid 3. Although no timing studies were performed here, the speedup in using hybrid 3 versus the benchmark values could be at least 10 times faster. Thus, using the knowledge that the many body expansion can be truncated at order 3 with minimal loss in accuracy as well as the data presented here to determine the loss in accuracy upon replacing expensive calculations with calculations using lower levels of theory, there is promise that the VMFC can be used to compute accurate interaction energies for large systems in less time than through alternative approaches.

Figure 3.6: Interaction energies are calculated using various quantum chemistry methods in tandem as compound approaches. They are presented versus the VMFC(5) benchmark values; the values shown are X-$\mathcal{E}$ where $\mathcal{E}$ is the VMFC(5) value. B2PLYP-D3BJ interaction energies were included to show the accuracy of DFT alone, with this functional standing out in Figure 3.5 as being exemplary. Hybrid 1 is the 2-body VMFC term from the benchmark data with 3- and 4-body B3LYP-D3. Hybrid 2 is SCF/aQZ + δ^{MP2}/a[T,Q]Z + $\delta_{MP2}^{CCSD(T)}$/aDZ, with the simple interaction energy shown. Hybrid 3 is also SCF/aQZ + δ^{MP2}/a[T,Q]Z + $\delta_{MP2}^{CCSD(T)}$/aDZ, but only for the VMFC 2-body term; the three body VMFC term is only MP2/aDZ, and the VMFC series is truncated here. Hybrid 4 is the benchmark 2-body VMFC term with 3- and 4-body VMFC terms from MP2/aTZ and the 5-body term truncated.



3.4 Conclusion

A truncation order of $m = 3$ has been shown to be appropriate for reducing truncation errors in the VMFC form of the many body expansion to below 0.5 kcal mol^{-1} for general chemical systems across SCF, MP2, FNO-CCSD(T), and DFT levels of theory. The generality of this finding is supported by the wide variety of intermolecular interaction motifs found in our test set and confirmed by symmetry adapted perturbation theory (3.2). Based on the accuracy of various reduced cost methods such as DFT or performing calculations with smaller basis set sizes, several hybrid methods are proposed. These methods often maintain the sub 0.5 kcal mol^{-1} accuracy with respect to benchmark values while reducing the computational cost by a figure on the order of magnitude of 10-fold. Finally, all of the data from this study is provided as Python data structures which enables easy analysis for other applications.[25] This data includes gold-standard level interaction energies for pentamer systems; to our knowledge, no benchmark set of pentamer structures at this level of theory currently exists.

CHAPTER 4
CONCLUSION

One critical problem motivating both of these projects is the dissolution of a crystalline solid in solution; controlling this physical behavior is critical for modulating the kinetics of the transition of drug molecules *in vivo* from their solid form to the active molecular form. In order to reach such a goal, atomistic solvation models and methods to compute single point energies of large systems will be necessary. This thesis presents studies into both of these areas.

In order to compute free energy, entropy is needed. To this end, a new approach for linking molecular mechanics and quantum mechanics calculations was explored. Instead of embedding potentials between MM into QM calculations and vice versa, MM was used as a tool to assess the distribution of conformations of a solute molecule in atomistic solvent. This molecular mechanics trajectory then was used to compute accurate single point energies using DFT in order to correct the potential energy surface. Although still underdeveloped, the progress made here will hopefully inspire future developments. The creation of an interface between the programs Psi4 and OpenMM will make integration of QM and MM calculations into a uniform workflow more facile, and its development was driven by this project.

With regards to lattice energy, using the Valiron and Mayer functional counterpoise correction to reduce computational cost for computing the interaction energy of a system of fragments was studied. Although it had been hypothesized and often taken for granted that the VMFC could be truncated to $m = 3$, here this was demonstrated as true for a system of 14 chemically diverse small molecule pentamer crystal structures. Truncating the VMFC can enable enormous reductions in computational cost as behemoth calculations are replaced by a larger number of small, readily tractable calculations. The VMFC formalism

enables these smaller computations on a reduced set of fragments from the supersystem to be computed in the basis set of only these fragments. Therefore, for a system of 1,000 fragments, if a truncation order of 3 is used, the largest computation required is one where the basis set functions are only contributed to by 3 fragments versus 1,000.

In addition to establishing this truncation order, hybrid methods that accurately reproduce the benchmark VMFC(5) values are tested. For one hybrid method, SCF/aQZ + δ^{MP2}/a[T,Q]Z + $\delta_{MP2}^{CCSD(T)}$/aDZ for the 2-body term and MP2/aDZ for the 3-body term, an average error of less than 0.5 kcal mol^{-1} versus the benchmark was found. For this hybrid method, it is expected that the reduction in computational cost is on the order of 50-fold. Therefore, when combining an appropriately reduced level of theory and a well chosen truncation order, computing the interaction energy for large systems becomes possible.

It is the hope that developments such as the parallel MM and QM simulations as well as the VMFC benchmarking study presented here will contribute to attaining the goal of predicting critical physical properties from first principles.

Appendices

APPENDIX A

GENERATION OF PENTAMER STRUCTURES

The main test set comprises the 14 numbered systems used for the testing of the many body expansion. In cases where the system name may be ambiguous (the ice systems, for example), the system name used in calculations is provided. The clusters of acetic acid, ammonia, low pressure carbon dioxide, cyanamide, formamide, and urea are taken from the C21 dataset.[27] All of the other geometries were generated using crystallographic information files (CIF) found using the crystallography open database (COD). Using the provided CIF files, the free program Mercury was used in order to expand the crystal structure using the fragments closest to the origin. Typically, this means that an entire unit cell was not generated as the unit cell often contains more than 5 molecules.

For four structures from the COD, hydrogens had to be added to the structures as the CIF files did not contain them (often hydrogens are not included for structures generated using X-ray diffraction because hydrogens interact weakly with X-rays). These were ethylene, methane, hydrogen fluoride, and water in the IC crystal structure. Hydrogens were added using the babel program.[28]

REFERENCES

(1) Eastman, P. et al. *Journal of Chemical Theory and Computation* **2013**, *9*, PMID: 23316124, 461–469.

(2) Parrish, R. M. et al. *Journal of Chemical Theory and Computation* **2017**, *13*, PMID: 28489372, 3185–3197.

(3) Mennucci, B.; Tomasi, J.; Cammi, R.; Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Gabriel, S.; Stephens, P. J. *The Journal of Physical Chemistry A* **2002**, *106*, 6102–6113.

(4) Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(5) K., V.; E., H.; C., A.; S., K.; S., Z.; J., S.; E., D.; O., G.; P., L.; I., V.; D., M. A. *Journal of Computational Chemistry* **2010**, *31*, 671–690.

(6) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.

(7) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247 –260.

(8) Mobley, D.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. *bioRxiv* **2018**.

(9) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.

(10) Araz, J.; L., B. B.; B., J. D.; I., B. C. *Journal of Computational Chemistry* **2002**, *21*, 132–146.

(11) Carroll, W. R.; Zhao, C.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D. *Organic Letters* **2011**, *13*, PMID: 21797218, 4320–4323.

(12) Li, P.; Parker, T. M.; Hwang, J.; Deng, F.; Smith, M. D.; Pellechia, P. J.; Sherrill, C. D.; Shimizu, K. D. *Organic Letters* **2014**, *16*, PMID: 25238038, 5064–5067.

(13) Richard, R. M.; Bakr, B. W.; Sherrill, C. D. *Journal of Chemical Theory and Computation* **2018**.

(14) Raghavachari, K.; Trucks, W. G.; Pople, A. J.; Head-Gordon, M. *Chemical Physics Letters* **1989**, *157*, 479–483.

(15) DePrince, E. A.; Sherrill, D. C. *Journal of Chemical Theory and Computation* **2013**, *9*, 2687–2696.

(16) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1995**, *106*, 9639–9646.

(17) Becke, D. A. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

(18) Lee, C.; Yang, W.; Parr, R. G. *Physical Review B* **1988**, *37*, 785–789.

(19) Grimme, S. *The Journal of Chemical Physics* **2006**, *124*.

(20) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, PMID: 26626525, 364–382.

(21) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*.

(23) Grimme, S.; Ehrlich, S.; Goerigk, L. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(24) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerellJr., A. D.; MerzJr., K. M.; Sherrill, C. D. *The Journal of Chemical Physics* **2017**, *147*, 161727.

(25) Zott, M. D.; Richard, R. M.; Sherrill, C. D. *In preparation* **2018**.

(26) Richard, M. R.; Herbert, M. J. *Journal of Chemical Theory and Computation* **2013**, *9*, 1408–1416.

(27) De-la Roza, A. O.; Johnson, E. R. *The Journal of Chemical Physics* **2012**, *137*, 054103.

(28) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*, 33.