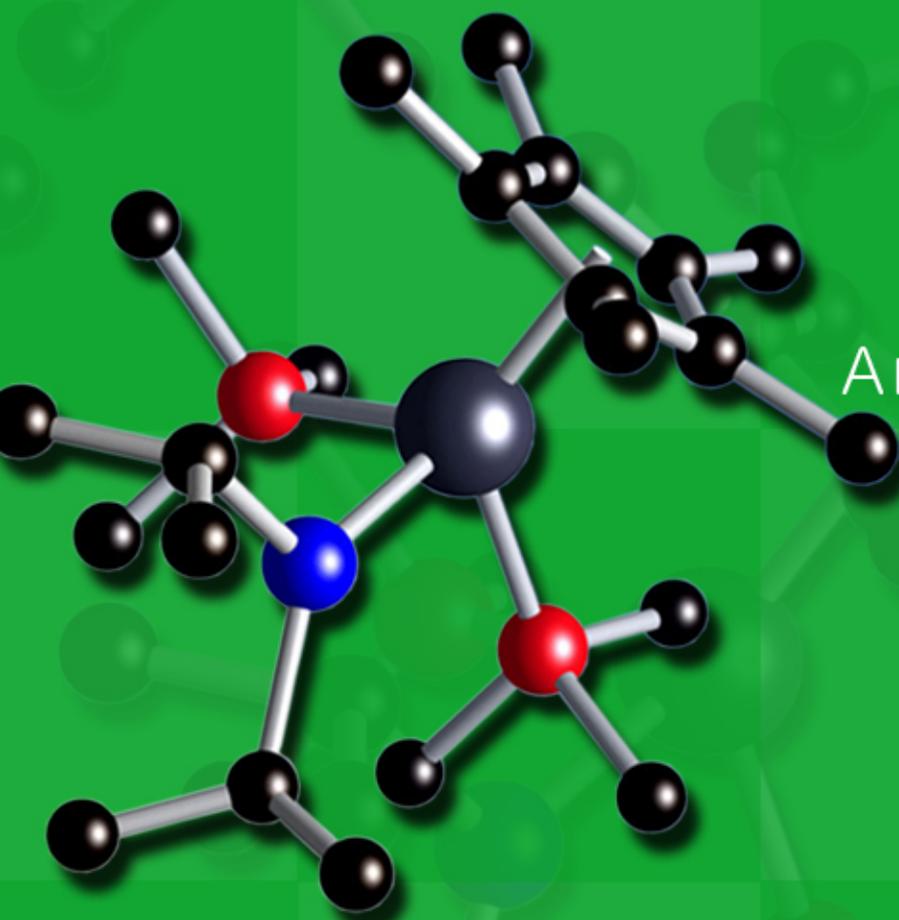


# Studies in Computational Quantum Chemistry



Amy Jean Austin

# Studies in Computational Quantum Chemistry

**Author**

Amy Jean Austin

Faculty Advisor: George A Petersson

USA

**Published By:**

MedCrave Group LLC

March 08, 2016

## Disclaimer

A Dissertation submitted to the Faculty of Wesleyan University in Partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry.

## Acknowledgement

At the end of my first year in the Ph.D. program in the department of Chemistry at Wesleyan University I was speaking to someone about how the experience was coming along... I tried to refer to the Chemistry faculty (the reader is assured, in a positive way), but I had a "slide of tongue." What actually came out was "The Chemistry family." Now, of course, I felt very silly at the time and corrected myself promptly, but later I realized this was honestly how I felt.

One of the qualities that I found so attractive of Wesleyan is the small and intimate nature of the department. With less than twenty professors, around thirty graduate students, and ten current undergraduates, we are a tiny but fierce department, striving for academic excellence in all areas of chemical research as a common goal. We are such a small department that we get to know one another quite well (whether we want to or not), and even by the end of my first year I had begun to think of the chemistry department as a whole: faculty, staff, and all students as a type of extended family. In fact, much like families, we don't even always get along! (The author inserts a chuckle here.) However, in the end, we will always pull ranks to be there for each-other. There was a time when I needed to go on a medical leave of absence, and everyone was very supportive to help me get done what needed to be in place. When I was ready to come back, they welcomed me with open arms, and helped me get re-oriented so that I could get back on the path that leads to graduation.

It is with all of my heart that I thank the chemistry family of Wesleyan University. I would like to give special thanks to my research advisor Professor George Petersson labmates Ericka Barnes, Frank Dobek, MiKyung Lee, and Sonya Tulyani. My time in the department of chemistry has been one of the most challenging, exciting, and happiest times of my life. Thank you for allowing me to be a part of the Wesleyan experience.

<b>Contents</b>	
<b>Abstract</b>	<b>1</b>
<b>Chapter 1</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
An Overlap Criterion for the Selection of Molecular Core Orbitals.	
A Spherical Atom Model for Dispersion Forces	
Vibrational Frequencies of Transition States	
The Mechanism and Enantioselectivity of the Fiest-Bénary Reaction	
<b>Theoretical Background</b>	<b>3</b>
Introduction	
Background	
The Hartree-Fock Approximation and Self-Consistent Field (SCF) Method.	
Perturbation Theory.	
Coupled Cluster Singles and Doubles (CCSD) Method with Perturbative Triples (CCSD(T)).	
Quadratic Configuration Interaction Singles and Doubles (QCISD) Method with Perturbative Triples (QCISD(T)).	
Density Functional Theory.	
The Complete Basis Set Extrapolation.	
Figure 1	
The ONIOM Hybrid Method.	
Force Field (Molecular Mechanics, MM) Methods.	
Figure 2	
<b>Chapter 2</b>	<b>16</b>
<b>An Overlap Criterion for Selection of Core Orbitals</b>	<b>17</b>
Introduction and Background.	
Table 1	
Results and Discussion.	
Figure 3	
Table 2	
Figure 4	
Figure 5	
Figure 6	
Figure 7	
Table 3	
Figure 8	
Figure 9	
Table 4	
Table 5	
Table 6	
Table 7	
Conclusions	

**A Spherical Atom Model for Dispersion Forces**

Introduction

The Spherical Atom Model for Dispersion Forces.

Figure 10

Figure 11

Figure 12

Figure 13

Figure 14

Comparisons with other Models

Numerical Tests

Figure 15

Figure 16

Figure 17

Figure 18

Figure 19

Table 8

Figure 20

Figure 21

Figure 22

Figure 23

Table 9

DFT Functional Comparisons

Figure 24

Figure 25

Figure 26

Figure 27

Figure 28

Figure 29

Figures 30

Figures 31

Table 10

The APF-SAM Hybrid Density Functional Theory.

Figure 32

Figure 33

Figure 34

Table 11

Figure 35

Figure 36  
Figure 37  
Table 12  
Conclusions

## Chapter 4

50

### Vibrational Frequencies of Transition States

51

Introduction and Goal of Research

Figure 38

Calibration with Experimental Harmonic Frequencies of Stable Molecules.

Table 13

Table 14

Table 15

Symmetric Transition States of Degenerate Chemical Reactions.

Table 16

Table 17

Table 18

Table 19

Table 20

Asymmetric Transition States of Non-Degenerate Chemical Reactions

Figure 39

Table 21

Table 22

Table 23

Figure 40

Figure 41

Figure 42

Figure 43

Definitive Tests for the Intrinsic Reaction Coordinate Maximum (IRCM<sub>ax</sub>) Method for Predicting Frequencies of Transition States.

Table 24

Table 25

Table 26

Table 27

Limitations of the IRC<sub>Maximum</sub> Method for obtaining accurate frequencies of transition states and zero point energy contributions to the activation energy.

Figure 44

Figure 45

Table 28

Conclusions

**The Mechanism and Enantioselectivity of the Interrupted Feist-Bénary Reaction**

Introduction

Figure 46

Figure 47

Figure 48

General Considerations

Figure 49

Figure 50

Figure 51

Difficulties Encountered.

Figure 52

Table 29

Figure 53

Figure 54

Figure 55

Figure 56

Figure 57

Figure 58

Figure 59

Figure 60

Figure 61

Figure 62

Final Results

Figure 63

Figure 64

Figure 65

Figure 66

Figure 67

Figure 68

Table 30

Figure 69

Figure 70

Figure 71

Conclusions

## Abstract

An overlap criterion is defined, such that the selection of core orbitals in molecular systems is based on similarity to their constituent atom core orbitals. We test the overlap criterion for problematic systems, including molecules containing third-row main group elements, and it is shown to remove errors in atomization energies up to 100kcal/mol arising from core orbital inversions between the molecular system and isolated atoms. We present a novel atom-atom pair wise potential function for describing dispersion forces, based on the interaction of spherical polarizable atoms. Using the atomic radius and  $c_6$  parameters from  $He_2$ ,  $Ne_2$ ,  $Ar_2$ ,  $H_2$  (triplet S) and  $C_2$  (nonet S)(triplet S) and (nonet S), this model predicts dispersion interactions which are in excellent agreement compared to benchmark potential energy curves for HeNe, HeAr, and NeAr, and HC (sextet S). This spherical atom model for dispersion forces can be integrated into density functionals and molecular mechanics potentials. For four exothermic, hydrogen transfer reactions examined in this work, it was found that high levels of correlation treatment with large basis sets were necessary. We have shown that the IRC Max method of Petersson and coworkers offers a cost-efficient means of attaining reasonable frequencies of transition states and zero point energy changes compared to reactants since it locates the correct position of the transition state in a cost-effective way. The limitations of the IRCMaximum method for obtaining accurate frequencies of transition states and zero point energy changes compared to reactants are discussed.

The mechanism and enantioselectivity of the Feist-Bénary reaction was studied. The flatness of the potential energy surface imposes several limitations. We predict a 5.2 kcal/mol selectivity that is in favor of the S enantiomer, which is an overestimate. The selectivity for the S enantiomer may be attributed to steric interactions within the side chains of the catalysts for the two enantiomers, as well as steric and electrostatic interactions between the catalyst and aldol complex. The repulsion between the pyrimidinyl group and the approaching nucleophile may inhibit attack present in the R enantiomer is not observed in the S enantiomer. A modification in the side chain is proposed to enhance the selectivity.

# Chapter 1

## Introduction

### An overlap criterion for the selection of molecular core orbitals

The mixing that occurs among orbitals in the valence of atoms leading to chemical bonding does not occur for orbitals in the core set which are tightly bound and lie close to the nucleus. Therefore, one approximation for improving computational expense would be to neglect the core electrons in the correlation treatment. The usual approach for partitioning molecular orbitals into valence and core sets involves an energy criterion. The lowest energy molecular orbitals are selected for the core. The number of such core orbitals is set equal to the sum of the core sizes for the constituent atoms. The use of an energy criterion for the selection of molecular core orbitals may result in the inversion of molecular orbitals with respect to the atoms, and thus can create inconsistency between the treatment of orbitals in the atoms and in the molecule. Recent studies reveal that large errors in atomization energies are obtained when methods that correlate only the valence electrons are applied to certain molecules containing third-row elements. It is the goal of this work to develop an overlap criterion for the selection of core orbitals in molecular systems that is based on mathematical similarity to their constituent atom core orbitals.

### A spherical atom model for dispersion forces

Dispersion forces play a significant role in the structures and properties of large molecular structures, such as biological systems, studies of which are therefore inherently computationally demanding. Molecular mechanics empirical potentials, which include London dispersion terms based on point-dipole interactions, are often used. However, quantitative predictions of dispersion forces require high levels of correlation treatment. When practical, Density functional Theory (DFT) has become the method of choice for studies of large molecules since it is less computationally demanding than high-level *ab initio* methods. A major shortcoming of functionals in wide use is their failure to properly describe long range correlation effects, such as dispersion forces. Recent successful attempts to incorporate the fundamental physics of dispersion forces into GGA DFT have been achieved, but these functionals requires six-dimensional numerical integration over pairs of points, which is impractical. We seek to develop an atom-atom pair wise potential function that can be appended to DFT methods without significantly affecting their computational cost. In the spherical atom model for dispersion forces, atoms are treated as thin shells of polarizable media. We propose that the primary deviation from the  $\frac{C_6}{R^6}$  interaction of two point dipoles is due to the fact that atoms have volume.

### Vibrational frequencies of transition states

Zero Point Energy (ZPE) changes from reactant to transition states raise or lower the activation energy of reaction, which affects its rate; and therefore, it is important to obtain these changes accurately. We have shown for

four non-degenerate chemical reactions that high levels of correlation treatment are necessary (QCISD(T) with large basis sets). The Intrinsic Reaction Coordinate Maximum (IRC Max) method of Petersson and co-workers was previously developed to improve the geometries of transition states and barrier heights of chemical reactions. It was the goal of this work to test whether the IRC Max method can also improve the normal mode frequencies of transition states and ZPE changes from reactant to transition state of chemical reactions.

### The mechanism and enantioselectivity of the Feist-Bénary reaction

The Feist-Bénary synthesis is used to synthesize furans from  $\alpha$ -halogenated ketone and  $\beta$ -dicarbonyl compounds. A modification is the interrupted Feist-Bénary reaction (IFB), in which the synthesis is stopped at the hydroxydihydrofuran Cinchona alkaloid quinine catalysts that are derivatized to contain substituted pyrimidinyl groups are highly enantioselective. Enantioselective Feist-Bénary reactions are synthetically important. For example, zaragonic acid is a fungal metabolite that is a cholesterol reducing agent. A better understanding of the catalytic action facilitates enhancing the enantioselectivity. It is the goal of this work to understand through computation how these asymmetric catalysts interact with reactants in order to confer the observed enantioselectivity.

## Theoretical Background

### Introduction

The goal of quantum chemistry is to provide accurate mathematical predictions of chemical properties in the most time-efficient manner possible. The ability to predict the outcome of a chemical reaction, for example, is a notion that is fundamental for any chemist right from their “teeth cutting” days of basic organic chemistry. We are taught to think about molecules in terms of regions of high and low electron density in a qualitative sense in order to predict reasonable reaction mechanisms and therefore products. We might even think of this as a completely qualitative density functional theory. In quantum chemistry, we strive to mathematically represent an accurate physical model of the desired system. It turns out that this is a process that is somewhat *ad hoc*. An equation is derived from the physics of the model, and then once data have been collected, the equation can be tested. Of course, we always find that our model was too simple (or that in reality chemistry is just really hard!?). Therefore, we have to amend the model to include some effects that were neglected, etc. The cycle is repeated until benchmark data can be accurately reproduced, and then it can be used for predictive purposes, at least until new information emerges that proves a discrepancy. Then the model must be revised again. That new models are developed to refine previous ones in light of new information has always been a part of the progress of science, and is present throughout the history of quantum chemistry. Nowadays the Bohr model of the atom is accepted, originally proposed in 1913, in which negatively charged electrons in stationary states with

quantized energy levels surround a compact nucleus (that contains most of the mass of the atom) of positive charge. This is in stark contrast to J. J. Thompson's 1904 formulation, in which the atom was seen as a spherical cloud of positive charge with embedded negative electrons. Sounds a bit like plum pudding. In 1911, Rutherford proposed the "planetary model" in which the electrons would 'orbit' the nucleus, but of course, then the electrons would be accelerating and continuously radiating energy. They would lose their ability to maintain 'orbit' and spiral into the nucleus. Of course, Bohr's inclusion of quantization of angular momentum refined this model. The idea of energy being quantized was a result of blackbody experiments in the 1800s. Blackbodies are objects that absorb all light falling on them. Max Planck assumed that the atoms of the blackbody could emit light energy in packets, or discrete amounts based on their frequency. He was able to re-produce experimental blackbody curves, and the notion of quantization was born.

The Schrödinger wave equation is a useful tool for describing chemistry both qualitatively, for having a conceptual picture of chemical bonding, and for providing accurate quantitative predictions of chemical properties, especially in situations where experimental results are not sufficiently reliable or unobtainable (e.g. transition state properties). The angular portion of the atomic orbital wave function can be described accurately from solutions to the Schrödinger equation for a central field, and a linear combination of the constituent atomic orbitals (AOs) is taken to form the molecular wave function in a procedure called LCAO (linear combination of atomic orbitals). The Woodward-Hoffman rules are extremely important for predicting the regioselectivity of conjugated pi bonding systems in cyclization reactions, and these are based on a qualitative picture of the highest energy occupied molecular orbital (HOMO) of one species relative to the lowest unoccupied molecular orbital (LUMO) of the second species in the chemical reaction. Since the Schrödinger equation itself is too complex to be soluble for any but a few specialized potentials, (e.g., particle in a box, harmonic oscillator, hydrogen atom) modern electronic structure theory has created a hierarchy of approximations to the Schrödinger equation (with density functional theory being a bit of a wild card here). Better approximations to the Schrödinger equation do require more computational time. We have a sign in our lab predating any of the current Petersson-crew that reads "Methane by Christmas!" However, in this age of technology the notion of "ready-set-upgrade!" has never been more descriptive of the times. Computing speed is developing at rates that make our best methods more accessible for increasingly larger systems. Additionally, hybrid methods, such as the onion-like ONIOM method, and Intrinsic Reaction Coordinate Maximum (IRC Max) methodologies provide cost-efficient means of predicting properties as though they are calculated with one better methodology which is itself impractical, when in fact lower methods were utilized as well.

Another way that quantitative descriptions of chemical properties and reactions are improving is simply through wider use. John Pople released the Gaussian 70 program

[1], named for its use linear combinations of Gaussian type orbitals for the description of atomic orbitals for faster two-electron integrations [2]. This is in contrast to Slater type orbitals which are exponential, and while they give better cusp-behavior they are much more difficult to integrate. The Gaussian code automates the predictive power of electronic structure theory and the user-friendly interface (Gaussview) extends the use to chemists in any subfield.

## Background

Experimental studies conducted in the early twentieth century revealed that electrons gave diffraction patterns similar to light waves when passed through a double slit. It is reasonable that electrons should therefore behave not just as particles, but also as waves, and are found to obey a wave equation, specifically the fundamental equation for all of quantum mechanics, the Schrödinger equation. The time dependent Schrödinger Equation has the form of Equation (2.2.1).

$$\hat{H}\Psi(\vec{r},t)=\frac{i\hbar}{2\pi}\frac{\partial\Psi(\vec{r},t)}{\partial t} \quad (2.2.1),$$

Where  $\Psi$  the particle wave function,  $\vec{r}$  is its coordinate vector,  $\hbar$  is Planck's constant, and  $\hat{H}$  is the Hamiltonian operator, given by Equation (2.2.2),

$$\hat{H}=\frac{-\hbar^2}{8\pi^2m}\nabla^2+V,\nabla=\frac{\partial}{\partial x}\vec{i}+\frac{\partial}{\partial y}\vec{j}+\frac{\partial}{\partial z}\vec{k} \quad (2.2.2)$$

In Equation (2.2.2), the first term describes the kinetic energy of the particle, and  $V$  is the potential energy.

When the potential energy,  $V$ , is not a function of time, the Schrödinger equation can be simplified to its time-independent form, given in Equation (2.2.3),

$$\left(\frac{-\hbar^2}{8\pi^2m}\nabla^2+V(\vec{r})\right)\Psi(\vec{r})=E\Psi(\vec{r}) \quad (2.2.3)$$

The solutions to Equation 2.2.3 are the eigenvector wave functions,  $\psi$ , known as the stationary states, and their accompanying eigen value energies. The lowest energy is the ground state.

Chemists are interested in using the wave equation for the calculation of atomic and molecular properties such as energies, geometries, vibrational frequencies, thermo chemistry, etc which involve time-independent potentials, and therefore we focus future attention solely on the time-independent Schrödinger equation of Equation (2.2.3). The goal of ab initio quantum mechanics is to determine the wave function,  $\psi$ , and the observable property, or eigen value, corresponding to the operator at hand (e.g., the energy in Equation (2.2.3) when the Hamiltonian operator is used).

The general Hamiltonian operator for a molecular system in a vacuum is given by equation (2.2.4),

$$\hat{H}_{molecular} = \frac{-\hbar^2}{8\pi^2} \left( \sum_{i=1}^{N_e} \frac{\nabla_i^2}{m_i} + \sum_{A=1}^{N_n} \frac{\nabla_A^2}{m_A} \right) + \frac{1}{4\pi\epsilon_0} \left( \sum_{i < j}^{N_e} \frac{e^2}{\Delta r_{ij}} + \sum_{A < B}^{N_n} \frac{Z_A Z_B e^2}{\Delta R_{AB}} - \sum_{i < A}^{N_e} \sum_{A}^{N_n} \frac{Z_A e^2}{\Delta r_{iA}} \right) \quad (2.2.4)$$

Here,  $m$  is the mass of the particle,  $\hbar$  is Planck's constant,  $-e$  is the charge of an electron, and  $Ze$  is the charge of a nucleus, where  $Z$  is the atomic number. The first two terms of equation (2.2.4) correspond to the kinetic energies of the electrons,  $i$ , and nuclei,  $A$ , respectively. The last three terms are potential energy terms, the first of describes the electron-electron repulsion, followed by the nuclear-nuclear repulsion, and finally the electron-nuclear attraction. Already we get the feeling the "simple" wave equation (2.2.3) is very complex. In fact, the only chemical system in which the Schrödinger equation is analytically tractable without approximation is the Hydrogen atom, and in this case, we won't be predicting much of any use. The first of many approximations is the Born-Oppenheimer approximation. In this approximation, pseudo separability is employed since the nuclear motion is so slow relative to electronic velocities. The first step in the Born-Oppenheimer

approximation is to assume that the electrons move in a field of fixed nuclei, and this allows the kinetic energy of the nuclei to be neglected and the repulsion between nuclei to be constant so that an electronic Hamiltonian,  $\hat{H}_{elec}$ , is formed. Although the electronic wave function does not explicitly depend on the nuclear coordinates, the electronic wave function as well as the electronic energy differs with varying arrangements of the nuclei. After solving the electronic Schrödinger Equation, the electronic energy,  $E_{elec}$ , along with the nuclear repulsion become the effective potential for the nuclear Hamiltonian,  $\hat{H}_{nuc}(\vec{R})$ . Solving the nuclear Schrodinger equation allows for description of the vibrational, rotational, and translational degrees of freedom. The electronic and nuclear Hamiltonians are presented in atomic units (1 Hartree =  $\frac{e^2}{a_0}$ , and  $a_0 = \frac{\hbar^2}{4\pi^2 m_e e^2}$ ), in Equation 2.2.5.

$$\hat{H}_{elec}(\vec{r}, \vec{R}) = \frac{-1}{2} \left( \sum_{i=1}^{N_e} \nabla_i^2 \right) + \left( \sum_{i < j}^{N_e} \frac{1}{\Delta r_{ij}} - \sum_{i < A}^{N_e} \sum_{A}^{N_n} \frac{Z_A}{\Delta r_{iA}} \right), \hat{H}_{nuc}(\vec{R}) = \frac{-1}{2} \left( \sum_{A=1}^{N_n} \frac{\nabla_A^2}{m_A} \right) + \sum_{A=1}^{N_n} \sum_{B>A}^{N_n} \frac{Z_A Z_B}{R_{AB}} + E_{elec} \hat{H}_{nuc}(\vec{R}) \Psi_{Nuc} = E_{total} \Psi_{Nuc} \quad (2.2.5)$$

Another approximation to the Schrödinger Equation is the Independent Partical Approximation (IPA) called the "bare nucleus" model is a rather extreme further restriction in which the electron-electron repulsion is neglected from the Born-Oppenheimer Hamiltonian of Equation (2.2.5). It is important because each electron behaves independently, and a new Hamiltonian can be written as a simple sum of one-electron Hamiltonians in Equation (2.2.6),

$$\hat{H}_{Bare-Nuc} = \frac{-1}{2} \sum_{i=1}^{N_e} \nabla_i^2 - \sum_{i < A}^{N_e} \sum_{A}^{N_n} \frac{Z_A}{\Delta r_{iA}} \quad (2.2.6)$$

The solution to the bare nucleus equation (2.2.6) is a product of one-electron wave functions, Equation (2.2.7)

$$\Psi(\vec{r}) = \prod_{i=1}^{N_e} \psi_i(\vec{r}_i) \quad (2.2.7)$$

and a sum of one electron orbital energies. These solutions are the spatial components of electron wave functions, but spin  $b(w)$  and  $b(w)$  must also be considered. One electron spin orbitals,  $\phi_i(x)$ , molecular orbital, MO, describe the spatial and spin components, Equation (2.2.8)

$$\phi_i(x) = \psi_i(\vec{r}) \alpha(\omega) = \psi_i(\vec{r}) \beta(\omega) \quad \text{or} \quad (2.2.8)$$

These one electron MO spin orbitals are constructed from linear combinations of atomic orbital, AO,  $\chi_\mu$ . The atomic orbitals are linear combinations of Gaussian Type Orbital (GTOs) (e.g. due to computational expediency) basis functions, as shown in Equation (2.2.9),

$$\phi_i(x) = \sum_{\mu} a_{\mu i} \chi_{\mu}(x) \quad (2.2.9)$$

$$\chi_l^m(x) = N Y_l^m(\theta, \varphi) r^l \sum_{j=1}^{N_{prim}} C_j e^{\alpha_j r^2}$$

Since electrons are Fermions, the electron wave function must change sign when the spatial and spin coordinates of any two electrons are interchanged. The Anti symmetrizer of Equation (2.2.10) corrects the one electron wave functions in this way, such that a Slater determinant is formed from  $N$  electrons occupying  $N$  spatial orbitals. Since the rows are labeled by electrons and columns by spin orbitals, interchange of any two rows institutes a sign change, thereby ensuring anti symmetry of the wave function. If two columns are equal, the determinant is zero, and no two electrons can occupy the same spin orbital simultaneously.

$$\Phi(\vec{r}) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\vec{r}_1)\alpha(1) & \phi_1(\vec{r}_1)\beta(1) & \phi_2(\vec{r}_1)\alpha(1) & \phi_2(\vec{r}_1)\beta(1) & \dots & \phi_2^n(\vec{r}_1)\alpha(1) & \phi_2^n(\vec{r}_1)\beta(1) \\ \phi_1(\vec{r}_2)\alpha(2) & \phi_1(\vec{r}_2)\beta(2) & \phi_2(\vec{r}_2)\alpha(2) & \phi_2(\vec{r}_2)\beta(2) & \dots & \phi_2^n(\vec{r}_2)\alpha(2) & \phi_2^n(\vec{r}_2)\beta(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_1(\vec{r}_i)\alpha(i) & \phi_1(\vec{r}_i)\beta(i) & \phi_2(\vec{r}_i)\alpha(i) & \phi_2(\vec{r}_i)\beta(i) & \dots & \phi_2^n(\vec{r}_i)\alpha(i) & \phi_2^n(\vec{r}_i)\beta(i) \\ \phi_1(\vec{r}_j)\alpha(j) & \phi_1(\vec{r}_j)\beta(j) & \phi_2(\vec{r}_j)\alpha(j) & \phi_2(\vec{r}_j)\beta(j) & \dots & \phi_2^n(\vec{r}_j)\alpha(j) & \phi_2^n(\vec{r}_j)\beta(j) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_1(\vec{r}_n)\alpha(n) & \phi_1(\vec{r}_n)\beta(n) & \phi_2(\vec{r}_n)\alpha(n) & \phi_2(\vec{r}_n)\beta(n) & \dots & \phi_2^n(\vec{r}_n)\alpha(n) & \phi_2^n(\vec{r}_n)\beta(n) \end{vmatrix} \quad (2.2.10)$$

We can now re-introduce the exact  $r_{ij}^{-1}$  electronic repulsion term back into the Hamiltonian if we construct our final many electron wave function as a linear combination of configurations,  $\Phi_\mu$ , as shown in Equation (2.2.11),

$$\Psi = \sum_{\mu} c_{\mu} \Phi_{\mu} \quad (2.2.11)$$

There are two requirements for an “exact” solution to the Schrödinger equation. The one electron expansion of equation 2.2.9 must be complete (i.e.,  $\{\Phi_{\mu}\}$  is a basis for the one-electron Hilbert space). If this is true, then there is a complete basis set (CBS) expansion. The second requirement is that the n-electron expansion of equation 2.2.11 be complete (i.e.,  $\{\Phi_{\mu}\}$  is a basis for the many-electron Hilbert space). This is said to be a full configuration interaction expansion (FCI). Of course, the one electron and n-electron expansions must be truncated, which limits the accuracy to which we can predict the energy. It is important to approximate to the CBS limit since including higher levels of angular momentum provides flexibility in describing bonding that occurs within molecules. It is also important to approximate the FCI by incorporation of higher excitations since electrons use the virtual (unoccupied) orbitals to correlate their motions. The FCI defined in general for a finite basis set. The exact solution to the Schrödinger equation is unobtainable except for a few potentials. In the next sections, we will describe widely used approximations to the Schrodinger equation, beginning with Hartree-Fock, Møller-Plesset, Coupled Cluster, and Density Functional theories.

### The Hartree-Fock approximation and self-consistent field (SCF) method.

Unlike in the Independent Particle Approximation discussed previously, where the electrons behave entirely independently, in the Hartree-Fock [3-4] approach, each electron interacts with an “average” field from all the other electrons present. We will show that Hartree-Fock theory includes the average Coulombic repulsion between all electrons, and that since the wave function is anti symmetrized, as shown in equation 2.2.10, part of the electron-electron correlation is included for same spin electrons. This type of correlation is also known as exchange correlation. The Fock operator is not unlike the

general molecular Hamiltonian of equation 2.2.5. Using the spin orbital basis set,  $\{\phi_i\}$ , the Fock operator can be expressed as Equation 2.3.1,

$$\hat{F}(1) = \hat{h}_i + \sum_b^{N/2} 2\hat{J}_b - \hat{K}_b, \quad (2.3.1)$$

Where  $\hat{h}_i$  includes the core effects, specifically the electron kinetic energy and the electron-nuclear attraction, given in Equation 2.3.2,

$$\hat{h}_i = \frac{-1}{2} \nabla^2 - \sum_A \frac{Z_A}{r_{A1}} \quad (2.3.2)$$

$\hat{J}_b$  in Equation 2.3.1 is called the Coulomb Operator, and it describes the repulsion between all electrons. It is expressed as equation 2.3.3,

$$\hat{J}_b(1) = \left[ \int dr_2 \phi_b(2) \frac{1}{r_{12}} \phi_b(2) \right] \langle a(1) | \tilde{J}_b(1) | a(1) \rangle = \langle ab | ab \rangle \quad (2.3.3)$$

$\hat{K}_b(1)$  in Equation 2.3.1 is the exchange operator which results from the fact that the wave function is anti symmetric, since electrons are Fermions. It is defined by its action to exchange the coordinates of electrons, thereby introducing exchange correlation effects. Unlike the Coulomb operator, the exchange operator is nonlocal since its effect on cannot be defined uniquely by a discrete point in space but depends on its value throughout all space.

$$\hat{K}_b(1)\phi_a(1) = \left[ \int dr_2 \phi_b(2) \frac{1}{r_{12}} \phi_a(2) \right] \phi_b(1) = \langle a(1) | \hat{K}_b(1) | a(1) \rangle = \langle ab | ba \rangle \quad (2.3.4)$$

The orbital energies are found from the diagonalization of the Fock operator, as shown in Equation 2.3.5,

$$E_i = \langle i | \hat{F}(1) | i \rangle = \langle i | \hat{H}^{core} | i \rangle + \sum_a^{N/2} 2 \langle ab | ab \rangle - \langle ab | ba \rangle \quad (2.3.5)$$

The Hartree-Fock equations will now be derived using Lagrange's method of undetermined multipliers. Then, we expand in the AO basis set  $\{\chi_{\mu}\}$  to arrive at the Roothaan-Hall equations, and describe the iterative Self Consistent

Field (SCF) method for obtaining the Hartree-Fock molecular orbitals, energies, and properties.

We hope to optimize the set of coefficients for a trial function,  $|\phi'\rangle$ , where the trial function is defined by Equation (2.3.6)

$$|\phi_i(x)\rangle = \sum_{\mu} a_{\mu i} |\chi_{\mu}(x)\rangle \quad (2.3.6)$$

These basis functions are normalized but not necessarily orthogonal, and the overlap integral and Hamiltonian operator notations are given in Equation (2.3.7),

$$\langle \chi_{\mu} | \chi_{\nu} \rangle = S_{\mu\nu}, \langle \chi_{\mu} | \hat{H} | \chi_{\nu} \rangle = H_{\mu\nu} \quad (2.3.7)$$

We seek the set of coefficients for which  $\langle \phi' | \hat{H} | \phi' \rangle$  is a minimum. Since the coefficients are not independent, Lagrange's method of undetermined multipliers. The Lagrangian is defined in Equation (2.3.8),

$$L(c_1, \dots, c_N) = \langle \phi' | \hat{H} | \phi' \rangle - E (\langle \phi' | \phi' \rangle - 1) = \sum_{\mu\nu} a_{\mu} a_{\nu} H_{\mu\nu} - E (a_{\mu} a_{\nu} S_{\mu\nu} - 1) \quad (2.3.8)$$

$$\hat{F}_{\mu\nu} = \hat{H}_{\mu\nu}^{Core} + \sum_a \sum_{\lambda\sigma} a_{\lambda a} a_{\sigma a}^* [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\lambda | \sigma\nu \rangle] = \hat{H}_{\mu\nu}^{Core} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[ \langle \mu\lambda | \nu\sigma \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right] P_{\lambda\sigma} = 2 \sum_a \sum_{\lambda\sigma} a_{\lambda a} a_{\sigma a}^* \quad (2.3.11)$$

Here,  $P_{\lambda\sigma}$ , is called the density matrix, or charge-density

Since the Fock matrix  $F$  depends on the molecular orbital coefficients, the Roothaan-Hall equations are non-linear and must be solved iteratively. This is accomplished through the Self-Consistent-Field (SCF) method. In the SCF method, an initial guess is made to the molecular orbitals and the integrals are evaluated. The Fock matrix is then constructed and diagonalized to obtain the density matrix. The density matrix is obtained from molecular orbital coefficients. They are compared to those used to generate the Fock matrix. By the Variational principle, the expectation value for the energy for any approximate wave function will always be greater than that of the exact wave function solution. If convergence is not achieved, the Fock matrix is reconstructed.

## Perturbation theory

Electrons of opposite spin also instantaneously correlate their motions, and this energy lowering effect is not included in the Hartree-Fock[3-4] approach. Møller-Plesset [5-7] (MPn) theory adds higher excitations to the Hartree-Fock through Perturbation theory as a first attempt at recovering electron correlation in a non-iterative correction. In Perturbation theory, a small perturbation is applied to

$$(H = H_0 + \lambda V)(\psi^{(0)} + \lambda \psi^{(1)} + \lambda \psi^{(2)} + \dots) = (E^{(0)} + \lambda E^{(1)} + \lambda E^{(2)} + \dots)(\psi^{(0)} + \lambda \psi^{(1)} + \lambda \psi^{(2)} + \dots) \quad (2.4.3)$$

We differentiate the Lagrangian with respect to the  $k^{\text{th}}$  coefficient and set equal to zero, as shown in Equation (2.3.9),

$$\frac{\partial L}{\partial a_k} = 0, \quad k = 1, 2, \dots, N$$

$$0 = \sum_{\nu} a_{\nu} H_{\nu k} + \sum_{\mu} a_{\mu} H_{\mu k} - E \left( \sum_{\nu} a_{\nu} S_{\nu k} + \sum_{\mu} a_{\mu} S_{\mu k} \right) \quad (2.3.9)$$

We now have the Hartree-Fock matrix equations given presented in Equation (2.3.10),

$$Ha = ESa$$

In order to arrive at the molecular orbital coefficients,  $a_{\mu i}$  of equation 2.2.9, we must employ the Roothaan-Hall equations. To do this, we work in the AO basis set of  $\{\chi_{\mu}\}$  of equation 2.2.11. When we apply the Fock operator of to the linear combination of molecular orbitals, we arrive at the Roothaan-Hall equations, as shown in equation (2.3.11),

a Hamiltonian that is known,  $H_0$ , and which is small in comparison to  $V$ , as shown in equation 2.4.1

$$H = H_0 + \lambda V \quad (2.4.1)$$

In Equation 2.4.1,  $V$  is not the potential energy, but simply the perturbed Hamiltonian. Because  $V$  is small compared to  $H_0$ , the perturbed wave function and energy can be expressed as a power series in  $V$ . Usually, this is accomplished by expanding in the parameter,  $\lambda$ , as shown in equation 2.4.2,

$$\psi = \psi^{(0)} + \lambda \psi^{(1)} + \lambda \psi^{(2)} + \dots \quad (2.4.2)$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda E^{(2)} + \dots$$

We then operate on the perturbed wave function by the perturbed operator, according to the Schrödinger equation, to give equation 2.4.3,

We can then collect equations corresponding to the powers of  $\lambda$ . After re arrangement, the first three equations corresponding to the powers of 0, 1, and 2 for  $\lambda$  are presented in equations 2.4.4, 2.4.5, and 2.4.6, respectively:

$$(H_0 - E^{(0)})\psi^{(0)} = 0 \quad (2.4.4)$$

$$(H_0 - E^{(0)})\psi^{(1)} = (E^{(1)} - V)\psi^{(0)} \quad (2.4.5)$$

$$(H_0 - E^{(0)})\psi^{(2)} = (E^{(1)} - V)\psi^{(1)} + E^{(2)}\psi^{(0)} \quad (2.4.6)$$

We define  $H_0$  as the sum of the one-electron Fock operators. Forming the inner product on both sides of equations 2.4.4-2.4.6 with  $\langle \psi^{(0)} |$  yields the MP0, MP1, and MP2 energy expressions. We will derive each in turn,

$$\langle \psi^{(0)} | H_0 | \psi^{(1)} \rangle - E^{(0)} \langle \psi^{(0)} | \psi^{(1)} \rangle = E^{(1)} \langle \psi^{(0)} | \psi^{(0)} \rangle - \langle \psi^{(0)} | V | \psi^{(0)} \rangle + E_0 \langle \psi^{(0)} | \psi^{(1)} \rangle - E^{(0)} \langle \psi^{(0)} | \psi^{(1)} \rangle = E^{(1)} \langle \psi^{(0)} | \psi^{(0)} \rangle - \langle \psi^{(0)} | V | \psi^{(0)} \rangle \quad 2.4.8$$

$$E^{(1)} = \langle \psi^{(0)} | V | \psi^{(0)} \rangle$$

Since  $H_0$  is Hermitian,  $\langle \psi^{(0)} | H_0 | \psi^{(1)} \rangle = \langle \psi^{(1)} | H_0 | \psi^{(0)} \rangle^* = E_0 \langle \psi^{(1)} | \psi^{(0)} \rangle^* = E_0 \langle \psi^{(0)} | \psi^{(1)} \rangle$ , and since the wave functions are orthonormal, the entire left side of equation 2.4.8 goes to zero, and the first order MP1 correction is the expectation value for the perturbation. It is important to note that since the entire Hamiltonian is given by equation 2.4.1, the Hartree-Fock energy is the sum of the zero and first order corrections, as given by equation 2.4.9,

$$E_0 + E^{(1)} = \langle \psi^{(0)} | H_0 | \psi^{(0)} \rangle + \langle \psi^{(0)} | V | \psi^{(0)} \rangle = E^{\text{Hartree-Fock}}$$

**Second-order correction:** We take the inner product of the both sides of equation 2.4.6 with  $\langle \psi^{(0)} |$  in order to obtain the second order correction, as shown in equation (2.4.10) [5],

$$\langle \psi^{(0)} | H_0 | \psi^{(2)} \rangle - E^{(0)} \langle \psi^{(0)} | \psi^{(2)} \rangle = E^{(1)} \langle \psi^{(0)} | \psi^{(1)} \rangle - \langle \psi^{(0)} | V | \psi^{(1)} \rangle + E^{(2)} \langle \psi^{(0)} | \psi^{(0)} \rangle \quad (2.4.10)$$

$$E^{(2)} = \langle \psi^{(0)} | V | \psi^{(1)} \rangle$$

Again, the left side of the Schrodingers equation is

zero since  $H_0$  is Hermitian, and since the wave functions are orthonormal, only the term containing the perturbed Hamiltonian remains. We still must find an expression for

$|\psi^{(1)}\rangle$  in terms of virtual orbitals in order to determine the

completed expression for  $|\psi^{(1)}\rangle$ . We will express  $|\psi^{(1)}\rangle$  as a linear combination of virtual orbitals, as given in equation 2.4.11,

$$|\Psi^{(1)}\rangle = \sum_s a_s^{(1)} \Psi_s^{(1)} \quad (2.4.11)$$

We then substitute into the equation 2.4.5 and take the inner product of both sides with an arbitrary virtual orbital,

and then consider the results for MP3 and MP4.

**Zerth-order correction:** We take the inner product of the both sides of equation 2.4.4 with  $\langle \psi^{(0)} |$  in order to obtain the zero<sup>th</sup> order correction, as shown in equation 2.4.7,

$$\langle \psi^{(0)} | H_0 | \psi^{(0)} \rangle = E^{(0)} \langle \psi^{(0)} | \psi^{(0)} \rangle = E^{(0)} \quad (2.4.7)$$

Since  $H_0$  was defined to be the sum of the one electron Fock operators,  $E_0$  is then the sum of the orbital energies.

**First-order correction:** We take the inner product of the both sides of equation 2.4.5 with  $\langle \psi^{(0)} |$  in order to obtain the zero<sup>th</sup> order correction, as shown in equation 2.4.8,

$|\Psi_t\rangle$ , as shown in equation 2.4.12,

$$\sum_s a_s \left( \langle \Psi_t | H_0 | \Psi_s \rangle - E^{(0)} \langle \Psi_t | \Psi_s \rangle \right) = E^{(1)} \langle \Psi_t | \Psi^{(0)} \rangle - \langle \Psi_t | V | \Psi^{(0)} \rangle \quad (2.4.12)$$

$$\sum_s a_s \left( \langle \Psi_t | H_0 | \Psi_s \rangle - E^{(0)} \langle \Psi_t | \Psi_s \rangle \right) = E^{(1)} \langle \Psi_t | \Psi^{(0)} \rangle - \langle \Psi_t | V | \Psi^{(0)} \rangle$$

We now require that  $s=t$  since this is the only condition that equation 2.4.12 is nonzero, allowing us to solve for the coefficient,  $a_t$  of the arbitrary virtual orbital, presented in equation 2.4.13,

$$a_i = \frac{\langle \Psi_i | V | \Psi^{(0)} \rangle}{E^{(0)} - E_i} \quad (2.4.13)$$

The coefficients of  $|\psi^{(1)}\rangle$  are now known, and we can substitute for  $|\psi^{(1)}\rangle$  in Equation 2.4.10 in order to obtain the second order energy expression, as given in equation 2.4.14,

$$E^{(2)} = - \sum_t \frac{\langle \Psi^{(0)} | V | \Psi_t \rangle \langle \Psi_t | V | \Psi^{(0)} \rangle}{E_t - E^{(0)}} \quad (2.4.14)$$

The numerator and the denominator will always be positive quantities and since there this is multiplied by a factor of negative one, the MP2 energy is energy lowering. The total second order energy can also be expressed as the sum of individual electron-electron pair energies, as shown in equation (2.4.15) [5],

$$E_0^{(2)} = \sum_{i < j} e_{ij}^{(2)} = \sum_{i < j, a < b} \frac{\langle ij | H | ab \rangle \langle ab | H | ij \rangle}{E_i + E_j - E_a - E_b} \quad (2.4.15)$$

This second order energy corresponds to a double excitation of electrons in their ground state orbitals to virtual orbitals in order to correlate their motions. The second order correction does allow for simultaneous electron pair excitations to the same virtual orbital, which is physically impossible. A correction for this is made by the third order correction.

**Third-order correction:** The third-order Møller-Plesset energy is derived similarly to the second order, presented previously, and is given by equation (2.4.16) [6].

$$E_0^{(3)} = \sum_{ij} \sum_{kl} \left\{ \sum_{abcd} \frac{\langle ij | H | ab \rangle \langle ab | H^{(1)} - E^{(1)} | cd \rangle \langle cd | H | ij \rangle}{\Delta e_{ij}^{ab} \Delta e_{kl}^{cd}} \right\} \quad (2.4.16)$$

While the second order energy expression allows for unphysical simultaneous pair excitations to the same virtual orbital, the third order energy eliminates this problem.

**Fourth-order correction:** The fourth order Møller-Plesset energy includes triple and quadruple excitations, and is given by equation (2.4.17) [7].

$$E_0^{(4)} = \sum \frac{\langle \Psi_0 | H | \Psi_{ij}^{ab} \rangle \langle \Psi_{ij}^{ab} | H^{(1)} - E^{(1)} | \Psi_{ijkl}^{abcd} \rangle \langle \Psi_{ijkl}^{abcd} | H^{(1)} - E^{(1)} | \Psi_{mn}^{ef} \rangle \langle \Psi_{mn}^{ef} | H | \Psi_0 \rangle}{\Delta e_{ij}^{ab} \Delta e_{kl}^{cd} \Delta e_{mn}^{ef}} - E^{(2)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle \quad (2.4.17)$$

### Coupled cluster singles and doubles (CCSD) method with perturbative triples (CCSD (T))

Like the MPn methods discussed previously, the coupled-cluster (CC) methodology developed by Bartlett and coworkers in the 1970's [8-12], is a *configuration interaction* (CI) approach, which asserts that the exact molecular wave function is not a single Slater determinant, as in Hartree-Fock theory. Rather, it is a linear combination of the HF determinant and all possible substituted determinants, in which virtual orbitals have replaced occupied orbitals in the determinant. The resulting electron excitations from occupied orbitals to virtual orbitals results in their strategic ability to “avoid” one another, and hence allows them to dynamically correlate their motions.

In CC theory, the molecular wave function, or exponential ansatz, is given by equation 2.5.1,

$$\Psi = e^{\hat{T}} \Phi_0 \quad (2.5.1)$$

Where  $\hat{O}_0$  is the ground state Hartree-Fock wave function, and  $\hat{T}$  is the excitation operator.  $e^{\hat{T}}$  is given as a Taylor-series expansion in equation 2.5.2,

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} \quad (2.5.2),$$

The excitation operator,  $\hat{T}$ , is defined in equation 2.5.3,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_n \quad (2.5.3)$$

Where n is the number of electrons excited.  $\Psi'$  is defined by equation 2.5.4,

$$\hat{T}_1 \Phi_0 = \sum_{a=n+1}^{\infty} \sum_{i=1}^n t_i^a \phi_i^a \quad (2.5.4)$$

Where  $\phi_i^a$  is the singly excited Slater determinant, and  $t_i^a$  is the coefficient (or amplitude) we wish to determine. The  $\hat{T}_2$  expression is defined in equation 2.5.5,

$$\hat{T}_2\Phi_0 = \sum_{b=a+1}^{\infty} \sum_{a=n+1}^{\infty} \sum_{j=i+1}^n \sum_{i=1}^{n-1} t_{ij}^{ab} \phi_{ij}^{ab} \quad (2.5.5)$$

Where  $t_{ij}^{ab}$  is the doubly excited Slater determinant, and we wish to find the amplitude,  $t_{ij}^{ab}$ . The limits of the sums in Equations 2.5.4 and 2.5.5 are chosen such that all possible single and double excitations, respectively, are included once. The operator  $\hat{O}_0$  takes a linear combination

$$\Psi^{CCD} = e^{\hat{T}_2}\Phi_0 = \left(1 + \hat{T}_2 + \frac{\hat{T}_2^2}{2} + \dots\right)\Phi_0 = \Phi_0 + \sum_{b=a+1}^{\infty} \sum_{a=n+1}^{\infty} \sum_{j=i+1}^n \sum_{i=1}^{n-1} t_{ij}^{ab} \Phi_{ij}^{ab} + \frac{1}{32} \sum_{i,j,k,l} t_{ij}^{ab} t_{kl}^{cd} \Phi_{ijkl}^{abcd} + \dots \quad (2.5.6)$$

$a,b,c,d$

Where  $\hat{T}_2$  is the HF Slater determinant, and  $\hat{T}_2$  is defined in Equation 2.5.5. The goal is to find the  $\hat{T}_1$  amplitudes of Equation 2.5.5. It is noteworthy to point out that the third term contains quadruple excitations. Indeed, the CCD includes contributions from the quadruple, hexuple, and higher order excitations into the ground state wave function, but they are expressed as products of simultaneous doubles. To improve the CCD approximation, we include  $\hat{T}_1$  so that  $\hat{T} = \hat{T}_1 + \hat{T}_2$ . This is the coupled cluster singles and doubles [9-12] (CCSD) approximation. The CCSD wave function is given by equation 2.5.7,

$$\hat{T} = \hat{T}_1 + \hat{T}_2$$

$$e^{\hat{T}_1 + \hat{T}_2} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \left(\frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_2\hat{T}_1^2 + \frac{1}{24}\hat{T}_1^4\right) + \dots \quad (2.5.7)$$

$$\Psi = C_0 \Phi_{HF} + \sum_i^{occ} \sum_a^{vir} C_i^a \Phi_i^a + \sum_{i<j}^{occ} \sum_{a<b}^{vir} C_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{i<j}^{occ} \sum_{a<b}^{vir} C_{ijk}^{abc} \Phi_{ijk}^{abc} \dots$$

The  $\hat{T}_1$  term generates the single excitations, while the  $(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2)$  term generates the double excitations. The triple and quadruple excitations are generated by  $(\hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3)$  and  $(\frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_2\hat{T}_1^2 + \frac{1}{24}\hat{T}_1^4)$ , respectively. Calculation of the CCSD energy requires only calculation of the double excitation amplitudes,  $t_{ij}^{ab}$ , but due to coupling between excitations, calculation of  $t_{ij}^{ab}$  requires calculation of the single,  $t_{ijk}^{ab}$ , triple,  $t_{ijk}^{abc}$ , and quadruple excitations amplitudes,  $t_{ijkl}^{abcd}$ . However, the quadruples and higher excitations are formed from the products of simultaneous doubles. The coefficients are determined by projecting the Schrödinger's equation on the left with the N-electron configurations generated by the truncated  $\hat{T}$  operator.

of the Hartree-Fock  $\hat{O}_0$  determinant, as well as all possible excitations. If we were also using a complete one-electron basis set, the coupled cluster approach would yield an exact solution to the Schrodinger equation. First of all, the basis set will be truncated, and this will result in a source of error. In the coupled-cluster doubles [8] (CCD) method, we make an additional approximation that the most important excitations are the double excitations from occupied to

virtual orbitals, and that  $\hat{T} \approx \hat{T}_2$ . Therefore, the CCD wave function is given by equation 2.5.6,

The resulting non-linear system of equations is solved iteratively. If we allow  $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$  in the  $e^{\hat{T}}$  expression, we include triple excitations. This is the coupled cluster singles, doubles, and triples (CCSDT) method. However, CCSDT is very computationally demanding. An approximation is the CCSD(T)[13] method, in which non iterative perturbative triples are added to the CCSD energy.

### Quadratic configuration interaction singles and doubles (QCISD) method with perturbative triples (QCISD(T))

The non-variational Quadratic Configuration Interaction [13-16] (QCI) methodology developed by Pople similar to the coupled cluster (CC) approach of Bartlett discussed previously but lacks higher order single excitation terms (e.g  $\hat{T}_1^3$ ). The QCI and CC techniques are nearly identical, but the QCI contains only the quadratic terms necessary to make it size-consistent. In most cases, the difference between QCISD and CCSD is negligible, but in species such as  $O_3$  for which single excitations are important, CCSD (T) is a significant improvement over QCISD (T). The CCSD [14] energy can be written in the form of Equation 2.6.1,

$$E_{CCSD} a_{ij}^{ab} = \langle \Phi_{ij}^{ab} | \hat{H} | (e^{\hat{T}_1 + \hat{T}_2} \Phi_0) \rangle \quad (2.6.1)$$

Where  $t_{ij}^{ab}$  are the double excitation amplitudes,  $t_{ij}^{ab}$ .  $\Phi_{ij}^{ab}$  is the doubly excited Slater determinant, and  $\Phi_0$  is the Hartree-Fock ground state Slater determinant.

The QCISD adds a quadratic term to the CISD that is necessary for size-consistency. The QCISD energy is expressed in Equation 2.6.2,

$$E_{QCISD} a_{ij}^{ab} = \langle \Phi_{ij}^{ab} | \hat{H} | (1 + \hat{T}_1 + \hat{T}_2 + \frac{\hat{T}_2^2}{2}) \Phi_0 \rangle \quad (2.6.2)$$

One can also include triple excitations to obtain the Quadratic Configuration Interaction Singles, Doubles, and Triples (QCISDT) energy, but like the CCSDT it is too cost-inefficient. The most commonly used approximation is the QCISD(T) [15,16], in which non-iterative perturbative triples are added to the QCISD energy.

### Density functional theory

The traditional, wave function-based electronic structure theory approaches involve solving approximations to the Schrödinger equation for an unknown wave function, as discussed previously. In 1964, Hohenberg & Kohn [17] suggested that the problem of finding an  $n$ -electron wave function, which is complex because it involves calculation of a function of  $3N$  variables could be simplified by considering the total electron density as the primary variable. In this formulation, the ground state of the system is defined by that electron density distribution which minimizes the total energy. Of course, this is the familiar variational principle. In their famous proof of existence theorem, Hohenberg and Kohn show, *reductio ad absurdum*, that the electron density uniquely determines the Hamiltonian, and thus all properties of the system. We begin by assuming that there exist two

different external potentials,  $V_{ext}$  and  $\rho(r)$  that give rise to the same electron density  $\rho(r)$ , where the electron density is defined in Equation 2.7.1,

$$\rho(\vec{r}) = \int |\Psi|^2 d\tau \quad (2.7.1)$$

and  $\Psi$  is the ground state wave function. By the variational theorem, the ground state energy,  $E$  must be less than the expectation value for the energy of a trial wave function,  $\Psi'$ , chosen such that it is not also the exact wave function of the unperturbed system,  $V_{ext}'$ .  $\Psi'$  corresponds to the ground state wave function when  $V_{ext}'$  is applied (i.e.,

$E_0'$  is the ground state energy when  $\Psi'$  is applied to  $\Psi'$ ). Equation 2.7.2 shows this variational condition,

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \quad (2.7.2)$$

$$\hat{H} = \hat{T} + \hat{V}_{elec} + \hat{V}_{ext}$$

$$\hat{H}' = \hat{T} + \hat{V}_{elec} + \hat{V}'_{ext}$$

Where  $\hat{H}$  is the Hamiltonian involving  $V_{ext}$  and  $\hat{H}'$  involves  $V_{ext}'$ .  $\hat{T}$  is the kinetic energy of the system. Substituting in for  $\hat{H}$  and  $\hat{H}'$  into Equation 2.7.2, we have equation 2.7.3,

$$E_0 < E_0' - \int \rho(\vec{r}) \{V_{ext}' - V_{ext}\} d\vec{r} \quad (2.7.3)$$

We then consider interchanging positions of the potentials, to form Equation 2.7.4,

$$E_0 < E_0' - \int \rho(\vec{r}) \{V_{ext}' - V_{ext}\} d\vec{r} \quad (2.7.4)$$

If we subtract Equations 2.7.3 from 2.7.4, we arrive at Equation 2.7.5,

$$E_0 + E_0' < E_0' + E_0 \quad (2.7.5)$$

or

$$0 < 0$$

Equation 2.7.5 is obviously contradictory, and therefore we have shown that there cannot exist two different external potentials that produce a unique ground state electron density. Therefore, the electron density is uniquely connected to the Hamiltonian for a system, and hence if we know the ground state electron density uniquely, the ground state energy, and all other properties are defined, as well. Note that the Hohenberg-Kohn theorem of 1964 only offers proof of existence for a unique functional of the electron density capable of describing molecular properties. It does not help us define the form of this functional. Actually, density functional theory has been around almost since the birth of quantum mechanics, and the earliest attempts began well before the Hohenberg-Kohn theorems. Some of the most fundamental work was done by Thomas and Fermi in 1927, then Slater's Hartree-Fock exchange approximation in 1951, and in 1965 the Kohn-Sham approach paved the way for more modern functionals by partitioning the electronic energy into several terms. The local density approximation (LDA) of the uniform electron gas approximation was employed initially to approximate to the electron exchange-correlation component, and modern functional use more sophisticated forms for this term. These early attempts and some popular current functional will be discussed, but first we must examine an important concept of density functional theory. Density functional themselves are mathematical models of "hole functions." For example, Slater's approximation to the Hartree-Fock exchange makes assumptions about one of the two types of hole functions, the exchange hole. Therefore, brief mention will be made here to define the terminology. If we examine an electron that is spin-up, the Pauli exclusion principle dictates that other spin-up electrons in the vicinity are repelled since no two electrons with the same quantum number can occupy

the same region of space. Therefore, spin-down electrons are unaffected. The spin-up electron under consideration is surrounded by a region deficient in like spin electrons, and therefore this region is positively charged due to the abundance of background positive charge, referred to as jellium A, exchange hole [18] (also sometimes called a Fermi hole) is said to have been created, and it has a positive charge. The boundaries of the hole must be modeled mathematically. This exchange hole correlates to ensuring the anti symmetry of the wave function. There is a second type of hole, which is called the Coulomb hole [18], which accounts for Coulomb repulsion and also correlation effects. In any given region around an electron, there will be fewer electrons surrounding it due to electrostatic repulsion, and instantaneous correlation effects.

The Thomas-Fermi Model [19] of 1927 is presented mainly because it is of historical interest. Here, the kinetic energy is expressed as a functional of the electron density, and the nuclear-electron attraction and electron-electron repulsion are treated classically. The Thomas-Fermi Model expression for the energy in atomic units is given in Equation 2.7.6,

$$E_{TF} = \frac{3}{10} \left( \frac{3\pi^2}{8} \right)^{2/3} \int \rho^{5/3}(\vec{r}) d\vec{r} - Z \int \frac{\rho(\vec{r})}{r} d\vec{r} + \frac{1}{2} \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r} \quad (2.7.6)$$

Where the first term describes the kinetic energy, the second term describes the electron-nuclear attraction, and the third term describes the electron repulsion.

In 1951, Slater [20] approximated to the Hartree-Fock exchange by mathematically modeling the exchange hole. Slater assumed the exchange hole would be spherically

symmetric and centered about the reference electron,  $\rho(\vec{r})$ . It was further assumed that the exchange hole density was uniform and equal but opposite in sign to the electron

density of that reference electron,  $\rho(r_{ij})$ . The radius of the sphere is given in equation 2.7.7,

$$r_s = \left( \frac{3}{4\pi} \right)^{1/3} \rho(\vec{r}_i)^{-1/3} \quad (2.7.7)$$

The approximate energy expression for the exchange energy is then given by equation 2.7.8,

$$E_x = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \int \rho(\vec{r}_i)^{4/3} d\vec{r} \quad (2.7.8)$$

The work of Kohn & Sham [21] of 1965 was pivotal for extending the applicability of density functional theory to calculations of atoms and molecules. In the Kohn-Sham approach, the energy of the system is partitioned into components, as in Equation 2.7.9,

$$E = E^T + E^V + E^J + E^{XC} \quad (2.7.9)$$

where  $E^T$  is the kinetic energy of the electrons,  $E^V$  is the potential energy arising from the nuclear-electron attraction,  $E^{XC}$  is the electron-electron Coulomb repulsion, and  $E^{XC}$  accounts for the exchange-correlation energy (arising due to the anti symmetry of the wave function), as well as the instantaneous correlation of all electrons.

The total energy is given by Equation 2.7.10,

$$E = \sum_i E_i + \sum_i \left[ -\frac{1}{2} \langle \psi_i | \nabla^2 | \psi_i \rangle - \sum_i \sum_A \frac{Z_A}{r_{iA}} |\psi_i|^2 + \frac{1}{2} \sum_i \sum_j \iint |\psi_i|^2 \frac{1}{r_{12}} |\psi_j|^2 d\vec{r}_1 d\vec{r}_2 + E^{XC} \right] \quad (2.7.10)$$

Kohn and Sham proved it possible that the form of the kinetic energy is the same as for Hartree-Fock theory, such that we have a non-interacting N-electron reference system, where the ground state wave function,  $\psi_0$ , is given by a single Slater determinant as in Equation 2.2.10. It is important to note in Equation 2.7.10 that the  $E^{XC}$  term is undefined and unknown in the Kohn-Sham approach!! Otherwise, it would be in principle exact! Kohn and Sham did demonstrate that  $E^{XC}$  is a functional of only the electron density. The goal of modern Density Functional Theory is to better approximate to the exchange-correlation functional that would produce this energy term.

The Kohn-Sham operator can be defined as the Hartree-Fock kinetic energy term plus an effective non-local potential,  $V_s(r)$ , and it is presented in Equation 2.7.11

$$\hat{f}^{KS} = -\frac{1}{2} \nabla^2 + V_s(\vec{r}) \quad (2.7.11)$$

The potential differs from the Hartree-Fock potential in that the density resulting from the summation of the of the Kohn-Sham orbital densities exactly equals the ground state density of the system. The Kohn-Sham approach is similar then to the self-consistent-field technique seen previously. First, the initial Kohn Sham orbitals are obtained by constructing the Slater determinant from an initial guess. This allows determination of the electron density. We solve the integrals of equation 2.7.10, and reconstruct the Slater determinant. We then apply the Kohn-Sham operator of equation 2.7.11 and if the energy not converged, we must repeat these steps. Today, Peter Pulay's DIIS algorithm is used. Usually the exchange-correlation functional is approximated as the sum of an exchange functional,  $E^X$  and a correlation functional,  $E^C$  [18]. Purely "local functionals" depend only on the electron density, while gradient corrected functionals (generalized gradient-correction approximation,

GGA) depend also on its gradient. One early addition to the Kohn-Sham technique was the introduction of the local density approximation (LDA) as the  $E^{XC}$  term. This uses the known exchange potential and correlation energy of the homogeneous electron gas with density equal to the local density of the true system (which is inhomogeneous). The LDA exchange energy is given by equation 2.7.12

$$E_{LDA}^X = -\frac{3}{2} \left( \frac{3}{4\pi} \right)^{1/3} \int \rho^{4/3}(\vec{r}) d^3\vec{r} \quad (2.7.12)$$

This  $\rho^{4/3}(\vec{r})$  dependence was also derived independently in Slater's approximation to the Hartree-Fock exchange.

In 1988, Becke & coworkers [22] formulated their gradient corrected exchange functional based on the LDA functional. This functional is presented in equation 2.7.13,

$$E_{B88}^X = E_{LDA}^X - \gamma \int \frac{\rho^{4/3} s^2}{1 + 6s\beta \sinh^{-1} s} d^3\vec{r} \quad (2.7.13)$$

$$s = \frac{|\nabla\rho(\vec{r})|}{\rho^{4/3}}$$

$$E_{B3LYP}^{XC} = E_{LDA}^X + c_0 (E_{HF}^X - E_{LDA}^X) + c_x \Delta E_{B88}^X + E_{VWN3}^C + c_C (E_{LYP}^C - E_{VWN3}^C) \quad (2.7.17)$$

The  $c_i$  parameters were obtained by non-linear least squares fitting to the atomization energies, ionization energies, proton affinities, and electron affinities in the G2 test set.

The B3LYP functional and other hybrid functionals such as PBE1PBE [30] and BPW91 are known to give reliable thermal chemistry, to an accuracy of 5kcal/mol, and are also good for obtaining geometries of stable molecules (e.g., reactants and products). They are known to underestimate barrier heights of chemical reactions. In addition, these density functionals do not incorporate dispersion forces correctly as they do not decay as  $\frac{1}{r^6}$ . Several new functionals have attempted to remedy these problems, and several examples are discussed in Chapter 5: A Spherical Atom Model for Dispersion Forces.

### The complete basis set extrapolation

Previous sections have focused on specific techniques which have been designed to cost-effectively incorporate higher levels of configuration in the N-electron molecular orbital wave function,  $\Psi^{(M-\Sigma_i, \delta^m)}$ , in order to better approximate to the full configuration interaction (FCI). In other words, we were focused on improving the level of theory. Recall that there are two requirements for achieving an exact solution to the Schrödinger's equation, and the second

Here,  $\beta$  was determined by least squares fit to exactly known energies of rare gas atoms for helium through radon [23-26]. This functional was designed to recover the

exchange energy density asymptotically as  $-\frac{1}{r}$

A popular correlation functional was developed by Lee et al. [27]. It is often paired with the Becke88 exchange correlation functional, and is presented in equation 2.7.16,

$$E_{LYP}^C = -a \int \frac{1}{1+d\rho^{-1/3}} \left\{ \rho + b\rho^{-2/3} \left[ C_F \rho^{-5/3} - 2t_w + \left( \frac{1}{9}t_w + \frac{1}{18}\nabla^2\rho \right) \right] \right\} e^{-c\rho^{-1/3}} d\vec{r} \quad (2.7.16)$$

The correlation functional by Colle & Salvetti [28] in which the correlation energy density is expressed in terms of the total electron density and a Laplacian of the Hartree-Fock density matrix was revised to include only the total electron density,  $c_F$ , and local-kinetic energy density, as well as gradient corrections. The a,b,c and  $C_F$  parameters are obtained from a fit to He atom, and  $t_w$ , is the Thomas-Fermi kinetic energy density discussed previously. Hybrid density functionals include some of the exact exchange correlation of Hartree-Fock theory.

The popular B3LYP [29] energy is given by equation 2.7.17,

one is that the one-electron basis function expansion of the molecular orbitals in Gaussian type orbitals must also be complete, to give the complete basis set limit. Of course, a basis set must be truncated somewhere, and the demands of the computational system can introduce sizable basis set truncation errors. Therefore, it is necessary to have a means of reliably extrapolating to the complete basis set (CBS) limit.

For a Helium-like atom with infinite nuclear charge, Schwartz [31] has shown that the contribution to the second order Møller-Plesset energy from the angular momentum component,  $\ell$ , is approximated by Equation 2.8.1,

$$\lim_{\ell \rightarrow \infty} \Delta E_{\ell}^{(2)} \approx \frac{-45}{256} (\ell+a)^{-4} \quad (2.8.1)$$

$$Z \rightarrow \infty$$

Where  $a = \frac{1}{2}$  but can be varied as this will not change the asymptotic behavior of the angular momentum expansion. Equation 2.8.1 shows that as s, p, d, and f orbitals are added to spherical atoms, the angular dependence goes as  $\ell^{-4}$ .

Petersson & Nyden [32,33] extended the two-electron Helium-like expression to N-electron atoms. They note that the Møller-Plesset second order pair correlation energy can be written as a sum of pair energies, as previously discussed in Section 14 (see Equation 2.4.15). The asymptotic convergence of the angular momentum is given by Equation 2.8.2

$$\Delta E_l^{(2)} = \frac{-45}{256}(l+a)^{-4} + \frac{-225}{1024}(l+a)^{-6} + \dots \quad (2.8.2)$$

Pair natural orbitals (PNO's) are the orbitals used in CBS extrapolations because those corresponding to filled shells fall on a straight line given by Equation 2.8.3

$${}^{a\beta}e_{ij}^{(\infty)}(CBS) = {}^{a\beta}e_{ij}^{(\infty)}(N) - \left[ \sum_{a=1}^N C^{a_i, a_j} \right]^2 \frac{225}{4068} |S_{ij}|^2 (N_{ij} + \delta_{ij})^{-1} |S_{ij}| = \int |\phi_i \phi_j| d\tau \quad (2.8.3)$$

Here, the sum of the CI coefficients is the interference factor, and describes higher order effects beyond MP2. The absolute over lab overlap factor  $|S_{ij}|$  accounts for the attenuation of inter orbital electron pair energies compared to intra orbital pair energies. Figure 1 shows a CBS extrapolation for Helium atom, in which the circles are pair natural orbitals. The 1s, 1s alpha-beta intra orbital pair energy is plotted as a function of the second term in Equation 2.8.3, and therefore the CBS limit is the y-intercept. The

circles are the pair natural orbital energies, and filled circles correspond to filled shells. Notice that these fit exactly to Equation 2.8.3 as a straight line. Petersson and co-workers have developed a family of methods based on this technique including the CBS-4M [34,35], CBS-q [36] and CBS-Q [37], CBS-QB3 [38] and CBS-APNO [38]. Currently, Petersson and Ericka Barnes are developing the CBS-Wes1p, which involves extrapolation of the triples component of CCSD(T).

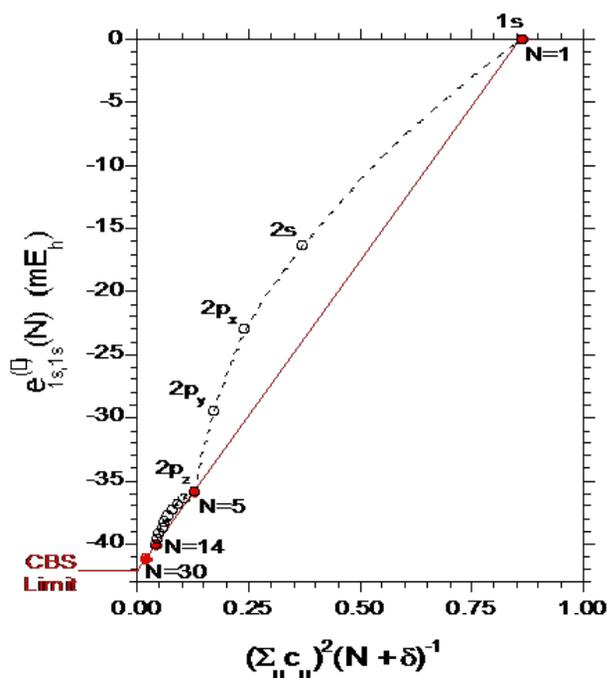


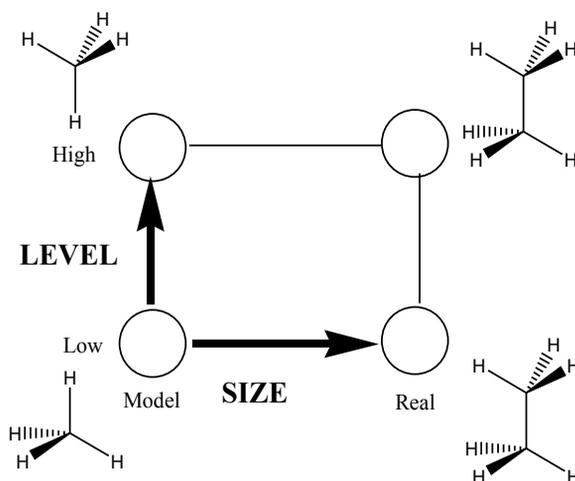
Figure 1: CBS extrapolation for Helium Atom.

### The ONIOM hybrid method

The onion-like ONIOM model of Morokuma & coworkers [39-45] is a general hybrid scheme in which a molecule or complex (called the “real system”) can be partitioned into layers which are treated with different model chemistries. In a three layer ONIOM partition, the highest level (“high level”) of correlation treatment is used on the layer in which bond-formation or breaking is taking place (“high model layer”). An intermediate level of theory (“medium level”) is used to describe a second layer (“intermediate model

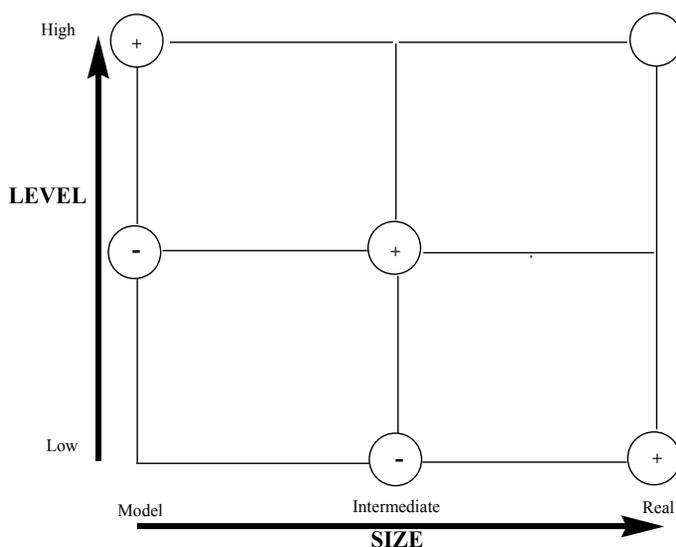
layer”) to describe electronic effects on the high model layer. The lowest level of theory (“low level”) describes environmental effects on the high model layer and is used on the remainder of the system (“low model system”). For a two layer ONIOM partition of a molecule or complex (“real system”), the same scheme is adopted, in which a high level of theory (“high level”) is utilized to describe the model portion (“model layer”) where the chemistry is occurring. The remainder of the molecule is treated with a lower level of theory (“low level”). The ONIOM approximation to the energy is given in Equation 2.9.1,

$$E(\text{ONIOM}) = E(\text{low level, real system}) + E(\text{high level, model system}) - E(\text{low level, model system})$$



For a three-layer ONIOM calculation, the energy is given by equation 2.9.2,

$$E(\text{ONIOM}) = E(\text{low level, real system}) - E(\text{low level, intermediate system}) + E(\text{medium level, intermediate system}) - E(\text{medium level, model system}) + E(\text{high level, model system}) \quad (2.9.2)$$



In the ONIOM method, “linker” hydrogens are placed along the bonds where the partitions are made.

## Force field (molecular mechanics, MM) methods

Electronic structure methods solve the electronic Schrödinger's equation which is a parametric function of the nuclear coordinates, as in the Born-Oppenheimer approximation, discussed previously. In the Force Field, also called Molecular Mechanics (MM) methods, this quantum mechanical treatment is replaced with writing the electronic energy,  $E_{MM}$ , as classical functions of the nuclear positions.  $E_{MM}$  is partitioned into five major terms, as shown in Equation (2.10.1),

$$E_{MM} = E_{\text{stretch}} + E_{\text{bend}} + E_{\text{torsion}} + E_{\text{disp}} + E_{\text{electrostatic}} \quad (2.10.1)$$

Where the  $E_{\text{stretch}}$  is the bond stretching function,  $E_{\text{bend}}$  provides the energy for bending,  $E_{\text{torsion}}$  estimates the energy for rotating through a dihedral,  $E_{\text{disp}}$  estimates the dispersion energy, and  $E_{\text{electrostatic}}$  provides the non bonded electrostatic attractions between partially positive and negative regions within a molecule. The function for calculating the stretching energy is often taken to be the classical harmonic oscillator, given in Equation (2.10.2).

$$E_{\text{stretch}}(R^{AB} - R_0^{AB}) = \frac{1}{2} k^{AB} (R^{AB} - R_0^{AB})^2 \quad (2.10.2)$$

Where  $R^{AB}$  is the stretching force constant for a bond between atoms A and B.  $R_0^{AB}$  is the distance between the two atoms, and  $R_0^{AB}$  is the equilibrium distance between them.  $R_0^{AB}$  and  $R_0^{AB}$  are parameters which are unique to specific MM programs. While the Amber force [46] field uses the expression of Equation 2.10.2, in other MM programs, such as the Goddard and co-worker's Universal Force Field [47] (UFF) Molecular Mechanics program, a Morse potential is substituted for the stretching function. The Morse potential is given in Equation (2.10.3)

$$E_{\text{stretch}}^{\text{Morse}}(R^{AB} - R_0^{AB}) = D [1 - e^{-\alpha(R^{AB} - R_0^{AB})}]^2 \quad (2.10.3)$$

$$\alpha = \sqrt{\frac{k}{2D}}$$

D is the dissociation energy and  $\alpha$  is a function of the stretching constant and dissociation energy.

The energy for bending through an angle formed by atoms A-B-C is given in Equation (2.10.4)

$$E_{\text{bend}}(\hat{e}^{ABC}) = \frac{1}{2} k^{ABC} (\hat{e}^{ABC} - \hat{e}_0^{ABC})^2 \quad (2.10.4)$$

Again, the harmonic expression is used, and  $\hat{e}_0^{ABC}$  is the angle between atoms A-B-C,  $\hat{e}_0^{ABC}$  is the equilibrium bond angle, and  $k^{ABC}$  is the bending force constant. Two parameters are entered for this function,  $\hat{e}_0^{ABC}$  and  $\hat{e}_0^{ABC}$ .

The torsional energy function,  $E_{\text{torsion}}$ , estimates the energy for rotation about an A-B-C-D dihedral angle. The torsion function is written as a Fourier series in the angle of rotation, as shown in Equation (2.10.5).

$$E_{\text{torsion}}(\theta) = \sum_{a=1}^n \frac{V_a}{2} \cos(a\theta) \quad (2.10.5)$$

$V$  is a parameter that determines the size of the barrier to rotation. Three terms are usually included in the torsional energy function, as given in Equation (2.10.6).

$$E_{\text{torsion}}(\theta) = \frac{1}{2} V_1^{ABCD} [1 + \cos(\theta)] + \frac{1}{2} V_2^{ABCD} [1 - \cos(2\theta)] + \frac{1}{2} V_3^{ABCD} [1 + \cos(3\theta)] \quad (2.10.6)$$

The third term alone is the torsional energy function for rotation of ethane through any H-C-C-H dihedral angle, in which minima exist at 60 and 180 degrees between terminal hydrogens, as seen in Figure 2.

If one considers rotation of n-butane through the C-C-C-C dihedral angle through 360 degrees, the two gauche conformers have different rotational barriers than the anti configuration. This can be accounted for by adding in the first term of Equation (2.10.6). Three parameters enter into the torsional energy function,  $V_1^{ABCD}$ ,  $V_2^{ABCD}$ , and  $V_3^{ABCD}$ .

The dispersion energy function,  $E_{\text{disp}}$ , is usually taken as the Lennard-Jones potential, and can be written as Equation (2.10.7)

$$E_{LJ}(R) = \epsilon \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^6 \right] \quad (2.10.7)$$

Where  $\epsilon$  is the well-depth,  $R_0$  is the equilibrium bond distance, and R is the inter nuclear separation. The UFF method utilizes an exponential repulsive term, which is in agreement with how orbitals decay.

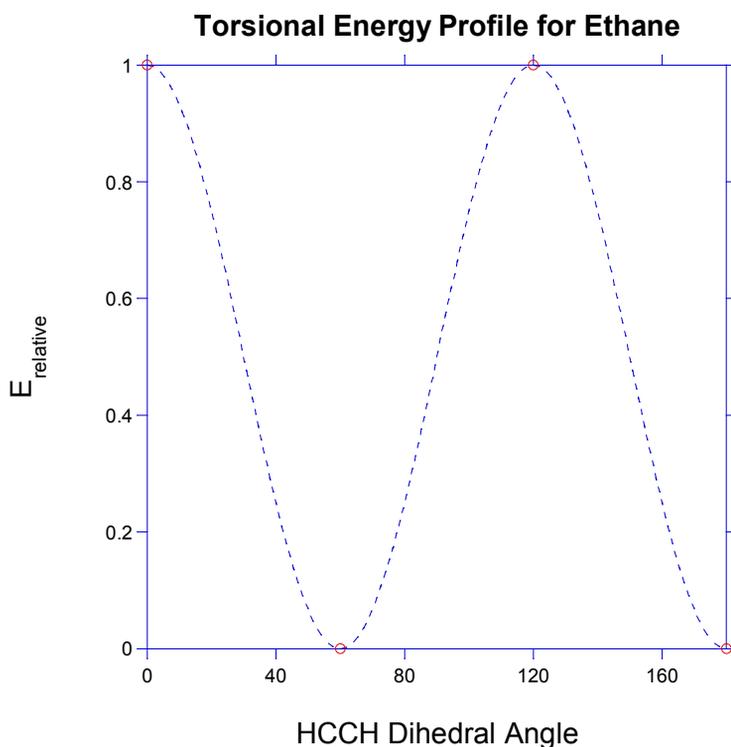
The Coulombic attraction between permanent dipoles within molecules are estimated with the electrostatic energy function, given in Equation (2.10.8)

$$E_{\text{electrostatic}}(R^{AB}) = \frac{Q_A Q_B}{\epsilon R^{AB}} \quad (2.10.8)$$

The charges are obtained by fitting to electronic structure potential energy calculations,  $\epsilon$  is the dielectric constant, and  $Q_A$  is the internuclear distance. The  $Q_A$ ,  $Q_B$ , and  $E_{MM}$  values are parameters which must be supplied.

The  $E_{MM}$  electronic energy is a measure of the steric energy of the molecule compared to no steric interaction. When the heat of formation parameters are replaced by strainless heat of formation parameters, the  $E_{MM}$  electronic energy becomes a measure of the strain in the molecule. The advantage of molecular mechanics over electronic structure methods is the drastically reduced computational expense. For large molecules, the time for calculating  $E_{MM}$

is quadratic with the number of atoms. However, generality is forfeited for computational speed. Both the functional forms used to calculate the components of  $E_{MM}$ , as well as the parameters used limit the ability of Molecular Mechanics methods to accurately describe molecular interactions and properties for structures different from those used to obtain the parameters.



**Figure 2:** Torsional Energy Potential for Rotation through any H-C-C-H dihedral angle.

# Chapter 2

## An Overlap Criterion for Selection of Core Orbitals

### Introduction and background

The core-valence partition is employed almost universally in qualitative descriptions of chemical bonding, and also provides one of the more common approximations helping to make quantitative calculations practical. The mixing that occurs among orbitals in the valence of atoms leading to chemical bonding does not occur for orbitals in the core set which are tightly bound and lie close to the nucleus. Therefore, one approximation for improving computational expense would be to neglect the core electrons in the correlation treatment. The usual approach for partitioning molecular orbitals into valence and core sets involves an energy criterion. The lowest energy molecular orbitals are selected for the core. The number of such core orbitals is set equal to the sum of the core sizes for the constituent atoms. The use of an energy criterion for the selection of molecular core orbitals may result in the inversion of molecular orbitals with respect to the atoms, and thus can create inconsistency between the treatment of orbitals in the atoms and in the molecule. Recent studies reveal that large errors in atomization energies are obtained when methods that correlate only the valence electrons are applied to certain molecules containing third-row elements. In their examination of gallium fluorides, Bauschlicher, Melius, and Allendorf report that G2 atomization energies disagree dramatically with experiment [48]. Table 1 shows that the difference between theory and experiment is 40 kcal/mol for GaF [49], growing to 108.0 kcal/mol for GaF<sub>3</sub> [50]. This error greatly exceeds the 2 kcal/mol error typical of the G2 [51] and CBS-QB3 [52,53] approaches. Thus, it is necessary to examine carefully the problems contributing to these effects and present a general solution that can be implemented in systematic model chemistries. It is the goal of this work to develop an overlap criterion for the selection of core orbitals in molecular systems that is based on mathematical similarity to their constituent atom core orbitals.

**Table 1:** Summary of atomization energies for GaF<sub>n</sub> in kcal/mol. Geometries were calculated at the B3LYP/6-31G\* level.

Method	E (GaF)	E (GaF <sub>2</sub> )	E (GaF <sub>3</sub> )
G2	97.2	135.7	231.6
Exp	138.6	224.6	339.6
ΔE	41.4	88.9	108

### Results and discussion: an energy criterion versus an overlap criterion for the selection of molecular core orbitals

The usual approach for partitioning molecular orbitals into valence and core sets involves an energy criterion. The lowest energy molecular orbitals are selected for the core. The number of such core orbitals is set equal to the sum of the core sizes for the constituent atoms. An alternative to this usual treatment is to employ an energy ceiling, above which all orbitals are included in the valence set.

However, it is evident from Figure 3 in which the occupied atomic orbital energies for the elements hydrogen through lanthanum are plotted that there is no one energy gap separating core orbitals from valence orbitals. A second alternative would require the user to select which molecular orbitals are excluded from the correlation space. This hand selection of molecular orbitals would not provide systematic model chemistry and is also not feasible for the general user. There are two fundamental problems pertinent to the use of the frozen-core approximation based on orbital energies in systematic model chemistries. Since the sum of the atomic core sizes provides the number of molecular orbitals to freeze in the usual approach to the energy criterion, the first problem is which definition to adopt for the constituent atomic cores. Three reasonable definitions include: principal quantum number [54], noble gas core, and inner noble gas core. Table 2 shows the core-valence definitions for selected atoms in this study. <sup>a</sup>In accord with standard G2 theory, potassium has the  $1s^2 2s^2 2p^6$  in the atomic core, while the  $3s^2 3p^6 4s^1$  are valence orbitals [52]. To be consistent with the majority of examples in the literature, our first core definition employs the variant of the principal quantum number definition used in the G2 model. Under the principal quantum number definition, the highest shell is taken to be valence while shells with lower principal quantum number are frozen [55]. In order to evaluate the effects of including only the 3d orbitals in the correlation space for third-row main group element containing molecules, our second definition freezes the noble gas core. The noble gas core definition freezes the highest rare gas configuration. Finally, our third definition freezes only the inner noble gas core, with the intention of including the principal core correlation effects. Figure 4 & 5 provide visual illustrations for two examples of the three atomic core definitions for gallium and potassium, respectively. Figure 4 shows that for gallium, in the principal quantum number definition, shown in red, the  $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$  orbitals are frozen while the  $4s^2 4p^1$  orbitals are in the valence set. The noble gas core definition, shown in blue, does not include the 3d electrons in the atomic valence, and the inner noble gas core, shown in red, freezes only the  $1s^2 2s^2 2p^6$  electrons. Figure 4 shows that in accord with standard G2 theory, potassium has the  $3s^2 3p^6 4s^1$  orbitals in the atomic core, while the  $3s^2 3p^6 4s^1$  are valence orbitals under the principal quantum number definition and the inner noble gas core definition. The noble gas core definition freezes the  $1s^2 2s^2 2p^6 3s^2 3p^6$  orbitals while including the  $4s^1$  in the atomic valence set. Figure 5 shows an example of the Core-Valence definition for Potassium Atom.

A second problem with the use of an energy criterion is the appropriate identification of molecular core orbitals for molecules compared to atoms containing third-row main group elements. Here, inversion of molecular orbitals with respect to the atoms can create inconsistency between treatment of orbitals in the atoms and in the molecule. Figure 6 shows a plot of the lowest energy atomic valence orbital and highest energy atomic core orbital through lanthanum. It is clear that an energy criterion for the core-valence partitioning of the gallium fluorides will include the fluorine

$2s$  atomic valence orbital in the molecular core while placing one of the gallium  $3d$  atomic core orbitals in the molecular valence space. Other effects from implementing the energy criterion [56-59] are intrinsic effects such as orbital mixing and dispersion effects, which will also be discussed. The highest-energy atomic valence orbital (i.e. filled shell) to lanthanum is the cesium  $5p$  orbital. All atomic orbitals in this range are shown.

As an example of such orbital inversion, Figure 7 shows that when the principal quantum number definition is employed for the partitioning of molecular orbitals of gallium fluoride an energy criterion freezes the first fifteen molecular orbitals since on Ga fourteen orbitals are frozen and on the F atom 1 orbital is frozen. The fifteen lowest energy occupied molecular orbitals of GaF include the fluorine  $2s$  (molecular orbital 1), the gallium  $1s^2 2s^2 2p^6 3s^2 3p^6$  (molecular orbitals 1-10), the fluorine  $2s$  (molecular orbital 11), and four of the gallium  $3d$  orbitals (molecular orbitals 12-16). The highest energy gallium  $3d$  orbital is thus forced into the valence. However, all five  $3d$  orbitals are frozen in the gallium atom, leading to inconsistencies. These inconsistencies give rise to errors of 41.4kcal/mol for this case and grow to 108.0kcal/mol for GaF<sub>3</sub>. We will show that use of an overlap criterion for the selection of molecular core orbitals will allow the molecular orbital corresponding to the fluorine  $2s$  to be recognized as a valence orbital and the molecular orbital corresponding to the gallium  $3d$  orbital to be recognized as a core orbital based on their similarity to known atomic solutions. It has been suggested that the problems with the gallium fluorides can be resolved by including the  $3d$  orbitals of the third-row main-group elements in the valence space. Table 3 shows the errors in atomization energies for MP2 (FC)/6-311+G (2df,p) compared to MP2(full)/6-311+G(2df,p). This core definition gives more reliable thermo chemistry as the range in error decreases from 1.93-4.11 kcal/mol to 0.61-1.07 kcal/mol when the  $3d$  orbitals of Ga are included in the correlation treatment. However, this remedy creates a new inversion problem if the energy criterion is used to select molecular core orbitals. Figure 8 shows that under the energy criterion and a noble gas core definition, the fourteen lowest energy molecular orbitals of sodium bromide are frozen. These orbitals are the bromine  $1s^2$  (molecular orbital 1), the sodium  $2s^2 2p^6 3s^2 3p^6$  (molecular orbital 2), the bromine  $2s^2 2p^6 3s^2 3p^6$  and four of the  $3d$  orbitals (molecular orbitals 3-14). The sodium  $2p^6$  and  $2p^6$  orbitals are included in the valence shell of sodium bromide. However, these orbitals are treated as core in the sodium atom. The deviation induced from substituting the sodium  $2s$  and  $2p$  orbitals for four of the bromine  $3d$  orbitals in the correlation space is 20.79 kcal/mol compared to 0.13 kcal/mol when an overlap criterion is used and consistent orbitals are frozen.

Figure 9 shows that these energetic inversions are mirrored in the selection of core molecular orbitals for potassium fluoride. On the application of an inner noble gas core definition, an energy criterion freezes the five lowest

energy molecular orbitals of potassium fluoride. These include the potassium  $2s^2$  (molecular orbital 1), the fluorine  $2s^2$  (molecular orbital 2), the potassium  $2s^2$  (molecular orbital 3), and two of the potassium  $2p$  orbitals (molecular orbitals 4 and 5). The remaining potassium  $2p$  orbital is included in the correlation; however, all three  $2p$  orbitals are treated as core in the potassium atom. As demonstrated in Table 4, this inconsistency gives rise to a deviation of 5.82kcal/mol compared to 0.16 kcal/mol when the overlap criterion properly identifies all three potassium  $3p$  orbitals as valence. It is clear that there are no truncations for the core-valence partition in which orbital inversion does not pose a problem when an energy criterion is used.

Our goal is to identify which orbitals in a molecular system are most like the known core orbitals of the isolated atoms. We present an overlap criterion for the selection of molecular core orbitals, in which identification of orbitals in a molecular system are chosen based on similarity to the known constituent atomic solutions. The core space is formed from the union of known atomic cores. The molecular orbital eigenvectors are each projected onto this core space by a symmetric orthogonalization projector  $P^{core}$ , given in Equation (3.2.1),

$$P^{core} = OO^T S \quad (3.2.1),$$

Here,  $P^{core}$  is formed from  $O$  core orbitals having coefficients  $O$  where  $u$  goes from 1 to  $N$  (the number of atomic orbital basis functions in the molecule),  $S$  is the overlap of atomic orbitals in the molecular system with core orbitals of constituent atoms, and  $O$  is defined by Equation (3.2.2),

$$O = O(O^T S O)^{1/2} \quad (3.2.2).$$

The core-valence mixing given by the projected norm squared from the overlap integral of the projector should be close to either one or zero after projection depending on whether the orbital is assigned to the core or valence space, respectively. Cases of large core-valence mixing which persist after projection have been identified and will be discussed as a complication intrinsic to  $3d$  orbitals. The overlap based core-valence partition was tested against several published cases for which the energy criterion selection is known to fail, including: GaF, GaF<sub>2</sub>, GaF<sub>3</sub>, GaO, GaOH, HGaO, GeF<sub>2</sub>, KF, and NaF. Table 4 shows the total energies of the molecules and atoms calculated at the MP2 level of theory with all electrons correlated and the energy changes arising from the use of each of the three frozen core definitions along with either the energy or overlap criterion for selection of the core orbitals.

The 6-311+G (2df,p) basis and MP2(full)/6-31G\* optimized geometries were used throughout. The resulting

MP2 (full) atomization energies were compared with MP2 (FC) approximations. The non-relativistic MP2 (full)/6-311+G (2df,p) model is not an adequate approximation to experiment but we use it as a benchmark for the corresponding frozen core calculations. Table 3 shows the errors in atomization energies for MP2 (FC)/6-311+G (2df,p) compared to MP2(full)/6-311+G(2df,p). All calculations were performed with a developmental version of Gaussian 01 [60].

Revisiting the problem of the gallium fluorides, the use of an energy criterion for partitioning the molecular orbitals results in deviations from full correlation atomization energies calculations of 41.60 kcal/mol for GaF to 100.99 kcal/mol for GaF<sub>3</sub>, as shown in Table 3. The use of an overlap criterion for selecting core orbitals removes the inconsistency leading to these errors. Projection of the fluorine 2s molecular orbital onto the space spanned by the atomic core orbitals reveals that this orbital is valence-like (norm squared after projection=0.038) and it is permuted to the valence of gallium fluoride. The gallium 3d orbital is identified as core (norm squared after projection = 0.956) and not included in the correlation space. The result is a reduction of the error from 41.60kcal/mol to 2.28 kcal/mol. Similarly, two and three gallium 3d orbitals are included in the correlation space for GaF<sub>2</sub> and GaF<sub>3</sub>. For the NaBr case, we saw that the sodium 2s<sup>2</sup> and 2p<sup>6</sup> orbitals are included in the valence shell of sodium bromide. However, these orbitals are treated as core in the sodium atom. The deviation induced from substituting the sodium 2s and 2p orbitals for four of the bromine 3d orbitals in the correlation space is 20.79 kcal/mol compared to 0.13 kcal/mol, as shown in Table 3, when an overlap criterion is used and consistent orbitals are frozen.

Even with too large of a correlation space, we have shown that an energy criterion cannot sufficiently describe KF, since the potassium 2p orbital is included in the correlation space of the molecule. However, all three 2p orbitals are treated as core in the potassium atom. As demonstrated in Table 3, this inconsistency gives rise to a deviation of 5.82kcal/mol compared to 0.16kcal/mol when the overlap criterion properly identifies all three potassium 3p orbitals as valence. Even with consistent selection of core and valence orbitals, a definition which includes too large a core space can lead to substantial errors in calculated energy differences. The gallium oxides provide an example of such errors. As observed in the gallium fluorides cases, orbital inversion was found to be present between one of the gallium 3d orbitals and the 2s orbital of oxygen for both GaO and GaOH. When the principal quantum number definition is applied using an energy criterion, one of the gallium 3d orbitals is substituted for the oxygen 2s orbital in the correlation space of GaO and GaOH. Table 3 shows that the errors in these atomization energies are 24.76 and 39.66 kcal/mol, respectively. As with the gallium fluorides, the inversion effects are corrected by employing an overlap criterion, with which the errors in atomization energy fall to

15.60 and 9.03 kcal/mol, respectively. Table 3 also shows that HGaO, which has no orbital inversions, exhibits an atomization energy error of 19.04kcal/mol. Clearly, in these cases sizeable errors remain even after errors due to inversion have been remedied. These large errors arise due to significant mixing between one of the gallium 3d orbitals and the oxygen 2s orbital of the gallium oxides, molecular orbitals eleven and sixteen having substantial contributions from both. When the principal quantum number definition is applied, the largest core-valence mixings for GaO, GaOH, and HGaO are 23.8%, 15.5%, and 24.4%, respectively, and are presented in Table 4 & 5.

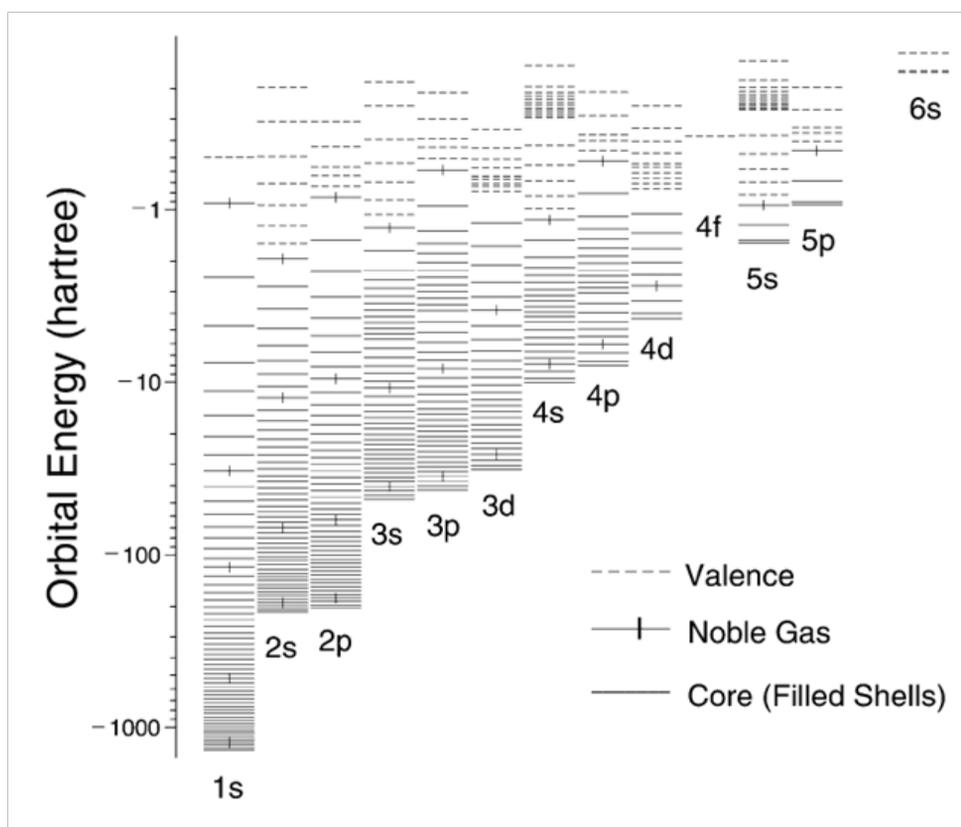
In contrast to the gallium fluorides, the gallium 3d orbitals of the gallium oxides do not remain essentially atomic-like in the molecule. Rather, the gallium 3d orbitals participate in the bonding. The chemistry inherent in the 3d-2s hybrid orbitals of the gallium oxides cannot be physically removed, and necessitates their inclusion in the correlation treatment. Table 3 shows that when a noble gas core definition is applied, errors in atomization energy range from 0.25 to 0.97 kcal/mol. Germanium difluoride and sodium fluoride provide examples for cases in which effects of several kcal/mol are present when the principal quantum number definition is employed in the absence of either inversion or the dramatic mixing exhibited by the gallium fluorides and gallium oxides, respectively. The results presented in Table 3 suggest that the sodium 2s and 2p electrons have very little effect on the binding of species such as NaF because errors are only reduced by 1 kcal/mol. A more detailed analysis given in Table 6 of the components of the MP2/6-311+G (2df,p) correlation energy contribution to the atomization energy of NaF demonstrates that this is misleading.

This component of the correlation energy can be determined through the second-order pair correlation energies for NaF partitioned into intra-and interatomic (dispersion) sets. Given in red are the sodium interatomic pair energies. Given in green are the fluorine interatomic effects, and given in black are the sodium interatomic (dispersion) effects. Sodium fluoride exists as an ion pair, making a partition of the second-order correlation energy into intraatomic and interatomic pair energies unambiguous. The sodium 3s electron is transferred to the fluorine thereby markedly increasing the fluorine valence shell correlation energy, as seen in Table 7. This transfer also eliminates the sodium (core, 3s valence) correlation energy in NaF, making the difference between Na atom and the NaF molecule necessarily negative when the Na 2s and 2p electrons are correlated. The correlation energy within the sodium core is reduced in the NaF molecule compared to the atom by a similar amount. These two intraatomic effects are very nearly equal, but opposite in sign to the dispersion attraction (i.e. the sum of the interatomic pair energies) between the sodium (2s,2p) core and the fluorine (2s,2p) valence electrons, the difference between them being about 1 kcal/mol.. This result is in good agreement with the errors in atomization energy from Table 3. The

dispersion attraction in NaF is considerably larger than that for Ne<sub>2</sub> (1.52 kcal/mol) or NeF<sup>-</sup> (2.80 kcal/mol) at the same bond length. The loosely bound valence electrons of F<sup>-</sup> are more easily polarized, and the Na<sup>+</sup> ion pulls the F<sup>-</sup> electrons closer.

The dispersion and sodium intraatomic effects are each large compared to the accuracy sought in models such as CBS-QB3 [57,58] and G2 [56], and there is no reason for these errors to cancel in general. Germanium difluoride provides such an example. The dispersion effect is responsible for the sizeable core contribution to the bond

energy of GeF<sub>2</sub>. If we exclude the two bonding pairs, the remaining Ge (3d)-F (2s,2p) lone pair correlation energies account for 6.43 kcal/mol of the GeF<sub>2</sub> binding energy, comparable to the 6.36 kcal/mol atomization energy error found for this species. The covalent bonds formed by valence electrons will generally bring the cores within their van der Waals radii, making dispersion forces comparable to those of van der Waals complexes and providing a significant component of bond energies. The dispersion effects inherent to NaF and GeF<sub>2</sub> are included only by the application of the inner noble gas core definition.



**Figure 3:** Occupied atomic orbital energies for the elements from hydrogen through lanthanum.

**Table 2:** Core-Valence definitions for selected atoms.

Atom	Principal Q#		Noble Gas Core		Inner Noble Gas Core	
	Core	Valence	Core	Valence	Core	Valence
O <sub>2</sub>	Core	Valence	Core	Valence	Core	Valence
O, F	1s <sup>2</sup>	2s <sup>2</sup> 2p <sup>6</sup>	1s <sup>2</sup>	2s <sup>2</sup> 2p	None	1s <sup>2</sup> 2s <sup>2</sup> 2p
Na	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup>	3s <sup>1</sup>	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup>	3s <sup>1</sup>	1s <sup>2</sup>	2s <sup>2</sup> 2p <sup>6</sup> 3s <sup>1</sup>
Ka	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup>	3s <sup>2</sup> 3p <sup>6</sup> 4s <sup>1</sup>	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup> 3s <sup>2</sup> 3p <sup>6</sup>	4s <sup>1</sup>	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup>	3s <sup>2</sup> 3p <sup>6</sup> 4s <sup>1</sup>
Ga-Kr	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup> 3s <sup>2</sup> 3p <sup>6</sup> 3d <sup>10</sup>	4s <sup>2</sup> 4p	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup> 3s <sup>2</sup> 3p <sup>6</sup>	3d <sup>10</sup> 4s <sup>2</sup> 4p	1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>6</sup>	3s <sup>2</sup> 3p <sup>6</sup> 3d <sup>10</sup> 4s <sup>2</sup> 4p

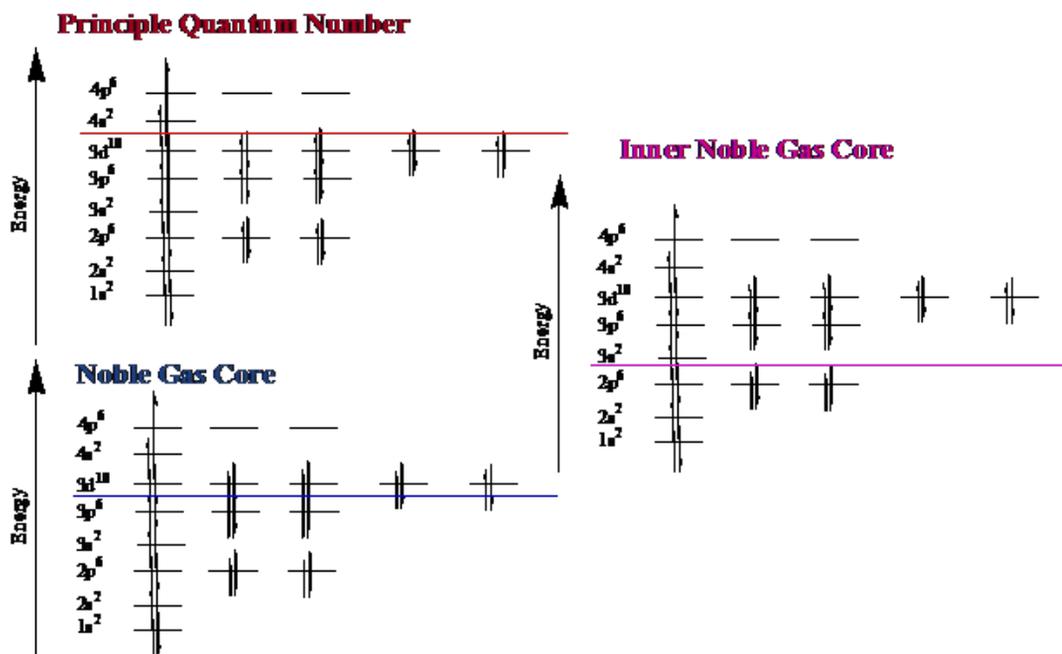


Figure 4: Example of Core-Valence Definition for Gallium Atom.

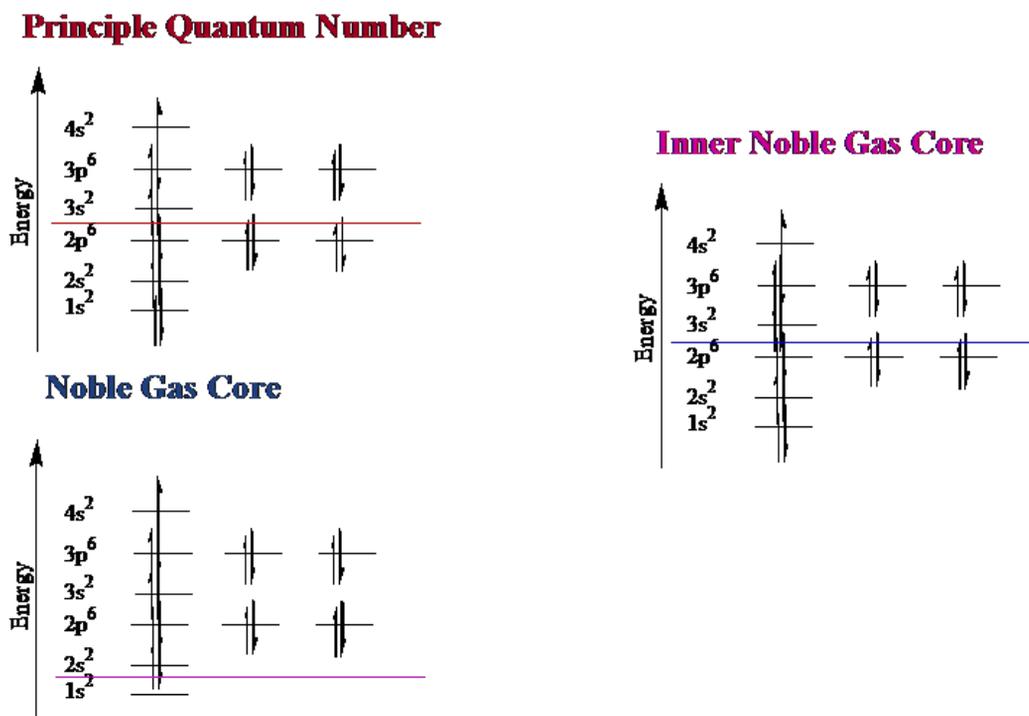


Figure 5: Example of Core-Valence definition for Potassium Atom

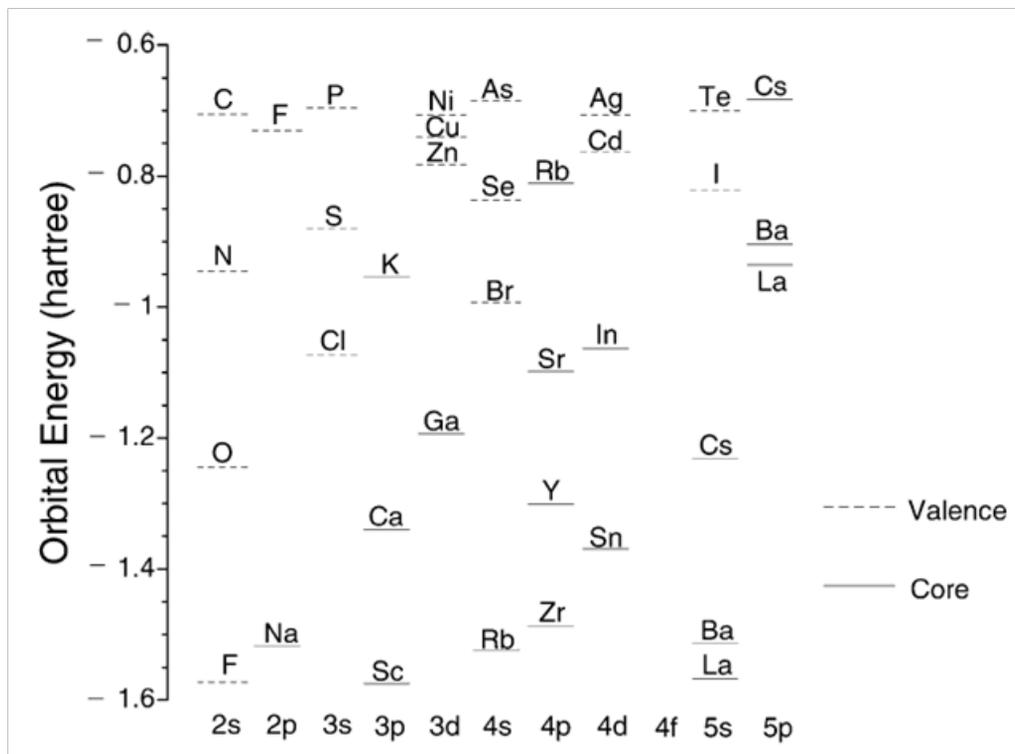


Figure 6: The lowest-energy atomic valence orbital (i.e.: partially filled shell) is the fluorine 2s orbital.

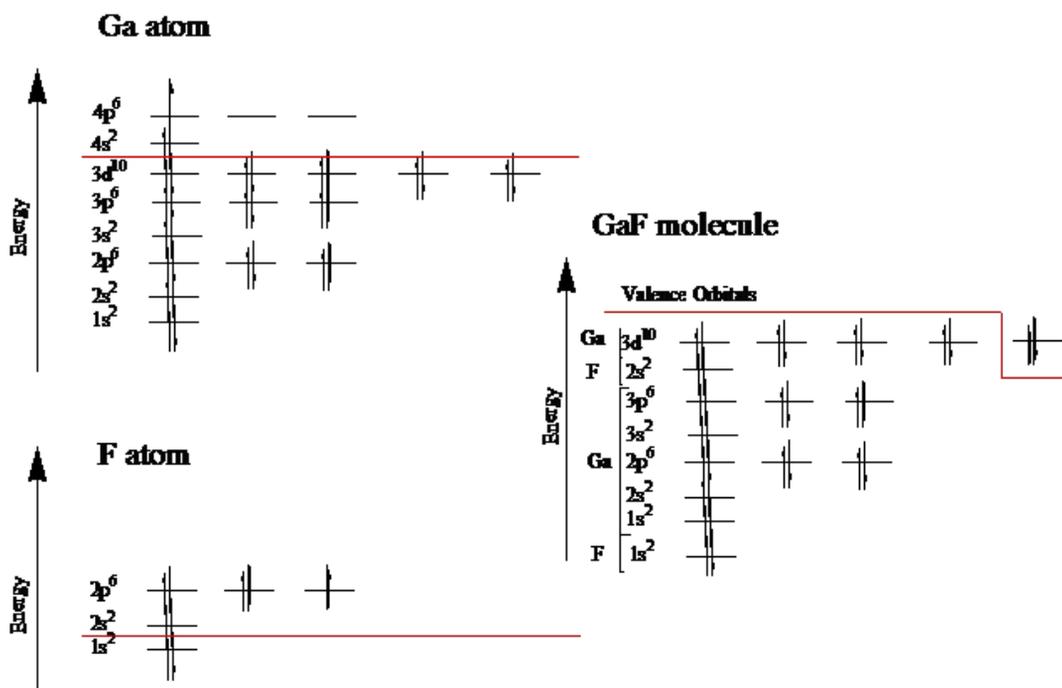
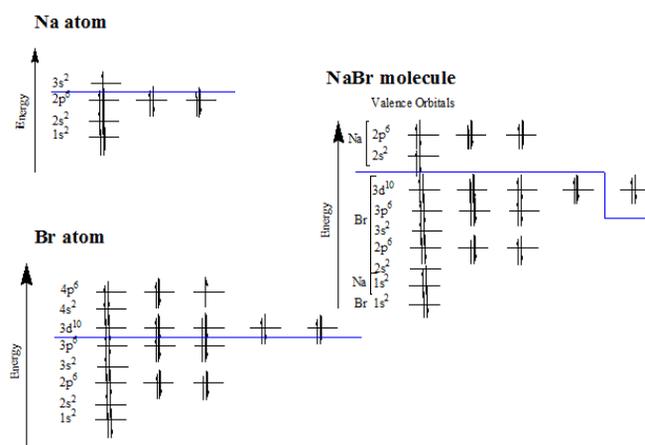
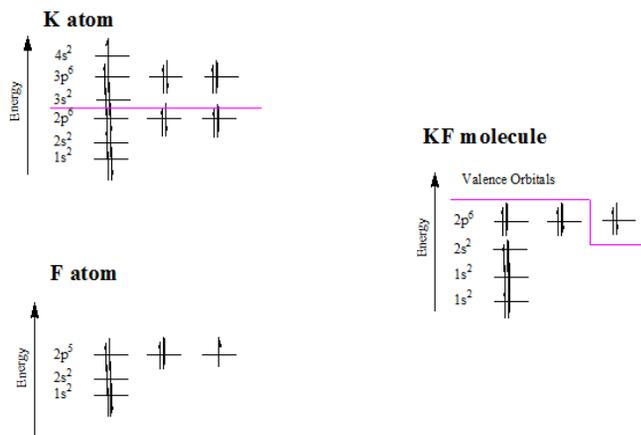


Figure 7: Example of Orbital Inversion when an Energy Criterion is employed for selecting core orbitals in GaF molecule when the Principle Quantum Number is selected for the core-valence partitioning of constituent atoms.

**Table 3:** Errors in atomization energies for MP2(FC)/6-311+G(2df,p) compared to MP2(full)/6-311+G(2df,p).

Molecule	AE (kcal/mol = E( full) -E(frozen))						E( full) (kcal/mol)
	Principal Q#		Noble Gas Core		Inner Noble Gas Core		
	Energy	Overlap	Energy	Overlap	Energy	Overlap	
GaF	41.6	2.28	0.61	0.61	0.12	0.12	147.2
GaF <sub>2</sub>	72.84	1.93	0.65	0.65	0.24	0.24	229.2
GaF <sub>3</sub>	100.99	4.11	1.07	1.07	0.35	0.35	360.9
GaO	24.76	15.6	0.77	0.77	0.15	0.15	93.23
GaOH	39.66	9.03	0.97	0.97	0.15	0.15	221.32
HGaO	19.04	19.04	0.25	0.25	0.21	0.21	184.91
GeF <sub>2</sub>	6.36	6.36	1.73	1.73	0.26	0.26	267.55
NaBr	1.25	1.25	20.79	-0.13	0.11	0.11	89.83
KF	0.38	0.38	32.4	0.44	5.82	0.16	118.73
NaF	1.15	1.15	1.15	1.15	0.04	0.04	117.44

**Figure 8:** Example of Orbital Inversion when an Energy Criterion is employed for selecting core orbitals in NaBr molecule when the Noble Gas Core definition is selected for the core-valence partitioning of constituent atoms.**Figure 9:** Example of Orbital Inversion when an Energy Criterion is employed for selecting core orbitals in KF molecule when the Inner Noble Gas Core definition is selected for the core-valence partitioning of constituent atoms.

**Table 4:** Energy differences for MP2(FC)/6-311+G(2df,p) compared to MP2(full/6-311+G2df,p).

Molecule	$\Delta E$ (Hartree) = Effroeen ) - E(full)						E( full) Lhartreesl
	Principal Q#		Noble gas care		Inner noble gas core		
	Energy	Overlap	Energy	Overlap	Energy	Overlap	
GiF	0.51300	45033	0.27655	0.27655	0.08638	0.08638	-2023.49166
GaF <sub>2</sub>	0.58384	0.47083	0.29768	0.29768	0.08657	0.08657	-2123.24192
GaF <sub>3</sub>	0.64975	49537	0.31941	0.31941	0.08675	0.08675	-2223.07147
GaO	0.48531	0.47070	0.27596	0.27596	0.08643	0.08643	-1998.75628
GaOH	0.50904	0.46024	0.27628	0.27628	0.08642	0.08642	-1999.46021
HGaO	0.47619	0.47619	0.27513	0.27513	0.08653	0.08653	-1999.40219
Ge F <sub>2</sub>	0.46644	0.46644	0.29793	0.29793	0.08838	0.08838	-2275.40609
NaBr	0.55436	0.55436	0.42090	0.38756	0.10754	0.10754	-2735.02225
K F	0.13642	0.13642	0.40655	35562	0.12402	0.11500	-699.29172
NaF	0.15911	0.15911	0.15911	0.15911	0.01521	0.01521	-261.78896
Ga	0.42563	0.42564	0.25452	0.25452	0.08619	0.08619	-1923.63746
F	0.02106	0.02106	0.02106	0.02106	0.00000	0.00000	-99.61962
O	0.02021	0.02021	0.0921	0.02021	0.00000	0.00000	-74.97025
H	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.49981
Ge	0.41417	0.41417	0.25305	0.25305	0.08796	0.08796	-2075.744147
Na	0.13621	0,13621	0.13621	0.13621	0.01515	0.01515	-161.98219
Br	0.41615	41615	0.25156	0.25156	0.09221	0.09221	-2572.89691
K	0.11475	0.11475	0.33386	33386	0.11475	11475	-599.48290

**Table 5:** Largest percentage of core-valence mixing for any occupied orbital.

Molecule	Principal Q#	Noble Gas Core	Lnnr Noble Gas Core
GaF	4.42,3.80	0.012	0.006
GaF <sub>2</sub>	9.17	0.012	0.006
GaF <sub>3</sub>	6.83	0.009	0.006
GaO	23.8	0.012	0.006
GaOH	15.5	0.012	0.006
HGaO	24.4	0.012	0.006
GeF <sub>2</sub>	6.97	0.014	0.005
NaBr	0.068	0.068	0.037
KF	0.026	3.46	0.026
NaF	2.15	2.15	0.002

**Table 6:** Moller-Plesset 6-311+G(2df,p) Second-Order Pair Correlation Energies (Hartrees) for NaF Partitioned into Intra- and Interatomic (Dispersion) sets.

	Na 1s	F 1s	Na 2s	Na 2p $\sigma$	Na 2 $\pi$	Na 2 $\pi$	F 2s	F 2p $\sigma$	F 2p $\pi$	F 2p $\pi$
Na 1s	-0.008602	0.000000	-0.000028	-0.000008	-0.000007	-0.000007	0.000000	0.000000	0.000000	0.000000
F 1s	0.000000	-0.01241	0.000000	-0.000004	0.000000	0.000000	-0.000287	-0.000380	-0.000548	-0.000548
Na 2s	-0.002070	0.000000	-0.003709	-0.001612	-0.001735	-0.001735	-0.000012	-0.000114	-0.000256	-0.000256
Na 2 $\sigma$	-0.000227	-0.000022	-0.00332	-0.008424	-0.003686	-0.003686	-0.000012	-0.000109	-0.000110	-0.000110
Na 2p $\pi$	-0.000249	0.000000	-0.003548	-0.005036	-0.009382	-0.009325	-0.000038	-0.000257	-0.000023	-0.000027
Na 2p $\sigma$	-0.000249	0.000000	-0.003548	-0.005036	-0.005353	-0.009382	-0.000038	-0.000257	-0.000027	-0.000023
F 2s	-0.000002	-0.000798	-0.000032	-0.000250	-0.000052	-0.000052	-0.018520	-0.002659	-0.005353	-0.005353
F 2p $\sigma$	-0.000014	-0.000698	-0.000187	-0.000842	-0.000320	-0.000320	-0.004470	-0.020959	-0.006860	-0.00686
F 2p $\pi$	-0.000001	-0.000632	-0.000013	-0.000272	-0.000041	-0.000030	-0.009904	-0.010770	-0.022760	-0.009623
F 2p $\pi$	-0.000001	-0.000632	-0.000013	-0.000272	-0.000030	-0.000041	-0.009904	-0.010770	-0.012214	-0.022760

**Table 7:** Components of MP2/6-311+G(2df,p) correlation energy contribution to the dissociation energy of NaF.

Correlation Orbital Set Changes in Components of Correlation Energy (Kcal/Mol)					Total
	F (val, val)	Na (core, val)	Na (core, core)	Na (core) F (val)	
FC	48.70	0.00	0.00	0.00	48.7
FC + Na (2p)	48.70	-1.91	-1.88	4.37	49.27
FC + Na (2s, 2p)	48.70	-2.09	-105	5.00	49.57
Full	48.95 <sup>a</sup>	-2.09	-2.06	5.06 <sup>b</sup>	49.86

<sup>a</sup> Includes F (core, core) and F (core, val.)

<sup>b</sup> Includes Na (core) F (core)

## Conclusion

Two fundamental issues pertinent to the use of the frozen-core approximation in systematic model chemistries are: 1) the definition of constituent atomic core orbitals and 2) the consistent identification of molecular core orbitals between molecule and constituent atoms such that orbital inversions do not arise. We have shown that no definition of the atomic core orbitals permits a consistent identification of molecular core orbitals based on orbital energies, except for the trivial case of defining all occupied orbitals as valence. In contrast, the overlap criterion presented here provides a consistent identification of core molecular orbitals for any atomic core atomic core definition based on filled shells.

We have seen that even with consistent selection of core and valence orbitals, a definition which includes a large (eg, inner noble gas core) core space can lead to substantial errors in calculated energy differences. We have identified two sources of error in thermo chemistry that are inherent:

- I. inclusion of orbitals in the core that mix significantly with valence orbitals upon bonding, and
- II. neglect of dispersion interactions between the filled shell of one atom and the electrons of neighboring atoms. Orbital mixings are known to give rise to errors in atomization energy of 10 to 20kcal/mol for the gallium oxides, and, the dispersion interactions between filled shells in covalent systems that can contribute 5kcal/mol or more to binding energies. The core-valence mixings computed during the application of the overlap criterion for the identification of core orbitals provide a useful diagnostic for the possible presence of either of these sources for systematic error. Mixings larger than 10% suggest very serious mixing between designated core and valence orbitals, and the associated errors are greater than 10kcal/mol for some of the cases considered here. Mixings larger than 1% suggest that the designated core orbitals may be providing significant dispersion contribution to the thermo chemistry of the molecule.

# Chapter 3

## A Spherical Atom Model for Dispersion Forces

### Introduction

It is well known that dispersion forces play a significant role in the structures and properties of large molecular structures, such as biological systems. Structural studies of these very large systems generally employ molecular mechanics empirical potentials, which include London dispersion terms based on point-dipole interactions. When electronic structure calculations are required, Density functional Theory (DFT) has become the method of choice for studies of large molecules since it is less computationally demanding than high-level *ab initio* methods. A detailed history, development, and some functionals in popular use today has already been presented in Chapter 1. As a reminder, however, Hohenburg & Kohn [61] point out that a wave function has both an amplitude and phase, whereas only the amplitude need be considered in Density Functional Theory. A major shortcoming of the local density approximation [62,63] (LDA) and generalized gradient approximation [64,65] (GGA) functionals with and without "exact exchange" density approximations is their failure to properly describe long range correlation effects, such as dispersion forces.

Much effort has been made to incorporate the fundamental physics of dispersion forces into GGA DFT. Recently, Hirao & coworkers [66] have applied their long-range exchange correction (LRXC) scheme [67-69] and the Andersson-Langreth-Lundqvist (ALL) van der Waals long-range correlation functional [70] to several conventional GGA functionals which account for short-range electron correlation. Qualitatively correct dissociation energies were obtained for the He, Ne, and Ar homonuclear rare gas diatomics. However, the ALL long-range functional requires six-dimensional numerical integration over pairs of points, which significantly increases the computer time required to calculate the interaction and has not yet been implemented self-consistently. In other recent work, Becke and co-workers have devised a method of calculating the  $C_6$  coefficients using DFT based on the instantaneous dipole moment of the exchange hole [71]. This method of calculating the  $C_6$  coefficient produces qualitatively correct dissociation energies for rare gas dimers. However, since the  $C_6$  coefficient is also determined from a six-dimensional integral, the method also increases the cost for DFT studies.

Other recent methods take a more empirical approach to introducing dispersion interactions into DFT (e.g., DFT-D, in which a  $C_6/R^6$  dispersion energy term is added for each pair of atoms to the Hartree Fock or DFT electronic energy) [72-76]. These methods are no more costly than standard DFT calculations, but when Wu & Yang [75] appended a damped  $C_6/R^6$  term to the B3LYP [77] functional, using  $C_6$  values obtained from molecular polarizabilities, only 40% of the binding energy of Ne dimer was recovered, and the He dimer was not bound at all. Another recent empirical attempt to incorporate van der Waals interactions into DFT was made by Goddard et al. [78,79] with their X3LYP functional, which

is linear combination of exchange functionals constructed to empirically simulate the long-range behavior of a Gaussian-type electron density. However, both the exchange functionals used in the linear combination, B88 [65] and PW91 [64], and the Lee Yang Parr [80] (LYP) correlation functional are local in nature and consequently not known to treat long-range correlation effects properly. In this work, we present a novel atom-atom pairwise potential function for describing dispersion forces based the interaction of two spherical shells of polarizable media. The primary deviation from the  $C_6/R^6 + C_8/R^8$  interaction of two point dipoles is due to the fact that atoms have volume. We will show that the resulting spherical atom potential is more accurate for describing dispersion interactions than either damped or undamped  $C_6/R^6$  potentials, while not significantly affecting the computational cost of DFT methods. The spherical atom model for dispersion forces can also be incorporated into molecular mechanics force fields.

### The spherical atom model for dispersion forces

The total interaction energy between two atoms can be partitioned into the Hartree-Fock plus the correlation energy, Equation (4.2.1)

$$V(R) = V_{\text{HF}} + V_{\text{corr}} \quad (4.2.1)$$

For the diatomic species considered in this work ( $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ ,  $\text{HeNe}$ ,  $\text{HeAr}$ , and  $\text{NeAr}$ , and  $\text{H}_2$ ,  $\text{C}_2$ , and  $\text{HC}$ )  $V_{\text{corr}}$  is the dispersion energy,  $V_{\text{disp}}$ , which is usually expressed as a multipole expansion as shown in Equation (4.2.2)

$$V_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} \dots \quad (4.2.2)$$

in which the coefficients,  $C_n$ , are related to the multi pole polarizabilities. For example, the well-known perturbation theory approximations for the first two terms are given in Equation (4.2.3)

$$C_6 \approx \frac{3}{2} \frac{I_a I_b}{(I_a + I_b)} \alpha_a \alpha_b \quad (4.2.3)$$

Where  $I$  and  $\alpha$  are the first ionization energies and dipole polarizabilities, respectively, of the interacting atoms, and  $C_8$  is given by Equation (4.2.4),  $\alpha$

$$C_8 \approx \frac{15}{2} \frac{I_a I_b}{(I_a + I_b)} (\alpha_a \alpha_b + \alpha_b \alpha_a) \quad (4.2.4)$$

Where  $\varrho_a$  and  $\varrho_b$  are the quadrupole-quadrupole polarizabilities of the interacting atoms. In the Andersson-Langreth-Lundqvist (ALL) correlation correction scheme [70], the dispersion attraction is calculated for pairs of

infinitesimal volume elements between two atoms. The ALL long-range exchange correlation energy is given by Equation (4.2.5) [70],

$$E_{xc}^{l-r} = \frac{6e}{4(4\pi)^{3/2} m^{1/2}} \int_{V_1} d^3r_1 \int_{V_2} d^3r_2 \frac{\sqrt{n_1(r_1)n_2(r_2)}}{\sqrt{n_1(r_1)+n_2(r_2)}} \frac{1}{|r_1-r_2|^6} \quad (4.2.5)$$

Where  $n(r_i)$  is the charge density of fragment I,  $e$  is the electron charge, and  $m$  is the electron mass. However, as noted previously, the ALL long-range correlation scheme involves six-dimensional integrals over pairs of points, which significantly adds to the computational demand of the DFT. If we model the dispersion interaction of two atoms as two interacting spherical shells of polarizable media, then the dispersion potential,  $V_{disp}$ , can still be expressed as a simple, algebraic function of the distance between the nuclei. A spherical atom is clearly a more realistic model than a point atom. In order to find the dispersion potential function,  $V_{disp}$ , between two atoms modeled as spherical shells, we first consider the dispersion interaction between one spherical shell of a polarizable medium and an infinitesimal surface area element on the second sphere, as depicted in Figure 10. The area element,  $dA_1$ , is located on the surface of a sphere 1 of radius  $r_1$ , and is a distance of  $t$  away from the center of sphere 2 of radius  $r_2$ .  $\rho$  is the interaction distance between  $dA_1$  on sphere 1 and another area element  $dA_2$ , on the surface of the sphere 2. The distance between the centers of the two spheres is  $R$ .

The force,  $d\vec{F}_1$ , experienced by surface element,  $dA_2$ , on sphere 2 from interaction with  $dA_1$  on sphere 1 has a magnitude that is proportional to the derivative of the familiar  $C_6/R^6$  dispersion energy as given in Equation (4.2.6)

$$\|d\vec{F}_1\| = \frac{6 * dC_6}{\rho^7} \quad (4.2.6)$$

Where  $dC_6$  is the contribution toward  $C_6$  from the two interacting surface area elements as shown in Equation (4.2.7),

$$dC_6 = C_6 \left( \frac{dA_1}{A_1} \right) \left( \frac{dA_2}{A_2} \right) \quad (4.2.7)$$

and  $C_6$  is the lowest coefficient from the multi pole expansion given in Equation (4.2.2).

The contribution of  $dA_1$  and  $dA_2$  to the dispersion attraction between the two spheres is the component  $d\vec{F}_1$  in the direction of the internuclear axis. We project  $d\vec{F}_1$  onto the inter nuclear axis by first projecting it onto  $\tau$ , the axis adjoining  $dA_2$  and the center of sphere 1. The cylindrical symmetry allows the cancellation of all components of the potential interaction not along  $\tau$  by adding the forces between pairs of points located on opposite sides of the circle swept out on Sphere 1 by rotating  $\phi_1$  from 0 to  $2\pi$

for any fixed value of  $\theta_1$  (ie: forces  $d\vec{F}_2$  and  $d\vec{F}_1$  in Figure 1). Averaging these forces produces a net force,  $d\vec{F}_{net}(a)$ , that is entirely along  $\tau$ . The magnitude of  $d\vec{F}_{net}$  is found by projecting  $d\vec{F}_1$  onto  $\tau$  as shown in Equation (4.2.8),

$$\|d\vec{F}_{net}\| = 6 \left( \frac{dC_6 * \cos \lambda_1}{\tilde{n}^7} \right), \quad (4.2.8.)$$

Where  $dC_6$  is given in Eq. (6), and  $\lambda_1$  is the angle between  $d\vec{F}_1$  (also  $d\vec{F}_{net}(a)$ ) and the net force,  $d\vec{F}_{net}(a)$ . Substituting into Equation (4.2.9) for  $\cos \lambda_1$ :

$$\tilde{n} \cos \lambda_1 = \frac{(\tilde{n} - r_1 \cos \theta_1)}{\tau} \quad (4.2.9)$$

and the interaction distance,  $\rho$ , as shown in Equation (4.2.10).

$$\tilde{n}^2 = r_1^2 \sin^2 \theta_1 + (\tau - r_1 \cos \theta_1)^2 = r_1^2 + \tau^2 - 2\tau r_1 \cos \theta_1 \quad (4.2.10)$$

We have that the magnitude of the net force,  $d\vec{F}_{net}(a)$ , is given in Equation (4.2.11).

$$\|d\vec{F}_{net}(a)\| = 6 \frac{dC_6 * (\tau - r_1 \cos \theta_1) r_1 \cos \theta_1}{\tilde{n}^8} = \frac{6 * \tau * r_1 * \cos \theta_1}{(\tau^2 + r_1^2 - 2\tau r_1 \cos \theta_1)^4} \quad (4.2.11)$$

Assuming a hollow sphere with an infinitesimally thin shell (such that there is no need for radial integration), the total force experienced by the surface area element  $dA_2$  from the entire first sphere can be obtained by integrating over the surface of sphere 1 as given in Equation (4.2.12),

$$\|F_{total(a)}\| = 6 \int_{\phi_1=0}^{2\pi} \int_{\theta_1=0}^{\pi} \frac{C_6 \left( \frac{dA_2}{A_2} \right) \left( \frac{dA_1}{A_1} \right) \tau r_1 \cos \theta_1}{(\tau^2 + r_1^2 - 2\tau r_1 \cos \theta_1)^4} r_1^2 \sin \theta_1 d\theta_1 d\phi_1 = \frac{2C_6 \left( \frac{dA_2}{A_2} \right) r_1 (5r_1^2 + 3\tau^2)}{(r_1 - \tau)^5 (r_1 + \tau)^5}$$

In order to evaluate the total force between the two spherical shells of polarizable media, we must integrate over the surface of the sphere 2. Once again, the contribution of the total force to the dispersion attraction between the two spheres is along the inter nuclear axis. Cancellation of all components of the total force not along the inter nuclear axis is accomplished by adding the forces between pairs of points (ie: forces  $d\vec{F}_{net}(a)$  and  $d\vec{F}_{net}(b)$  in Figure 1) located on opposite sides of the circle swept out on sphere 2 by rotating  $\phi_2$  from 0 to  $2\pi$  for any fixed value of  $\theta_2$ . Averaging these forces produces a net force,  $d\vec{F}_{net}(a,b)$ , that is entirely along  $R$ . The magnitude of  $d\vec{F}_{net}(a)$  can be found by taking the projection of  $d\vec{F}_{net}(a)$  onto the inter nuclear axis as given in Equation (4.2.13)

$$\|dF_{net(a,b)}\| = \frac{C_6 \left( \frac{dA_2}{A_2} \right) \left( \frac{dA_1}{A_1} \right) \left( \frac{dA_3}{A_3} \right)}{(r_1 - (r^5 + \bar{r}))^5} * \cos \lambda_2 \quad (4.2.13)$$

$$\|dF_{net(a,b)}\| = \frac{C_6 \left( \frac{dA_2}{A_2} \right) \left( \frac{dA_1}{A_1} \right) \left( \frac{dA_3}{A_3} \right)}{(r_1 - (r^5 + \bar{r}))^5} \left( \frac{R - r_2 \cos \theta_2}{\rho} \right) \quad (4.2.14)$$

where  $\lambda_2$  is the angle between  $d\vec{F}_{net}(a)$  (also  $d\vec{F}_{net}(b)$ ) and the internuclear axis. Substituting into Eq. (10) for  $\cos \lambda_2$  as given in Equation 4.2.14

Again, assuming a hollow sphere with an infinitesimally thin shell (such that there is no need for radial integration), the total force experienced by the two spheres is given by integrating over the surface of sphere 2 as shown in Equation (4.2.15)

The dispersion energy from the interaction between the two spheres is the integral of the total force in Eq. (11) with respect to the internuclear distance  $R$  as given in Equation (4.2.16)

$$V_{\text{attract}} = -C_6 \int_R^\infty F_{\text{total}}(R) dR = -C_6 \left( \frac{1}{[R^2 - (r_1 + r_2)^2]^3} \right) \left( 1 + \frac{2/3 \bar{a}}{[R^2 - (r_1 - r_2)^2]^3} \right) \quad (4.2.16)$$

$$\|F_{\text{total}}\| = 2 \int_{\phi_2=0}^{2\pi} \int_{\theta_2=0}^{\pi} \frac{C_6 r_1^2 \tau (5r_1^2 + 3\tau^2)}{A_2 (r_1 - \tau)^5 (r_1 + \tau)^5} \left( \frac{R - r_2 \cos \theta_2}{\tau} \right) (r_2^2 \sin \theta_2) d\theta_2 d\phi_2 =$$

$$C_6 \left( \frac{15R^4 + 10R^2 r_1^2 - r_1^4 + 48R^3 r_2 + 8R r_1^2 r_2 + 54R^2 r_2^2 - 2r_1^2 r_2^2 + 24R r_2^3 + 3r_2^4}{(24R^2 r_2 (R^2 - r_1^2 - 2R r_2 + r_2^2))^4} \right.$$

$$\left. \frac{15R^4 + 10R^2 r_1^2 - r_1^4 - 48R^3 r_2 - 8R r_1^2 r_2 + 54R^2 r_2^2 - 2r_1^2 r_2^2 - 24R r_2^3 + 3r_2^4}{(24R^2 r_2 (R^2 - r_1^2 + 2R r_2 + r_2^2))^4} \right)$$

Where  $\delta$  is given by Equation (4.2.17)

$$\delta(r_1, r_2, R) = 3R^4 (r_1^2 - 3r_1 r_2 + r_2^2) + 3(r_1 - r_2)^2 (r_1^4 - r_1^3 r_2 + 4r_1^2 r_2^2 - r_2^3 r_1 + r_2^4) - 2R^2 (3r_1^4 - 9r_1^3 r_2 + 10r_1^2 r_2^2 - 9r_2^3 r_1 + 3r_2^4) \quad (4.2.17)$$

In the case of homonuclear diatomics, Equation (4.2.16) reduces to Equation (4.2.18),

$$V_{\text{attract}}(r, R) = -C_6 \left( \frac{1}{[R^2 - (2r)^2]^3} \right) \left\{ 1 - \frac{2}{3} \left( \frac{r}{R} \right)^2 \left[ 3 - 4 \left( \frac{r}{R} \right)^2 \right] \right\}$$

$$= \frac{-C_6}{R^6} \left\{ \frac{1}{\left[ R^2 - \left( \frac{2r}{R} \right)^2 \right]^3} \right\} \left\{ 1 - \frac{1}{2} \left( \frac{2r}{R} \right)^2 + \frac{1}{6} \left( \frac{2r}{R} \right)^4 \right\} \quad (4.2.18)$$

$$= \frac{-C_6}{R^6} \left\{ 1 + 3 \left( \frac{2r}{R} \right)^2 + 6 \left( \frac{2r}{R} \right)^4 + \dots \right\} \left\{ 1 - \frac{1}{2} \left( \frac{2r}{R} \right)^2 + \frac{1}{6} \left( \frac{2r}{R} \right)^4 \right\}$$

The first term of the attractive potential energy function given in Equation (4.2.18) suggests that unlike the multipole expansion, the primary deviation from the  $C_6/R^6$  attraction arising from two point-dipoles is a consequence of the fact that atoms have volume. The dispersion interaction of two atoms is amplified compared to two interacting point dipoles centered at the nuclei since the electron density is closer than the internuclear distance. The first term in Equation (4.2.18) shows that the extent of this amplification is determined by the radii of the interacting atoms, and therefore is quite physically intuitive. The second term in the product of Equation (4.2.18) provides a modest correction to the first term, attenuating the attraction predicted by the first term. This correction is physically reasonable since the true interaction distance needed to correctly describe the dispersion attraction ought not to be either the internuclear distance at one extreme, or the point of closest approach of the electron density at the other extreme, but rather at an intermediate distance. If we assume a simplified attractive potential given by Equation (4.2.19),

$$V_{\text{attract}}(r_{1s}, r_{2s}, R) = \frac{-C_6}{[R^2 - (r_{1s} + r_{2s})^2]^3} \quad (4.2.19)$$

where  $r_{1s}$  and  $r_{2s}$  are the radii of the interacting spheres, then we can approximate to these radii for various atoms by comparing a Taylor series expansion of the spherical atom model potential to the multipole expansion as shown in Equation (4.2.20),

$$\begin{aligned} \frac{1}{[R^2 - 4(r)^2]^3} &= \frac{1}{R^6} * \frac{1}{\left[1 - 4\left(\frac{r}{R}\right)^2\right]^3} \\ &= \frac{1}{R^6} * \left[1 + 12\left(\frac{r}{R}\right)^2 + 96\left(\frac{r}{R}\right)^4 + \dots\right] \\ &= \frac{C_6}{R^6} + \frac{C_8}{R^8} + \dots \quad (4.2.20). \\ \Rightarrow \frac{C_8}{C_6} &= 12r^2 \\ \therefore r &= \frac{1}{2} \sqrt{\frac{C_8}{3C_6}} \end{aligned}$$

This approximation for  $r_s$  allows us to assess the importance of the error term. Using  $C_8$  and  $C_6$  values provided by Cybulski et al. [81] the spherical atom radii for  $\text{Ar}_2$ ,  $\text{Ne}_2$ , and  $\text{He}_2$  are 1.426, 1.093, and 0.897 bohr, respectively. This error term provides a modest correction to the first term, 7%, 6%, and 5% at the equilibrium geometries for  $\text{Ar}_2$ ,  $\text{Ne}_2$ , and  $\text{He}_2$ , respectively. However, Equation (4.2.20) is not the best choice for selecting the spherical atom radii for atoms since  $C_8$  values are often not known with sufficient accuracy. Instead we obtain spherical atom radii through nonlinear least squares fitting to benchmark curves for homonuclear diatomics, and predict them for hetero nuclear diatomics through combination rules. We

will show that through parameterization of the first term in Equation (4.2.18) benchmark potential curves can be simulated to high accuracy for our test systems. Therefore, we will neglect the second term of Equation (4.2.18), such that the simplified dispersion attractive potential energy function is given in Equation (4.2.19).

Using the definition of Equation (4.2.20) for the spherical atom radius,  $r_s$ , we can assess how the spherical atom model for dispersion forces compares to the multipole expansion. In Figure 11-13, we compare the ratio of the spherical atom model potential function with the ratio of the multipole expansion (truncated to one, two, three, and four terms) to high accuracy benchmark calculations of the correlation energy for He, Ne, and Ar dimers.

Figure 11-13 show that the attractive potential energy predicted by the spherical atom model is in better agreement with the three and four term multipole expansion than even the two term truncated multipole expansion. The multipole expansion attractive potential function diverges in the united atom limit, and the spherical atom model attractive potential function diverges when the two spheres touch, ie: when  $R = 2r$ . This divergence is unphysical since the correlation energy of the dimer must converge to that of the united atom limit as the internuclear separation goes to zero. Therefore, like the multipole expansion, the spherical atom model must be damped to correct for this unphysical divergence for short internuclear separations. As two atoms approach, the dispersion attraction they experience becomes attenuated due to overlapping of the electron densities. Atoms are not spherical shells with fixed radii, but rather species in which the electron density dies off exponentially since the atomic orbitals do. Therefore, there can be overlap of the electron densities even though the internuclear distance is not shorter than the spherical atom radii. The overlapped electron density is not available for dispersion. The spherical atom model does not take this effect into account, and therefore we must damp out the electron density not contributing to the dispersion.

This effect is similar to the attenuation in the  $1/R$  Coulombic attraction that a bare proton experiences as it approaches an atom, as depicted in Figure 14. The electron density of the atom dies off exponentially, and as the internuclear distance approaches zero, an increasing portion of the electron density is outside the internuclear distance,  $R$ , and therefore does not contribute to the coulombic attraction between the proton and the atom. If we subtract the portion of the electron density outside of the internuclear distance from the part of the density within the internuclear distance, we can obtain an analytic expression for the damping factor that attenuates the proton-atom Coulombic attraction, as shown in Equation (4.2.21),

$$\left( \frac{1}{R\pi} \int_0^R 4\pi r^2 e^{-2r} dr \right) - \left( \frac{1}{\pi} \int_\infty^R \frac{4\pi r^2 e^{-2r}}{r} dr \right) = \frac{1}{R} [1 - (1+R)e^{-2R}] \quad (4.2.21)$$

Where  $r$  is the radial coordinate of the atom in question, and  $a_0R$  is the internuclear separation in Bohr.

Equation (4.2.21) shows that as the internuclear distance approaches 0, the damping function for the  $1/R$  Coulombic attraction will become zero, eliminating the unphysical  $1/R^6$  divergence in the united atom limit. However, in the spherical atom model, the potential energy diverges before the united atom limit, when the two spheres touch. Therefore, we must shift divergence to  $R = 2r$ . The function we use to damp the spherical atom model is given in Equation (4.2.22),

$$\left[ 1 - \left( 1 + \frac{R-2r_d}{r_d} \right) e^{-2\left( \frac{R-2r_d}{r_d} \right)} \right] \quad (4.2.22)$$

The radius of damping is given as  $r_d$  and is a separate

quantity from  $r_s$ , the spherical atom radius, given in Equation (4.2.19). Like the spherical atom radius,  $r_s$ , we will obtain the radii of damping for homonuclear diatomics through nonlinear least squares fitting, and from them predict the  $r_d$  values for heteronuclear diatomics. Using the damping function of Equation (4.2.22), and the attractive potential energy expression of Equation (4.2.19), the spherical atom model potential function for describing the dispersion potential of two interacting atoms is given by Equation (4.2.23)

$$V^{\text{disp}}(r_1, r_2, R) = \frac{-C_6}{\left\{ R^2 - \left[ 2 \left( \frac{r_{s,1} + r_{s,2}}{2} \right) \right]^2 \right\}^3} \left[ 1 - \left( 1 + \frac{R-2r_d}{r_d} \right) e^{-\frac{R-2r_d}{r_d}} \right]^2 \quad (4.2.23)$$

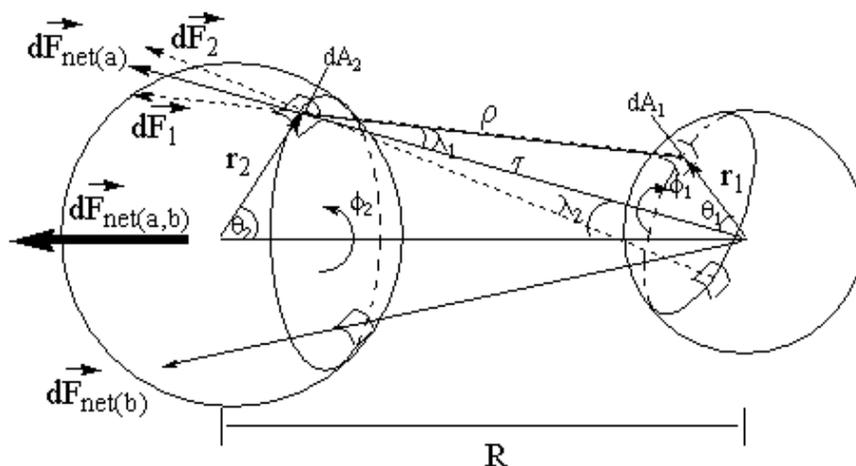


Figure 10: Dispersion interaction between a two spherical shells of polarizable media.

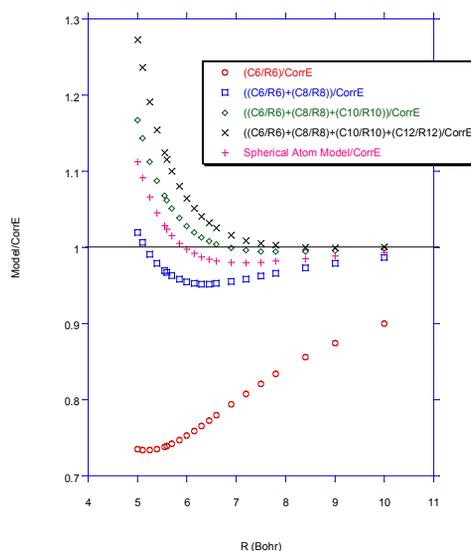
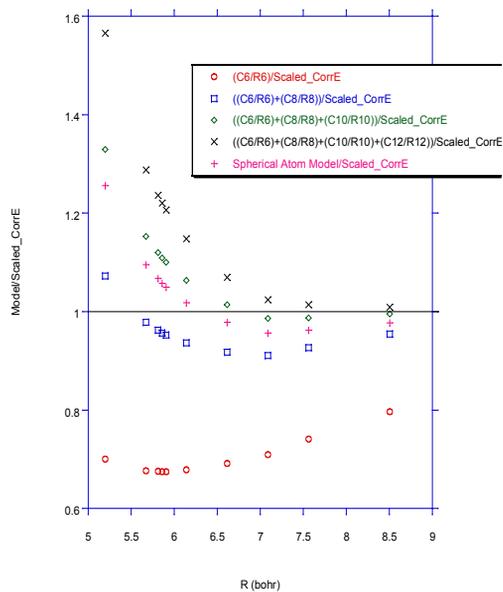
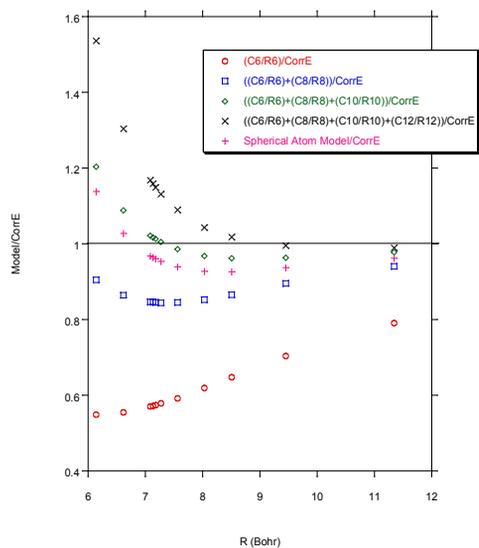


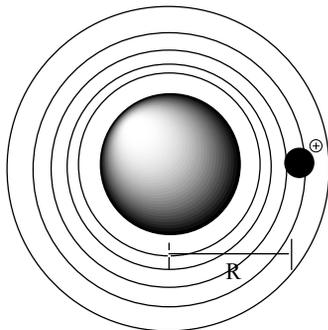
Figure 11: Ratio of one, two, three, and four term truncated multipole expansion to the correlation energy of He dimer.



**Figure 12:** Ratio of one, two, three, and four term truncated multipole expansion to the correlation energy of Ne dimer.



**Figure 13:** Ratio of one, two, three, and four term truncated multipole expansion to the correlation energy of Ar dimer.



**Figure 14:** Coulombic attraction between an atom and a bare proton.

## Comparisons with other models

**Energy comparison:** The Spherical Atom Model attractive term of Equation (4.2.19) resembles simpler potentials of the form  $C_6/R^6$  that were added originally to the Hartree-Fock energy by Alrichs and recently extended to DFT to provide an empirical correction for van der Waals interactions (ie: the HF-D and DFT-D approaches, respectively). However, rather than the amplification of the  $C_6/R^6$  energy being described in terms of atoms having volume, in HF-D, the higher order interactions beyond the  $C_6$  are introduced as a truncated three-term multipole expansion (given in Equation (4.2.2)). Figure 11-13 show that the truncated three term expansion gives quantitatively correct results for rare-gas atoms. However, the accuracy of the method is compromised due to the difficulty in obtaining accurate  $C_8$  and  $C_{10}$  values either experimentally or computationally. As discussed previously, while current DFT-D models extend the applicability of DFT to dispersion complexes with only limited success [75], we will show that the spherical atom model can be incorporated into DFT with better results. Similar functional forms for the dispersion energy have been used to describe the total van der Waals potential energy curves for atom-atom pairs in conventional molecular mechanics (MM) force fields [82], such as MM2 [83], CHARMM [84], or AMBER [85]. The van der Waals interaction energy is modeled as a 6-12 Lennard-Jones potential function given by Equation (4.3.1) [82].

$$V_{LJ} = D_0 \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^6 \right] \quad (4.3.1)$$

where the well-depth is given by  $D_0$  and  $R_0$  is the equilibrium geometry. The exponential-6 is another common functional form used in MM force fields, and is given by Equation (4.3.2) [82].

$$V_{LJ} = D_0 \left[ \frac{6}{\xi-6} e^{\xi} e^{-\xi \frac{R}{R_0}} - \frac{\xi}{\xi-6} \left( \frac{R_0}{R} \right)^6 \right], \quad (4.3.2)$$

Where the well-depth is given by  $D_0$  and  $R_0$  is the equilibrium geometry and  $\xi$  is a parameter introduced to provide flexibility.  $\xi$  is obtained through non-linear least square fitting to a reference potential.

**Damping function comparison:** In 1952, Brooks [86] pointed out that the usual multipole expansion is only valid at internuclear separations where the charge distributions do not overlap, and that the correct expansion is Equation (4.3.3),

$$V_{disp} = \sum_{n=1}^{\infty} f_{2n}(R) C_{2n} R^{-2n}, \quad (4.3.3)$$

Where  $f_{2n}(R)$  are the damping functions for the individual dispersion coefficients. The electrostatic

expansions for overlapping charge distributions can be numerically calculated for various atom-atom systems [87,88] but algebraic expressions cannot be achieved for atoms beyond hydrogen. In 1967, Musher & Amos [89] showed that the damping functions should be a product of a polynomial in the nuclear coordinate,  $R$ , times an exponential in  $R$ . The universal damping functions proposed by Tang & Toennies [90] express the damping functions as incomplete gamma functions, consistent with the ideas of Musher & Amos [89] as given in Equation (4.3.4):

$$f_{2n} = 1 - \sum_{k=0}^{2n} \frac{\left[ b \left( \frac{R}{R_m} \right) \right]^k}{k!} e^{-b \left( \frac{R}{R_m} \right)}, \quad (4.3.4)$$

Where  $R$  is the inter nuclear separation and  $R_m$  is the sum of the van der Waals radii of the two interacting atoms. Beginning with the HF-D approach of Alrichs et al. [72] attempts were made to model the dispersion in a more empirical way by assuming that each of the terms in the multi pole expansion could be damped by the same term, and therefore this damping term could be factored out as in Equation (4.3.5):

$$V_{disp} = f_{damp}(R) \sum_{n=1}^{\infty} C_{2n} R^{-2n}. \quad (4.3.5)$$

The damping factor Alrichs found to work well for his test set was Equation (4.3.6):

$$f_{damp}(R) = e^{-(1.28R_m/R-1)^2} \text{ for } R < 1.28R_m \quad (4.3.6)$$

$$= 1 \text{ for } R \geq 1.28R_m$$

Where  $R$  is the inter nuclear separation, and  $R_m$  is the sum of the van der Waals radii for the interacting atoms. His function, however, has a discontinuity, and therefore is undesirable. Several completely empirical damping functions that multiply the entire multi pole expansion were proposed [13-16] that mimic the function of Alrichs, while being everywhere continuous.

Wu & Yang [75] propose two damping functions, each damping the  $C_6/R^6$  dispersion at different rates (with Function I damping at shorter inter nuclear separations than Function II), Equation (4.3.7) and Equation (4.3.8), respectively,

$$f_d I(R/R_M) = \left( 1 - e^{-3.54 \left( \frac{R}{R_M} \right)^3} \right)^2, \quad (4.3.7)$$

$$f_d^{II}(R/R_M) = \frac{1}{1+e^{-23\left(\frac{R}{R_M}-1\right)}}, \quad (4.3.8)$$

where  $R$  is the internuclear separation and  $R_M$  is the summation of the van der Waals radii of the interacting atoms. Elstner et al. [73] propose a damping function that works well in their tight-binding method as given in Equation (4.3.9)

$$f_d(R) = \left[ 1 - e^{-3.0\left(\frac{R}{R_O}\right)^7} \right]^4 \quad (4.3.9)$$

While the damping functions above are entirely empirical and not based on any physical model, the Spherical Atom damping function of Equation (4.2.22) is based on the damping to the  $1/R$  Coulombic attraction between a bare nucleus and an atom. We note that this simple function is of the form of a polynomial in  $R$  times an exponential in  $R$ , and this is physically sound while working well in practice.

## Numerical tests

### Rare gas homonuclear diatomics with hartree-fock:

We compare the spherical atom model for dispersion forces, given in Equation (4.2.23), to accurate CCSD (T) potential curves reported by Cybulski et al. [81] for the rare gas dimers,  $\text{Ne}_2$ ,  $\text{Ar}_2$ ,  $\text{HeNe}$ ,  $\text{NeAr}$ , and  $\text{HeAr}$ . For the case of He dimer, we compare the spherical atom model for to highly accurate full CI calculations [91]. We also examine how the spherical atom model compares to CBS extrapolated MP4 calculations for  $\text{H}_2^3$ ,  $\text{H}_2^9$ ,  $\text{C}_2^+$ , and  $\text{HC}^+$ . The size parameter for the spherical atom,  $r_s$ , and the damping radius,  $r_d$ , of Equation (4.2.23), are obtained by non-linear least squares fitting to the benchmark potential curves for the homonuclear diatomic systems,  $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ ,  $\text{H}_2^3$ , and  $\text{H}_2^9$ , and  $\text{C}_2$ .

The spherical atom model dispersion potential energy, given by Equation (4.2.23) compared to benchmark potential energy calculations for  $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ , triplet  $\text{H}_2$ , and nonet  $\text{C}_2$  are shown in Figure 15 through Figure 19. Here, the  $r_s$  and  $r_d$  values reported were obtained from non-linear least squares fitting to accurate benchmark [81,91] potential curves (Figure . The spherical atom model for dispersion forces, Equation (4.2.23), shows excellent agreement with benchmark potential energy curves. Table 8 compares the Spherical Atom Model+HF/aug-cc-pV5Z dissociation energies and equilibrium geometries with benchmark studies [81,92] for the homo nuclear diatomics under investigation:  $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ , triplet  $\text{H}_2$ , and nonet  $\text{C}_2$ . The equilibrium geometries calculated using the Spherical Atom Model + HF/aug-cc-pV5Z are in agreement with benchmark [81] geometries to all significant figures. The

largest deviation in calculated dissociation energy is  $1 \mu$  Eh for  $\text{Ar}_2$  and nonet  $\text{C}_2$ .

**Combination rules:** The parameters  $C_6$ ,  $r_s$ , and  $r_d$  once determined for the homo nuclear diatomics can then be predicted for the hetero nuclear diatomics by the use of combining rules. The spherical atom model dispersion function, Equation (4.2.23), suggests that the appropriate combination rule for the hetero nuclear  $r_s$  value is the arithmetic mean of the constituent homo nuclear values, as in Equation (4.4.1),

$$r_s^{ij} = \frac{r_s^{i,i} + r_s^{j,j}}{2}, \quad (4.4.1)$$

Where  $r_s^{ij}$  is the spherical atom radius for the hetero nuclear diatomic, and  $r_s^{i,i}$  and  $r_s^{j,j}$  are the spherical atom radii for the constituent homo nuclear diatomics. The combination rule that we shall adopt which works well in practice for the hetero nuclear damping radius,  $r_d$ , is the geometric mean of the constituent homo nuclear values, given in Equation (4.4.2),

$$r_d^{i,j} = \sqrt{r_d^{i,i} * r_d^{j,j}}, \quad (4.4.2)$$

Where  $r_d^{ij}$  is the damping radius for the hetero nuclear diatomic, and  $r_d^{i,i}$  and  $r_d^{j,j}$  are the damping radii for the constituent homo nuclear diatomics. Different combination rules for the  $C_6$  coefficients of unlike atom-atom pairs have been carefully tested for many complexes [75]. In the well-known London approximation [92], shown in Equation (4.4.3),

$$C_6^{ij} = \frac{3 \alpha_i \alpha_j I_i I_j}{2 I_i + I_j}, \quad (4.4.3)$$

Where  $C_6^{ij}$  is the  $C_6$  coefficient of the hetero nuclear diatomic,  $\alpha_i$  and  $\alpha_j$  are the atomic polarizabilities of the interacting atoms, and  $I_i$  are the first ionization energies of the interacting atoms. A substantial improvement to the London approximation was made by Slater & Kirkwood [93], as given in Equation (4.4.4)

$$C_6^{ij} = \frac{2\alpha_i \alpha_j C_6^{ii} C_6^{jj}}{\alpha_i^2 C_6^{jj} + \alpha_j^2 C_6^{ii}} \quad (4.4.4)$$

Here, the  $C_6$  coefficient of the hetero nuclear diatomic is related to the atomic polarizabilities of the interacting atoms,  $\alpha_i$  and  $\alpha_j$ , and also to the  $C_6$  coefficients for the constituent homo nuclear atom-atom pairs. The Slater-Kirkwood approximation gave drastic improvements over the London approximation [93,94] and it is the predominant combination rule used in modern molecular mechanics force fields,

[82-85,93-95] and DFT-D techniques.

In 1968, Tang [96] showed that the exact upper bound for the mixed  $C_6^{ij}$  coefficient is the geometric mean, as given in Equation (4.4.5)

$$C_6^{ij} \leq \sqrt{C_6^{ii} C_6^{jj}} \quad (4.4.5)$$

In the present work we shall estimate the hetero nuclear  $C_6^{ij}$  value by the geometric mean of the constituent homo nuclear  $C_6$  values. This combination rule has been shown to work well for rare gases.

**Rare gas heteronuclear diatomics with hartree-fock:** The spherical atom model dispersion potential energy, given in Equation (4.2.23), compared to benchmark calculations for HeNe, NeAr, HeAr, and  ${}^6\text{O}_g^+$  HC dimers are shown in Figure 20 through Figure 23 Here, the  $C_6$ ,  $r_s$  and  $r_d$  values used to predict these potentials were obtained from the

combination rules given in Equations (23), (24), and (27) for  $r_s$ ,  $r_d$ , and  $C_6$  respectively. The spherical atom model potential curves predicted for the hetero nuclear systems from their constituent homo nuclear cases, shows excellent agreement with benchmark potential energy curves. Table 9 compares the Spherical Atom Model + HF/aug-cc-pV5Z dissociation energies and equilibrium geometries with benchmark studies [21] for the hetero nuclear diatomics under investigation: HeNe, NeAr, HeAr, and  ${}^6\text{O}_g^+$  HC.

When the combination rules are implemented, while the equilibrium geometries calculated using the Spherical Atom Model + HF/aug-cc-pV5Z remain in agreement with benchmark<sup>21</sup> geometries to all significant figures, largest deviation in the calculated dissociation energy triples for HeNe and doubles for HeAr. The Spherical Atom Model predicts dissociation energies to within 4% and 1% for these dimers, respectively.

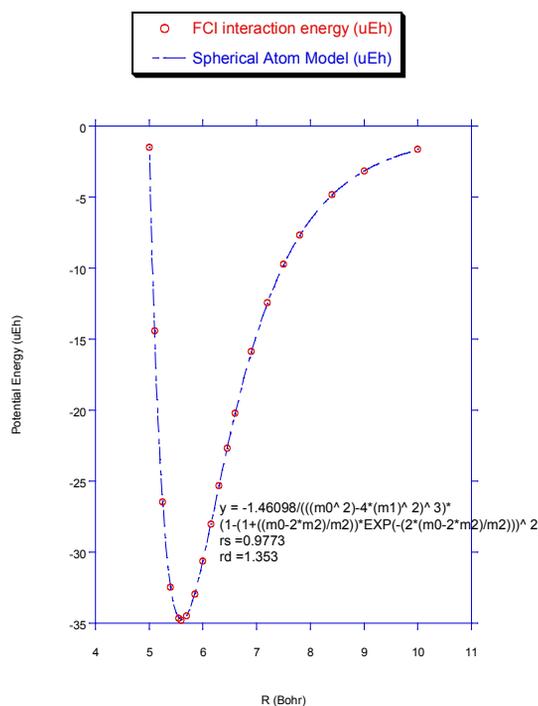


Figure 15: He Dimer Potential Energy Curve.

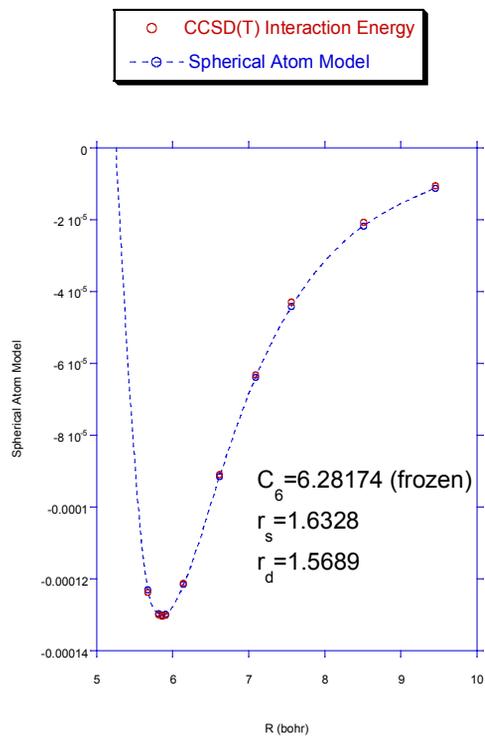


Figure 16: Ne Dimer Potential Energy Curve.

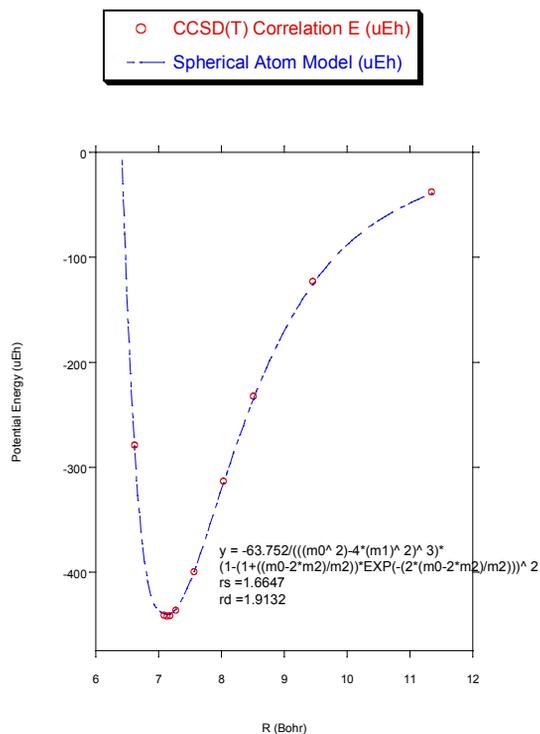


Figure 17: Ar Dimer Potential Energy Curve.

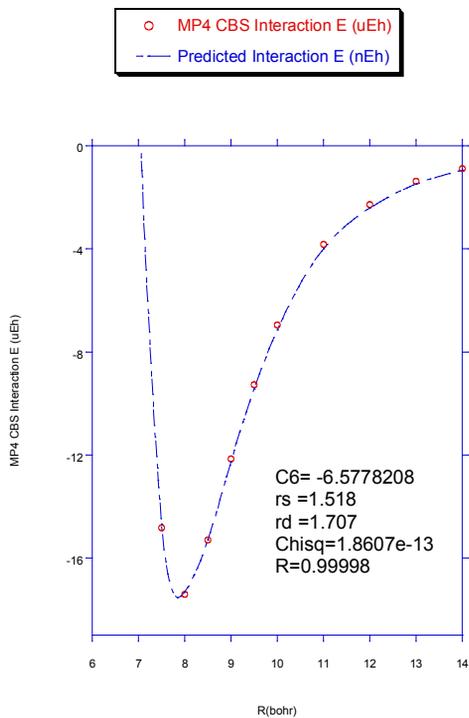


Figure 18: 4.4.4:  $6^1_u$   $H_2$  Potential Energy Curve.

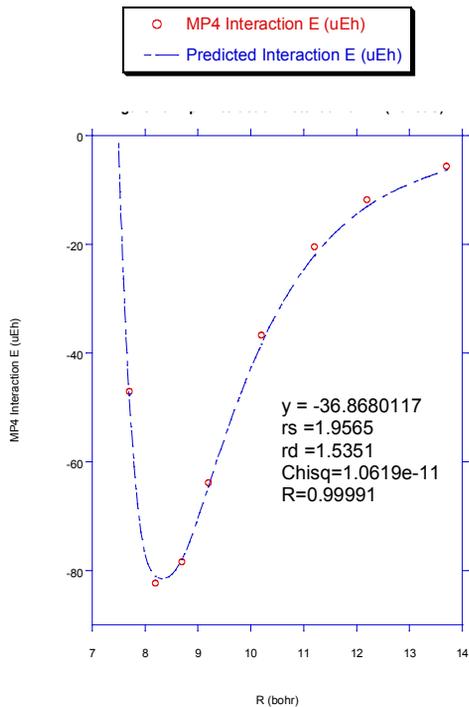


Figure 19:  $9^1_g$   $C_2$  Potential Energy Curve.

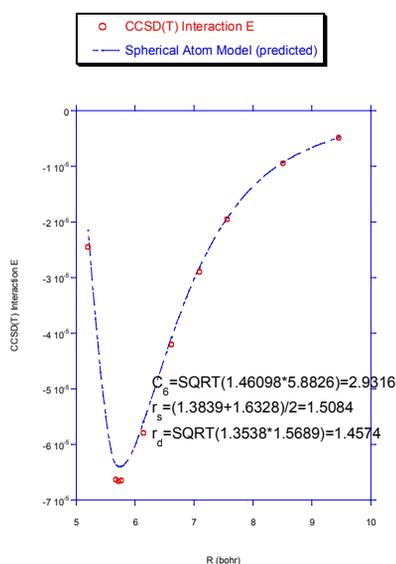
**Table 8:** Comparison of the Spherical Atom Model + HF/aug-cc-pV5Z dissociation energies and equilibrium geometries with benchmark studies [21-32] for the homonuclear diatomics under investigation: He<sub>2</sub>, Ne<sub>2</sub>, Ar<sub>2</sub>, triplet H<sub>2</sub>, and nonet C<sub>2</sub>.

Homonuclear Diatomic	Equilibrium Geometry, R <sub>e</sub> (Bohr)	Spherical Atom Model + HF/Aug-Cc-Pv5z Dissociation Energy, D <sub>e</sub> (Eh)	Benchmark Dissociation Energy, D <sub>e</sub> (Eh)	Deviation Dissociation Energy, D <sub>e</sub> (Predicted Benchmark) (Meh)
He <sub>2</sub>	5.6	-0.000035	-0.000035 a)	0
Ne <sub>2</sub>	5.86	-0.00013	-0.000130 b)	0
Ar <sub>2</sub>	7.136	-0.000441	-0.000442 b)	1
(triplet S) H <sub>2</sub>	8.5	-0.000017	-0.000017 c)	0
(nonet S) C <sub>2</sub>	8.2	-0.000081	-0.000082 c)	1

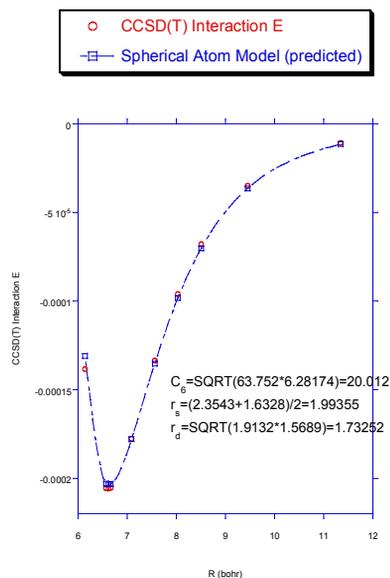
a) FCI, reference 31

b) CCSD(T)/aug-cc-pV5Z(332221),

c) MP4/aug-cc-pV5Z(332221) bond functions



**Figure 20:** HeNe Dimer Potential Energy Curve.



**Figure 21:** NeAr Dimer Potential Energy Curve.

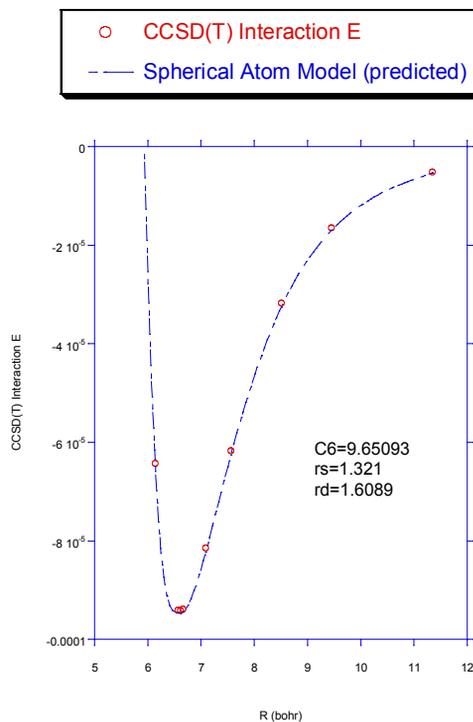


Figure 22: HeAr Dimer Potential Energy Curve.

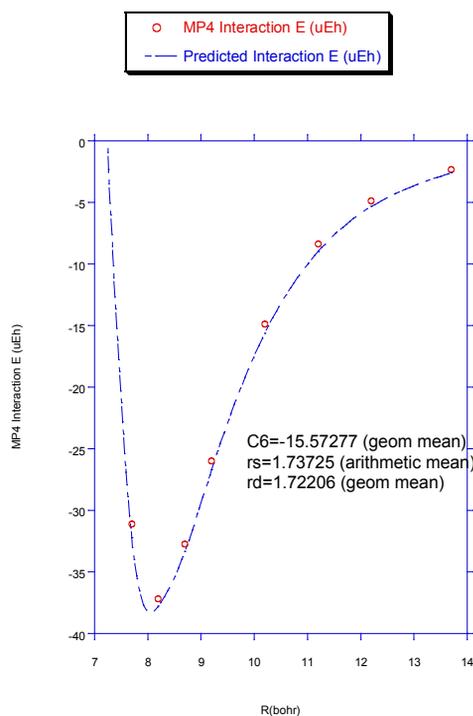


Figure 23: Sextet HC Potential Energy Curve.

**Table 9:** Comparison of the Spherical Atom Model + HF/aug-cc-pV5Z dissociation energies and equilibrium geometries with benchmark studies [80, 91] for the homonuclear diatomics under investigation: HeNe, NeAr, HeAr, and  ${}^6\tilde{\text{O}}_g^-$  HC.

Homonuclear Diatomic	Equilibrium Geometry, $R_e$ (Bohr)	Spherical Atom Model + Hf/Aug-Cc-Pv5z Dissociation Energy, $D_e$ (Eh)	Benchmark Dissociation Energy, $D_e$ (Eh)	Deviation Dissociation Energy, $D_e$ (Predicted-Benchmark) (Meh)
HeNe	5.718	-0.000064	-0.000067 a)	3
NeAr	6.616	-0.000203	-0.000206 a)	2
HeAr	6.614	-0.000095	-0.000094 a)	-1
(sextet S) HC	8.2	-0.000038	-0.000037 b)	-1

a) CCSD(T)/aug-cc-pV5Z(332221),

b) MP4/aug-cc-pV5Z(332221) bond functions

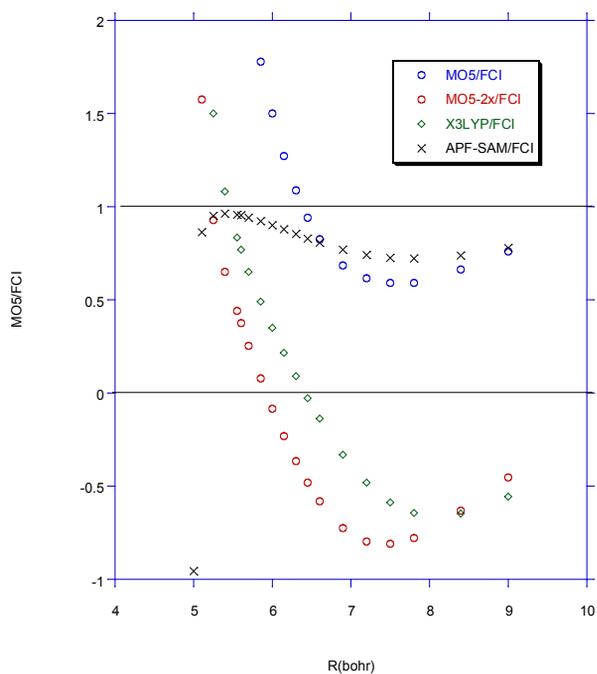
## DFT functional comparisons

Several density functionals have been proposed with the goal of treating dispersion properly. Three recent density functionals of this type are the MO5 [97] and MO5-2x [98] functionals of Truhlar and coworkers, and the X3LYP [99] of Goddard et al. [99]. The X3LYP functional has been shown to give qualitatively good binding energies for He and Ne dimers, giving 5% and 25% errors in the dissociation energies compared to experimental benchmarks. However, the X3LYP gave only 2% of the binding in Argon dimer. Truhlar's MO5 functional gives a 30% error in dissociation energy for the ethene dimer compared to MP2/6-311G(2df,2p) calculations. We have computed MO5, MO5-2x, and X3LYP potential energy curves for the 6 rare gas dimers  $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ , HeNe, HeAr, and NeAr, using a large basis set that reduces the counterpoise correction to 1  $\mu\text{Eh}$  error. Figure 24 through Figure 29 show the ratio of the MO5, MO5-2x, and X3LYP interaction energies to the benchmark CCSD(T) interaction energies of Cybulski et al. for  $\text{He}_2$ ,  $\text{Ne}_2$ ,  $\text{Ar}_2$ , HeNe, HeAr, and NeAr, respectively.

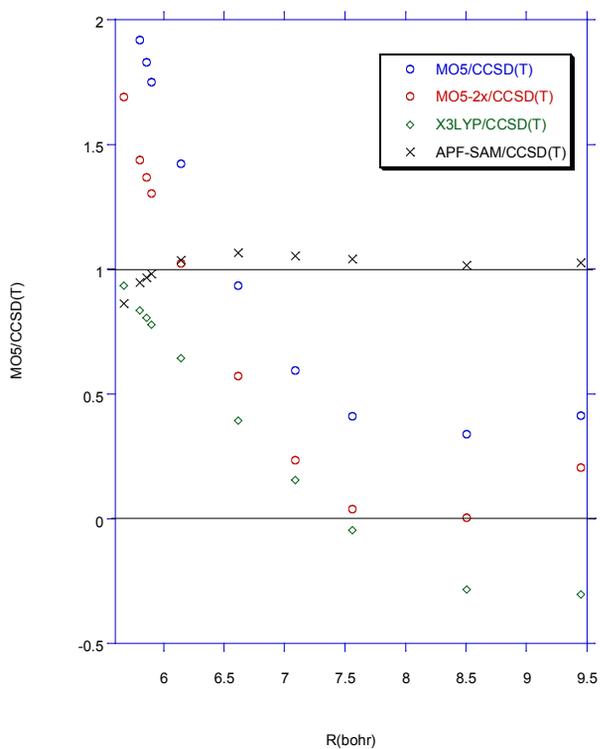
Figure 24 through Figure 29 illustrate that for all six rare-gas cases, even when the density functional is giving qualitatively reasonable dissociation energies, the ratio of the DFT to the CCSD(T) interaction energy drops rapidly to zero as the internuclear distance increases. This shows that whatever bonding is present in the MO5, MO5-2x, or the X3LYP density functionals decays much more rapidly than

$\text{R}^{-6}$  and hence is not dispersion. Figure 16 in particular shows that even though the X3LYP functional gives qualitatively reasonable binding energies for Ne dimer, dispersion is not treated properly in this functional.

Many other density functionals have been developed with the goal of improving weakly bound interactions such as hydrogen bonding, but are not expected necessarily to include dispersion or improve the description of van der Waals interactions in DFT. Among these are the hcth [100-102] thcth [103], and the tpssh [104] density functionals. Like the MO5, MO5-2x, or the X3LYP density functionals, the hcth, thcth, and the tpssh DFT's can also give spurious binding for some van der Waals systems. Figures 30 & 31 show that although the tpssh gives a qualitatively reasonable dissociation energy for Neon dimer, the ratio of the tpssh to the CCSD(T) interaction energy falls to zero rapidly with increasing internuclear separation. Therefore, as expected, the spurious attraction given by the tpssh functional does not incorporate dispersion, although this "pseudo-dispersion" may be sufficient for most purposes. The hcth and thcth functionals give interaction energies that are drastically overbound compared with CCSD(T) benchmark potentials. Table 10 shows the Neon dimer Density Functional theory deviations in equilibrium geometry (Bohr) and Energies (kcal/mol) at the benchmark [81] CCSD(T)/aug-cc-pV5Z(332221) Equilibrium geometry of 5.86 Angstroms (Eh) compared with the benchmark Equilibrium dissociation energy, -0.000130Eh.



**Figure 24:** Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies for He<sub>2</sub>.



**Figure 25:** Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies Ne<sub>2</sub>.

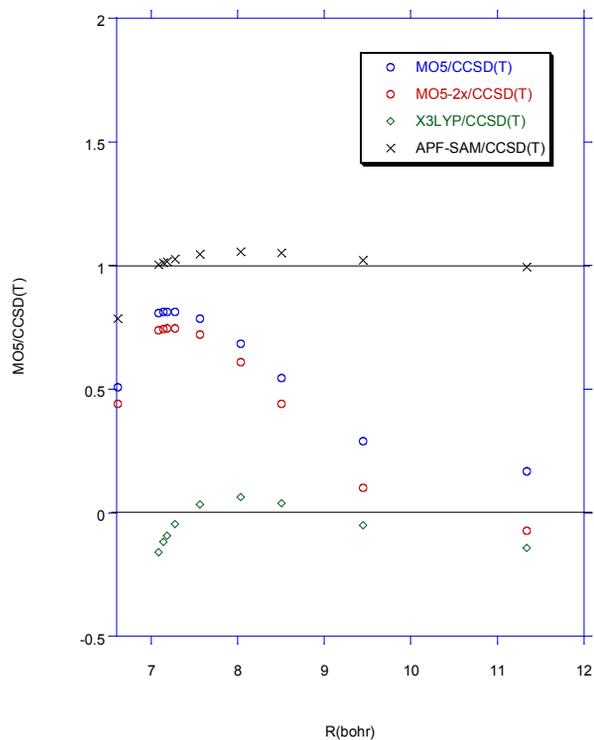


Figure 26: Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies  $\text{Ar}_2$

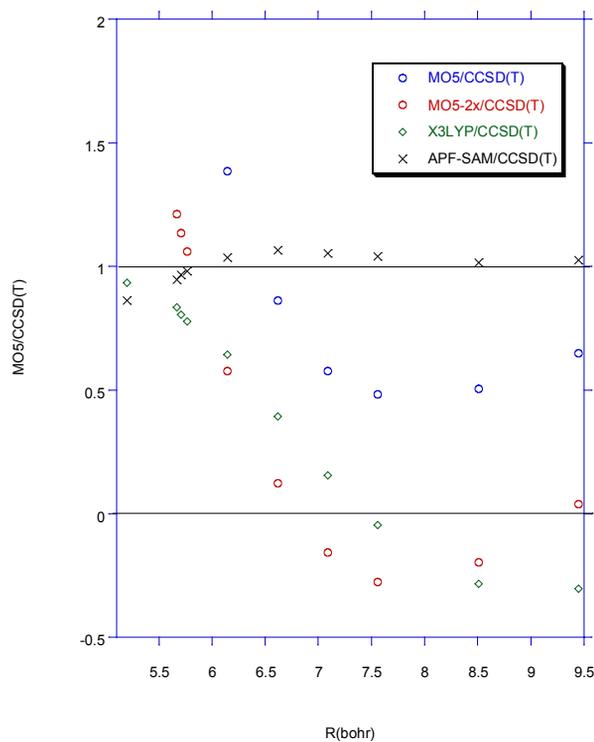
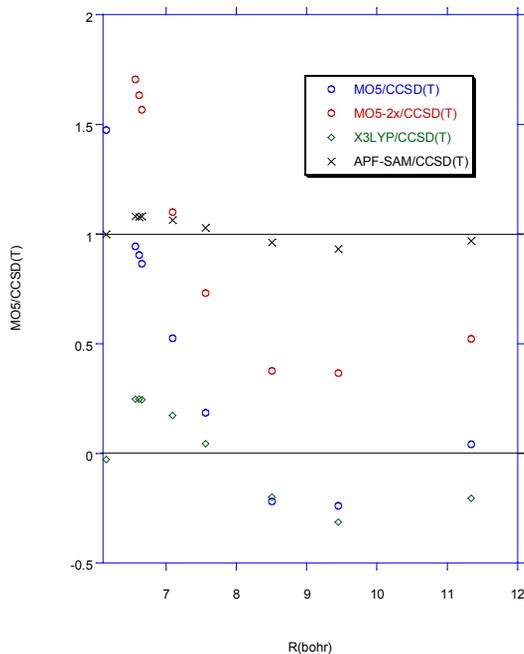
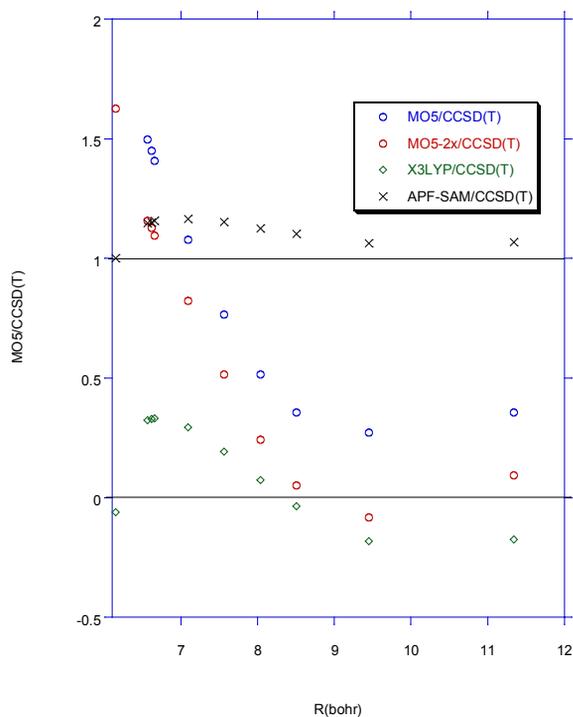


Figure 27: Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies  $\text{HeNe}$ .



**Figure 28:** Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies HeAr.



**Figure 29:** Ratio of the MO5, MO5-2x, and X3LYP interaction energies to benchmark CCSD(T) interaction energies NeAr.

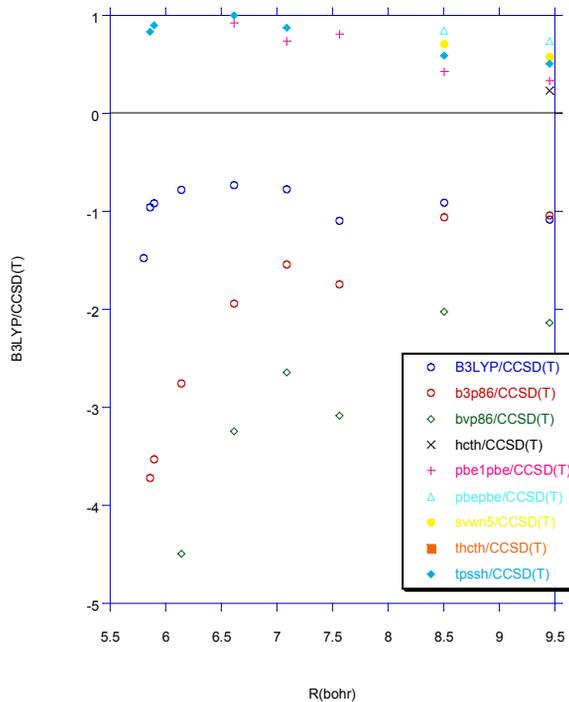


Figure 30: Ratio of DFT to benchmark CCSD(T) Interaction Energy for Ne dimer.

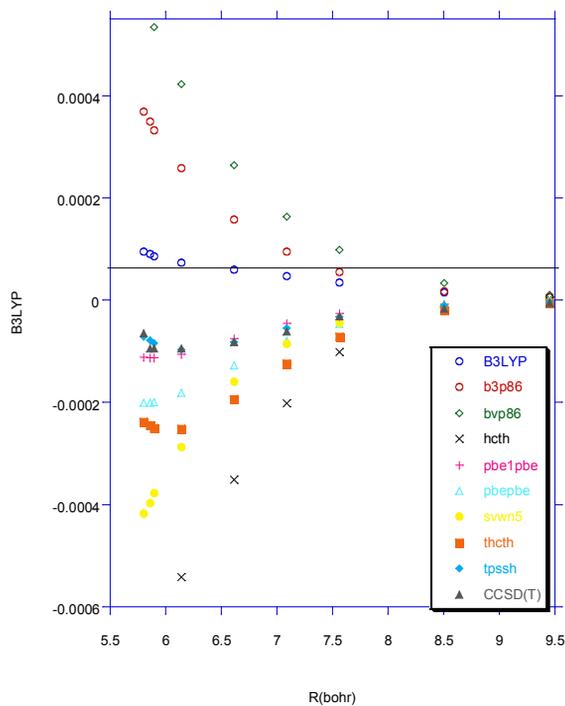


Figure 31: DFT and benchmark CCSD(T) Potential Energy curves for Ne Dimer.

**Table 10:** Neon dimer Density Functional theory deviations in equilibrium geometry (Bohr) and Energies (Eh) at the benchmark [80] CCSD(T)/aug-cc-pV5Z(332221) Equilibrium geometry of 5.86 Bohr compared with the benchmark Equilibrium dissociation energy, -0.000130Eh.

Dft Functional	Dft Equilibrium Geometry, $R_e$ (Bohr)	Dft Energy At $R_e$ (Eh)
B3 LYP	a	a
b3p86	a	a
bvp86	a	a
hcth	5.671	-0.000691
pbe1pbe	5.898	-0.000113
pbepbe	5.898	-0.000201
svwn5	5.198	-0.000699
thcth	6.144	-0.000251
tpssh	6.144	-0.000097

### The APF-SAM hybrid density functional theory

Popular density functionals in current use give drastically dissimilar behaviors for dispersion interaction potentials [75], some functionals predicting purely repulsive potential energy curves for the rare-gas diatomics, such as the B3LYP hybrid DFT, while others predicting spurious binding, often drastically over-estimating the binding in rare-gases, such as the hcth functional, as observed in Figure 31 for Neon dimer. The B3LYP functional, while being overly repulsive, and the PBE1PBE [105] functional, while giving spurious binding, have been shown to give good thermo chemistry for covalent molecules [106]. In taking a normalized linear combination of these two functionals, spurious interactions between weakly bound systems can be reduced, while the resulting hybrid functional is guaranteed to have comparable accuracies for geometries, irrational frequencies, and thermo chemistry as the parent methods. The B3LYP coefficient was optimized through simultaneous non-linear least squares fits to the ratio of the potential energy curves weighted by the dissociation energy for Ar, Ne, and He dimers, and was found to be 34.97% B3LYP (and therefore 65.03% PBE1PBE). We shall refer to this hybrid functional as the Austin-Petersson-Frisch (APF) hybrid functional.

In order to incorporate the spherical atom model (SAM) into the APF functional, the spherical atom radii,  $r_s$  values, found through non-linear least squares optimization of SAM plus the Hartree-Fock interaction curve to benchmark potentials were used since the size of the interacting spherical shells should not change on going from Hartree-Fock to DFT. However, the damping radii,  $r_d$  values, will change since there is short range correlation energy present in the DFT that is not present in Hartree-Fock (and therefore the damping required for DFT will be increased). We have re-optimized the  $r_d$  values for Ar, Ne, and He homo nuclear rare-gas diatomics by non-linear least squares fitting of the APF hybrid functional plus SAM to benchmark CCSD (T) curves. The new  $r_d$  values are 2.408, 2.382, 2.198 for Ar, Ne, and He dimers, respectively. The results of the APF-SAM approach are shown in Figure 32 through Figure 34 for He, Ne, and Ar dimers, respectively.

Although the APF-SAM approach describes benchmark

CCSD(T) potential curves with decreased accuracy compared to the HF-SAM approach, it is still in good agreement with the benchmark potentials, the largest error in dissociation energy being a mere 2% for Ne dimer. Table 11 presents the Austin-Petersson-Frisch (APF) Density Hybrid Functional with appended Spherical Atom Model (APF-SAM) of Equation (4.2.23) predicted equilibrium geometries,  $R_e$ , (Bohr), APF-SAM Dissociation Energies,  $D_e$ , (Eh) and deviations from benchmark calculations [81,91] for homo nuclear rare gas diatomics. When the Spherical Atom Model is appended to the Austin-Petersson-Frisch Density Functional, errors upto 50 mBohr are observed in the equilibrium geometry of He and Ar dimers, while 38 mBohr error is present in Ne dimer. These errors are within 1% of benchmark [81] values. While the largest deviation in dissociation energy is 6  $\mu Eh$  for Ar dimer, the largest percent error is 5% and is in the case of He<sub>2</sub>. Using literature  $C_6$  values,  $r_s$  values obtained from HF-SAM, and the  $r_d$  values obtained from APF-SAM for the homonuclear rare gas pairs, the potential energy curves for the heteronuclear rare gas curves were predicted using the combination rules given in Equations (4.4.1), (4.4.2), and (4.4.5) for  $r_s$ ,  $r_d$ , and  $C_6$  respectively. The results of the APF-SAM approach are shown in Figure 35 through Figure 37 for HeNe, NeAr, and HeAr dimers, respectively.

The largest error of the APF-SAM approach for predicting hetero nuclear rare-gas pair dissociation energies is 15% for NeAr, which is still in good qualitative agreement with high level CCSD(T) calculations [81]. Table 12 presents the Austin-Petersson-Frisch (APF) Density Hybrid Functional with appended Spherical Atom Model (APF-SAM) of Equation (4.2.23) predicted equilibrium geometries,  $R_e$ , (Bohr), APF-SAM Dissociation Energies,  $D_e$ , (Eh) and deviations from benchmark calculations [81,91] for hetero nuclear rare gas diatomics. The largest deviation in bond length is for the HeAr dimer, 57 mBohr. However, this is only a 1% error in equilibrium geometry and is the largest of these three cases. The NeAr dimer displays the largest deviation in calculated dissociation energy of 32  $\mu Eh$ , which is also the largest percent error of 15%. Therefore, for these cases, when the Spherical Atom Model is appended to the Austin-Petersson-Frisch density functional, qualitative accuracy is achieved.

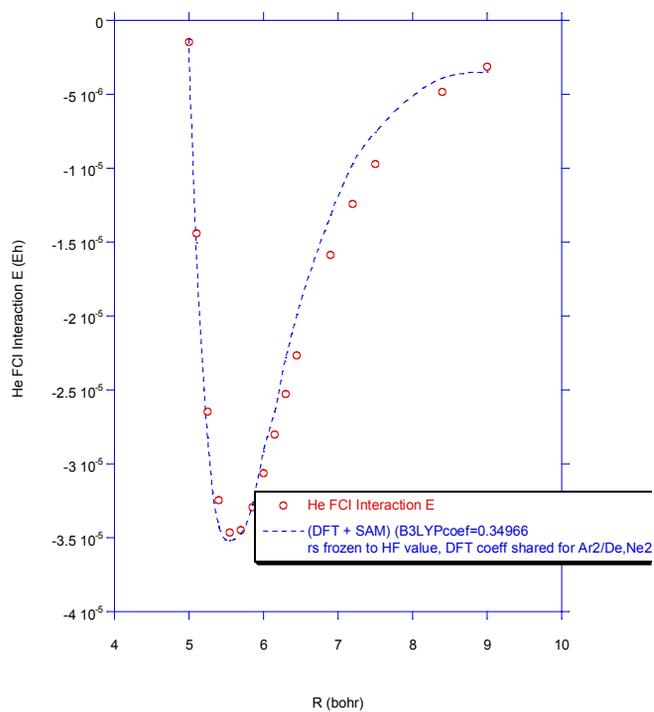


Figure 32: He Dimer Potential Energy Curve.

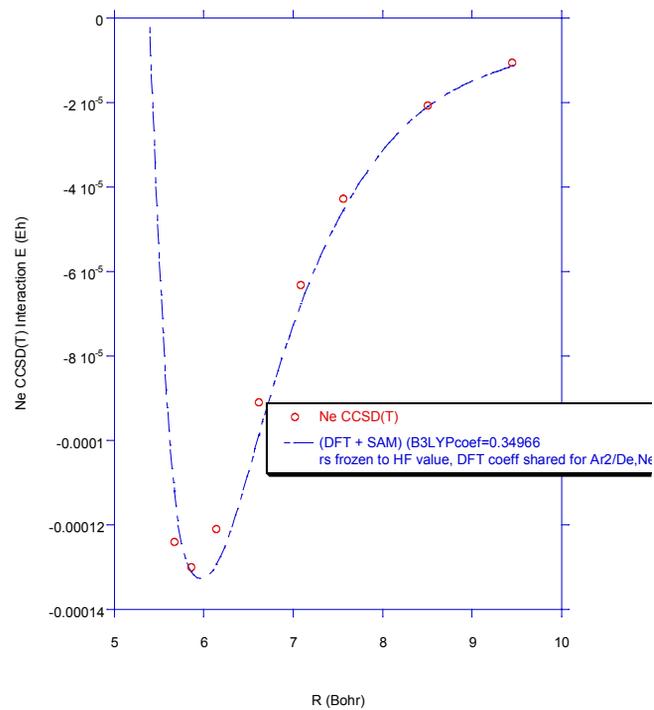
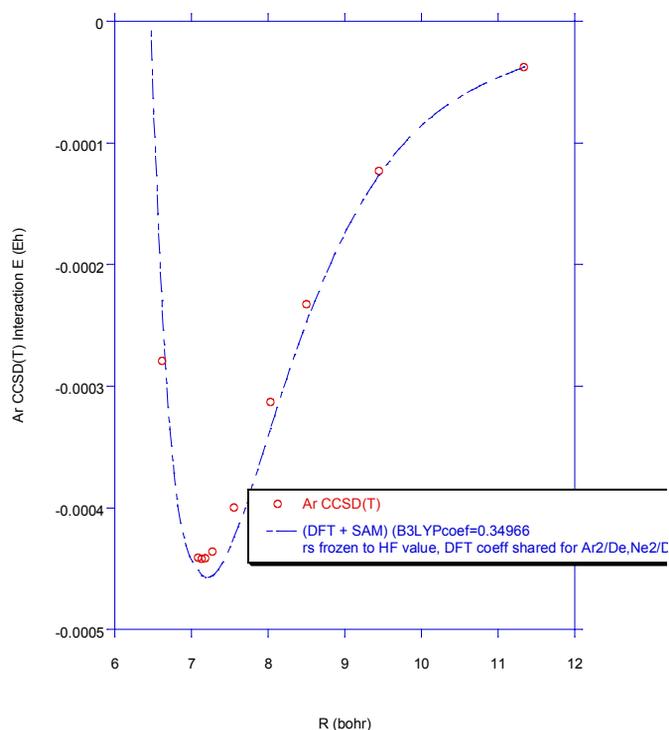


Figure 33: Ne Dimer Potential Energy Curve.



**Figure 34:** Ar Dimer Potential Energy Curve.

**Table 11:** Austin-Petersson-Frisch (APF) Density Hybrid Functional with appended Spherical Atom Model (APF-SAM) of Equation (4.2.23) predicted equilibrium geometries,  $R_e$ , (Bohr), APF-SAM Dissociation Energies,  $D_e$ , (Eh) and deviations from benchmark calculations [80,91] for homonuclear rare gas diatomics.

Homonuclear Diatomic	APF-SAM Equilibrium Geometry, $R_e$ (Bohr)	APF-SAM Dissociation Energy, $D_e$	Deviation Equilibrium Geometry(Predicted-Benchmark) (Bohr)	Deviation Dissociation Energy, $D_e$ (Predicted-Benchmark) ( $\mu$ Eh)
He <sub>2</sub>	5.55	-0.000033	-0.05	2
Ne <sub>2</sub>	5.898	-0.000128	0.038	3
Ar <sub>2</sub>	7.183	-0.000448	-0.05	-6

**Table 12:** Austin-Petersson-Frisch (APF) Density Hybrid Functional with appended Spherical Atom Model (APF-SAM) of Equation (4.2.23) predicted equilibrium geometries,  $R_e$ , (Bohr), APF-SAM Dissociation Energies,  $D_e$ , (Eh) and deviations from benchmark Calculations for homonuclear rare gas diatomics.

Homonuclear Diatomic	APF-SAM Equilibrium Geometry, $R_e$ (Bohr)	APF-SAM Dissociation Energy, $D_e$	Deviation Equilibrium Geometry(Predicted-Benchmark) (Bohr)	Deviation Dissociation Energy, $D_e$ (Predicted-Benchmark) ( $\mu$ Eh)
HeNe	5.766	-0.000066	0.047	1
NeAr	6.654	-0.000237	0.038	-32
HeAr	6.560	-0.000102	-0.057	-7

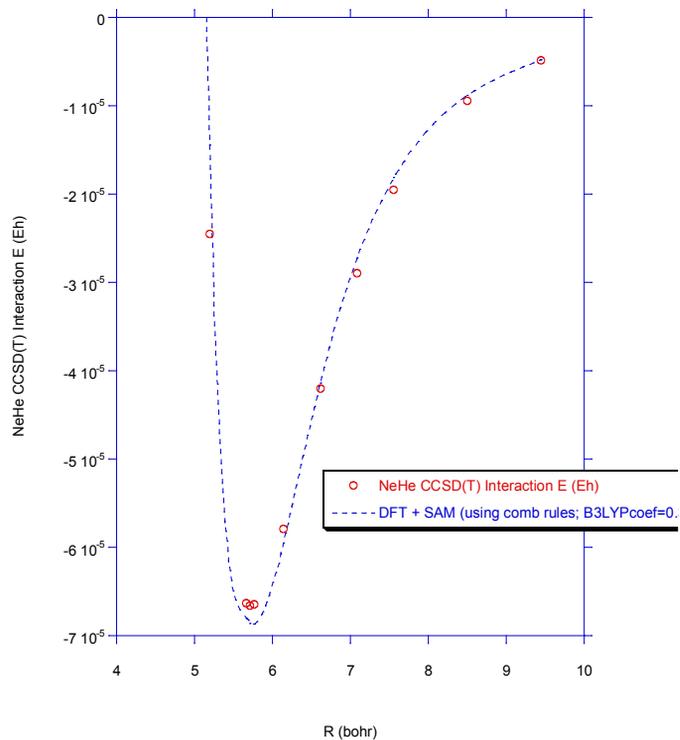


Figure 35: HeNe Dimer Potential Energy Curve.

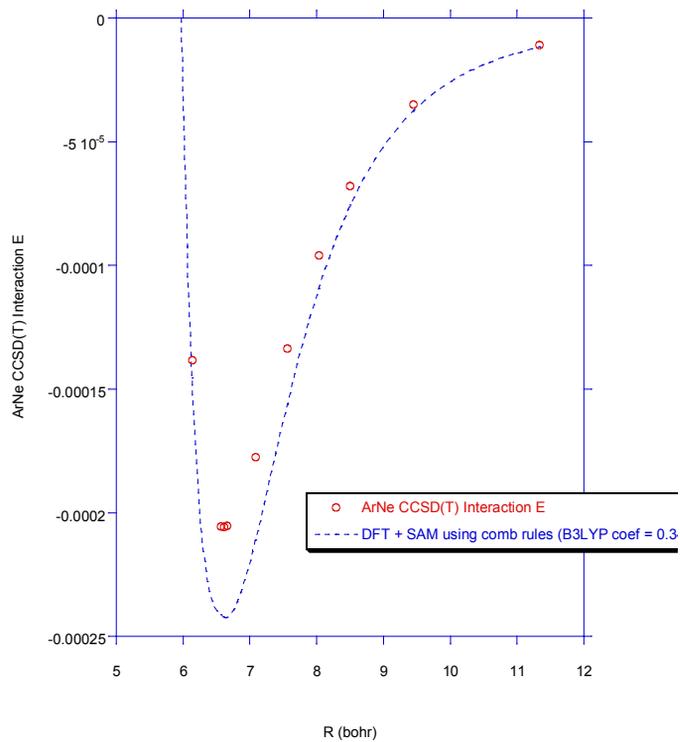
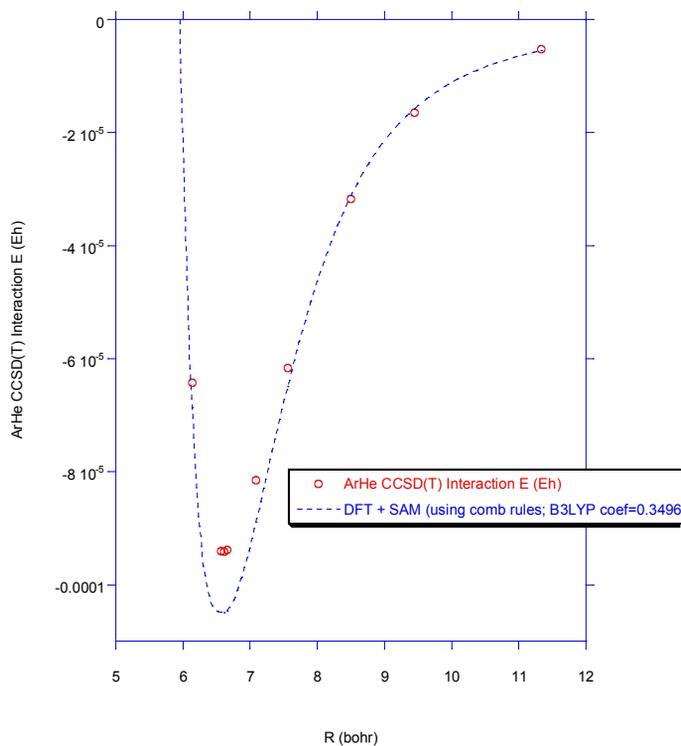


Figure 36: NeAr Dimer Potential Energy Curve.



**Figure 37:** HeAr Dimer Potential Energy Curve.

## Conclusion

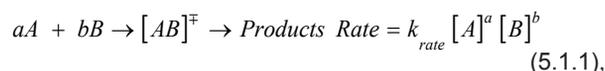
In this work we have presented the spherical atom model for empirically estimating dispersion interactions that works well and is significantly more accurate than damped or undamped  $C_6/R^6$ . The spherical atom model can be appended to a linear combination of popular density functionals which minimizes spurious interactions in order to incorporate dispersion forces into density functional theory. Some recent functionals such as the MO5, MO5-2x, and X3LYP which have been developed with the goal of including dispersion forces do not provide consistently improved interaction energies for the rare gas dimers, and even in cases where some binding is produced, the interaction is spurious and shown not to be due to dispersion.

# Chapter 4

## Vibrational Frequencies of Transition States

### Introduction and goal of research

Zero Point Energy (ZPE) changes from reactant to transition states raise or lower the activation energy of reaction, which affects its rate; and therefore, it is important to obtain these changes accurately. We have shown for four non-degenerate chemical reactions that high levels of correlation treatment are necessary (QCISD (T) with large basis sets). The Intrinsic Reaction Coordinate Maximum (IRCMMax) method of Petersson and co-workers was previously developed to improve the geometries of transition states and barrier heights of chemical reactions. It was the goal of this work to test whether the IRCMax method can also improve the normal mode frequencies of transition states and ZPE changes from reactant to transition state of chemical reactions. The rate of a chemical reaction is governed by the concentration of reactants in the rate limiting step of the chemical mechanism, their orders in the chemical mechanism, and a rate constant, as given in the chemical rate equation, Equation (5.1.1)



where [A] and [B] are the concentrations of reactants A and B, a and b are their stoichiometric coefficients in the elementary step, and  $\Delta E_0^{\ddagger}$  is the rate constant for the reaction. In 1935, Henry Eyring postulated that this rate constant depends on the energy of the transition state relative to the reactants of the chemical reaction, as given in Equation. (5.1.2),

$$\frac{\Delta E_0^{\ddagger}}{k_B T} = \ln \left( \frac{Q_A^a Q_B^b}{Q_{AB}^{\ddagger}} \right) + \ln \left( \frac{h}{k_B T} \right) \quad (5.1.2).$$

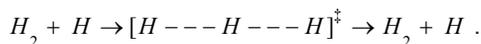
Here,  $k_B$  is the Boltzman constant, T is the Kelvin temperature, h is Plank's constant,  $Q_A/V$ ,  $Q_B/V$ , and  $Q_{AB}^{\ddagger}/V$  are the partition functions per volume for the reactants and transition states for all normal coordinates except the reaction path (the partition function for this coordinate is included in the factor  $k_B T/h$ ), R is the universal gas constant, and  $\Delta E_0^{\ddagger}$  is the barrier height of the chemical reaction [107]. The barrier height is defined as the difference in total energies of the transition state and reactants, as shown in Equation (5.1.3) [107].

$$\Delta E_0^{\ddagger} = E_0^{\ddagger} - (E_0^A + E_0^B) \quad (5.1.3).$$

The total energies used in Equation (5.1.3) for obtaining the barrier height of the chemical reaction, includes the zero point energies of the transition state and reactants. The zero-point energy is the smallest amount of energy that a system can possess, and is given by the energy of the harmonic oscillator where the quantum number n is zero, as given by Equation (5.1.4),

$$(5.1.4).$$

State for the degenerate chemical reaction



The effect of the zero point energy change from reactant to transition states will be to raise or lower the activation energy of the reaction, and which effect is observed depends on where the transition state lies along the reaction path (Figure 38). For this degenerate chemical reaction, the transition state is constrained by symmetry to be in the center of the reaction path. The zero point energy change from reactant to transition state is negative, producing smaller activation energy for the chemical reaction. One can see that for exothermic reactions in which the transition state is located close to the reactants on the reaction path, as well as for endothermic reactions, in which the transition state is close products along the reaction path, the zero-point energy contribution serves to raise the activation energy compared to the classic barrier height. Therefore, the precise location of the transition state along the reaction coordinate is essential for predicting chemical kinetics. We will show that it is impractical to reach this correct point along the reaction path by conventional theoretical methods. The IRC Max of Petersson and co-workers has been shown to reach the correct location for the transition state [108]. Thus in this work, we have tested the potential for the IRC Max method to improve normal mode frequencies of the reactants, products, transition states, and zero point energy contributions to the activation energies of six degenerate and four non degenerate chemical reactions. We will show while the IRC Max method can improve the accuracy of vibrational normal modes and zero-point energy contributions to the activation energy it is limited to do so by having a reasonably good IRC. In the majority of cases, the IRC Max method fails to improve normal mode frequencies and ZPE changes.

First, we must assess the accuracy of conventional computational methodologies for vibrational frequencies and zero point energies of stable molecules in order to select a benchmark for calibration of transition state frequencies and zero point energies where no experimental data is available. Once we have selected our benchmark, we test the accuracies of the computational methodologies alone by comparing the normal modes and zero point energies of transition states of degenerate chemical reactions, where the location of the transition state along the chemical reaction is constrained by symmetry. Next, we test the accuracy of non-degenerate chemical reactions in order to assess the importance that the position of the transition state along the reaction path plays in the estimation of zero point energy changes from reactant to transition state. We then shall show that the IRCMax<sup>2</sup> method improves the accuracy of zero point energy changes relative to the selected benchmark by locating the correct position of the transition state.

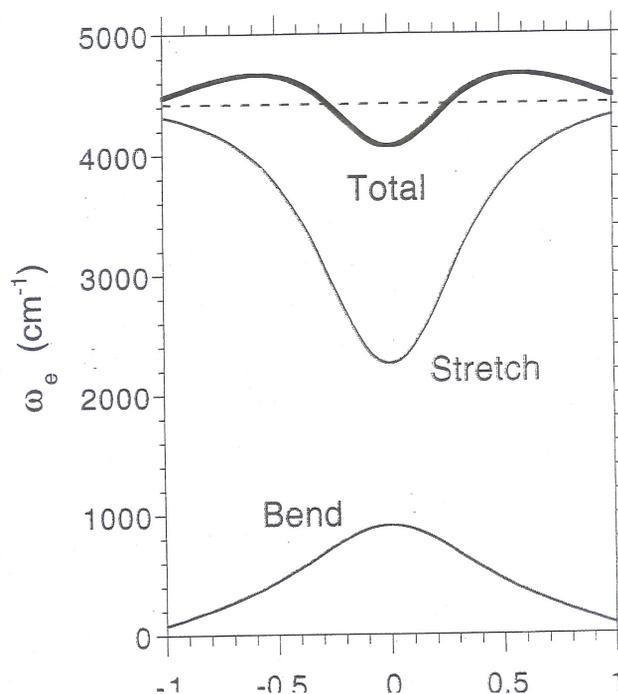


Figure 38:  $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ . Zero Point Energy change from reactant to transition state as a function of reaction coordinate.

### Calibration with experimental harmonic frequencies of stable molecules

In order to select a benchmark for comparison of vibrational frequencies and zero point energies for transition states, it is necessary to compare calculated frequencies of stable molecules with experimental frequencies. This was originally done by John Keith. Since anharmonic effects generally exceed the computational errors for the most demanding large basis set QCISD (T) [109] calculations, we will limit our test set to species for which experimental harmonic constants are available. Anharmonic effects cannot be recovered without explicit calculation of third- and fourth-derivatives, and current computational methodologies have only numerical third and fourth derivatives available, making their determination impractical. Therefore, we limit the calibration to harmonic analyses of the vibrational frequencies.

Our test set includes the reactants and products from the ten hydrogen atom exchange reactions examined in the original IRCMax paper, and we focus on a few of the most important single-reference computational methodologies, Hartree-Fock [110,111] and Møller-Plesset second order [112] (MP2), MP3 [113] B3LYP [114,115], QCISD [116-118] and QCISD(T) [109]. We examine the convergence with basis sets: /3-21G(\*), /6-31G(d), /6-311G(d,p), /6-311G(2d,p), /6-311+G(2df,p), and /6-311+G(2df,2pd) [119-121]. The calculated and experimental harmonic frequencies [122-126] for  $\text{H}_2$ , CH, NH, OH, HF, CN,  $\text{N}_2$ ,

$\text{H}_2\text{O}$ , HCN,  $\text{CH}_3$ , and  $\text{CH}_4$  are given in Table 13. In order to uniformly compare the various computational methods, all calculated harmonic frequencies reported were scaled by factors, also reported in Table 13, that were optimized for these stable species excluding CN, in which unrestricted HF and MP2 exhibit spin-contamination, and also excluding  $\text{CH}_3$  radical for which no experimental harmonic constant is available.

The convergence of calculated frequencies with basis set and level of theory is summarized in Table 14. The spin-contaminated cyano radical was not included in this summary and will be considered separately to prevent this one pathological case from obscuring the results. While the 6-31G\* basis set is adequate to achieve the limit of Hartree-Fock accuracy, the 6-311G\*\* basis set is required for the MP2. The MP2 level of theory gives a slight improvement in calculated frequencies and a substantial improvement is seen for MP3 and B3LYP frequencies. The B3LYP frequencies achieve the same accuracy as MP2, while being much less computationally demanding. In order to realize the limit of accuracy for the highest level of correlation treatment in this study, the QCISD and QCISD (T), the large 6-311+G (2df,p) basis set is required.

The quadratic CI frequencies produce the smallest errors in calculated, scaled harmonic frequencies, and therefore the QCISD (T) is selected for the benchmark for comparison of simpler computational methodologies. The total RMS error compared with experiment for QCISD(T)/6-311+G(2df,2pd) ( $15\text{cm}^{-1}$ ) is 2.5 times smaller than the

B3LYP can offer with its limit of accuracy in the 6-311G\*\* basis set ( $40\text{ cm}^{-1}$ ). Therefore, we will select the QCISD (T)/6-311+G(2df,2pd) as an adequate reference point where experimental data is not known. The data in Table 14 show that the non-perturbative triples do not play a large role in the accuracy of frequency calculations. However, previous studies have shown that omitting triple excitations can double the errors in systems with occupied antibonding orbitals (eg:  $F_2$  and  $O_2$ , which were not included in this work) [127].

The errors in the calculated values of the harmonic

vibrational frequency for the cyano radical are presented in Table 15. Almost every method has an error for this case that is at least twice as large as the corresponding RMS error from Table 15. Since the transition states studied in this work are radical species with the potential for spin contamination, the use of the QCISD (T) level of theory is necessary to provide benchmarks for transition states. Although Table 15 indicates that the 6-311+G(2df,p) basis set achieves maximal accuracy for this level of theory for the small number of stable molecules examined here, we adopt the QCISD(T)/6-311+G(2df,2pd) level of theory and basis set as the benchmark for our study of transition states.

**Table 13:** Sample of Root-Mean Square Errors in Scaled Harmonic Frequencies of Reactants  $H_2$ , CH, NH, OH, HF,  $N_2$ ,  $H_2O$ , HCN,  $CH_4$ .

Molecule	Mode	HF/ 3-21G*	HF/ 6-31G*	MP2/ 6-311G**	B3LYP/ 6-311G**	QCISD/6- 311G(2d,p)	QCISD(T)/6- 311+G(2df,2pd)	Experiment
$H_2$	$\nu_1$	4531.37	4339.53	4454.16	4428.01	4400.21	4389.58	4401.21 [14]
CH	$\nu_1$	2802.87	2856.42	2887.86	2809.94	2827.26	2854.58	2858.5 [14]
NH	$\nu_1$	3077.16	3295.39	3327.73	3244.41	3239.52	3284.19	3282.27 [14]
OH	$\nu_1$	3511.09	3733.15	3786.50	3709.54	3736.35	3751.32	3737.76 [14]
HF	$\nu_1$	3951.56	4070.26	4177.63	4126.86	4181.49	4166.17	4138.32 [14]
CN	$\nu_1$	1765.68	1850.93	2807.09	2156.55	2152.12	2096.30	2068.59 [14]
$N_2$	$\nu_1$	2542.01	2575.97	2141.62	2451.89	2374.73	2328.15	2358.57 [14]
$H_2O$	$\nu_1(2a_1)$	3709.45	3801.85	3836.55	3817.44	3842.14	3845.60	3832.17 [15]
	$\nu_2(1a_1)$	1750.71	1706.00	1638.43	1641.88	1661.10	1658.33	1648.47 [15]
	$\nu_3(1b_2)$	3839.29	3912.29	3942.45	3914.59	3942.67	3951.25	3942.53 [15]
HCN	$\nu_1(2\sigma)$	3591.63	3436.92	3431.09	3464.87	3423.33	3434.71	3442.3 [16]
	$\nu_2(1\sigma)$	2330.14	2277.37	1986.23	2204.67	2139.31	2102.08	2129.1 [16]
	$\nu_3(1\pi)$	963.05	830.49	744.26	787.30	742.29	726.49	727.0 [16]
$CH_3$	$\nu_1(1a_1')$	3005.14	2929.93	3037.03	3037.24	3106.81	3113.93	
	$\nu_2(1a_2'')$	561.18	464.75	499.65	570.83	420.27	509.78	
	$\nu_3(2e')$	3167.19	3084.12	3220.96	3211.70	3282.47	3295.36	
	$\nu_4(1e')$	1522.14	1460.25	1438.20	1419.76	1440.48	1422.66	
$CH_4$	$\nu_1(1a_1)$	3100.72	2986.21	3022.33	3032.33	3028.82	3029.21	3030 [17]
	$\nu_2(1e)$	1692.81	1590.23	1552.43	1564.38	1587.01	1565.97	1567 [18]
	$\nu_3(2t_2)$	3191.56	3083.83	3157.54	3138.73	3137.74	3147.34	3158 [18]
	$\nu_4(1t_2)$	1479.28	1389.74	1339.88	1344.43	1375.45	1338.71	1357 [18]
Scale		0.973	0.934	0.983	1.002	0.996	0.997	
RMS Error	f							
Bend		155.7	62.5	14.7	31.0	16.8	10.4	
Stretch		168.3	101.7	80.3 g	47.6	31.5	16.3	
Total		165.4	94.0	69.9 h	44.3	28.7	15.1	

f. We omit the methyl radical,  $\cdot CH_3$ , since the harmonic constants have not been experimentally determined.

g. This value increases to  $219.2\text{ cm}^{-1}$  if we include the highly spin contaminated case of the  $\cdot C=N:$  radical.

h. This value increases to  $191.8\text{ cm}^{-1}$  if we include the highly spin contaminated case of the  $\cdot C=N:$  radical.

**Table 14:** Root-Mean-Square (RMS) errors in scaled calculated harmonic frequencies of Reactants H<sub>2</sub>, CH, NH, OH, HF, N<sub>2</sub>, H<sub>2</sub>O, HCN, CH<sub>4</sub>. (cm<sup>-1</sup>).

Motion	Basis Set	UHF	MP2	Method MP3	B3LYP	QCISD	QCISD(T)
Scale	3-21G(*)	0.973	1.04	1.035	1.052	1.062	1.068
Factors	6-31G*	0.934	0.994	0.989	1.007	1.007	1.013
	6-311G**	0.933	0.983	0.973	1.002	0.989	0.996
	6-311G(2d,p)	0.935	0.989	0.98	1.003	0.996	1.003
	6-311+G(2df,p)	0.935	0.99	0.979	1.003	0.995	1.002
	6-311+G(2df,2pd)	0.934	0.985	0.974	1.003	0.989	0.997
Bend	3-21G(*)	157.7	154.1	149.5	164.6	170.9	171.1
	6-31G*	62.5	51.5	47.7	49.4	64.8	69.7
	6-311G**	53.3	14.7	27.5	31	19.8	15.8
	6-311G(2d,p)	44.2	13.2	11.1	18.1	16.8	16.8
	6-311+G(2df,p)	51	31.7	29.7	26.1	14.6	13.6
Stretch	6-311+G(2df,2pd)	48.5	19.8	26	24	12.7	10.4
	3-21G(*)	151.8	220.8	189.3	221.1	197.7	202.7
	6-31G*	85.2	93.2	74	77.7	54	50.5
	6-311G**	92.8	80.3	46.3	42.6	31.5	34.2
	6-311G(2d,p)	90.3	79.3	41.2	39.2	22.2	26.6
Total	6-311+G(2df,p)	89.4	76.5	42.8	39.5	14.6	12.4
	6-311+G(2df,2pd)	90	79	42.2	40.7	19.7	16.3
	3-21G(*)	152.8	206.1	180.2	208.4	191.3	195.3
	6-31G*	80.2	84.7	68.4	71.7	56.9	55.9
	6-311G**	84.7	69.9	42.4	40	29.1	30.7
	6-311G(2d,p)	81.3	69	36.7	35.2	21	24.5
	6-311+G(2df,p)	81.5	68.1	40	36.6	14.6	12.7
	6-311+G(2df,2pd)	81.6	69.2	38.8	37.2	18.2	15.1

**Table 15:** Errors (theory-experiment) in scaled calculated harmonic frequencies (cm<sup>-1</sup>) for the •C=N: radical.

Basis Set	UHF	MP2	Method MP3	B3LYP	QCISD	QCISD(T)
3-21G(*)	-302.9	649.5	468.8	131.4	80.9	81.4
6-31G*	-217.7	776.8	582.4	106.1	122.8	84.4
6-311G**	-205.0	739.8	525.8	88.0	62.3	23.4
6-311G(2d,p)	-189.5	776.1	565.6	90.3	83.5	36.4
6-311+G(2df,p)	-175.9	777.9	565.6	86.7	83.8	27.7

## Symmetric transition states of degenerate chemical reactions

Now that a benchmark has been selected, we turn our attention to transition states for chemical reactions. John Keith had begun to examine the symmetric transition states. First, we examine symmetric transition states for the six degenerate atom transfer reactions that were considered in the original IRCMax paper: H-H-H, H-F-H, F-H-F, HO-H-OH, NC-H-CN, and CH<sub>3</sub>-H-CH<sub>3</sub>. Since the position of these transition states along the reaction path is determined by the symmetry, they provide a simple test of the inherent accuracy of each computational method for the vibrational frequencies of transition states and zero point energy contributions to the activation energy.

The calculated harmonic frequencies for the symmetric transition states are given in Table 16 & 17. The RMS errors in the Hartree-Fock frequencies for the transition states in Table 16 & 17 are twice that found for the stable molecules in Table 14. This shows that electron correlation plays an increased roll in transition states of chemical reactions compared to reactants and products. A systematic study of the convergence of calculated frequencies with basis set and level of correlation treatment is summarized in Table 18. Total RMS errors over all modes of 264.4cm<sup>-1</sup> persist for Hartree-Fock frequencies with even the largest basis set 6-311G(2d,p) although the smaller 6-31G\* basis set gives a better result, 178.4cm<sup>-1</sup> error compared to the quadratic CI benchmark. This poor result may be attributed to qualitatively incorrect geometries predicted for the H-F-H transition state. The small basis set HF/3-21G\* calculations give a linear H-F-H transition state with a low frequency symmetric stretch; however, all higher calculations in Table 18 give a bent transition state with a symmetric stretching frequency over 1,000 cm<sup>-1</sup>. The total RMS errors observed

in Table 18 indicate that as observed with the stable molecules, the MP2, MP3, and B3LYP reach their limits of accuracy with the 6-311G(2d,p) basis set, as does the QCISD in this case.

It is important to note here that the B3LYP performs three times as well as MP2 and two times better than MP3 for describing the frequencies of transition states, (28.0cm<sup>-1</sup> compared to 88.1 and 52.1cm<sup>-1</sup>, respectively) and is comparable to QCISD (21.6cm<sup>-1</sup>). In the absence of spin contamination, B3LYP hybrid DFT frequencies are more accurate than MP3 frequencies, and they are considerably less expensive than MP2 frequencies. Therefore, so long as the correct location of the transition state along the reaction path is known (and here it is constrained by symmetry), it is computationally practical to achieve good frequencies of transition states. The total zero point energies (ZPE) in reactants (kcal/mol), ZPE changes from the symmetric transition states to their reactants (kcal/mol), and errors in calculated ZPE changes from reactant to transition state compared to the quadratic CI benchmark (kcal/mol) compared to the quadratic CI benchmark are presented in Table 19 & 20.

The errors in ZPE change from reactant to transition state in Table 19 show that the B3LYP with 6-311G\*\* gives errors for the symmetric transition states that are smaller than MP2 by an order of magnitude for H-H-H, H-F-H, F-H-F, HO-H-OH, NC-H-CN, and H<sub>3</sub>C-H-CH<sub>3</sub>. It is comparable to the QCISD with large basis set 6-311G (2d,p), and is faster than MP2. Therefore, as we have observed with the frequencies of transition states, we can conclude that so long as the correct location of the transition state along the reaction path is known, it is also computationally practical to achieve good zero point energy contributions to the activation energy of chemical reactions.

**Table 16:** Scaled harmonic frequencies (cm<sup>-1</sup>) for transition states of degenerate reactions.

Transition State	Mode	Motion	HF/3-21G*	HF/6-31G*	MP2/6-311G**	B3LYP/6-311G**	QCISD/6-311G(2d,p)	QCISD(T)/6-311+G(2df,2pd)
H...H...H	v <sub>1</sub> (σ <sub>v</sub> )	H-H-H stretch	1999.41	1887.94	2116.08	2060.85	2052.3	2043.53
	v <sub>2</sub> (π <sub>u</sub> )	H-H-H bend	1091.12	1050.22	957.13	945.4	947.7	914.73
H...F...H	v <sub>1</sub> (σ <sub>v</sub> )	H-F-H stretch	1910.14	1799.35	2361.28	2111.44	2094.04	2117.8
	v <sub>2</sub> (π <sub>u</sub> )	H-F-H bend	635.28	559.94	456.53	259.84	274.18	250.39
F...H...F	v <sub>1</sub> (2a <sub>1</sub> )	F-H-F stretch	545.42	1077.06	1810.8	1635.22	1652.92	1666.85
	v <sub>2</sub> (1a <sub>1</sub> )	F-H-F bend	264.3	291.59	395.31	432.91	403.79	418.06
HO...H...OH	v <sub>1</sub> (2a)	O-H-O stretch	386.52	395.91	493.33	454.11	477.12	461.57
	v <sub>2</sub> (1b)	O-H-O bend	530.22	535.16	613.6	585.94	589.09	585.59
	v <sub>3</sub> (1a)	HOOH torsion	347.38	325.48	388.86	388.76	333.28	323.59

	$v_4(3a)$	H-O-H- bend	899.37	873.43	1033.44	906.06	933.67	932.7
	$v_5(2b)$	H-O-H- bend $_{\perp}$	1486.62	1391.91	1396.7	1413.08	1396.27	1391.53
	$v_6(4a)$	H-O-H- bend $_{\perp}$	1568.2	1439.36	1710.76	1594.51	1615.28	1633.73
	$v_7(3b)$	-O-H stretch	3657.57	3782.22	3809.07	3795.4	3798.09	3799.48
	$v_8(5a)$	-O-H stretch	3658.14	3788.13	3816.54	3801.01	3804.24	3804.15
NC...H...CN	$v_1(1\sigma_v)$	C-H-C stretch	431.07	407.3	456.03	415.5	422.94	425.15
	$v_2(1\pi_u)$	C-H-C bend	182.17	125.24	91.4	109.49	108.9	103.46
	$v_3(1\pi_v)$	:N=C-H bend	379.33	249.69	347.81	243.05	247.19	242.69
	$v_4(2\pi_u)$	:N=C-H bend $_{\perp}$	1953.41	1024.02	1089.73	902.15	859.67	921.61
	$v_5(1\sigma_u)$	:N=C stretch	1948.61	1983.25	3119.51	2250.78	2205.12	2168.81
	$v_6(2\sigma_v)$	:N=C stretch	1981.24	2012.26	3151.48	2286.8	2243.91	2212.68

**Table 17:** Continued) Scaled harmonic frequencies ( $\text{cm}^{-1}$ ) transition states of degenerate reactions.

Transition State	Mode	Motion	HF/3-21G*	HF/6-31G*	MP2/6-311G**	B3LYP/6-311G** 6-311G**	QCISD/6-311G(2d,p)	QCISD(T)/6-311+G(2df,2pd)
H <sub>3</sub> C...H...CH <sub>3</sub>	$v_1(1a_g)$	C-H-C stretch	483.14	466.72	532.91	506.91	510.50	510.05
	$v_2(1e_u)$	C-H-C bend	380.07	327.48	297.88	332.67	315.43	306.80
	$v_3(1a_u)$	H <sub>3</sub> C-CH <sub>3</sub> torsion	39.78	38.34	51.16	43.22	39.66	17.68
	$v_4(1e_g)$	H-C-H- bend	780.92	717.03	697.83	695.60	713.69	692.83
	$v_5(2e_u)$	H-C-H- bend $_{\perp}$	1492.49	1387.71	1351.09	1360.09	1383.64	1345.06
	$v_6(2a_g)$	H-C-H umbrella	1260.31	1196.29	1169.40	1150.91	1186.42	1144.60
	$v_7(1a_{2u})$	H-C-H umbrella	1303.11	1215.94	1191.34	1181.34	1219.25	1184.98
	$v_8(2e_g)$	H-C-H bend	1562.69	1479.11	1443.94	1442.91	1473.45	1445.29
	$v_9(3e_u)$	H-C-H bend	1613.77	1497.87	1459.26	1460.02	1494.64	1462.49
	$v_{10}(2a_{2u})$	-C-H stretch	3131.54	3019.78	3054.44	3059.90	3059.02	3056.83
	$v_{11}(3a_g)$	-C-H stretch	3133.62	3022.10	3056.02	3061.34	3060.39	3058.23
	$v_{12}(3e_g)$	-C-H stretch	3248.77	3135.48	3199.73	3189.19	3183.98	3190.67
	$v_{13}(4e_u)$	-C-H stretch	3250.35	3136.67	3200.33	3190.20	3184.79	3191.84
Scale			0.973	0.934	0.983	1.002	0.996	0.997
<b>RMS Error</b>								
	Bend		190.4	149.2	87.4	18.1	18.1	-
	Stretch		467.1	282.8	120.9	15.8	13.4	-
	Other		354.1	77.9	82.2	30.2	27.3	-
	Total		355.6	178.4	96.2	23.6	21.6	-

$\perp$  Denotes motion of H perpendicular to the A-H-A axis.

**Table 18:** Root-Mean-Square (RMS) errors in scaled calculated harmonic frequencies ( $\text{cm}^{-1}$ ) for symmetric transition states.

Motion	Basis Set	UHF	MP2	Method MP3	B3LYP	QCISD	QCISD(T)
Stretch	3-21G(*)	467.1	198.3	141.9	159.2	91.4	106.0
	6-31G*	282.8	246.6	131.6	48.9	78.2	84.6
	6-311G**	459.9	120.9	99.9	15.8	26.5	9.0
	6-311G(2d,p)	459.0	130.6	82.7	15.5	13.4	-
Bend	3-21G(*)	190.4	166.4	172.6	112.4	132.5	136.0
	6-31G*	149.2	126.7	111.9	80.0	98.7	101.0
	6-311G**	136.6	87.4	24.7	18.1	14.5	26.2
	6-311G(2d,p)	106.0	56.9	30.2	16.7	18.1	-
Other	3-21G(*)	354.1	465.3	426.7	441.7	419.0	410.4
	6-31G*	77.9	146.7	70.3	116.1	61.7	61.6
	6-311G**	106.5	82.2	40.6	30.2	26.1	25.8
	6-311G(2d,p)	122.9	67.7	34.1	38.5	27.3	-
Total TS	3-21G(*)	355.6	334.6	303.8	307.4	287.5	284.1
	6-31G*	178.4	176.6	103.2	91.1	78.5	81.2
	6-311G**	265.8	96.2	61.1	23.6	23.5	22.4
	6-311G(2d,p)	264.4	88.1	52.1	28.0	21.6	-
Molecular	3-21G(*)	111.2	176.0	162.1	168.0	176.1	180.2
	6-31G*	39.2	68.1	53.6	44.5	59.2	63.2
	6-311G**	63.6	10.4	17.8	3.4	14.6	11.3
	6-311G(2d,p)	79.7	22.2	14.4	3.9	22.0	-

**Table 19:** Total Zero Point Energies (ZPE) in Reactants (kcal/mol). ZPE Changes from Reactant to Transition State (kcal/mol). Errors in calculated ZPE changes from Symmetric TS to Reactant compared to the quadratic CI benchmark (kcal/mol).  $\Delta E = E_{(\text{QCISD(T)/6-311+G(2df,2pd)})} - E_{(j)}$ .

Reactant	HF/3-21G*	HF/6-31G*	MP2/6-311G**	B3LYP/6-311G**	QCISD/6-311G(2d,p)	QCISD(T)/6-311+G(2df,2pd)
H <sub>2</sub>	6.483	6.209	6.267	6.348	6.273	6.265
HF	5.654	5.823	5.878	5.912	5.962	5.944
H <sub>2</sub> O	13.305	13.478	13.480	13.412	13.521	13.532
OH	5.024	5.341	5.328	5.444	5.327	5.353
CN	2.526	2.648	3.950	3.095	3.085	2.991
HCN	10.124	10.025	8.820	9.235	9.025	8.973
CH <sub>3</sub>	12.037	11.550	11.621	11.917	11.762	11.900
CH <sub>4</sub>	29.328	28.024	28.083	28.057	28.232	28.279

**Table 20:** Zero Point Energy Changes from Reactant to Transition State (kcal/mol).

Transition State	HF/3-21G*	HF/6-31G*	MP2/6-311G**	B3LYP/6-311G**	QCISD/6-311G(2d,p)	QCISD(T)/6-311+G(2df,2pd)
H...H...H	-0.500	-0.502	-0.501	-0.694	-0.625	-0.724
H...F...H	-1.103	-1.647	-1.193	-2.148	-2.181	-2.198
F...H...F	-4.117	-3.448	-2.156	-2.334	-2.441	-2.363
HO...H...OH	-0.396	-0.890	0.167	-0.344	-0.325	-0.382
NC...H...CN	0.786	-2.371	1.230	-1.654	-1.661	-1.459
H <sub>3</sub> C...H...CH <sub>3</sub>	7.293	6.670	6.588	6.303	6.611	5.951

Errors in Calculated Zero Point Energy Changes from Reactant to Transition State  
(kcal/mol)  $\Delta E = E_{(\text{QCISD(T)/6-311+G(2df,2pd)})} - E_{(j)}$

H...H...H	-0.223	-0.221	-0.223	-0.029	-0.098	0.000
H...F...H	-1.095	-0.551	-1.005	-0.050	-0.017	0.000
F...H...F	1.754	1.085	-0.207	-0.030	0.078	0.000
HO...H...OH	0.014	0.507	-0.549	-0.038	-0.058	0.000
NC...H...CN	-2.245	0.912	-2.689	0.194	0.202	0.000
H <sub>3</sub> C...H...CH <sub>3</sub>	-1.342	-0.719	-0.637	-0.352	-0.660	0.000

### Asymmetric transition states of non-degenerate chemical reactions

From studies of symmetric transition states, it has become clear that when the position of the transition state is known, DFT methods such as B3LYP that give a reasonable estimate of the correlation energy provide a practical means of obtaining good frequencies of transition states and zero point energy contributions to the activation energy. Now we test the accuracy of these methods (HF, MP2, MP3, B3LYP, QCISD and QCISD (T)) for four non-degenerate chemical reactions to assess the importance that the position of the transition state along the reaction path plays in the estimation of zero point energy changes from reactant to transition state and normal mode frequencies of transition states. We will then show that the IRCMax method improves the accuracy of zero point energy changes relative to the selected benchmark by locating the correct position of the transition state with cost-effective methodologies. The four non-degenerate chemical reactions are presented in Figure 39. Scaled harmonic frequencies of four exothermic hydrogen transfer reactions are presented in Table 21.

The RMS errors in these frequencies compared to our quadratic CI benchmark are given in Table 22. The total errors for the RMS frequencies of asymmetric transition states are large and do not show smooth convergence as was seen for the symmetric transition states. The smallest error (66.2cm<sup>-1</sup>) is achieved only by utilizing the highest level of correlation treatment, QCISD, with the largest basis set 6-311+G(2df,2pd). This is triple the smallest error observed for the symmetric transition states of Table 18, which could be obtained with B3LYP/6-311G\*\* (23.6cm<sup>-1</sup>). For this set of non-degenerate chemical reactions, the B3LYP performs the worst, giving errors near 1000cm<sup>-1</sup>. The B3LYP hybrid DFT is known to underestimate barrier heights, and doesn't predict a transition state for the reaction of H<sub>2</sub> + F → H +

H-F for several basis sets, including 6-31G\*, 6-311G\*\*, 6-311G(2d,p), 6-311+G(2df,p), and 6-311+G(2df,2pd). Also, it fails to predict a transition state in the reaction of H<sub>2</sub> + CN → H + HCN for the 3-21G\* and 6-31G\* basis sets. Hartree-Fock, MP2, and MP3 also perform poorly, giving errors of 684.2, 625.3, and 400.0cm<sup>-1</sup> even with the largest basis set, 6-311+G(2df,2pd). That only the QCISD/6-311+G(2df,2pd) level is adequate for describing frequencies of transition states, suggests when transition state are not constrained by symmetry along the reaction path, limitations of the available methodologies and basis sets make accurate predictions of frequencies of transition states computationally impractical. Table 23 presents errors in the changes in zero point energy (ZPE) from reactant to transition state (kcal/mol) for the four non-degenerate hydrogen transfer reactions under investigation.

The behaviors observed in the frequencies of the asymmetric transition states are reflected in the contribution of the zero point energy toward the activation energy. The B3LYP/6-311G\*\* gives a small error of -0.051 kcal/mol for the reaction of CH<sub>4</sub> + OH → H<sub>2</sub>O + CH<sub>3</sub>, but performs poorly for the other three reactions, and predicts no transition state for H<sub>2</sub> + F → H + H-F. Hartree-Fock and MP2 give errors in ZPE contribution that are large compared to QCISD/6-311G(2d,p). The convergence of change in ZPE from reactant to transition state is also not smooth as it was for the symmetric transition states (Table 19). It appears that when the location of a transition state can vary with computational method, prediction of the zero point energy contribution to the activation energy is restricted to one of the best correlation treatments currently available, the quadratic CI, making such calculations computationally demanding. In non-degenerate chemical reactions, different methods of electron correlation can predict different locations for the transition state along the reaction path, as well as yield different estimations of barrier heights, and exothermicities.

For example, in Figure 40 energy profiles for the reaction of  $\text{H}_2 + \text{OH} \rightarrow \text{H-H-O-H} \rightarrow \text{H} + \text{H}_2\text{O}$  are presented using three different computational methodologies. Without correlation energy, the HF/3-21G\* reaction profile is endothermic, with a transition state that is much later than for the MP2/6-31G\* or B3LYP/6-311G\*\*. The B3LYP gives the smallest estimation of the barrier height and also predicts the earliest transition state.

In the reaction path for  $\text{H}_2 + \text{OH} \rightarrow \text{H-H-O-H} \rightarrow \text{H} + \text{H}_2\text{O}$  is defined by molecular coordinates, O-H inter nuclear distance Figure 41 distance as a function of H-H inter nuclear distance for HF/3-21G\*, MP2/6-31G\*, and B3LYP/6-311G\*\*. Although the energy profiles of Figure 40 show drastically dissimilar energetics along the path, Figure 41 illustrates that all three paths are quite similar with the location of the transition state varying with different methodologies. In the Intrinsic Reaction Maximum Coordinate (IRC Max) Method of Petersson & co-workers [108], geometries of transition states and barrier heights of chemical reactions were shown to be improved by locating the correct point for the transition state along the reaction path. Figure 5.4.4 depicts the IRC Max method for the reaction of  $\text{H}_2 + \text{OH} \rightarrow \text{H-H-O-H} \rightarrow \text{H} + \text{H}_2\text{O}$ .

The unrestricted Hartree-Fock gives an energy profile that is endothermic with the 3-21G basis set, whereas the MP2/6-31G\* profile is exothermic and has an earlier transition state. As we have seen for this reaction in Figure 41, the reaction paths defined in terms of reaction coordinate are similar, but differ in the location of the transition state. In the IRCMax method, we select the maximum of a high level energy (MP2/6-31G\*) along a lower-level reaction path (HF/3-21G). When this is done, Figure 42 reveals that the MP2/6-31G\* energies along the HF/3-21G path (denoted [MP2/6-31G\*/HF/3-21G\*]) are in quantitative agreement with the MP2/6-31G\* energy profile. Therefore, by moving toward the correct transition state along the reaction path,

we are able to improve geometries of transition states and barrier heights. We will now show that since the zero point energy change along the reaction path changes similarly for different levels of theory, that the IRCMaximum method can be used to improve our ability to predict ZPE changes by taking us to the correct point along the path.

In Figure 43 we examine the ZPE change as a function of reaction coordinate for the reaction  $\text{H}_2 + \text{OH} \rightarrow \text{H} + \text{H}_2\text{O}$ . The HF/3-21G, MP2/6-31G\*, and B3LYP/6-311G\*\* methodologies have qualitatively similar changes in zero point energy compared to reactants along the reaction path, and the behavior is similar to that seen for the triatomic degenerate reaction,  $\text{H}_2 + \text{H} \rightarrow \text{H}_2 + \text{H}$ , shown in Figure 38. Also plotted are the QCISD/6-311G\*\* and QCISD (T)/6-311+G (2df,2pd) transition state locations for reference. However, since no analytical second derivatives are available for quadratic CI, intrinsic reaction pathways are not currently attainable. The methods predict transition states that are in different regions of the curve where the zero point energy is changing rapidly. Therefore, being at the correct location of the transition state is requisite for accurate prediction of zero point energy contributions toward the activation energy. Figure 43 shows unphysical "shoulders" for the MP2 and B3LYP methods, occurring at +/- 0.4 Angstroms in the reaction coordinate. This behavior is present also for the triatomic degenerate reaction,  $\text{H}_2 + \text{H} \rightarrow \text{H}_2 + \text{H}$ . We will discuss this point further in Section 4.6 as a limitation of the IRCMax method for predicting frequencies and zero point energy changes. We will show from the triatomic reaction of  $\text{H}_2 + \text{H}$  that because the reaction path is defined in Cartesian rather than polar coordinates, small errors in geometry along the reaction path far from the transition state result in unphysical bending modes. These unphysical modes sum to give unphysical zero point energies, and hence erroneous zero point energy changes along the reaction path.

**Table 21:** Scaled Harmonic Frequencies of Asymmetric Transition States ( $\text{cm}^{-1}$ ).

Transition State	Mode	Motion	HF/3-21G*	HF/6-31G*	MP2/6-311G**	B3LYP/6-311G**	QCISD/6-311G(2d,p)	QCISD(T)/6-311+G(2df,p)
HHF	$\nu_1(2a')$	HHF stretch	1478.1	1661.1	2941.4		3146.6	3603.5
	$\nu_2(1a')$	HHF "rotation"	636.8	629.6	148.2		158.0	232.0
HHOH	$\nu_1(2a')$	HHO stretch	1316.0	1521.1	1236.9	3117.1	2227.6	2506.3
	$\nu_2(1a')$	HHO bend	658.0	662.5	1114.8	1001.7	507.3	607.9

Transition State	Mode	Motion	HF/ 3-21G*	HF/ 6-31G*	MP2 6-311G**	B3LYP 6-311G**	QCISD/6- 311G(2d,p)	QCISD(T)/6- 311+G(2df,p)
	$\nu_3(1a'')$	HHO bend	719.1	770.7	907.8	778.6	671.6	543.8
	$\nu_4(3a')$	HOH bend	1622.1	1351.3	3079.7	1057.3	1165.5	1054.9
	$\nu_5(4a')$	OH stretch	3638.4	3762.9	3568.1	3713.0	3756	3746.6
HHCN	$\nu_1(2\sigma)$	HHC stretch	2507.5	2120.7	3486.3	4190.2	2819.0	3084.5
	$\nu_2(1\pi)$	:NCH bend	240.9	180.6	86.4	54.2	145.1	117.6
	$\nu_3(2\pi)$	HHC bend	1012.3	887.9	399.8	204.1	657.3	508.8
	$\nu_4(1\sigma)$	NC stretch	1856.9	1896.4	2698.6	2166.6	2153.8	2127.4
H3CHOH	$\nu_1(2a')$	CHO stretch	531.5	518.2	714.1	710.2176	635	688.8
	$\nu_2(2a'')$	CHO bend	401.7	362.5	261.2	317.1	337.9	325.0
	$\nu_3(1a')$	CHO bend	392.8	358.3	338.3	376.0	420.4	333.1
	$\nu_4(1a'')$	HCHOH torsion	41.8	45.4	8.7	32.3	60.6	24.4
	$\nu_5(3a')$	HCH-bend	927.7	907.2	862.5	880.2	903.0	891.7
	$\nu_6(3a'')$	HCH-bend	1179.6	1147.2	1138.3	1155.1	1208.6	1205.3
	$\nu_7(3a')$	HCH umbrella	1285.722	1238.0	1206.1	1254.9	1357.4	1283.0
	$\nu_8(5a')$	HCH bend	1541.816	1456.4	1311.1	1381.5	1429.7	1366.6
	$\nu_9(1a'')$	HCH bend	1553.784	1475.8	1376.7	1440.5	1485.8	1462.4
	$\nu_{10}(1a')$	HOH bend	1647.484	1511.7	1410.2	1467.1	1506.9	1481.7
	$\nu_{11}(1a')$	(C-H) <sub>3</sub> stretch	3143.86	3032.7	2906.7	3063.8	3070.4	3067.7
	$\nu_{12}(2a'')$	(C-H) <sub>2</sub> stretch	3358.8	3153.1	3032.3	3184.0	3186.7	3196.3
	$\nu_{13}(5a'')$	(C-H) <sub>1</sub> stretch	3268.988	3378.3	3035.6	3189.7	3192.0	3199.3
	$\nu_{14}(9a')$	OH stretch	3640.188	3774.5	3849.5	3761.2	3769.2	3757.7

No transition structure was found for B3LYP/6-311G\*\*

Table 21a

Transition State	Mode	Motion	HF/ 3-21G*	HF/ 6-31G*	HF 6-311G**	HF/ 6-311G(2d,p)	HF/ 6-311+G(2df,p)	HF/ 6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1478.1	1661.1	1758.7	1762.3	1779.1	1776.6
	$v_2(1a')$	HHF "rotation"	636.8	629.6	576.4	564.6	571.6	527.7
HHOH	$v_1(2a')$	HHO stretch	1316.0	152.1	1402.4	1251.4	1399.4	1385.6
	$v_2(1a')$	HHO bend	658.0	662.5	1225.6	652.4	1234.1	1228.7
	$v_3(1a'')$	HHO bend	719.1	770.7	873.3	644.1	841.0	865.8
	$v_4(3a')$	HOH bend	1622.1	1351.3	2177.4	1591.3	2142.5	2131.9
	$v_5(4a')$	OH stretch	3638.4	3762.9	3785.8	3818.8	3777.0	3785.0
HHCN	$v_1(2\sigma)$	HHC stretch	2507.5	2120.7	2288.0	2253.8	2312.0	2299.6
	$v_2(1\pi)$	:NCH bend	240.9	180.6	167.2	160.4	163.6	163.1
	$v_3(2\pi)$	HHC bend	1012.3	887.9	601.5	597.1	595.5	574.0
	$v_4(1\sigma)$	NC stretch	1856.9	1896.4	1916.0	1928.5	1944.9	1942.7
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	531.5	518.2	538.2	536.8	459.0	530.7
	$v_2(2a'')$	CHO bend	401.7	362.5	371.0	374.3	474.0	396.5
	$v_3(1a')$	CHO bend	392.8	358.3	366.9	368.5	522.7	367.9
	$v_4(1a'')$	HCHOH torsion	41.8	45.4	51.4	45.3	586.2	27.8
	$v_5(3a')$	HCH-bend	927.7	907.2	894.7	899.3	942.0	898.3
	$v_6(3a'')$	HCH-bend	1179.6	1147.2	1141.8	1141.5	1102.1	1140.3
	$v_7(3a')$	HCH umbrella	1285.7	1238.0	1222.3	1226.0	1178.8	1228.7
	$v_8(5a')$	HCH bend	1541.8	1456.4	1415.2	1425.5	1440.5	1412.6
	$v_9(1a'')$	HCH bend	1553.8	1475.8	1447.8	1454.2	1442.1	1452.6
	$v_{10}(1a')$	HOH bend	1647.5	1551.7	1470.1	1480.0	1533.2	1470.3
	$v_{11}(1a')$	(C-H)3 stretch	3143.8	3032.7	2980.9	2985.5	2990.6	2986.1
	$v_{12}(2a'')$	(C-H)2 stretch	3358.8	3152.1	3100.1	3102.0	3115.5	3103.0
	$v_{13}(5a'')$	(C-H)1 stretch	3268.9	3378.1	3101.0	3102.4	3116.8	3103.1
	$v_{14}(9a')$	OH stretch	3640.2	3774.5	3826.7	3826.207	3453.9	3832.7
Scale Factors			0.973	0.934	0.933	0.935	0.935	0.934

Table 21b

Transition State	Mode	Motion	MP2/ 3-21G*	MP2/ 6-31G*	MP2 6-311G**	MP2/ 6-311G(2d,p)	MP2/ 6-311+G(2df,p)	MP2/ 6-311+G(2df,2pd)
HHF	$\nu_1(2a')$	HHF stretch	2019.5	2628.6	2941.4	3145.5	3361.8	3454.6
	$\nu_2(1a')$	HHF "rotation"	606.4	366.1	148.2	76.5	122.2	176.7
HHOH	$\nu_1(2a')$	HHO stretch	1634.2	1464.4	1236.9	1279.3	1280.6	1058.7
	$\nu_2(1a')$	HHO bend	1398.5	1242.4	1114.8	1103.9	1105.0	590.8
	$\nu_3(1a'')$	HHO bend	1084.5	936.6	907.8	885.8	886.7	507.8
	$\nu_4(3a')$	HOH bend	2562.5	3044.2	3079.7	3356.5	3359.9	2698.5
	$\nu_5(4a')$	OH stretch	3508.4	3674.3	3568.1	3728.6	3732.4	3771.6
HHCN	$\nu_1(2\sigma)$	HHC stretch	3715.6	3417.6	4190.2	3683.6	3791.7	3779.5
	$\nu_2(1\pi)$	:NCH bend	194.4	117.0	86.4	76.6	50.6	48.4
	$\nu_3(2\pi)$	HHC bend	865.4	700.3	399.8	403.6	297.9	349.6
	$\nu_4(1\sigma)$	NC stretch	2814.0	2890.2	2698.6	2874.3	2884.3	2869.5
	H <sub>3</sub> CHOH	$\nu_1(2a')$	CHO stretch	636.2	693.7	710.2	752.8	850.8
$\nu_2(2a'')$		CHO bend	380.1	256.6	317.1	298.6	181.8	320.2
$\nu_3(1a')$		CHO bend	426.9	348.2	376.0	349.8	298.2	328.7
$\nu_4(1a'')$		HCHOH torsion	7.1	14.5	32.3	50.4	478.5	30.2
$\nu_5(3a')$		HCH-bend	976.6	952.0	880.2	913.4	1030.3	919.7
$\nu_6(3a'')$		HCH-bend	1219.4	1172.0	1155.1	1224.0	1377.8	1203.8
$\nu_7(3a')$		HCH umbrella	1369.2	1315.5	1254.9	1296.4	1416.7	1284.0
$\nu_8(5a')$		HCH bend	1598.8	1476.4	1381.5	1403.3	1463.2	1369.4
$\nu_9(1a'')$		HCH bend	1609.6	1505.0	1440.5	1480.8	1575.5	1464.7
$\nu_{10}(1a')$		HOH bend	1714.6	1538.8	1467.1	1515.2	1598.0	1483.4
$\nu_{11}(1a')$		(C-H) <sub>3</sub> stretch	3249.7	3135.4	3063.8	3076.4	3119.0	3074.4
$\nu_{12}(2a'')$		(C-H) <sub>2</sub> stretch	3379.3	3270.8	3184.0	3203.1	3221.1	3206.9
$\nu_{13}(5a'')$		(C-H) <sub>1</sub> stretch	3385.6	3276.2	3189.7	3208.2	3224.5	3210.2
$\nu_{14}(9a')$		OH stretch	3577.0	3714.6	3761.2	3762.7	3796.2	3781.0
Scale Factors			1.040	0.994	0.983	0.989	0.990	0.985

Table 21c

Transition State	Mode	Motion	MP2/ 3-21G*	MP2/ 6-31G*	MP2/ 6-311G**	MP2/ 6-311G(2d,p)	MP2/ 6-311+G(2df,p)	MP2/ 6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	2019.5	2628.6	2941.4	3145.5	3361.8	3454.6
	$v_2(1a')$	HHF "rotation"	606.4	366.1	148.2	76.5	122.2	176.7
HHOH	$v_1(2a')$	HHO stretch	1634.2	1464.4	1236.9	1279.3	1280.6	1058.7
	$v_2(1a')$	HHO bend	1398.5	1242.4	1114.8	1103.9	1105.0	590.8
	$v_3(1a'')$	HHO bend	1084.5	936.6	907.8	885.8	886.7	507.8
	$v_4(3a')$	HOH bend	2562.5	3044.2	3079.7	3356.5	3359.9	2698.5
	$v_5(4a')$	OH stretch	3508.4	3674.3	3568.1	3728.6	3732.4	3771.6
HHCN	$v_1(2\sigma)$	HHC stretch	3715.6	3417.6	4190.2	3683.6	3791.7	3779.5
	$v_2(1\pi)$	:NCH bend	194.4	117.0	86.4	76.6	50.6	48.4
	$v_3(2\pi)$	HHC bend	865.4	700.3	399.8	403.6	297.9	349.6
	$v_4(1\sigma)$	NC stretch	2814.0	2890.2	2698.6	2874.3	2884.3	2869.5
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	636.2	693.7	710.2	752.8	850.8	783.0
	$v_2(2a'')$	CHO bend	380.1	256.6	317.1	298.6	181.8	320.2
	$v_3(1a')$	CHO bend	426.9	348.2	376.0	349.8	298.2	328.7
	$v_4(1a'')$	HCHOH torsion	7.1	14.5	32.3	50.4	478.5	30.2
	$v_5(3a')$	HCH-bend	976.6	952.0	880.2	913.4	1030.3	919.7
	$v_6(3a'')$	HCH-bend	1219.4	1172.0	1155.1	1224.0	1377.8	1203.8
	$v_7(3a')$	HCH umbrella	1369.2	1315.5	1254.9	1296.4	1416.7	1284.0
	$v_8(5a')$	HCH bend	1598.8	1476.4	1381.5	1403.3	1463.2	1369.4
	$v_9(1a'')$	HCH bend	1609.6	1505.0	1440.5	1480.8	1575.5	1464.7
	$v_{10}(1a')$	HOH bend	1714.6	1538.8	1467.1	1515.2	1598.0	1483.4
	$v_{11}(1a')$	(C-H) <sub>3</sub> stretch	3249.7	3135.4	3063.8	3076.4	3119.0	3074.4
	$v_{12}(2a'')$	(C-H) <sub>2</sub> stretch	3379.3	3270.8	3184.0	3203.1	3221.1	3206.9
	$v_{13}(5a'')$	(C-H) <sub>1</sub> stretch	3385.6	3276.2	3189.7	3208.2	3224.5	3210.2
	$v_{14}(9a')$	OH stretch	3577.0	3714.6	3761.2	3762.7	3796.2	3781.0
Scale Factors			1.040	0.994	0.983	0.989	0.990	0.985

Table 21d

Transition State	Mode	Motion	MP3/ 3-21G*	MP3/ 6-31G*	MP3/ 6-311G**	MP3/ 6-311G(2d,p)	MP3/ 6-311+G(2df,p)	MP3/ 6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1871.6	1788.4	2821.6	2842.9	3025.3	2951.6
	$v_2(1a')$	HHF "rotation"	629.7	406.6	315.3	268.3	144.8	284.1
ZP (kcal/mol)			3.579	3.140	4.488	4.451	4.536	4.629
HHOH	$v_1(2a')$	HHO stretch	1803.6	1956.0	2918.7	2241.8	3025.7	3250.2
	$v_2(1a')$	HHO bend	735.5	669.1	1181.9	673.8	1110.5	1302.2
	$v_3(1a'')$	HHO bend	735.8	619.2	913.8	532.5	799.9	0.0
	$v_4(3a')$	HOH bend	1452.4	1273.2	1348.8	1170.2	1329.9	1303.1
ZP (kcal/mol)			6.763	6.392	9.104	6.608	8.965	8.378
	$v_5(4a')$	OH stretch	3556.6	3695.1	3743.0	3751.4	3720.6	3874.4
HHCN	$v_1(2\sigma)$	HHC stretch	3472.4	3249.6	3524.2	3533.7	3845.7	3629.1
	$v_2(1\pi)$	:NCH bend	211.8	128.9	95.7	96.5	60.4	74.0
	$v_3(2\pi)$	HHC bend	903.5	696.3	426.9	408.3	257.0	364.3
	$v_4(1\sigma)$	NC stretch	2635.7	2693.9	2626.7	2661.4	2668.5	2654.2
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	609.2	634.9	661.5	660.6	669.5	669.9
	$v_2(2a'')$	CHO bend	383.0	271.8	306.5	330.6	333.3	345.2
	$v_3(1a')$	CHO bend	417.1	349.8	368.2	364.4	353.8	355.2
	$v_4(1a'')$	HCHOH torsion	19.8	0.0	29.5	47.8	0.0	0.0
	$v_5(3a')$	HCH-bend	956.3	943.9	909.4	914.5	900.7	910.9
	$v_6(3a'')$	HCH-bend	1196.1	1132.4	1161.0	1189.2	1202.9	1188.8
	$v_7(3a')$	HCH umbrella	1339.4	1287.2	1245.2	1268.4	1257.2	1241.4
	$v_8(5a')$	HCH bend	1579.5	1469.8	1389.8	1411.1	1364.9	1364.1
	$v_9(1a'')$	HCH bend	1587.0	1484.4	1432.4	1463.7	1444.6	1437.3
	$v_{10}(1a')$	HOH bend	1707.4	1523.8	1467.1	1496.3	1468.5	1464.0
	$v_{11}(1a')$	(C-H)3 stretch	3218.2	3112.7	3030.9	3046.5	3041.5	3037.2
	$v_{12}(2a'')$	(C-H)2 stretch	3344.2	3241.4	3157.6	3166.6	3170.0	3163.6
	$v_{13}(5a'')$	(C-H)1 stretch	3349.9	3248.1	3159.5	3170.1	3173.8	3167.3
	$v_{14}(9a')$	OH stretch	3562.5	3712.6	3781.7	3762.6	3769.1	3781.0
Scale Factors			1.035	0.989	0.973	0.98	0.979	0.974

Table 21e

Transition State	Mode	Motion	B3LYP/ 3-21G*	B3LYP/ 6-31G*	B3LYP/ 6-311G**	B3LYP/ 6-311G(2d,p)	B3LYP/ 6-311+G(2df,p)	B3LYP/ 6-311+G(2df,2pd)
HHF	$\nu_1(2a')$	HHF stretch	2365.6					
	$\nu_2(1a')$	HHF "rotation"	265.2					
HHOH	$\nu_1(2a')$	HHO stretch	1844.8	2292.1	3123.3	2882.0	3084.1	3748.7
	$\nu_2(1a')$	HHO bend	686.9	686.1	1003.7	636.4	575.6	558.5
	$\nu_3(1a'')$	HHO bend	429.2	513.2	780.1	400.3	500.9	474.8
	$\nu_4(3a')$	HOH bend	1441.0	1192.9	1059.4	1022.6	932.1	913.2
ZP (kcal/mol)			6.298	6.702	8.536	7.070	7.286	8.148
	$\nu_5(4a')$	OH stretch	3599.0	2691.8	3720.4	3742.4	3741.6	3748.7
HHCN	$\nu_1(2\sigma)$	HHC stretch			4198.6	4221.0	4320.9	
	$\nu_2(1\pi)$	:NCH bend			54.3	53.4	35.2	
	$\nu_3(2\pi)$	HHC bend			204.5	186.7	57.4	
	$\nu_4(1\sigma)$	NC stretch			2171.0	2167.6	2162.0	
H <sub>3</sub> CHOH	$\nu_1(2a')$	CHO stretch	654.2	605.4	704.2	709.0	760.8	778.4
	$\nu_2(2a'')$	CHO bend	312.0	284.2	317.1	324.7	354.5	355.0
	$\nu_3(1a')$	CHO bend	416.2	371.4	376.0	378.7	375.2	366.2
	$\nu_4(1a'')$	HCHOH torsion	28.7	43.7	32.3	53.2	13.1	-10.9
	$\nu_5(3a')$	HCH-bend	910.8	907.5	880.2	892.1	889.9	910.1
	$\nu_6(3a'')$	HCH-bend	1018.1	1111.9	1155.1	1157.5	1194.9	1197.8
	$\nu_7(3a')$	HCH umbrella	1321.9	1257.6	1254.9	1259.1	1280.4	1286.5
	$\nu_8(5a')$	HCH bend	1561.5	1450.1	1381.5	1391.9	1349.1	1350.8
	$\nu_9(1a'')$	HCH bend	1564.0	1475.4	1440.5	1448.0	1445.9	1448.5
	$\nu_{10}(1a')$	HOH bend	1755.3	1503.7	1467.1	1474.3	1464.1	1464.1
	$\nu_{11}(1a')$	(C-H)3 stretch	3251.2	3112.4	3063.8	3068.0	3072.2	3072.2
	$\nu_{12}(2a'')$	(C-H)2 stretch	3394.2	3240.0	3184.0	3185.2	3189.7	3189.7
	$\nu_{13}(5a'')$	(C-H)1 stretch	3398.8	3243.5	3189.7	3191.2	3194.5	3194.5
	$\nu_{14}(9a')$	OH stretch	3599.1	3718.2	3761.2	3765.5	3771.6	3771.6
Scale Factors			1.052	1.007	1.002	1.003	1.003	1.003

Table 21 f

Transition State	Mode	Motion	QCISD/ 3-21G*	QCISD/ 6-31G*	QCISD/ 6-311G**	QCISD/ 6-311G(2d,p)	QCISD/ 6-311+G(2df,p)	QCISD/ 6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1754.2	2568.0	3085.2	3134.0	3487.8	3559.2
	$v_2(1a')$	HHF "rotation"	602.3	395.7	215.8	147.4	212.2	204.9
HHOH	$v_1(2a')$	HHO stretch	1725.3	1881.2	2559.2	2227.9	2363.6	2405.4
	$v_2(1a')$	HHO bend	664.9	605.3	638.0	507.3	551.5	553.6
	$v_3(1a'')$	HHO bend	717.2	672.2	1023.8	671.6	620.4	612.8
	$v_4(3a')$	HOH bend	1436.1	1278.0	1144.4	1165.5	1086.6	1104.4
	$v_5(4a')$	OH stretch	3557.8	3697.6	3743.2	3756.0	3755.3	3765.2
HHCN	$v_1(2\sigma)$	HHC stretch	2548.0	2536.5	2822.4	2807.7	3023.6	2970.3
	$v_2(1\pi)$	:NCH bend	269.5	171.3	141.8	144.5	128.4	128.2
	$v_3(2\pi)$	HHC bend	1152.9	888.2	676.4	654.6	556.2	613.8
	$v_4(1\sigma)$	NC stretch	2167.9	2179.8	2131.6	2145.1	2168.4	2155.1
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	580.3	591.8	636.8	635.0	644.7	647.5
	$v_2(2a'')$	CHO bend	366.3	249.1	289.8	337.9	332.3	356.6
	$v_3(1a')$	CHO bend	413.1	344.8	368.4	420.4	351.1	347.4
	$v_4(1a'')$	HCHOH torsion	8.1	0.0	25.8	60.6	0.0	0.0
	$v_5(3a')$	HCH-bend	948.2	925.9	906.1	903.0	893.6	904.9
	$v_6(3a'')$	HCH-bend	1173.7	1143.9	1180.1	1208.6	1199.8	1187.8
	$v_7(3a')$	HCH umbrella	1336.0	1284.4	1255.5	1357.4	1270.1	1255.2
	$v_8(5a')$	HCH bend	1598.9	1471.7	1407.7	1429.7	1375.6	1379.1
	$v_9(1a'')$	HCH bend	1605.3	1495.7	1448.2	1485.8	1459.9	1451.8
	$v_{10}(1a')$	HOH bend	1729.8	1528.2	1479.6	1506.9	1478.6	1473.3
$v_{11}(1a')$	(C-H)3 stretch	3252.1	3130.6	3050.8	3070.4	3063.1	3056.9	
$v_{12}(2a'')$	(C-H)2 stretch	3382.5	3261.9	3176.9	3186.7	3191.5	3182.8	
$v_{13}(5a'')$	(C-H)1 stretch	3384.8	3264.2	3169.5	3192.0	3194.4	3185.3	
$v_{14}(9a')$	OH stretch	3560.1	3714.6	3742.7	3769.2	3770.9	3780.7	
Scale Factors			1.062	1.007	0.989	0.996	0.995	0.989

Table 21g

Transition State	Mode	Motion	QCISD/3-21G*	QCISD/6-31G*	QCISD/6-311G**	QCISD/6-311G(2d,p)	QCISD/6-311+G(2df,p)	QCISD/6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1754.2	2568.0	3085.2	3134.0	3487.8	3559.2
	$v_2(1a')$	HHF "rotation"	602.3	395.7	215.8	147.4	212.2	204.9
HHOH	$v_1(2a')$	HHO stretch	1725.3	1881.2	2559.2	2227.9	2363.6	2405.4
	$v_2(1a')$	HHO bend	664.9	605.3	638.0	507.3	551.5	553.6
	$v_3(1a'')$	HHO bend	717.2	672.2	1023.8	671.6	620.4	612.8
	$v_4(3a')$	HOH bend	1436.1	1278.0	1144.4	1165.5	1086.6	1104.4
HHCN	$v_5(4a')$	OH stretch	3557.8	3697.6	3743.2	3756.0	3755.3	3765.2
	$v_1(2\sigma)$	HHC stretch	2548.0	2536.5	2822.4	2807.7	3023.6	2970.3
	$v_2(1\pi)$	:NCH bend	269.5	171.3	141.8	144.5	128.4	128.2
	$v_3(2\pi)$	HHC bend	1152.9	888.2	676.4	654.6	556.2	613.8
	$v_4(1\sigma)$	NC stretch	2167.9	2179.8	2131.6	2145.1	2168.4	2155.1
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	580.3	591.8	636.8	635.0	644.7	647.5
	$v_2(2a'')$	CHO bend	366.3	249.1	289.8	337.9	332.3	356.6
	$v_3(1a')$	CHO bend	413.1	344.8	368.4	420.4	351.1	347.4
	$v_4(1a'')$	HCHOH torsion	8.1	0.0	25.8	60.6	0.0	0.0
	$v_5(3a')$	HCH-bend	948.2	925.9	906.1	903.0	893.6	904.9
	$v_6(3a'')$	HCH-bend	1173.7	1143.9	1180.1	1208.6	1199.8	1187.8
	$v_7(3a')$	HCH umbrella	1336.0	1284.4	1255.5	1357.4	1270.1	1255.2
	$v_8(5a')$	HCH bend	1598.9	1471.7	1407.7	1429.7	1375.6	1379.1
	$v_9(1a'')$	HCH bend	1605.3	1495.7	1448.2	1485.8	1459.9	1451.8
	$v_{10}(1a')$	HOH bend	1729.8	1528.2	1479.6	1506.9	1478.6	1473.3
	$v_{11}(1a')$	(C-H)3 stretch	3252.1	3130.6	3050.8	3070.4	3063.1	3056.9
	$v_{12}(2a'')$	(C-H)2 stretch	3382.5	3261.9	3176.9	3186.7	3191.5	3182.8
	$v_{13}(5a'')$	(C-H)1 stretch	3384.8	3264.2	3169.5	3192.0	3194.4	3185.3
	$v_{14}(9a')$	OH stretch	3560.1	3714.6	3742.7	3769.2	3770.9	3780.7
Scale Factors			1.062	1.007	0.989	0.996	0.995	0.989

Table 21h

Transition State	Mode	Motion	QCISD/321G*	QCISD/6-31G*	QCISD/6-311G**	QCISD/6-311G(2d,p)	QCISD/6-311+G(2df,p)	QCISD/6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1754.2	2568.0	3085.2	3134.0	3487.8	3559.2
	$v_2(1a')$	HHF "rotation"	602.3	395.7	215.8	147.4	212.2	204.9
HHOH	$v_1(2a')$	HHO stretch	1725.3	1881.2	2559.2	2227.9	2363.6	2405.4
	$v_2(1a')$	HHO bend	664.9	605.3	638.0	507.3	551.5	553.6
	$v_3(1a'')$	HHO bend	717.2	672.2	1023.8	671.6	620.4	612.8
	$v_4(3a')$	HOH bend	1436.1	1278.0	1144.4	1165.5	1086.6	1104.4
HHCN	$v_5(4a')$	OH stretch	3557.8	3697.6	3743.2	3756.0	3755.3	3765.2
	$v_1(2\sigma)$	HHC stretch	2548.0	2536.5	2822.4	2807.7	3023.6	2970.3
	$v_2(1\pi)$	:NCH bend	269.5	171.3	141.8	144.5	128.4	128.2
	$v_3(2\pi)$	HHC bend	1152.9	888.2	676.4	654.6	556.2	613.8
H <sub>3</sub> CHOH	$v_4(1\sigma)$	NC stretch	2167.9	2179.8	2131.6	2145.1	2168.4	2155.1
	$v_1(2a')$	CHO stretch	580.3	591.8	636.8	635.0	644.7	647.5
	$v_2(2a'')$	CHO bend	366.3	249.1	289.8	337.9	332.3	356.6
	$v_3(1a')$	CHO bend	413.1	344.8	368.4	420.4	351.1	347.4
	$v_4(1a'')$	HCHOH torsion	8.1	0.0	25.8	60.6	0.0	0.0
	$v_5(3a')$	HCH-bend	948.2	925.9	906.1	903.0	893.6	904.9
	$v_6(3a'')$	HCH-bend	1173.7	1143.9	1180.1	1208.6	1199.8	1187.8
	$v_7(3a')$	HCH umbrella	1336.0	1284.4	1255.5	1357.4	1270.1	1255.2
	$v_8(5a')$	HCH bend	1598.9	1471.7	1407.7	1429.7	1375.6	1379.1
	$v_9(1a'')$	HCH bend	1605.3	1495.7	1448.2	1485.8	1459.9	1451.8
	$v_{10}(1a')$	HOH bend	1729.8	1528.2	1479.6	1506.9	1478.6	1473.3
	$v_{11}(1a')$	(C-H)3 stretch	3252.1	3130.6	3050.8	3070.4	3063.1	3056.9
	$v_{12}(2a'')$	(C-H)2 stretch	3382.5	3261.9	3176.9	3186.7	3191.5	3182.8
	$v_{13}(5a'')$	(C-H)1 stretch	3384.8	3264.2	3169.5	3192.0	3194.4	3185.3
$v_{14}(9a')$	OH stretch	3560.1	3714.6	3742.7	3769.2	3770.9	3780.7	
Scale Factors			1.062	1.007	0.989	0.996	0.995	0.989

Table 21i

Transition State	Mode	Motion	QCISDT/3-21G*	QCISDT/6-31G*	QCISDT/6-311G**	QCISDT/6-311G(2d,p)	QCISDT/6-311+G(2df,p)	QCISDT/6-311+G(2df,2pd)
HHF	$v_1(2a')$	HHF stretch	1787.9	2625.0	3173.5	3247.3	3603.5	3702.7
	$v_2(1a')$	HHF "rotation"	595.1	391.3	142.3	39.7	232.0	215.0
HHOH	$v_1(2a')$	HHO stretch	1736.7	1913.9	2367.0	2349.6	2506.3	2582.3
	$v_2(1a')$	HHO bend	722.6	671.2	672.3	664.9	607.9	592.7
	$v_3(1a'')$	HHO bend	660.0	600.5	469.2	482.2	543.8	529.1
	$v_4(3a')$	HOH bend	1443.8	1274.5	1130.5	1138.9	1054.9	1068.3
	$v_5(4a')$	OH stretch	3548.0	3692.7	3764.6	3749.1	3746.6	3760.4
HHCN	$v_1(2\sigma)$	HHC stretch	2634.6	2548.8	2873.6	2474.3	3084.5	3027.9
	$v_2(1\pi)$	:NCH bend	260.8	166.1	147.3	153.4	117.6	125.4
	$v_3(2\pi)$	HHC bend	1132.7	889.6	681.9	661.5	508.8	616.8
	$v_4(1\sigma)$	NC stretch	2198.2	2167.0	2113.5	2128.2	2127.4	2119.1
H <sub>3</sub> CHOH	$v_1(2a')$	CHO stretch	581.7	597.9	666.2	663.5	688.8	701.0
	$v_2(2a'')$	CHO bend	350.5	222.3	264.1	294.9	325.0	340.8
	$v_3(1a')$	CHO bend	400.4	343.9	364.8	361.6	333.1	342.0
	$v_4(1a'')$	HCHOH torsion	0.0	0.0	0.0	52.9	24.4	20.2
	$v_5(3a')$	HCH-bend	941.3	926.5	905.3	911.0	891.7	906.7
	$v_6(3a'')$	HCH-bend	1339.1	1137.6	1153.8	1180.3	1205.3	1194.7
	$v_7(3a')$	HCH umbrella	1339.1	1287.3	1263.2	1287.1	1283.0	1269.1
	$v_8(5a')$	HCH bend	1602.6	1473.7	1402.0	1424.7	1366.6	1368.1
	$v_9(1a'')$	HCH bend	1608.5	1498.9	1449.0	1481.9	1462.4	1454.6
	$v_{10}(1a')$	HOH bend	1730.6	1530.0	1480.5	1511.0	1481.7	1476.0
	$v_{11}(1a')$	(C-H) <sub>3</sub> stretch	3266.7	3134.2	3055.6	3072.8	3067.7	3062.7
	$v_{12}(2a'')$	(C-H) <sub>2</sub> stretch	3399.9	3266.4	3181.8	3191.4	3196.3	3189.5
	$v_{13}(5a'')$	(C-H) <sub>1</sub> stretch	3401.3	3268.5	3183.6	3195.5	3199.3	3192.6
	$v_{14}(9a')$	OH stretch	3575.9	3706.3	3776.1	3758.4	3757.7	3771.9
Scale Factors			1.068	1.013	0.996	1.003	1.002	0.997

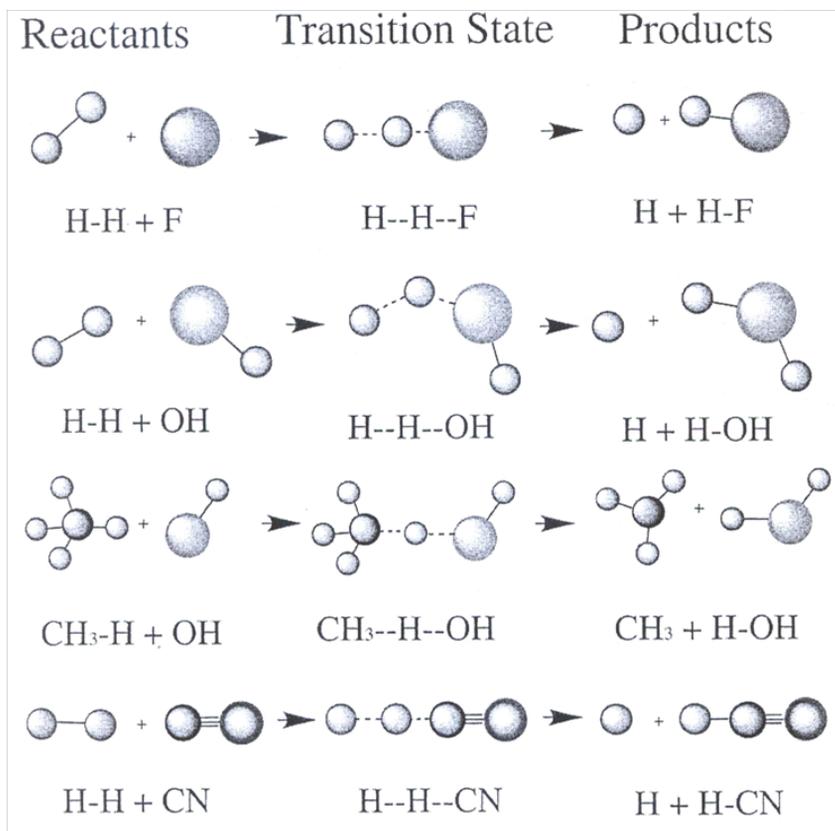


Figure 39: Four Non-Degenerate Chemical Reactions.

Table 22: Root Mean Square (RMS) Errors in scaled calculated harmonic frequencies of asymmetric transition states (cm<sup>-1</sup>).

Motion	Basis Set	Method					
		HF	MP2	MP3	B3LYP	QCISD	QCISD(T)
Scale	3-21G*	0.973	1.040	1.035	1.052	1.062	1.068
Factors	6-31G*	0.934	0.944	0.989	1.007	1.007	1.013
	6-31G**	0.933	0.983	0.973	1.002	0.989	0.996
	6-311G(2d,p)	0.935	0.989	0.98	1.003	0.996	1.003
	6-311+G(2df,p)	0.935	0.99	0.979	1.003	0.995	1.002
	6-311+G(2df,2pd)	0.934	0.985	0.974	1.003	0.989	0.997
Bend	3-21G*	241.2	602.9	242.8	256.4	233.8	231.4
	6-31G*	156.6	707.8	139.6	217.3	128.7	132.4
	6-31G**	441.7	708.2	254.1	213.7	169.9	96.9
	6-311G(2d,p)	181.5	794.1	47.3	154.2	70.9	84.3

Motion	Basis Set	Method					
	6-311+G(2df,p)	437.8	804.5	216.5	194.7	40.0	38.0
	6-311+G(2df.2pd)	429.4	551.4	317.2	217.4	33.6	0.0
Stretch	3-21G*	1176.5	968.7	893.6	1788.4	977.6	954.6
	6-31G*	1481.9	793.7	917.2	2343.6	638.2	587.5
	6-31G**	1075.8	902.9	424.9	1753.6	292.7	303.7
	6-311G(2d,p)	1113.5	776.1	416.2	1745.0	317.0	356.3
	6-311+G(2df,p)	1067.8	774.9	375.1	1768.5	141.2	61.7
	6-311+G(2df.2pd)	1070.8	838.8	576.0	1835.3	108.9	0.0
Other	3-21G*	298.7	277.0	293.3	36.0	274.0	269.2
	6-31G*	293.7	106.9	136.2	152.9	128.6	125.5
	6-31G**	256.5	48.0	71.3	152.2	4.0	53.3
	6-311G(2d,p)	247.8	100.2	42.5	153.8	55.7	126.0
	6-311+G(2df,p)	473.1	330.7	51.6	152.1	14.4	12.4
	6-311+G(2df.2pd)	221.2	28.0	50.9	152.7	15.9	0.0
Molecular	3-21G*	126.0	198.3	176.1	184.4	194.6	197.6
	6-31G*	75.8	74.5	57.6	359.1	64.6	140.2
	6-31G**	58.9	64.8	23.9	16.4	20.5	155.7
	6-311G(2d,p)	58.1	25.5	20.3	11.0	38.9	147.4
	6-311+G(2df,p)	122.3	86.2	18.5	11.7	3.7	9.4
	6-311+G(2df.2pd)	57.1	12.2	42.5	10.6	7.9	0.0
All Freq For TS	3-21G*	690.2	712.2	541.7	1018.2	582.0	569.1
	6-31G*	843.2	692.9	525.8	1321.5	372.6	346.2
	6-31G**	692.5	733.0	305.6	994.8	207.4	185.6
	6-311G(2d,p)	643.2	737.7	235.8	983.8	186.1	213.6
	6-311+G(2df,p)	701.5	751.9	265.8	1000.8	84.6	45.0
	6-311+G(2df.2pd)	684.2	625.3	400.7	1040.2	66.2	0.0

**Table 23:** Zero Point Energy Changes from Reactant to Transition State (kcal/mol). Errors (kcal/mol) in Calculated Zero Point Energy Changes from Reactant to transition state.  $\Delta E = E(\text{QCISD}(T)/6\text{-}311\text{+G}(2\text{df},2\text{pd})) - E(j)$ .

Zero Point Energy Changes from Reactant to Transition State (kcal/mol)						
HF Methods for Unsymmetric Transition States						
Transition State	HF/3-21G*	HF/6-31G*	HF6-311G**	HF/6-311G(2d,p)	HF/6-311+G(2df,p)	HF/6-311+G(2df,2pd)
HHF	-3.457	-2.932	-2.792	-2.817	-2.783	-2.831
HHOH	-0.128	-1.965	1.995	-0.171	1.889	1.908
HHCN	0.821	-0.052	-0.588	-0.683	-0.591	-0.654
H <sub>3</sub> CHOH	-1.566	-1.320	-1.549	-1.525	-0.948	-1.533
MP2 Methods for Unsymmetric Transition States						
Transition State	MP2/3-21G*	MP2/6-31G*	MP26-311G**	MP2/6-311G(2d,p)	MP2/6-311+G(2df,p)	MP2/6-311+G(2df,2pd)
HHF	-3.288	-1.623	-1.847	-1.734	-1.372	-1.082
HHOH	2.304	4.150	2.580	3.133	3.136	0.734
HHCN	1.285	1.806	1.030	0.388	0.229	0.363
H <sub>3</sub> CHOH	-1.881	-1.029	-1.629	-1.480	0.227	-1.317
MP3 Methods for Unsymmetric Transition States						
Transition State	MP3/3-21G*	MP3/6-31G*	MP3/6-311G**	MP3/6-311G(2d,p)	MP3/6-311+G(2df,p)	MP3/6-311+G(2df,2pd)
HHF	-2.979	-2.999	-1.593	-1.717	-1.620	-1.439
HHOH	0.185	0.417	3.166	0.579	2.915	2.770
HHCN	1.615	0.974	0.603	0.447	0.442	0.505
H <sub>3</sub> CHOH	-1.780	-1.774	-1.352	-1.333	-1.408	-1.172
B3LYP Methods for Unsymmetric Transition States						
Transition State	B3LYP/3-21G*	B3LYP/6-31G*	B3LYP/6-311G**	B3LYP/6-311G(2d,p)	B3LYP/6-311+G(2df,p)	B3LYP/6-311+G(2df,2pd)
HHF	-3.315					
HHOH	-0.735	-1.204	2.068	0.661	0.873	1.692
HHCN			0.410	0.355	0.048	
H <sub>3</sub> CHOH	-0.541	-1.916	-1.690	-1.522	-1.443	-1.420
QCISD Methods for Unsymmetric Transition States						
Transition State	QCISD/3-21G*	QCISD/6-31G*	QCISD/6-311G**	QCISD/6-311G(2d,p)	QCISD/6-311+G(2df,p)	QCISD/6-311+G(2df,2pd)
HHF	-3.700	-2.096	-1.463	-1.579	-0.967	-0.778
HHOH	-0.680	0.006	1.543	0.315	0.405	0.606
HHCN	0.416	0.373	0.197	0.014	0.063	0.248
H <sub>3</sub> CHOH	-2.127	-2.090	-1.640	-1.277	-1.672	-1.585
QCISDT Methods for Unsymmetric Transition States						
Transition State	QCISDT/3-21G*	QCISDT/6-31G*	QCISDT/6-311G**	QCISDT/6-311G(2d,p)	QCISDT/6-311+G(2df,p)	QCISDT/6-311+G(2df,2pd)
HHF	-3.743	-2.097	-1.530	-1.660	-0.863	-0.660
HHOH	-0.775	-0.075	0.403	0.269	0.401	0.591
HHCN	0.464	0.236	0.251	-0.468	-0.122	0.232
H <sub>3</sub> CHOH	0.613	-2.209	-1.730	-2.632	-2.739	-1.741

**Table 23a:** Errors in Calculated Zero Point Energy Changes from Reactant to Transition State (kcal/mol)  $\Delta E = E_{(\text{QCISD(T)}/6-311+G(2df,2pd))} - E_{(j)}$ 

HHF						
Basis Sets	HF	MP2	MP3	B3LYP	QCISD	QCISD(T)
3-21G*	2.797	2.628	2.319	2.656	3.04	3.083
6-31G*	2.272	0.963	2.339		1.436	1.438
6-311G**	2.132	1.187	0.933		0.803	0.871
6-311G(2d,p)	2.157	1.074	1.057		0.919	1.000
6-311+G(2df,p)	2.123	0.712	0.961		0.307	0.203
6-311+G(2df,2pd)	2.171	0.422	0.779		0.118	0.000

HHOH						
Basis Sets	HF	MP2	MP3	B3LYP	QCISD	QCISD(T)
3-21G*	0.718	-1.714	0.406	1.325	1.271	1.366
6-31G*	2.556	-3.56	0.174	1.794	0.585	0.666
6-311G**	-1.404	-1.989	-2.576	-1.477	-0.952	0.188
6-311G(2d,p)	0.762	-2.542	0.012	-0.07	0.276	0.321
6-311+G(2df,p)	-1.298	-2.545	-2.324	-0.282	0.186	0.190
6-311+G(2df,2pd)	-1.317	-0.143	-2.179	-1.102	-0.015	0.000

**Table 23b:** Errors in Calculated Zero Point Energy Changes from Reactant to Transition State (kcal/mol)  $\Delta E = E_{(\text{QCISD(T)}/6-311+G(2df,2pd))} - E_{(j)}$ 

HHCN						
Basis Sets	HF	MP2	MP3	B3LYP	QCISD	QCISD(T)
3-21G*	-0.588	-1.053	-1.383		-0.184	-0.232
6-31G*	0.284	-1.573	-0.742		-0.141	-0.004
6-311G**	0.82	-0.798	-0.371	-0.178	0.036	-0.019
6-311G(2d,p)	0.916	-0.155	-0.215	-0.122	0.218	0.701
6-311+G(2df,p)	0.823	0.004	-0.209	0.185	0.169	0.354
6-311+G(2df,2pd)	0.886	-0.131	-0.272		-0.016	0.000

H <sub>3</sub> CHOH						
Basis Sets	HF	MP2	MP3	B3LYP	QCISD	QCISD(T)
3-21G*	-0.175	0.14	0.039	-1.199	0.386	-2.353
6-31G*	-0.421	-0.711	0.033	0.175	0.349	0.468
6-311G**	-0.192	-0.112	-0.389	-0.051	-0.101	-0.011
6-311G(2d,p)	-0.216	-0.26	-0.407	-0.219	-0.464	0.891
6-311+G(2df,p)	-0.793	-1.967	-0.333	-0.297	-0.069	0.999
6-311+G(2df,2pd)	-0.208	-0.423	-0.569	-0.321	-0.156	0.000

**Table 23c:** Summary Errors in Calculated Zero Point Energy Changes from Reactant to Transition State (kcal/mol)  $\sigma E = E_{(\text{QCISD(T)}/6-311+G(2df,2pd))} - E_{(j)}$ 

	HF/ 3-21G*	HF/ 6-31G*	MP2/ 6-311G**	B3LYP/ 6-311G**	QCISD/ 6-311G(2d,p)	QCISD(T)/ 6-311+G(2df,2pd)
HHF	2.797	2.272	1.187	No TS	0.919	0.000
HHOH	0.718	2.556	-1.989	-2.576	-0.070	0.000
HHCN	-0.588	0.284	-0.798	-0.178	0.218	0.000
H <sub>3</sub> CHOH	-0.175	-0.421	-0.112	-0.051	-0.464	0.000

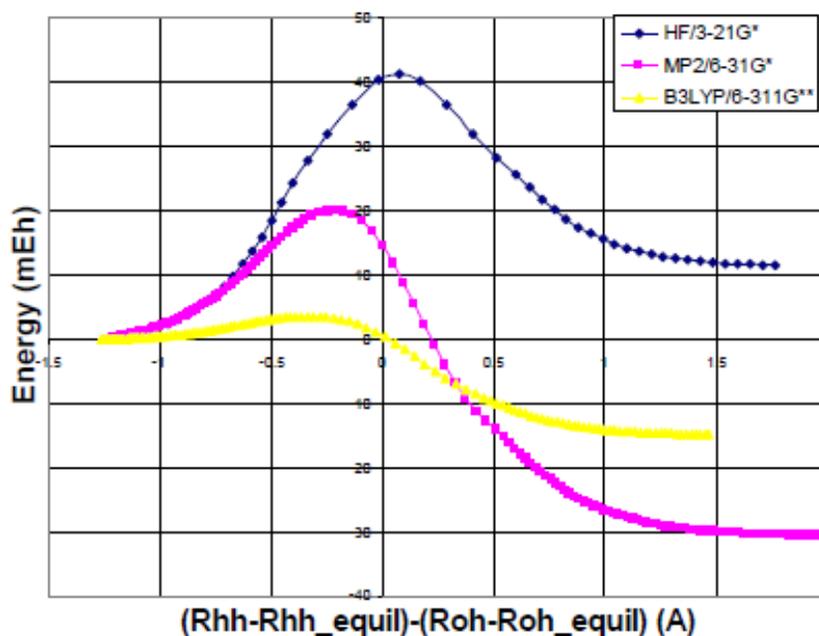
Energy Profiles for  $\text{H}_2 + \text{OH} \rightarrow \text{H-H-O-H} \rightarrow \text{H} + \text{H}_2\text{O}$ 

Figure 40: Energy Profile for  $\text{H}_2 + \text{OH} \rightarrow \text{H} + \text{H}_2\text{O}$ .

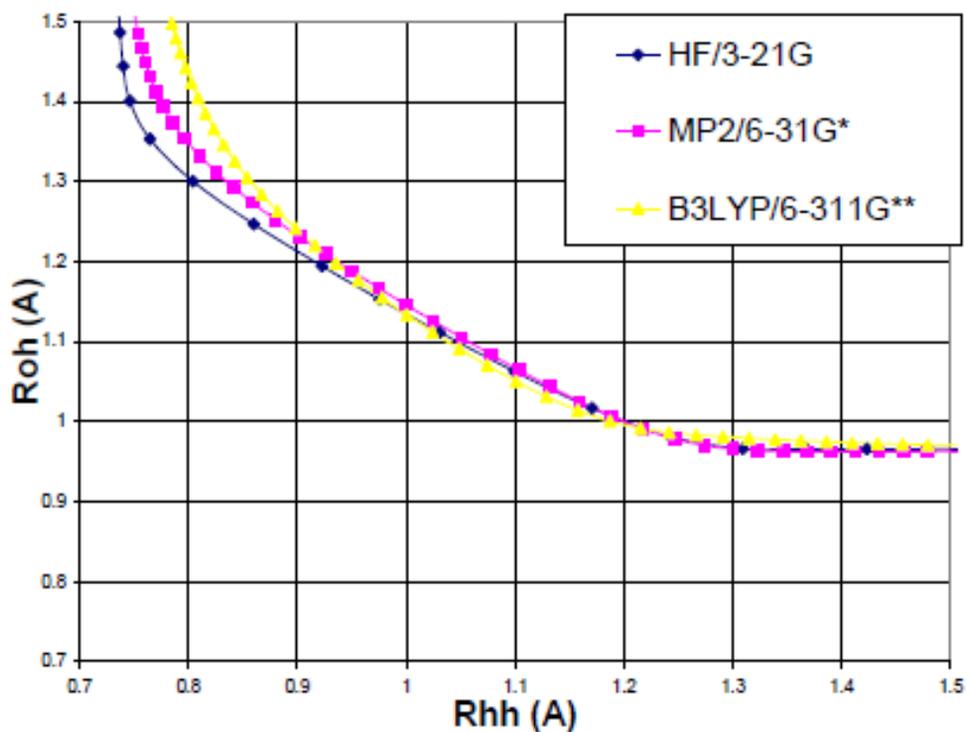


Figure 41: Reaction Path for  $\text{H}_2 + \text{OH} \rightarrow \text{H-H-O-H} \rightarrow \text{H} + \text{H}_2\text{O}$  defined in molecular coordinates. Roh = O-H internuclear distance (Angstroms), Rhh = H-H internuclear distance (Angstroms).

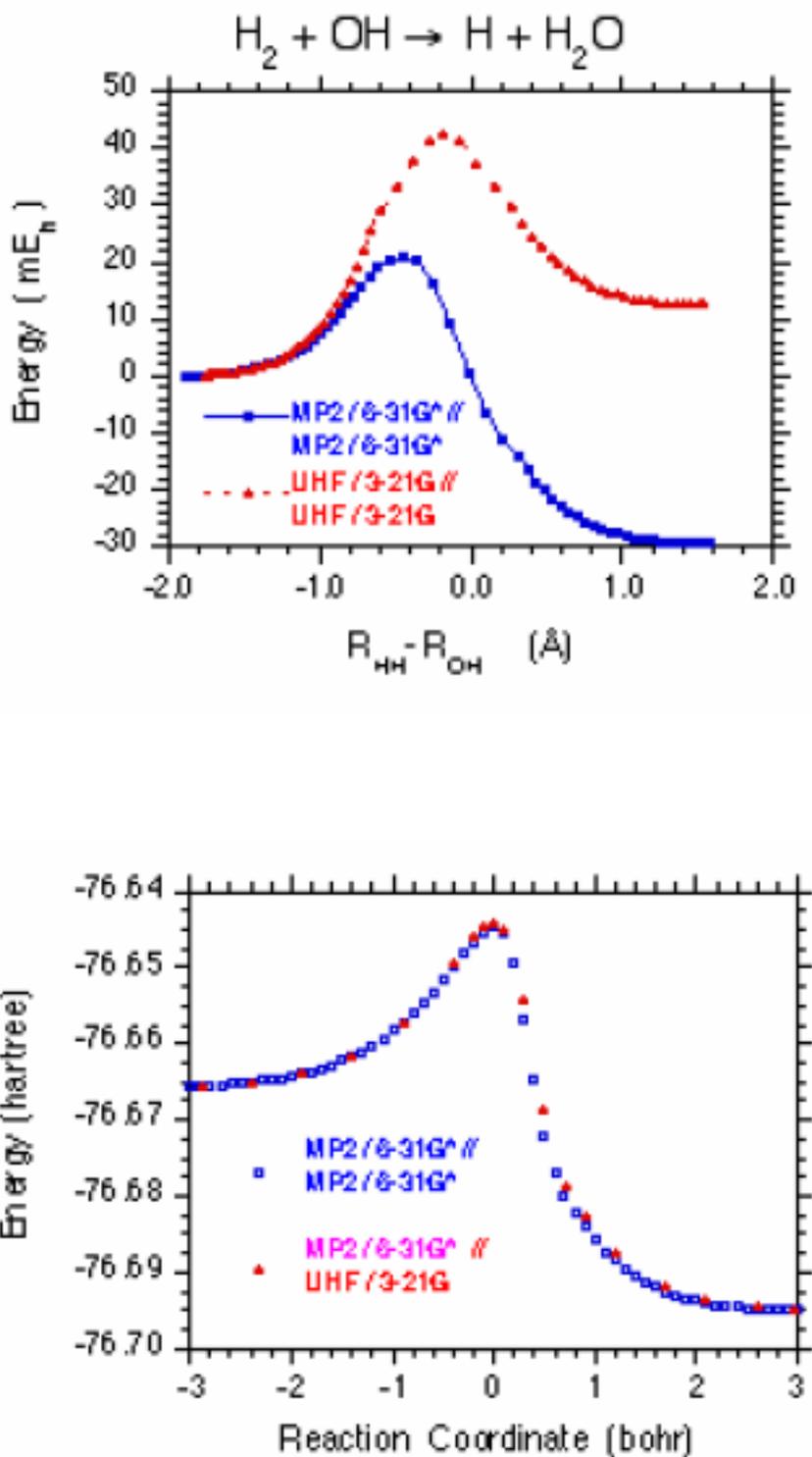


Figure 42: Depicts the IRCMax method for the reaction of  $\text{H}_2 + \text{OH} \rightarrow \text{H}-\text{H}-\text{O}-\text{H} \rightarrow \text{H} + \text{H}_2\text{O}$ .

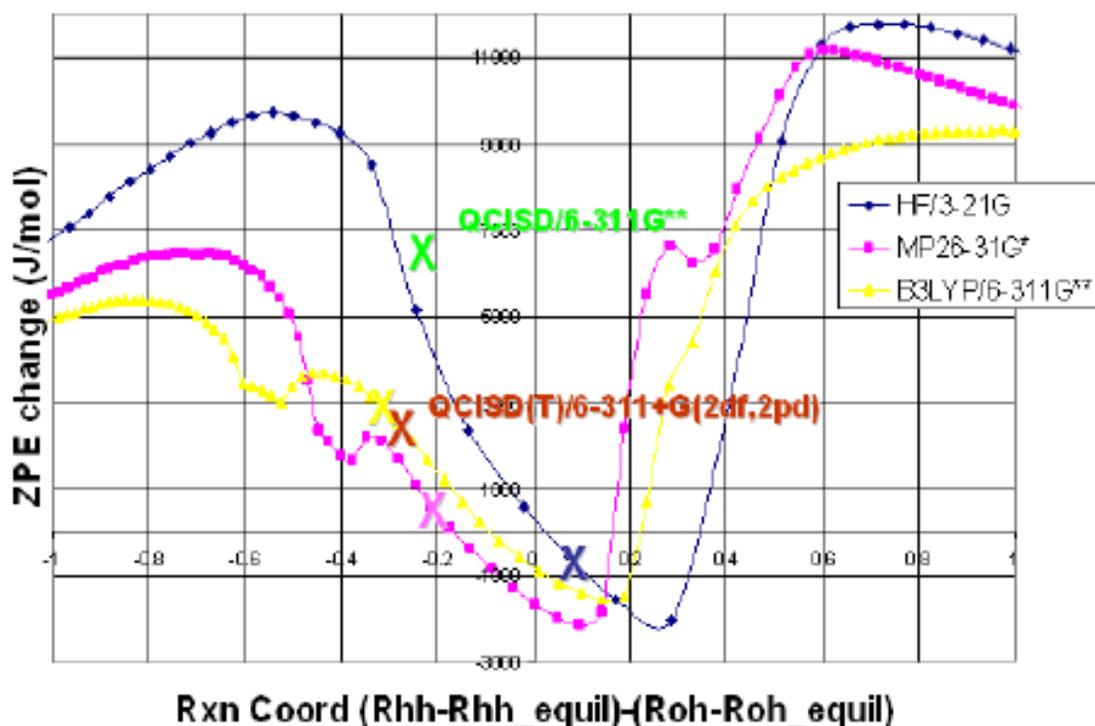


Figure 43: Zero Point Energy Change as a function of Reaction Coordinate for  $\text{H}_2 + \text{OH} \rightarrow \text{H} + \text{H}_2\text{O}$ .

### Definitive tests for the intrinsic reaction coordinate maximum (IRCM<sub>Max</sub>) method for predicting frequencies of transition states

Four definitive tests have been constructed to determine whether the IRC<sub>Maximum</sub> method improves normal mode frequencies of transition states and zero point energy contributions to the activation barrier compared to computationally practical methodologies (eg, Hartree-Fock, MP2, CBS-4M [129,130], B3LYP, CBS-QB3) [131]. The following IRC<sub>Maximum</sub> high-level energies along lower-level paths were located, denoted Max{high level//IRC{low level}: Max{MP2/6-311+G(2df,2pd)//IRC{HF/3-21G\*}, Max{CBS-4M//IRC{HF/3-21G\*}, Max{QCISD(T)/6-311+G(2df,2pd)//IRC{B3LYP/6-311G\*\*}, Max{CBS-QB3//IRC{B3LYP/6-311G\*\*}. The corresponding low-level frequencies were calculated on each of the IRC<sub>Maxima</sub>. These IRC<sub>Max</sub> frequencies and the lower level frequencies were compared to the high level frequencies in order to assess improvement.

**Max{MP2/6-311+G(2df,2pd)//IRC{HF/3-21G\*}**: Table 24 presents the IRC<sub>Maximum</sub> frequencies and zero point energy contributions to the activation energies of the four asymmetric hydrogen transfer reactions, Max{MP2/6-311+G(2df,2pd)//IRC{HF/3-21G\*} (see Table 19 for Zero Point Energies of reactants).

The HF/3-21G\* stretching frequency for transition state

H-H-F is erroneously small due to qualitatively incorrect geometry. With this basis set, it is predicted to be linear, whereas higher levels of correlation treatment all predict a bent geometry. The Hartree-Fock stretching frequency calculated at the IRC<sub>Maximum</sub> geometry is much improved since the bent geometry is recovered. The small HHH bend is overestimated with HF/3-21G\*, and goes slightly negative ( $-20.2 \text{ cm}^{-1}$ ). Future treatment of small negative modes will be reported as  $0 \text{ cm}^{-1}$  for calculation of zero point energy changes from reactant to transition state. The total RMS error in the frequencies is reduced by more than a factor of two when the IRC<sub>Max</sub> technique is applied. The HF/3-21G\* also overestimates the energy lowering effect of the ZPE contribution to the activation energy, which the IRC<sub>Maximum</sub> improves (2.375 compared with  $-0.772 \text{ kcal/mol}$  deviation). For the transition state H-H-O-H the HF/3-21G\* predicts a zero point energy change from reactant to transition state that is energy lowering by  $-0.128 \text{ kcal/mol}$ ; however large basis set MP2/6-311+G(2df,2pd) predicts that the energy will be raised by 0.734. The IRC<sub>Maximum</sub> corrects the Hartree-Fock behavior, giving a zero point energy contribution of 0.508. The total RMS error in the frequencies decreases by a factor of four when IRC<sub>Max</sub> method is applied.

The HF/3-21G\* underestimates the stretching frequencies of transition state H-H-C-N by  $1000 \text{ cm}^{-1}$  and overestimates the bending frequencies by  $670 \text{ cm}^{-1}$  for the HHC bending

modes, and  $200\text{cm}^{-1}$  for the NCH bending modes. The normal modes calculated at the IRCMaximum geometry are much improved. The total RMS error in the frequencies calculated at the IRCMax geometry is reduced by a factor of 1.5 compared with the Hartree-Fock frequencies. Although the NC stretching frequency is still underestimated by  $1000\text{cm}^{-1}$ , the HC stretching frequency is overestimated by  $482\text{cm}^{-1}$ , half the error observed for Hartree-Fock. The HHC bending frequency is overestimated by  $125\text{cm}^{-1}$ , which is also less than that of the HF. Because the HF/3-21G\* underestimates the stretches and overestimates the bending frequencies, there is a fortuitous cancellation of errors that leads to a slightly improved estimation of the ZPE change ( $0.821\text{ kcal/mol}$ ) from reactant to transition states than the IRCMaximum ZPE change ( $1.068\text{ kcal/mol}$ ) compared to MP2/6-311+G(2df,2pd) ( $0.363\text{ kcal/mol}$ ). The IRCMaximum method does overestimate the ZPE contribution to the activation energy, but it does improve the normal mode frequencies, and therefore does not rely on the fortuitous cancellation of positive and negative errors. For the transition state H<sub>3</sub>C-H-OH, the IRCMaximum ZPE change from reactant to transition state shows only modest improvement over the HF/3-21G\* ( $0.248$  and  $0.148\text{ kcal/mol}$  deviation, respectively) compared to the MP2/6-311+G(2df,2pd). The normal mode frequencies calculated at the IRC Maximum geometry give roughly the same RMS error as the Hartree-Fock frequencies.

**Max{CBS-4M//IRC{HF/3-21G\*}**: Table 25 presents the IRCMaximum frequencies and zero point energy contributions to the activation energies of the four asymmetric hydrogen transfer reactions, Max{CBS-4M//IRC{HF/3-21G\*} (see Table 19. for reactant Zero Point Energies).

The total RMS errors in frequencies are reduced by a factor of three when the IRCMax method is applied for the HHF transition state. The stretching frequency is underestimated by  $2225\text{cm}^{-1}$  while the bend is overestimated by  $422\text{cm}^{-1}$ . These errors are halved with the IRCMax method. The stretching frequency is overestimated by  $720\text{cm}^{-1}$  and the bend is underestimated by  $215\text{ cm}^{-1}$ . This improvement is reflected in the ZPE change. The Hartree-Fock method overestimates the energy lowering effect ( $-3.457\text{ kcal/mol}$ ) compared to the QCISD (T)/6-311+G (2df,2pd) ( $-0.660\text{ kcal/mol}$ ), but the frequencies at the IRCMaximum geometry are in better agreement ( $-0.154\text{ kcal/mol}$ ). For the transition state HHOH, the HF/3-21G\* has a zero point energy contribution that is energy lowering by  $-0.128\text{ kcal/mol}$ , whereas large basis set QCISD(T)/6-311+G(2df,2pd) predicts an energy raising of  $0.591\text{kcal/mol}$ .

The IRCMaximum method corrects the bending and stretching frequencies such that the zero point energy change from reactant to transition state is also energy raising,  $0.738\text{ kcal/mol}$ . The RMS errors in the frequencies remain roughly the same when the IRCMax method is applied. The HHCN transition state suffers from spin contamination, and in this case, the error from the mixture of doublet and quartet states

from the Hartree-Fock method cannot be removed. The error in the zero point energy change for the IRCMaximum geometry is 1.5 times as large compared with QCISD(T)/6-311+G(2df,2pd) as the HF/3-21G\*. Due to the unphysical spin contamination present in the Hartree-Fock method, it is a poor choice for the lower level. The total RMS errors in the frequencies remain roughly the same when the IRCMax technique is implemented. In Section 4.5.3, we will show that when the B3LYP/6-311G\*\* is chosen as the lower level, the IRC Max technique does not exacerbate irremovable errors. For the transition state H<sub>3</sub>C-H-OH, the ZPE change calculated at the IRCMaximum predicts an energy raising effect, but both the benchmark QCISD (T) and the HF give an energy lowering effect of equal magnitude. In this case the IRCMaximum technique is overestimating many of the large normal mode frequencies of this transition state compared to the benchmark. The total RMS error in the calculated frequencies is increased by a factor of 4.5 when the IRC Max method is introduced in this case.

**Max{QCISD(T)/6-311+G(2df,2pd)//IRC(B3LYP/6-311G\*\*)}**: Table 26 shows the IRCMaximum frequencies and zero point energy contributions to the activation energies of the four asymmetric hydrogen transfer reactions, Max{QCISD(T)/6-311+G(2df,2pd)//IRC{B3LYP/6-311G\*\*} (see Table 19 for Reactant Zero Point Energies). The B3LYP hybrid density functional is known to underestimate barrier heights, and in the reaction with the 6-311G\*\* basis set fails to predict a transition state in the reaction of  $\text{H}_2 + \text{F} \rightarrow \text{HF} + \text{H}$ . However, the frequencies calculated at the IRCMaximum are in excellent agreement with benchmark frequencies (6.1% deviation for the stretch and 5.1% deviation for the bend). The zero point energy “change” from reactant to transition states for the B3LYP/6-311G\*\* is merely the total zero point energy of reactants since there is no predicted transition state, and this drastically overestimates the energy lowering effect. The IRC Max technique corrects this error, and gives a zero point energy change that is in qualitative agreement with the benchmark ( $-1.080$  compared to  $-0.660\text{kcal/mol}$ , respectively).

Figure 43 shows that when we look at how the ZPE changes as a function of the reaction coordinate, the QCISD(T) and B3LYP transition states are closer than the MP2 and HF transition states for the reaction of  $\text{H}_2 + \text{OH} \rightarrow \text{H}_2\text{O} + \text{H}$ . For the MP2 Maximum along the Hartree-Fock path (Section 4.5.1), we observed that the HF/3-21G\* method emulated the MP2 normal modes for the H-H-O-H transition state better than the frequencies at the IRCMaximum geometry. In this case, where the transition states are closer together along the reaction profile, the frequencies at the IRCMaximum geometry are in better agreement with the benchmark than B3LYP frequencies. The total RMS error in the frequencies is reduced by a factor of four when the IRCMax method is employed. The zero point energies are improved as well when the IRCMax technique is employed. B3LYP/6-311G\*\* overestimates the energy raising effect of the zero point energy change ( $2.068$

kcal/mol), but when the IRCMax technique corrects this error and gives a zero point energy change that is in qualitative agreement with the benchmark (0.274 compared to 0.591 kcal/mol, respectively). For the HHCN transition state, we see from Table 26 that the B3LYP/6-311G\*\* serves as a better choice for the lower level reaction path than HF/3-21G\* for the lower level due to spin contamination. The frequencies and zero point energy contribution calculated at the IRCMaximum geometry are qualitatively similar to the B3LYP frequencies. The B3LYP zero point energy change from reactant to transition state is moderately better than that of the IRCMaximum ZPE change compared to the benchmark.

Table 26 shows that for the H<sub>3</sub>C-H-OH transition state, the IRCMaximum method offers a substantial improvement for frequencies and the zero point energy contribution to the activation energy. The total RMS error in the frequencies does roughly double with use of the IRCMax method. However, with the exception of the  $\nu_1$  (2a') stretching and  $\nu_2$  (2a'') bending frequencies, all errors in the normal modes were below 20cm<sup>-1</sup>. This improvement in the frequencies provided by the IRCMax technique is also observed in the zero point energy change from reactant to transition state. The ZPE change calculated at the IRCMax geometry (-1.714kcal/mol) has reduced the error from the B3LYP ZPE change (-1.690kcal/mol) compared to the QCISD (T) benchmark (-1.740kcal/mol) by a factor of two.

**Max{CBS-QB3//IRC(B3LYP/6-311G\*\*)}:** Table 27 shows the IRCMaximum frequencies and zero point energy contributions to the activation energies of the four asymmetric hydrogen transfer reactions, Max{CBS-QB3//IRC(B3LYP/6-311G\*\*)} (see Table 19 for reactant Zero Point Energies). Since the B3LYP/6-311G\*\* method fails to predict a transition state in the reaction of H<sub>2</sub> + F → HF + H, the IRCMaximum technique again offers a substantial improvement for calculation of frequencies of the H-H-F transition state and zero point energy for the contributions to the activation energy. The RMS error in the frequencies is 155.2cm<sup>-1</sup>. Again, the IRCMax technique corrects the large overestimate of the energy lowering effect of the zero point energy change from reactant to product seen in the B3LYP since the B3LYP is merely the total zero point energy of reactants. The zero point energy change calculated at the IRCMaximum geometry gives a zero point energy change that is -0.843kcal/mol compared to -0.660kcal/mol for the QCISD (T) benchmark.

The data presented in Table 27 shows that the IRCMax

technique also improves the normal mode frequencies and zero point energy changes from reactant to transition state for the H-H-O-H transition state. The RMS error for the frequencies calculated at the IRCMaximum geometry is three times less than that the B3LYP/6-311G\*\* frequencies (227.4 cm<sup>-1</sup> compared to 725.4 cm<sup>-1</sup>) compared to the QCISD (T) benchmark. The zero point energy change predicted with the B3LYP method overestimates the energy raising effect toward the activation energy by an order of magnitude, but the IRCMax technique corrects this, and the zero point energy change error compared to the QCISD(T) benchmark is reduced to a factor of two (0.278kcal/mol versus 0.591 kcal/mol, respectively). For the asymmetric transition state HHCN, Table 27 shows that the IRCMax method offers an improvement for calculation of frequencies. The RMS error has been reduced compared to the B3LYP/6-311G\*\* by a factor of six (226.7 versus 1312.8 cm<sup>-1</sup>). The B3LYP overestimates the stretches while underestimating the bending modes, leading to a fortuitous cancellation of errors, and a zero point energy change that is in better agreement with the QCISD (T) benchmark. The normal modes calculated at the IRCMaximum geometry are all in better agreement with the benchmark than the B3LYP frequencies, but there is no cancellation of errors, and the IRCMax method overestimates the energy raising effect of the zero point energy change by a factor of 2.5 from reactant to transition state for this case (0.63kcal/mol versus 0.23 kcal/mol). For the case of H<sub>3</sub>C-H-OH, the Table 27 indicates that the IRCMax method is not leading to an improvement in the frequencies of this transition state or the zero point energy change to reactants compared to the B3LYP/6-311G\*\*. The RMS error in the frequencies calculated with the IRC Maximum geometry are an order of magnitude larger (684cm<sup>-1</sup> for IRC Max versus 70cm<sup>-1</sup> for B3LYP). This is due to large errors in bending frequencies: The  $\nu_2$  (2a'') CHO bend is underestimated by 220.5cm<sup>-1</sup>, the  $\nu_5$  (3a') HCH bend is overestimated by 332.3cm<sup>-1</sup>, and the  $\nu_{10}$  (1a') HOH bend is overestimated by 525.8cm<sup>-1</sup>. These errors in the frequencies lead to an estimation of the zero point energy that underestimates the energy lowering effect of the zero point energy change. The zero point energy change calculated at the IRC Maximum geometry obtains only 30% of the QCISD (T) ZPE change (-0.510 versus -1.741 kcal/mol, respectively). In this case, the geometry is closer to the CBS-APNO geometry, which may indicate that our benchmark, the QCISD (T)/6-311+G(2df,2pd) is not converged.

**Table 24:** IRCMaximum frequencies ( $\text{cm}^{-1}$ ) and zero point energy contributions (DZPE) (kcal/mol) to the activation energies of the four asymmetric hydrogen transfer reactions,  $\text{Max}\{\text{MP2/6-311+G(2df,2pd)}/\text{IRC}\{\text{HF/3-21G}^*\}$ .

Transition State	Mode	Motion	MP2/ 6-311+G(2df,2pd)	HF/3-21G*	Deviation	HF/3-21G* freq @ MP2/6- 311+G(2df,2pd) max along HF/3-21G* path	Deviation
HHF	$v_1(2a')$	HHF stretch	3454.6	1478.1		4315.1	
	$v_2(1a')$	HHF "rotation"	176.7	636.8		0.0	
RMS error frequencies					2029.3		878.5
$\Delta\text{ZPE}$ (kcal/mol)			-1.082	-3.457	2.375	-0.309	-0.772
HHOH	$v_1(2a')$	HHO stretch	2698.5	1316.0		3574.8	
	$v_2(1a')$	HHO bend	590.8	658.0		480.3	
	$v_3(1a'')$	HHO bend	507.8	719.1		0.0	
	$v_4(3a')$	HOH bend	1058.7	1622.1		852.8	
	$v_5(4a')$	OH stretch	3771.6	3638.4		3489.4	
RMS error frequencies					1515.1		1077.0
$\Delta\text{ZPE}$ (kcal/mol)			0.734	-0.128	0.861	0.508	0.226
HHCN	$v_1(2\sigma)$	HHC stretch	3779.5	2507.5		4262.4	
	$v_2(1\pi)$	HHC bend	349.6	1012.3		474.8	
	$v_3(2\pi)$	HHC bend	349.6	1012.3		474.8	
	$v_4(3\pi)$	:NCH bend	48.4	240.9		0.0	
	$v_5(4\pi)$	:NCH bend	48.4	240.9		0.0	
	$v_6(1\sigma)$	NC stretch	2869.5	1856.9		1831.9	
RMS error frequencies					1896.3		1160.1

Transition State	Mode	Motion	MP2/ 6-311+G(2df,2pd)	HF/3-21G*	Deviation	HF/3-21G* freq @ MP2/6- 311+G(2df,2pd) max along HF/3-21G* path	Deviation
$\Delta ZPE$ (kcal/mol)			0.363	0.821	-0.458	1.068	-0.705
H <sub>3</sub> CHOH	$\nu_1(2a')$	CHO stretch	783.0	531.5		811.2	
	$\nu_2(2a'')$	CHO bend	320.2	401.7		191.0	
	$\nu_3(1a')$	CHO bend	328.7	392.8		282.2	
	$\nu_4(1a'')$	HCHOH torsion	30.2	41.8		0.0	
	$\nu_5(3a')$	HCH bend	919.7	927.7		851.0	
	$\nu_6(3a'')$	HCH bend	1203.8	1179.6		1434.6	
	$\nu_7(3a')$	HCH umbrella	1284.0	1285.7		1451.5	
	$\nu_8(5a')$	HCH bend	1369.4	1541.8		1517.7	
	$\nu_9(1a'')$	HCH bend	1464.7	1553.8		1607.4	
	$\nu_{10}(1a')$	HOH bend	1483.4	1647.5		1664.0	
	$\nu_{11}(1a')$	(C-H)3 stretch	3074.4	3143.8		3127.5	
	$\nu_{12}(2a'')$	(C-H)2 stretch	3206.9	3358.8		3235.5	
	$\nu_{13}(5a'')$	(C-H)1 stretch	3210.2	3268.9		3237.2	
	$\nu_{14}(9a')$	OH stretch	3781.0	3640.2		3592.6	
RMS error frequencies					436.5		470.6
$\Delta ZPE$ (kcal/mol)			-1.317	-1.566	0.248	-1.466	0.148
Scale Factors			0.997	0.973		0.973	

**Table 25:** IRC Maximum frequencies ( $\text{cm}^{-1}$ ) and zero point energy contributions (DZPE) (kcal/mol) to the activation energies of the four asymmetric hydrogen transfer reactions,  $\text{Max}\{\text{CBS-4M}/\text{IRC}\{\text{HF/3-21G}^*\}$ .

Transition State	Mode	Motion	QCISDT/ 6-311+G(2df,2pd)	HF/3-21G*	% Deviation	HF/3-21G* @ CBS-4M max along HF/3- 21G*path	% Deviation
HHF	$\nu_1(2a')$	HHF stretch	3702.7	1478.1		4423.5	
	$\nu_2(1a')$	HHF "rotation"	215.0	636.8		0	
RMS error frequencies					2264.2		752.2
$\Delta\text{ZPE}$ (kcal/mol)			-0.660	-3.457	2.797	-0.154	-0.506
HHOH	$\nu_1(2a')$	HHO stretch	2582.3	1316		4040.0	
	$\nu_2(1a')$	HHO bend	592.7	658.0		207.3	
	$\nu_3(1a'')$	HHO bend	529.1	719.1		0.0	
	$\nu_4(3a')$	HOH bend	1068.3	1622.1		754.3	
	$\nu_5(4a')$	OH stretch	3760.4	3638.4	3.24	3556.7	5.42
RMS error frequencies					1402.0		1641.2
$\Delta\text{ZPE}$ (kcal/mol)			0.591	-0.128	0.719	0.738	
HHCN	$\nu_1(2\sigma)$	HHC stretch	3027.9	2507.5		3847.7	
	$\nu_2(1\pi)$	HHC bend	616.8	1012.3		674.4	
	$\nu_3(2\pi)$	HHC bend	616.8	1012.3		674.4	
	$\nu_4(3\pi)$	:NCH bend	125.4	240.9		49.1	
	$\nu_5(4\pi)$	:NCH bend	125.4	240.9		49.1	
	$\nu_6(1\sigma)$	NC stretch	2119.1	1856.9		1849.0	

Transition State	Mode	Motion	QCISDT/ 6-311+G(2df,2pd)	HF/3-21G*	% Deviation	HF/3-21G* @ CBS-4M max along HF/3- 21G*path	% Deviation
RMS error frequencies					824.0		873.7
$\Delta ZPE$ (kcal/mol)			0.232	0.821	-0.589	1.211	-0.979
H <sub>3</sub> CHOH	$\nu_1(2a')$	CHO stretch	701.0	531.5		628.5	
	$\nu_2(2a'')$	CHO bend	340.8	401.7		448.9	
	$\nu_3(1a')$	CHO bend	342.0	392.8		84.4	
	$\nu_4(1a'')$	HCHOH torsion	20.2	41.8		0.0	
	$\nu_5(3a')$	HCH-bend	906.7	927.7		1436.6	
	$\nu_6(3a'')$	HCH-bend	1194.7	1179.6		1438.3	
	$\nu_7(3a')$	HCH umbrella	1269.1	1285.7		1488.4	
	$\nu_8(5a')$	HCH bend	1368.1	1541.8		1657.7	
	$\nu_9(1a'')$	HCH bend	1454.6	1553.8		1704.5	
	$\nu_{10}(1a')$	HOH bend	1476.0	1647.5		1166.2	
	$\nu_{11}(1a')$	(C-H)3 stretch	3062.7	3143.8		3197.4	
	$\nu_{12}(2a'')$	(C-H)2 stretch	3189.5	3358.8		3201.0	
	$\nu_{13}(5a'')$	(C-H)1 stretch	3192.6	3268.9		3383.4	
	$\nu_{14}(9a')$	OH stretch	3771.9	3640.2		3551.0	
RMS error frequencies					405.3		904.2
ZP (kcal/mol)			-1.741	-1.722	-0.018	-0.918	-0.822
Scale Factors			0.997	0.973		0.973	

**Table 26:** IRCMaximum frequencies ( $\text{cm}^{-1}$ ) and zero point energy ( $\Delta\text{ZPE}$ ) (kcal/mol) contributions to the activation energies of the four asymmetric hydrogen transfer reactions,  $\text{Max}\{\text{QCISD(T)/6-311+G(2df,2pd)}\}/\text{IRC}\{\text{B3LYP/6-311G}^{**}\}$ .

Transition State	Mode	Motion	QCISDT/ 6-311+G(2df,2pd)	B3LYP/6-311G**	Deviation	B3LYP/6-311G** @ QCISD(T)/6 311+G(2df,2pd) max along B3LYP/6-311G** path	Deviation
HHF	$\nu_1(2a')$	HHF stretch	3702.7			3478.3	
	$\nu_2(1a')$	HHF "rotation"	215.0			203.9	
RMS error frequencies					3708.9		224.6
$\Delta\text{ZPE}$ (kcal/mol)			-0.660	-6.348	5.688	-1.080	0.420
HHOH	$\nu_1(2a')$	HHO stretch	2582.3	3123.3		2338.6	
	$\nu_2(1a')$	HHO bend	592.7	1003.7		689.1	
	$\nu_3(1a'')$	HHO bend	529.1	780.1111		524.3	
	$\nu_4(3a')$	HOH bend	1068.3	1059.415		1134.9	
	$\nu_5(4a')$	OH stretch	3760.4	3720.426		3746.4	
RMS error frequencies					725.4		270.8
$\Delta\text{ZPE}$ (kcal/mol)			0.591	2.068	-1.477	0.274	0.317
HHCN	$\nu_1(2s)$	HHC stretch	3027.9	4198.58		2775.3	
	$\nu_2(1p)$	HHC bend	616.8	204.5182		757.7	
	$\nu_3(2p)$	HHC bend	616.8	204.5182		757.7	
	$\nu_4(3p)$	:NCH bend	125.4	54.26832		213.7	
	$\nu_5(4p)$	:NCH bend	125.4	54.26832		213.7	
	$\nu_6(1s)$	NC stretch	2119.1	2170.953		2187.8	

Transition State	Mode	Motion	QCISDT/ 6-311+G(2df,2pd)	B3LYP/6-311G**	Deviation	B3LYP/6-311G** @ QCISD(T)/6 311+G(2df,2pd) max along B3LYP/6-311G** path	Deviation
RMS error frequencies					1312.8		351.8
$\Delta$ ZPE (kcal/mol)			0.232	0.410	-0.178	0.437	-0.205
H <sub>3</sub> CHOH	$\nu_1(2a')$	CHO stretch	701.0	704.2056		797.9	
	$\nu_2(2a'')$	CHO bend	340.8	317.133		278.5	
	$\nu_3(1a')$	CHO bend	342.0	375.9504		355.4	
	$\nu_4(1a'')$	HCHOH torsion	20.2	32.2644		0.0	
	$\nu_5(3a')$	HCH-bend	906.7	880.1568		897.4	
	$\nu_6(3a'')$	HCH-bend	1194.7	1155.106		1174.7	
	$\nu_7(3a')$	HCH umbrella	1269.1	1254.905		1280.0	
	$\nu_8(5a')$	HCH bend	1368.1	1381.457		1364.7	
	$\nu_9(1a'')$	HCH bend	1454.6	1440.475		1441.3	
	$\nu_{10}(1a')$	HOH bend	1476.0	1467.128		1468.4	
	$\nu_{11}(1a')$	(C-H)3 stretch	3062.7	3063.815		3056.6	
	$\nu_{12}(2a'')$	(C-H)2 stretch	3189.5	3183.955		3173.5	
	$\nu_{13}(5a'')$	(C-H)1 stretch	3192.6	3189.667		3179.5	
	$\nu_{14}(9a')$	OH stretch	3771.9	3761.207		3749.6	
RMS error frequencies					70.4		125.2
$\Delta$ ZPE (kcal/mol)			-1.741	-1.690	-0.051	-1.714	-0.027
Scale Factors			0.997	1.002		1.002	

**Table 27:** IRCMaximum frequencies ( $\text{cm}^{-1}$ ) and zero point energy ( $\Delta\text{ZPE}$ ) (kcal/mol) contributions to the activation energies of the four asymmetric hydrogen transfer reactions,  $\text{Max}\{\text{CBS-QB3//IRC}(\text{B3LYP/6-311G}^{**})\}$ .

Transition State	Mode	Motion	QCISD(T)/6-311+G(2df,2pd)	B3LYP/6-311G**	Deviation	B3LYP/6-311G** freq @ CBS-QB3 max along B3LYP/6-311G** path	Deviation
HHF	$\nu_1(2a')$	HHF stretch	3702.7			3566.3	
	$\nu_2(1a')$	HHF "rotation"	215.0			140.8	
RMS error frequencies					3708.9		155.2
ZP (kcal/mol)			-0.660	-6.348	-5.688	-0.843	-0.183
HHOH	$\nu_1(2a')$	HHO stretch	2582.3	3123.3		2719.8	
	$\nu_2(1a')$	HHO bend	592.7	1003.7		623.4	
	$\nu_3(1a'')$	HHO bend	529.1	780.1		375.9	
	$\nu_4(3a')$	HOH bend	1068.3	1059.4		979.8	
	$\nu_5(4a')$	OH stretch	3760.4	3720.4		3737.1	
RMS error frequencies					725.4		227.3
ZP (kcal/mol)			0.591	2.068	1.477	0.278	-0.313
HHCN	$\nu_1(2\sigma)$	HHC stretch	3027.9	4198.6		3231.1	
	$\nu_2(1\pi)$	HHC bend	616.8	204.5		639.9	
	$\nu_3(2\pi)$	HHC bend	616.8	204.5		639.9	
	$\nu_4(3\pi)$	:NCH bend	125.4	54.3		172.7	
	$\nu_5(4\pi)$	:NCH bend	125.4	54.3		172.7	
	$\nu_6(1\sigma)$	NC stretch	2119.1	2171.0		2186.9	
RMS error frequencies					1312.8		226.7

Transition State	Mode	Motion	QCISD(T)/6-311+G(2df,2pd)	B3LYP/6-311G**	Deviation	B3LYP/6-311G** freq @ CBS-QB3 max along B3LYP/6-311G** path	Deviation
ZP (kcal/mol)			0.232	0.4	0.178	0.633	0.401
H <sub>3</sub> CHOH	v <sub>1</sub> (2a')	CHO stretch	701.0	704.2		770.5	
	v <sub>2</sub> (2a'')	CHO bend	340.8	317.1		120.4	
	v <sub>3</sub> (1a')	CHO bend	342.0	376.0		252.9	
	v <sub>4</sub> (1a'')	HCHOH torsion	20.2	32.3		0.0	
	v <sub>5</sub> (3a')	HCH-bend	906.7	880.2		1239.0	
	v <sub>6</sub> (3a'')	HCH-bend	1194.7	1155.1		1245.0	
	v <sub>7</sub> (3a')	HCH umbrella	1269.1	1254.9		1335.3	
	v <sub>8</sub> (5a')	HCH bend	1368.1	1381.5		1462.8	
	v <sub>9</sub> (1a'')	HCH bend	1454.6	1440.5		1493.8	
	v <sub>10</sub> (1a')	HOH bend	1476.0	1467.1		2001.8	
	v <sub>11</sub> (1a')	(C-H) <sub>3</sub> stretch	3062.7	3063.8		3060.8	
	v <sub>12</sub> (2a'')	(C-H) <sub>2</sub> stretch	3189.5	3184.0		3163.7	
	v <sub>13</sub> (5a'')	(C-H) <sub>1</sub> stretch	3192.6	3189.7		3170.9	
	v <sub>14</sub> (9a')	OH stretch	3771.9	3761.2		3742.6	
RMS error frequencies					70.4		684.1
ZP (kcal/mol)			-1.741	-1.690	0.051	-0.510	1.231
Scale Factors			0.997	1.002		1.002	

### Limitations of the IRCMaximum method for obtaining accurate frequencies of transition states and zero point energy contributions to the activation energy.

The IRCMaximum method has been shown to remove substantial errors compared to traditional techniques consistently only when the best B3LYP/6-311G\*\* IRC is used. In this case the IRCMax method provides a drastic improvement especially for the reaction of  $H_2 + F \rightarrow HF + H$ . Here, the B3LYP hybrid DFT fails to predict a transition state with the 6-311G\*\* basis set. Yet, the B3LYP reaction path can still be utilized in the IRCMax technique to obtain geometries, barrier heights, frequencies, and zero point energy changes compared to reactant that are in good agreement with large basis set quadratic CI (Table 26 & 27). The IRCMax technique is based on the fact that the reaction path defined in molecular coordinates is similar for the constituent high and low levels. When they are dissimilar, the IRCMax is limited in its capacity to improve the frequencies of transition states and zero point energy changes. The transition state for the reaction  $H_2 + CN \rightarrow HCN + H$ , HHCN, involves cyano radical, and spin contamination can be problematic. Hartree-Fock theory gives a mixture of doublet and quartet states, whereas MP2 overcorrects for this spin contamination. Table 24 & 25 reveal that using the HF path as the lower level leads to nearly doubled RMS errors in the frequencies of this transition state, as well as a predicted zero point energy change that over estimates the energy raising effect towards the energy of activation by an order of magnitude. However, as discussed in Section 5.5, if we define the B3LYP/6-311G\*\* as the lower level path, improvement is seen in the frequencies and zero point energy changes (Table 26 & 27). A second limitation of the IRCMaximum method for improving frequencies of transition states and zero point energy contributions of the energy of activation of chemical reactions stems from an inability to achieve sufficient accuracy in calculating geometries along the reaction path, which cause bending modes to behave unphysically since the path is defined in Cartesian rather than polar coordinates. We will show this from the triatomic reaction of  $H_2 + H \rightarrow H_2 + H$  momentarily. Presently, let's turn our attention to Figure 43. Here, the zero point energy change compared to reactants is plotted along the reaction path for  $H_2 + CN \rightarrow HCN + H$ . Note the unphysical "shoulders" present at +/- 0.4 Angstroms in the reaction coordinate.

In Figure 44, the QCISD (T)/6-311G\*\* individual normal mode frequencies for this reaction are plotted along the reaction path. Figure 43:  $H_2 + CN \rightarrow HCN + H$  Normal Mode Frequencies along the QCISD/6-311G\* Reaction Path. While the large frequencies are well-behaved, the smaller frequencies become unphysical as they approach zero, and also have unphysical "shoulders" at -1 and 0.5 Angstroms in the reaction coordinate.

In Figure 45, we examine the normal mode frequencies

of the triatomic degenerate chemical reaction  $H_2 + H \rightarrow H_2 + H$ . Even in this simple reaction the small frequencies become unphysical as they approach zero, while the large frequencies are well-behaved. If we consider the rotational frequencies, presented in Table 5.6.1, for the further simplified system,  $H_2$ , very small deviations on the order of 10 and 100uA from the equilibrium geometry,  $R_e$ , cause the rotational frequencies to become either large positive frequencies when the deviation from  $R_e$  is positive, or large negative frequencies when the deviation from  $R_e$  is negative. In fact, a deviation of 1mA from  $R_e$ , which is the limit of accuracy for polyatomic molecular geometries, causes the rotational frequency of  $H_2$  to be 170cm<sup>-1</sup>!

Consider the rotation of  $H_2$ . In polar coordinates, the potential is parabolic in the magnitude of the deviation from  $R_e$ , as given in Equation (5.6.1) below,

$$V(r) = V_0 + \frac{k_{stretch}}{2}(r-r_e)^2, \quad (5.6.1)$$

Where  $k_{stretch}$  is the force constant for the H-H stretch. If we were to move in the polar coordinate system, the rotational derivatives would be identically zero. However, defining a polar coordinate system for the general polyatomic case can be done with Z-matrices, but is not uniquely defined and analytical derivatives are terrible to calculate. Therefore, current algorithms move in Cartesian coordinates, which are uniquely defined for large systems and analytical derivatives are easy. If we express the potential for rotation, given in Equation (5.6.1) in Cartesian coordinates, we have Equation (5.6.2)

$$V(r) = V_0 + \frac{k_{stretch}}{2}(\sqrt{x^2+y^2}-r_e) \quad (5.6.2)$$

Here, x is in the direction of the internuclear axis. By moving in the Cartesian frame, a rotation results by a deviation in the direction perpendicular to the internuclear axis, the y direction in Equation (5.6.2). The rotational force constant, then, is given by the second derivative of the potential with respect to the y coordinate. The first derivative, given by Equation (5.6.3)

$$\begin{aligned} \frac{\partial V}{\partial y} &= k_{stretch}(x-r_e)(1/x)(0) = 0 \\ \Rightarrow \left( \frac{\partial V}{\partial y} \right)_{y=0} &= 0 \end{aligned} \quad (5.6.3)$$

is identically zero in the limit of an infinitely small displacement in the y-direction, which corresponds to moving along the arc of a circle, which corresponds to movement in polar coordinates. However, the second derivative is not zero in the limit of an infinitely small displacement, as shown in Equation (5.6.4).

$$\frac{\partial^2 V}{\partial y^2} = k_{\text{stretch}} - \frac{k_{\text{stretch}} r_e}{\sqrt{x^2 + y^2}} + \frac{k_{\text{stretch}} r_e y^2}{(x^2 + y^2)^{3/2}} \quad (5.6.4)$$

$$\Rightarrow \left( \frac{\partial^2 V}{\partial y^2} \right)_{y=0} = k_{\text{stretch}} \left( 1 - \frac{r_e}{r} \right)$$

Eq. (8) shows that the rotational force constant will be zero if and only if we are at the exact equilibrium geometry. If  $r > r_e$ , then the rotational force constant will be positive. If  $r < r_e$ , then the rotational force constant will be negative. Table 28 shows that force constants calculated by Equation

(5.6.4) are in good agreement with calculated numerical force constants. The unphysical bending frequencies seen in the degenerate reaction  $\text{H}_2 + \text{H} \rightarrow \text{H}_2 + \text{H}$  are explained by an inability to achieve sufficient accuracy in calculating geometries along the reaction path. While the stretching frequency is so large that it is unaffected by not being exactly on the reaction path, the small H-H-H bending frequency is drastically affected as it approaches zero. Figure 45 shows that the H-H-H bending frequency becomes negative; therefore, it is evident from the discussion above that for each individual step along the path we are deviating from the equilibrium geometry such that  $r < r_e$ .

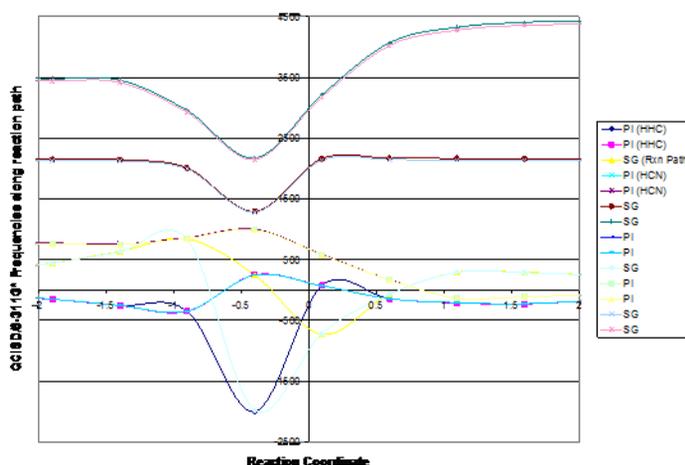


Figure 44: The QCISD(T)/6-311G\*\* individual normal mode frequencies for this reaction are plotted along the reaction path.

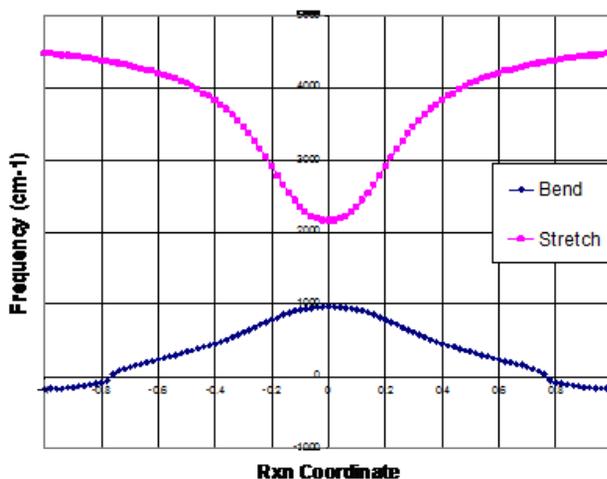


Figure 45:  $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$  Normal Modes as a Function of Reaction Coordinate.

**Table 28:** The effect of deviation from Equilibrium geometry,  $R_e$ , on rotational frequencies of  $H_2$ .

H2 Rotational Frequencies..comparison of cartesian freqs to modeled freqs		
	Gaussian-Numer Deriv. for H2 Rot	$k_{rot} = d^2E/dy^2(N/m)_{y=0} = k_{stretch}(1-r_e/r)$
Re+ $\epsilon$	4.553656971	4.553656971
Re+10uA+ $\epsilon$	17.60133836	17.61070455
Re-10uA+ $\epsilon$	-16.40822754	-16.39259052
e+100uA+ $\epsilon$	53.97435823	53.96672184
Re+1000uA+ $\epsilon$	169.8042087	169.4251476

Re (A)	0.367740625
$\epsilon$ (A)	7.16438E-07

## Conclusion

Being at the correct location of the transition state along the path of a chemical reaction is essential for accurate calculation of frequencies of transition state and zero point energy contributions to the free energy of activation. For symmetric transition states of degenerate chemical reactions in which the position of the transition state is constrained by symmetry to be in the center of the reaction path, traditional techniques such as the B3LYP hybrid density functional provide cost-efficient means of obtaining reasonable frequencies of transition states (Table 18, 23  $cm^{-1}$  error for 6-311G\*\* basis set) and zero point energy changes compared to reactants (Table 18, largest error with  $H_3C-H-CH_3$ -0.352 kcal/mol).

However, more interesting chemical reactions involve chemical changes, in which products *differ* from reactants. For the four, non-degenerate, exothermic, hydrogen transfer reactions examined in this work, it was found that high levels of correlation treatment, quadratic CI, with large basis sets were necessary in order to achieve reasonable frequencies of transition states (Table 22, QCISD(T) with largest basis 6-311+G(2df,2pd) gives 45 $cm^{-1}$  error, but this is an order of magnitude larger with a next smaller basis, 6-311+G(2d,p), 213.6 $cm^{-1}$ ). Similarly, large basis QCISD is necessary to obtain adequate zero point energy changes for the four transition states (Table 23). The different methodologies predict different locations for the transition state along each reaction path. We have shown that while the Intrinsic Reaction Coordinate Maximum Method of Petersson and coworkers can offer a cost-efficient means of attaining reasonable frequencies of transition states and zero point energy changes from reactants to transition state, a good treatment of the low level IRC is needed. Noteworthy examples of the success of the IRCMax are for reactions such as  $H_2 + F \rightarrow HF + H$  and  $H_2 + CN \rightarrow HCN + H$  where the B3LYP hybrid density functional

underestimates an already small barrier height such that there is no predicted transition state for many basis sets. Yet, the B3LYP reaction pathway can still be utilized as the lower level in the IRCMax technique to give reasonable frequencies and zero point energy contributions to the free energy of activation compared to benchmark calculations (Table 26 & 27). However, the IRCMax method fails to improve frequencies of transition states and ZPE changes in the majority of cases examined.

We have also observed that one must be judicious in selecting the constituent high and low levels for spin contaminated cases. For example, in the case of HHCN, Hartree-Fock and MP2 were not good choices since they have unphysical behavior and in opposite ways. However, if the B3LYP is used to define the lower level reaction path, improved frequencies and zero point energy changes from reactants are observed (Table 26 & 27). Additionally, the high and low levels must be chosen so that the IRCMaximum will not be too far away from the low level transition state because bending modes become unphysical as we move away from the maximum on the reaction path. We have shown that this unphysical behavior in the bending modes arises from an inability to achieve sufficient accuracy in calculating geometries along the reaction path, since the path is defined in Cartesian rather than polar coordinates. This result had important consequences for the reaction of  $H_2 + OH \rightarrow H_2O + H$ . The combination of MP2 and Hartree-Fock could not be used for the IRCMax technique because the IRCMaximum fell within a region of the reaction path where the bending modes were behaving unphysically (Figure 43-45). However, because the QCISD(T) maximum is much closer to the B3LYP transition state, RMS errors in the frequencies of this transition state were reduced by a factor of three, and the corresponding zero point energy contributions to the free energy of activation were also improved (Table 26).

# Chapter 5

## The Mechanism and Enantioselectivity of the Interrupted Feist-Bénary Reaction

### Introduction

The Feist-Bénary reaction is used to synthesize furans from  $\beta$ -halogenated ketone and  $\beta$ -dicarbonyl compounds, as shown in Figure 46 [132]. A modification is the interrupted Feist-Bénary reaction (IFB), in which the presence of proton sponge inhibits elimination to form the furan, thereby stopping the reaction at the hydroxydihydrofuran. The first step is the Knoevenagel condensation (related to the Aldol condensation). An amine catalyst abstracts a proton from a  $\beta$ -dicarbonyl enol, assists a proton transfer between the  $\beta$ -dicarbonyl and the pyruvate by migration in protonated form between the two groups. A nucleophilic attack occurs between the  $\beta$ -dicarbonyl enolate anion and the electrophilic pyruvate, but it is unclear at which stage of catalyst migration that this attack is complete. Once the new C-C bond is formed, the pyruvate deprotonates the catalyst. In a subsequent step, an enolate displaces an alkyl halogen to close the ring, as shown in Figure 47.

As we will show in supporting calculations, the aldol forms a complex with the derivatized quinidine asymmetric catalyst through hydrogen bonding, as shown in Figure 48. While a quinine catalyst gives low Enantioselectivity, cinchona alkaloid quinine catalysts that are derivatized to contain substituted pyrimidinyl groups are highly Enantioselective [133-135]. Enantioselective IFB reactions are synthetically important. For example, zaragonic acid is a fungal metabolite that is a cholesterol reducing agent. Recently, Calter et al. [135] designed the synthesis of a novel

zaragonic acid core, in which the first step is an interrupted Feist-Bénary reaction. A better understanding of the catalytic action facilitates enhancing the enantioselectivity. It is the goal of this work to understand through computation how these asymmetric catalysts interact with reactants in order to confer the observed enantioselectivity.

### General considerations

The aldol bond forming step is thought to be the stereochemistry determining one, since it is the ketone carbon of the alpha halogenated pyruvate that is pro-chiral. The R- and S- enantiomers are shown in Figure 49, neglecting the side chain on quinidine for the moment. Where the side chain is located with respect to the aldol is another variable, since there are multiple conformational minima, six for each diastereomer (the carbon alpha to the quinuclidine backbone nitrogen of the side chain is itself chiral), as shown below. The staggered structures correspond to the hydrogen bonding between the aldol-catalyst complex and hydrogens of two different carbons alpha to the quinidine nitrogen. Three structures are possible for this case, depending on the alignment of the side chain. The eclipsed structures correspond to hydrogen bonding between the aldol-catalyst complex and hydrogens on the same carbon alpha to the quinidine nitrogen (Figure 50). The derivatized quinidine asymmetric catalyst is shown in Figure 51. It was observed at the ONIOM(B3LYP/6-31G\*:HF/3-21G\*) level that all three eclipsed structures for the R-enantiomer simply rotated into their corresponding staggered structures, presumably due to steric. Therefore, further discussion will be limited to the 6 possible staggered structures.

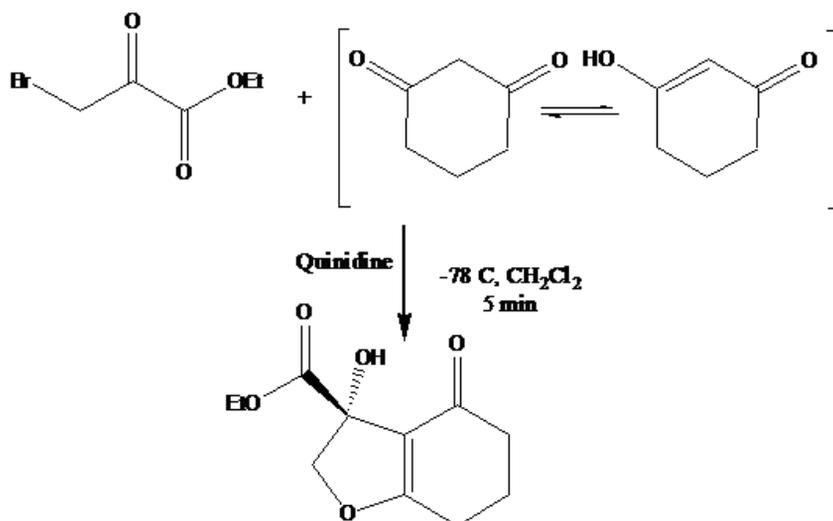
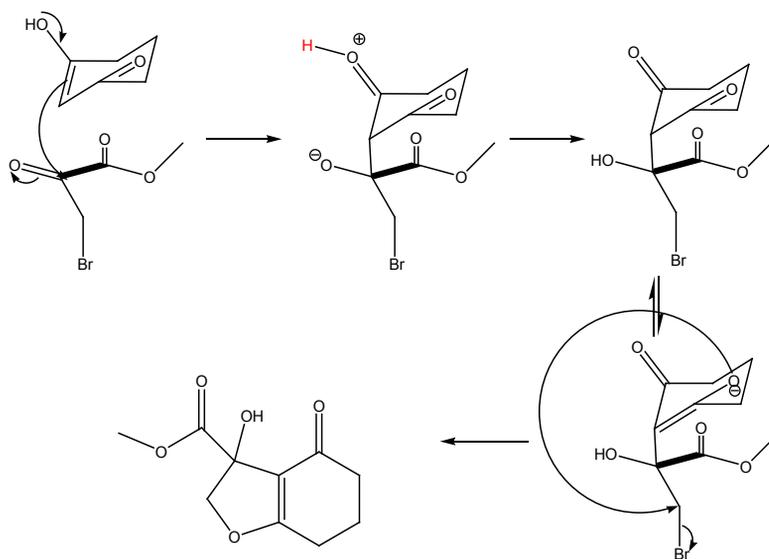
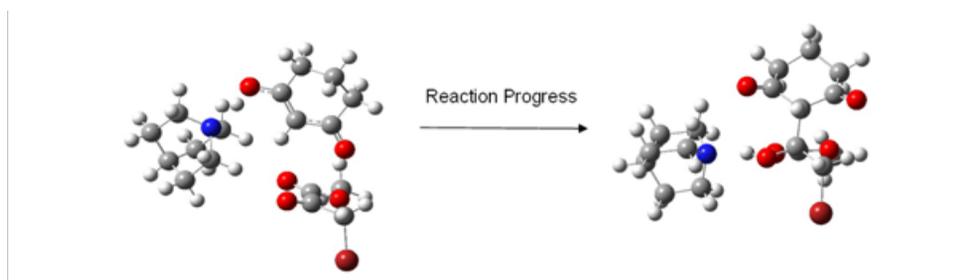


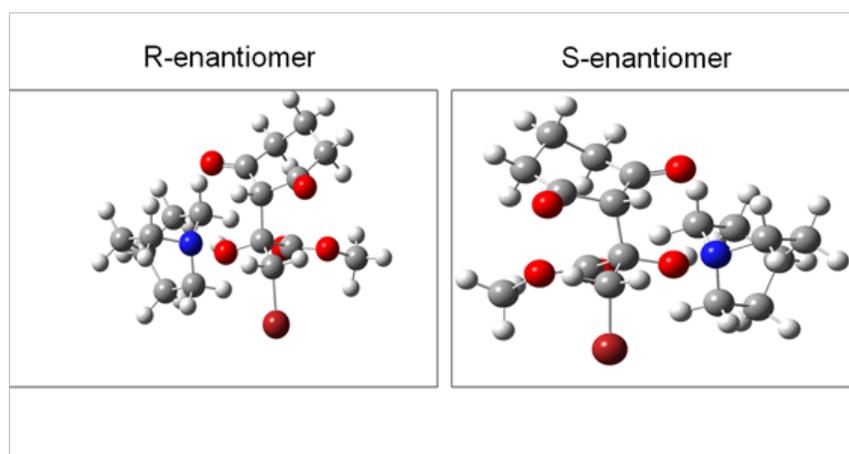
Figure 46: The Interrupted Feist-Bénary Synthesis.



**Figure 47:** General Mechanism of the Uncatalyzed Feist-Bénary Reaction.



**Figure 48:** Hydrogen Bonding Interaction between catalyst and aldol complex through reaction progress.



**Figure 49:** R and S Enantiomers Excluding Side Chain and Vinyl Moiety (Model Layer).

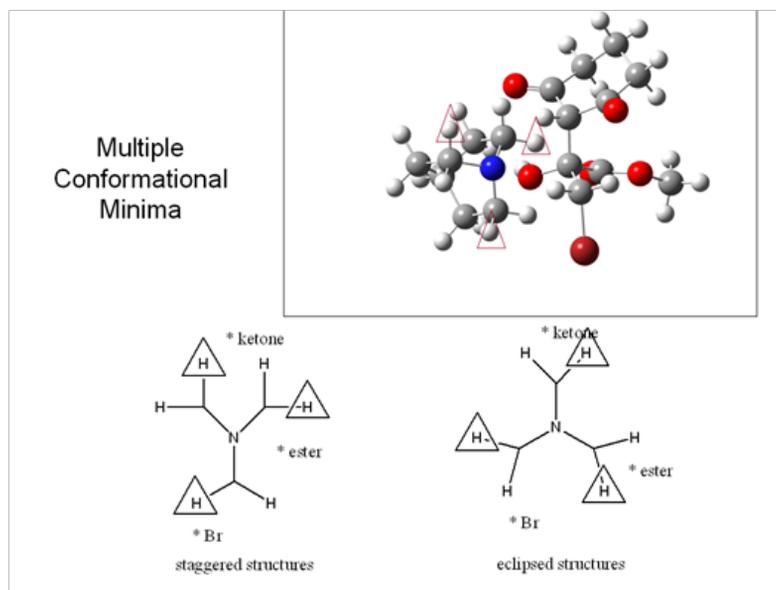


Figure 50: Multiple Conformational Minimum Energy Structures.

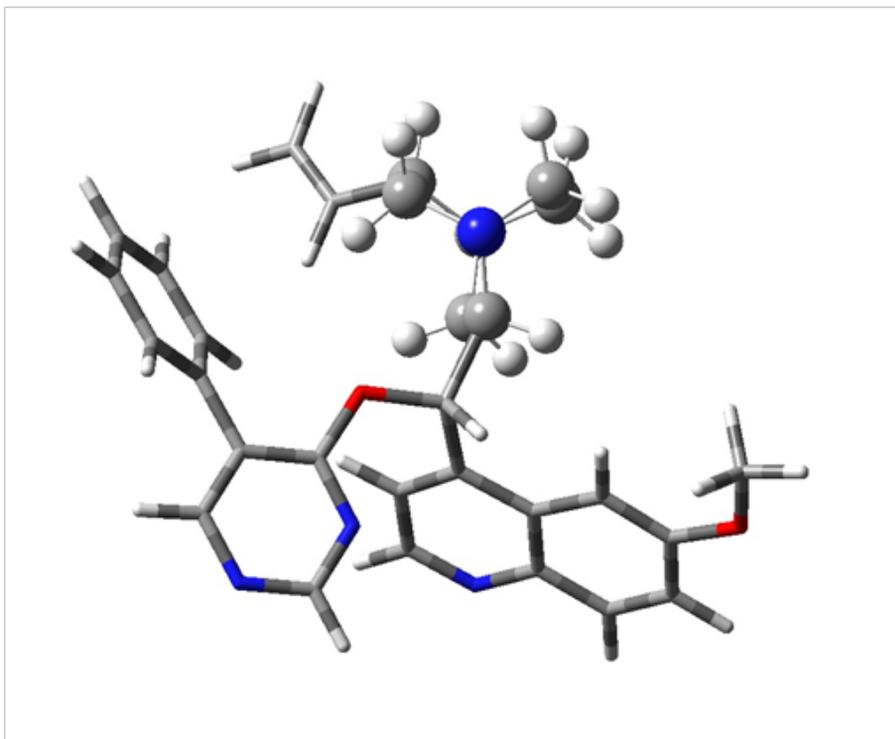


Figure 51: Derivatized Quinidine Asymmetric Catalyst.

## Difficulties encountered

In order to estimate the enantioselectivity, it is ideal to compare the complete reaction profiles for each of the six rotomers. We will briefly examine an ONIOM(B3LYP/6-31G\*:HF/3-21G\*) reaction profile for the preferred R and S enantiomers, and show that due to basis set superposition error, this level of theory has insufficient resolution in the thermochemistry to explain the enantioselectivity. We will next show an ONIOM calibration study that was performed in order to minimize the basis set superposition error. However, the final ONIOM level of theory was found to be impractical for this system. In Section 5.4, we will present the final results for the six possible transition states with the best methodologies practical to perform. This will allow us to obtain an estimate of the enantioselectivity.

The ONIOM(B3LYP/6-31G:HF/3-21G\*) Free Energy of Activation of Stationary Points are shown in Figure 52. The reactant-catalyst complex exhibits substantial binding (14.3kcal/mol for R-enantiomer and 5.1kcal/mol for the S-enantiomer) relative to the free reactants. The momentum gained from the binding affinity after formation of the reactant-catalyst complex for the R-enantiomer “sees” a 5.6kcal/mol forward barrier, while the S-enantiomer “sees” only a 1.98 kcal/mol barrier. At this level of theory, the S-mechanism is favored. However, these surfaces are flat, and therefore approximations were necessary. First, a true transition state could not be located using the full system. The transition state stand-in was the minimum energy structure corresponding to equal bond distances between the catalyst nitrogen and oxygens of the pyruvate and cyclohexadione. Frequency calculations revealed small negative modes for the reactant and product catalyst complexes, which means that we cannot be confident we are truly at energy minima. In light of the inherent difficulties of describing accurately such a large system that has a flat potential surface, we began investigating the model system alone, but with a larger basis set. We used the B3LYP method combined with the larger basis set employed for the B3LYP geometries in the CBS-QB3 method, 6-311+G(2d,p).

However, as shown in Table 29, since the basis set superposition error is changing by 2.6 kcal/mol along the reaction path, this indicates there is insufficient resolution in even for this larger basis set for describing the thermochemistry. An EE of 94% requires that the basis set superposition error change by less than 1kcal/mol along the path. Therefore, we abandoned further studies with the ONIOM(B3LYP/6-31G:HF/3-21G\*) level and embarked on an ONIOM calibration study intended to reduce the basis set superposition error. We began by further restricting the model system to include only the 6 heavy atoms in the highest layer (shown in Figure 54 but with ammonia rather than thrimethylamine), and performed CBS-APNO on this layer. Basis set superposition error is negligible with the APNO method. When we compared the CBS-QB3 reactant-

complex dissociation and the CBS-QB3 endothermicity with the APNO results, we found that them to be in good agreement. The reactant-complex was 1.84 kcal/mol more tightly bound for the APNO, and the endothermicity was 0.5 kcal/mol lower for the APNO. Therefore, in the true ONIOM calculation for the full molecule, we would use the CBS-QB3 for the highest 6-heavy atom layer. We also compared the CBS-4M model chemistry to the CBS-APNO and CBS-QB3. Since we have a closed shell singlet all along the path, this method is performing as well as the QB3 for the highest layer. Its reactant-complex dissociation energy is 0.5kcal/mol lower for the QB3 and 2.3 kcal/mol lower for the APNO, and the endothermicity is actually lower for the 4M by 0.48 kcal/mol compared with the APNO and 0.98 kcal/mol compared with the QB3. Next, we introduced a medium layer that included 18 more heavy atoms. We have calculated the CBS-QB3 for the product-catalyst complex medium layer and compare it with the CBS-4M. The CBS-4M alone approximates the QB3 well, indicating that only a 2-layer ONIOM partitioning is needed. The dissociation energy of the reactants is 3.2kcal/mol more strongly bound in the QB3, and the error in the barrier height is only 0.57kcal/mol with 4M being the smaller barrier.) For the final step of this calibration step was to decide which method is adequate for the lowest layer. Since electrostatics of the side chain of the catalyst are important for structure, we performed a B3LYP/6-31+G\* calculation to extract CHELP charges, and read these into a Molecular Mechanics, Universal Force Field (UFF) calculation for the low layer. This will accomplish incorporating better electrostatics than HF/3-21G\* while retaining computational speed.

The final partitioning, in which a CBS-4M single point energy is performed on the model system, and the low layer is treated with UFF molecular mechanics with embedded B3LYP/6-31+G\* Chelp charges will be designated: ONIOM(CBS-4M:UFF(B3LYP/6-31+G\*Chelp\_charges)). We then embarked on a transition state search for the model system. We first sought a HF/3-21G\* transition state for the model layer shown in Figure 53, but found that we needed to begin the search with the more limited nine heavy atom layer system of Figure 54.

Figure 54 shows the nine heavy atom system transition state found at the HF/3-21G\* level. In the reaction profile of Figure 53, the HF/3-21G\* electronic energy profile for 9 heavy atom system is shown. The first three points are obtained from an Intrinsic Reaction Coordinate Scan (IRC) obtained from a true first order saddle point, with reaction path frequency approximately 300cm<sup>-1</sup>. The remainder must be obtained from an optimization to product-catalyst complex since there exists a broad plateau where the gradient is too shallow to follow by IRC. In the first portion of the reaction path, the energy rises sharply (1), where the hydrogen transfers from the cyclohexadione enol oxygen to the nitrogen of the H-transfer catalyst. This serves to build electron density on the nucleophilic carbon beta to this

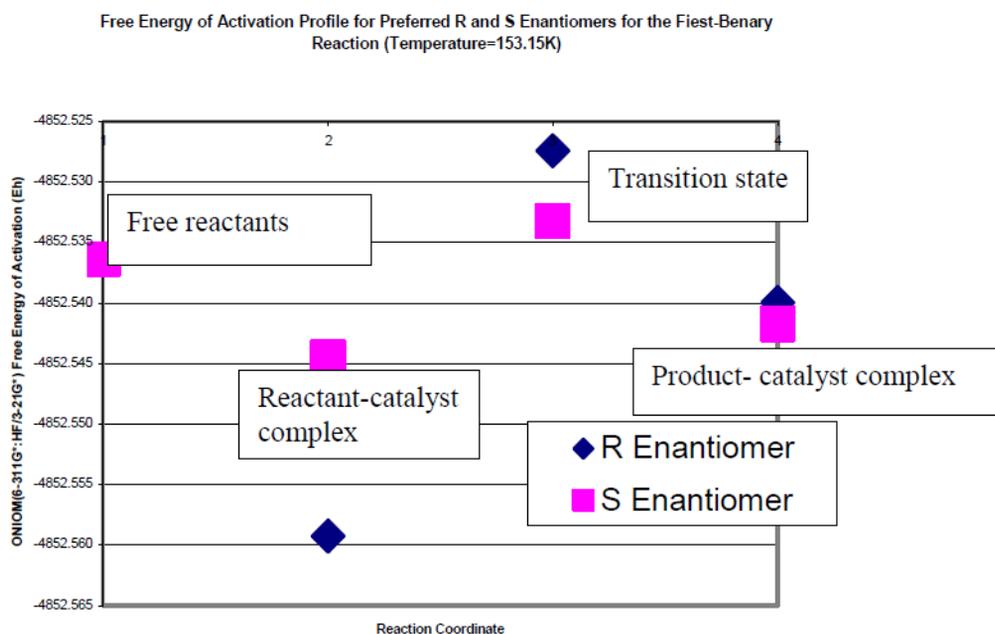
enol oxygen. Once the hydrogen is transferred, the energy proceeds downhill (2) as the aldol C-C bond is formed in a broad intermediate (3) where the protonated nitrogen catalyst shuffles between the oxygens of the nucleophile and electrophile species. Once the N-H protonated catalyst has found the oxygen of the electrophile, the energy proceeds downhill (4) to product-catalyst complex. Therefore, at the HF/3-21G level, the reaction path is highly decoupled, with hydrogen transfer preceding nucleophilic attack for aldol C-C bond formation.

The remainder of the model layer of the full system was added, and a HF/3-21G\* transition state was located for this system. The HF/3-21G\* transition state for the model layer is shown in Figure 55. The model layer of the full system is quite different than the 9 heavy atom system since the nucleophilic carbanion is stabilized by the presence of the second flanking carbonyl moiety. The Intrinsic Reaction Coordinate calculation of the model system at the HF/3-21G\* level is shown in Figure 56. This stabilization changes the reaction mechanism drastically, and is calculated from the Intrinsic Reaction Coordinate as follows: Hydrogen transfer to the nitrogen occurs rapidly to form an anion that is stabilized by conjugation; therefore, it need not attack straight away. Instead, the shuttling of the N-H catalyst between oxygens precedes nucleophilic attack, and the subsequent energy lowering when the N-H catalyst finds the pyruvate oxygen. Electron correlation becomes more important in this system than seen previously due to the conjugation of electrons in the cyclohexadione, and therefore a new method must be considered for the geometry method besides HF/3-21G\* where electron correlation effects are considered only in an average sense. Therefore, the final partitioning, chosen to minimize basis set superposition error, was not appropriate for the system under investigation since the HF/3-21G\* geometry method is devoid of instantaneous electron correlation.

The first attempt to include electron correlation was to use the B3LYP/6-31G\* method (with p polarization function on the transferring hydrogen). In order to do this, the B3LYP IRCMaximum was located along the HF path, and is shown in Figure 57. The B3LYP maximum was shifted dramatically toward products, and therefore points flanking the IRCMaximum were used as guesses to locate the B3LYP/6-31G\* (with p polarization function on transferring hydrogen) shown in Figure 58. A frequency calculation performed on the B3LYP transition state above reports a  $-142.9\text{cm}^{-1}$  negative eigenvalue. We searched for a second B3LYP transition state using the points employed for the HF/3-21G\* guess, but could find no second B3LYP transition state. Next, we searched for an MP2/6-31G\* (with p polarization function on transferring hydrogen) transition state, again beginning with locating the MP2 IRCMaximum along the HF/3-21G\* path. This IRCMaximum is shown in Figure 59.

The MP2/6-31G\* (with p polarization function on transferring hydrogen) IRCMaximum was this time shifted toward reactant in terms of the position of the pyruvate compared to the HF/3-21G\*. Points flanking the IRCMaximum were used to set the search for the MP2 transition state. Because points in the IRCMaximum region were used, this transition state is designated "MP2/6-31G\* IRCMax region." Figure 60 shows the MP2/6-31G\* IRCMax region Transition State. A second, later transition state was located at the MP2/6-31G\* (with p polarization on transferring hydrogen) using the original points for HF/3-21G\* transition state guess. Since the HF points were used, this transition state is designated "MP2/6-31G\* HF region". The MP2/6-31G\* HF region transition state is shown in Figure 61. The primary way in which the HF/3-21G\*, B3LYP/6-31G\* (with p polarization function on transferring hydrogen), MP2/6-31G\* (with p polarization function on transferring hydrogen) IRCMax region, and MP2/6-31G\* (with p polarization function on transferring hydrogen) HF region transition states differ is the position of the catalyst between the cyclohexadione and the pyruvate. Figure 62 plots the difference in O-H distance between the cyclohexadione and pyruvate as a function of aldol C-C bond distance for these transition states, as well as reactant-catalyst and product catalyst complex (for reference).

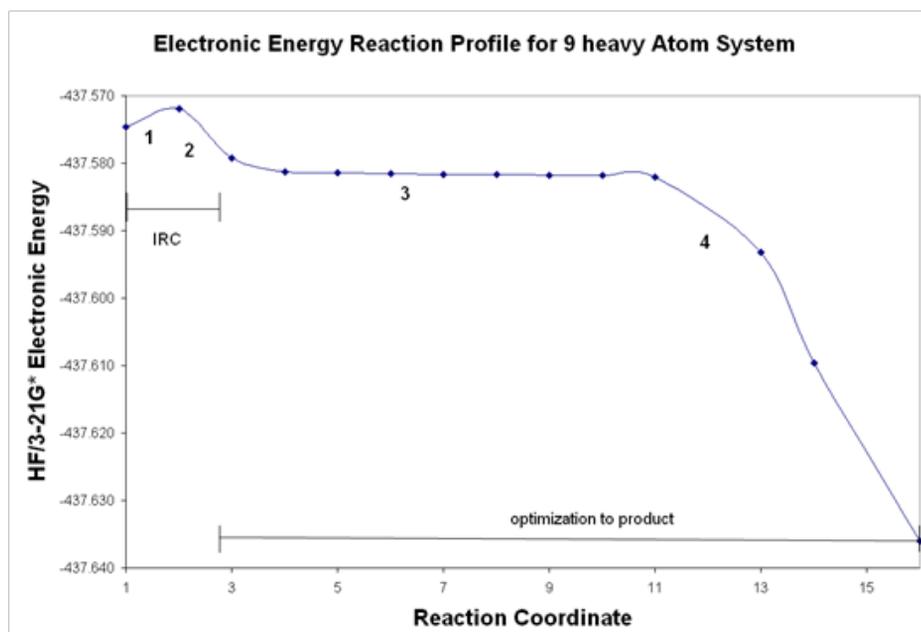
The blue dashed lines of Figure 62 indicate the displacement from IRCMax structures for the B3LYP/6-31G\* (with p polarization function for transferring hydrogen) and MP2/6-31G\* (with p polarization function for transferring hydrogen) IRCMax region. In both cases, the transition state optimizations shifted away from the HF/3-21G\* path substantially. Also shown, falling between the product-catalyst complex, and the MP2/6-31G\* IRCMax region transition state are a sequence of blue dots representing the HF/3-21G\* IRC for the model system. The IRC was not extended all the way to reactants and products, but the aldol C-C bond was 1.67 Angstroms in the furthest point towards product catalyst complex, and the catalyst is still associated with the pyruvate. Therefore, at the HF/3-21G\* level of theory, if the catalyst is ever associated with the cyclohexadione, deprotonation and migration to the pyruvate occurs immediately, while C-C bond formation is the predominate variable. A variety of transition states were found that vary in the position of the catalyst between the cyclohexadione and the pyruvate. The best approximate we can make is to say that the true transition state lies somewhere in the region along the diagonal between reactant and catalyst complex. Therefore, we select for our model layer the one most closely in the middle, the MP2/6-31G\* (with p polarization on transferring hydrogen) HF region.



**Figure 52:** The ONIOM(B3LYP/6-311G\*:HF/3-21G\*) Free Energy of Activation Along the Interrupted Feist-Bényary Reaction Path.

**Table 29:** Basis Set Super Position Error (BSSE) corrections along B3LYP/6-311+G(2d,d,p) reaction path.

BSSE correction reactants	7.053840
BSSE correction products	4.409513
$\Delta E$ BSSE along path(Kcal/mol)	-2.644327



**Figure 53:** HF/3-21G\* Electronic Energy Profile for Nine Heavy Atom System. The first three points are obtained from an Intrinsic Reaction Coordinate Calculation. The remainder is obtained from an optimization to product-catalyst complex.

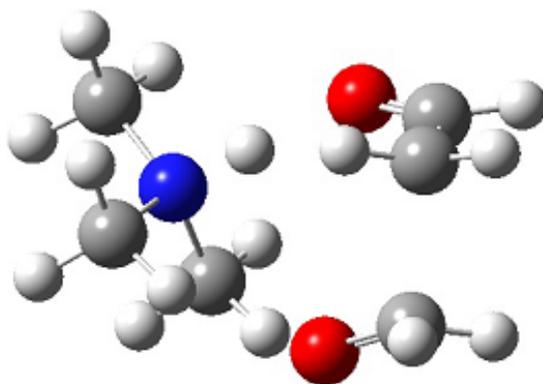


Figure 54: Nine Heavy Atom HF/3-21G\* Transition State.

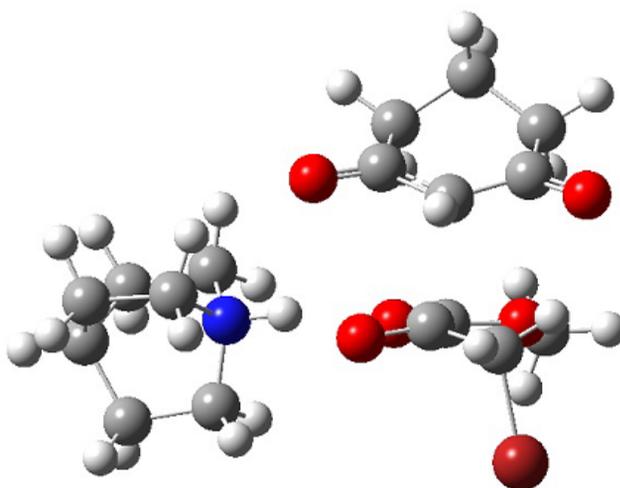


Figure 55: HF/3-21G\* Transition State for Model Layer.

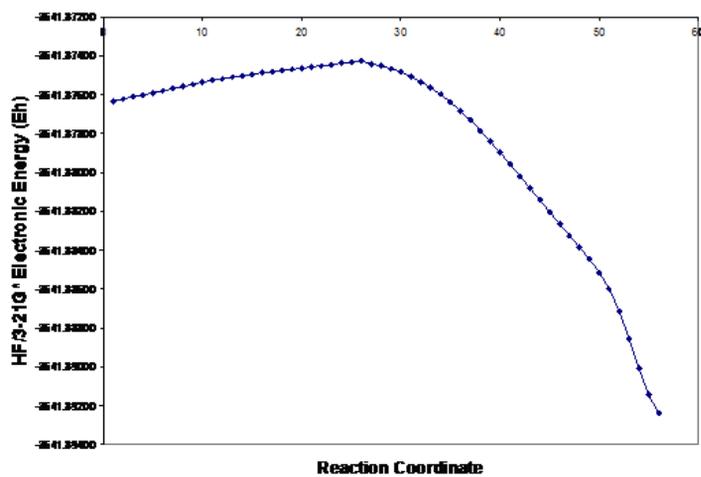
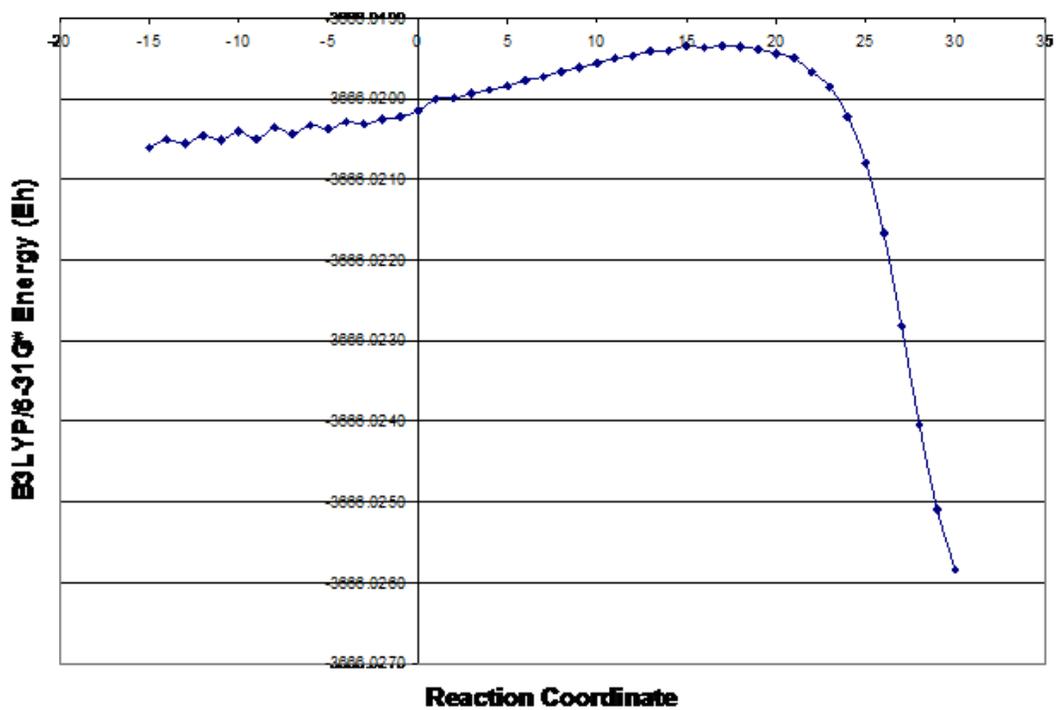
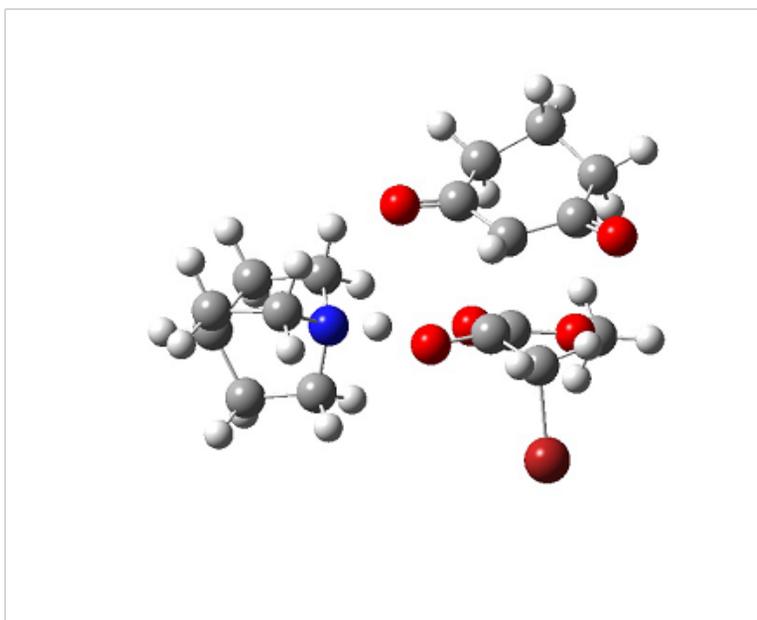


Figure 56: HF/3-21G\* Intrinsic Reaction Coordinate Calculation on the Model Layer.



**Figure 57:** B3LYP/6-31G\* (with p polarization function on transferring hydrogen) Intrinsic Reaction Coordinate Maximum along HF/3-21G\* path.



**Figure 58:** B3LYP/6-31G\* (with p polarization function on transferring hydrogen) Transition State for Model Layer.

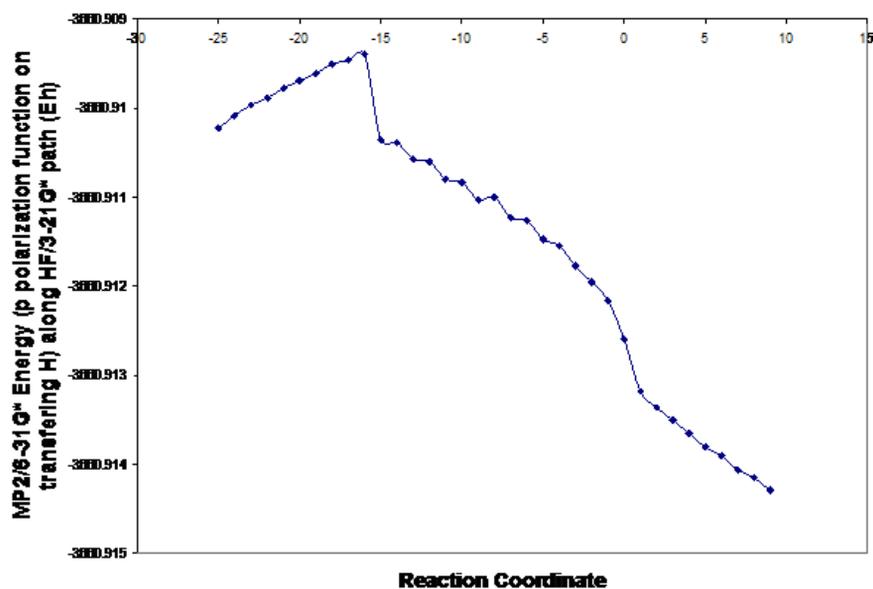


Figure 59: MP2/6-31G\* (with p polarization function) IRCMaximum along HF/3-21G\* path for model layer.

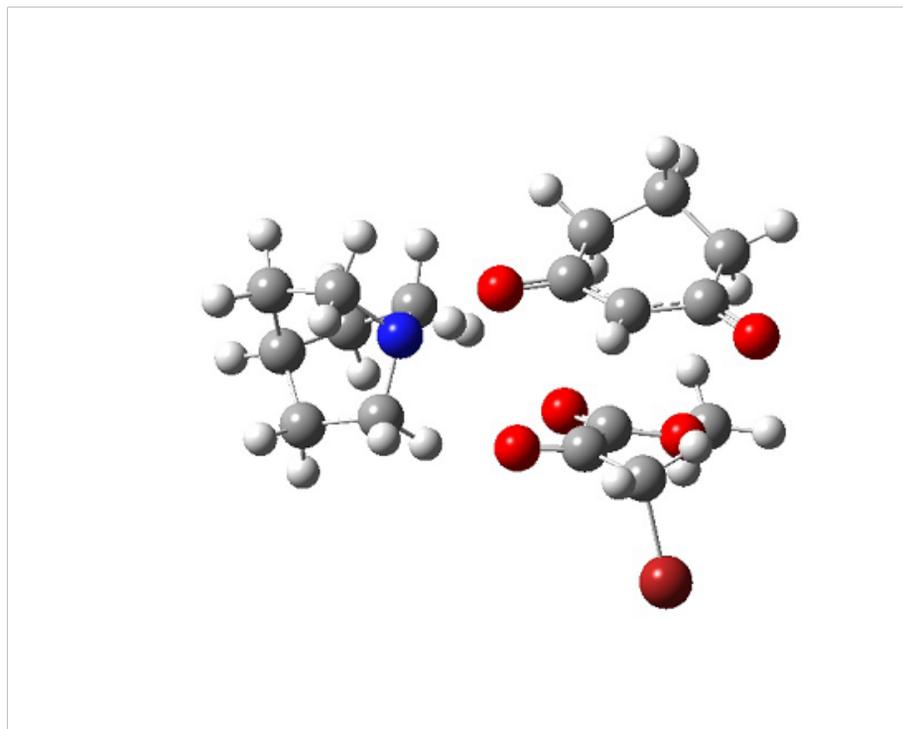


Figure 60: MP2/6-31G\* (with p polarization function on transferring hydrogen) IRCMax Region Transition State for Model Layer.

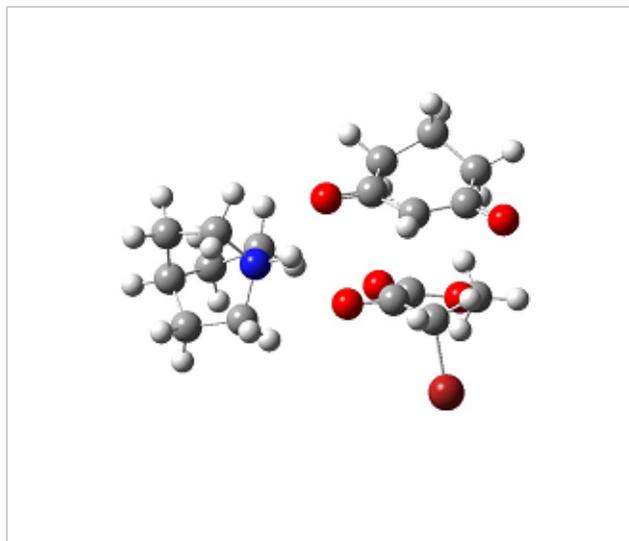


Figure 61: MP2/6-31G\* (with p polarization on transferring hydrogen) Hartree-Fock region Transition State for the Model Layer.

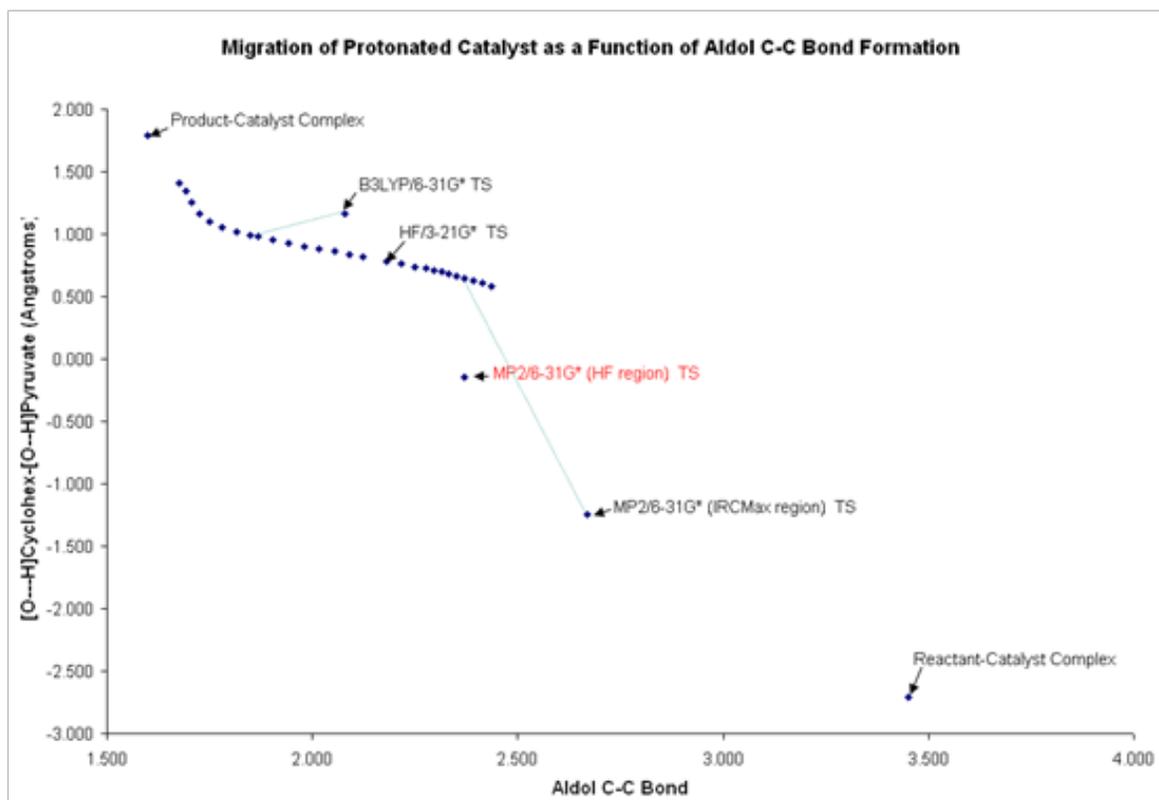


Figure 62: O-H distance of Cyclohexadione minus O-H distance of Pyruvate as a function of Aldol C-C bond distance.

## Final results

The MP2/6-31G\* (with p polarization function on the transferring hydrogen) model layer was frozen to the MP2 structure and the derivatized side chain was added in the three possible conformations for each R and S enantiomer and optimized to a minimum using the UFF molecular mechanics method with embedded B3LYP/6-31+G\* chelp charges. The three possible transition state structures for the R enantiomer are presented in Figure 63 through Figure 65, and the three rotamers for the S enantiomer transition states are given in Figure 66 through Figure 68. In Figure 63 through Figure 68, the transferring hydrogen has been replaced with a fluorine for emphasis. The R-Br and S-ketone structures have less steric interactions between the side chain of the catalyst and the aldol complex. This is reflected in the ONIOM(MP2/6-31G\*\_extra\_polar:UFF\_ChelpCharge) energies presented in Table 30.

The enantioselectivity is estimated to be 5.2kcal/mol, which is an overestimate for 94% ee at -78 °C (1.2 kcal/mol is needed to explain the selectivity). Figure 69 shows the preferred R and S enantiomers in orientations that expose potential reasons for the selectivity. The difference in energy between the two conformations of the catalyst in the absence of the substrate is about half (4mEh vs 8mEh) of the observed between the two enantiomers. Therefore, there are steric interactions within the side chains of the catalysts that account for some of the observed selectivity. The close association of the pyrimidinyl ring with the bromine and forming aldol C-C bond may cause a repulsion that is present in the R enantiomer but avoided in the S enantiomer. The selectivity may have been overestimated because the model system was frozen when the side chains were added. Only the side chains were allowed to optimize to minimum energy structures. If it were feasible to re-optimize the entire system at the ONIOM(MP2/6-31G\*\_extra\_polar:UFF\_ChelpCharge) level, the catalyst may shift toward the pyruvate, thereby alleviating some of the repulsion in the R enantiomer. In order to improve the enantioselectivity of the catalyst, one might consider modifying the side chain to provide rigidity between the pyrimidinyl-phenyl group to create a derivatized "diazaphenanthrene" moiety within the side chain, as shown in Figure 70. This rigid phenanthrene group provides additional hindrance in the R-enantiomer, but this strain is easily relieved in the S enantiomer, as shown in Figure 71.

Other approximations that will have affected our ability to predict the observed enantioselectivity are that the calculations were all gas phase, yet the solvent used in experiments was methylene dichloride. The high dielectric constant of this solvent will have effects that can be estimated. This must be left for a future student to estimate through available computational techniques. Additionally, MP2 frequencies required larger amounts of disk than available to the author (290 GB of localscratch but the hosts on Swallowtail cluster computing system ehwfd queue have only 230 GB available). Incorporating the Gibb's Free

Energy, provided by the partition functions calculated during the frequency calculations, would have allowed for inclusion of entropic effects which are sizeable for ter-molecular reactions. When the disk needs can be met, these effects must be considered.

## Conclusion

The mechanism and enantioselectivity of the Interrupted Feist-Bénary reaction was studied. Due to the flatness of the potential energy surface, several limitations were introduced. Even stable structures (i.e., reactants and products) could not be calculated with sufficient accuracy to be confident that we have truly located a minimum on the potential energy surface. The transition states were challenging to obtain, and we were restricted to rigorous studies on the model system. The side chains that provide enantioselectivity had to be added in a second step, and further transition state optimizations were not feasible. Six transition states were located with the ONIOM(MP2/6-31G\*\_extra\_polar:UFF\_ChelpCharge) method, which is the best choice for balancing accuracy and computational expense. We predict a 5.2 kcal/mol selectivity that is in favor of the S enantiomer, which is an overestimate. The selectivity for the S enantiomer was attributed to steric interactions within the side chains of the catalysts for the two enantiomers, as well as steric and electrostatic interactions between the catalyst and aldol complex. The repulsion between the pyrimidinyl group and the approaching nucleophile may inhibit attack present in the R enantiomer is not observed in the S enantiomer. A modification in the side chain is proposed to enhance the selectivity.

As the reader is by now aware, there were several difficulties encountered during the study of the Feist-Bénary reaction mechanism and enantioselectivity, owing to the flatness of the potential surface. Many insights were gained that will benefit the future student seeking to elucidate the reaction profile for this system. In this numbered list of steps, I share my advice to those conducting further studies:

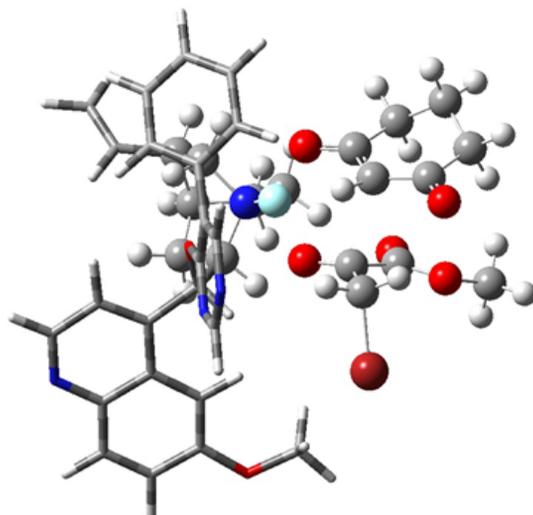
1. A common error (and one my advising professors made at the outset!) is to simply "pick" level(s) of theory to run the calculations. The first step should be to perform a calibration of the level(s) of theory against a benchmark known to minimize potential sources of error (e.g., basis set superposition error) which may be problematic. In identifying sources of error, it is important to be ever mindful of the reaction at hand, and this I did not do when choosing the CBS-4M as the optimized level for the model layer. The geometry method for the CBS-4M is the HF/3-21G\* level, which neglects correlation effects. Yet, correlation energy is important for when the nucleophilic attack occurs in the IFB reaction. Of course, with current computational limitations within the group, the CBS-QB3 method would not be practical for the model layer, but it ought to have been clear the best choice would be the MP2 level since B3LYP underestimates barrier

heights. In summary, the correct ONIOM partitioning is: ONIOM(MP2/6-31G\*: UFF) where a p polarization function is on the transferring hydrogen, and B3LYP/6-31+G\* Chelp Charges are embedded in the UFF (to ensure the best treatment of electrostatics that is practical). Only use the MP2 level if frequencies are practical. If not, then the B3LYP level must be employed instead.

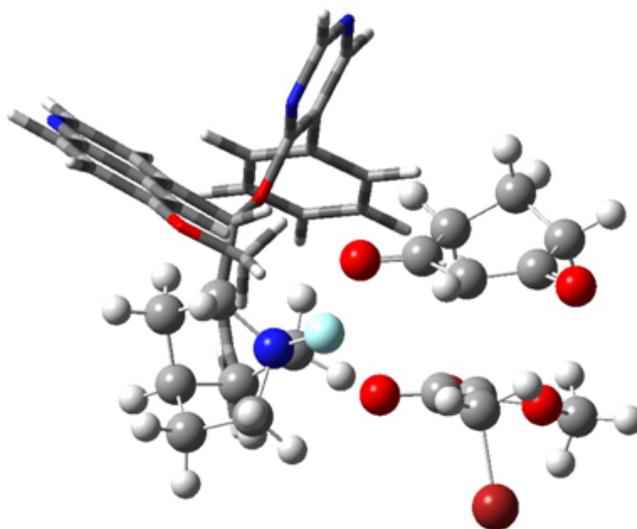
II. Solvent effects are important to include in all optimizations (e.g., to stable minimum energy

structures, transition states, and during calculation of IRC's). It is important to obtain frequency calculations at  $-78^{\circ}\text{C}$  since this is the temperature the reaction achieves best enantioselectivity.

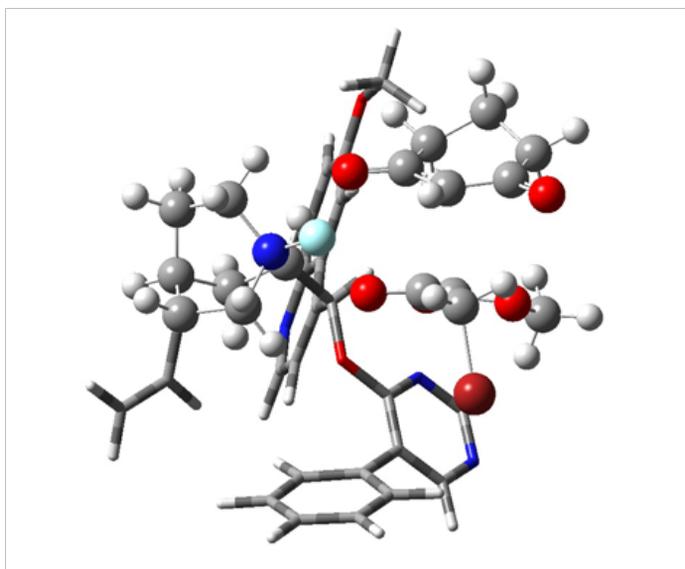
III. In estimating the enantioselectivity, it is necessary to compare barrier heights, whereas in this work the absolute energies of the preferred enantiomers were compared.



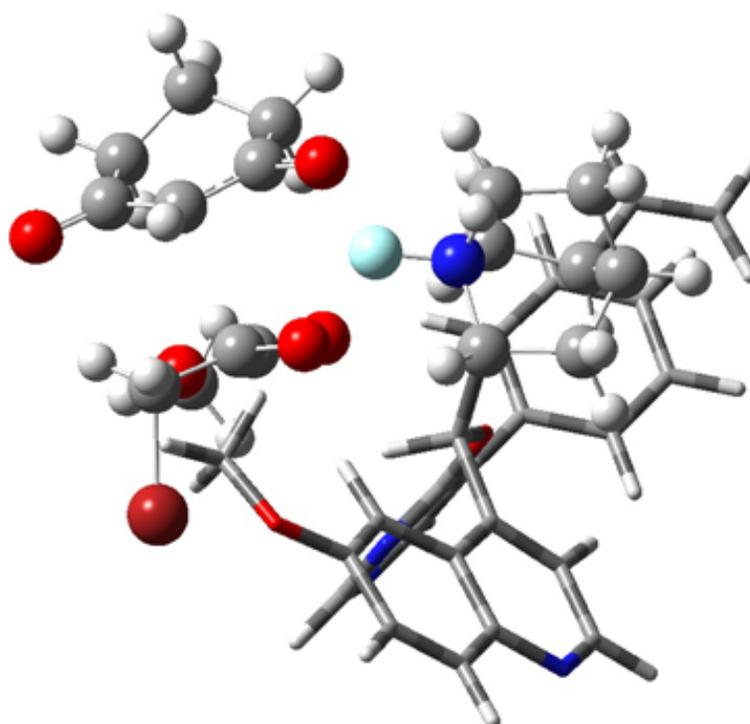
**Figure 63:** R Enantiomer, Side Chain by Br. "R-Br" (Lowest Energy).



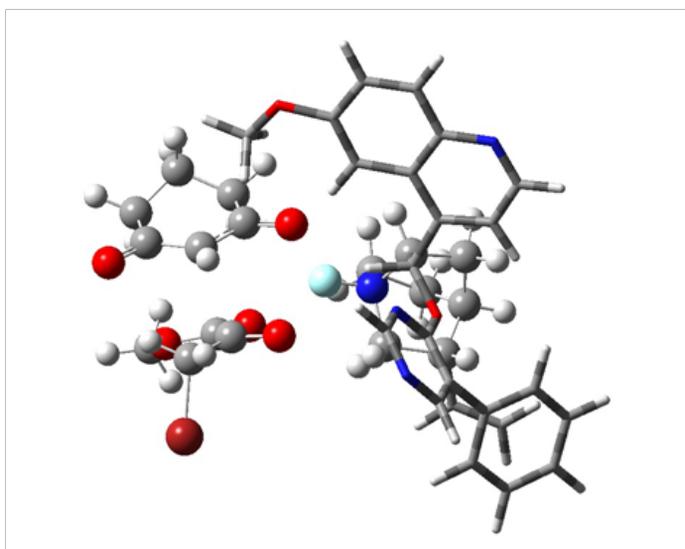
**Figure 64:** R Enantiomer. Side Chain by Ketone. "R-Ketone" (2<sup>nd</sup> Lowest Energy).



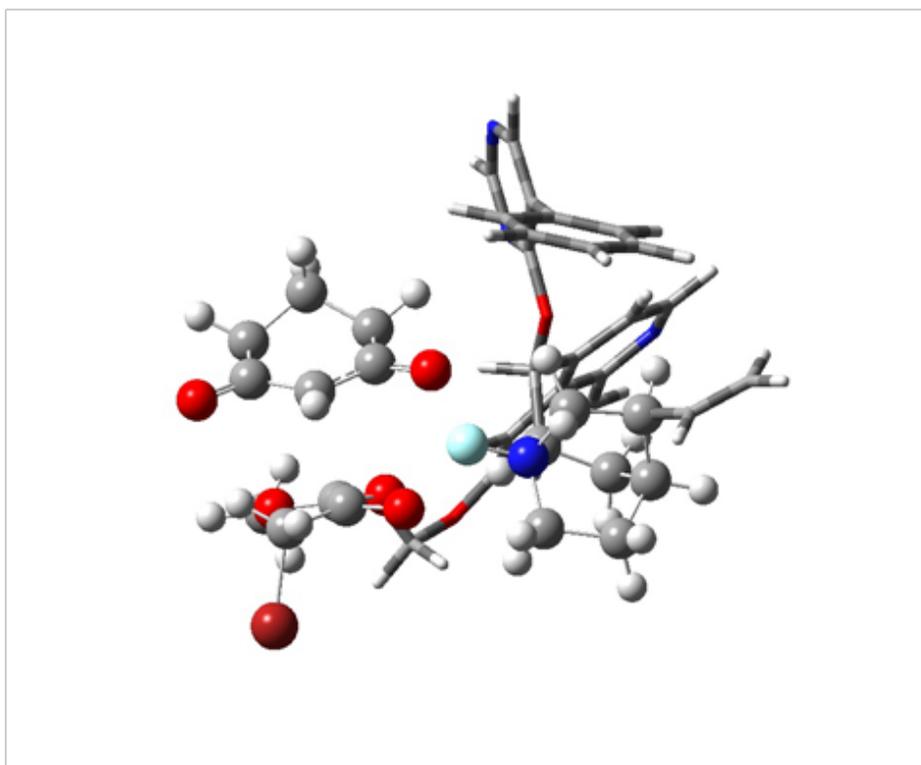
**Figure 65:** R enantiomer. Side chain by acetate. "R-Acetate" (Highest Energy).



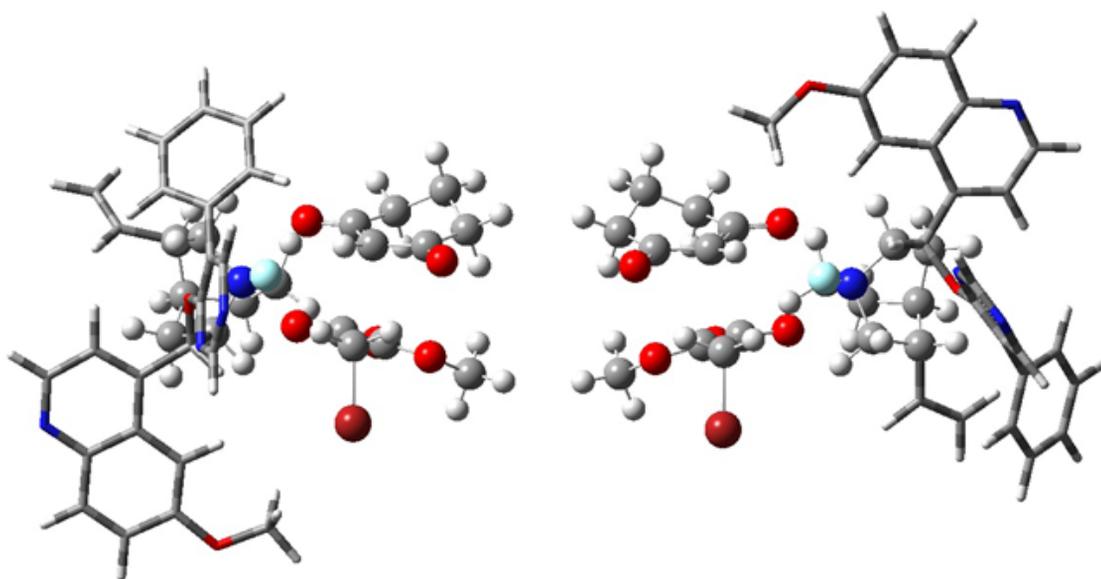
**Figure 66:** S enantiomer. Side chain by Br. "S-Br" (2<sup>nd</sup> Lowest Energy).



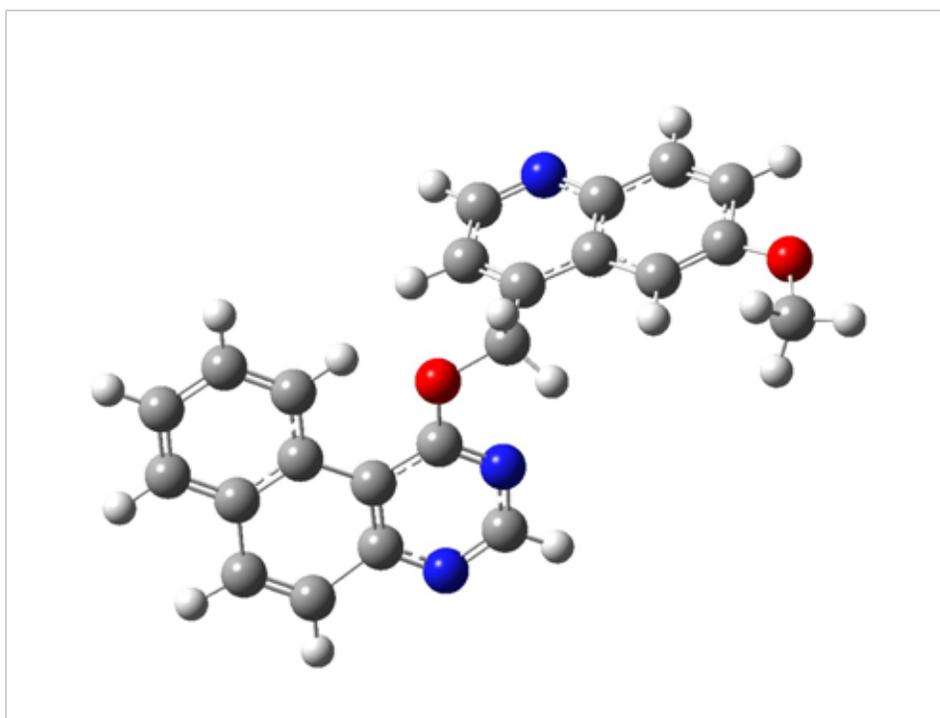
**Figure 67:** S enantiomer. Side Chain by Ketone. "S-Ketone" (Lowest Energy).



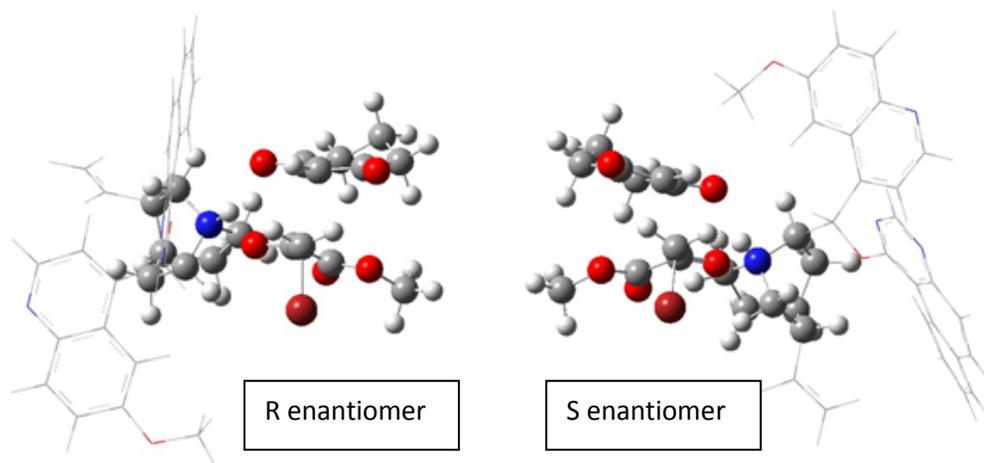
**Figure 68:** S enantiomer Transition States. Side Chain by Acetate. "S-Acetate" (Highest Energy).



**Figure 69:** Preferred R (R-Br) and S (S-Ketone) Enantiomer Transition States.



**Figure 70:** Quinuclidine Side Chain with diaza phenanthrene group.



**Figure 71:** ONIOM(MP2/6-31G\*\_extra\_polar:UFF\_ChelpCharge) R and S Transition States with Modified Quinuclidine Side Chain (dinitro "phenanthrene" moiety).

**Table 30:** ONIOM(MP2/6-31G\*\_extra\_polar:UFF\_Chelp Charge) energies (Eh) for the Six Transition States.

Transition State	Energy (Eh) New Chelp Charges (Side Chain And Model Layer Separate) ONIOM(MP2/6-31G*_extra_polar:UFF_ChelpCharge)
R-Br	-3661.267619
R-Ketone	-3661.258536
R-acetate	-3661.105597
S-Br	-3661.243631
S-Ketone	-3661.275908
S-acetate	-3661.212290

# References

## References

- Gaussian 70
- Boys SF (1950) Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proceedings of the Royal Society* 200(1063): 542-554.
- Roothan CCJ (1951) New Developments in Molecular Orbital Theory. *Reviews of Modern Physics* 23: 68-69.
- Pople JA, Nesbet RK (1954) Self Consistent Orbitals for Radicals. *The Journal of Chemical Physics* 22: 571.
- Møller C, Plesset MS (1934) Note on an Approximation Treatment for Many-Electron Systems. *Phys Rev* 46(7): 618.
- Pople JA, Seeger R, Krishnan R (1977) Variational Configuration Interaction Methods and Comparison with Perturbation Theory. *International Journal of Quantum Chemistry* 12(S11): 149-163.
- Krishnan R, Pople JA (1978) Approximate fourth-order perturbation theory of the electron correlation energy. *International Journal of Quantum Chemistry* 14(1): 91-100.
- Pople JA, Krishnan R, Schlegel HB, Binkley JS (1978) Electron correlation theories and their application to the study of simple reaction potential surfaces *International Journal of Quantum Chemistry* 14(5): 545-560.
- Cizek J (1969) On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules. *Adv Chem Phys* 14: 35.
- Purvis GD, Bartlett RJ (1982) A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics* 76(4): 1910-1918.
- Scuseria GE, Janssen CL, Schaefer III HF (1988) An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *J Chem Phys* 89(12): 7382.
- Scuseria GE, Schaefer HF (1989) Is coupled cluster singles and doubles (CCSD) more computationally intensive than quadratic configuration interaction (QCISD)? *The Journal of Chemical Physics* 90: 3700.
- Pople JA, Head-Gordon M, Raghavachari K (1987) Quadratic configuration interaction. A general technique for determining electron correlation energies. *J Chem Phys* 87: 5968.
- Pople JA, Head-Gordon M, Raghavachari K (1987) Quadratic configuration interaction. A general technique for determining electron correlation energies *J Chem Phys* 87(10): 5968.
- Gauss J, Cremer C (1988) Analytical evaluation of energy gradients in quadratic configuration interaction theory. *Chemical Physics Letters* 150(3-4): 280-286.
- Salter EA, Trucks GW, Bartlett RJ (1989) Analytic energy derivatives in many-body methods. I. First derivatives. *J Chem Phys* 90(3): 1752.
- Hohenberg P, Kohn W (1964) Inhomogeneous Electron Gas. *Phys Rev* 136: B864.
- Koch W, Holthausen MC (2001) *A Chemists's Guide to Density Functional Theory*. (2<sup>nd</sup> edn), Wiley-VCH, Germany, pp. 313.
- Thomas LH (1927) The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society* 23(5): 542-548.
- Slater JC (1951) A Simplification of the Hartree-Fock Method. *Phys Rev* 81: 385.
- Kohn W, Sham LJ (1965) Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev* 140: A1133.
- Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38(6): 3098-3100.
- Dobson JF, Vignale G, Das MP (1998) *Electronic Density Functional Theory: Recent Progress and New Directions*. Springer.
- Perdew JP (1991) In *Electronic Structure of Solids '91*. In: Ziesche P and Eschrig H (Eds.), Akademie Verlag, Berlin, Germany, p11.
- John PP, Chevary JA, Vosko SH, Kobljar AJ, Mark RP (1992) Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys Rev B* 46: 6671.
- John PP, Chevary JA, Vosko SH, Kobljar AJ, Mark RP (1993) Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys Rev B* 48: 4978.
- Chengteh L, Weitao Y, Robert GP (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37: 785.
- Colle, Salvetti Theor (1975) A correlation-energy formula due to Colle and Salvetti. *Chim Acta* 37: 329.
- Axel DB (1993) Density-functional thermo chemistry. III. The role of exact exchange *J Chem Phys* 98(7): 5648.
- Perdew P, Burke K, Ernzerhof M (1996) Generalized Gradient Approximation Made Simple. *Physical Review Letters* 77: 3865.
- Schwartz C (1962) Importance of Angular Correlations between Atomic Electrons. *Phys Rev* 126: 1015.
- Nyden MR, Petersson GA (1981) Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. *J Chem Phys* 75(4): 1843.
- Petersson GA, Nyden MR (1981) Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. *J Chem Phys* 75(4): 3423.
- Ochterski JW, Petersson GA, Montgomery JA (1996) A complete basis set model chemistry. V. Extensions to six or more heavy atoms *J Chem Phys* 104(7): 2598.
- Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. *J Chem Phys* 112(15): 6532.
- Petersson GA, Tensfeldt TG, Montgomery JA (1991) A complete basis set model chemistry. III. The complete basis

- set-quadratic configuration interaction family of methods. *J Chem Phys* 94(9): 6091.
37. Ochterski JW, Petersson GA, Montgomery JA (1996) A complete basis set model chemistry. V. Extensions to six or more heavy atoms *J Chem Phys* 104: 2598.
  38. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* 110(6): 2822.
  39. Maseras F Morokuma K (1995) IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J Comp Chem* 16(9): 1170-1179.
  40. Humbel S, Sieber S, Morokuma K (1996) The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and S N 2 reaction: RCl+Cl-. *J Chem Phys* 105(5): 1959.
  41. Matsubara T, Sieber S, Morokuma K (1996) A test of the new "integrated MO + MM" (IMOMM) method for the conformational energy of ethane and n-butane. *Int J Quant Chem* 60(6): 1101-1109.
  42. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, et al. (1996) ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and Pt(P(t-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> Oxidative Addition. *J Phys Chem* 100(50): 19357-19363.
  43. Svensson M, Humbel S, Morokuma K (1996) Energetics using the single point IMOMO (integrated molecular orbital+molecular orbital) calculations: Choices of computational levels and model system. *J Chem Phys* 105(9): 3654.
  44. Stefan D, István K, Suzie BK, Keiji M, Michael J F (1999) A new ONIOM implementation in Gaussian 98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J Mol Struct (Theochem)* 461-462: 1-21.
  45. Vreven T, Morokuma K, Farkas O, Schlegel HB, Frisch MJ (2000) *J Comp Chem* 24(6): 1419.
  46. Wendy DC, Piotr C, Christopher IB, Ian RG, Kenneth MM, et al. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 117: 5179.
  47. Rappé AK, Casewit CJ, Colwell KS, Goddard III WA, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations *J Am Chem Soc* 114(25): 10024.
  48. Møller C, Plesset MS (1934) Note on an Approximation Treatment for Many-Electron Systems. *Phys Rev* 46(7): 618.
  49. Roothaan CCJ (1951) Perspective on "New developments in molecular orbital theory". *Rev Mod Phys* 23: 69-89.
  50. Pople J A, Nesbet R K (1954) Self-Consistent Orbitals for Radicals. *J Chem Phys* 22: 571.
  51. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *J Chem Phys* 94: 7221-7230.
  52. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) A Complete Basis Set Model Chemistry. VI. Use of Density Functional Geometries and Frequencies. *J Chem Phys* 110: 2822-2827.
  53. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. *J Chem Phys* 112(15): 6532-6542.
  54. Fischer C (1977) The Hartree-Fock method for atoms. A numerical approach. John Wiley and Sons, Inc, New York, USA.
  55. Bauchlicher CW, Melius CF, Allendorf MD (1999) Gallium compounds, a possible problem for the G2 approaches. *J Chem Phys* 110(4): 1879-1881.
  56. Duke BJ, Radom L (1998) Gaussian-2 (G2) theory for third-row elements: A systematic study of the effect of the 3d orbitals. *J Chem Phys* 109: 3352.
  57. Yamaguchi, Y. Schaefer H.F. (1996) The GeOH - HGeO System. Are the 3d Electrons Core or Valence? *J Chem Phys* 104.
  58. Richards CA, Yamaguchi Y, Kim SJ, Schaefer HF (1996) The GaOH - HGaO Potential Energy Hypersurface and the Necessity of Correlating the 3d Electrons. *J Chem Phys* 104(24): 9841.
  59. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. (2001) Gaussian, Inc. Gaussian 01, development version (revision A.01) Gaussian, Pittsburg, USA.
  60. Hohenberg P, Kohn W (1964) Inhomogeneous Electron Gas. *Phys Rev* 136: B864.
  61. John Clarke Slater (1974) Quantum Theory of Molecules and Solids. (Vol 4), McGraw-Hill, New York, USA.
  62. S H Vosko, L Wilk, M Nusair (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 58(8): 1200-1211.
  63. Perdew JP (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B Condens Matter* 33(12): 8822-8824.
  64. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38(6): 3098-3100.
  65. Muneaki Kamiya, Takao Tsuneda, Kimihiko Hirao (2002) A density functional study of van der Waals interactions. *J Chem Phys* 117: 6010.
  66. Iikura H, Tsuneda T, Yanai T, Hirao K (2001) A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J Chem Phys* 115: 3540-3544.
  67. Savin A (1996) Recent Developments and Applications of Modern DFT. Texas, USA.
  68. Leininger T, Stoll H, Werner HJ, Savin A (1997) Combining long-range configuration interaction with short-range density functionals. *Chem Phys Lett* 275(3-4): 151-160.
  69. Andersson Y, Langreth DC, Lundquist BI (1995) Van der

- Waals Interactions in Density-Functional Theory. *Phys Rev Lett* 72: 102.
70. Johnson ER, Becke AD (2005) A post-Hartree-Fock model of intermolecular interactions. *J Chem Phys* 123: 024101.
  71. Ahlrichs R, Penco R, Scoles G (1977) Intermolecular forces in simple systems. *Chem Phys* 19(2): 119-130.
  72. Elstner M, Hobza P, Suhai S, Kaxiras E (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J Chem Phys* 114: 5149.
  73. Xueyuan Wu, MC Vargas, S Nayak, V Lotrich, G Scoles (2001) Towards extending the applicability of density functional theory to weakly bound systems. *J Chem Phys* 115(19): 8748-8757.
  74. Qin Wu, Yang W (2002) Empirical correction to density functional theory for van der Waals interactions. *J Chem Phys* 116: 515.
  75. Zimmerli U, Parrinello M, Koumoutsakos P (2004) Dispersion corrections to density functionals for water aromatic interactions. *J Chem Phys* 120: 2693.
  76. Stephens PJ, Devlin FJ, Chablowski CF, Frisch M (1994) Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J Phys Chem* 98(45): 11623.
  77. Xin Xu, Goddard W III (2003) The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *PNAS* 101(9): 2673-2677.
  78. Xu X, Zhang Q, Moller R, Goddard W III (2005) An extended hybrid density functional (X3LYP) with improved descriptions of nonbond interactions and thermodynamic properties of molecular systems. *J Chem Phys* 122: 014105.
  79. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37(2): 785-789.
  80. Cybulski SM, Toczyłowski RR (1999) Ground state potential energy curves for a coupled-cluster study. *J Chem Phys* 111: 10520.
  81. Hart JR, Rappe AK (1992) van der Waals functional forms for molecular simulations. *J Chem Phys* 97: 1109.
  82. Allinger NL (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J Am Chem Soc* 99 (25): 8127-8134.
  83. Brooks BR, Broccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4(2): 187-217.
  84. Weiner SJ, Kollman PA, Nguyen Dt, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem*. 7(2): 230-252.
  85. Weiner SJ, Kollman PA, Nguyen Dt, Case DA, Singh UC, et al. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106(3): 765-784.
  86. Brooks FC (1958) *Phys Rev* 110: 661.
  87. Koide A (1976) A new expansion for dispersion forces and its application. *J Phys Chem B* 9(18): 3173.
  88. Krauss M, Neumann DB (1979) Charge overlap effects in dispersion energies. *J Chem Phys* 71: 107.
  89. Musher JI, Amos AT (1967) Theory of Weak Atomic and Molecular Interactions. *Phys Rev* 31: 164.
  90. Tang KT, Toennies JP (1984) The damping function of the van der Waals attraction in the potential between rare gas atoms and metal surfaces. *J Chem Phys* 80: 3726.
  91. Van Mourik T, Dunning TH (1999) A new ab initio potential energy curve for the helium dimer. *J Chem Phys* 111(20): 9248-9258.
  92. London FZ (1930) *Phys Chem* 11: 222.
  93. Slater JC, Kirkwood JG (1931) The Van Der Waals Forces in Gases. *Phys Rev* 37: 682.
  94. Halgren TA (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J Am Chem Soc* 114(20): 7827-7843.
  95. Kang YK, Jhon MS, Theor (1982) *Chem Acc* 61: 41.
  96. Tang KT (1968) Upper Bounds for van der Waals Interactions of Two and Three Atoms. *J Chem Phys* 49: 4727.
  97. Zhao Y, Schultz NE, Truhlar DG (2005) Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J Chem Phys* 123: 131103-1.
  98. Ref MO5-2x.
  99. Xin Xu, William A Goddard III (2004) The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc Natl Acad Sci* 101(9): 2673-2677.
  100. Hamprecht FA, Cohen AJ, Tozer DJ, Handy NC (1998) Development and assessment of new exchange-correlation functionals. *J Chem Phys* 109: 6264.
  101. Boese AD, Handy NC, Sprick M (2000) New generalized gradient approximation functionals. *J Chem Phys* 112: 1670.
  102. Boese AD, Handy NC (2001) A new parametrization of exchange-correlation generalized gradient approximation functionals. *J Chem Phys* 114: 5497.
  103. ref – thcth.
  104. Staroverov VN, Scuseria GE, Tao J, Perdew JP (2003) Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J Chem Phys* 119: 12129.
  105. Perdew JP, Burke K, Ernzerhof M (1996) Generalized Gradient Approximation Made Simple. *Phys Rev Lett* 77: 3865.
  106. Rabuck AD, Scuseria GE (1999) Assessment of recently developed density functionals for the calculation of enthalpies of formation in challenging cases. *Chem Phys Letters* 309(5-6): 450-456.

107. Eyring H (1935) The Activated Complex in Chemical Reactions. *J Chem Phys* 3: 107-115.
108. Malick DK, Petersson GA, Montgomery JA (1998) *J Chem Phys* 108: 5704.
109. Pople JA, Head-Gordon M, Raghavachari K (1987) Quadratic configuration interaction. A general technique for determining electron correlation energies. *J Chem Phys* 87: 5968.
110. Roothan CCJ (1951) New Developments in Molecular Orbital Theory. *Rev Mod Phys* 23: 69-89.
111. Pople JA, Nesbet RK (1954) Self-Consistent Orbitals for Radicals. *J Chem Phys* 22(3): 571.
112. Møller C, Plesset MS (1934) Note on an Approximation Treatment for Many-Electron Systems. *APS Journals Archive* 46: 618.
113. Pople JA, Binkley JS, Seeger R (1976) Theoretical models incorporating electron correlation. *Int J Quant Chem Symp* 10(S10): 1-19.
114. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98: 5648.
115. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37: 785.
116. Pople JA, Head-Gordon M, Raghavachari K (1987) Quadratic configuration interaction. A general technique for determining electron correlation energies. *J Chem Phys* 87: 5968.
117. Salter EA, Trucks GW, Bartlett RJ (1989) Analytic energy derivatives in many-body methods. I. First derivatives. *J Chem Phys* 90: 1752.
118. Binkley JS, Pople JA, Hehre WJ (1980) Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *JACS* 102(3): 939-947.
119. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ (1982) Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *JACS* 104(10): 2797-2803.
120. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J Chem Phys* 72: 650.
121. Frisch MJ, Pople JA, Binkley JS (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J Chem Phys* 80: 3265.
122. Huber K, Herzberg G (1979) *Molecular Spectra and Molecular Structure 4. Constants of Diatomic Molecules*. Van Nostrand.
123. Strey G (1967) *J Mol Spectrosc* 24: 87.
124. Strey G, Mills IM (1973) *Mol Phys* 26: 129.
125. Dennison DM (1940) *Rev Mod Phys* 12: 175.
126. Duncan JL, Mills IM The calculation of force constants and normal coordinates-IV  $\text{XH}_4$  and  $\text{XH}_3$  molecules. *Spectrochim Acta* 20(3): 523-546.
127. Petersson G (2001) Complete Basis Set Models for Chemical Reactivity: from Helium Atom to Enzyme Kinetics. *Springer* 22: 99-130.
128. Ochterski JW, Petersson GA, Montgomery JA (1996) A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J Chem Phys* 104: 2598.
129. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. *J Chem Phys* 112: 6532.
130. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* 110: 2822.
131. (a) Feist F (1902) Feist-Bénary Synthesis. *Chem Ber* 35(2): 1537-1544. (b) Bénary E (1911) Feist-Bénary Synthesis. *Chem Ber* 44: 489.
132. Dunlop AP, Hurd CD (1950) *J Org Chem* 15: 1160.
133. Canton IJ, Cocker W, McMurry TB (1961) *Tetrahedron* 15: 45.
134. Calter MA, Phillips RM, Flaschenriem C (2005) *J Am Chem Soc* 127: 14566.
135. Calter MA, Zhu C, Lachicotte R (2002) *Organic Letters* 4: 209.